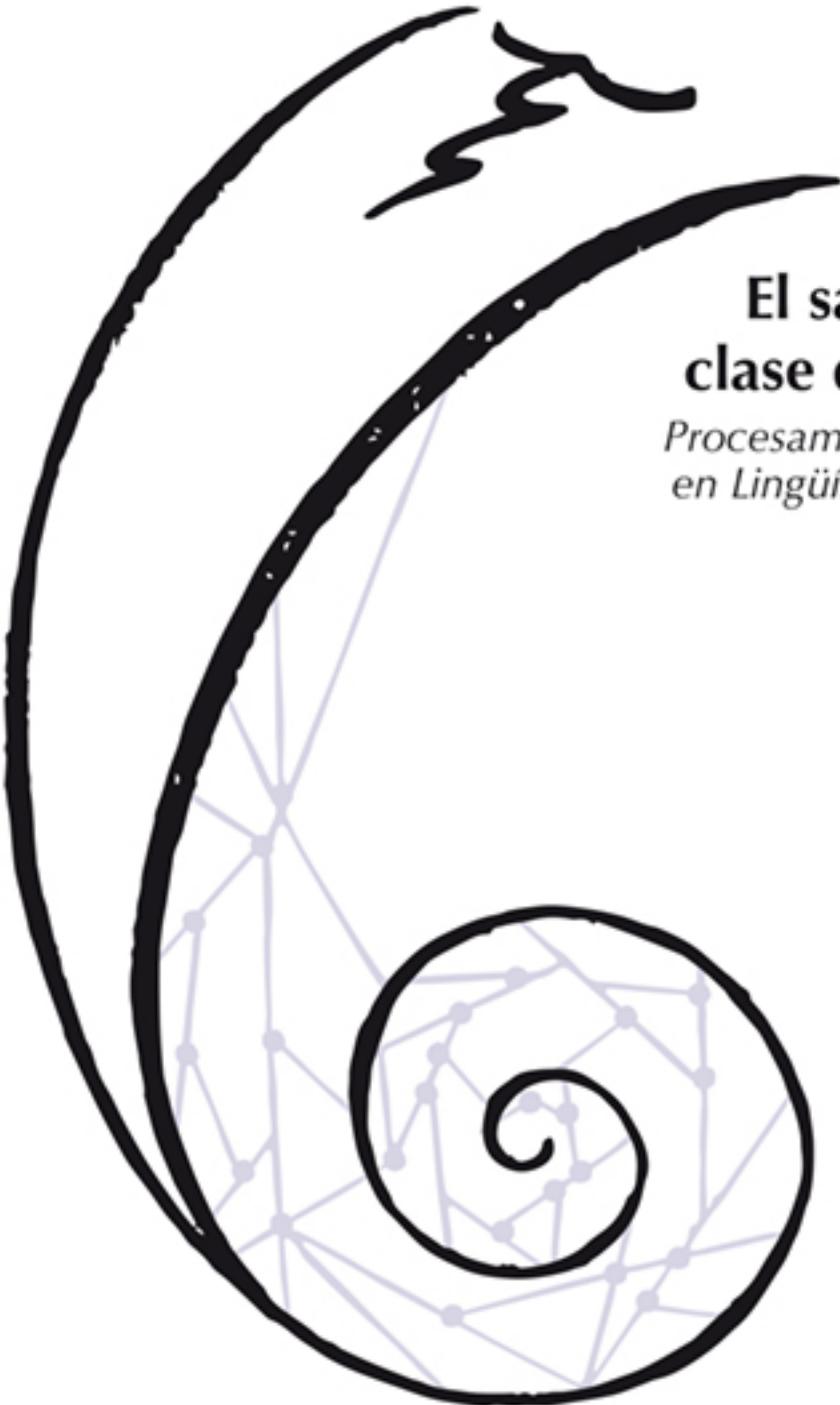




UAEM | Universidad Autónoma
del Estado de México



El salón de clase de lengua

*Procesamiento de datos
en Lingüística Aplicada*

Clara Cecilia
Uribe Hernández



iu
investigación
universitaria





El salón de clase de lengua

*Procesamiento de datos
en Lingüística Aplicada*



UAEM | Universidad Autónoma
del Estado de México

Dr. en D. Jorge Olvera García
Rector

Dra. en Est. Lat. Ángeles Ma. del Rosario Pérez Bernal
Secretaria de Investigación y Estudios Avanzados

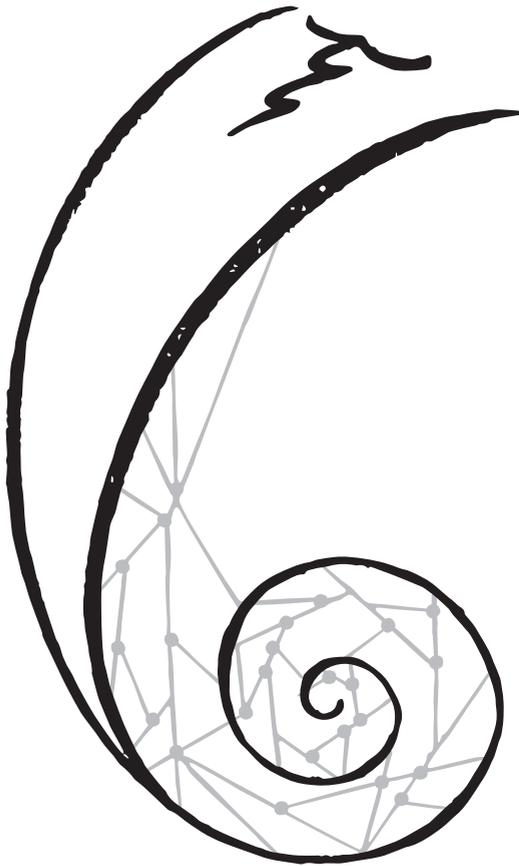
Dr. en Hum. Víctor Alonso Galeana Estrada
Director de la Facultad de Lenguas

L. C. C. María del Socorro Castañeda Díaz
*Directora de Difusión y Promoción de la Investigación
y los Estudios Avanzados*

El salón de clase de lengua

*Procesamiento de datos
en Lingüística Aplicada*

Clara Cecilia Uribe Hernández



El salón de clase de lengua.
Procesamiento de datos en Lingüística Aplicada

Libro de investigación e interés académico y sin fines de lucro.

1a edición, enero 2017

ISBN: 978-607-422-786-4

D.R. © Universidad Autónoma del Estado de México
Instituto Literario núm. 100 Ote., Centro, C.P. 50000,
Toluca, México
<http://www.uaemex.mx>

Impreso y hecho en México
Printed and made in Mexico

El contenido de esta publicación es responsabilidad de la autora.

En cumplimiento a la normatividad sobre el acceso abierto de la investigación científica, esta obra se pone a disposición del público en su versión electrónica en el repositorio de la UAEM (<http://ri.uaemex.mx>) para su uso en línea con fines académicos y no de lucro, por lo que se prohíbe la reproducción parcial o total, directa o indirecta del contenido de esta presentación impresa sin contar previamente con la autorización expresa y por escrito de los editores, en términos de lo así previsto por la *Ley Federal del Derecho de Autor* y, en su caso, por los tratados internacionales aplicables.

Índice

Introducción	11
1. Investigación en lingüística aplicada; los salones de clase de lengua	15
1.1 Las ciencias del lenguaje	18
1.2 Áreas de estudio de la lingüística aplicada.....	22
1.3 Relación de la lingüística aplicada con otras ciencias.....	30
1.4 Salones de clase de lengua.....	32
1.4.1 Personajes en el salón de clase de lengua.....	35
1.4.2 Procesos en el salón de clase de lengua	36
1.5 Áreas de estudio en un salón de clase de lengua ...	39
1.6 La psicolingüística	41
1.7 Investigación en las áreas de estudio en un salón de clase de lengua.....	42
2. Procesamiento de datos en lingüística aplicada.....	49
2.1 Documentación de un proyecto de investigación .	52
2.2 Diseño de un proyecto de investigación	53
2.2.1 Diseños en investigación cuantitativa	55
2.2.2 Diseños en investigación cualitativa.....	59
2.3 Instrumento de levantamiento de datos.....	60
2.4 Tratamiento	62
2.5 Datos levantados	63
2.6 Procesamiento de los datos	63
2.6.1 Datos cuantitativos.....	64
2.6.2 Datos cualitativos	64
2.6.3 Transformación de datos cualitativos en cuantitativos.....	65

3. Elementos metodológicos y estadísticos	67
3.1 Elementos metodológicos	69
3.1.1 Variables y componentes o dimensiones, entidades	70
3.1.2 Tipos de entidades	70
3.1.2.1 Entidades según el nivel de medición....	71
3.1.2.2 Entidades según la posición del investigador y la relación entre ellas.....	72
3.1.2.3 Entidades en lingüística aplicada	74
3.1.2.4 Entidades según los valores que pueden asumir	75
3.1.3 Hipótesis.....	76
3.1.4 Puntajes totales y parciales	78
3.1.5 Confiabilización de los reactivos de un instrumento de levantamiento de datos.....	78
3.1.5.1 Diseño inicial	79
3.1.5.2 Piloteo.....	79
3.1.5.3 Diseño final	80
3.2 Elementos estadísticos.....	80
3.2.1 Tipos de estadística	80
3.2.1.1 Estadísticas descriptiva e inferencial	81
3.2.1.2 Estadísticas no paramétrica y paramétrica	82
3.2.2 Medidas descriptivas	82
3.2.2.1 Medidas de posición	82
3.2.2.2 Medidas de tendencia central.....	82
3.2.2.3 Medidas de dispersión	84
3.2.2.4 Medidas de forma	84
3.2.3 Tipos de muestras	87
3.2.4 Relaciones causa-efecto entre las variables.....	88
3.2.4.1 Relación entre una variable independiente y una variable dependiente.....	88
3.2.4.2 Relaciones entre varias variables independientes y una o más dependientes	89

4. Auxiliares estadísticos	91
4.1 Parámetros básicos	93
4.1.1 Nivel de significancia	94
4.1.2 Intervalo de confianza.....	95
4.1.3 Grados de libertad	95
4.1.4 Prueba de normalidad de la curva.....	96
4.1.5 Prueba de Levene.....	98
4.1.6 Tabla de valores de prueba	100
4.2 Coeficientes	100
4.2.1 Alfa de Cronbach.....	100
4.2.2 Correlación	101
4.2.3 Determinación	104
4.2.4 Regresión lineal	104
4.2.5 Regresión múltiple	108
4.3 Toma de decisiones	110
4.3.1 Elección de la prueba estadística	111
4.3.1.1 Criterios de elección de la prueba estadística	111
4.3.1.2 Curva de una o de dos colas	112
4.3.2 Análisis de acuerdo con el nivel de medición de las variables	113
4.4 Confiabilización de los datos levantados y capturados	114
4.5 Gráficas	114
5. Pruebas estadísticas	117
5.1 Prueba t de Student	122
5.2 ANOVA	128
5.3 La prueba Chi-cuadrada.....	133
5.4 U. de Mann Whitney	135
5.5 Kruskal-Wallis	136
5.6 McNemar	138
5.7 Wilcoxon	139
5.8 Q de Cochran	141
5.9 Friedman	142

6. Uso de escalas	145
6.1 Escalas	147
6.2 Escala ordinal	151
6.3 Escalas de Likert	151
7. Grados de dificultad y discriminación de reactivos de exámenes de opción múltiple.....	157
7.1 Preparación de los datos	160
7.2 Grado de dificultad de los reactivos	162
7.3 Grado de discriminación de los reactivos	162
7.4 Evaluación de los distractores	163
7.5 Ejemplo	163
8. El procesamiento de datos mediante el <i>Statistical Package for Social Sciences</i> (SPSS)	171
8.1 Paquetes estadísticos	173
8.2 Definición de códigos	176
8.3 Creación de la plantilla de datos para el SPSS....	178
8.4 Carga de los datos en el archivo electrónico....	182
8.5 Procesamiento de los datos en el archivo electrónico	182
8.6 Pruebas en general	185
8.7 Procesamiento de datos de varios grupos.....	186
8.8 Calcular el valor de una variable	186
8.9 Transformar variable por rangos de sus valores..	188
8.10 Gráficas	195
Bibliografía	199
Anexos	211

Introducción

Este libro es el resultado del contacto cotidiano con necesidades en materia de investigación de estudiantes de facultades de lenguas, tanto del nivel de licenciatura como de maestría. Ha sido escrito con la finalidad de que se cuente con una obra que conjunte aspectos someros de carácter teórico con elementos más precisos y detallados acerca del procesamiento de datos en el área de la lingüística aplicada; particularmente en salones de clase de lengua.

Son muchos los procesos que viven los estudiantes para lograr conducir estos trabajos a un feliz término, desde la sorpresa y el temor que provoca en los estudiantes la exigencia de investigar y elegir un tema de investigación, la concepción misma de su proyecto, su composición, sus paradigmas, hasta su desarrollo completo. Un gran reto para ellos es la elección del punto de partida teórico y la forma de manejar los datos de una investigación, principalmente el levantamiento de los mismos y los diversos procesos a los que tienen que someterlos.

Esta obra pretende esencialmente dar respuesta a la falta de lineamientos que den soporte al levantamiento de datos en conjunción con el manejo de herramientas estadísticas y de procesamiento de datos que sustenten el análisis para la derivación de resultados.

Las propuestas que se dirigen a los estudiantes y a los docentes no proceden precisamente de una investigación particular. Por el contrario, surgen de las experiencias metodológicas vividas a través del desarrollo de investigaciones llevadas a cabo en salones de clase de inglés y francés como lengua extranjera.

Al hablar de procesamiento de datos, estadística y paquetes de computadora, el libro parece dirigirse más hacia las investigaciones de corte cuantitativo que a las cualitativas. Sin embargo, los estudios en el salón de clase de lengua se diseñan con base en los dos paradigmas, cuantitativo y cualitativo; incluso bajo consideraciones mixtas. La realidad

es que, en nuestros días, también existe el procesamiento de datos de investigaciones cualitativas.

La estadística, como las matemáticas, siempre ha sido un área temida. Sin embargo, basta con conocer un poco de aritmética para comprender sus métodos y la utilidad de los productos numéricos que se obtienen al aplicarla. Una buena parte de esta obra contiene éstos métodos, desmenuzados, con la finalidad de que sea más clara su elección y uso en los diversos tipos de investigaciones que pueden realizarse en un salón de clase de lengua. No obstante, para obtener un mayor provecho del libro, es recomendable poseer algunas nociones básicas de estadística.

Contiene ocho capítulos. El primero enmarca el contexto y las áreas de estudio de las investigaciones en un salón de clase de lengua; abarca siete temas.

El primer tema, acerca de las ciencias del lenguaje, como una base para contextualizar la lingüística aplicada, introduce una reflexión sobre su concepción en general.

El segundo tema aborda las áreas de estudio de la lingüística aplicada. El tercero, su relación con otras ciencias. El cuarto trata los salones de clase de lengua, destacando dos elementos, los personajes que interactúan en ellos y los procesos que se realizan en su interior.

El quinto tema complementa las áreas de estudio en los salones de clase de lengua. El tema seis versa, específicamente, acerca de la importancia de la psicolingüística en la investigación en estas áreas. El séptimo establece que en la actualidad los datos cualitativos, para su procesamiento, pueden convertirse en cuantitativos. Da algunos lineamientos teóricos para el manejo de los objetos de estudio abordados desde el punto de vista de sus componentes.

El segundo capítulo incorpora consideraciones básicas para levantar datos y su procesamiento: documentación y diseño de un proyecto, instrumentos para el levantamiento de datos, tratamientos efectuados con sujetos de investigación y procesamiento de datos bajo los paradigmas tanto cuantitativo como cualitativo.

Los capítulos tres, cuatro y cinco constituyen un ensamble que engloba los temas sobre procesamiento de datos. El capítulo tres proporciona información relativa a la conveniencia de manejar sus características generales y componentes para enriquecer el aprovechamiento de los resultados, presenta elementos básicos para la estadística descriptiva.

El capítulo cuatro, denominado Auxiliares estadísticos, presenta otros mecanismos estadísticos diferentes a los mostrados en el capítulo tres, éstos son utilizados para procesos que no son tan elementales como el cálculo de la media. Los hay considerados como básicos, coeficientes y algunos para la toma de decisiones en cuanto al uso de pruebas esta-

dísticas. El último inciso presenta información relativa a la elaboración de gráficas.

Es importante resaltar que los capítulos tres y cuatro contienen el tema de la confiabilización aplicada a dos análisis diferentes. En el tres se habla de la confiabilización de los reactivos incluidos en los instrumentos de levantamiento de datos, en cuanto a la certeza de que siempre miden con los mismos resultados; su construcción y contenido; mientras que el cuatro contempla el nivel de confiabilidad de los datos levantados.

El capítulo cinco indica las pruebas estadísticas que se aplican a los datos levantados para obtener resultados, denominadas como pruebas de hipótesis, básicamente de corte cuantitativo. Presenta un cuadro con las pruebas paramétricas y no paramétricas y un cuadro en el que se relacionan las pruebas con consideraciones tanto de la hipótesis nula como de las pruebas estadísticas. Precisa solamente aquellas pruebas que se utilizan con mayor frecuencia, con indicaciones relativas a su uso, la forma adoptada por la hipótesis nula, las instrucciones en el programa SPSS (*Statistical Package for Social Sciences*) y los resultados que se obtienen al procesar los datos con éste.

Para los incisos en los que se indica la secuencia de comandos teclados para el procesamiento de los datos, se utiliza el programa estadístico SPSS, versión 18, se eligió por ser el más completo. Está dirigido al procesamiento de datos en las ciencias sociales, tiene una vida actual de 47 años; su estructura es de fácil acceso, facilita el aprendizaje del funcionamiento de otros paquetes. Los ejemplos y comandos incluidos proporcionan la oportunidad de conocer un paquete estadístico específico, conocimiento que puede traspolarse a otros paquetes. Si se estima conveniente, puede revisarse el capítulo ocho, sobre el uso del SPSS, para familiarizarse con él antes de leer el capítulo tres y los subsecuentes.

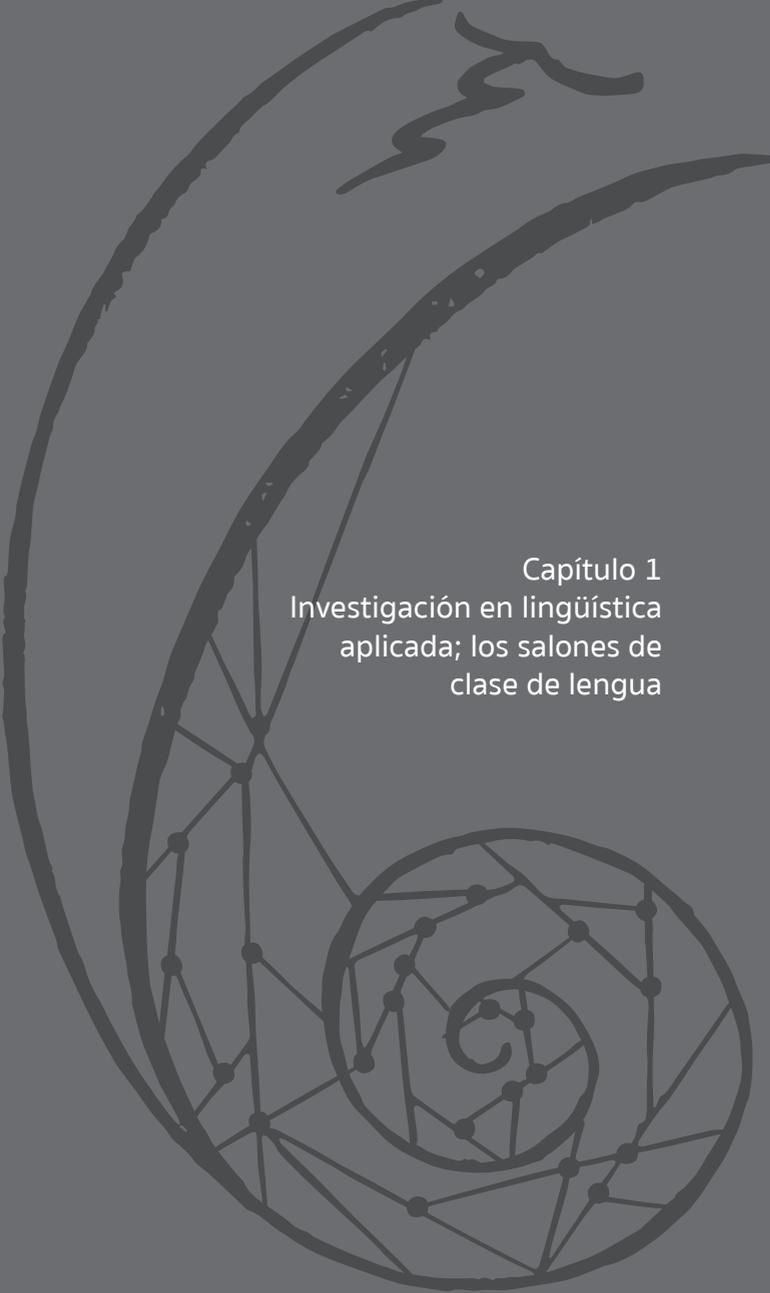
El capítulo seis expone el uso de las escalas como instrumentos de levantamiento de datos. Se presenta una distinción entre escalas en general, escalas ordinales y, detalladamente, escalas Likert. Se enfatiza en que el uso de escalas del tipo Likert conlleva una serie de consideraciones específicas que no pueden dejarse de lado.

El capítulo siete se refiere a la confiabilización de las pruebas cuyos reactivos se construyen con respuestas de opción múltiple. Se muestra el procedimiento conocido como la Prueba de Madsen, con la que se miden los grados de dificultad y discriminación de cada uno de los reactivos de la prueba en proceso.

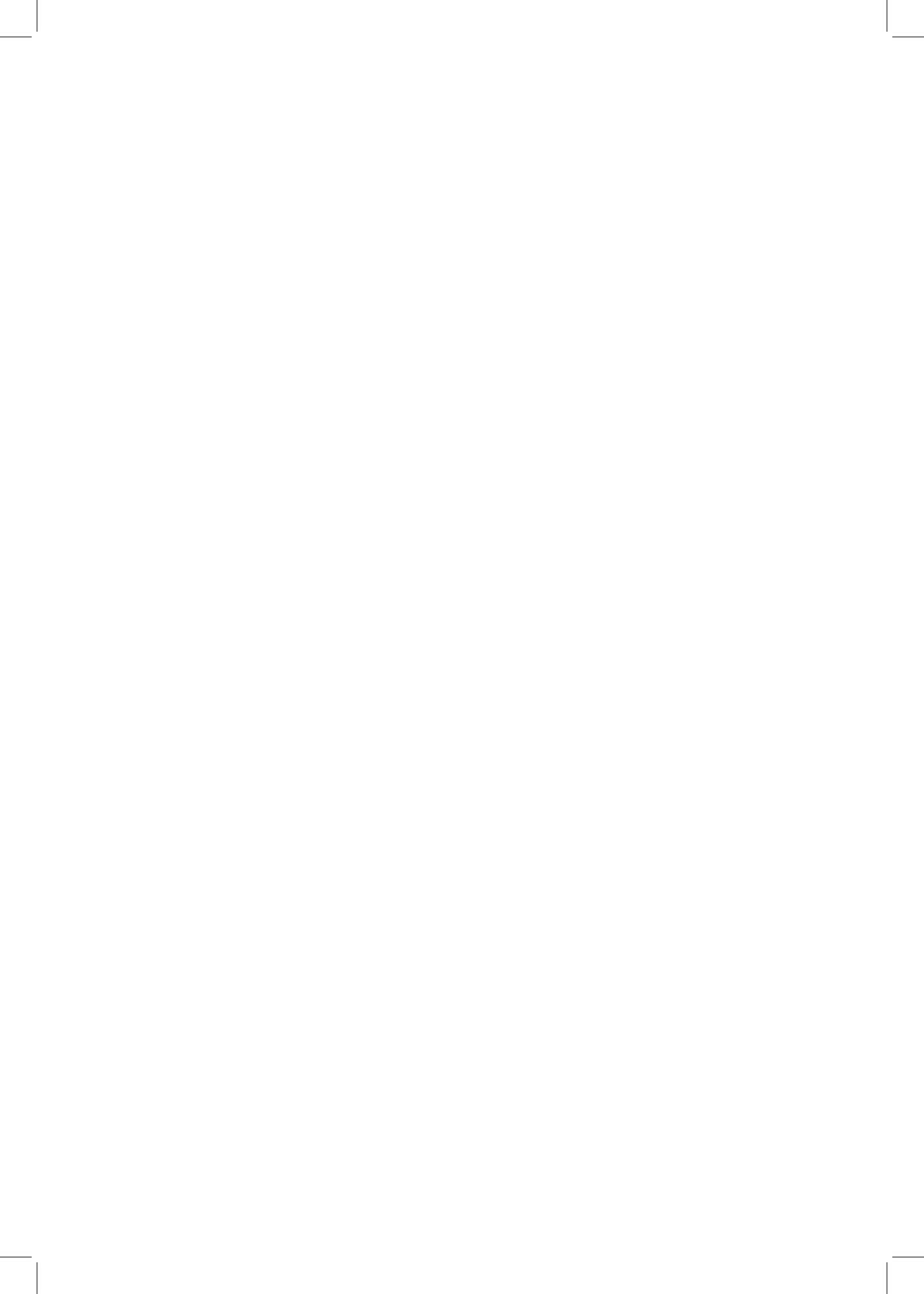
El capítulo ocho constituye una guía general para el manejo del SPSS como una de las posibles opciones para el proceso de datos en forma electrónica, a través de los así denominados paquetes estadísticos. Se introducen sucintamente las opciones que hay en este momento

en el mercado, así como comentarios relacionados con el uso de este tipo de paquetes. El segundo y tercer incisos exponen el trabajo previo de diseño que hay que llevar a cabo antes de la captura de los datos en cualquier paquete estadístico, la forma de asignarles códigos para que se manejen en forma numérica y las plantillas que van a crearse. Se indica las plantillas básicas del SPSS, una para variables y otra para datos. Los últimos cuatro incisos indican el procesamiento general de datos en el SPSS y casos específicos como el procesamiento de datos de varios grupos; el cálculo del valor de una variable, como el puntaje total dado por la suma de los puntajes de varios reactivos; la transformación del valor de una variable en rangos; y la elaboración de gráficas.

En la parte final, se presentan las referencias y los anexos. Una particularidad de los anexos es que contienen herramientas complementarias para el análisis de los datos procesados y tablas utilizadas en la búsqueda de valores de parámetros con la finalidad de compararlos con valores calculados. El último anexo contiene fórmulas estadísticas, de tal forma que el proceso de datos pueda llevarse a cabo a pesar de que no se cuente con un paquete estadístico.



Capítulo 1
Investigación en lingüística
aplicada; los salones de
clase de lengua



La investigación en lingüística aplicada relativa a los salones de clase de lengua enfrenta un horizonte amplio en cuanto a las posibilidades para su desarrollo.

Alan Davies¹ (2007) en su libro *Introducción a la lingüística aplicada* (LA) señala que ésta surgió en los años cincuenta como estudios de posgrado, con el objetivo de analizar la enseñanza de lenguas, con un carácter aplicado y para la creación de políticas.

Al hablar de la vinculación entre la lingüística aplicada y la enseñanza de lengua en un salón de clase, surge la inquietud de saber si esta ciencia antecedió la enseñanza de lenguas. Sin embargo, es esta última la que despertó la necesidad de un análisis profundo de la forma en la que se llevaba a cabo y la que dio origen a la lingüística aplicada.

En la actualidad, la lingüística aplicada plantea dos campos de estudio primordiales, el proceso de enseñanza/aprendizaje de una lengua y la lengua aunada a su uso. El desarrollo mismo de estos dos procesos conduce a la apertura de su campo de investigación, al relacionarse con otras ciencias o disciplinas que se conjugan para resolver problemas en materia de lengua.

Schmitt y Celce-Murcia afirman que “la lingüística aplicada utiliza lo que se conoce de la lengua, de su aprendizaje, de cómo se usa, para lograr un propósito o resolver un problema del mundo real” (2000, p.1). Consideran que los propósitos de la lingüística aplicada son muchos y variados; para ilustrarlo presentan la siguiente definición:

¹ La traducción de las citas y resúmenes fue hecha por la autora del trabajo.

En un sentido amplio, corresponde a la lingüística aplicada incrementar la comprensión del papel de la lengua en los asuntos humanos y, por lo tanto, proveer el conocimiento necesario para aquellos que son los responsables de tomar decisiones relativas a la lengua, ya sea que las necesidades surjan en el salón de clase, en el lugar de trabajo, en el tribunal o en el laboratorio (Wilkins, 1999, p. 7 citado por Schmitt y Celce-Murcia, 2000, p.1).

En este capítulo se aborda la noción de ciencias del lenguaje como un contexto de partida para la presentación de un panorama de las áreas de estudio de la lingüística aplicada y la mención de otras ciencias con las que se relaciona. Se consideran elementos presentes en un salón de clase de lengua, los personajes, los procesos y las áreas que son materia de estudio en la enseñanza de lengua. Se enfatiza, en un inciso particularizado, la relevancia de los estudios de psicolingüística en estas áreas y se dan algunos lineamientos para la consideración de los datos cuantitativos y cualitativos así como las formas de abordar un objeto de estudio. Al presentar una gama amplia de posibilidades de investigación, el contenido de este capítulo soporta la elección del área de investigación en un salón de clase de lengua y las bases para descomponer el objeto de estudio en entidades medibles.

1.1. Las ciencias del lenguaje

Es de particular interés conocer el área del conocimiento nominada, ciencias del lenguaje, como una base para contextualizar la lingüística aplicada. Es necesario puntualizar que al hablar de ciencias del lenguaje en realidad se habla de lingüística;

bajo el rubro de ciencias del lenguaje no se esconde, necesariamente, otra manera de ver la lingüística, sino se trata de la traducción de la palabra 'lingüística', en alemán *Sprachwissenschaft*, que los autores franceses del Diccionario Enciclopédico de las Ciencias del Lenguaje hicieron suya con su interpretación muy personal de lo que entienden por ello (M. Akerberg, comunicación personal, 13 de noviembre de 2015).²

Si bien, se indicó que el origen de la lingüística aplicada obedeció a la necesidad de estudiar la enseñanza de lenguas, es relevante conocer lo que ha acontecido desde el punto de vista de los planteamientos teóricos acerca del estudio de las lenguas; conocer que, a pesar de la existencia de estudios o bien asignaturas con la denominación de ciencias

² Marianne Akerberg Afzelius es doctora en Lingüística, maestra e investigadora en los programas de Maestría en Lingüística Aplicada y Doctorado en Lingüística en la UNAM. Actualmente está jubilada.

del lenguaje, no hay un acuerdo general aceptado en relación con sus componentes. Su definición es arbitraria.

Estamos habituados a hablar de “ciencias del lenguaje”. Sin embargo, es un campo, un conjunto de ciencias, que no ha sido precisado.

A continuación se analizan dos puntos de vista; el de los autores del diccionario enciclopédico antes mencionado, Ducrot y Todorov (2006) y el de Álvarez (2005).

El *Diccionario enciclopédico de las ciencias del lenguaje* de Ducrot y Todorov (2006, 1974), contiene los siguientes rubros: las escuelas, los dominios, los conceptos metodológicos, los conceptos descriptivos y un apéndice. El contenido del primer capítulo es un resumen de la historia de la lingüística. El segundo capítulo se acerca un poco más a su concepción de las ciencias del lenguaje, bajo la denominación de dominios. Tanto el tercero como el apéndice contienen elementos del lenguaje.

Ducrot y Todorov (2006) introducen el segundo capítulo, el denominado Dominios, con el tema *Componentes de la descripción Lingüística*. Los definen como ámbitos. Destacan tres, establecidos en un principio como la tradición occidental, para llevar a cabo esta tarea y las críticas al respecto, desde la lingüística del siglo XX. El siguiente cuadro los muestra.

Cuadro 1.1. Ámbitos para la descripción de una lengua: tradición occidental

Los medios materiales de expresión	<ul style="list-style-type: none"> • Pronunciación • Escritura 	
La gramática, compuesta por:	<ul style="list-style-type: none"> • La morfología, las palabras • La sintaxis, combinación de palabras en la frase 	<ul style="list-style-type: none"> • Partes del discurso (nombre, verbo) • Variaciones de la palabra (género, número, declinación y conjugación) • Orden de las palabras • Fenómenos de rección, concordancia o régimen
Diccionario o léxico, sentidos que poseen las palabras	<ul style="list-style-type: none"> • La parte semántica por excelencia 	

Fuente: elaborado a partir de Ducrot y Todorov (2006).

Se incluye la morfología y la sintaxis como componentes de la gramática. Palpablemente, en este inciso, consideran las tareas para el estudio del lenguaje mismo. En sus palabras, “desde lo exterior hasta la significación” (Ducrot y Todorov, 2006, p. 67), critican planteamientos, señalando que “la palabra ya no se considera la unidad significativa fundamental”. Su obra contiene un análisis de la consideración de la evolución de la lengua y su descripción, de acuerdo con las escuelas de lingüística a través del tiempo. Concluyen que las áreas relevantes en el estudio y descripción de la lengua son: la gramática, la sintaxis, el componente semántico y el componente fonológico o morfofonológico. También señalan como resumen de los niveles de análisis comunes con los elementos suprasegmentales y segmentales los siguientes:

Cuadro 1.2. Niveles de análisis comunes con los elementos suprasegmentales y segmentales

Niveles	Segmental	Suprasegmental
1 ^{er} Nivel: fonemas (oposición)	Fonología	Tonología
2 ^o Nivel: morfemas (contraste)	Morfología	Acentología
3 ^{er} Nivel: proposiciones (oposición)	Sintaxis	Análisis de la entonación
4 ^o Nivel: ? (contraste)	Estilística	Análisis de las connotaciones expresivas

Fuente: Ducrot y Todorov (2006, p. 216).

Por otra parte, Álvarez (2005) en su artículo sobre las ciencias del lenguaje y la comunicación, considera que son la psicolingüística, la semiótica, la retórica, la pragmática, la enunciación, la sociolingüística, la etnografía de la comunicación, el análisis del discurso y la lingüística textual. En el anexo A se presentan estas ciencias y sus componentes, de acuerdo con la propia concepción de Álvarez (2005), divididas en cuadros para cada una, basados en el análisis componencial.

El siguiente cuadro muestra la comparación de la designación como ciencias del lenguaje de Ducrot y Todorov (2006) y de Álvarez (2005).

Cuadro 1.3. Comparativo de la consideración de las Ciencias del lenguaje

Ciencia	Álvarez	Ducrot y Todorov
Análisis del discurso	X	
Enunciación	X	
Estilística		X
Etnografía de la comunicación	X	
Filosofía del lenguaje		X
Geolingüística		X
Lingüística textual	X	
Poética		X
Pragmática	X	
Psicolingüística	X	
Retórica	X	X
Semiótica	X	X
Sociolingüística	X	X

Fuente: Álvarez (2005), Ducrot y Todorov (2006).

Como puede apreciarse en el cuadro 1.3, las coincidencias entre los autores son sólo en cuanto a la Retórica, la Semiótica y la Sociolingüística.

Finalmente, la clasificación de Álvarez (2005) es una clasificación más completa de las ciencias del lenguaje, aunque no coincide con otros autores.

Si se consideran las disciplinas y ciencias que tienen como objeto de estudio la lengua, se tienen:

Cuadro 1.4. Disciplinas y ciencias cuyo objeto de estudio es la lengua

Análisis del discurso
Escritura
Fonética y fonología

Gramática
Lingüística, lenguaje y lengua
Lexicología
Lexicografía
Literacidad (traducida anteriormente como alfabetización)
Literatura
Manejo de corpus lingüístico
Morfología
Pragmática
Redacción
Retórica
Semántica
Sintaxis
Terminología
Vocabulario

Por otra parte, a través del análisis de los Programas de doctorado en ciencias del lenguaje de seis instituciones universitarias, tres mexicanas, dos españolas y una argentina, se corrobora que no hay una definición unánime de la denominación ciencias del lenguaje. Cada institución establece sus propias concepciones (ver Anexo B).

1.2. Áreas de estudio de la lingüística aplicada

Hemos visto que las investigaciones de lingüística aplicada se dirigen primordialmente a dos áreas: la enseñanza de lenguas y el uso de las mismas. En este inciso se resaltan áreas de estudio sobre la lengua, su enseñanza y aprendizaje, así como algunas disciplinas alrededor de ella; áreas de estudio en las que la lengua es la materia prima, como el caso del análisis del discurso en español y en lenguas extranjeras.

Schmitt y Celce-Murcia (2000), en el libro de *Introducción a la lingüística aplicada*, editado para establecer un panorama de la lingüística aplicada, se comenta que “tradicionalmente la primera preocupación de

la lingüística aplicada ha sido la teoría sobre la adquisición de una segunda lengua, la pedagogía de una segunda lengua y la interfaz entre las dos” (Schmitt y Celce-Murcia, 2000, p. 2), por lo que el libro cubre esas áreas. Establecen cinco posibilidades para reportar las áreas en esta disciplina: las grandes áreas; las subdisciplinas; la descripción de la lengua y su uso; las áreas esenciales para la investigación en lingüística aplicada; y las habilidades de lengua y su evaluación.

En el primer grupo están áreas de estudio de elementos relacionados con la lengua, tomados de la convocatoria para el congreso 2002 de la *American Association of Applied Linguistics (AAAL)*, en la que se consideran 18: adquisición, cerebro, cognición, comprensión escrita, comprensión oral, cultura, evaluación, expresión escrita, expresión oral, ideología, instrucción, interacción, medios, metodología de la investigación, políticas, sociedad, traducción/interpretación y tecnología.

Para el segundo grupo, el de las subdisciplinas, toman la opinión de Carter y Nunan (2001, p.2). Consideran nueve áreas: literacidad; patología del discurso; educación de mudos; interpretación y traducción, ya contemplado en el primer caso; prácticas de comunicación; lexicografía y adquisición de la lengua materna. Agregan que:

además de la formación en lengua materna, la planeación de lengua y el bilingüismo/multilingüismo, hay otras dos áreas no consideradas por Carter y Nunan (2001), la identificación de autoría y la lingüística forense. Estas áreas ejemplifican cómo puede utilizarse el conocimiento en lingüística aplicada en forma práctica, en áreas que no pertenecen a la educación (Schmitt y Celce-Murcia, 2000, p. 2).

Consideran que la lingüística aplicada también tiene interés en los casos en los que el lenguaje funciona mal, como los desórdenes en el discurso de los afásicos, los esquizofrénicos y los autistas o aquellos pacientes con hemisferectomía; en la creencia de que se comprende mejor la función del cerebro cuando se daña o funciona mal el sistema del lenguaje del hablante (Schmitt y Celce-Murcia, 2000, p.3).

En la descripción de la lengua y su uso, los temas tratados son: gramática, vocabulario, análisis del discurso, pragmática y lingüística del corpus. En las áreas esenciales para la investigación en LA incluyen la adquisición de una segunda lengua, la psicolingüística, la sociolingüística y el enfoque en el aprendiente de lengua: motivación, estilos y estrategias. Por último, en lo que se refiere a las habilidades de lengua y su evaluación, los capítulos del libro incluyen comprensión oral, expresión oral y pronunciación, comprensión escrita, expresión escrita y evaluación.

Pennycook (2001) contempla la lingüística aplicada como un área de trabajo que tiene que ver con el uso de la lengua en diversos ámbi-

tos, como los profesionales, la traducción, la patología del lenguaje, la literacidad y la educación en el área de lenguas, en una relación que no solamente es la aplicación de conocimiento lingüístico en esos ámbitos sino un dominio de trabajo, semiautónomo, interdisciplinario, que hace uso de otras áreas como la sociología, la educación, la antropología, los estudios culturales y la psicología, pero que no depende de ellas.

Kaplan (2002, p. 4), aporta un estudio de las áreas en los años setenta y ochenta del siglo veinte. En los setenta, el objetivo primordial sigue siendo la enseñanza de lenguas y se incluyen: evaluación de lengua (*language assessment*), adquisición de segunda lengua, literacidad, multilingüismo, derechos de las lenguas minoritarias, planeación y políticas de lengua y capacitación de profesores.

En los años ochenta:

la lingüística aplicada se desarrolló verdaderamente en forma sistemática, más allá de la enseñanza y el aprendizaje de lenguas para abarcar: evaluación de lengua; planeación y políticas de lengua; uso de la lengua en ámbitos profesionales; traducción; lexicografía; multilingüismo; lengua y tecnología; y lingüística de corpus, la cual ha atraído constantemente más atención para la lingüística aplicada que para la lingüística teórica (Kaplan, 2002, p. 4).

Menciona la relación de la lingüística aplicada con otras disciplinas, hacia el final de la década de los ochenta, además de la lingüística: psicología; educación; antropología; sociología; ciencias políticas; estudios de políticas y administración pública; y estudios del inglés: composición, retórica y estudios literarios (Kaplan, 2002, pp. 4 y 5).

Kaplan (2002, p. 10), además de sugerir áreas de investigación de la LA, plantea algunos problemas a los que se enfrenta: en el aprendizaje de la lengua, surgimiento, conciencia, reglas, uso, contexto, automaticidad, actitudes y destreza; en la enseñanza de lenguas: recursos, capacitación, práctica, interacción, comprensión, uso, contextos, desigualdades, motivaciones y resultados; de literacidad: elementos lingüísticos y de aprendizaje; en contacto de lenguas: lengua y cultura; en desigualdad de lenguas: etnicidad, clase, región, género y edad; en planeación y políticas lingüísticas: planeación de estatus y planeación de corpus y ecología de la lengua; en evaluación de lengua: validez, confiabilidad, usabilidad, responsabilidad; en uso de la lengua: dialectos, registros, comunidades de discurso, situación de *gate-keeping*, lo cual significa situaciones en las que hay control y limitación del acceso general a algo y acceso a los servicios limitado; en lengua y tecnología: aprendizaje, evaluación, acceso y uso; en traducción e interpretación: en línea, fuera de línea y tecnología asistida; por último, patologías del discurso: afasias, dislexias e incapacidades físicas.

Wiśniewski (2007) concibe la lingüística aplicada como un paraguas que abarca muy diversas áreas de estudio, todas ellas dirigidas hacia la lengua usada actualmente. Considera que el énfasis está en los usuarios de la lengua y en las maneras en las que la usan; un énfasis opuesto al de la lingüística teórica que estudia la lengua en forma abstracta, sin hacer referencia a contextos específicos. Aun en la lingüística aplicada, señala que hay una diferencia de opinión en cuanto a su alcance, sus dominios y sus límites, pero concede que existen muchos elementos investigados por esta disciplina, como el análisis del discurso, el lenguaje de signos, la estilística y la retórica así como el aprendizaje de lengua por niños y adultos, ya sea como lengua materna o como segunda lengua o extranjera. Resalta los estudios de correlación entre lenguaje y género así como de fenómenos de transferencia de información y la comunicación interpersonal. También, la lingüística forense, la interpretación y la traducción, así como la metodología de la enseñanza de lenguas y los cambios en las lenguas. Para él, en un principio, los estudios de la LA se asociaron principalmente con la enseñanza de la lengua materna, de la segunda y de la extranjera, pero ahora, la LA es considerada como una rama de ciencia más interdisciplinaria.

Wiśniewski (2007), en su *blog*, establece las siguientes áreas como campo de trabajo de la lingüística aplicada: adquisición de L1, adquisición de L2, análisis del discurso, análisis de corpus, características de la lengua, desarrollo de la escritura, esperanto, estilística, etimologías, fonética, fonología, formación de palabras, funciones de la lengua, gramática, gramática generativa, historia de la lengua, inglés antiguo, inglés de la edad media, inglés moderno temprano, inglés moderno tardío, lengua, lengua y cultura, lengua y género, lengua y medio ambiente, lenguaje de signos, lexicología, lexicografía, lingüística antropológica, lingüística aplicada crítica, lingüística cognitiva, corpus lingüístico, morfología, neuro-lingüística, origen de la lengua, pragmática, psicolingüística, semántica, semántica lexical, sintaxis, sociolingüística y variaciones regionales.

Este mismo autor afirma que en la unión europea, "el enfoque de la lingüística aplicada se dirige a los elementos relacionados con políticas de lenguas de esta comunidad multilingüe"(Wiśniewski, 2007). Agrega que en esta unión hay necesidad de una lengua franca y de mantener las lenguas para que no se degraden ya que hay una mezcla de ellas cada vez más compleja, ocasionada por la migración de personas dentro de la unión y de los lugares limítrofes. Todo ello conduce a que el foco de atención se dirija hacia el análisis de actitudes hacia la lengua, a la adopción de políticas de lengua comunes y a la creación de libros de texto y materiales para la enseñanza de lenguas. Por último, hace énfasis en la lingüística aplicada

crítica, “interesada en problemas sociales relacionados con la lengua, tales como el desempleo, el analfabetismo y la pedagogía” (Wiśniewski, 2007). Reyes (2003), estudiante del Centro de enseñanza de lenguas extranjeras, CELE-UNAM, menciona las áreas de la lingüística aplicada a partir del reporte de Payrató (1998), quien presenta la consideración de esas áreas, conforme transcurrieron coloquios y congresos internacionales en esta materia; en un inicio, en estos congresos consideraron tres temas.

- La automatización en lingüística.
- La pedagogía de las lenguas vivas.
- La coordinación de las investigaciones en el dominio europeo.

Señala cómo fueron incrementando las áreas de estudio de la lingüística aplicada, conforme transcurrieron otros congresos. Son cuatro los congresos que presenta a partir del libro de Payrató (1998). En 1969, el II Congreso Internacional de Lingüística Aplicada se realizó en Cambridge; en 1987, se celebró en Sydney, bajo el auspicio de la AILA (Asociación Internacional de Lingüística Aplicada); en 1990, el coloquio organizado por la AILA en Tesalónica y, en 1993, en el congreso de la misma asociación, celebrado en Ámsterdam.

Los temas de cada uno de los congresos son los siguientes. Del II Congreso Internacional de Lingüística Aplicada celebrado en Cambridge en 1969: lingüística aplicada a textos literarios, análisis computacional de textos, investigaciones en la psicología de aprendizaje de L1 y L2, investigación sobre el habla y su aplicación, tecnología del aprendizaje de una lengua, metodología de la enseñanza de una lengua, aspectos lingüísticos de las perturbaciones del habla y su terapia, lexicografía, materiales para pruebas lingüísticas, medida y corrección de errores en L2, teoría de la traducción, lingüística contrastiva y sociolingüística.

En el congreso internacional celebrado en Sydney, en 1987, introdujeron los temas bajo el auspicio de la AILA. Los temas nuevos fueron: desarrollo de L1 y lenguaje infantil, enseñanza lingüística: L1 y L2, desarrollo de L2 en niños y adultos, dominio de L2 evaluación y examen, enseñanza bilingüe: enseñanza de los inmigrantes, planificación y política lingüística, multilingüismo y mantenimiento de la lengua: minorías lingüísticas, contacto de lenguas: pidgins y criollos, interpretación y traducción, lenguaje y administración: lingüística legal, lenguaje y orden social: lenguaje e ideología, neurolingüística, lenguaje y signos (incluido el lenguaje de los sordos), análisis de discurso y estilística, lexicografía, problemas de terminología, métodos cuantitativos en lingüística, tecnología del aprendizaje de la lengua: aprendizaje asistido por computado-

ra, lingüística y ciencias de la información: generación de textos y *parsing* (análisis gramatical).

En el coloquio organizado por la AILA en 1990 en Tesalónica, hubo cuatro temas primordiales: lenguaje y pensamiento; lenguaje, cultura y sociedad; enseñanza y aprendizaje de lenguas; e interacción y comunicación. En el congreso de la misma asociación, celebrado en Ámsterdam en 1993, se incluyó: aprendizaje de lenguas por parte de adultos, lingüística contrastiva y análisis de errores, metodología de la enseñanza de lenguas extranjeras y formación de profesorado, lenguaje y género, lenguas para finalidades específicas, terminología, psicolingüística, adquisición de L2, adquisición de vocabulario, muerte de lenguas y sustitución lingüística, variación lingüística y adquisición de habilidades lingüísticas. Estos temas ya se habían establecido con anterioridad.

El consenso de las áreas de la lingüística aplicada, a partir de las propuestas de los autores analizados, se presenta en el siguiente cuadro, en el cual se incluyeron áreas adicionales marcadas con un asterisco.

Cuadro 1.5. Áreas de estudio de la lingüística aplicada

• Adquisición de habilidades lingüísticas
• Adquisición de L1, L2, adquisición de vocabulario
• Análisis del discurso, análisis del discurso y estilística
• Análisis computacional de textos
• Análisis de corpus
• Aprendizaje de LE* y de lenguas por parte de los adultos
• Aspectos lingüísticos de las perturbaciones del habla y su terapia
• Cerebro y lengua
• Composición
• Contacto de lenguas, pidgins y criollos
• Corpus, corpus lingüístico, manejo de corpus
• Derechos de las lenguas minoritarias
• Desarrollo de L2 en niños y adultos
• Dominio de L2, evaluación y examen
• Educación en el área del lenguaje, educación de discapacitados

• Enseñanza y aprendizaje de lenguas
• Enseñanza bilingüe: enseñanza de los inmigrantes
• Enseñanza lingüística: L1, L2 y LE*
• Escritura
• Estilística
• Estudios literarios
• Evaluación de lengua
• Fonética y fonología
• Gramática , Gramática generativa
• Habilidades en lengua: comprensión oral, comprensión escrita, expresión oral, expresión escrita
• Identificación de autoría
• Instrucción
• Interacción, interacción y comunicación
• Interpretación
• Lengua, características, esperanto, etimologías, funciones, historia, lengua y cultura, lengua y género, lengua e ideología, lengua y medio ambiente, origen, lengua y sociedad, lengua y uso, lengua y uso en diferentes ámbitos, ámbitos profesionales
• Lengua inglesa, inglés antiguo, inglés de la edad media, inglés moderno temprano, inglés moderno tardío
• Lenguaje de signos, lenguaje de los sordos, lenguaje y género
• Lenguaje infantil, desarrollo de L1
• Lenguaje y orden social: lenguaje e ideología, lenguaje y pensamiento, lenguaje, cultura y sociedad
• Lenguas para finalidades específicas
• Lexicología , problemas de terminología
• Lexicografía



• Lingüística, lingüística afectiva, lingüística antropológica, lingüística cognitiva, lingüística contrastiva, lingüística contrastiva y análisis de errores, lingüística de corpus, lingüística forense, lingüística y género*, lingüística legal, lingüística y tecnología, lingüística y ciencias de la información: generación de textos y análisis gramatical, lingüística sincrónica y diacrónica
• Lingüística aplicada crítica
• Literacidad (traducida anteriormente como alfabetización)
• Literatura, textos literarios
• Materiales para pruebas lingüísticas
• Medición y corrección de errores en L1*, L2 y LE*
• Medios de comunicación
• Métodos cuantitativos en lingüística
• Metodología de la enseñanza de una lengua, metodología de la enseñanza de lenguas extranjeras y formación de profesores
• Metodología de la investigación
• Morfología
• Muerte de lenguas y sustitución lingüística
• Multilingüismo, mantenimiento de la lengua: minorías lingüísticas
• Neurolingüística
• Palabras, formación
• Patología del discurso
• Planeación lingüística
• Perturbaciones del habla
• Políticas lingüísticas
• Prácticas de comunicación
• Pragmática
• Profesores de lengua, capacitación
• Psicolingüística
• Psicología del aprendizaje en L1, en L2 y en LE*
• Redacción

• Retórica
• Semántica , semántica lexical
• Sintaxis
• Sociolingüística
• Tecnología, Tecnología del aprendizaje de una lengua, aprendizaje asistido por computadora
• Terapia del habla
• Terminología
• Traducción, traducción e interpretación, teoría de la traducción
• Uso de la lengua en ámbitos profesionales
• Variación lingüística, variaciones regionales
• Vocabulario

En esta lista están marcados con negritas elementos componentes, también denominados como niveles, dimensiones o subáreas de la lingüística en los estudios de lengua, como la pragmática, y las ciencias que estudian el lenguaje por sí mismo, sin enlace con otras ciencias. De acuerdo con este cuadro, pueden establecerse dos grandes áreas de aplicación de la Lingüística: la enseñanza de lenguas y tópicos relacionados (con 65 áreas de estudio) y estudios de carácter sociolingüístico (6 áreas); éstas últimas: sociolingüística en general; contacto de lenguas, pidgins y criollos; derechos de las lenguas minoritarias; muerte de lenguas y sustitución lingüística; multilingüismo, mantenimiento de la lengua: minorías lingüísticas y variación lingüística, variaciones regionales.

Las 65 áreas señaladas son fuentes de investigación en el salón de clase de lengua.

1.3. Relación de la lingüística aplicada con otras ciencias

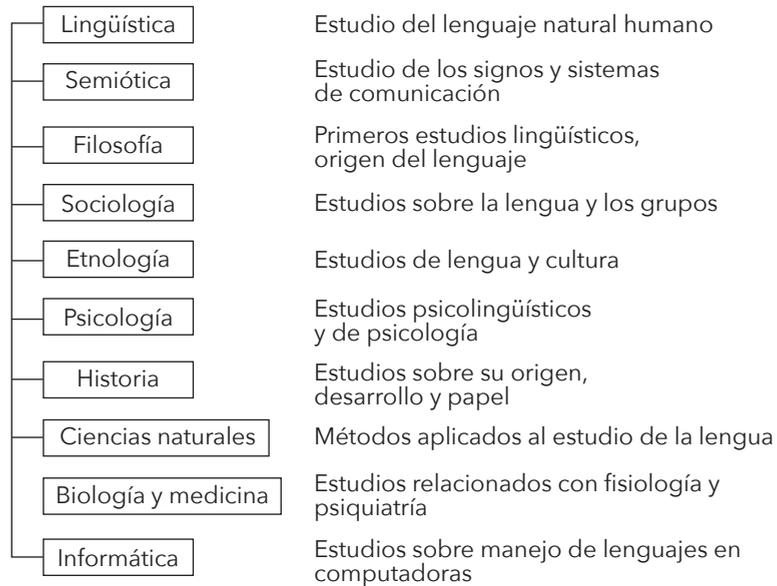
En lo que concierne a otras ciencias con las que se relaciona la lingüística aplicada, puede consultarse el libro *Fundamentos de Lingüística Aplicada* de Jesús Antonio Collado (1986).

A continuación se presenta un esquema elaborado a partir de esa obra, se rediseñó para mostrar algunas relaciones de la lingüística aplicada con otras ciencias. En los recuadros del lado izquierdo se encuentran algunas de las ciencias con las que tiene relación la lingüística aplicada.



En el lado derecho están las áreas de estudio correspondientes.

Esquema 1.1. Contacto de la lingüística aplicada con otras ciencias



Fuente: adaptado de Collado (1986).

Con base en los autores analizados y en el esquema 1.1, algunas ciencias con áreas de estudio comunes con la lingüística aplicada son las siguientes.

Cuadro 1.6. Ciencias con áreas de estudio comunes con la lingüística aplicada

- Administración pública
- Antropología
- Biología
- Ciencias naturales
- Ciencias políticas
- Computación
- Derecho
- Educación
- Etnología
- Etnografía
- Filosofía
- Historia
- Informática
- Medicina
- Psicología
- Semiótica
- Sociología

Estas 17 ciencias no son las únicas que tienen áreas de estudio comunes con la lingüística aplicada pero sí son las más relevantes.

1.4. Salones de clase de lengua

Una pregunta ya tradicional es ¿el salón de clase, particularmente el salón de clase de lengua, es un escenario para llevar a cabo investigaciones?

En el salón de clase de lengua ocurren cotidianamente relaciones humanas actuadas y vividas por personajes muy propios de su medio ambiente. Ello permite ser una fuente de múltiples investigaciones.

Con el fin de contrastar los puntos de vista de lo que significa un salón de clase de lenguas expresados por alumnos de la Licenciatura en Lenguas de la Facultad de Lenguas de la Universidad Autónoma del Estado de México, se presenta el análisis de las encuestas que respondieron, como personajes mismos de los sucesos en su interior, y la opinión de Ruth Amanda Salcedo (2002).

Esta opinión fue tomada de la introducción a la *Memoria del V Congreso de Investigación Educativa e Innovación Pedagógica*, con el tema: *la investigación en el aula: y la innovación pedagógica*, como una autoridad en el ámbito educativo. Para ella, “el salón de clase es el lugar deseable en el que se produce conocimiento”.

En la introducción a la memoria mencionada, inicia diciendo que “El aula de clase es descrita y vivida por muchos como un espacio jerarquizado y regulado en el que se evidencian estructuras de poder que determinan, por tanto, la dinámica de las relaciones sicosociales de los allí involucrados”; hace resaltar elementos externos al salón de clase, las estructuras de poder y su relación con los internos, como las relaciones sicosociales.

Contempla el salón de clase como un fenómeno social al afirmar que “el aula es algo más que el espacio en el que sólo se transmiten ideas o pautas de comportamiento” (Salcedo, 2002), en el salón de clase hay una socialización que es consecuencia de dos elementos: las prácticas sociales y las interacciones que surgen y se desarrollan. En el salón de clase “la relación dialéctica en la que conviven paralelamente tendencias conservadoras de reproducción con corrientes renovadoras que impulsan el cambio como condición también de supervivencia y humanización está atravesada por el conocimiento” (Salcedo, 2002).

Lo más relevante en su afirmación es que concibe el conocimiento como una herramienta educativa que debe dar a los educandos los insumos necesarios para la comprensión y explicación de los fenómenos sociales a través de una reconstrucción crítica de los mismos.

En síntesis, las categorías de los elementos de un salón de clase que pueden establecerse a partir del discurso de Salcedo son: el espa-

cio; la presencia del conocimiento; los elementos externos; el fenómeno social de las prácticas sociales y las interacciones; y la relación dialéctica entre las tendencias en el salón de clase, conservadoras y renovadoras, entre las que “atraviesa el conocimiento” (Salcedo, 2002).

Por otra parte, el punto de vista de los alumnos de la Facultad de Lenguas se recabó durante varios semestres, mediante una pregunta hecha el primer día de clase de la asignatura o unidad de aprendizaje Metodología de la Investigación, cursada en el primer semestre. La pregunta es: para ti, ¿qué es un salón de clase?

A continuación se presenta el resultado del análisis de las respuestas. Las respuestas se categorizaron como: el salón de clase (naturaleza y características), objetivos en el salón de clase, los personajes y su papel, afecto en el salón de clase y sucesos en el salón de clase. Los siguientes datos se tomaron de las recopilaciones de los semestres 2009B, 2011B, 2012B y 2013B.

Cuadro 1.7. Concepciones de los alumnos acerca del salón de clase

Categoría	Componentes
El salón de clase, su naturaleza	<ul style="list-style-type: none"> • Es un lugar, un espacio, un espacio físico, un aula, un lugar académico, un cuarto o una sala, un espacio aislado • No es un aula sino los elementos para lograr un ambiente de aprendizaje • Espacio con elementos que actúan de acuerdo con normas establecidas • Es una estructura • Es el “mayor” sitio para aprender • Primordial y esencial para el desarrollo • Son los estudiantes y el profesor • Es como un segundo hogar • Es un grupo donde hay apoyo, unión y respeto • Es un conjunto de personas que se reúnen en un lugar para aprender • Es un grupo de personas con intereses en común • Brinda respuestas • Hay mobiliario
Objetivos en el salón de clase	<ul style="list-style-type: none"> • Impartir clases • Desarrollar y lograr el aprendizaje, la concentración y el aprendizaje • Expresar ideas • Aclarar dudas • Adquirir, compartir y reforzar conocimientos • Intercambiar opiniones • Compartir cualidades y capacidades • Estudiar • Perfeccionar habilidades adquiridas • Incrementar la cultura personal • Entablar comunicación

Personajes y su papel	<ul style="list-style-type: none"> • Son dos, el profesor y los alumnos • Todas las personas tienen la voluntad y las ganas de aprender • Los que quieren aprender y los que quieren compartir sus conocimientos • El papel del profesor. Es concebido como el poseedor del conocimiento, lo comparte con los alumnos. Es capaz de ver las capacidades y habilidades de cada alumno • El papel del alumno es compartir sus conocimientos, tiene que poner de su parte con respeto e interés y es el principal actor del proceso enseñanza/aprendizaje
Afecto en el salón de clase	<ul style="list-style-type: none"> • Es necesario que haya buena comunicación • Es necesario que haya comprensión entre el maestro y los alumnos • Se conocen amigos
Sucesos en el salón de clase	<ul style="list-style-type: none"> • Hay relaciones entre las personas, profesor-alumno y alumno-alumno • Hay convivencia entre el profesor y los alumnos. Se convive en forma interactiva • Por la comunicación de ambos se obtiene conocimiento • Hay contacto verbal y físico • Hay un trabajo en equipo • Los estudiantes desarrollan una parte de su vida • Hay actividades de aprendizaje y convivencia y de enseñanza/aprendizaje • Hay intercambio de ideas, conocimiento, dudas y puntos de vista • Se realizan trabajos de aprendizaje • Hay formación cada día para ser mejores • Hay interacción entre los maestros y los alumnos • Se adquieren procedimientos y herramientas para el aprendizaje • Se enseña y se aprende por parte de los dos • Se viven nuevas experiencias • Hay que inculcar el valor de la confianza y el respeto • Se enseñan y aprenden cosas diferentes

En el análisis destacaron los siguientes comentarios: el salón de clase “es como mi 2ª casa porque en la forma que yo trate a los demás es lo que voy a recibir”, comentario que conlleva mucha carga afectiva, y “un salón de clase para mí es como un ‘calmecac mexicatl’³, donde el pupilo encuentra y/o forma su vocación con ayuda del maestro preparándolo para el mundo exterior”.

³ “el calmecac era el colegio sacerdotal para educar a los jóvenes de la nobleza mexicana” (Dehouve, 2013).



Llama la atención el hecho de que los alumnos consideran al maestro como el poseedor y transmisor del conocimiento.

Al contrastar las categorías señaladas por Salcedo (2002) con las que surgieron del análisis de las encuestas, se encuentra que sus categorías revelan un interés muy alto en los elementos externos al salón de clase como generadores de situaciones al interior del mismo; factores como: las estructuras de poder, las prácticas sociales y las tendencias educativas hacia la conservación o la renovación. Las categorías obtenidas con la información de los alumnos, por el tipo de pregunta que se les hizo, se concentran en el aula y, particularmente, en las experiencias de los maestros y los alumnos, en la interacción entre ellos, donde destaca el afecto y el conocimiento.

Hay una coincidencia, en los dos casos, en cuanto a la naturaleza del salón de clase, como un espacio en donde hay interacción entre los sujetos inmersos en él y en donde está presente el conocimiento.

Los alumnos pasan muchas horas cautivos en los salones de clase; así también los maestros. El salón de clase es un pequeño universo cotidiano en el que tienen lugar un sin fin de experiencias entre ellos que no son sólo académicas. Sus propios avatares están ahí. Sus aportaciones personales de todo tipo, académicas, sociales, afectivas, cognitivas, su practicidad, su pensamiento, se conjugan e interactúan para propiciar o detener el aprendizaje. Es así como surge la necesidad de estudiar los muy numerosos contextos y sus elementos que pueden presentarse en ese medio ambiente tan especial y específico. Son diversos los temas que pueden abordarse en los estudios de salón de clase; particularmente en los salones de enseñanza de lengua, en los que la lengua misma es un instrumento para la comunicación y materia de estudio. Qué decir cuando se trata de aprender una lengua extranjera, con el juego entre la lengua materna, la lengua que se aprende y las demás que puedan estar presentes.

El aprendizaje de una lengua se desarrolla en dos ambientes principales: informal y formal. Este último se asocia con los salones de clase. Las dimensiones que pueden abordarse implican la consideración de distintos componentes. Entre ellos, el más importante es el de los personajes; también están los procesos, los materiales y los instrumentos pedagógicos, enlazados entre sí y con otros factores externos al salón de clase.

1.4.1. Personajes en el salón de clase de lengua

La denominación de personajes para los actores de la educación en un salón de clase de lengua y fuera de él nace de un contacto temprano con el teatro que condujo a hacer conciencia de que en él, finalmente, par-

ticipan personas que conjugan en su experiencia su propio papel como individuos, del que no pueden desprenderse, con el papel, también propio, de la obra con la que se han comprometido.

Los personajes principales en un salón de clase de lengua son el maestro y los alumnos así como todas aquellas personas que no se encuentran físicamente ahí pero están presentes en las mentes de ambos, en sus participaciones. El maestro y los alumnos son una gran fuente de objetos de estudio a investigar en un salón de clase de lengua que son aportados por cada uno de ellos al ingresar al aula; encontramos elementos de su persona y de su comportamiento individual y grupal así como de la relación entre ellos.

Para ambos, maestro y alumnos, entre los personajes externos están principalmente los padres y otros miembros de la familia, todas aquellas personas que ejercen alguna influencia en ellos, un novio o novia, el esposo o la esposa, los amigos. También son personajes externos la institución misma en la que se trabaja y los que laboran en ella; las políticas educativas, los planes y programas de estudio, los textos predeterminados, los requisitos de enseñanza/aprendizaje, las normas de la institución. Y, en un peldaño más alto, que pareciera lejano pero que puede ejercer una presión educativa, están también las políticas nacionales e internacionales en materia de educación, como autoridades que dictaminan los senderos de la educación en el país y en el mundo. Se habla de los empleadores, los que darán fuentes de trabajo a los egresados de las universidades. También ellos establecen exigencias y necesidades que marcan las consideraciones educativas y académicas para la formación de los egresados de una Licenciatura en el área de la enseñanza/aprendizaje de una lengua.

¿Qué aportan los profesores y los alumnos al salón de clase de lengua? Ambos aportan sus propios perfiles físicos y psicológicos, como personas y como protagonistas de los papeles que desempeñan, así como sus objetivos, conocimiento y expectativas.

Los papeles más relevantes en un sentido educativo se refieren a su actuación e interacción en el binomio enseñanza-aprendizaje, en el que la investigación gira a su alrededor, para analizar primordialmente los resultados de aprendizaje de los alumnos frente a las estrategias de enseñanza de los profesores.

1.4.2. Procesos en el salón de clase de lengua

La presentación de Liliana López Sorzano (2011) en la que relata un ejercicio de Nicolás Paris, aprendiz de maestro en ese momento, ilustra una experiencia educativa de lo que acontece en un salón de clase.

Nicolás Paris trabajó con el modelo de educación Escuela Nueva en una escuela rural con 16 alumnos de diferentes edades y niveles académicos, todos ellos estuvieron en el mismo salón de clase, se utilizó el dibujo como instrumento de aprendizaje. Los resultados le permitieron pensar en rechazar “el proceso de aprendizaje tradicional en donde las materias se toman por separado y entendió que en el salón de clase como en la vida todo está en constante relación, lejos de los procesos fragmentados” (López, 2011). También cambió su posición como profesor, dejó de ser un maestro que “deposita saber”, para convertirse en “un mediador, un facilitador, alguien que construye conocimiento al lado de sus pupilos, no desde arriba, sino a su lado” (López, 2011).

El interés sustancial de este inciso es hablar de los procesos que ocurren al interior del salón de clase de lengua, procesos que se ven afectados o modificados por los de carácter externo; los procesos, como aquellos conjuntos de pasos o etapas que se presentan en los salones de clase de lengua para dar lugar al desarrollo de la enseñanza y aprendizaje de una lengua.

En una conjugación de elementos de los personajes, principalmente el maestro y los alumnos, en la interacción entre ellos surgen diversos procesos que ocurren paralelamente en el salón de clase de lengua, los que son muy vastos. En este inciso, los procesos en el salón de clase de lengua se clasifican en las áreas generales: afectiva; cognitiva; de interacción; educativa y académica, de enseñanza y aprendizaje; de investigación y administrativa; desde luego, el rubro más extenso es el del área educativa y académica. Las áreas: afectiva, cognitiva y de interacción corresponden a los personajes maestro y alumnos y al intercambio que acontece entre ellos.

Se incluyó la actuación de los personajes, sus expectativas y sus objetivos en el área educativa y académica porque, a la vez que combinan elementos afectivos y cognitivos de los profesores y alumnos, implican su actuación en el salón de clase, en un ámbito educativo; profesores y alumnos son portadores de entidades para su estudio en el salón de clase de lengua; también se incluyen la actuación, expectativas y objetivos de los personajes fuera del salón de clase.

El área administrativa incluye básicamente los elementos manejados por el profesor.

Con respecto a la investigación y la innovación en el aula, existen dos formas de abordar la investigación: a través de la participación de un investigador externo o mediante la participación del profesor mismo como en los casos en los que se usa la forma particular de la investigación cualitativa conocida como investigación-acción. El siguiente cuadro muestra estos procesos.

Cuadro 1.8. Procesos en el salón de clase de lengua

Área general	Procesos
Afectiva	<ul style="list-style-type: none"> • Manejo del afecto • Expresión del afecto • Recepción del afecto
Cognitiva	<ul style="list-style-type: none"> • Memoria • Pensamiento • Lenguaje • Metacognición
De interacción	<ul style="list-style-type: none"> • Intercambio de afecto • Intercambio de conocimientos, ideas, pensamientos y opiniones • Comunicación • Reflexión (profesores-alumnos, alumnos-alumnos)
Educativa y académica	<ul style="list-style-type: none"> • Actuación de los personajes • Expectativas de los personajes • Objetivos de los personajes • Competencia docente • Aprendizaje, estilos y estrategias • Enseñanza, estrategias, métodos y modelos • Discurso áulico • Práctica educativa: la docencia, la pedagogía y la formación • Uso de la lengua en el salón de clase: lengua materna, segunda lengua y lengua extranjera • Evaluación de conocimientos, del aprendizaje • Lectura y escritura • Práctica en el aula • Materiales • Uso de tecnología
Administrativa	<ul style="list-style-type: none"> • Papel gerencial del docente • Orden y disciplina • Funcionamiento del medio ambiente • Organización del aula • Creación de escenarios • Normas en el salón de clase • Logística
De investigación	<ul style="list-style-type: none"> • Investigación en el aula • Innovación

Si bien existen procesos externos al salón de clase, éstos no constituyen un punto de interés primordial de este trabajo. Esos serían los que competen a los personajes que actúan y ejercen una influencia pero que no están físicamente presentes, mencionados en el inciso 1.3.1, como pueden ser los de las autoridades administrativas, en cuanto a políticas

educativas públicas y privadas, sociales y gubernamentales. A estos procesos puede agregarse el que define Mendoza (2009) como el proceso de cambio en el salón de clase, el cual es más bien un proceso externo que tiene repercusiones en los procesos internos. Abarca la forma de concebir los procesos de trabajo en las escuelas, la evolución de los planes de trabajo, las dinámicas propias de cada escuela y de cada grupo y los acontecimientos sociales que repercuten en los cambios escolares. En este momento, puede agregarse el tema de actualidad mundial, el *bullying* en las instituciones educativas; tema que pertenece a los ámbitos interno y externo al salón de clase.

1.5. Áreas de estudio en un salón de clase de lengua

Estas áreas, mostradas en el siguiente cuadro, se eligieron a partir de las áreas de la lingüística aplicada que tienen relación con el estudio de una lengua y con la docencia de la misma, y con base en los elementos analizados para el salón de clase. En el siguiente cuadro se muestran dos columnas: la primera indica el área de estudio y la segunda, las subáreas.

Cuadro 1.9. Áreas de estudio en el salón de clase de lengua

Adquisición y aprendizaje	<ul style="list-style-type: none"> • L1, L2, LE y habilidades lingüísticas: comprensión y expresión oral y escrita y gramática • Estrategias de aprendizaje • Estilos de aprendizaje • Medición y corrección de errores en L1, L2 y LE • Lengua para propósitos específicos • Bilingüismo
Enseñanza	<ul style="list-style-type: none"> • Estilos de enseñanza • Estrategias de enseñanza • Los mismos rubros que en adquisición y aprendizaje
Evaluaciones	<ul style="list-style-type: none"> • Exámenes y certificaciones de lengua
Enseñanza/aprendizaje	<ul style="list-style-type: none"> • Manejo y producción de materiales • Prácticas de comunicación • Tecnología
Lengua	<ul style="list-style-type: none"> • Características, funciones

Fonética y fonología	
Morfología	
Sintaxis	
Semántica	
Pragmática	
Uso de lengua en el aula	<ul style="list-style-type: none"> • L1, L2, LE, bilingüismo • Profesores • Alumnos
Discurso áulico	
Lengua	• Cultura, ideología y pensamiento
Lectura	• Textos de diversos géneros
Escritura	• Generación de textos de diversos géneros
Lexicología y lexicografía	
Terminología	
Manejo de corpus lingüístico	
Traducción	• En términos de enseñanza/aprendizaje, no como énfasis terminal
Interpretación	
Interacción	• Interrelación, intercambio y reflexión
Psicolingüística	• Afectividad y cognición
Formación de profesores	
Planeación Lingüística	

En este cuadro, las áreas de estudio se presentan bajo un esquema de salón de clase. Son doce grandes áreas, separadas en el cuadro por una doble raya. Se inicia por los aspectos académicos: la adquisición y el aprendizaje y la enseñanza. Son las dos áreas primordiales en los estudios de salón de clase de lengua por ser las que corresponden al objetivo más relevante, la investigación de los resultados del proceso de enseñanza-aprendizaje. El tercer grupo corresponde a las evaluaciones del aprendizaje de los alumnos por los profesores, un campo de estudio estrechamente enlazado con la investigación de los resultados de la enseñanza.

El cuarto se refiere a elementos que sustentan la enseñanza/aprendizaje. El quinto grupo está constituido por los estudios de lengua, sea una lengua nativa, una segunda lengua o extranjera; las disciplinas complementarias de la lingüística que analizan las lenguas (Florit, 2013); el



uso de lengua o lenguas en el aula; el discurso en el aula y la lengua y sus relaciones con la cultura, la ideología y el pensamiento.

El sexto grupo incluye la lectura y la escritura, dos procesos enlazados que son fundamentales para el aprendizaje en general y de las lenguas.

El séptimo grupo enmarca dos ciencias que apoyan el estudio y descripción del léxico de una lengua y la terminología, importantes para la elaboración de diccionarios, glosarios y vocabularios. En este sentido, se llevan a cabo estudios para una lengua con propósitos específicos.

El octavo, sobre el manejo de corpus lingüístico, desde los pequeños hasta los que contienen registros que no son manejables manualmente.

El noveno grupo está destinado a la traducción e interpretación. Este rubro, también dirigido a los procesos de enseñanza y aprendizaje.

El décimo grupo abarca las áreas propias de la relación entre los personajes del salón de clase de lengua: la interacción y la psicolingüística. Las dos últimas áreas, formación de profesores y planeación lingüística, son externas en relación con el salón de clase de lengua; se incluyen dada su relevancia en el proceso general.

1.6. La psicolingüística

Se incluye un inciso para tratar el tema específico de la psicolingüística, para mostrar que, si bien surge a través del enlace de la lingüística con la psicología, hasta el momento, se ha puesto énfasis solamente en una parte del área cognitiva de la psicología, para mostrar la ubicación del lenguaje en el cerebro, su construcción, su manejo, su relación con el pensamiento. Es vital la consideración de otras áreas de gran relevancia en psicología, el área afectiva, la que comprende las actitudes, la motivación, las creencias, la ansiedad, los estilos de aprendizaje; el área cognitiva con la aptitud, las estrategias de aprendizaje (Gardner y MacIntyre, 1993); elementos todos ellos que, al fusionarse con la lingüística, también conducen a estudios de psicolingüística que son indispensables principalmente para las investigaciones que se desarrollan para conocer la relación de los personajes y los procesos en el salón de clase.

La literatura acerca de la psicolingüística habla de la forma en la que el ser humano adquiere y usa el lenguaje, de las áreas cerebrales en las que se ubica. Así, se le atribuye como materia de estudio: el lenguaje, su uso y comprensión, el pensamiento, la memoria, los procesos cognitivos, la adquisición de la lengua por los niños o el aprendizaje de la lengua por los adultos, bajo el punto de vista de la psicología cognitiva.

Si se considera que la psicolingüística es una ciencia que resulta de la fusión de estudios comunes a la lingüística y a la psicología, además de los fenómenos cognitivos, es necesario considerar los afectivos, campo de estudio muy vasto, para dar cabida a una visión más completa de las propuestas de estudio de la psicología y la lingüística. En esta línea de trabajo, están precisamente los personajes y los procesos en el salón de clase de lengua, con todas esas características que no son solamente de la esfera cognitiva.

Dentro de esta parte de la psicología, se estudian temas como: personalidad, temperamento, carácter, actitudes, percepción, motivación, rasgos de comportamiento, sentimientos, afecto, creencias, comportamiento y ansiedad, entre otros. Esto significa que, al abordar el estudio de estas características en su relación con aspectos del área cognitiva, como el aprendizaje de una lengua, estamos inmersos en estudios de corte psicolingüístico.

Por ejemplo, cuando se llevan a cabo estudios con protocolos verbales, en los que se pide a los sujetos que simultáneamente resuelvan los cuestionamientos de una prueba y narren la forma en la que lo hacen, están presentes elementos psicológicos tanto cognitivos como afectivos. Es decir, se analiza el proceso de la lengua cuando el sujeto revela en voz alta, en protocolos, la manera en la que responde o procesa una respuesta y sus reacciones ante las preguntas.

1.7. Investigación en las áreas de estudio en un salón de clase de lengua

Para analizar la investigación en las áreas de estudio en un salón de clase de lengua, es menester revisar algunos elementos de la investigación en general en la actualidad. Hacia los años 90 y al inicio del siglo XXI se hablaba de los debates originados por las diferencias entre las investigaciones cuantitativa y cualitativa y de cómo cada uno de los defensores de cada paradigma pretendía demostrar la validez de los resultados de sus investigaciones, algunos todavía aferrándose al método científico, los otros en una reacción contra el uso de una metodología originada en las ciencias naturales, no aplicable a los estudios en las ciencias sociales.

Finalmente, algunas características establecidas para diferenciar ambas metodologías, como el número de sujetos o el número de variables estudiadas, con el tiempo han sido dejadas de lado. Prácticamente, las diferencias principales residen en el manejo cuantitativo de los grandes números, por un lado, en la investigación cuantitativa y la categorización para la descripción de los fenómenos, por la otra, en la investigación cualitativa. No obstante, una forma de establecer categorías es el conteo

de regularidades y características compartidas. Hoy en día, se analizan resultados de investigaciones cualitativas mediante la cuantificación.

En la investigación contemporánea en los salones de clase destaca el artículo de Quezada y Canessa "La complejidad de los procesos educativos en el aula de clases"(2008) en el que hablan de "una serie de procesos tanto individuales como sociales, los cuales están integrados dentro de otro de mayor envergadura: educación". Analizan un aula como un sistema de individuos y dinámicas que no pueden explicarse como comportamientos aislados entre ellos. Su artículo tiene como objetivo "entregar elementos que permiten comprender a [sic] los fenómenos educativos de aula como comportamientos propios de sistemas complejos, sistemas que se componen de unidades interactuantes y excitables de las que emergen comportamientos de los cuales la predictibilidad exacta es imposible" (Quezada y Canessa, 2008, p. 103). Lo importante es rescatar que los sistemas complejos constan de elementos interactuantes y excitables entre sí, denominados agentes que cambian sus estados internos a través de la interacción con otros agentes y con el medio ambiente. Citan a Holland y Miller (1991) para decir que el comportamiento individual de los agentes puede ser simple pero "la interacción de éstos produce comportamientos y características del sistema que como un todo son diferentes a una agregación lineal de sus conductas individuales" (Quezada y Canessa, 2008, p. 106). Por último concluyen que:

Los Sistemas Complejos no sólo explican mejor sino que pueden dar pistas para diseñar intervenciones educativas más certeras. Si los procesos educativos corresponden a fenómenos complejos, esto no hará necesariamente más difícil la intervención o reducirá las expectativas de éxito, pero reconocer eso permitirá que dicha intervención sea menos ingenua y pondrá límites realistas a los resultados esperados. Este marco permite abrazar la complejidad inherente al fenómeno y, así, promover las acciones más adecuadas para los fines del proyecto educativo (Quezada y Canessa, 2008, p. 116).

Por un lado, introducen el elemento de la complejidad y, por el otro, coinciden con Davis (2007), cuando afirman que no se espera que los resultados de las investigaciones sean respaldados con certeza.

En el sentido de la complejidad, es necesario comentar acerca de Edgar Morin y el pensamiento complejo, quien al hablar de este tema en una entrevista que le hizo la periodista Elda Molina (2008), afirma que la cuestión fundamental en el pensamiento complejo son las relaciones de las partes al todo y del todo a las partes. Esto "significa que no es únicamente una parte que se encuentra en el todo y las cosas complejas, es el todo que se encuentra en las partes" (Molina, 2008). Por ello, considera como un defecto la separación de las disciplinas que impide la visualización y concepción de los problemas fundamentales y globales. Por lo tan-

to, señala como importante la transdisciplinariedad que “no es solamente la colaboración de varias disciplinas sino una concepción capaz de atravesar la decisión de poner la correlación en un modo coherente organizativo de los elementos de conocimiento de las disciplinas” (Molina, 2008).

Otro paradigma importante en estas investigaciones es la manera en la que se definen los elementos a estudiar en relación con el objeto de estudio. La investigación tradicional, marcadamente cuantitativa, ha trabajado con variables, primordialmente con las variables independiente y dependiente. Sin embargo, en la actualidad y precisamente en términos de la complejidad, las entidades abordadas ya no pueden estudiarse solamente como un todo. Surge entonces la necesidad de estudiar la entidad desglosada, dividida en partes; lo que ha dado lugar a otras denominaciones de los elementos estudiados.

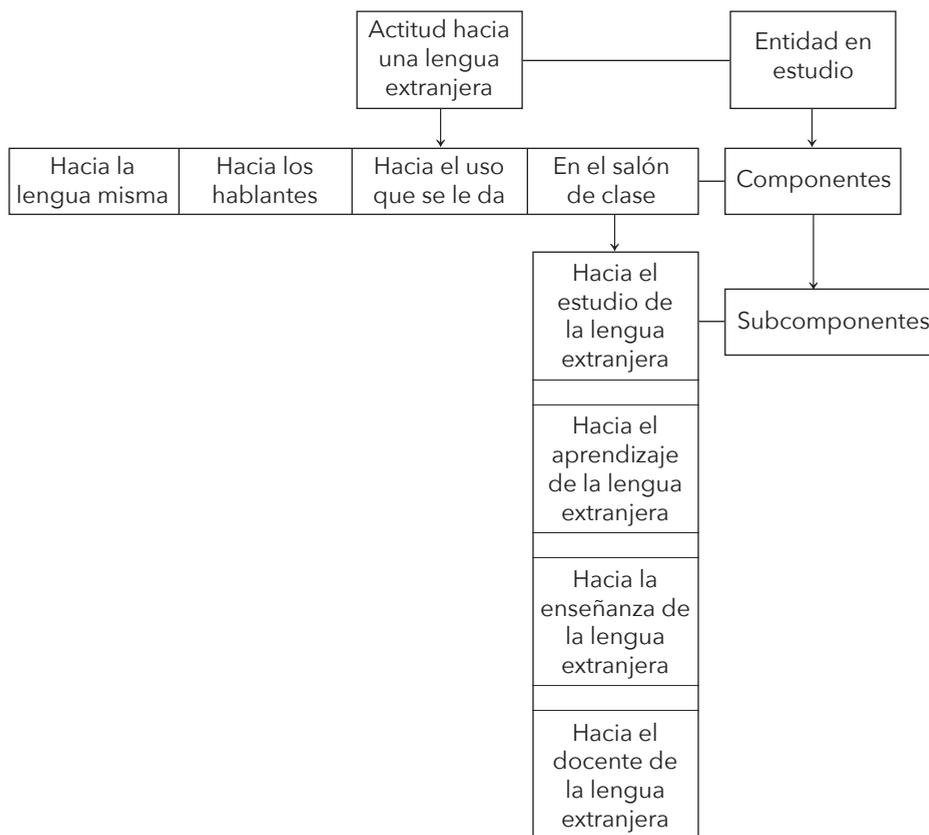
Una manera de abordar los fenómenos bajo este tipo de concepción es el análisis componencial. Este análisis estudia las entidades materia de las investigaciones a través de sus componentes. La separación en componentes permite contar con mejores resultados, más reveladores; por ejemplo, en una medición cuantitativa de nivel de manejo de la lengua inglesa, del rendimiento en el aprendizaje de la misma, existe una diferencia significativa si en vez de contar con un puntaje de rendimiento global se obtiene un puntaje separado para cada una de las habilidades: comprensión oral, expresión oral, comprensión escrita y expresión escrita.

Podrían agregarse otros componentes más, como conocimiento de gramática o de vocabulario. El panorama acerca del objeto de estudio es más amplio si se trabaja con sus componentes.

Cuando se estudia la actitud hacia una lengua extranjera, particularmente en relación con el salón de clase, una entidad que forma parte del objeto de estudio es esa actitud; sus componentes son los tipos de actitudes que pueden observarse.

Estos componentes también se dividen en componentes. El esquema 1.2 los muestra. Aún los componentes desglosados en este caso para la actitud hacia una lengua extranjera en el salón de clase pueden subdividirse en sus propios componentes. Lo importante es arribar a elementos del objeto de estudio que puedan ser medidos u observados. Como se mencionó, a mayor número de componentes medidos, mayor amplitud para el análisis de resultados.

Esquema 1.2. Componentes de la actitud hacia una lengua extranjera

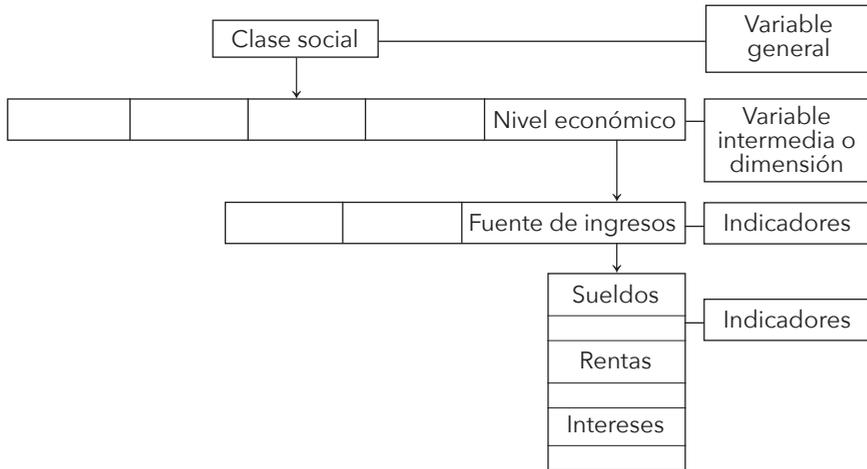


Fuente: elaborado a partir de Gardner y Lambert (1972, p. 50), Hohental (s/f) y Uribe (2002, p. 31).

Otros autores manejan equivalentes de los componentes del término original variable, como dimensiones e indicadores de estudio de las entidades. Rivera (2011) maneja que “una variable es un concepto menos abstracto del concepto del cual proviene y se puede observar en la realidad. Contiene indicadores a través de los cuales se pueden contar y/o medir aspectos concretos del objeto”. Utiliza indistintamente la denominación de variable y de dimensión. Explica la operacionalización de las variables como “el paso de variables generales, no observables directamente a variables empíricas o indicadores” (Rivera, 2011). Hay, entonces, una variable o dimensión general, variables intermedias o dimensiones de la variable general e indicadores, que son a su vez dimensiones de las variables intermedias. Pone como ejemplo el que se expresa en el esquema

1.3, utilizando el mismo esquema presentado para el análisis componencial, con una terminología diferente.

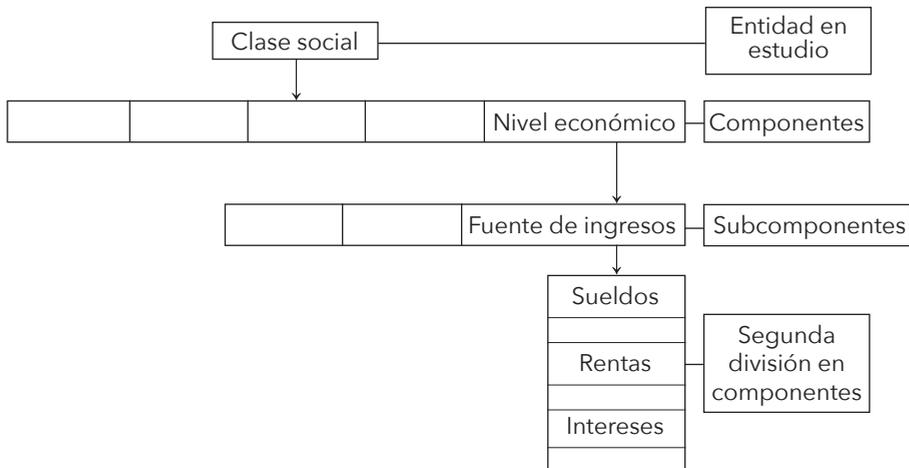
Esquema 1.3. Dimensiones e indicadores de la clase social



Fuente: elaborado a partir de la clasificación de Rivera (2011).

El mismo ejemplo, expresado a través del análisis componencial, se encuentra en el esquema 1.4.

Esquema 1.4. Componentes de la clase social expresados en el análisis componencial



Fuente: elaborado a partir de la clasificación de Rivera (2011).

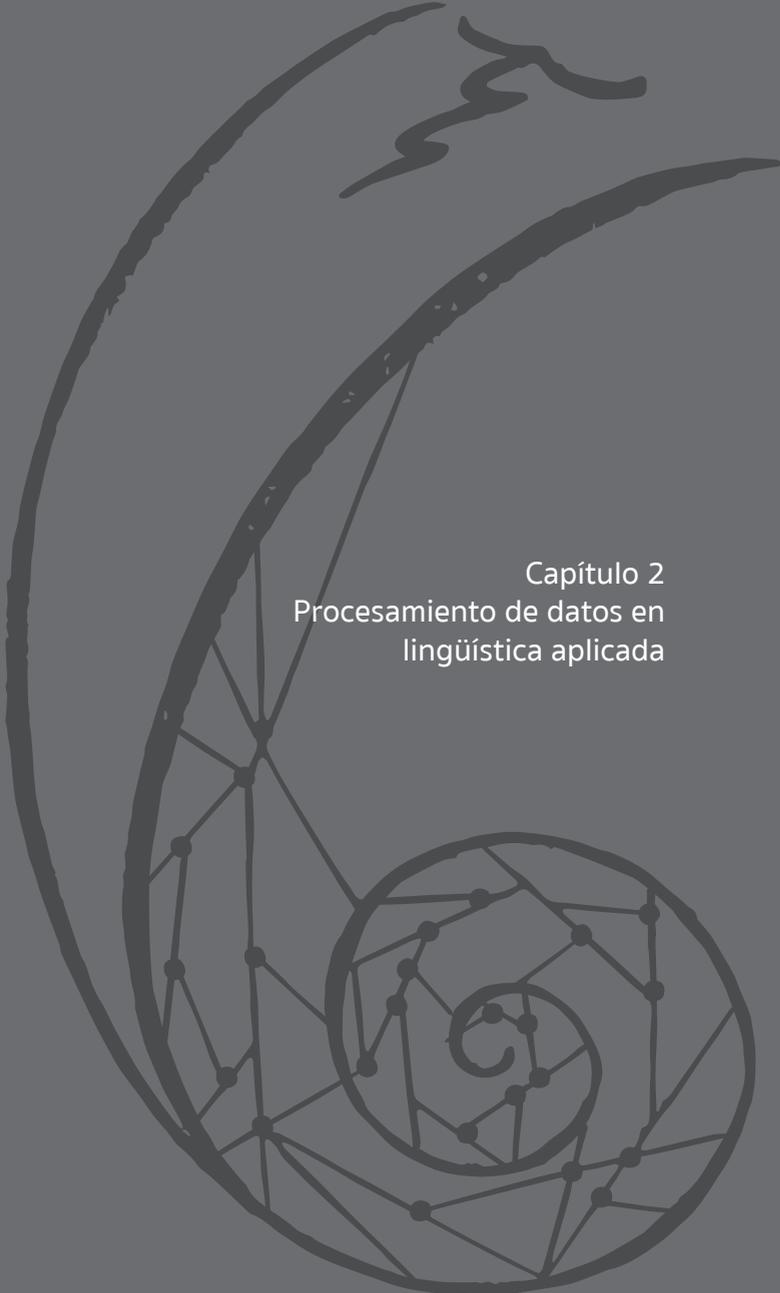
La expresión de este ejemplo en los esquemas 1.3 y 1.4 muestra que la forma de contemplar los elementos bajo estudio y medición u observación es equivalente. En cada caso general o particular es posible plantear componentes o bien dominios de estudio, dimensiones o indicadores, para dar pie a la aplicación de diversas metodologías en la consideración de las entidades bajo estudio.

En este capítulo se comentó la falta de un acuerdo que defina las ciencias denominadas como ciencias del lenguaje: las áreas que conforman la lingüística aplicada, como un contexto punto de partida para los estudios en un salón de clase de lengua. Se delineó el salón de clase de lengua como un espacio generador de investigaciones del proceso enseñanza-aprendizaje de una lengua, que puede ser la lengua materna, la segunda o extranjera. Se encuentran las áreas y subáreas de estudio en ese espacio y se dieron lineamientos para las consideraciones acerca del objeto de estudio y sus componentes.

Finalmente, como recomendaciones adicionales a lo tratado, hay que enfatizar que cuando se llevan a cabo estudios en salón de clase de lengua, uno de los elementos de partida es el establecimiento de las características de los sujetos bajo estudio en la investigación; las que siempre deben tomarse en cuenta son: la edad de los sujetos y su nivel de dominio de la lengua estudiada. Es esencial considerar que las personas poseen capacidades individuales distintas.

Los profesores y los alumnos llevan a cabo investigaciones en áreas distintas a las que se refieren los estudios en un salón de clase de lengua. Lo que se ha presentado en este capítulo, a pesar de centrarse en investigaciones de esta naturaleza, no invalida las bondades por el uso de pruebas estadísticas y procesamiento de datos en otras áreas de la lingüística aplicada que no son precisamente en el salón de clase de lengua.





Capítulo 2
Procesamiento de datos en
lingüística aplicada



El segundo capítulo presenta el tema de la documentación, no en el sentido del manejo de documentos sino de la necesidad de llevar a cabo un registro cotidiano de lo que va aconteciendo en la investigación y de los elementos que la conforman, como los enlaces de ideas y relaciones que se producen durante su transcurso, a través de las lecturas y los hallazgos de campo; se señala la relevancia y las características de los instrumentos para el levantamiento de datos, su significación y el procesos de ellos en los paradigmas tanto cuantitativos como cualitativos.

Como en otras ciencias, el procesamiento de datos en lingüística aplicada se contempla en varios momentos: cuando se planean y se eligen o se crean los instrumentos para su levantamiento, cuando se registran y cuando se analizan. El primer momento es aquel en el que se decide qué instrumentos son necesarios para el levantamiento de datos y se elige entre el uso de instrumentos hechos con anterioridad o la creación de instrumentos exprofeso para la investigación. El registro se lleva a cabo en momentos precisos o durante el desarrollo de la investigación y, finalmente, la tercera fase es para analizarlos y obtener resultados a partir de ellos. Este capítulo consta de la documentación de un proyecto de investigación a través de su desarrollo, el diseño de una investigación, los instrumentos, el tratamiento, los datos levantados y el procesamiento de los mismos.

2.1. Documentación de un proyecto de investigación

Al hablar de un proyecto de investigación, necesariamente surge el tema de la "documentación". Esta documentación no se refiere a la consulta y manejo de documentos, indica que es indispensable documentar un proyecto, desde su inicio y durante su desarrollo. Esto significa que cotidianamente hay que registrar los hechos, acontecimientos, actividades, datos e información alrededor del mismo; es vital para la fluidez de su desarrollo. Algunos investigadores utilizan un diario de campo; otros simplemente llevan consigo una libreta para anotar todo lo mencionado y, principalmente, las ideas generadas alrededor de la teoría y la práctica de la investigación.

Cuando el investigador está inmerso en un proyecto de investigación, posee mucha información, información generada a través de la lectura, de la reflexión, de las conversaciones con otras personas. Se cree que se recuerda todo y que la información surge de inmediato. La realidad es que no sucede así. Cuando el investigador asume otras responsabilidades o toma alguna pausa para realizar otras actividades, la calidad de la información en el cerebro disminuye. Por eso, es importante registrar, escribir, documentar todo lo que se ha mencionado como soporte del trabajo.

Esta actividad de documentar el proyecto incluye el manejo de las fuentes de información. Esto significa que, paralelamente, durante el desarrollo del proyecto, las referencias tanto del marco teórico como del inciso destinado para ello deben documentarse en algún medio. Lo más recomendable es la creación de un archivo de referencias con el que se trabaja todo el tiempo.

Algunos autores denominan esta actividad como el diario de campo. Medina (2012) considera que los diarios de campo representan una estrategia de suma importancia porque "al documentar los hechos relevantes y cuestiones de menor importancia o hechos perdidos en la interpretación, generalización, evaluación o presentación de los resultados, vistos desde la perspectiva del investigador, se suscitan, inevitablemente, procesos de reflexión" (Medina, 2012).

Estos procesos de reflexión conducen, como ya se mencionó, a la generación de ideas, de conocimiento teórico y práctico. Este autor señala que los diarios de campo son utilizados para la investigación cualitativa; en realidad, se utilizan independientemente de la metodología seguida.

Por la existencia de archivos electrónicos, es importante reflexionar sobre el hecho de que las computadoras y otros medios de almacenamiento electrónico de datos e información no son infalibles. Existe la posibilidad de que se presenten errores de grabación de la información.

Por otra parte, está el riesgo de que el medio electrónico se pierda o sea robado. Todo esto implica una dolorosa pérdida de archivos de in-

formación. Por ello, es necesario pensar en medios alternos de almacenamiento de datos. Una posibilidad es el uso de varias unidades USB y del procesador central de la computadora; simultáneamente puede utilizarse el sistema de versiones de los archivos, controladas con los nombres de los mismos con la indicación de la fecha de creación y actualización del archivo. Si se perdiera la última versión (denominada versión n), puede recurrirse a la versión anterior (llamada versión $n-1$) y actualizarla. Por último, puede crearse un correo electrónico al que se manden periódicamente los archivos del proyecto de investigación o utilizarse alguna otra forma de almacenamiento en la red como, por ejemplo, “la nube”.

2.2. Diseño de un proyecto de investigación

Hay diversas formas de visualizar el diseño de las investigaciones. Principalmente en la investigación cualitativa, los investigadores lo consideran como la forma de concebir todo el proyecto y su desarrollo. Salamanca y Martín-Crespo (2007) señalan que, en la investigación cualitativa, “el diseño puede cambiar según se va desarrollando la investigación, el investigador va tomando decisiones en función de lo que ha descubierto”. No obstante, el diseño, ya sea cuantitativo o cualitativo, siempre soporta y es soportado por los sujetos o informantes y su relación con los instrumentos de levantamiento de datos o información y el levantamiento mismo.

En este inciso se consideran algunos diseños de investigaciones que se reportan como parte de la metodología y se configuran en función de la relación que existe entre los sujetos o informantes que participan en ella, los tratamientos y los instrumentos utilizados para el levantamiento de datos.

Es importante saber que uno de los diseños que permite dar más solidez a los resultados consiste en trabajar con dos grupos, con los integrantes de cada uno de ellos con las mismas condiciones y características de estudio. Los grupos se denominan experimental y control. El primero recibe un tratamiento especial del cual se desea saber si modifica los resultados obtenidos, el grupo control no lo recibe.

Para expresar los diseños, existe la notación de Campbell y Stanley (1963) y Cook y Campbell (1979), quienes definen que “X = Aplicación de la Variable Independiente; O = Observación o medida de la Variable Dependiente; A = Aleatorización, E = Equiparación; — = Grupo comparable, pero no-equivalente”. Esta notación ya no se ajusta completamente a los proyectos de investigación actual.

Para actualizar esta notación y la forma de denominar las características del objeto de estudio, se presenta la siguiente propuesta:

- G_{Ei} para indicar el grupo experimental
- G_{Ci} para indicar el grupo control
- G_{ir} para indicar un grupo de sujetos en cuestión
- S_{ir} para indicar un sujeto específico
- I_{ir} para indicar un informante específico
- D_{ir} para indicar que tiene lugar un levantamiento de datos
- T_{ir} para indicar que ocurre un tratamiento, con $i = 1, 2, 3, 4, \dots, n$

Para indicar el levantamiento de datos, podría usarse L_i . Se propone D_i para evitar la confusión que podría surgir con la letra L , interpretada en lingüística aplicada como lengua.

Esta simbología puede usarse en ambos tipos de investigación, cuantitativa y cualitativa, ya que contempla la posibilidad de que haya sujetos o informantes.

Comparativamente con la forma tradicional de la notación, ésta agrega S_i e I_{ir} para que incluya sujetos o informantes individuales y D_i y T_{ir} para que refleje que se habla de levantamiento de datos y de tratamiento. Para indicar el número de levantamiento de datos, denotado como D_{ir} , se asigna el valor a la i cronológicamente y respetando el orden G_E, G_C .

Alaminos y Castejón definen que “el diseño de investigación establece el procedimiento concreto de obtención de datos” (2006, p. 8) y clasifican los métodos de investigación en ciencias sociales como: método experimental, método selectivo o correlacional y método observacional, con las características presentadas en el siguiente esquema.

Esquema 2.1. Método de investigación en ciencias sociales

Manejo de variables o dimensiones	Método experimental	Método selectivo o correlacional	Método observacional
Grado de control interno			
Naturalidad			
Manipulación o nivel de intervención			
	Menor manipulación		

Fuente: esquematizado a partir de Alaminos y Castejón (2006).

En este esquema, las investigaciones de corte cuantitativo se ubican en el extremo izquierdo, el del método experimental. En el extremo derecho, el del método observacional, se encuentran las cualitativas. El método experimental utiliza un mayor control de los elementos observados, una menor naturalidad y una mayor manipulación. El observacional no controla los elementos, hay gran naturalidad y no hay manipulación o intervención.

2.2.1. Diseños en investigación cuantitativa

Los diseños de índole cuantitativa tienen su origen en el método científico, su estructura obedece en gran medida a este método. Alaminos y Castejón opinan que:

Una de las clasificaciones más citadas es la que realizan Campbell y Stanley (1963) y Cook y Campbell (1979). Estos autores diferencian tres grandes categorías de diseños, según el grado de validez interna que garantiza cada una de ellas, los diseños pre-experimentales, los diseños experimentales auténticos y los diseños cuasi-experimentales (2006, p. 16).

Igualmente, mencionan que “Bernia (1979) y Morales (1981) marcan que el diseño experimental establece cómo y en qué condiciones concretas se pone a prueba la hipótesis, de forma que se trate de asegurar la validez interna de los resultados” (Alaminos y Castejón, 2006, p. 16). En general, los estudios más comunes que se realizan en relación con un salón de clase de lengua son los cuasiexperimentales de grupos intactos. La gran diferencia entre los estudios experimentales y los cuasiexperimentales es la manera en la que se seleccionan los sujetos que participan en el levantamiento de datos. En los experimentales, es requisito que los sujetos sean seleccionados por algún método aleatorio; en los cuasiexperimentales, no es necesario. Por ejemplo, si se quisiera conocer el nivel de manejo de una lengua extranjera por alumnos de sexto semestre de una escuela preparatoria pública, en el diseño experimental se tomaría una muestra significativa (calculada de tal forma que los resultados sean generalizables a todos los alumnos de sexto semestre) mediante una selección al azar, como asignar un número a todos y tener una urna como la de la Lotería Nacional, con igual número de bolitas y estudiantes e ir extrayendo una por una hasta tener el tamaño planeado para la muestra. En cambio, para la investigación cuasiexperimental no habría una selección sujeto por sujeto; se dice que es de grupos intactos porque en el área de educación y, particularmente en un salón de clase de lengua, en la mayoría de las veces los estudios se realizan con grupos de alumnos

integrados por la institución educativa; el investigador acude al aula y levanta los datos con los que estén presentes.

Además de aseverar que los diseños van de la mano con los sujetos o informantes y su relación con los instrumentos y el levantamiento de datos, en las investigaciones cuantitativas es necesario hablar de la existencia de diseños que requieren la confrontación de datos de varios grupos que se componen por sujetos con las mismas características pero que no experimentan el mismo proceso educativo a través del desarrollo del proyecto de investigación. En un principio, como se mencionó, se hablaba de un grupo experimental, al que se le aplicaba un tratamiento en dicho desarrollo y un grupo control que no lo recibía. En la actualidad también se utilizan más de dos grupos. Por ejemplo: si se lleva a cabo un estudio para ver qué método de enseñanza propicia más el aprendizaje de una lengua extranjera, pueden tenerse tres grupos de sujetos, el primero con el uso del método comunicativo, el segundo con el método colaborativo y el tercero con un método tradicional. La notación en este caso sería:

- $G_{1,}$ para indicar el grupo experimental al que se le aplica el método comunicativo
- $G_{2,}$ para indicar el grupo experimental al que se le aplica el método colaborativo
- $G_{3,}$ para indicar el grupo control al que se le aplica el método tradicional

Algunos de los diseños cuantitativos utilizados en investigaciones relacionadas con el salón de clase de lengua son los siguientes; se indica la configuración del diseño y la notación para expresarla.

- Diseño preprueba, tratamiento, posprueba, con grupo experimental, un solo instrumento y tratamiento. El siguiente cuadro muestra su configuración.

Cuadro 2.1. Diseño preprueba, tratamiento, posprueba, con grupo experimental

Grupo	Levantamiento de datos de la preprueba	Tratamiento	Levantamiento de datos de la posprueba
G_E	D_1	T_1	D_2



Éste es un diseño en el que se trabaja solamente con un grupo denominado experimental. Se aplica un tratamiento (por ejemplo: la enseñanza de una lengua extranjera con el método comunicativo) para conocer si ejerce alguna influencia en el aprendizaje de la misma. Por ello, debe haber una medición de este aprendizaje antes y después del tratamiento, para conocer si hay un diferencial de aprendizaje entre las dos pruebas, antes y después del tratamiento.

- Diseño preprueba, tratamiento, posprueba, con grupo experimental y grupo control, un solo instrumento y tratamiento

Cuadro 2.2. Diseño preprueba, tratamiento, posprueba, grupos experimental y control con un instrumento y un tratamiento

Grupo	Levantamiento de datos de la preprueba	Tratamiento	Levantamiento de datos de la posprueba
G_E	D_1	T_1	D_3
G_C	D_2	-	D_4

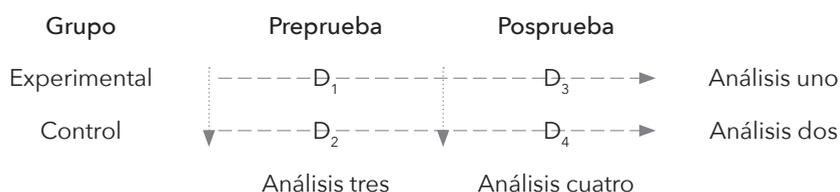
En la notación tradicional, se asignaba la numeración de las observaciones y tratamientos de acuerdo con los grupos, grupo por grupo. Como parte de la actualización de la notación, se asigna la numeración en función del momento de los levantamientos, para conservar el orden cronológico antes mencionado.

En este diseño se trabaja con dos grupos, experimental y control. Se aplica un tratamiento al grupo experimental (por ejemplo, la enseñanza de una lengua extranjera con el método comunicativo) para conocer si ejerce alguna influencia en el aprendizaje de la misma. Por ello, debe haber una medición de este aprendizaje antes y después del tratamiento. En este caso, se cuenta con cuatro diferenciales, uno para el grupo experimental y otro para el grupo control y dos entre el grupo experimental y el grupo control como un conjunto.

Hay cinco formas de analizar los resultados. El primer análisis es a través de la comparación de las dos pruebas del grupo experimental (D_1 y D_3), para conocer el cambio en ese grupo; el segundo análisis, la comparación de las dos pruebas del grupo control (D_2 y D_4), para conocer su cambio. Éstos son dos análisis diacrónicos. El tercero y el cuarto implican la comparación del grupo experimental con el grupo control, para conocer el diferencial entre los dos grupos; la comparación de las prepruebas de los dos grupos, para saber si parten del mismo nivel de aprendizaje

(D_1 y D_2), y la comparación de las pospruebas de los dos grupos (D_3 y D_4), para conocer si alcanzan el mismo nivel de aprendizaje, como lo muestra el siguiente esquema.

Esquema 2.2. Comparaciones con el uso de preprueba-posprueba y grupos experimental y control



En este caso, son dos levantamientos de datos para el grupo experimental, D_1 y D_3 , y dos para el grupo control, D_2 y D_4 . El tratamiento T_1 sólo es para el grupo experimental, el grupo control no recibe ningún tratamiento. Para determinar la similitud o diferencia entre los dos grupos, es importante considerar los puntos de partida y alcances de ambos grupos. Pudiera resultar que el grupo experimental aparentemente haya logrado un puntaje final más bajo que el del grupo control pero también puede ser que el grupo control haya partido de un nivel más alto que el del grupo experimental. De ahí la importancia de considerar los puntos de partida y los diferenciales logrados por ambos grupos. Esto constituye el quinto análisis.

- Diseño preprueba, tratamiento, posprueba, con grupo experimental, y grupo control, dos instrumentos y tratamiento.

Cuadro 2.3. Diseño preprueba, tratamiento, posprueba, grupos experimental y control, con dos instrumentos y un tratamiento

Grupo	Levantamiento de datos de la preprueba	Tratamiento	Levantamiento de datos de la posprueba
G_E	D_1 D_2	T_1	D_5 D_6
G_C	D_3 D_4	-	D_7 D_8

Este diseño trabaja con dos grupos, experimental y control. Se aplican un tratamiento al grupo experimental (por ejemplo: la enseñanza de una lengua extranjera con el método comunicativo) para conocer si ejerce alguna influencia en el aprendizaje de la misma y en la autonomía del estudiante.

La diferencia entre el caso anterior y este segundo diseño radica en que en este caso hay dos instrumentos aplicados antes y después del tratamiento; uno de ellos, una prueba de rendimiento en el aprendizaje de una lengua extranjera; el segundo, una prueba de autonomía del estudiante. Se generan dos conjuntos de cinco análisis, uno para cada instrumento y las posibles relaciones entre los datos levantados con cada instrumento.

En lo sucesivo, estos ejemplos pueden tomarse como base para la expresión de una gran variedad de diseños, mediante la combinación de grupos de sujetos, levantamientos de datos y tratamientos.

2.2.2. Diseños en investigación cualitativa

Salamanca y Martín-Crespo (2007) consideran que en la investigación cualitativa, el diseño cambia a través del desarrollo de la investigación. Esto no significa que carece de teoría o de estructura, "se necesita cierta flexibilidad para permitir que el diseño se adapte al fenómeno que se está estudiando, pudiendo ser necesario modificarlo una vez iniciada la investigación para obtener un conocimiento más profundo del sujeto/objeto de estudio". Resaltan como características del diseño de una investigación cualitativa que se amolda a lo que va descubriéndose, es holista (pretende conocer la totalidad del fenómeno), exige que el investigador permanezca en el campo durante períodos largos, el propio investigador es un instrumento de investigación, requiere un análisis de datos continuos y provoca la construcción de un modelo del comportamiento del fenómeno. Salgado también opina que en la investigación cualitativa "el diseño es más flexible y abierto [...] se va ajustando a las condiciones del escenario o ambiente" (2007, p. 72). Toma como diseños de la investigación cualitativa los establecidos por Hernández *et al.* (2006): teoría fundamentada, diseños etnográficos, diseños narrativos, diseños de investigación-acción y diseños fenomenológicos.

Salamanca y Martín-Crespo (2007) establecen, con base en Lincoln y Guba (1995), que hay tres fases generales para el desarrollo de una investigación cualitativa: orientación y panorama general, exploración concentrada y confirmación y cierre. Al respecto, Taylor y Bogdan (1990, citado en Salgado, 2007, p. 74) proponen tres momentos para un enfo-

que de análisis en progreso en investigación cualitativa: descubrimiento, codificación y relativización, dirigidos a la búsqueda del desarrollo de una comprensión profunda de los escenarios o personas que se estudian. En estas tres fases, destaca la segunda, destinada por ellos para desarrollar categorías de codificación y codificar los datos, separándolos de acuerdo con dichas categorías. Miles y Huberman (1994, p. 10) consideran tres etapas para el análisis: reducción de datos por selección, resumen o paráfrasis, asumidos en un gran modelo; presentación de datos o presentación concentrada, sistemática y poderosa y conclusiones/verificación, la extracción de significado de los datos.

También en la investigación cualitativa pueden bosquejarse diseños de investigación. Las diferencias en cuanto a levantamiento de datos, dependiendo del tipo de metodología empleada, están dadas por el instrumento o instrumentos que se utilizan, algunos de ellos indicados en el cuadro 2.5. En última instancia, los dos tipos de metodología se aplican para analizar información proporcionada por personas.

De cualquier forma, al hacer el reporte de investigación, siempre habrá un esquema del diseño seguido que puede presentarse; la notación propuesta lo permite.

2.3. Instrumento de levantamiento de datos

Algo muy delicado en las investigaciones es el levantamiento de datos. Alrededor de ese levantamiento giran muchos elementos metodológicos. A su vez, la filosofía que sustenta la metodología de una investigación marca el tipo de levantamiento de datos que se lleva a cabo. Es el levantamiento el que muestra una de las diferencias entre las investigaciones de carácter cuantitativo y cualitativo. Por ello, la planeación y el bosquejo y diseño del levantamiento de datos son cruciales en las investigaciones.

Salkind (2008, p. 298) establece recomendaciones para el proceso de levantamiento de datos, desde su planeación hasta la inclusión del análisis que tendrá que hacerse y de las implicaciones que se presentarán. Primero establece una relación entre el tipo de datos que van a obtenerse y la pregunta de investigación, no sólo recomienda buscar lo que se realizó antes, sino también algunas cosas de las que han hecho ellos. Recomienda que, mientras se piensa en el tipo de datos a levantar, se contemple el lugar en el que van a recopilarse. Sugiere que podrían usarse datos levantados con anterioridad pero la realidad es que, en general, estos datos son muy propios de la investigación en curso. También es necesario planear el lugar o lugares y las fechas para el levantamiento,

el tiempo que requerirá, así como reflexionar acerca de la viabilidad del acceso a los posibles sujetos que participarán en la investigación, teniendo presente su disponibilidad en cuanto a la demanda de ellos por otras investigaciones. No hay que delegar el levantamiento, a menos que sea a personas capacitadas. En resumen, el siguiente cuadro contiene algunas preguntas que hay que responder para conducir el levantamiento de datos en condiciones óptimas.

Cuadro 2.4. Preguntas para la preparación del levantamiento de datos

- ¿Existe una relación entre el tipo de datos que van a obtenerse y la pregunta de investigación?
- ¿Se realizó una búsqueda de los trabajos hechos en la misma línea que el que se está planeando?
- ¿Qué tipo de datos va a levantarse?
- ¿Es conveniente y posible hacer algunas cosas llevadas a cabo por otros anteriormente?
- ¿Existen datos levantados que son de utilidad para esta investigación?
- ¿En qué lugar se levantarán los datos?
- ¿En qué fecha o momento se levantarán los datos?
- ¿Cuánto tiempo será requerido?
- ¿Qué recursos materiales se necesitarán para el levantamiento de datos?
- ¿Qué recursos financieros se requerirán para el levantamiento de datos?
- ¿Habrá necesidad de la colaboración de otras personas en el levantamiento de datos?
- ¿Qué sujetos o informantes participarán en el levantamiento?
- ¿Se tendrá un acceso fácil para abordar a los sujetos o informantes de la investigación?
- ¿Se requiere algún trámite o permiso para levantar los datos?
- ¿Puede obtenerse una retroalimentación a través de los sujetos o informantes de la investigación?

Al pensar en las formas de obtener los datos nos encontramos precisamente con los instrumentos de levantamiento de ellos. Salkind (2008, p. 299) recomienda que es conveniente tener claridad acerca de la forma de usarlos y de la facilidad para ello.

El instrumento de levantamiento de datos es toda aquella herramienta que nos conduce precisamente al levantamiento de datos, tomado ya sea de sujetos para la investigación cuantitativa o de informantes para la investigación cualitativa. Uno de los requisitos de su construcción es el establecimiento de esquemas de enlace entre el marco teórico y la creación de los instrumentos. La verdadera riqueza del análisis de los datos es la discusión de los mismos frente a las teorías de los autores presentados en el marco teórico; ello conduce a que los instrumentos de levantamiento de datos deben construirse sentando sus bases en el marco teórico, de tal forma que pueda existir ese regreso.

La elección del instrumento o instrumentos dependerá de la metodología empleada para el desarrollo de todo el proyecto, de tal forma que los datos levantados y procesados den respuesta a las interrogantes de la investigación. En un proyecto de investigación pueden usarse varios instrumentos; en muchas ocasiones se combinan instrumentos de corte cuantitativo y cualitativo. El siguiente cuadro muestra algunos de los instrumentos tradicionales utilizados para el levantamiento de datos, separados de acuerdo con las metodologías cuantitativa y cualitativa.

Cuadro 2.5. Instrumentos tradicionales de levantamiento de datos

Tipo de investigación	Instrumentos
Cuantitativa	Cuestionarios de preguntas cerradas
	Escalas
	Escalas de Likert
	Diferencial semántico
Cualitativa	Cuestionarios con preguntas abiertas
	Entrevistas estructuradas y no estructuradas, personales, por teléfono o por medio electrónico
	Observación simple
	Observación participativa

En lingüística aplicada, se utilizan, además, otros instrumentos, tales como grupo focal, protocolos verbales (protocolo de pensamiento en voz alta), juicios de gramaticalidad, juicios de aceptabilidad, grados de dificultad y de discriminación de reactivos de opción múltiple, historias de vida, pruebas diseñadas específicamente para medir una dimensión del objeto de estudio.

2.4. Tratamiento

En ocasiones algunos denominan el tratamiento como un instrumento. Si bien, en un reporte de investigación el tratamiento se describe en el inciso de instrumentos, hay una diferencia entre ellos.

El instrumento sirve para medir u observar una entidad, variable o dimensión del objeto de estudio, para levantar datos.

El tratamiento es aquel procedimiento que se aplica a los sujetos con la finalidad de ver si produce cambios en ellos. Por ejemplo, si investigamos para conocer si el método comunicativo para la enseñanza de una lengua extranjera influye en su aprendizaje, el tratamiento se compone de las sesiones en las que se imparte la lengua extranjera con el método comunicativo.

2.5. Datos levantados

Los datos levantados son aquellos datos recopilados en una o varias etapas, mediante los instrumentos de levantamiento. Los datos simplemente organizados y procesados sin ninguna transformación, es decir, derivados de un conteo, para datos cuantitativos o de un proceso de frecuencias y categorización, sin ningún análisis o transformación, para datos cualitativos, se denominan datos fríos, en el sentido de que hasta ese momento no han sido agrupados, transformados o interpretados. Se dice que son datos que “no hablan”.

En una segunda etapa los datos son transformados para obtener porcentajes, para ser agrupados o para obtener puntajes parciales y totales de los sujetos participantes o para categorizarlos.

2.6. Procesamiento de los datos

Existen diferencias en cuanto a los datos provenientes de una investigación cuantitativa o cualitativa. Sin embargo, en la actualidad existe el procesamiento de datos para ambos tipos de investigación.

El procesamiento de datos consiste en capturarlos en un medio que haga posible su manejo, llevar a cabo operaciones numéricas, de agrupamiento o estadísticas que los transformen en datos analizables o en información. Si se trata de pocos datos éstos pueden procesarse manualmente; si son voluminosos se busca un medio electrónico. Si hay necesidad de llevar a cabo procesos estadísticos o análisis textuales, lo recomendable es hacerlo a través de un medio electrónico, independientemente del volumen manejado.

Salkind aporta las siguientes recomendaciones para el procesamiento de datos:

- Hacer una copia o copias de los datos
- Guardar en otra parte una copia de los datos levantados
- Hacer un respaldo, lo que ya se mencionó en el inciso sobre documentación

- Nunca descartar por completo los datos levantados o tirarlos, ya que, en el futuro, otros investigadores podrían necesitarlos (2008, pp. 299-300).

2.6.1. Datos cuantitativos

La característica más importante en el procesamiento de los datos cuantitativos es que se desarrolla en varias etapas, las cuales se muestran en el siguiente cuadro.

Cuadro 2.6. Etapas del procesamiento de datos cuantitativos

Captura o registro de los datos	<ul style="list-style-type: none"> • Obtención de datos fríos
Etapas previas al proceso de cálculos y transformaciones, para datos procesados estadísticamente	<ul style="list-style-type: none"> • Confiabilización de los datos levantados • Tipificación estadística de los mismos, su distribución en comparación con la curva normal • Estado de la media o medias y la varianza o varianzas
Etapas de procesamiento para análisis de resultados	<ul style="list-style-type: none"> • Cálculos y transformaciones de los datos • Producción de tablas, gráficas y reportes de información de los resultados del procesamiento

Para la obtención de datos fríos y el procesamiento en general, es recomendable el uso de un paquete estadístico. El uso del SPSS (*Statistical Package for Social Sciences*) es explicado en el capítulo ocho. La confiabilización de los datos levantados se lleva a cabo mediante la prueba denominada Alfa de Cronbach, la distribución con respecto a la curva normal, con la prueba diseñada con ese objetivo, así denominada y la prueba de Levene para homogeneidad de medias y varianzas. Estas pruebas se especifican en el capítulo cuatro. Los cálculos y transformaciones de los datos así como la producción de tablas, gráficas y reportes de información también pueden producirse con un paquete estadístico.

2.6.2. Datos cualitativos

En cuanto al procesamiento de los datos de una investigación cualitativa, anteriormente había un rechazo a hablar de ello, como si sólo correspondiera a la investigación cuantitativa. Chernobilsky (2007, p. 240) afirma

que ha habido un desarrollo paralelo entre la aceptación de la investigación cualitativa por investigadores de diversas disciplinas y una serie de herramientas informáticas que han intentado ayudar al investigador cualitativo para el procesamiento de sus datos. Éste se lleva a cabo no sólo a través de procesadores de texto o bases de datos sino también mediante programas de computadora específicos como el CAQDAS (*Computer Assisted Qualitative Data Analysis Software*). Esta autora resalta los programas “creados para asistir al investigador en el proceso de construcción de conceptos cualitativos, tipologías y desarrollos teóricos” (2007, p. 241) y puntualiza que, al haber una diversidad creciente, surge la necesidad de crear patrones comunes que armonicen e intercambien datos y resultados del análisis cualitativo, sin alterar las características propias de la investigación cualitativa (2007, pp. 241 y 242).

Atribuye a los programas una serie de capacidades limitadas que apoyan al investigador pero que no lo sustituyen; por ello sirven para integrar y estructurar los datos, para buscar textos, facilitar la codificación, recuperar segmentos codificados, consultar bases de datos, facilitar la escritura de textos pequeños y escribir el informe final pero no pueden elegir la forma de analizar los datos. Su importancia reside en que llevan a cabo funciones que no pueden efectuarse con un procesador de texto, como “vincular datos para formar redes de información, vincular categorías para formar redes semánticas que especifiquen relaciones jerárquicas o no, probar hipótesis a través de búsquedas complejas que recuperen datos para construir teoría fundamentada ni hacer diagramas de redes semánticas y desarrollos teóricos” (2007, pp. 244 y 245).

Tanto las entrevistas como las observaciones se preparan con una guía de entrevista o de observación. Pueden llevarse a cabo con grabaciones visuales o auditivas. Existe la posibilidad de que se requiera la transcripción de los datos levantados, para ello, hay procedimientos y notaciones establecidas.

Es importante saber que las transcripciones consumen un promedio de diez minutos por minuto grabado. Una posibilidad para estos casos es grabar y tomar notas, de tal forma que después se documenten las notas tomadas y se complementen con las grabaciones sin necesidad de transcribirlas.

2.6.3. Transformación de datos cualitativos en cuantitativos

Los datos procedentes de investigaciones de corte cualitativo también pueden procesarse mediante un paquete estadístico. Este apartado contiene algunos procesos para convertirlos en datos cuantitativos.

Rodríguez, en su artículo "Cómo analizar cuantitativamente datos cualitativos" (2003), presenta información muy completa acerca de varios procesos para efectuar estas transformaciones. Primeramente señala que hay un procedimiento interpretativo que "comprende: reducción de los datos, selección de palabras claves, agrupamientos de frases en dimensiones, edición de categorías exhaustivas y codificación de categorías". En segundo término, hay un "análisis que se transforma en una cuantificación de códigos numéricos, el recuento de códigos y la obtención de distribuciones de frecuencias; independientemente de la estructura y significación del contenido de las categorías".

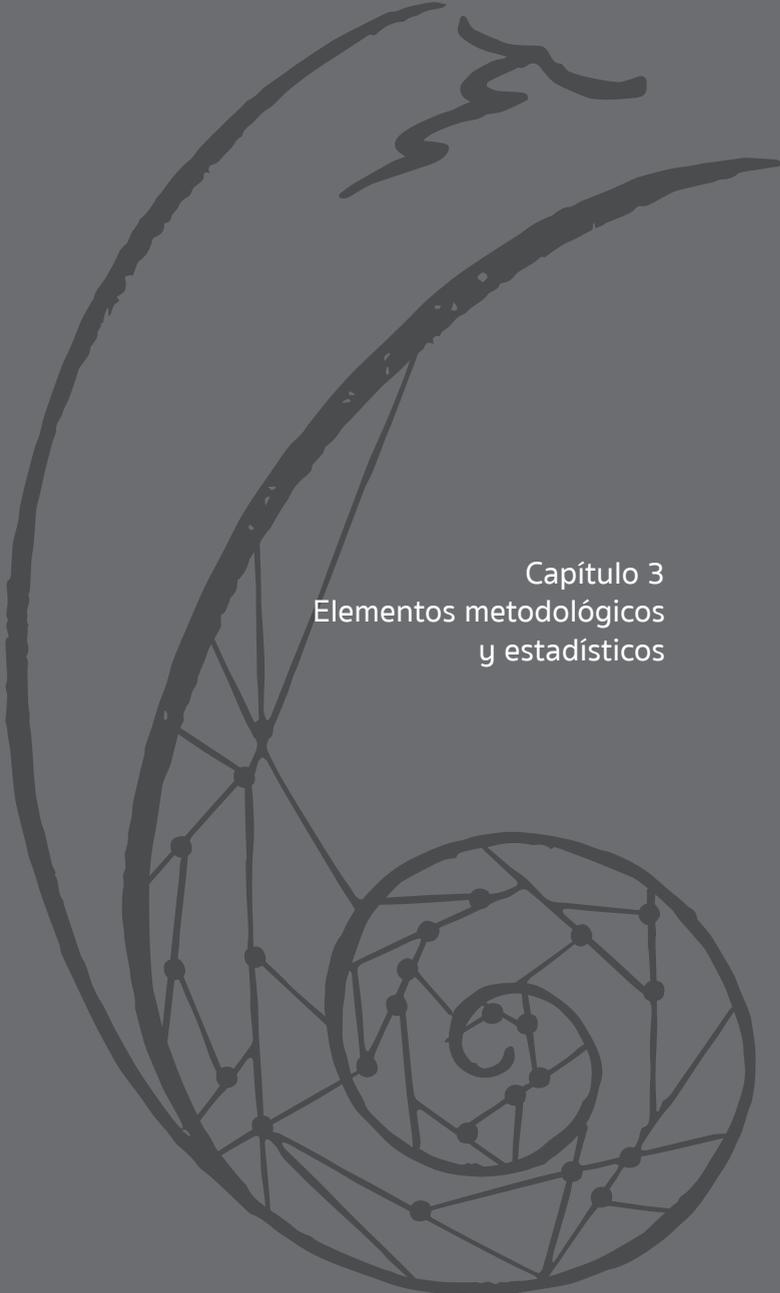
También afirma que "el procedimiento tradicional de la cuantificación de datos cualitativos es la categorización, la codificación y la tabulación", en donde "interesa más la frecuencia de los códigos que el propio contenido de las categorías". En cambio G. Rodríguez *et al.* (1999) sólo tratan la categorización y la codificación.

Rodríguez (2003) sugiere el uso de los enfoques lexicométricos o de la estadística textual apoyados en las técnicas estadísticas desarrolladas por la Escuela Francesa de Análisis de Datos (Benzécri, 1973). En resumen, el proceso consiste en un conteo de ocurrencias que puede someterse a un análisis estadístico. Este autor enfatiza que en tal caso hay un conteo directo sin necesidad de categorizar.

Para terminar el capítulo, hay que hacerse la gran pregunta, ¿cómo saber que en el fenómeno en estudio, la variable o dimensión analizada como un efecto se debe a la variable o a la dimensión específica bajo análisis como la causa y no a otras variables o dimensiones?

Una manera de saberlo sería, en primer término, utilizando el coeficiente de correlación y después el coeficiente de determinación, ambos se analizan en el inciso 4.2. El coeficiente de determinación indica el porcentaje de participación de una variable o dimensión en la otra o sobre sus dimensiones. Todos los fenómenos bajo estudio reciben la influencia de varias variables o dimensiones; cuando estudiamos una de ellas, simplemente reportamos su porcentaje de participación.

Los datos levantados en las investigaciones cuantitativa y cualitativa se someten al procesamiento de datos para el análisis de resultados. Anteriormente, se pensaba que los datos solamente se procesaban para la investigación cuantitativa. En la actualidad, los datos de la cualitativa son transformados para que puedan someterse, igualmente, al procesamiento de datos.



Capítulo 3
Elementos metodológicos
y estadísticos



En este capítulo se revisan separadamente los elementos metodológicos y los estadísticos. En los primeros, se consideran: entidades, hipótesis, manejo de puntajes y confiabilización de reactivos de un instrumento. Los temas de las entidades y de los puntajes se aplican a los dos tipos de metodología: cuantitativa y cualitativa. En teoría, la investigación cualitativa no plantea hipótesis. Sin embargo, el investigador siempre tiene expectativas de lo que puede encontrar. La confiabilización de los reactivos de un instrumento de levantamiento de datos es un proceso básicamente de tipo cuantitativo, lo que no impide que los instrumentos de tipo cualitativo puedan pilotarse para confiabilizarlos.

Los elementos estadísticos analizados para el procesamiento de datos son: tipos de estadística, a partir de dos criterios: descriptiva e inferencial, no paramétrica y paramétrica; medidas descriptivas; tipos de muestras y relaciones causa-efecto. Los tipos y medidas descriptivas se usan en ambos paradigmas de investigación. Los tipos de muestras y las relaciones causa-efecto se dirigen más hacia la investigación cuantitativa.

3.1. Elementos metodológicos

En el inciso de elementos metodológicos se analizan concepciones y clasificaciones de las variables, componentes, dimensiones o indicadores y la hipótesis.

El manejo de puntajes refiere que es mejor tener mediciones de componentes, además de mediciones generales.

La confiabilización de los reactivos del instrumento de levantamiento de datos es un proceso que se lleva a cabo con dos objetivos:

analizar que conducen al levantamiento de los mismos datos siempre que se aplican y verificar que son comprensibles por los sujetos en estudio, en cuanto a su construcción, con respecto a las instrucciones y al contenido de cada reactivo.

3.1.1. Variables y componentes o dimensiones, entidades

Para dar pie a la aplicación de diversas metodologías en el manejo de las entidades en estudio, es necesario tomar las consideraciones ya comentadas en el inciso 1.6. Se toma la posibilidad de asumir la denominación de variables, de componentes, o bien, de dimensiones e indicadores; en el entendido de que en las investigaciones hay un objeto de estudio, entidades que se analizan, de donde surgen precisamente esas variables, componentes, dimensiones o indicadores que permiten al investigador contar con unidades medibles u observables. Estas unidades medibles u observables son entidades que pueden asumir diversos valores. Por ejemplo, el nivel de manejo de una lengua extranjera puede asumir los valores: básico, preintermedio, intermedio y avanzado. Para usar una denominación genérica se habla de las entidades.

3.1.2. Tipos de entidades

Bajo criterios diferentes Villavicencio (2008) hace diversas distinciones entre tipos de entidades. Utiliza cinco, según la naturaleza de la entidad; según la posición del investigador; según la escala de medición; según la amplitud y según la abstracción. Enlaza la primera definición con las metodologías cualitativa y cuantitativa, señala que las entidades pueden mostrarse como categorías o atributos o bien en forma numérica; en el primer caso se denominan cualitativas, en el segundo, cuantitativas. Por ejemplo: el color de los ojos o cantidades, respectivamente.

Las entidades cuantitativas pueden ser discretas y continuas; la entidad discreta es aquella que “presenta separaciones o interrupciones en la escala de valores que puede tomar. Estas separaciones o interrupciones indican la ausencia de valores entre los distintos valores específicos que puede asumir. Un ejemplo es el número de hijos” (Villavicencio, 2008). La entidad continua puede asumir cualquier valor dentro del intervalo de valores establecido (Villavicencio, 2008). Larson-Hall (2010, p. 33) las clasifica en continuas y categóricas; las continuas, como el caso de una prueba en la que los sujetos obtienen diversos puntajes, sus valores son números que representan una medición y abarcan un rango de

respuestas elevado; por ejemplo: la calificación de un examen de comprensión oral. Las categóricas no tienen un valor numérico inherente, hay categorías en las que una de ellas no se toma como mayor que otra; en general su posible rango de valores es restringido.

Atendiendo a la posición del investigador, las entidades se clasifican en independientes, dependientes y extrañas. Según el enfoque metodológico se denominan como variables, dimensiones, indicadores o componentes; elementos sinónimos que dependen precisamente del enfoque. Dependiendo de la escala de medición de las entidades específicas pueden ser nominales, ordinales, de intervalo y de razón. Existen dos posibilidades en cuanto a su amplitud, las individuales y las colectivas. Finalmente, Villavicencio (2008) señala que, según la abstracción, pueden ser generales, intermedias y de indicación; sin embargo no las define.

En este trabajo se precisan solamente las entidades manejadas como variables bajo los criterios de la escala de medición (presentadas en general y en el programa estadístico como nivel de medición), la posición del investigador y consideraciones en la lingüística aplicada.

3.1.2.1. Entidades según el nivel de medición

En este inciso se usa el término variable o variables, que bien puede manejarse como entidad o variable. En un momento dado, cada investigador podría extrapolar al componente, la dimensión o el indicador que, al adquirir la posibilidad de ser medido u observado, se denomina como entidad.

El nivel de medición de las variables se establece con el criterio de cómo es la variable en sí misma, los valores que puede asumir y si hay algún orden o no.

Las variables, según el nivel de medición de los datos, se clasifican en: nominales, ordinales, de intervalo y de razón (Salkind, 2008, pp. 100 y 101). Es un criterio que se refiere a características de cada uno de estos tipos de variables, con base en Salkind (2008), se definen a continuación.

Variable nominal

Se refiere a aquella variable cuyas características pertenecen a una sola categoría. Esta variable puede asumir diversos valores que, por su naturaleza, son denominados como subcategorías. En general, el ejemplo más utilizado es la variable "sexo", cuyas subcategorías son: masculino y femenino; también es nominal el dato del grupo al que pertenece el sujeto en estudio o el nivel de estudio de una lengua extranjera.

Salkind precisa que “los niveles de medición nominales abarcan categorías mutuamente exclusivas, por ejemplo, la afiliación política no puede ser republicana y demócrata” (2008, p. 100). Un alumno no podría reportar como nivel máximo de estudio de una lengua extranjera básico e intermedio. Álvarez acota que las variables nominales asumen “valores que son palabras, nombres o categorías mutuamente excluyentes. También se puede asignar números a las categorías, pero no tienen un sentido matemático” (2008, p. 1).

Variable ordinal

Es aquella variable en la que destaca que la característica es medida en objetos ordenados (Salkind, 2008, p. 101). Por ejemplo: los lugares ocupados en un grupo de aprendizaje de una lengua extranjera, primero, segundo, tercero.

Variable de intervalo

La variable de intervalo puede asumir valores dentro de un rango numérico. Por ejemplo: los valores pueden oscilar entre 1 y 7, “los intervalos de la escala de medición son iguales entre sí” (Salkind, 2008, p. 100). Es el caso de una escala para medir actitudes hacia una lengua extranjera.

Variable de razón

La variable de razón es como la anterior con la diferencia de que se considera el cero como parte del rango de sus valores posibles. Es el caso de las calificaciones de una prueba de aprendizaje de una lengua extranjera en la que el alumno puede obtener una calificación de cero.

3.1.2.2. Entidades según la posición del investigador y la relación entre ellas

Variable independiente

Es aquella variable manejada por el investigador. Generalmente el interés del investigador se centra en el análisis del efecto de esta variable

sobre la variable dependiente; por ejemplo, un método de enseñanza de una lengua extranjera sobre el aprendizaje.

Larson-Hall (2010, p. 392) en ocasiones las denomina como variables explicativas, en los casos en los que intenta explicar las variaciones de la variable dependiente a través de varias variables independientes; cuando se investiga, por ejemplo, la relación entre varios métodos de enseñanza de una lengua extranjera, comunicativo y colaborativo y el aprendizaje de dicha lengua.

Variable dependiente

Es una variable que no es manipulada por el investigador. Su interés se dirige hacia la posibilidad de que la variable independiente ejerza una influencia en el comportamiento de ella.

Por ejemplo, el aprendizaje de una lengua extranjera, para el que se investigan variables independientes como el estilo de enseñanza del profesor o sus métodos de enseñanza.

Variable extraña

Es toda aquella circunstancia o suceso que se presenta en el momento en el que se lleva a cabo el levantamiento de los datos de la investigación. Por ejemplo, el caso en el que se está aplicando una prueba de rendimiento en lengua extranjera y hay un simulacro de temblor en el que los alumnos deben abandonar el salón de clase y regresar a él después de quince minutos o cuando hay una aplicación de un instrumento de medición y los alumnos tendrán examen en la siguiente hora de clase. Es importante prever y evitar este tipo de circunstancias que pueden contaminar el levantamiento.

Alaminos y Castejón (2006) se basan en Campbell y Stanley (1963) para resaltar algunas variables extrañas que pueden afectar la validez de los resultados: la historia, factores externos a los sujetos que surgen durante el levantamiento sin ser previstos; los cambios internos de los sujetos como el cansancio o la práctica; la aplicación de pruebas antes del levantamiento; la regresión estadística o lo que sucede cuando se aplica una segunda medición; la mortalidad experimental, cuando hay sujetos que abandonan el levantamiento; el uso de dos grupos de sujetos que no poseen las mismas características, entre otras. También señalan algunas formas para contrarrestar estas variables extrañas: eliminar estas variables, mantener las mismas condiciones para todos los sujetos,

balancear las condiciones si existe una variable que puede afectar el levantamiento, aleatorización y control estadístico.

Variable espúrea

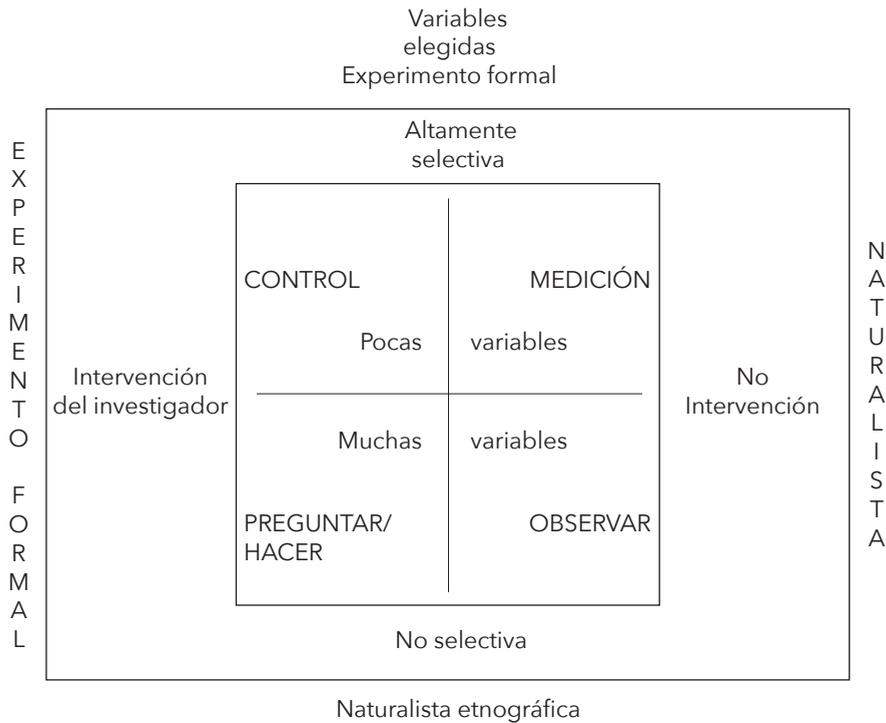
Este es un tipo de variable señalado por Hernández *et al.* (2006) de la que podemos concluir que más bien se trata de una recomendación en el sentido de ser muy cuidadoso para no establecer relaciones entre variables que no son factibles. Por ejemplo, la relación entre una característica física (el color de los ojos) y un resultado de tipo cognitivo, como el rendimiento en una lengua extranjera.

3.1.2.3. Entidades en lingüística aplicada

Para comprender la consideración de las variables en el área de la lingüística aplicada, David Nunan (1999) retoma los parámetros para el diseño de una investigación de Van Lier (1988) y comenta que él “argumenta que la investigación en lingüística aplicada puede analizarse en términos de dos parámetros: un parámetro intervencionista y un parámetro de selectividad” (Nunan, 1999, p. 5). El parámetro intervencionista se refiere a la forma en la que interviene el investigador en el medio ambiente y el de selectividad, de acuerdo con el grado de preespecificación del fenómeno que va a estudiar, es decir, con la selección de las variables. La combinación de los dos parámetros da lugar a cuatro espacios semánticos: control, medición, preguntar/hacer y observar.

La visualización de la relación entre estos dos parámetros se presenta en el siguiente esquema, tomado y adaptado de Van Lier (1988).

Esquema 3.1. Relación entre intervencionismo y selectividad en las investigaciones



Fuente: Van Lier (1988). Tomado y adaptado de Nunan (1999, p. 7).

Brown (2004, p. 482) reporta que en el área de control, el investigador conduce un experimento planeado cuidadosamente restringido en cuanto a los participantes y en relación con el enfoque del contenido. En el área de medición, la intervención es mínima pero el enfoque hacia los datos es muy condicionado; en el área de preguntar/hacer, el investigador puede intervenir para pedir a los sujetos que hablen de lo que están pensando. Por último, el área de observar implica que el investigador lo hace con un mínimo de intervención y de selectividad.

3.1.2.4. Entidades según los valores que pueden asumir

Una variable que sólo puede asumir dos valores es denominada dicotómica. Por ejemplo, el caso de las preguntas cuya respuesta puede ser solamente si o no. Este tipo de variable es común en la aplicación de exámenes, por ejemplo las llamadas "Yes/No questions". En la actualidad, se diseñan encuestas con preguntas cuyas respuestas son sólo dos.

Al existir encuestas grabadas por teléfono, para emitir la respuesta, el encuestado teclea solamente un 1 o un 2. Por ejemplo ¿Ha vivido en un país cuya lengua oficial es el inglés? Si sí, teclee uno; si no, teclee dos.

3.1.3. Hipótesis

En general, las hipótesis se definen en los proyectos de investigación cuantitativa. No se plantean abiertamente hipótesis en una investigación cualitativa; se señala que no abiertamente porque la realidad es que el investigador, al iniciar un proyecto, como se mencionó, siempre tiene una expectativa acerca de lo que puede encontrar.

En una forma tradicional, las hipótesis plantean relaciones entre variables; generalmente, entre las variables independiente y dependiente.

Salkind (2008, pp. 121-142) señala lo siguiente acerca de lo que es y significa establecer una hipótesis. Su papel es reflejar el planteamiento del problema general o lo que inicialmente motivó la pregunta de investigación. Existe entonces una relación entre la pregunta de investigación y la hipótesis.

la pregunta de investigación es una guía para la creación de la hipótesis y, en cambio, la hipótesis determinará las técnicas que se usarán para probarla y responder a la pregunta... la hipótesis traduce el planteamiento del problema o la pregunta de investigación a una forma más viable para probarla (Salkind, 2008, p. 121).

Las hipótesis son declarativas, afirmativas, establecen la relación entre las variables estudiadas. Salkind agrega que reflejan una teoría o marco teórico en los que se basan, son breves y puntuales y son probables, es decir, pueden someterse a prueba (2008, pp. 131 y 132).

En general, se habla de dos tipos de hipótesis, la hipótesis de investigación y la hipótesis nula; se entiende por hipótesis de investigación aquella que plantea las expectativas del investigador, su concepción acerca del comportamiento del objeto de estudio. Para que esta hipótesis tenga una existencia válida, necesariamente debe existir una hipótesis denominada la hipótesis nula, la cual plantea lo opuesto a la hipótesis de investigación; para que pueda haberla, la hipótesis nula da lugar a que el fenómeno se comporte en forma opuesta. La hipótesis de una investigación planteada como: *la aptitud para el aprendizaje de una lengua extranjera influye en su aprendizaje*, tiene como hipótesis nula: *la aptitud para el aprendizaje de una lengua extranjera no influye en su aprendizaje*. Salkind, al definir estos dos tipos de hipótesis, utiliza dos criterios: el de la relación de las variables y el de la diferencia o igualdad entre ellas (2008,

p. 123). El siguiente cuadro presenta las características de estas hipótesis, en función de los dos criterios mencionados.

Cuadro 3.1. Diferencias entre la hipótesis de investigación y la hipótesis nula

Hipótesis	Relación entre las variables	Igualdad o diferencia entre las variables
De investigación	Si la hay	No hay igualdad, Si hay diferencia
Nula	No la hay	Si hay igualdad, No hay diferencia

Fuente: elaborado a partir de Salkind (2008, p. 123).

Salkind (2008, pp. 126-127) precisa un poco más la hipótesis de investigación, dependiendo de la forma de plantear la naturaleza de la desigualdad entre las variables, clasificándola en direccionada y no direccionada. La direccionada establece que existe una diferencia entre las variables y el sentido o dirección, la no direccionada no establece magnitudes, simplemente declara que hay una relación o diferencia entre las variables. Por ejemplo:

Hipótesis direccionada

El promedio de las calificaciones de lengua inglesa de los estudiantes de una preparatoria privada es mayor que el de los estudiantes de una preparatoria pública.

Hipótesis no direccionada

El promedio de las calificaciones de lengua inglesa de los estudiantes de una preparatoria privada es diferente del de los estudiantes de una preparatoria pública.

En el manejo de las hipótesis, es necesario hacer la distinción entre el planteamiento de la hipótesis de investigación como tal y la hipótesis nula y el uso de la hipótesis nula cuando se utiliza una prueba estadística para determinar el comportamiento de los datos, desde el punto de vista estadístico. Cuando se habla de las pruebas estadísticas para este segundo caso, la hipótesis nula se establece puntualmente bajo el criterio de la igualdad de las variables en estudio. Por ejemplo:

Hipótesis de investigación

El promedio de las calificaciones de lengua inglesa de los estudiantes de una preparatoria privada es diferente del de los estudiantes de una preparatoria pública.

Hipótesis nula

No hay diferencia entre el promedio de las calificaciones de lengua inglesa de los estudiantes de una preparatoria privada y el de los estudiantes de una preparatoria pública.

Hipótesis nula para comportamiento de los datos

No existe diferencia entre las medias de los puntajes de la preprueba y la posprueba de rendimiento en lengua inglesa.

3.1.4. Puntajes totales y parciales

Cuando se analizan los datos, si la variable o entidad se considera solamente en general, como una sola dimensión, se habla de un puntaje total. El puntaje total es muy importante y puede marcar un perfil de la entidad analizada, pero es más rico el análisis si se consideran puntajes parciales, es decir, mediciones u observaciones de varias dimensiones o de varios componentes de la entidad. Por ejemplo: se hacen estudios de ansiedad frente a una lengua extranjera en los que el puntaje muestra un perfil de esta ansiedad. Si se desglosa en componentes, se tiene ansiedad frente a la producción oral, ansiedad frente a la producción escrita o frente al maestro de lengua extranjera, entre otras; la obtención de puntajes específicos por componente permite conocer mejor las dimensiones en las que se presenta.

3.1.5. Confiabilización de los reactivos de un instrumento de levantamiento de datos

La confiabilización de los reactivos de un instrumento de levantamiento de datos, como se mencionó, significa que se someten a un proceso para

garantizar que conducen al levantamiento de los mismos datos siempre que se aplican. Ésta es la característica que acredita la confiabilidad del instrumento. También se busca que su estructura y su contenido sean comprensibles para los sujetos bajo estudio.

Hay dos opciones en cuanto al uso de un instrumento de levantamiento de datos. Éste puede construirse o utilizarse uno ya hecho. En los dos casos hay que confiabilizarlo, a menos que el ya elaborado haya sido sometido previamente al proceso mencionado. La forma de hacerlo depende del tipo de instrumento en cuestión.

En general, este proceso se lleva a cabo mediante tres etapas: un diseño inicial, un piloteo y un diseño final. A continuación se presentan estas etapas.

3.1.5.1. Diseño inicial

El diseño inicial consiste, como se señaló, en construir el instrumento de levantamiento de datos deseado o bien en utilizar uno hecho con anterioridad. El tipo de instrumento dependerá de los objetivos de la investigación y del tipo de datos por levantar. En general, se pondrá énfasis en estos elementos.

3.1.5.2. Piloteo

La etapa de piloteo consiste en aplicar el instrumento a sujetos con las mismas características de los sujetos reales de la investigación. Es un proceso de simulación del levantamiento de datos. El objetivo es determinar su confiabilidad y si su contenido e instrucciones son accesibles para los sujetos.

Este proceso se realiza utilizando una hoja de registro en la que se anotan datos del mismo, datos de los sujetos participantes y, principalmente, los comentarios recibidos acerca del instrumento (ver anexo C). El objetivo es conocer las dificultades o facilidades para utilizarlo.

Después de aplicarlo, se harán pruebas con los datos de la aplicación. Se debe tener presente que en este caso lo más importante es que las pruebas no se dirigen hacia el contenido de las respuestas de los sujetos como descriptores de los mismos sino para probar que el instrumento es confiable.

El contenido de las respuestas sólo sirve para conocer si se obtiene la información deseada, a través de los reactivos. No existe interés en

conocer resultados acerca del comportamiento de los sujetos revelado en la prueba.

Para probar la confiabilidad de cada instrumento existen diversos procedimientos, de acuerdo con su construcción.

En los capítulos 6 y 7 se presentan dos instrumentos y se indica el procedimiento para confiabilizarlos.

Salkind (2008, p. 299) recomienda no comisionar a otros para el levantamiento de datos, a menos que se trate de personas que hayan sido capacitadas por los propios investigadores y que se tenga la certeza de que comprenden lo que significa el proceso de recolección de datos.

3.1.5.3. Diseño final

El diseño final del instrumento consiste en efectuar las modificaciones necesarias derivadas del análisis de la prueba piloto, para obtener un instrumento depurado.

Existen diversos tipos de instrumentos en el área de la lingüística aplicada. Entre ellos están: el uso de escalas para los casos de petición de la opinión de los sujetos o de medición de sus actitudes y cuestionarios de opción múltiple, para indagar el conocimiento y uso de elementos específicos de una lengua nacional o extranjera.

Estos dos casos se presentan en los capítulos seis y siete.

3.2. Elementos estadísticos

En este inciso se abordan los temas: tipos de estadística, medidas descriptivas, tipos de muestras en cuanto a procesamiento de datos y relaciones causa-efecto entre variables. En los tipos de estadística están las distinciones entre estadística descriptiva e inferencial y la subdivisión de ésta en estadística paramétrica y no paramétrica.

3.2.1. Tipos de estadística

El tipo de estadística que va a usarse depende de los objetivos y del diseño de la investigación así como de los sujetos o informantes y de las entidades manejadas.

3.2.1.1. Estadísticas descriptiva e inferencial

Una primera clasificación de la estadística establece dos tipos, la descriptiva y la inferencial. La estadística descriptiva es utilizada “para organizar y describir las características de un conjunto de datos” (Salkind, 2008, p. 8). Este mismo autor también establece que la estadística inferencial se utiliza para hacer inferencias a partir de un conjunto de datos pequeño o más grande. En esta opción:

se pueden probar relaciones entre variables, comparar grupos con respecto a cierta característica y hacer inferencias [...] la estadística inferencial generaliza a la población de la cual fue extraída, lo que significa que los resultados de una prueba estadística pueden generalizarse a un grupo más amplio de personas, que solamente a los que participaron en el experimento (Juárez *et al.*, 2002, p. 8).

En la estadística descriptiva, los datos se muestran como tablas de frecuencia y porcentajes, gráficas, medidas de tendencia central y medidas de dispersión (Juárez *et al.*, 2002, p. 5). En los incisos 3.2.2.2 y 3.2.2.3, se presentan estas medidas.

La obtención de los elementos de la estadística descriptiva en el SPSS se hace a través de los comandos:

Analizar, Estadísticos descriptivos, Descriptivos,
Opciones, Continuar, Aceptar

La estadística inferencial puede clasificarse en univariada y multivariada. La univariada contempla una variable dependiente y una independiente. Puede haber varias independientes pero sólo habrá una dependiente. La multivariada contempla dos o más variables dependientes, sin importar el número de variables independientes (Juárez *et al.*, 2002, p. 8).

En el caso de la estadística inferencial, Juárez *et al.* (2002, p. 5) señalan que hay dos formas para el trabajo con los datos, dependiendo de si la investigación es univariada o multivariada. En la univariada se calcula la correlación y se hace comparación a través de una prueba de hipótesis; en la multivariada hay métodos de agrupamiento y modelos de predicción.

3.2.1.2. Estadísticas no paramétrica y paramétrica

A su vez, la estadística inferencial se clasifica en dos tipos, la estadística paramétrica y la estadística no paramétrica. La estadística paramétrica se usa cuando los grupos de sujetos del trabajo de campo se eligen en forma aleatoria, los datos recabados para la variable dependiente se distribuyen como una curva normal (ver inciso 4.1.4), el nivel de medición es intervalar o de razón y hay homocedasticidad (ver inciso 4.1.5), es decir, homogeneidad de varianzas, la misma dispersión con respecto a la media de la variable dependiente (ver inciso 3.2.2.3). La estadística no paramétrica se utiliza cuando los datos recabados no se ajustan a los parámetros de distribución de datos como la curva normal y nivel de medición por intervalo o de razón (Juárez *et al.*, 2002, p. 10). Se basa en frecuencias, porcentajes, modas y rangos (ver incisos 3.2.2.2 y 3.2.2.3).

Para comprender las dos formas de contemplar la estadística, desde este criterio, es necesario conocer lo que significa la distribución de los datos recabados y los niveles de medición de los mismos. La distribución de los datos recabados es una representación gráfica de las frecuencias de presentación de cada valor asumido por la entidad en estudio. Los niveles de medición se analizan en el inciso 3.1.2.1.

3.2.2. Medidas descriptivas

Álvarez considera que “Las medidas descriptivas son valores numéricos calculados a partir de la muestra y que nos resumen la información contenida en ella” (s/f, p. 9). Considera cuatro tipos de medidas descriptivas, de posición, de centralización, de dispersión y de forma.

3.2.2.1. Medidas de posición

Las medidas de posición, según Álvarez, “dividen un conjunto ordenado de datos en grupos con la misma cantidad de individuos. Las divisiones se denominan cuantiles, percentiles, cuartiles y deciles (s/f, p. 9).

3.2.2.2. Medidas de tendencia central

Las medidas de tendencia central se refieren a la tendencia de los datos a presentarse en el sentido de una forma promedio. A continuación se describen las tres formas en las que se calculan.

Media

La media es un promedio, es el resultado de dividir el valor de la suma total de las respuestas emitidas para una variable entre el número de respuestas dadas. Por ejemplo, en una prueba de nivel de lengua extranjera, la media es el resultado de dividir la suma de las calificaciones obtenidas por los alumnos entre el número de alumnos; lo cual arrojará la calificación promedio obtenida por los alumnos participantes. Por ejemplo, el siguiente cuadro muestra las calificaciones de una prueba de nivel de manejo de lengua extranjera de diez alumnos.

Cuadro 3.2. Calificaciones de nivel de manejo de lengua extranjera de diez sujetos

Sujeto	Calificación
1	8.0
2	7.5
3	4.0
4	9.5
5	10.0
6	6.5
7	7.8
8	8.7
9	9.0
10	8.0
Total	79.0

La media, el promedio de las calificaciones obtenidas se calcula dividiendo 79.0 entre 10, lo que da un resultado de 7.9. Se acostumbra denotar la media como: \bar{X}

Mediana

La mediana es el punto medio entre los puntajes de los participantes y significa que los valores de éstos se ordenan, en forma ascendente, de tal

forma que pueda precisarse el puntaje tal con respecto al cual la mitad de los demás puntajes se encuentra por encima de este valor y la otra mitad por debajo de él. Por ejemplo, tomando las mismas calificaciones que en el ejemplo anterior, se ordenan en forma descendente: 10.0, 9.0, 8.7, 8.0, 8.0, 7.8, 6.4 y 4.0 (un total del 10 sujetos), el valor de la mediana sería el promedio resultante de dividir $8.0 + 8.0$ (16.0), entre dos; el valor de la mediana es de 8.0.

Moda

La moda es el valor del puntaje que ocurre con mayor frecuencia. Con base en lo anterior, el valor más frecuente es 8.0.

Salkind (2008, p. 29) indica que el uso de la media, la mediana o la moda concuerda con el tipo de datos descritos. Para los datos cualitativos, de categorías o nominales, se utiliza solamente la moda; la media y la mediana son mejores para los datos cuantitativos.

3.2.2.3. Medidas de dispersión

Las medidas de dispersión "indican la mayor o menor concentración de los datos con respecto a las medidas de centralización." (Álvarez, s/f, p. 9). Analizamos las tres medidas de dispersión establecidas por Salkind (1008, p. 35-47), consideradas por él como variabilidad. Éstas son rango, desviación estándar y varianza. El rango indica "qué tan separados están los datos entre sí. Se calcula restando el valor menor de los puntajes del valor más alto" (Salkind, 2008, p. 36). La desviación estándar "representa la distancia promedio con respecto a la media en un conjunto de puntajes" (Salkind, 2008, p. 37). Entre más alta es la desviación estándar, más diferencia hay entre los puntajes (Salkind, 2008, p. 42).

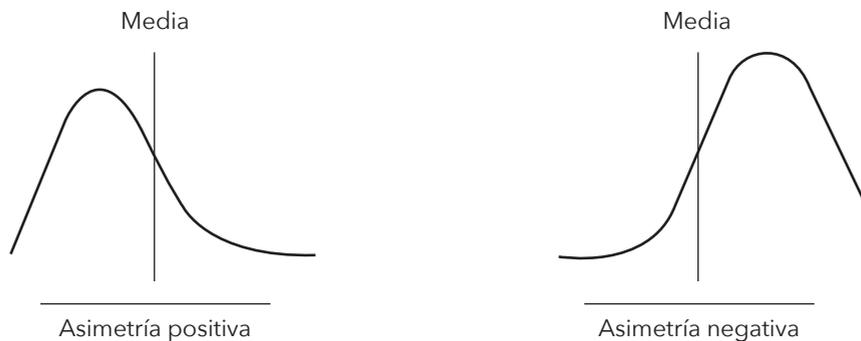
La varianza es el resultado de elevar al cuadrado la desviación estándar. Los valores de la varianza se utilizan para tomar decisiones en la elección de pruebas estadísticas (ver el inciso 4.1.5).

3.2.2.4. Medidas de forma

Las medidas de forma se refieren a la presentación de la curva de distribución de los datos. Son dos: la asimetría y el apuntamiento o curtosis. La asimetría implica si es simétrica o no con respecto a la media (Larson-Ha-

II, 2010, p. 69). La asimetría, según Salkind (2008, p. 58) significa que la curva que representa el comportamiento de los datos tiene una cola más larga que la otra. Así puede haber asimetría positiva, en cuyo caso la curva se desvía hacia la izquierda de la media o negativa, si se desvía hacia la derecha. Larson-Hall presenta lo siguiente.

Esquema 3.2. Distribuciones no-normales, asimétricas



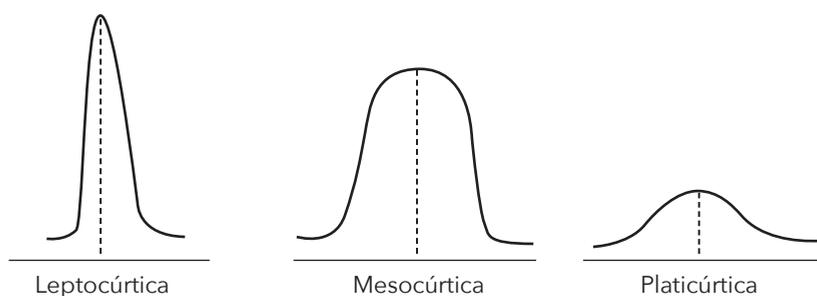
Fuente: Larson-Hall (2010, p. 79).

Para Salkind, “La curtosis revela si la curva es puntiaguda o achatada” (2008, p. 391). Refleja la concentración de los datos alrededor de la media. Según Suárez (2011a), “La curtosis mide el grado de agudeza o achatamiento de una distribución con relación a la distribución normal, es decir, mide cuán puntiaguda es una distribución”.

La curtosis determina el grado de concentración que presentan los valores en la región central de la distribución. La que tiene más concentración de puntajes en el centro es denominada curtosis delgada o “leptocúrtica”. La distribución de datos que es muy plana en la punta de la curva es denominada curtosis plana o “platicúrtica” (Larson-Hall, 2010, p. 80).

Suárez (2011a) considera además la mesocúrtica, donde los valores de las respuestas presentan una concentración normal. Tal idea se muestra a continuación.

Esquema 3.3. Las tres posibilidades de la curtosis



Fuente: Suárez (2011b).

Para Suárez (2011b), en la leptocúrtica existe una gran concentración, en la mesocúrtica, es una concentración normal y en la platicúrtica, la concentración es baja.

En cuanto al significado de la curtosis, Decarlo (1997, p. 302) señala que su importancia reside en que se utiliza para evaluar la normalidad de la distribución de datos, para describir el tipo y la desviación con respecto a la curva normal, para detectar valores atípicos, bimodalidad y mezclas, entre otras. El uso más importante es para determinar si la distribución de los datos se ajusta a la normal. Para Decarlo (1997, p. 297), el uso de la curtosis junto con la asimetría es básico para verificar la normalidad de los datos. Como se señaló, cuando la curva se ajusta a la distribución normal, se denomina mesocúrtica.

Recomiendan para su cálculo el uso de la prueba Shapiro-Wilk, la cual se encuentra en el SPSS.

Para interpretar los resultados de los cálculos de asimetría y curtosis, GraphPad (2014) recomienda lo siguiente.

Para la asimetría:

- Si la distribución es simétrica, la asimetría vale cero
- La asimetría con una larga cola hacia la derecha (los valores más altos), posee una asimetría positiva
- La asimetría con una larga cola hacia la izquierda (los valores más bajos), posee una asimetría negativa
- La asimetría no tiene unidades
- Si es mayor que 1 o menor que -1, la asimetría es relevante e indica que la distribución está muy lejos de ser simétrica

Para la curtosis:

- Una distribución Gaussiana (la normal) tiene una curtosis de cero
- La distribución plana tiene una curtosis negativa
- Una distribución más puntiaguda tiene una curtosis positiva
- La curtosis no tiene unidades
- El valor que reporta Prism¹ algunas veces es denominado **curtosis en exceso**, la curtosis esperada para una distribución Gausseana es de 0.0

Brown (2014) complementa.

- Si la asimetría es menor de -1 o mayor de 1, la distribución es completamente asimétrica
- Si su valor se encuentra entre -1 y -1/2 o entre +1/2 y +1, la distribución es moderadamente simétrica
- Si es entre -1/2 y +1/2, la distribución es aproximadamente simétrica
- Por ejemplo, con una asimetría de -0.1098, la muestra de datos es aproximadamente simétrica

El valor de la asimetría y la curtosis en el SPSS se obtiene con los comandos:

- Analizar, Estadísticos descriptivos, Frecuencias
- Se pasa la variable analizada al centro de la ventana, en el cuadro denominado:

Variabes:

- Estadísticos
- Se marca: Varianza, media, asimetría y curtosis, Continuar
- Gráficos
- Se marca: Histogramas, mostrar curva normal en el histograma, Continuar
- Aceptar

3.2.3. Tipos de muestras

En este inciso se establecen los tipos de muestras atendiendo al proceso de datos.

¹ Prism es un programa creado por GraphPad Software (2014). Esta compañía lo considera una poderosa combinación de bioestadística, ajustador de curva (regresión no lineal) y graficador científico.

Existen dos tipos principales de muestras, las muestras independientes y las muestras relacionadas. Las primeras se refieren a los casos en los que se efectúa una medición y hay dos grupos que son contrastados. Las segundas son cuando se hace más de una medición de los mismos sujetos.

3.2.4. Relaciones causa-efecto entre las variables

En la investigación, principalmente para el caso de la investigación cuantitativa, hay una forma de clasificar los niveles de desarrollo de la misma. Esta forma considera los niveles de investigación: exploratoria, descriptiva, correlacional y explicativa.

La explicación de estos niveles de investigación puede partir desde el punto de vista de las variables de la investigación. Si la investigación es exploratoria, sirve para descubrir las variables; si es descriptiva, las describe.

En cuanto al nivel de correlación, en la actualidad algunos autores lo omiten, considerando solamente tres. Este es un paso en investigación que no puede descartarse; para poder continuar hasta el nivel explicativo, en el que se establecen relaciones causa-efecto para explicar la participación de la variable independiente en el comportamiento de la variable dependiente, primero es necesario conocer si existe una relación entre las variables. Si el índice de correlación entre ellas indicara que no hay relación de variables, ya no se buscaría una relación causa-efecto.

Hay varios diseños de investigación, varias formas de procesamiento para establecer relaciones causa-efecto entre las variables.

Se analizan dos, para establecer la relación entre dos variables, una independiente y una dependiente y para conocer relaciones multivariadas, entre varias variables independientes y una o más dependientes. A continuación se encuentran los dos casos: relación entre una variable independiente y una variable dependiente; y relaciones entre varias variables independientes y una o más dependientes.

3.2.4.1. Relación entre una variable independiente y una variable dependiente

Un ejemplo es cuando se busca si existe una relación entre la aptitud para una lengua extranjera y el rendimiento en esa lengua. La variable independiente es la aptitud y la dependiente es el rendimiento en la lengua extranjera.

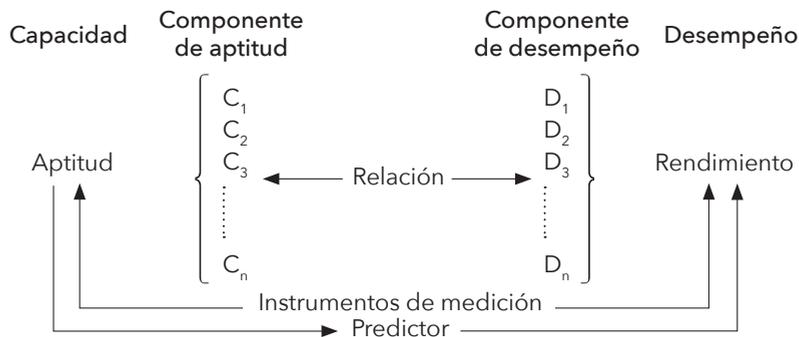
El proceso estadístico consiste en analizar la relación entre las dos variables, para conocer el grado de participación que tiene una variable independiente (la aptitud) en la ocurrencia de la variable dependiente (el rendimiento). Se utilizan los coeficientes de correlación y de determinación (ver incisos 4.2.2 y 4.2.3).

3.2.4.2. Relaciones entre varias variables independientes y una o más dependientes

Este segundo caso se refiere al hecho de descomponer las variables estudiadas en sus componentes; la aptitud, en hasta n componentes, denotados con C_i y el rendimiento a través de componentes de desempeño, denotados con D_i . El esquema 3.4 muestra la relación planteada para las variables.

En este caso, para conocer las relaciones multivariadas, entre varias variables independientes y dependientes, también se utilizan los coeficientes de correlación y de determinación, mediante el uso de la Regresión múltiple (ver inciso 4.2.5.), con la que se conocen los componentes de la aptitud que determinan los componentes del rendimiento.

Esquema 3.4. Componentes de la aptitud, forma de medirlos y su relación con los componentes del rendimiento



Fuente: Uribe (2010, p. 101).

En este capítulo se enfatizó la importancia de considerar las características de los objetos de estudio, denominadas de formas diversas dependiendo del enfoque del investigador, desde el planteamiento del análisis componencial; análisis que conduce a resultados más específicos que los de una sola medición general.

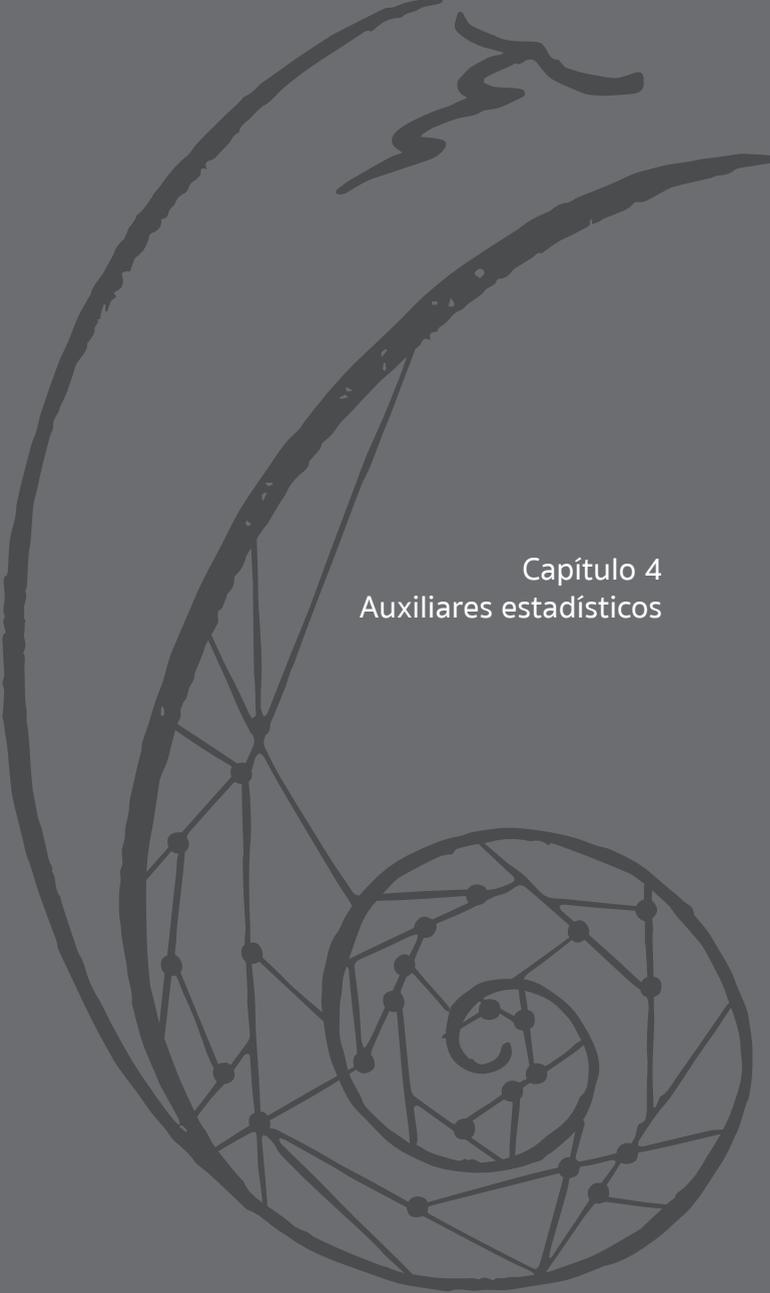
Se trató la forma de confiabilizar los reactivos de un instrumento, para garantizar el levantamiento de los mismos datos siempre que se

aplican, es decir, que proporcionan las mismas mediciones y que su estructura y contenido es comprensible por los sujetos en estudio.

Se revisaron los elementos estadísticos propios de la estadística descriptiva.

Finalmente, se revisaron los tipos de muestra de acuerdo con el procesamiento de datos y relaciones entre las variables independiente y dependiente.





Capítulo 4
Auxiliares estadísticos



El contenido del capítulo se dirige básicamente a proyectos de investigación de corte cuantitativo. El último inciso, sobre gráficas, es viable para ambos paradigmas. Presenta elementos estadísticos aquí denominados como auxiliares porque apoyan con parámetros para tomar decisiones acerca de la confiabilidad de los datos y la posibilidad de seguir adelante con la prueba estadística seleccionada o bien como base para consultar tablas de valores que son indicadores para conocer los resultados. Se divide en cinco incisos: parámetros básicos; coeficientes; toma de decisiones; confiabilización de datos levantados y capturados y gráficas.

Es importante resaltar que los capítulos tres y cuatro incluyen el tema de la confiabilización aplicada a dos análisis diferentes. En el capítulo tres se habla de la confiabilización de los reactivos incluidos en los instrumentos de levantamiento de datos, en cuanto a su precisión en la medición cada vez que se aplican, su construcción y su contenido; en cambio, el capítulo cuatro se refiere al nivel de confiabilidad de los datos levantados.

4.1. Parámetros básicos

Los parámetros básicos son índices y resultados de pequeñas pruebas para indicar la aceptación o rechazo del planteamiento para el análisis del resultado de un procesamiento de datos. Informan las bases para continuar el estudio, cuando los resultados lo permiten. Se presentan cinco parámetros: nivel de significancia, grados de libertad, prueba de normalidad de la curva, prueba de Levene y tabla de valores de prueba.

4.1.1. Nivel de significancia

El nivel de significancia, también denominado nivel de significación, es:

un número que expresa la probabilidad de que el resultado de un experimento o estudio haya ocurrido meramente por azar. Este número puede ser el margen de error (“Los resultados de esta encuesta de opinión pública son precisos al cinco por ciento”), o puede indicar un nivel de confianza (“si este experimento fuera repetido, hay una probabilidad de noventa y cinco por ciento de que las conclusiones se sostendrán”) (*The American Heritage New Dictionary of Cultural Literacy*, 2005).

Los niveles de significancia acostumbrados en el medio del manejo de datos son .01 y .05. Estos valores indican que la probabilidad de obtener una conclusión errónea es de .01 o de .05. Esto significa que los resultados son más confiables a la luz del nivel .01; sin embargo, comúnmente se acepta el nivel de .05.

Para hablar del nivel de significancia, hay que hablar paralelamente de las pruebas estadísticas. En el inciso 3.1.4, sobre las hipótesis, se señaló que es necesario hacer la distinción entre el planteamiento de la hipótesis de investigación como tal y la hipótesis nula y el uso de la hipótesis nula cuando se utiliza una prueba para determinar el comportamiento de los datos, desde el punto de vista estadístico.

Al llevar a cabo pruebas estadísticas, el nivel de significancia es un indicador que debe tomarse en cuenta para seguir adelante o no con otras pruebas. Para una hipótesis de investigación, si este nivel reporta un valor mayor a .05, significa que no es posible establecer conclusiones confiables o continuar con otras pruebas estadísticas. Si es el caso de una hipótesis nula, por el contrario, ese valor indica que ésta es aceptada.

Para el primer caso, cuando se habla de las pruebas estadísticas, el nivel de significancia indica la probabilidad de equivocarse en la interpretación de los resultados con respecto a la hipótesis de investigación manejada, hipótesis que establece la existencia de una relación entre las variables.

En el segundo caso, la hipótesis nula se establece puntualmente bajo el criterio de la igualdad de las variables en estudio. El nivel de significancia, igualmente, es un indicador de la probabilidad de equivocarse al determinar que las variables en estudio son iguales.

En ambos casos, cuando se acepta o se rechaza una hipótesis, el nivel de significancia indica la probabilidad de equivocarse al obtener conclusiones. Como consecuencia, en el análisis de los datos, pueden cometerse dos tipos de errores, denominados errores Tipo I y Tipo II. El error Tipo I consiste en rechazar una hipótesis nula que señala que las

variables son iguales cuando sí lo son. El error Tipo II significa aceptarla cuando no lo son. Igualmente, para la hipótesis de investigación, el error Tipo I se comete cuando se afirma que no hay relación entre las variables, cuando si la hay y el error Tipo II cuando se acepta que hay una relación entre las variables y en realidad no la hay.

4.1.2. Intervalo de confianza

Rada (2007) establece que:

en el contexto de estimar un parámetro poblacional, un intervalo de confianza es un rango de valores (calculado en una muestra) en el cual se encuentra el verdadero valor del parámetro, con una probabilidad determinada. La probabilidad de que el verdadero valor del parámetro se encuentre en el intervalo construido se denomina **nivel de confianza**, y se denota $1-\alpha$. La probabilidad de equivocarnos se llama **nivel de significancia** y se simboliza α . Generalmente se construyen intervalos con confianza $1-\alpha=95\%$ (o significancia $\alpha=5\%$). Menos frecuentes son los intervalos con $\alpha=10\%$ o $\alpha=1\%$.

También afirma que: “Los intervalos de confianza permiten verificar hipótesis planteadas respecto a parámetros poblacionales”.

4.1.3. Grados de libertad

Los grados de libertad constituyen una cifra que forma parte de las decisiones tomadas al llevar a cabo pruebas estadísticas. Significa “la libertad de variaciones que puede tener una variable” se calcula teniendo el tamaño de la muestra menos uno ($gl = n-1$)” (Juárez *et al.*, 2002, p. 15); es decir, el número de sujetos en estudio menos uno.

En esta obra se manejan las pruebas T y F. La primera maneja los grados de libertad mencionados en el párrafo anterior; es decir, para n , número de sujetos o informantes, los grados de libertad son $n-1$.

El caso de la prueba F difiere porque “Existe una distribución F diferente para cada combinación de tamaño de muestra y número de muestras” (Suárez, 2012a).

El tamaño de la muestra o el número de sujetos o informantes se denomina como n y el número de muestras tomadas a cada uno de ellos se designa como k . Este autor señala que su característica es que “A diferencia de otras pruebas de medias que se basan en la diferencia existente entre dos valores, el análisis de varianza emplea la razón de las estimaciones, dividiendo la estimación intermediente [sic] entre la estimación interna” (Suárez,

2012a), estimaciones de la varianza. Por ello, se manejan dos grados de libertad simultáneamente, los grados de libertad del numerador de la razón F definidos como $k-1$ (o grados de libertad expresados en la tabla de F como la primera fila) y los grados de libertad para el denominador, calculados con $k(n-1)$ (o grados de libertad expresados en la tabla de F como la primera columna) (Suárez, 2012a).

Por ejemplo, si se tuvieran 7 sujetos con tres observaciones, el valor del numerador sería de $3-1=2$ y el del denominador de $3(7-1)$ o bien $3(6)$. Los grados de libertad para esta prueba serían: grados para la primera fila de la tabla, 2 y grados para la primera columna de la tabla, 18.

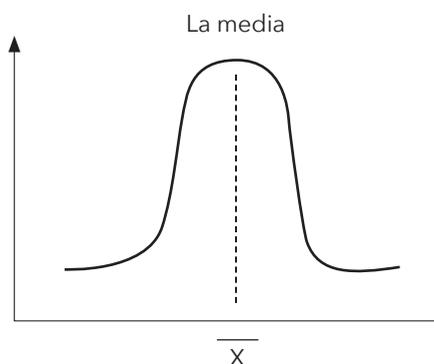
En general, se utilizan para indicar el valor en una tabla que corresponde al cálculo particular (ver inciso 4.1.6).

4.1.4. Prueba de normalidad de la curva

Para lograr una mejor comprensión del comportamiento de los datos recabados, puede hacerse una gráfica; ésta constituye una representación visual de los mismos.

Una de las gráficas más representativas del comportamiento de los datos y de las mediciones en general es la llamada curva normal, que tiene la siguiente forma:

Gráfica 4.1. La curva normal



Salkind (2008, pp. 135-136) señala que también es denominada como una curva en forma de campana, posee tres características, una distribución en la que la media, la mediana y la moda tienen el mismo valor, es una curva simétrica con respecto a la media y los extremos de la curva corresponden

a una curva asintótica, es decir que, en ambos extremos de la curva se acerca al eje horizontal pero nunca lo toca.

Álvarez (s/f, p. 11) agrega que la curva normal tiene las siguientes propiedades. El área total bajo la curva es igual a 1, su distribución tiene forma de montículo y es simétrica, tiene una media de cero y una desviación estándar de 1. La media divide el área a la mitad.

Para conocer si la distribución de los datos posee las características de una curva normal, se utiliza la prueba de Kolmogorov-Smirnov, una prueba no paramétrica. En ella, la hipótesis nula plantea que la variable en estudio sigue una distribución normal. En tal caso, el valor de la significancia debe ser mayor a .05, para poder aceptar la hipótesis nula. En el SPSS, las instrucciones usadas son:

- Analizar, estadística no paramétrica, cuadro de diálogos antiguos, K-S de una muestra
- Se pasa la variable en estudio a la ventana central
- En el cuadro de distribución de contraste, se elige la normal
- Opciones
- Descriptivos, Continuar
- Aceptar

Los resultados son los siguientes:

Cuadro 4.1. Resultados 1. Prueba de normalidad. Datos descriptivos de la variable
Pruebas no paramétricas
[Conjunto_de_datos1] F:\archivo para libro ejemplos.sav

Estadísticos descriptivos					
	N	Media	Desviación típica (estándar)	Mínimo	Máximo
Nivel de lengua	36	86.19	6.391	69	95

Los elementos descriptivos de la distribución de la variable son: la media y la desviación estándar y los valores mínimo y máximo alcanzados.

Cuadro 4.2. Resultados 2. Prueba de normalidad. Resultado de normalidad Prueba de Kolmogorov-Smirnov para una muestra

		Nivel lengua total
N		36
Parámetros normales ^{a,b}	Media	86.19
	Desviación típica	6.391
	Absoluta	.111
Diferencias más extremas	Positiva	.084
	Negativa	-.111
Z de Kolmogorov-Smirnov		.667
Significancia asintótica. (bilateral)		.765

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

El nivel de significancia es de .765, mayor de .05, por lo tanto se acepta la hipótesis nula de que la variable sigue una distribución normal. La tabla de valores para esta prueba se encuentra en el anexo D.

4.1.5. Prueba de Levene

La prueba de Levene es una prueba para verificar la homogeneidad de varianzas, también denominada homocedasticidad; es una prueba que se realiza antes de intentar la comprobación de la existencia de diferencias entre medias. Dependiendo del valor obtenido mediante esta prueba se elegirá el procedimiento para buscar las diferencias.

Por ejemplo, cuando se utiliza una prueba como la prueba t que se encuentra en el inciso 5.1, se solicita con:

- Analizar, Comparar medias, prueba t para muestras independientes
- Variables para contrastar, Variable de agrupación, definir los valores, Continuar
- Aceptar

En el programa estadístico SPSS, el resultado de la solicitud de comparación de medias para una prueba t con muestras independientes presenta dos cuadros. A continuación se presenta el primero.

Cuadro 4.3. Resultados 1. Datos de las variables de la comparación de dos muestras independientes con la prueba t

Estadísticos de grupo					
	Grupo	N	Media	Desviación típica (estándar)	Error típico de la media
Variable en estudio	Grupo control	17	17.41	1.543	.374
	Grupo experimental	19	17.68	1.797	.412

En este cuadro solamente se muestran las medias, las desviaciones estándar y el error típico de la media. Los datos más importantes son las medias y las desviaciones estándar. A continuación se encuentra el contenido parcial del segundo cuadro, el cual muestra los resultados de la prueba de Levene, (cabe destacar que el contenido relacionado con la prueba t no se presenta, pero puede verse en el inciso 5.1). En este caso, el interés se centra en los resultados de la prueba de Levene, donde la hipótesis nula supone que las varianzas son homogéneas.

Cuadro 4.4. Resultados 2. Prueba de Levene de la comparación de dos muestras independientes con la prueba t

Prueba de Levene para la igualdad de varianzas			
Nombre de la variable en estudio	Supuesto con respecto a las varianzas entre los dos grupos	Valor F	Significancia
Variable en estudio	Se han asumido varianzas iguales	1.208	.35
	No se han asumido varianzas iguales		

Como se mencionó, la hipótesis nula de esta prueba es que no existen diferencias entre las varianzas de los grupos comparados (es decir, que son iguales). Esta hipótesis se cumple cuando el valor F se acerca a uno y, principalmente, cuando el valor de la significancia es mayor a .05. Por lo tanto, el resultado más importante mostrado en este cuadro es el nivel de significancia, en este caso, de .35. Este nivel muestra que sí puede asumirse que las varianzas de los dos grupos son iguales.

4.1.6. Tabla de valores de prueba

Cuando se calculan cifras con base en pruebas estadísticas se utiliza una tabla de valores de donde se extrae el que corresponde a la investigación particular, conocido como valor de tabla. Este valor se encuentra utilizando los grados de libertad (vistos en el inciso 4.1.3) para poder comparar el valor de la tabla con el obtenido en el proceso de los datos, denominado valor calculado. Para aceptar la hipótesis establecida para la prueba, el valor calculado debe ser mayor al valor de la tabla. La tabla proporciona los valores para .05 y .01 niveles de significancia. Existen tablas para otros valores del nivel de significancia. En los anexos E y F se encuentran los valores que corresponden a la prueba t y a la prueba F, para .05 y .01.

4.2. Coeficientes

En este inciso se presentan tres coeficientes, Alfa de Cronbach, Correlación y Determinación; y dos formas de plantear relaciones entre variables, la regresión lineal y la regresión múltiple. Estos elementos permiten tomar decisiones en el procesamiento de datos. Se define un coeficiente como “expresión del grado o intensidad de una propiedad o característica” (WordReference, 2014). El valor de estos coeficientes puede variar dentro de un rango determinado (-1, +1); este valor es precisamente un indicador para la toma de decisiones.

4.2.1. Alfa de Cronbach

Salkind establece que “El Alfa de Cronbach es un coeficiente que indica la confiabilidad de la consistencia interna entre los reactivos de una prueba, en el sentido de que representan una y sólo una dimensión, constructo o área de interés” (2008, p. 105), agrega que este coeficiente muestra si el puntaje de cada reactivo se correlaciona con el puntaje total y que la prueba será consistente en todos los que la tomen. Se utiliza para confiabilizar el instrumento de levantamiento de datos o bien los datos levantados en una investigación. El Alfa de Cronbach es un coeficiente que puede asumir valores entre cero y uno. Su importancia será mayor conforme se acerque más al valor de uno. Un valor de .80 es muy aceptable. Para obtener el valor de Alfa de Cronbach con el SPSS se utilizan los siguientes comandos.

- Analizar, escala, análisis de fiabilidad
- Modelo Alfa, Pasar las variables al centro, escribir etiqueta para el nombre de la escala, Indicar modelo: Alfa
- Aceptar

Como resultado se obtienen dos cuadros. El primero contiene indicaciones de los casos manejados. El segundo es un cálculo de Alfa. A continuación se presentan los resultados del cálculo de Alfa para un caso en el que se consideraron seis totales de subescalas y un total de una escala de estrategias de aprendizaje de lengua, lo que da un total de siete elementos. Este cuadro muestra el total de los sujetos encuestados.

Cuadro 4.5. Resultados 1-1. Cálculo de Alfa de Cronbach

Escala: Estrategias de aprendizaje de lengua
Resumen del procesamiento de los casos

		N	%
Casos	Válidos	154	100.0
	Excluidos ^a	0	.0
	Total	154	100.0

a. Eliminación por lista basada en todas las variables del procedimiento.

El siguiente cuadro muestra el cálculo de Alfa de Cronbach para la consistencia interna de los puntajes totales de las subescalas y la escala completa. El Alfa de .773 se aproxima a .80, lo que indica que es un valor muy aceptable.

Cuadro 4.6. Resultados 1-2. Cálculo de Alfa de Cronbach

Estadísticos de fiabilidad

Alfa de Cronbach	N de elementos
.773	7

4.2.2. Correlación

El coeficiente de correlación es un indicador de la relación que existe o no entre variables. Este coeficiente asume valores en el rango (-1, 1). Cuando la correlación es positiva, significa que las variables correlacionadas varían en el mismo sentido; si es negativa, implica que varían en sentidos opuestos, si una aumenta, la otra disminuye y viceversa. Hay varios coeficientes de correlación. En el paquete estadístico, que se analiza en el capítulo ocho, se

maneja la correlación de Pearson, correlación Kendall's tau-b y correlación Spearman. Los más utilizados son Pearson y Spearman. Cuando los datos se ajustan a los parámetros de la estadística paramétrica, se utiliza la correlación de Pearson; si se ajustan a los de la no paramétrica, se usa la de Spearman, es decir, con una distribución de datos diferente a la curva normal y con un nivel de medición al menos ordinal. Larson-Hall opina que se utiliza para "probar la posibilidad de una relación entre variables de intervalo-escala" (2010, p. 401). Salkind considera la siguiente tabla para interpretarla.

Cuadro 4.7. Interpretación del valor del coeficiente de correlación

Correlación	Interpretación
.8 - 1.0	Relación muy fuerte entre las variables
.6 - .8	Relación fuerte
.4 - .6	Relación moderada
.2 - .4	Relación pobre o débil
.0 - .2	Relación débil o sin relación

Fuente: Salkind (2008, p. 85).

Vila *et al.* (2014) proponen el siguiente esquema para interpretar el coeficiente de correlación entre dos variables.

Esquema 4.1. Interpretación del coeficiente de correlación entre dos variables



Fuente: Vila *et al.* (2014).

La forma de calcular este coeficiente con el SPSS es como sigue: Analizar, Correlaciones, Bivariadas.

En la pantalla que aparece, se toman las dos variables de la lista de variables de la ventana del lado izquierdo y se pasan a la ventana de variables.

- Indicar el tipo de correlación que se desea
- Indicar si la curva de distribución de los datos es de dos o de una cola
- Marcar para que indique las correlaciones significativas
- Aceptar

En el siguiente cuadro se presentan los resultados de la correlación entre el puntaje total del rendimiento en lengua inglesa (denominado Puntaje

general) y el puntaje total de la aptitud para el aprendizaje de una lengua extranjera (denominado MLAT-total), para el caso en el que se pide una correlación de Pearson.

Cuadro 4.8. Correlación entre el puntaje total de rendimiento en lengua inglesa y el puntaje total de la aptitud para el aprendizaje de una lengua extranjera, según Pearson

Correlaciones

[Conjunto_de_datos3] F:\Datos cognitiva y afectiva\20121110Prueba MLAT.sav

		Puntaje general	MLAT-total
Puntaje general	Correlación de Pearson	1	.291**
	Sig. (bilateral)		.000
	N	142	142
MLAT-total	Correlación de Pearson	.291**	1
	Sig. (bilateral)	.000	
	N	142	162

**La correlación es significativa al nivel 0,01 (bilateral).

El cuadro muestra que se procesó la información de 142 sujetos y que entre las variables hay una correlación de .291, con un nivel de significancia de .01.

Los resultados de la correlación entre el puntaje total del rendimiento en lengua inglesa (denominado Puntaje general) y el puntaje total de la aptitud para el aprendizaje de una lengua extranjera (denominado MLAT-total), para el caso en el que se pide una correlación de Spearman, se presenta en el siguiente cuadro.

Cuadro 4.9. Correlación entre el puntaje total de rendimiento en lengua inglesa y el puntaje total de la aptitud para el aprendizaje de una lengua extranjera, según Spearman

Correlaciones no paramétricas

[Conjunto_de_datos3] F:\Datos cognitiva y afectiva\20121110Prueba MLAT.sav

		Puntaje general	MLAT-total
Rho de Spearman	Puntaje general	Coeficiente de correlación	1.000
		Sig. (bilateral)	.
		N	142
	MLAT-total	Coeficiente de correlación	.277**
		Sig. (bilateral)	.001
		N	142

**La correlación es significativa al nivel 0,01 (bilateral).

El cuadro muestra que se procesó la información de 142 sujetos y que entre las variables hay una correlación de .277, con un nivel de significancia de .01.

4.2.3. Determinación

El coeficiente de determinación, a diferencia del coeficiente de correlación, no sólo indica que hay una relación entre las variables sino también la proporción de la variación de la variable dependiente que es explicada por la variable independiente. Por ejemplo: si se estudia la relación entre la aptitud para una lengua extranjera y el rendimiento en su aprendizaje, el coeficiente de determinación indica la medida en la que la aptitud es causa del rendimiento.

El coeficiente de determinación se calcula elevando al cuadrado el valor del coeficiente de correlación. Al convertir su valor en porcentaje, indica la proporción mencionada. Si la correlación es de .60, el coeficiente de determinación es de .36, lo cual significa que la variable independiente participa en un 36% en el comportamiento de la variable dependiente. En el ejemplo del proceso para obtener la correlación, en el caso de una correlación de Pearson, el coeficiente es de .29, el coeficiente de determinación es de .29 al cuadrado, .084681. Esto significa que la variable independiente participa en la variable dependiente en un 8.5%.

En el mismo ejemplo, para la correlación de Spearman, el coeficiente de correlación es .277. El de determinación es .277 al cuadrado, .076729 o 7.676%.

Salkind (2008, p. 388), en relación con este coeficiente, establece el coeficiente de alienación, el cual define en un sentido opuesto al de determinación, como la carencia de variación de la variable dependiente en la relación de las variables, aquel porcentaje que no explica la relación entre las variables. En el ejemplo del párrafo anterior, este coeficiente tiene un valor de 64%.

4.2.4. Regresión lineal

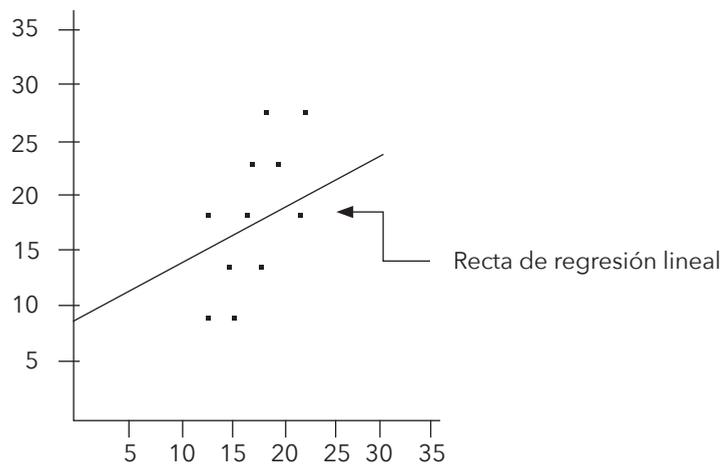
La regresión lineal es el proceso de encontrar una línea recta (como por ejemplo por cuadrados mínimos) que mejor se aproxima a un conjunto de puntos en una gráfica (Merriam-Webster, 2014).

Salkind enlaza el cálculo del coeficiente de correlación con la regresión lineal.

Entre más alto es el valor absoluto del coeficiente de correlación, más precisión hay en la predicción de una variable por la otra, con base en esa correlación, porque entre más compartan en común las variables, más se conocerá acerca de la segunda variable a través de la primera. Puede asumirse que cuando la correlación es perfecta (+1.0 o -1.0), la predicción también lo es (2008, p. 246).

También afirma que para llevar a cabo predicciones, se utiliza una línea de regresión de una variable Y en una variable X (2008, p. 248). Larson-Hall (2010, p. 400) la define como “La línea de regresión se diseña en un diagrama de dispersión que es el que se ajusta mejor a los datos. Cuando se inclina hacia arriba indica una relación positiva, si se inclina para abajo, es negativa. Si es una línea horizontal indica que no hay correlación”. Por ejemplo, en la siguiente gráfica se muestra la recta de regresión lineal y la dispersión de los datos con una relación positiva; la recta se inclina hacia arriba.

Gráfica 4.2. Recta de regresión lineal y dispersión de los datos, relación positiva



Los valores de la variable están representados por el símbolo \blacksquare . En el SPSS, la regresión se obtiene con el siguiente procedimiento:

1. Analiza, Regresión, lineales
2. Se pasan las variables dependiente e independiente al cuadro central
3. Aceptar

Los resultados se presentan como se muestra a continuación.

Cuadro 4.10. Resultado 1. Variables para el análisis de regresión lineal

Regresión

[Conjunto_de_datos3] F:\Datos cognitiva y afectiva\20121110Prueba MLAT.sav

Variables introducidas/eliminadas^b

Modelo	Variables introducidas	Variables eliminadas	Método
1	MLAT-total ^a		Introducir

a. Todas las variables solicitadas introducidas.

b. Variable dependiente: Puntaje general.

En la columna de variables introducidas está la variable independiente, puntaje total de la aptitud para una lengua extranjera. Se señala que la variable dependiente es el rendimiento en lengua inglesa (aquí denominado Puntaje general).

El siguiente cuadro contiene los datos de la variable independiente.

Cuadro 4.11. Resultado 2. Variable independiente para el análisis de regresión lineal

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
0 1	.291 ^a	.085	.078	1.13740

a. Variables predictoras: (Constante), MLAT-total.

El cuadro considera el puntaje total de la aptitud para una lengua extranjera como una variable predictora. Contiene el coeficiente de correlación (R) y el de determinación (R cuadrado); este último coeficiente indica que la variable independiente participa en la variable dependiente con un 8.5%.

El siguiente cuadro muestra datos de la relación de las dos variables en estudio y su relación.

Cuadro 4.12. Resultado 3. Variable independiente para el análisis de regresión lineal

ANOVA^b

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.	
1	Regresión	16.782	1	16.782	12.972	.000 ^a
	Residual	181.113	140	1.294		
	Total	197.895	141			

a. Variables predictoras: (Constante), MLAT-total.

b. Variable dependiente: Puntaje general.

La variable predictora si presenta una relación de causa-efecto.

Cuadro 4.13. Resultado 4. Valor de la t calculada con nivel de significancia de .000

Coeficientes^a

Modelo	Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
	B	Error típ.	Beta		
1 (Constante)	4.780	.442		10.812	.000
1 MLAT-total	.020	.006	.291	3.602	.000

a. Variable dependiente: Puntaje general.

El valor de la t calculada es de 3.602, con una significancia de .000. El valor de los grados de libertad es de 141. El valor de la t de tabla (ver anexo E), para 141 grados de libertad y un nivel de significancia de .01, es de 2.3264. Por lo tanto, el valor de t calculada es mayor que el de la tabla. Para obtener la gráfica de la dispersión de los datos, se indica lo siguiente:

SPSS

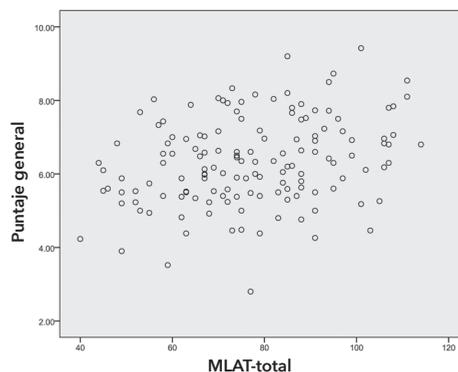
- Gráficos, Cuadros de diálogos antiguos, dispersión/puntos, Dispersión simple, Definir
- Pasar la variable dependiente al eje Y y la independiente al eje X
- Títulos
- Línea 1, nombre de la variable dependiente, Línea 2, el de la independiente
- Continuar
- Aceptar

Los resultados son los siguientes:

Gráfica 4.3. Relación de las dos variables. Puntaje general de rendimiento en lengua inglesa y aptitud para una lengua extranjera

Gráfico

[Conjunto_de_datos3] F:\Datos cognitiva y afectiva\20121110Prueba MLAT.sav



Para agregar la recta de regresión:

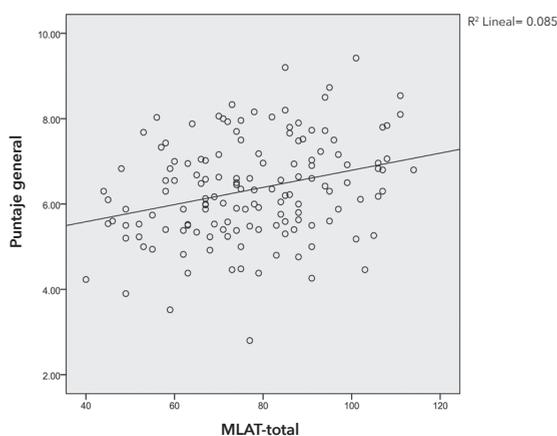
- Teclear dos veces sobre la gráfica
- Elementos, línea de ajuste total, cerrar

Se produce la siguiente gráfica de la relación entre las dos variables.

Gráfica 4.4. Relación de las dos variables. Puntaje general de rendimiento en lengua inglesa y aptitud para una lengua extranjera, con la recta de regresión

Gráfico

[Conjunto_de_datos3] F:\Datos cognitiva y afectiva\20121110Prueba MLAT.sav



4.2.5. Regresión múltiple

El análisis de regresión múltiple consiste en buscar la relación entre varias variables para tratar de predecir cómo influyen algunas de las variables independientes en la variable dependiente; es decir “qué tanto de la variable dependiente puede explicarse por la variación de los puntajes de las variables independientes” (Larson-Hall, 2010, p. 133). Hay algunos requisitos para el cálculo de la regresión:

- Hay dos o más variables
- Estas variables no tienen niveles entre ellas
- Si se toman promedios de las variables, se tienen tres o más
- Todas las variables son continuas; es decir, son categóricas
- A diferencia de la correlación, una variable debe ser dependiente y las otras independientes

En el SPSS, se obtiene como se indica a continuación.

- Analizar, Regresión, Escalamiento óptimo
- Pasar la variable dependiente y las independientes
- Aceptar

Los resultados son los siguientes.

Cuadro 4.14. Resultados 1. Casos incluidos en el proceso

CATREG: Regresión para datos categóricos
 [Conjunto_de_datos3] F:\Datos cognitiva y afectiva\20121110Prueba MLAT.sav
 Resumen del procesamiento de los casos

Casos activos válidos	132
Casos activos con valores perdidos ^a	30
Casos suplementarios	0
Total	162
Casos usados en el análisis	132

a. Casos excluidos: 7 11 14 17 18 19 22 40 44 79 85 107 112 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160.

Los siguientes cuadros (4.15 a 4.18) muestran los resultados 2 a 5 obtenidos en este proceso.

Cuadro 4.15. Resultados 2. Coeficientes de correlación y determinación

Resumen del modelo

R múltiple	R cuadrado	R cuadrado corregida	Error de predicción aparente
.545	.297	.206	.703

Variable dependiente: Puntaje general.

Predictores: MLAT-1 MLAT-2 MLAT-3 MLAT-4 MLAT-5 MLAT-total.

Cuadro 4.16. Resultados 3. Valor calculado para F y nivel de significancia

ANOVA

	Suma de cuadrados	Gl	Media cuadrática	F	Sig.
Regresión	39.202	15	2.613	3.267	.000
Residual	92.798	116	.800		
Total	132.000	131			

Variable dependiente: Puntaje general.

Predictores: MLAT-1 MLAT-2 MLAT-3 MLAT-4 MLAT-5 MLAT-total.

Cuadro 4.17. Resultados 4. Valor calculado para F y nivel de significancia para cada uno de los predictores

Coefficientes

	Coefficientes tipificados		gl	F	Sig.
	Beta	Bootstrap (1000) Estimación de error típico			
MLAT-1	-.152	.385	2	.155	.856
MLAT-2	.048	.205	1	.055	.815
MLAT-3	.237	.258	2	.846	.432
MLAT-4	-.292	.212	4	1.894	.116
MLAT-5	-.375	.187	2	4.039	.020
MLAT-total	.778	.569	4	1.870	.120

Variable dependiente: Puntaje general.

Cuadro 4.18. Resultados 5. Correlaciones entre la variable dependiente y las dependientes

Correlaciones y tolerancia						
	Correlaciones			Importancia	Tolerancia	
	Orden cero	Parcial	Semiparcial		Después de la transformación	Antes de la transformación
MLAT-1	.176	-.123	-.104	-.090	.467	.009
MLAT-2	.116	.054	.045	.019	.898	.063
MLAT-3	.372	.247	.214	.297	.813	.037
MLAT-4	.116	-.228	-.197	-.114	.454	.036
MLAT-5	.005	-.308	-.271	-.006	.521	.029
MLAT-total	.340	.358	.321	.892	.170	.003

Variable dependiente: Puntaje general.

4.3. Toma de decisiones

Larson-Hall (2010, p. 49) afirma que el investigador no sólo proclamará el valor de la significancia y declarará que la hipótesis es estadísticamente significativa o no. De hecho, dirá el tipo de prueba estadística que usó y

reportará el valor de esa estadística, los grados de libertad asociados con la prueba y el valor de la significancia.

4.3.1. Elección de la prueba estadística

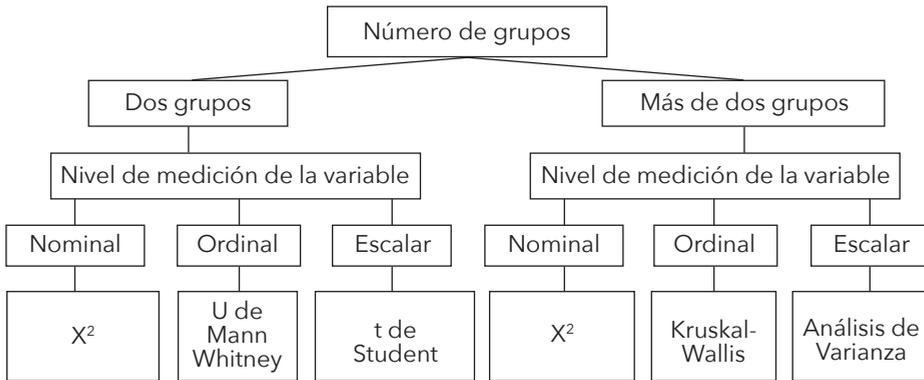
En este inciso se presentan dos rubros, los criterios de elección de las pruebas estadísticas y el significado de hablar de una curva con una o dos colas.

4.3.1.1. Criterios de elección de la prueba estadística

Larson-Hall (2010, p. 129-130) considera que en el área de investigación en segunda lengua, las pruebas básicas utilizadas con mayor frecuencia son: correlación, regresión, pruebas t, ANOVA y prueba Chi-cuadrada. Plantea dos tipos de investigaciones: de relaciones y de diferencias. Las de relaciones son para conocer qué tanto influye una variable en otra o cuáles pueden predecir puntajes de otra; para éstas se usa la correlación. Las de diferencias son para ver si una persona de una categoría particular o grupo influye en el resultado de alguna prueba; en este caso, se utilizan las pruebas Chi-cuadrada y t, y el análisis de varianza ANOVA. Se toman pruebas de relaciones en los casos en los que se usan puntajes de pruebas y de escalas, y variables nominales como la edad. Si se tiene una mezcla de tipos de variables, se usan las de diferencias.

Juárez *et al.* (2002) establecen criterios para la selección del tipo de prueba. Dividen el tipo de problemas en mano en dos grandes rubros: los estudios de comparación y los de asociación. Para la clasificación que hacen con respecto al primer grupo, que consiste en comparar grupos o mediciones y es para el caso de una variable dependiente y una variable independiente, el primer paso es determinar si se trabaja con muestras independientes o relacionadas. Las muestras independientes son aquellas en las que "la variable dependiente de interés es medida sólo una vez en la unidad de análisis" (Juárez *et al.*, 2002, p. 18). En cambio, en el caso de las relacionadas, "la variable dependiente es medida más de una ocasión en la unidad de análisis" (Juárez *et al.*, 2002, p. 18). A continuación se presenta un esquema que muestra las pruebas que deben elegirse para el caso de muestras independientes, de acuerdo con el número de grupos utilizados y el nivel de medición de las variables. X^2 es la prueba Chi-cuadrada.

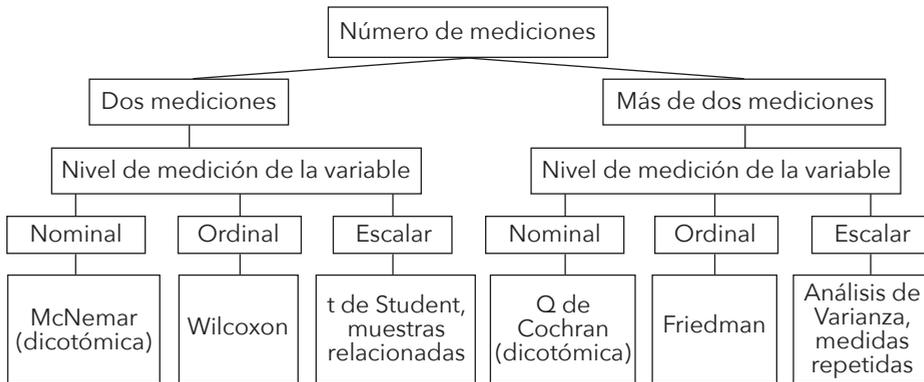
Esquema 4.2. Elección de pruebas estadísticas para muestras independientes



Fuente: adaptado de Juárez et al. (2002, p. 18).

El siguiente esquema muestra las pruebas que deben elegirse para el caso de muestras relacionadas, de acuerdo con el número de mediciones utilizadas y el nivel de medición de las variables.

Esquema 4.3. Elección de pruebas estadísticas para muestras relacionadas



Fuente: adaptado de Juárez et al. (2002, p. 18).

La presentación de las pruebas exhibidas en los esquemas 4.2 y 4.3 se encuentra en el capítulo cinco.

4.3.1.2. Curva de una o de dos colas

Al hacer las pruebas de hipótesis con los datos, hay dos alternativas. La primera consiste en considerar la hipótesis de que un grupo A tenga mejores resultados que un grupo B y la segunda, que el grupo A tenga peo-

res resultados que el grupo B. Si la decisión es trabajar solamente con una de las dos alternativas, puede usarse una prueba de una sola cola (Larson-Hall, 2010, p. 99).

4.3.2. Análisis de acuerdo con el nivel de medición de las variables

Juárez *et al.* (2002, p. 3) presentan un cuadro muy completo en cuanto a la relación de los niveles de medición de las variables y algunas características asociadas a ellos. A continuación está la transcripción.

Cuadro 4.19. Relación de los niveles de medición de las variables y algunas características asociadas a ellos

Nivel	Propiedades	Medidas de tendencia central y dispersión	Compara	Análisis	Ejemplo
Nominal	= o ≠	moda, porcentaje	Proporciones	Inferencial no paramétrica: χ^2 , McNemar, Coeficiente Phi	Presencia de estrés (si o no), Religión
Ordinal	≠, =, >, < .	moda, porcentajes, mediana	Medianas	Inferencial no paramétrica: U de Mann Whitney, Kolmogorov Smirnov, Friedman, Wilcoxon, prueba de los signos, correlación Spearman	Calidad (buena, mediana, mala), Nivel socioeconómico (alto, medio, bajo)
Intervalar	Distancias iguales, =, ≠, >, < 0 relativo	moda, porcentajes, mediana, media, desviación estándar, varianza, curtosis, sesgo, rango, deciles, cuartiles, percentiles	Medias	Inferencial paramétrica: t de Student, ANOVA, regresión, correlación Pearson	Temperatura, atributos psicológicos
Razón o proporción	Igual a intervalar, 0 absoluto	igual a intervalar	igual a intervalar	igual a intervalar	Altura, Peso, Edad en años, Grados Kelvin

Fuente: Juárez *et al.* (2002, p. 3).

4.4. Confiabilización de los datos levantados y capturados

Antes de llevar a cabo algún proceso con los datos recabados, introducidos a un procesador, es recomendable verificar que son confiables. La confiabilidad de los datos se define a través del cálculo del Alfa de Cronbach (ver inciso 4.2.1).

Adicionalmente, una vez confiabilizados los datos, en ocasiones es necesario llevar a cabo otros procesos para elegir la prueba estadística para analizarlos. Por ejemplo, en el inciso 3.2.1.2, se señaló que, para la elección de una prueba inferencial paramétrica, hay que asegurarse de que los datos posean las características de ser distribuidos como una curva normal y la homogeneidad de varianzas entre dos grupos que se comparan.

4.5. Gráficas

Existen diversos tipos de gráficas. Las más conocidas son las de barras y las de pastel. Las gráficas no se eligen al azar o por selección personal. También hay gráficas idóneas para cada tipo de datos. Bakieva, González, Jornet y Terol (2010) consideran los tipos de gráficas manejadas en el SPSS, versión 15. Ésas son ocho:

- de barras
- de líneas
- de áreas
- de sectores (conocida como gráfica de pastel)
- de máximos y mínimos
- de caja
- de dispersión
- histogramas

Las gráficas de barras, de líneas y de áreas se utilizan para variable categóricas; las de sectores, para la comparación de proporciones; las de máximos y mínimos son "las gráficas para las que se muestra el intervalo de datos comprendido entre dos valores" (Bakieva *et al.* 2010); de caja, cuando en la gráfica aparecen cuadros pequeños en la línea de datos, "pueden ser simples, agrupados y son útiles para mostrar la distribución de una variable escalar" (Bakieva *et al.* 2010). Las de dispersión son "observaciones individuales de las gráficas en los valores para una variables numérica"; "hay categorías amplias de gráficos que se crean con el elemento gráfico punto [...] los diagramas de dispersión son útiles para re-

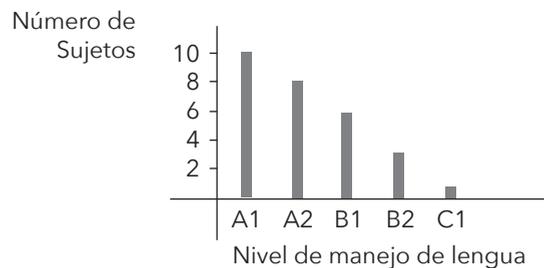
presentar datos multivariantes” (Bakieva *et al.* 2010). Por último, están los histogramas.

Los histogramas muestran el número de casos que se encuentran dentro de cada intervalo. Son útiles para mostrar la distribución de una única variable de escala. Los datos se agrupan y se resumen, utilizan estadístico de porcentaje o recuento. Una variante del histograma es el polígono de frecuencias, que es similar a un histograma normal pero con la diferencia de que se utiliza el elemento gráfico de área en vez del elemento gráfico de barra (Bakieva *et al.*, 2010).

Enseguida se presentan las dos gráficas más comunes: la de barras y la de sectores (conocida también como gráfica de pastel).

La gráfica de barras se utiliza para mostrar variables nominales (también denominadas variables categóricas), con el fin de comparar el comportamiento de las subcategorías de la variable. Por ejemplo, para graficar el número de sujetos en estudio que poseen diversos niveles de lengua. La variable o categoría en estudio es el nivel de lengua de los sujetos y las subcategorías o valores que puede asumir son los niveles: A1, A2, B1, B2 y C1. La gráfica se construye con una columna para cada uno de los cinco niveles, para indicar el número de sujetos que posee cada uno de ellos.

Gráfica 4.5. Número de sujetos por nivel de lengua



La gráfica 4.5 está respaldada por los siguientes datos:

Cuadro 4.20. Número de sujetos por nivel de lengua

Nivel de lengua	A1	A2	B1	B2	C1
Número de sujetos	10	8	6	3	1

Este tipo de gráfica sirve para visualizar las diferencias entre los puntajes para cada nivel de lengua. Mediante las barras se distinguen las similitudes y las diferencias entre los datos graficados.

La gráfica de sectores (gráfica de pastel) se prepara transformando los datos individuales a porcentajes de todo el conjunto. Sirve para comparar proporciones (Bakieva *et al.*, 2010).

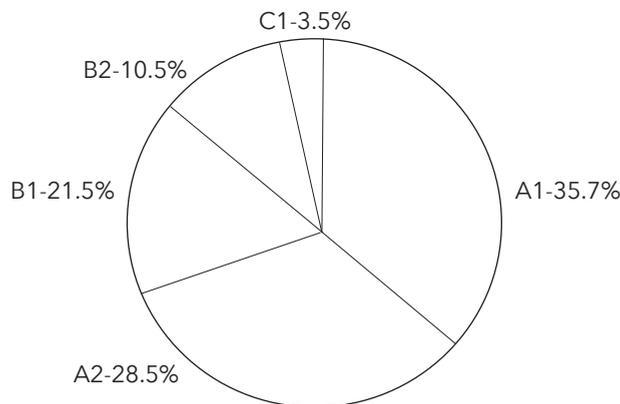
En este caso, el punto de partida para elaborar la gráfica es el siguiente cuadro.

Cuadro 4.21. Número y porcentaje de sujetos por nivel de lengua

Nivel de lengua	A1	A2	B1	B2	C1
Número de sujetos	10	8	6	3	1
Porcentaje de sujetos	35.7%	28.5%	21.5%	10.7%	3.6%

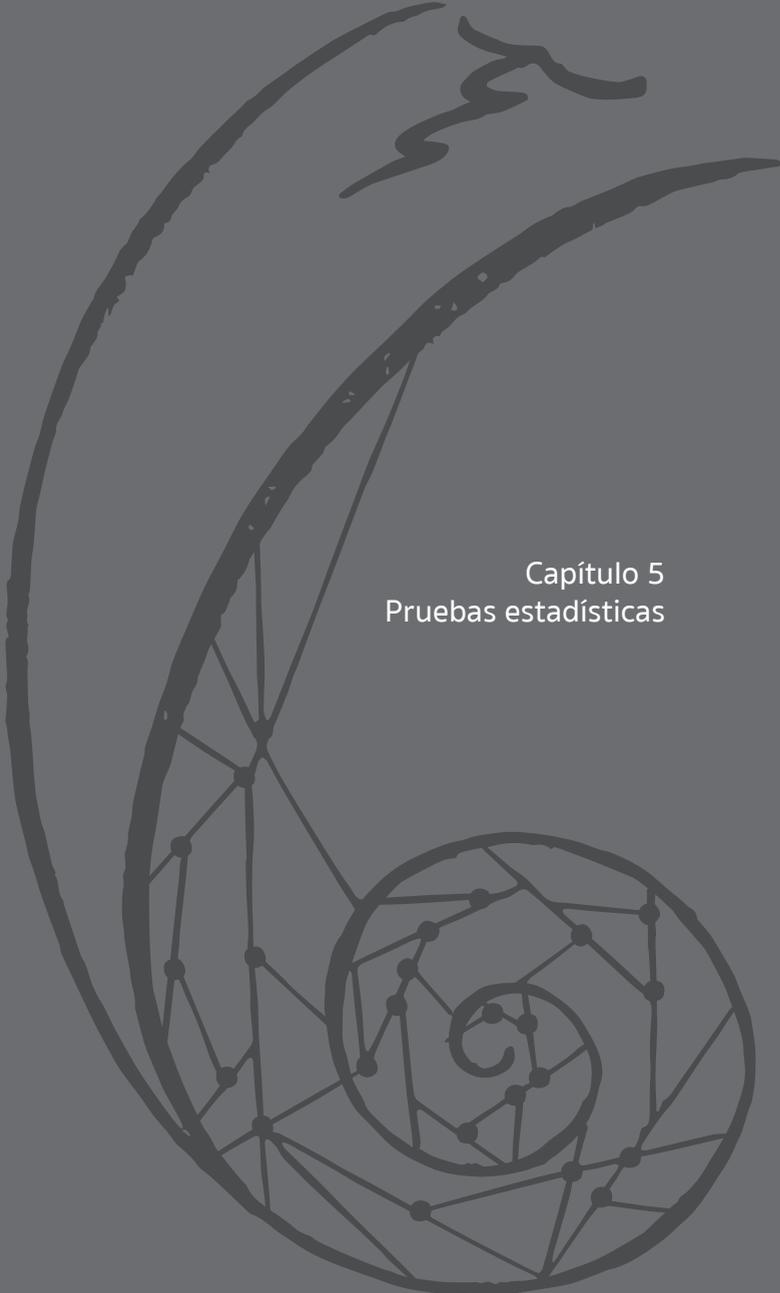
La expresión de los resultados en porcentajes favorece su análisis. Se visualizan con mayor claridad. La gráfica se muestra a continuación.

Gráfica 4.6. Porcentajes de sujetos por nivel de lengua



En la versión 18 se maneja la gráfica de ejes dobles (dos gráficas conjuntas). La forma de obtenerlas en el SPSS se encuentra en el inciso 8.10.

En este capítulo se revisaron los elementos estadísticos que corresponden al procesamiento de datos de tipo inferencial; algunos parámetros básicos para tener un punto de partida con datos confiables, criterios para la selección de las pruebas estadísticas idóneas e información general relativa al uso de gráficas.



Capítulo 5
Pruebas estadísticas



El capítulo cinco indica las pruebas estadísticas que se aplican a los datos levantados para obtener resultados, se denominan como pruebas de hipótesis, básicamente de corte cuantitativo. Este apartado inicia con un cuadro que resume las pruebas paramétricas y no paramétricas y un cuadro de la situación de la hipótesis nula en cada una de las pruebas. Precisa solamente aquellas pruebas que se utilizan con mayor frecuencia, con indicaciones relativas al uso de la prueba, la forma adoptada por la hipótesis nula, las instrucciones en el Programa SPSS y los resultados que se obtienen al procesar los datos con éste.

Como un resumen de los esquemas 4.2 y 4.3 y del cuadro 4.19, del capítulo anterior, se obtiene el siguiente cuadro, organizado dependiendo de si la prueba es o no paramétrica.

Cuadro 5.1. Pruebas de hipótesis paramétricas y no paramétricas

Tipo de prueba	Prueba	Número de grupos	Número de mediciones	Nivel de medición	Proceso
Paramétrica	t para una sola muestra	Una	Una	Escalar	Comparación de medias
	t de muestras independientes	Dos	Dos	Escalar	Comparación de medias

Paramétrica	t de muestras dependientes	Dos	Dos	Escalar	Comparación de medias
	t Student-Welch para varianzas no homogéneas	Dos	Dos	Escalar	Comparación de medias
	ANOVA	Más de dos	Más de dos	Escalar	Comparación de medias
No Paramétrica	χ^2 , Chi-cuadrada	Dos o más	Dos o más	Nominal	Relación entre dos variables categoriales
	U de Man Whitney	Dos	Dos	Ordinal	Comparación de medias
	Kruskal-Wallis	Más de dos	Más de dos	Ordinal	Comparación de medias
	McNemar	Dos	Dos	Nominal	Comparación de valores antes y después de un tratamiento
	Wilcoxon	Dos	Dos	Ordinal	Comparación de medias
	Q de Cochran	Más de dos	Más de dos	Nominal	Para k muestras de la variable independiente, con valores dicotómicos
	Friedman	Más de dos	Más de dos	Ordinal	Los grupos o medidas repetidas provienen de la misma población

La equivalencia entre las pruebas paramétricas y no paramétricas se establece como se indica a continuación. U de Mann Whitney es la alternativa no paramétrica para la prueba t de muestras independientes (Larson-Hall, 2010, p. 376). La Kruskal-Wallis y la Friedman son la contraparte no paramétrica de la ANOVA de una vía. McNemar y Wilcoxon son la alternativa no paramétrica para la prueba t de muestras dependientes o relacionadas (Larson-Hall, 2010, p. 381).

Se acostumbra hablar de muestras, no obstante, pueden expresarse como sujetos o bien como grupos.

El cuadro de las formas adoptadas para las hipótesis nulas en una investigación y en las pruebas estadísticas es el siguiente. En algunos

casos se plantea en términos de la falta de relación de las variables; en otros, en función de la igualdad o falta de diferencia entre ellas.

Cuadro 5.2. Usos de la hipótesis nula

Prueba de hipótesis	Hipótesis Nula, H_0	Valor del nivel de significancia, acepta H_0
Investigación	No hay relación entre variables	> .05
Investigación	No hay diferencia entre variables	> .05
Correlación	No existe correlación, es cero	
Prueba de Levene	Las varianzas son iguales	> .05
Prueba de normalidad	El conjunto de datos sigue una distribución normal	> .05
t de Student	No hay diferencia entre las medias	> .05
ANOVA	El poder predictivo del modelo es igual a cero (Larson-Hall, 2010, p. 192)	> .05
Chi-Cuadrada	Las variables no se relacionan	> .05
U. de Man Whitney	No hay diferencia entre las medias	> .05
Kruskal-Wallis	No hay diferencia entre las medias	> .05
McNemar	El tratamiento no produce cambios significativos en la variable en estudio	> .05
Wilcoxon	No hay diferencia entre las medias	> .05
Q de Cochran	No existe diferencia entre las medias	> .05
Friedman	Las muestras han sido sacadas de la misma población	> .05

Si el valor de la significancia es menor o igual a .05, la hipótesis nula se rechaza. Si es mayor, se acepta. Es decir, para todas las pruebas, la hipótesis nula se acepta si el valor de la significancia es mayor a .05. A partir del cuadro 5.2, se aprecia que en los catorce casos presentados predomina el uso de la hipótesis nula para indicar que no hay diferencia entre las medias (6 casos) y que no hay relación entre las variables (2 casos).

En los incisos siguientes se presentan las pruebas mencionadas en el cuadro 5.1. Se indica su uso, la hipótesis nula ya revisada en el cuadro 5.2 y la manera de procesar los datos en el SPSS.

5.1. Prueba t de Student

La prueba t de Student (prueba t) se utiliza para conocer si hay diferencias entre medias. Estas medias se refieren a las que se obtienen a partir de los datos levantados en uno o dos grupos de sujetos, según sea el caso. Si hay diferencias entre medias, se comprueba que los datos de los grupos difieren entre sí. Hay cuatro tipos de esta prueba, para una muestra, para dos muestras independientes, para dos muestras dependientes o relacionadas y la prueba t Student-Welch para varianzas no homogéneas.

A continuación se presentan los casos de las pruebas y ejemplos de su proceso en el SPSS para una y dos muestras y el caso de la t Student-Welch.

Prueba t para una muestra o grupo de sujetos

Hipótesis nula: "No hay diferencias significativas entre las medias de la muestra y un parámetro poblacional" (Visauta, 2007, p. 97).

SPSS

- Analizar, Comparar medias, prueba t para una muestra
- Se pasa la variable en estudio a la ventana central, se da un valor estimado para la variable
- Aceptar

Se acepta la hipótesis nula cuando el valor de la significancia es mayor a .05.

Cuadro 5.3. Resultados 1. Prueba t para una muestra, la variable

Prueba t

[Conjunto_de_datos3] F:\Datos cognitiva y afectiva\20121110Prueba MLAT.sav

Estadísticos para una muestra				
	N	Media	Desviación típ.	Error típico de la media
Puntaje general	142	6.3351	1.18470	.09942

Este cuadro muestra la media obtenida por el grupo único.

Cuadro 5.4. Resultados 2. Prueba t para una muestra, nivel de significancia

Prueba para una muestra						
	Valor de prueba = 6.5					
	T	Gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
Puntaje general	-1.659	141	.099	-.16493	-.3615	.0316

Los datos del cuadro 5.4 conducen a que no hay diferencias entre la media obtenida por el grupo y el parámetro poblacional; la significancia (.099) es mayor a .05.

Prueba t para dos muestras, grupos de sujetos con datos independientes

Para poder aplicar la prueba t de dos muestras o grupos de sujetos con datos independientes, se requiere que las variables sean de intervalo o de razón, una distribución normal o aproximada a la normal y que las varianzas sean homogéneas (ver inciso 3.2.1.2).

Hipótesis nula: No hay diferencias entre las medias de dos muestras, de dos grupos independientes.

SPSS

- Analizar, Comparar medias, prueba t para muestras independientes
- Se pasa la variable en estudio a la ventana central. Se indica la variable de agrupación, en la ventanilla denominada como tal definir grupos
- Indicar los valores de la variable que define los grupos, Continuar
- Aceptar

Se obtienen los siguientes resultados: Estadísticas de grupo, prueba de Levene para prueba de grupos independientes y prueba t para grupos independientes.

A continuación está un ejemplo en el que se muestran los resultados proporcionados por el SPSS, dos cuadros que contienen datos generales de las variables que identifican los datos de cada uno de los grupos comparados (denominado resultados 1) y los resultados de la prueba de Levene, para homocedasticidad de las varianzas y de la prueba t de diferencia de medias (denominados como Resultados 2-1 y Resultados 2-2 debido a que el SPSS produce un solo cuadro; éste se separó en dos para facilitar su lectura e interpretación).

Cuadro 5.5. Resultados 1. Medias de grupos independientes comparados con la prueba t

Estadísticas de grupo

	Grupo	N	Media	Desviación típica	Error típico de la media
Grado de conocimiento	Grupo control	17	2.41	.618	.150
	Grupo experimental	19	1.84	.765	.175

Este cuadro muestra las medias de cada uno de los grupos comparados; en este caso, el grado de conocimiento de los grupos experimental y control.

Cuadro 5.6. Resultados 2-1. Medias de grupos independientes comparados con la prueba t

Prueba de grupos independientes		
	Prueba de Levene para la igualdad de varianzas	
	F	Sig.
Grado de conocimiento Se han asumido varianzas iguales	.000	.994
No se han asumido varianzas iguales		

En este cuadro, se reconoce que las varianzas son iguales. Se acepta la hipótesis nula debido a que el nivel de significancia es mayor de .05.

En el siguiente cuadro, se muestra el valor de t, valor denominado t calculada y los grados de libertad, así como el valor de la significancia.

Cuadro 5.7. Resultados 2-2. Medias de grupos independientes comparados con la prueba t

Prueba de grupos independientes						
Prueba T para la igualdad de medias						
T	Gl	Sig. (bilateral)	Diferencia de medias	Error típico de la diferencia	95% intervalo de confianza para la diferencia	
					Inferior	Superior
2.439	.34	.020	.570	.234	0.94	1.044
2.468	.33684	.019	.570	.231	.100	1.039

Los pasos para aceptar o rechazar la hipótesis nula son los siguientes:

1. Se analiza el resultado de la prueba de Levene, de homogeneidad de varianzas (visto en el inciso 4.1.5). Si el valor de significancia de la prueba de Levene es mayor de .05, hay homogeneidad de varianzas.
2. Se procede a tomar el valor de la prueba t, los grados de libertad y el nivel de significancia.
3. Cuando el valor de la significancia para esta prueba t es mayor a .05, implica que se acepta la hipótesis nula y que, por lo tanto, no hay diferencias entre las medias de las dos muestras. En ese momento, el análisis termina.
4. Si el valor de la significancia de la prueba t es menor o igual a .05, implica que hay diferencias entre las medias de los grupos, en cuyo caso se sigue adelante con el análisis de los resultados de la prueba.
5. El siguiente paso consiste en consultar la tabla de valores de la prueba t, para determinar si se acepta la hipótesis de que existen diferencias entre los grupos analizados. El valor de t obtenido en este proceso, como se mencionó, se denomina t calculada.
6. Con la t calculada y los grados de libertad dados en la tabla de resultados, se consulta la tabla (ver anexo E). Si la t calculada es menor al valor de la tabla significa que no hay diferencias entre las medias de los grupos. Si es mayor o igual significa que sí hay diferencias entre los dos grupos.

Prueba t para dos grupos dependientes o relacionados

La prueba t con datos apareados es un prueba paramétrica a través de la cual se sabe si los puntajes de dos grupos en los que fueron encuesta-

das las mismas personas dos veces presentan diferencias estadísticamente significativas (Larson-Hall, 2010, p. 397).

Para aplicar la prueba t de dos grupos relacionados, se requiere que las variables sean de intervalo o de razón, una distribución normal o aproximada a la normal, no interesa conocer la varianza de las medias y hay dos mediciones, anterior y posterior a un tratamiento. Hipótesis nula: No hay diferencias entre las medias de dos grupos relacionados.

Los sujetos relacionados pueden ser los mismos sujetos en dos situaciones diferentes (Visauta, 2007, p. 103).

SPSS

- Analizar, comparar medias, prueba t para muestras relacionadas
- Se pasan al centro las variables en estudio, se establecen pares si hay necesidad; se van numerando y se señalan las variables que componen cada uno de los pares
- Aceptar

Como resultado, se obtienen tres cuadros. El primero contiene las medias de las variables comparadas. El segundo muestra la correlación de las dos variables. El tercero indica características de las variables comparadas y los resultados de la prueba t. A continuación se presenta un ejemplo en el que el tercer cuadro se separa en dos para facilitar su comprensión.

Cuadro 5.8. Resultados 1. Medias de grupos relacionados comparados con la prueba t

Estadísticas de muestras relacionadas					
		Media	N	Desviación típ.	Error típ. de la media
Par 1	Nivel conocimiento A	2.11	36	.747	.125
	Nivel conocimiento B	2.44	36	.504	.084

En esta ocasión, se comparan las variables relativas a nivel de conocimiento antes y después de un tratamiento. Se tienen 36 sujetos, con medias de 2.11 y 2.44, antes y después del tratamiento, respectivamente. El segundo cuadro de resultados es el siguiente.

Cuadro 5.9. Resultados 2. Correlación entre las muestras relacionadas

Correlaciones de muestras relacionadas				
		N	Correlación	Sig.
Par 1	Nivel de conocimiento A Nivel de conocimiento B	36	.320	.057

El cuadro muestra una correlación entre las dos muestras de .320. El tercer cuadro obtenido con el SPSS se muestra dividido en los siguientes dos cuadros. El primero muestra los valores de las variables comparadas, nivel de conocimiento A y nivel de conocimiento B. El último cuadro muestra los valores de la T, los grados de libertad y el nivel de significancia.

Cuadro 5.10. Resultados 3-1 de la prueba t de las muestras relacionadas

Prueba de muestras relacionadas					
	Diferencias relacionadas				
					95% Intervalo de confianza para la diferencia
	Media	Desviación estándar	Error estándar	Inferior	Superior
Par 1 Nivel conocimiento A Nivel conocimiento B	-.333	.756	.126	-.589	-.078

A continuación se muestra la segunda parte del tercer cuadro de los resultados.

Cuadro 5.11. Resultados 3-2 de la prueba t de las muestras relacionadas

Prueba de muestras relacionadas			
	T	Gl	Sig. (bilateral)
Par 1 Nivel conocimiento A Nivel conocimiento B	-2.646	35	.012

Se acepta la hipótesis nula cuando el valor de la significancia es mayor a .05. En este caso, la significancia es de .012; por lo tanto, se acepta la hipótesis de que existen diferencias entre las medias de las dos muestras relacionadas. El análisis termina con este resultado.

t de Student-Welch para varianzas no homogéneas

Para describir esta prueba se utilizó el texto de Raymundo (2014d). En relación con la prueba t de Student-Welch para dos muestras independientes con varianzas no homogéneas, Raymundo (2014d) afirma que, en este modelo estadístico, el agregado de Welch consiste en una ecuación para calcular los grados de libertad, de manera que disminuye el error por la no homogeneidad de las varianzas.

Los pasos a seguir son los siguientes:

1. Determinar el promedio, la varianza y el tamaño de la muestra de cada población en el estudio.
2. Aplicar la ecuación t.
3. Calcular los grados de libertad (gl) de acuerdo con la ecuación dada.
4. Comparar el valor de t calculado respecto a los grados de libertad con los valores de t críticos.
5. Decidir si se acepta o rechaza la hipótesis.

Esta prueba no se maneja como tal en la versión 18 del SPSS.

Prueba t de Student-Welch para dos muestras independientes con varianzas no homogéneas

SPSS

- Analizar, comparar medias. ANOVA de un factor
- Pasar al centro el nombre de la variable dependiente y el de la variable de agrupación (factor), Opciones
- Descriptivos, efectos fijos y aleatorios, prueba de homogeneidad de varianzas, Welch, Continuar
- Aceptar

5.2. ANOVA

Larson-Hall (2010, p. 396-397) solamente presenta la prueba ANOVA de una sola vía como una prueba paramétrica que tiene una variable dependiente y una variable independiente. Se utiliza cuando hay tres o más niveles (agrupamientos) en la variable independiente, pero puede usarse cuando hay solamente dos. Esta prueba pregunta si los grupos difieren en su desempeño en la variable dependiente.

En *Statistics Solutions* (2013) se indica que la prueba ANOVA es un método estadístico sustentado en el análisis de la varianza, desarrollado por Ronald Fisher en 1918 como una extensión de las pruebas T y Z. Esta es la razón por la que se incluye en este apartado. La prueba t sólo se utiliza para dos grupos, la ANOVA para comparar medias dentro y entre grupos, siempre que haya más de dos grupos y admite varias hipótesis nulas al mismo tiempo (*Statistics Solutions*, 2013). Por esta obra, se conoce que la ANOVA tiene tres variantes: de una sola vía, de dos vías y de n vías, una ANOVA multivariada. En la de una vía, se comparan varios grupos con un solo factor o variable independiente; por ejemplo, el rendimiento de tres grupos, de acuerdo con las horas de clase de lengua recibidas. En la de dos vías, hay dos variables independientes; por ejemplo, el rendimiento de tres grupos, de acuerdo con las horas de clase de lengua recibidas y el estilo de enseñanza de los profesores. Puede ser una ANOVA de dos vías balanceada, si el número de sujetos de cada grupo es el mismo y desbalanceada si no lo es. Se utiliza tanto para ver efectos de variables independientes por separado o en la interacción de las mismas. Por último, la ANOVA multivariada es aquella en la que se consideran factores múltiples, es decir, más de dos variables independientes.

En la página de *Statistics Solutions* (2013) pueden consultarse las alternativas para la determinación del grupo o grupos que provocan las diferencias.

La ANOVA es una prueba paramétrica. Los requisitos para utilizarla son similares a los estipulados para la prueba t, se requiere una distribución normal o aproximada a la normal y que las varianzas sean iguales. También es necesario que las variables dependientes sean de intervalo o de razón, continuas, las independientes sean categoriales y que los grupos sean independientes unos de otros (*Statistics Solutions*, 2013).

Larson-Hall (2010, p. 140) establece los siguientes atributos para la ANOVA de una sola vía:

- Tiene exactamente dos variables
- Una variable es categorial con tres o más valores y es la variable independiente
- La otra variables es continua y es la variable dependiente
- Si se toman promedios de las variables se tendrán tres o más

Un ejemplo de aplicación sería: ¿Hay diferencias entre un grupo de una preparatoria urbana privada, un grupo de una preparatoria urbana pública y un grupo de una preparatoria rural, en el aprendizaje de la gramática en lengua inglesa?

La hipótesis nula establece que no hay diferencias entre los grupos.

SPSS

- Analizar, comparar medias. ANOVA de un factor
- Pasar al centro los nombres de las variables dependientes (Lista de dependientes) y la variable que indica los grupos (factor), Opciones*
- Descriptivos, efectos fijos y aleatorios, prueba de homogeneidad de varianzas, Continuar
- Aceptar
 - * Las opciones en la ANOVA de un factor son:
 - Estadísticos
 - Descriptivos
 - Efectos fijos y aleatorios
 - Prueba de homogeneidad de las varianzas
 - Brown-Forsythe
 - Welch
 - Gráfica de las medias
 - Valores perdidos
 - Excluir casos según análisis
 - Excluir casos según lista

Como resultado se tiene la leyenda ANOVA DE UN FACTOR y tres cuadros. Uno con datos descriptivos de las variables, el segundo con la prueba de homogeneidad de las varianzas y el tercero con los resultados de la prueba ANOVA. Estos cuadros son los siguientes, el primero dividido en dos para fines de presentación.

**Cuadro 5.12. Resultados 1-1. ANOVA de una sola vía.
Datos descriptivos de las variables**

Descriptivos						
			N	Media	Desviación típica	Error típico
Grado de conocimiento Grupo uno	Masculino		12	2.17	.937	.271
	Femenino		24	2.08	.654	.133
	Total		36	2.11	.747	.125
	Modelo	Efectos fijos			.757	.126
		Efectos aleatorios				.126 ^a
Grado de conocimiento Grupo dos	Masculino		12	2.33	.492	.142
	Femenino		24	2.50	.511	.104
	Total		36	2.44	.504	.084
	Modelo	Efectos fijos			.505	.084
		Efectos aleatorios				.084 ^a

a. Advertencia. La varianza entre componentes es negativa. Ha sido reemplazada por 0,0 al calcular esta medida para los efectos aleatorios.

A continuación está la segunda parte del primer cuadro.

**Cuadro 5.13. Resultados 1-2. ANOVA de una sola vía.
Datos descriptivos de las variables**

Descriptivos							
			Intervalo de confianza para la media al 95%		Mínimo	Máximo	Varianza entre componentes
			Límite inferior	Límite superior			
Grado de conocimiento Grupo uno	Masculino		1.57	2.76	1	4	-.032
	Femenino		1.81	2.36	1	4	
	Total		1.86	2.36	1	4	
	Modelo		1.81	2.37			
		Efectos fijos		.51 ^a	3.71 ^a		
		Efectos aleatorios					
Grado de conocimiento Grupo dos	Masculino		2.02	2.65	2	3	-.002
	Femenino		2.28	2.72	2	3	
	Total		2.27	2.61	2	3	
	Modelo						
		Efectos fijos		2.27	2.62		
		Efectos aleatorios		1.38 ^a	3.51 ^a		

La segunda parte de los resultados se refiere a la prueba de homogeneidad de varianzas.

**Cuadro 5.14. Resultados 2. ANOVA de una sola vía.
Homogeneidad de varianzas**

Prueba de homogeneidad de varianzas				
	Estadístico de Levene	gl1	gl2	Sig.
Grado de conocimiento Grupo uno	3.145	1	34	.085
Grado de conocimiento Grupo dos	2.833	1	34	.101

Por último, está la tabla de la ANOVA.

Cuadro 5.15. Resultados 3. ANOVA de una sola vía.
Resultados de la prueba

ANOVA						
		Suma de cuadrados	Gl	Media cuadrática	F	Sig.
Grado de conocimiento Grupo uno	Inter-grupos	.056	1	.056	.097	.758
	Intra-grupos	19.500	34	.574		
	Total	19.556	35			
Grado de conocimiento Grupo dos	Inter-grupos	.222	1	.222	.872	.357
	Intra-grupos	8.667	34	.255		
	Total	8.889	35			

El procedimiento de cálculo es el siguiente:

- Comparar el valor crítico de la razón F con el valor de tabla (ver anexo F) o ver el valor de la significancia contra el alpha establecido
- Si el valor crítico calculado es mayor al valor en la tabla, se rechaza la hipótesis nula
- Se concluye que las medias de los grupos no son iguales
- Si el valor es menor, se acepta la hipótesis nula
- El resultado dirá que existen diferencias pero no dirá exactamente en qué grupos
- Para hacerlo, hay que llevar a cabo otros procesos

Para la ANOVA de dos vías o multivariada habrá dos o más variables independientes. Las variables independientes pueden asumir varios valores. Larson-Hall (2010, p. 141) señala los siguientes elementos como atributos de la ANOVA multivariada.

- Tiene más de dos variables
- Dos o más variables son categoriales y son las variables independientes
- Exactamente una variable es continua (de intervalo o de razón) y es la variable dependiente

Esta autora presenta el siguiente ejemplo para aplicar la ANOVA multivariada. Se desea saber si los factores de la L1 de los alumnos y el tipo de estructura de las oraciones afectan la precisión con la que son capaces de calificar las oraciones como gramaticales o agramaticales y si esos dos factores interactúan en estudiantes que provienen de cuatro L1 diferentes y a los que se les hizo analizar palabras Wh- con tres estructuras diferentes (Larson-Hall, 2010, p. 140).

La prueba t puede utilizarse, por ejemplo, para validar reactivos de una escala Likert.

5.3. La prueba Chi-cuadrada

Larson-Hall (2010, p. 135) define la prueba Chi-cuadrada como aquella que “se utiliza cuando se busca una relación entre dos variables categoriales”, para conocer si hay una relación entre el número de sujetos en cada categoría. Es una prueba no paramétrica. Las características que establece para esta prueba son:

- Hay exactamente dos variables
- Cada variable tiene dos o más presentaciones, de tal forma que su nombre es una palabra que indica la categoría; por ejemplo: nivel de manejo de lengua
- No se pueden calcular promedios de estas variables
- Todas las variables son categóricas
- No se definen necesariamente como independientes o dependientes
- Hipótesis nula: Las variables son independientes, no están relacionadas

SPSS

- Analizar, pruebas no paramétricas, cuadros de diálogos antiguos, Chi-cuadrada
- Rango esperado, obtenerlo de los datos, pasar las variables en cuestión a la ventana central, Valores esperados, todas las categorías iguales, Aceptar
- Opciones, descriptivos, continuar
- Lista: Contrastar variables

De acuerdo con Martínez (2005), quien analiza los estados civiles: casado, divorciado, unión libre y nunca casado y los sexos masculino y femenino, los resultados son:

El programa genera por defecto dos tablas; la primera de ellas corresponde a la tabla de contingencia, en ella aparecen las variables seleccionadas y los estadísticos que se hayan determinado en el procedimiento. La segunda tabla corresponde a la prueba de Chi-cuadrado de Pearson y en ella aparecen los resultados de las pruebas (valor del Chi-cuadrado, los grados de libertad (gl) y el valor de significación (Sig.)).

La primera se presenta en el siguiente cuadro.

Cuadro 5.16. Variables seleccionadas

		Género	
		Femenino	Masculino
		Recuento	Recuento
Estado civil	Casado	139	127
	Viudo	32	5
	Divorciado	12	13
	Unión libre	6	3
	Nunca casado	23	40

Fuente: Martínez (2005).

La segunda se encuentra en el siguiente cuadro.

Cuadro 5.17. Resultados de la prueba Chi-cuadrada

Pruebas de Chi-cuadrada de Pearson

		Género
Estado civil	Chi-cuadrada	24520
	gl.	4
	Sig.	.000*

Fuente: Martínez (2005).

Los resultados se basan en filas y columnas no vacías de cada subtabla más al interior.

*El estadístico de Chi-cuadrada es significativo en el nivel .05.

El valor de la significancia es menor a .05, por lo que se rechaza la hipótesis nula, lo que significa que sí hay relación entre las variables estado civil y género.

Esta información puede traspolarse al área de lenguas, considerando el nivel de lengua de los sujetos, de acuerdo con el Marco Común Europeo de Referencia (MCER) para la lengua y su sexo. Los resultados se presentarían como:

Cuadro 5.18. Variables seleccionadas

		Género	
		Femenino	Masculino
		Recuento	Recuento
Nivel de lengua según el MCER para la lengua	A1, acceso	139	127
	A2, plataforma	32	5
	B1, intermedio	12	13
	B2, intermedio alto	6	3
	C1 Dominio operativo eficaz	23	40

Fuente: traspolado de Martínez (2005).

La segunda se encuentra en el siguiente cuadro.

Cuadro 5.19. Resultados de la prueba Chi-cuadrada

Pruebas de Chi-cuadrada de Pearson

		Género
		Nivel de lengua según el MCER para la lengua
	Chi-cuadrada	
	gl.	4
	Sig.	.000*

Fuente: traspolado de Martínez (2005).

Los resultados se basan en filas y columnas no vacías de cada subtabla más al interior.

*El estadístico de Chi-cuadrada es significativo en el nivel .05.

Para este ejemplo sería un valor de la significancia menor a .05, por lo que se rechaza la hipótesis nula, lo que significa que sí hay relación entre las variables nivel de manejo de lengua y género. La tabla de valores de esta prueba se encuentra en el anexo G.

5.4. U. de Mann Whitney

La prueba U. de Mann Whitney “es una prueba alternativa no paramétrica para la prueba t de muestras independientes. Se usa cuando se tiene una variable independiente con justamente dos valores y una variable dependiente, también para dos grupos. Es esencialmente la misma que la prueba de suma de rangos de Wilcoxon” (Larson-Hall, 2010, p. 376, 395). Se utiliza cuando no se cumple la característica de normalidad.

Hipótesis nula: No hay diferencias entre las medias de dos muestras, de dos grupos independientes.

SPSS

- Analizar, estadística no paramétrica, cuadro de diálogos antiguos, 2 muestras independientes
- Se pasan las variables en estudio a la ventana central. Se indica la variable de agrupación, en la ventanilla denominada como tal. Definir grupos
- Indicar los valores de la variable que define los grupos, Continuar
- Se elige prueba U de Mann Whitney
- Aceptar

Como resultado se obtienen dos cuadros.

Cuadro 5.20. Resultados 1. Prueba U de Mann Whitney

Rangos				
	Grupo	N	Rango promedio	Suma de rangos
Puntaje total A	1	19	22.29	423.50
	2	26	20.86	479.50
	Total	42		

Cuadro 5.21. Resultados 2. Prueba U de Mann Whitney

Estadísticos de contraste ^a	
	Puntaje total IM
U de Mann-Whitney	203.500
W de Wilcoxon	479.500
Z	-.379
Sig. asintót. (bilateral)	.704

a. Variable de agrupación: grupo.

En este caso, el nivel de significancia es de .704. Por lo tanto, se acepta la hipótesis nula. No hay diferencia entre las medias de los dos grupos.

5.5. Kruskal-Wallis

La prueba Kruskal-Wallis es la contraparte no paramétrica de la ANOVA de una vía. Debe usarse cuando se tiene una variable independiente con

tres o más niveles y una variable dependiente (Larson-Hall, 2010, p. 395); esta prueba es una extensión de la prueba MNN-Whitney para el caso de más de dos grupos, el tipo de rango que tiene lugar en esta prueba y la primera tabla de los resultados mostrará los rangos de los grupos (Larson-Hall, 201, p. 378).

Hipótesis nula: No hay diferencias entre las medias de dos muestras, de dos grupos independientes.

SPSS

- Analizar, estadística no paramétrica, cuadro de diálogos antiguos, K muestras independientes
- Se pasan las variables en estudio a la ventana central. Se indica la variable de agrupación, en la ventanilla denominada como tal. Definir grupos (definir rango)
- Indicar los valores de la variable que define los grupos, rango, mínimo y máximo. Continuar
- Se elige prueba H de Kruskal-Wallis
- Aceptar

Como resultado se obtienen dos cuadros.

Cuadro 5.22 Resultados 1. Prueba Kruskal-Wallis en SPSS

Rangos				
	Grupo	N	Rango promedio	
Puntaje total IM	Dimensión 1	1	19	77.87
		2	23	71.09
		3	15	79.50
		4	23	96.26
		5	18	75.31
		6	18	81.36
		7	18	80.28
		8	20	57.45
	Total	154		

Cuadro 5.23. Resultados 2. Prueba Kruskal-Wallis

Estadísticos de contraste ^{a,b}	
	Puntaje total IM
Chi-cuadrada	8.873
Gl	7
Sig. asintót.	.262

a. Prueba de Kruskal-Wallis.

b. Variable de agrupación: grupo.

De igual manera, en el caso presentado, el nivel de significancia es de .262. Por lo tanto, se acepta la hipótesis nula, "No hay diferencia entre las medias de los dos grupos.

5.6. McNemar

Larson-Hall define que "La prueba McNemar es una prueba no paramétrica que puede usarse cuando todas las variables contienen datos categóricos (nominales)" (2010. p. 396).

La hipótesis nula plantea que no hay diferencias entre los valores de la variable en estudio, antes y después de un tratamiento. Por ejemplo: la actitud hacia la importancia de una lengua extranjera.

SPSS

- Analizar, estadística no paramétrica, cuadro de diálogos antiguos, 2 muestras relacionadas
- Se pasan las variables en estudio a la ventana central, estableciendo el par o los pares de variables,
- Se elige prueba McNemar
- Aceptar

Los resultados son los siguientes.

Cuadro 5.24. Resultados 1 de la prueba McNemar.
Prueba de McNemar. Tablas de contingencia

Actitud antes		Actitud después	
		a favor	en contra
Dimensión	a favor	8	0
	en contra	20	8

Los resultados del cuadro 5.24 significan que en la primera prueba ocho estaban a favor y en la segunda también. Igualmente, para 8 que estaban en contra en la primera y en la segunda. Ninguno de los que estaban a favor se cambió a en contra o viceversa y 20 que estaban en contra se convirtieron a favor en la segunda ocasión.

Cuadro 5.25. Resultados 2 de la prueba McNemar

Estadísticos de contraste ^b	
	Actitud antes y Actitud después
N	36
Sig. exacta (bilateral)	.000 ^a

a. Se ha usado la distribución binomial.

b. Prueba de McNemar.

El nivel de significancia es de .000, por lo cual se rechaza la hipótesis nula, sí hay diferencias antes-después.

5.7. Wilcoxon

Las pruebas Wilcoxon y McNemar son dos pruebas no paramétricas que se utilizan como alternativa de la prueba t para muestras relacionadas (Larson-Hall, 2010, p. 381). La prueba McNemar se encuentra en el inciso anterior.

Existen dos variantes para la prueba Wilcoxon: la prueba Wilcoxon Rank-sum y la Wilcoxon signed-ranks.

Larson-Hall (2010, p. 404) define la prueba Rank-sum como una prueba no paramétrica alterna a la prueba t para grupos independientes cuando se tiene una variable independiente con dos valores y una variable dependiente e indica que funcionalmente es la misma que la prueba U de Mann-Whitney.

Define la Wilcoxon signed-ranks como una prueba no paramétrica alterna a la prueba t de muestras relacionadas que se usa cuando se tienen dos medias de puntajes que se relacionan, porque pertenecen al mismo grupo en dos momentos diferentes o al mismo grupo con mediciones tomadas por medio de dos instrumentos diferentes. Es decir, "se usa para comparar los dos momentos en los que se llevaron a cabo las pruebas" (Larson-Hall, 2010, p. 381).

La hipótesis nula plantea que no hay diferencias entre las medias.

SPSS

- Analizar, estadística no paramétrica, cuadro de diálogos antiguos, 2 muestras relacionadas
- Se pasan las variables en estudio a la ventana central, estableciendo el par o los pares de variables
- Se elige prueba Wilcoxon
- Aceptar

Los resultados son los siguientes.

Cuadro 5.26. Resultados 1. Prueba de los rangos con signo Wilcoxon

Rangos				
		N	Rango promedio	Suma de rangos
Nivel lengua comprensión oral A - Nivel lengua comprensión oral B	Rangos negativos	0 ^a	.00	.00
	Rangos positivos	36 ^b	18.50	666.00
	Empates	0 ^c		
	Total	36		

a. Nivel lengua comprensión oral A < Nivel lengua comprensión oral B.

b. Nivel lengua comprensión oral A > Nivel lengua comprensión oral B.

c. Nivel lengua comprensión oral A = Nivel lengua comprensión oral B.

Los rangos negativos implican que los sujetos tuvieron puntajes más bajos en la segunda medición. Los rangos positivos implican que los puntajes fueron más altos en la segunda medición. Los denominados empates indican que no hubo diferencias. En el cuadro 5.26 se ve que los 36 casos tuvieron puntajes más altos en la segunda medición.

Cuadro 5.27. Resultados 2. Prueba de los rangos con signo Wilcoxon

Estadísticas de contraste^b

	Nivel lengua comprensión oral A - Nivel lengua comprensión oral B
Z	-5.346 ^a
Sig. asintót. (bilateral)	.000

a. Basado en los rangos negativos.

b. Prueba de los rangos con signo de Wilcoxon.

Al haber un nivel de significancia de .000, se rechaza la hipótesis nula. Por lo tanto, sí existen diferencias entre las variables comparadas.

5.8. Q de Cochran

Esta prueba no paramétrica se utiliza para k muestras de la variable independiente, cuando ésta tiene valores dicotómicos.

La hipótesis nula plantea que no existe diferencia significativa entre las muestras.

SPSS

- Analizar, estadística no paramétrica, cuadro de diálogos antiguos, k muestras relacionadas
- Se pasan las variables en estudio a la ventana central
- Se elige prueba Q de Cochran
- Estadísticos
- Descriptivos, Continuar
- Aceptar

Los resultados son los siguientes.

Cuadro 5.28. Resultados 1 de la prueba Q de Cochran

Pruebas no paramétricas

[Conjunto_de_datos1] F:\archivo para libro ejemplos.sav

Estadísticos descriptivos

	N	Media	Desviación típica	Mínimo	Máximo
Medición uno	36	1.78	.422	1	2
Medición dos	36	1.22	.422	1	2
Medición tres	36	1.78	.422	1	2
Medición cuatro	36	1.22	.422	1	2

Cuadro 5.29. Resultados 2 de la prueba Q de Cochran

Prueba de Cochran

Frecuencias

	Valor	
	1	2
Medición uno	8	28
Medición dos	28	8
Medición tres	8	28
Medición cuatro	28	8

Cuadro 5.30. Resultados 3 de la prueba Q de Cochran

Estadísticos de contraste

N	36
Q de Cochran	60.000 ^a
Gl	3
Sig. asintót.	.000

a. 2 se trata como un éxito.

5.9. Friedman

La prueba Friedman, denominada Prueba ANOVA RM de una vía, es una prueba no paramétrica utilizada en vez de la prueba Anova de una sola vía. En este caso, hay una variable independiente con mediciones repetidas y una variable dependiente (Larson-Hall, 2010, p. 393). Recomienda utilizarla cuando hay dos mediciones de la variable independiente (si estas fueran mediciones repetidas, podría usarse una prueba t de muestras relacionadas). “La prueba paramétrica Anova RM puede utilizarse con cualquier número de variables independientes pero la alternativa no paramétrica para cuando hay justamente una variable independiente es usar solamente la prueba Friedman” (p. 383).

Esta prueba puede utilizarse en aquellos casos en los que se cuenta con un número k de muestras relacionadas.

La hipótesis nula es que los grupos o las medidas repetidas provienen de la misma población.

SPSS

- Analizar, estadística no paramétrica, cuadro de diálogos antiguos, k muestras relacionadas
- Se pasan las variables en estudio a la ventana central
- Se elige prueba Friedman
- Estadísticos
- Descriptivos, Continuar
- Aceptar

Los resultados son los siguientes.

Cuadro 5.31. Resultados 1 de la prueba Friedman

Pruebas no paramétricas
[Conjunto_de_datos1] F:\archivo para libro ejemplos.sav
Estadísticos descriptivos

	N	Media	Desviación típica	Mínimo	Máximo
Medición uno	36	21.89	2.278	17	25
Medición dos	36	21.97	2.813	14	25
Medición tres	36	19.36	2.620	14	24
Medición cuatro	36	22.92	1.574	20	25

Cuadro 5.32. Resultados 2 de la prueba Friedman

Prueba de Friedman

	Rango promedio
Medición uno	2.63
Medición dos	2.78
Medición tres	1.42
Medición cuatro	3.18

Cuadro 5.33. Resultados 3 de la prueba Friedman

Estadísticos de contraste^a

N	36
Chi-cuadrada	40.146
Gl	3
Sig. asintót.	.000

a. Prueba de Friedman

En este caso, el nivel de la significancia es de .000. Por lo tanto, se rechaza la hipótesis nula; los datos provienen de poblaciones distintas. Si el valor calculado es mayor que el de tabla, se rechaza la hipótesis nula.

Como conclusión de este capítulo, para la interpretación de los resultados de las pruebas estadísticas, es fundamental tener presente que hay una hipótesis de investigación y una hipótesis nula.

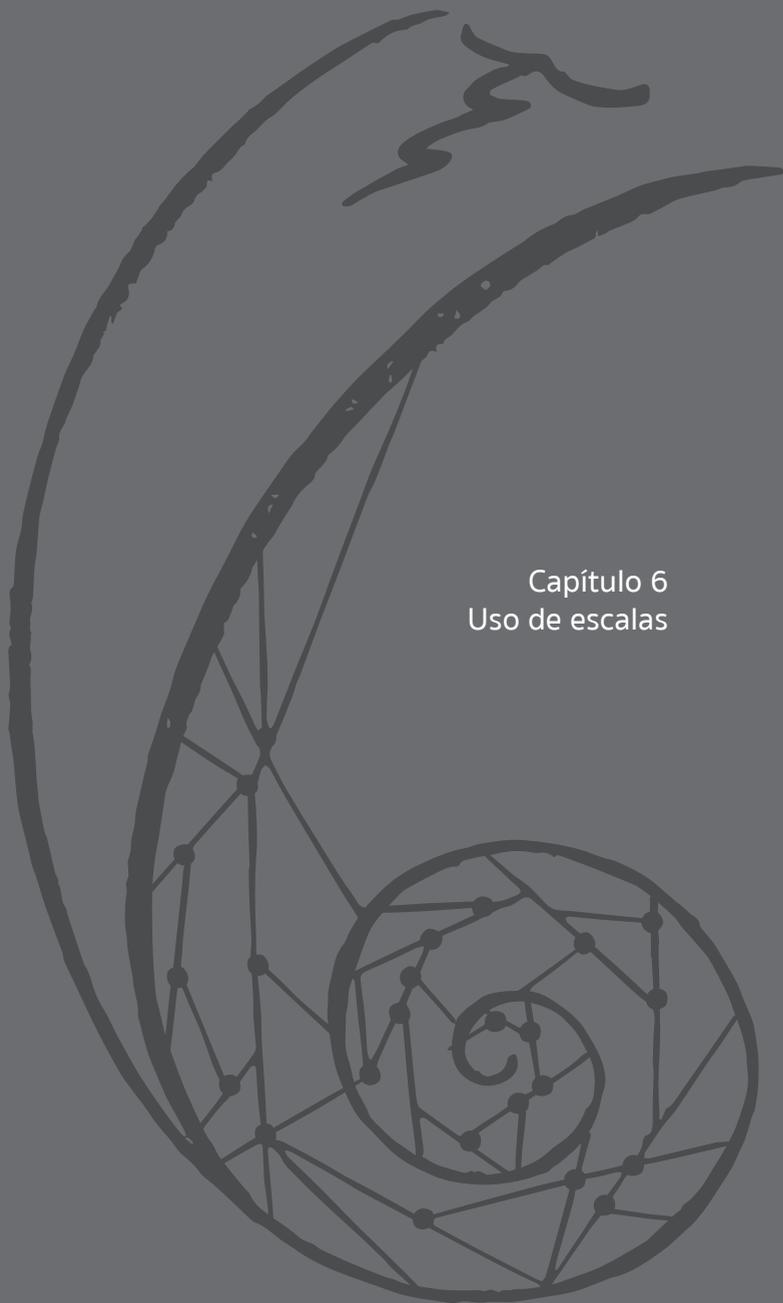
En los procesos de las pruebas analizadas, lo indicado es considerar el valor de la significancia, el cual indica si se acepta o se rechaza la hipótesis nula; hipótesis que plantea, según la prueba en proceso, que no hay relación ni correlación o diferencia entre variables o medias;

hay igualdad de varianzas; los datos siguen una distribución normal; un modelo no tiene valor predictivo o un tratamiento no produce cambios significativos en la variable en estudio. Ésta se acepta cuando el valor de la significancia es mayor a .05. La aceptación de la hipótesis nula significa que no se acepta la hipótesis de investigación.

Cuando se rechaza la hipótesis nula, para algunas pruebas, se tiene un valor de la prueba misma, denominado valor calculado, el cual se compara con un valor de tabla. Para aceptar la hipótesis de investigación el valor calculado en el proceso debe ser mayor que el valor en la tabla correspondiente.



Capítulo 6
Uso de escalas





La escala como instrumento de levantamiento de datos se denomina como tal porque las respuestas a sus reactivos se establecen en una escala de valores elegidos por los sujetos. Aunque la construcción de las escalas en general sigue los mismos patrones, es importante señalar que existe un tipo de escala que se diseña bajo algunas consideraciones particulares. Por ello, se incorporan las escalas en general y la construcción específica de las denominadas Escalas de Likert. A continuación se presentan ambos casos.

6.1. Escalas

Como se mencionó, las escalas contemplan valores elegidos por los sujetos. El número de respuestas puede variar en un rango de dos o más. En ocasiones se tiene hasta siete respuestas posibles. Por ejemplo: un reactivo cuyo contenido es "la lengua inglesa es útil para viajar por el mundo", en el que los sujetos pueden elegir como respuesta: totalmente en desacuerdo (TD), en desacuerdo (D), ni de acuerdo ni en desacuerdo (NAND), de acuerdo (A) y totalmente de acuerdo (TA) o bien, el reactivo "platico con hablantes nativos de la lengua francesa", con las respuestas: nunca, alguna vez, frecuentemente.

El sujeto que recibe la escala califica el contenido del reactivo de acuerdo con su propio criterio, eligiendo una respuesta o un código alfabético que corresponde a cada una de las elecciones de las respuestas posibles. Por ejemplo:

Cuadro 6.1. Ejemplos de reactivos de una escala con cinco posibles respuestas

No. de reactivo	Contenido	TD	D	NA ND	A	TA
1	La lengua inglesa es útil para viajar por el mundo					
2	La lengua inglesa es una lengua fácil de aprender					
3	Hay muchas oportunidades para practicar la lengua inglesa					

En este caso, los sujetos marcan la casilla de la respuesta elegida. Para el proceso de los datos recabados con la escala, cada elección tiene un valor numérico. Este valor numérico puede asignarse en diversas formas. La más común es dar valores a partir del uno. En este caso, se plantea como: 1= totalmente en desacuerdo, 2= en desacuerdo, 3= ni de acuerdo ni en desacuerdo, 4= de acuerdo y 5= totalmente de acuerdo. Esto significa que cada reactivo puede ser calificado con un mínimo de uno y un máximo de cinco. De esta forma, el puntaje que puede obtenerse para la prueba completa, tendrá un rango de valores entre el número total de reactivos por uno y el número total de reactivos por cinco. Por ejemplo: Si el número total de reactivos es de 30, el valor mínimo a alcanzar por cada sujeto es de 30 (un valor de uno para cada uno de los treinta reactivos) y el valor máximo es de 150 (el resultado de un valor de 5 para cada uno de los treinta reactivos, 5×30).

Existen dos modalidades en las que puede utilizarse una escala. Si se asignan todos los reactivos de la escala para la medición de un elemento específico de la variable medida, al final, el análisis de los datos levantados podrá hacerse por reactivo y por el puntaje total obtenido en la prueba. La otra modalidad consiste en diseñar la escala con la medición de varias categorías o bien, varios componentes de la variable analizada. En este caso se consideran grupos de reactivos organizados por subescalas, de tal forma que, al hacer el análisis de los datos habrá resultados por reactivo, por subescala y por la escala total, lo cual enriquece las posibilidades de análisis del fenómeno.

Para la etapa de piloteo y confiabilización de la escala se utiliza una prueba t. Esta prueba se aplica a cada uno de los reactivos del instrumento que se somete a prueba.

Procedimiento:

- Se calcula el puntaje obtenido por cada uno de los sujetos que participaron en el piloteo
- Se clasifican las pruebas en orden descendente del puntaje obtenido
- Se toma el 25% de las pruebas con los puntajes más altos. Este conjunto de pruebas constituye el denominado grupo alto
- Se toma el 25% de las pruebas con los puntajes más bajos. Este conjunto de pruebas constituye el denominado grupo bajo
- Se desecha el 50% de las pruebas ubicadas en el centro
- Se aplica la prueba t para grupos independientes, para cada uno de los reactivos

Procedimiento manual para hacerlo:

- Se obtiene la media para cada reactivo, considerando cada uno de los dos grupos, alto y bajo
- El resultado es la media del grupo alto, M_A , y la media del grupo bajo, M_B , para cada uno de los reactivos
- Se obtiene la varianza para cada reactivo considerando cada uno de los dos grupos, alto y bajo
- El resultado es la varianza del grupo alto, V_A , y la varianza del grupo bajo, V_B , para cada uno de los reactivos
- La fórmula de la prueba t para cada reactivo es:

$$t = \frac{M_A - M_B}{\sqrt{(V_A/N_1) + (V_B/N_2)}}$$

- N_1 se refiere al número de sujetos del grupo alto y N_2 a los sujetos del grupo bajo
- Los grados de libertad se calculan con $N_1 + N_2 - 2$
- Se busca el valor de la t de tabla (Ver anexo E)
La hipótesis para esta prueba es que existen diferencias entre los grupos alto y bajo.
- Si la t calculada es mayor o igual a la de la tabla significa que si hay diferencias entre el grupo alto y el grupo bajo. Por lo tanto, el reactivo en cuestión es confiable

- Si la t calculada es menor a la de la tabla, el reactivo no se utiliza en la prueba final
También puede utilizarse el SPSS. Por ejemplo, para probar la confiabilidad de los reactivos de una escala de 30 reactivos, aplicada a 24 sujetos.
- Se crea un archivo de datos en el SPSS con las respuestas de los 24 sujetos para los 30 reactivos
- Se calcula el puntaje logrado por cada sujeto y se ubica en una columna creada para ello (ver el inciso 8.8 Calcular el valor de una variable)
- Se clasifican los datos usando como parámetro de clasificación la columna del puntaje, en orden descendente

SPSS

- Datos
- Ordenar, descendente
- Se indica que la columna utilizada para clasificar los datos es la del puntaje

Una vez clasificado el archivo, se eliminan los sujetos que ocupan el 50% de ellos en la parte central. En el ejemplo, se guardan los sujetos clasificados que corresponden a las líneas 1 a 6 y 19 a 24 de la plantilla de datos. Se eliminan los sujetos 7 a 18. El archivo queda solamente con 12 sujetos, los seis más altos y los seis más bajos. Se eliminan los 12 sujetos del centro. Para poder aplicar la prueba t para muestras independientes, es necesario crear una columna en el archivo para indicar con un uno los del grupo alto y con un dos los del grupo bajo. Como la escala tiene 30 reactivos, hay que procesar datos para 30 pruebas t , una por cada uno de los reactivos, en la siguiente forma:

SPSS

- Analizar, Comparar medias, prueba t para muestras independientes
- Se pasa la variable que corresponde al reactivo en proceso a la ventana central. Se indica la variable de agrupación, en la ventanilla denominada como tal. Definir grupos
- Indicar los valores 1 y 2 de la variable que define los grupos, Continuar
- Aceptar

Como resultado, se obtiene el nivel de significancia y el valor de t que permiten aceptar o rechazar el reactivo, con la hipótesis de investigación

“Existen diferencias en las respuestas al reactivo por los sujetos del grupo alto y del grupo bajo”. La hipótesis nula establece que no hay diferencias (consultar el inciso 5.1). Una vez obtenidos los resultados para los 30 reactivos, la escala definitiva será construida solamente con los reactivos cuyo valor de la t calculada haya sido mayor o igual a la t de la tabla (ver anexo E).

6.2. Escala ordinal

“La escala ordinal es una escala que mide una propiedad con respecto a otras (*ranking*) pero en la que pueden no ser iguales los peldaños (*steps*) entre las mediciones. Por ejemplo mediciones de ansiedad hacia la lengua o cantidad de uso de la lengua” (Larson-Hall, 2010, p. 397).

6.3. Escalas de Likert

Es ineludible evitar la confusión que existe en el uso de escalas. El hecho de crearla con un rango de respuestas para los reactivos que la componen no significa que se esté haciendo uso de una escala Likert. En principio, las escalas Likert fueron creadas para la medición de actitudes. Siendo la actitud una disposición favorable o desfavorable hacia algo, se construyen con el mismo número de reactivos favorables y desfavorables hacia el objeto de estudio en cuestión.

La escala Likert es un tipo particular de escala creada para la medición de actitudes. También se conoce como “Método de rangos sumariados” (Nadelsticher, 1983, p. 17). Se hace siguiendo los mismos principios de una escala con algunas precisiones adicionales, principalmente en cuanto a la redacción y balance de los reactivos que componen el instrumento. Esta escala fue inventada por Rensis Likert, al inicio de los años treinta del siglo XX. Nadelsticher (1983, p. 16-17) puntualiza que al construirlas deben considerarse las características de una escala de actitudes, constituidas por afirmaciones, nunca por preguntas. Recomienda las características de estas afirmaciones que se mencionan a continuación.

- Evitar frases que se refieran al pasado en vez de presente
- Evitar frases que puedan ser interpretadas en más de un sentido
- Evitar afirmaciones que no tengan relación con el objeto psicológico medido
- Evitar oraciones en las que nadie o todos estarían de acuerdo

- Utilizar un lenguaje claro, simple y directo
- Los reactivos deben ser cortos, de no más de veinte palabras
- Cada reactivo debe contener sólo una idea
- Evitar frases que contengan universalidad, como todos, siempre, nunca, ninguno
- Debe evitarse el uso de palabras como simplemente
- Evitar palabras que puedan provocar equivocasiones
- Evitar el empleo de frases negativas complejas (Marín, 1975)

Nadelsticher comenta que “Aunque hay discusiones en torno a la escala de medición que emplean las escalas de actitudes, se ha optado por considerarlas como intervalares” (1983, p. 16). En relación con el contenido de las afirmaciones, Morosini (2012) establece que los reactivos deben ser juicios de valor o comportamiento, no hechos.

En cuanto al balance de los reactivos que componen una escala, hay que iniciar diciendo que las actitudes se definen como disposiciones favorables o desfavorables hacia algo. Por ello, las escalas Likert se diseñan tomando en consideración reactivos que contemplan actitudes favorables y desfavorables hacia el objeto de estudio. Esto significa que debe haber un equilibrio entre el número de reactivos con afirmaciones favorables y el número de ellos con afirmaciones desfavorables. Esta distinción entre estos dos tipos de reactivos, los favorables y los desfavorables, conduce a una asignación distinta de valores a las respuestas de cada tipo de reactivo, asignación que se muestra en el siguiente cuadro.

Cuadro 6.2. Asignación de valores a las respuestas de los reactivos de una escala Likert

Tipo de reactivo	Valores de las respuestas				
	Totalmente en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Totalmente de acuerdo
Favorable	1	2	3	4	5
Desfavorable	5	4	3	2	1

Si el reactivo favorable al objeto de estudio es calificado con “Totalmente de acuerdo”, su valor es el valor máximo por reactivo, el 5. El contenido favorable es completamente aceptado. En cambio, si el reactivo que es desfavorable al objeto de estudio, es calificado con “Totalmente de acuerdo”, su valor es solamente de 1. Respecto a las actitudes hacia la lengua inglesa se obtuvo lo siguiente.

Cuadro 6.3. Contraste entre los valores de un reactivo favorable al objeto de estudio y uno desfavorable

Reactivo	Valores de las respuestas				
	Totalmente en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Totalmente de acuerdo
Favorable La lengua inglesa es útil para viajar fuera de México	1	2	3	4	5
Desfavorable La lengua inglesa es inútil para viajar fuera de México	5	4	3	2	1

Al elegir la respuesta “Totalmente de acuerdo” para el reactivo “La lengua inglesa es inútil para viajar fuera de México”, no se concede valor a la lengua inglesa. Por lo tanto, el valor de esta respuesta es el mínimo posible para un reactivo, el valor de uno.

Toda esta asignación de valores según el tipo de reactivo respalda el equilibrio mencionado entre el número de reactivos de cada tipo. La ubicación de los reactivos es aleatoria. Se intercalan los reactivos favorables y desfavorables sin un orden particular. Por ejemplo, en el siguiente cuadro, se muestra una subescala de actitudes hacia la importancia concedida a la lengua inglesa, con cuatro reactivos. Los valores en términos del proceso de datos se asignan en la siguiente forma.

Cuadro 6.4. Subescala de actitudes hacia la importancia de la lengua inglesa, con cuatro reactivos

Reactivo	Valores de las respuestas				
	Totalmente en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Totalmente de acuerdo
(Favorable) La lengua inglesa es un medio de comunicación universal	1	2	3	4	5
(Desfavorable) La lengua inglesa nos quita nuestra identidad	5	4	3	2	1
(Desfavorable) El conocimiento de la lengua inglesa es innecesario	5	4	3	2	1
(Favorable) La lengua inglesa permite consultar bibliografía actualizada	1	2	3	4	5

Es preciso señalar que la escala se presenta a los sujetos en estudio con respuestas expresadas con letras. Los sujetos no conocen las equivalencias numéricas de las mismas. El siguiente cuadro muestra la misma subescala en la presentación que se entrega a los sujetos.

Cuadro 6.5. Subescala de actitudes hacia la importancia de la lengua inglesa con cuatro reactivos, en formato para su aplicación a los sujetos en estudio

Reactivo	Valores de las respuestas				
	Totalmente en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Totalmente de acuerdo
(Favorable) La lengua inglesa es un medio de comunicación universal	TD	D	NA-ND	A	TA
(Desfavorable) La lengua inglesa nos quita nuestra identidad	TD	D	NA-ND	A	TA
(Desfavorable) El conocimiento de la lengua inglesa es innecesario	TD	D	NA-ND	A	TA
(Favorable) La lengua inglesa permite consultar bibliografía actualizada	TD	D	NA-ND	A	TA

Para la confiabilización de la escala, Morosini (2012) recomienda tres cálculos de índices: de homogeneidad, de discriminación y de consistencia interna. El índice de homogeneidad utiliza la correlación entre reactivo y puntaje total, según Pearson y la correlación reactivo y puntaje total corregida. La correlación entre los reactivos y puntaje total se calcula como sigue:

SPSS

- Analizar, correlaciones bivariadas, Pearson
- En las variables, se indican todos los reactivos y la columna del puntaje total

Como resultado se tiene una matriz de correlaciones de los reactivos con ellos mismos y con el puntaje total. El índice de discriminación, basado en la consistencia interna de la escala, utiliza los puntajes de los sujetos, mediante el procedimiento de uso de la prueba t con el 25% de los puntajes más altos y el 25% de los más bajos. El índice de consistencia interna se evalúa mediante el alfa de Cronbach (ver el inciso 4.2.1.).

En el SPSS, se utiliza lo siguiente:

SPSS

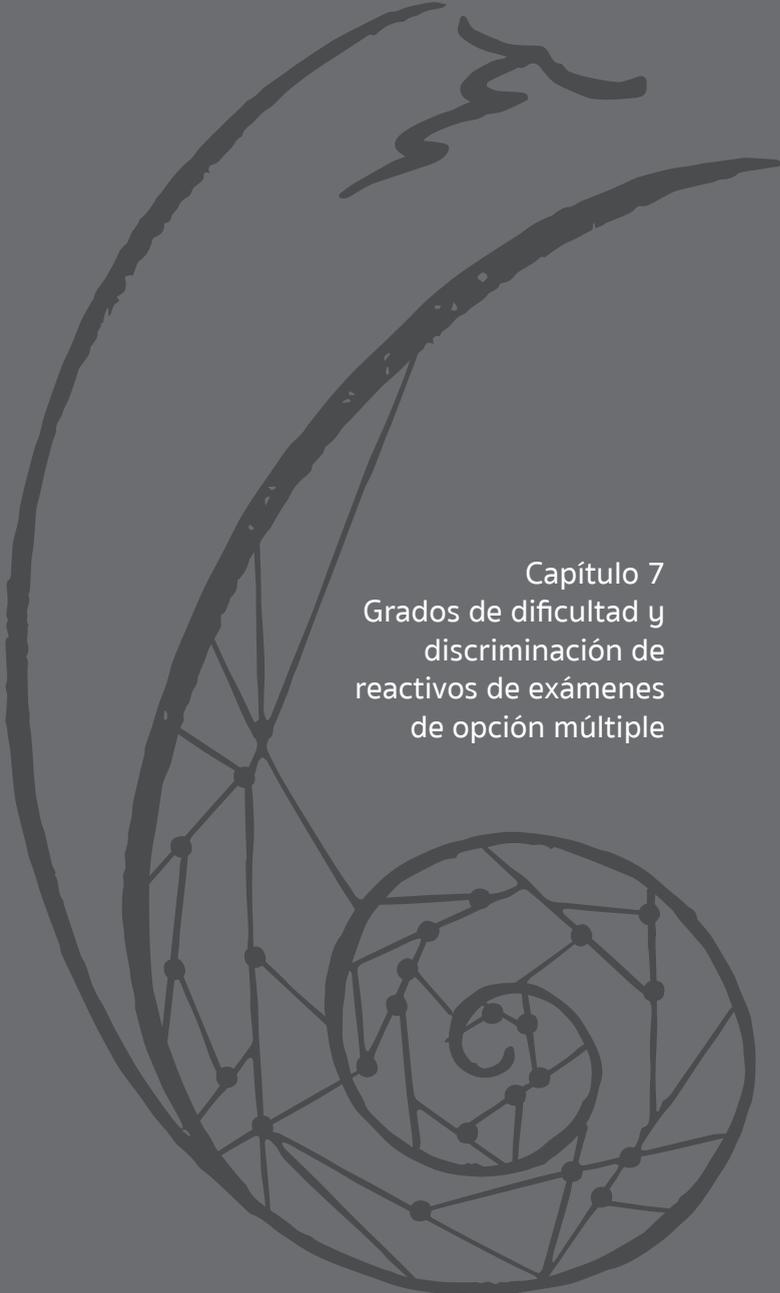
- Analizar, escala, análisis de fiabilidad
- Modelo Alfa, Pasar las variables al centro, escribir etiqueta para el nombre de la escala, Indicar modelo: Alfa
- Aceptar

Una vez capturados los datos de cada sujeto y reactivo, como en cualquier escala, lo que se usa para validar los resultados es la prueba t. En la misma forma que en el procedimiento mostrado en el inciso 6.1, se obtiene el puntaje de la prueba de cada sujeto. Se clasifica la columna de los puntajes en orden descendente, de puntaje mayor a menor. Para trabajar, se toma el 25% de los sujetos más altos y el 25 % de los sujetos más bajos. Si el número de sujetos es bajo entonces se toma el 33% de los más altos y el 33% de los más bajos. Es decir, se eliminan los sujetos centrales (50% ó 33%, según sea el caso).

En este capítulo lo más relevante, que merece toda nuestra atención, es el análisis de las características muy propias de la Escala de Likert. Como se señaló, el hecho de usar una escala como instrumento de levantamiento de datos no significa que se trata de una escala del tipo Likert.

En el capítulo siguiente se analiza el procedimiento de Madsen (1983) para el análisis de reactivos de opción múltiple. Este procedimiento y el utilizado para las escalas siguen el mismo proceso en cuanto a la selección de dos grupos, con los puntajes más altos y más bajos. Se describen en cada uno de los dos capítulos porque la segunda etapa de cada proceso es diferente.





Capítulo 7
Grados de dificultad y
discriminación de
reactivos de exámenes
de opción múltiple



Este capítulo se refiere a la confiabilización de pruebas cuyos reactivos se construyen con respuestas de opción múltiple. Al efectuar investigaciones en el salón de clase de lengua, con frecuencia, los instrumentos de medición de nivel de lengua; de establecimiento del perfil de los sujetos en estudio; o del avance en el conocimiento de una lengua, son de opción múltiple.

El uso de estos instrumentos implica un análisis de sus reactivos que garantice que miden lo que deben medir y que la prueba en general es confiable.

Los reactivos deben cumplir dos requisitos: grado de dificultad y grado de discriminación. El reactivo no debe caer en los extremos; es decir, que no sea tan fácil que todos los sujetos elijan la respuesta correcta ni tan difícil que nadie lo conteste. De esta manera, los grados de dificultad y discriminación de los reactivos de exámenes de opción múltiple son dos parámetros que, conjugados, constituyen indicadores de la confiabilidad de los reactivos que los conforman.

Para calcular estos parámetros, se utiliza el procedimiento establecido por Madsen (1983, pp. 178-184). Este autor propone una forma de confiabilizar los reactivos de un examen de tal forma que las evaluaciones que genera conduzcan a una medición más precisa de las habilidades de los alumnos. Señala que una de las características más importantes de los exámenes es que con ellos se obtenga un puntaje más alto por los estudiantes más avanzados y un puntaje menor por los más bajos.

Considera tres características que deben tener los exámenes: validez, confiabilidad y afecto. Define la validez en el sentido de que el examen mide lo que debe medir: la confiabilidad como "la prueba produce esencialmente los mismos resultados consistentes en diversas ocasiones

cuando las condiciones de la prueba permanecen iguales” (Madsen, 1983, p. 179); en el afecto, toma en cuenta el nivel en el que la prueba provoca ansiedad excesiva.

El procedimiento para calcular los grados de dificultad y de discriminación, propuesto por Madsen (1983, pp. 180-184), denominado como “Análisis de reactivos”, es el que se presenta a continuación, reportado paso a paso, el cual recomienda precisamente para exámenes de opción múltiple. Para favorecer un uso más sencillo, se reportan los pasos del procedimiento divididos en cuatro partes: la preparación de los datos, el grado de dificultad de los reactivos, el grado de discriminación y la evaluación de los distractores.

7.1. Preparación de los datos

- Obtener el puntaje total de todas las prueba aplicadas
- Ordenar las pruebas en orden descendente del puntaje total obtenido
- Dividir las pruebas ordenadas en tres grupos de igual número de sujetos. Separar el grupo de puntajes más altos y el de puntajes más bajos para tomar los más altos y los más bajos y desechar los del grupo central. Siempre se desecha el grupo central. La separación también podría hacerse tomando el 27% de los más altos y el 27% de los más bajos
- El grupo de puntajes más altos es denominado el grupo alto y el de los más bajos, el grupo bajo
- Hacer una cuadro de cuatro columnas con el número de cada reactivo, una columna para las respuestas, una columna para el grupo alto y una columna para el grupo bajo
- En las filas de cada reactivo se indica las posibles respuestas de opción múltiple; se agrega una fila para los casos sin respuesta

El cuadro es el siguiente:

Cuadro 7.1. Organización de los datos para el análisis de los reactivos

Número de reactivo	Respuestas	Grupo alto	Grupo bajo
Reactivo # 1	A		
	B		
	C		
	D		
(sin respuesta)	(X)		

Reactivo # 2 (sin respuesta)	A		
	B		
	C		
	D		
	(X)		
			
Reactivo # n (último reactivo de la prueba) (sin respuesta)	A		
	B		
	C		
	D		
	(X)		

Fuente: adaptado de Madsen (1983, p. 180).

- Para cada reactivo, se identifica la respuesta correcta y se marca en la tabla el número de elecciones hechas de cada una de las respuestas, contando por separado las respuestas del grupo alto y del grupo bajo. El cuadro queda de la siguiente manera.

Cuadro 7.2. Registro de los datos para el análisis de los reactivos

Número de reactivo	Respuestas	Grupo alto	Grupo bajo
Reactivo # 1 (sin respuesta)	A	1	3
	■ B	5	2
	C	2	1
	D	1	2
	(X)	1	2
Reactivo # 2 (sin respuesta)	A	1	2
	B	1	2
	■ C	6	2
	D	1	2
	(X)	1	2
			
Reactivo # n (último reactivo de la prueba) (sin respuesta)	■ A	4	2
	B	2	3
	C	2	1
	D	1	2
	(X)	13	2

Fuente: adaptado de Madsen (1983, pp. 180 y 181).

7.2. Grado de dificultad de los reactivos

Madsen (1983, p. 181) indica que el grado de dificultad de cada uno de los reactivos se calcula en la siguiente forma:

- Tomar el número de veces en las que el grupo alto seleccionó la respuesta correcta
- Sumarle el número de veces en las que el grupo bajo seleccionó la respuesta correcta
- Dividir el resultado de la suma entre el número total de sujetos, los del grupo alto más los del grupo bajo
- Se obtiene un porcentaje de los sujetos que contestaron correctamente
- Analizar si el reactivo es muy fácil o muy difícil

La forma de analizar la dificultad del reactivo se basa en que un reactivo es considerado muy fácil si más del 90% de los sujetos lo tienen bien, si es menor al 30%, se considera un reactivo muy difícil. "Este resultado del análisis proporciona información acerca de la instrucción y evalúa los propios reactivos de la prueba" (Madsen, 1983, p. 182).

7.3. Grado de discriminación de los reactivos

Para el grado o nivel de discriminación Madsen indica "qué tan bien se establece la diferencia entre aquellos que tienen un nivel de lengua más avanzado y aquellos con menor habilidad" (1983, p. 182). El grado de discriminación de cada reactivo se obtiene a partir del cuadro 7.2. En este caso, el procedimiento es:

- Tomar el número de veces en las que el grupo alto seleccionó la respuesta correcta
- Restarle el número de veces en las que el grupo bajo seleccionó la respuesta correcta
- Dividir el resultado de la resta entre el número total de sujetos, los del grupo alto menos los del grupo bajo
- Se obtiene un porcentaje de los sujetos que contestaron correctamente
- Analizar si el reactivo discrimina o no

La forma de interpretar el resultado es la siguiente: si el resultado es menor o igual al diez por ciento, el reactivo no es aceptable. Un 15% o más

es aceptable. Entre 10 y 15 es marginal o inaceptable. Madsen agrega que "cuando los reactivos no discriminan bien o son muy fáciles o muy difíciles, hay que analizar la escritura del reactivo, para encontrar la causa" (1983, p. 183).

7.4. Evaluación de los distractores

Por último, en el procedimiento para analizar la confiabilidad de los reactivos, Madsen analiza las características de los distractores. "Los distractores débiles provocan una discriminación pobre o un nivel de dificultad no deseado" (1983, p. 184). Un parámetro para decidir que un distractor no es útil es cuando no es elegido por al menos uno o dos de los sujetos a los que se les aplicó la prueba, en una muestra entre 10 y 20 sujetos.

También establece que los distractores no pueden "distracer", es decir, desviar la respuesta hacia el distractor, no puede ponerse un distractor en el formato de la respuesta correcta cuando los otros distractores se presentan en otro formato y no hay que utilizar distractores que claramente se ve que no lo son, es decir distractores que no tienen nada que ver con la respuesta.

7.5. Ejemplo

Se tienen diez reactivos de una prueba para medir el manejo de un punto gramatical de una lengua extranjera cuya respuesta puede ser A, B, C o D o sin respuesta (con la equivalencia numérica 1, 2, 3, 4, 5), respondidos por 39 sujetos. El proceso se lleva a cabo en dos etapas.

Etapas. Sirve para:

- Obtener el puntaje de cada uno de los sujetos participantes
- Clasificar los datos capturados de acuerdo con el puntaje obtenido por los sujetos, en orden descendente
- Seleccionar el 33% de los sujetos que obtuvieron los puntajes más altos
- Seleccionar el 33% de los sujetos que obtuvieron los puntajes más bajos
- Desechar el 34% de los sujetos que se situaron en medio de ellos

Etapas. En esta etapa se trabaja con los sujetos más altos y más bajos, denominados Grupo alto y Grupo bajo, respectivamente, para el cálculo de los grados de dificultad y de discriminación de cada uno de los reactivos.

Para la etapa uno, en el SPSS se crea la plantilla en la que la primera columna es el número de secuencia de los sujetos capturados, las siguientes 10 columnas pertenecen a cada uno de los aciertos en los reactivos; la penúltima, denominada PT, es el puntaje total obtenido en el examen de opción múltiple y la última, la columna que sirve para trabajar solamente con los sujetos seleccionados como grupo alto y grupo bajo.

En cada una de las columnas de los reactivos, se declaran los valores 0 y 1 para desaciertos y aciertos en las respuestas y el 2 para "sin respuesta". Para calcular el puntaje total de cada examen, se utiliza:

- Transformar, calcular variable
- Variable de destino, el nombre de la variable para el puntaje total (en este caso PT)
- Expresión numérica construida con la suma de cada reactivo
- En este caso, $R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 + R_{10}$, tomando cada reactivo de la ventana izquierda, arrastrado a la ventana de la expresión numérica
- Aceptar

El SPSS deja los valores de los puntajes en la columna destinada para ello (PT). A continuación está la tabla con una parte de los datos de la plantilla del ejemplo, los que corresponden a los primeros cinco sujetos. Las columnas indican los aciertos en los reactivos y el puntaje total y los renglones contienen cada uno de los sujetos. El máximo puntaje que puede alcanzarse es de 10.

Cuadro 7.3. Plantilla de respuestas a los reactivos y puntaje total

	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	R ₈	R ₉	R ₁₀	PT
S ₁	0	0	0	1	0	0	0	0	1	0	2
S ₂	0	0	0	0	0	1	0	0	0	0	1
S ₃	0	0	0	1	0	1	0	1	1	0	4
S ₄	0	0	0	0	0	0	0	0	0	0	0
S ₅	0	0	0	1	1	0	0	0	0	0	2

En este primer cálculo, los puntajes totales están ordenados por sujeto. Éstos son los siguientes, transformados en una tabla para su visualización:

Cuadro 7.4. Puntajes por sujeto

Sujeto	Puntaje	Sujeto	Puntaje	Sujeto	Puntaje
1	2	14	10	27	2
2	1	15	7	28	2
3	4	16	9	29	3
4	0	17	7	30	6
5	2	18	9	31	3
6	3	19	8	32	0
7	4	20	7	33	3
8	1	21	8	34	2
9	1	22	9	35	2
10	2	23	8	36	3
11	2	24	6	37	1
12	10	25	2	38	4
13	10	26	4	39	1

A continuación, se ordenan los resultados de acuerdo con el puntaje total obtenido.

SPSS

- Datos, ordenar casos
- Transferir del cuadro de la izquierda el nombre de la variable que identifica el puntaje total al cuadro central denominado Ordenar por
- Marcar orden descendente, Aceptar

El siguiente cuadro muestra los puntajes clasificados de esa forma.

Cuadro 7.5. Puntajes clasificados en orden descendente

Puntaje	Sujeto	Puntaje	Sujeto	Puntaje	Sujeto
10	12	6	30	2	11
10	13	4	3	2	25
10	14	4	7	2	27
9	16	4	26	2	38
9	18	4	38	2	34
9	22	3	6	2	35
8	19	3	29	1	2
8	21	3	31	1	8
8	23	3	33	1	9
7	15	3	36	1	37
7	17	2	1	1	39
7	20	2	5	0	4
6	24	2	10	0	32

La separación de los grupos alto y bajo es la siguiente. Se agrega una columna en la plantilla de datos del SPSS para indicar los sujetos que pertenecen a cada uno de los grupos alto, medio y bajo, con la indicación de 1 para el grupo alto, 2 para el grupo bajo y 3 para el grupo medio, de tal forma que posteriormente se pueda trabajar con los 13 sujetos con el puntaje más alto y los 13 con el puntaje más bajo. Los grupos quedan como se indica a continuación.

Cuadro 7.6. Puntaje total del grupo alto

Puntaje	Sujeto
10	12
10	13
10	14
9	16
9	18
9	22
8	19
8	21
8	23
7	15
7	17
7	20
6	24

Cuadro 7.7. Puntaje total del grupo medio

Puntaje	Sujeto
6	30
4	3
4	7
4	26
4	38
3	6
3	29
3	31
3	33
3	36
2	1
2	5
2	10

Cuadro 7.8. Puntaje total del grupo bajo

Puntaje	Sujeto
2	11
2	25
2	27
2	38
2	34
2	35
1	2
1	8
1	9
1	37
1	39
0	4
0	32

Para la etapa dos, dado que el procedimiento de análisis de reactivos establecido por Madsen (1983) requiere las respuestas para cada alternativa de cada reactivo, se cargan al SPSS las respuestas elegidas por los sujetos. Para preparar el cuadro de reactivos, respuestas y puntajes de los aciertos de cada reactivo se hace lo siguiente con el SPSS.

Paso 1

- Datos, segmentar archivo
- Organizar los resultados por grupos
- Transferir del cuadro de la izquierda el nombre de la variable que identifica el grupo de sujetos, alto, medio o bajo al cuadro central denominado Grupos basados en:
 - Ordenar archivo según variables de aplicación
 - Aceptar

Paso 2

- Analizar, estadísticos descriptivos, frecuencias
- Se transfieren todas las columnas de los reactivos y del puntaje total del cuadro de la derecha denominado Secuencia al cuadro central denominado Variables
- Aceptar

Se crea el cuadro de reactivos, respuestas y puntajes de los aciertos de cada reactivo para los sujetos de los grupos alto y bajo. El símbolo ■ indica la respuesta correcta. Los datos se toman de la columna de frecuencias en los resultados del SPSS para cada reactivo.

Cuadro 7.9. Puntajes de aciertos por reactivo, por grupos alto y bajo

Número de reactivo	Respuestas	Grupo alto aciertos	Grupo bajo aciertos
Reactivo # 1 (sin respuesta)	A	3	4
	■ B	8	1
	C	2	6
	D	0	2
	(X)	0	0
Reactivo # 2 (sin respuesta)	■ A	9	0
	B	2	7
	C	1	4
	D	1	2
	(X)	0	0
Reactivo # 3 (sin respuesta)	A	0	2
	B	1	5
	■ C	12	2
	D	0	4
	(X)	0	0
Reactivo # 4 (sin respuesta)	A	0	3
	B	0	2
	C	0	5
	■ D	13	2
	(X)	0	1
Reactivo # 5 (sin respuesta)	A	0	0
	■ B	12	4
	C	0	4
	D	1	4
	(X)	0	1
Reactivo # 6 (sin respuesta)	A	0	3
	B	0	1
	C	1	6
	■ D	12	3
	(X)	0	0
Reactivo # 7 (sin respuesta)	A	0	0
	B	2	8
	■ C	11	1
	D	0	3
	(X)	0	1
Reactivo # 8 (sin respuesta)	A	0	5
	B	2	3
	C	0	2
	■ D	11	3
	(X)	0	1

Reactivo # 9 (sin respuesta)	A	0	0
	B	2	4
	C	0	8
	■ D	11	1
	(X)	0	0
Reactivo # 10 (sin respuesta)	■ A	9	1
	B	3	5
	C	1	5
	D	0	2
	(X)	0	0

Para calcular los grados de dificultad y de discriminación de los reactivos, se construye un cuadro que muestra los cálculos según se indicó al inicio del capítulo. El grado de dificultad por reactivo, con la suma de aciertos del grupo alto y del grupo bajo entre el número total de sujetos, para este caso $13 + 13 = 26$ sujetos; el grado de discriminación por reactivo, a través de la sustracción de aciertos del grupo bajo y del alto, entre el número total de sujetos, para este caso, igualmente, $13 + 13 = 26$ sujetos.

Cuadro 7.10. Grados de dificultad y de discriminación, por reactivo

Reactivo	Puntaje alto	Puntaje bajo	Suma de puntajes	Grado de dificultad	Resta de puntajes	Grado de discriminación
1	8	1	9	34.61%	7	26.92%
2	9	0	9	34.61%	9	34.61%
3	12	2	14	53.84%	10	38.46%
4	13	2	15	57.69%	11	42.30%
5	12	4	16	61.53%	8	30.76%
6	12	3	15	57.69%	9	34.61%
7	11	2	13	50.00%	9	34.61%
8	11	3	14	53.84%	8	30.76%
9	11	1	12	46.15%	10	38.46%
10	9	1	10	38.46%	8	30.76%

Grado de dificultad

Reactivo muy fácil, mayor al 90%: muy difícil, menor al 30%. Aceptable, entre 30 y 90 por ciento.

Todos los reactivos se encuentran en un rango de aceptables (34.61, 61.53), por encima del 30% y por debajo del 90%.

Grado de discriminación

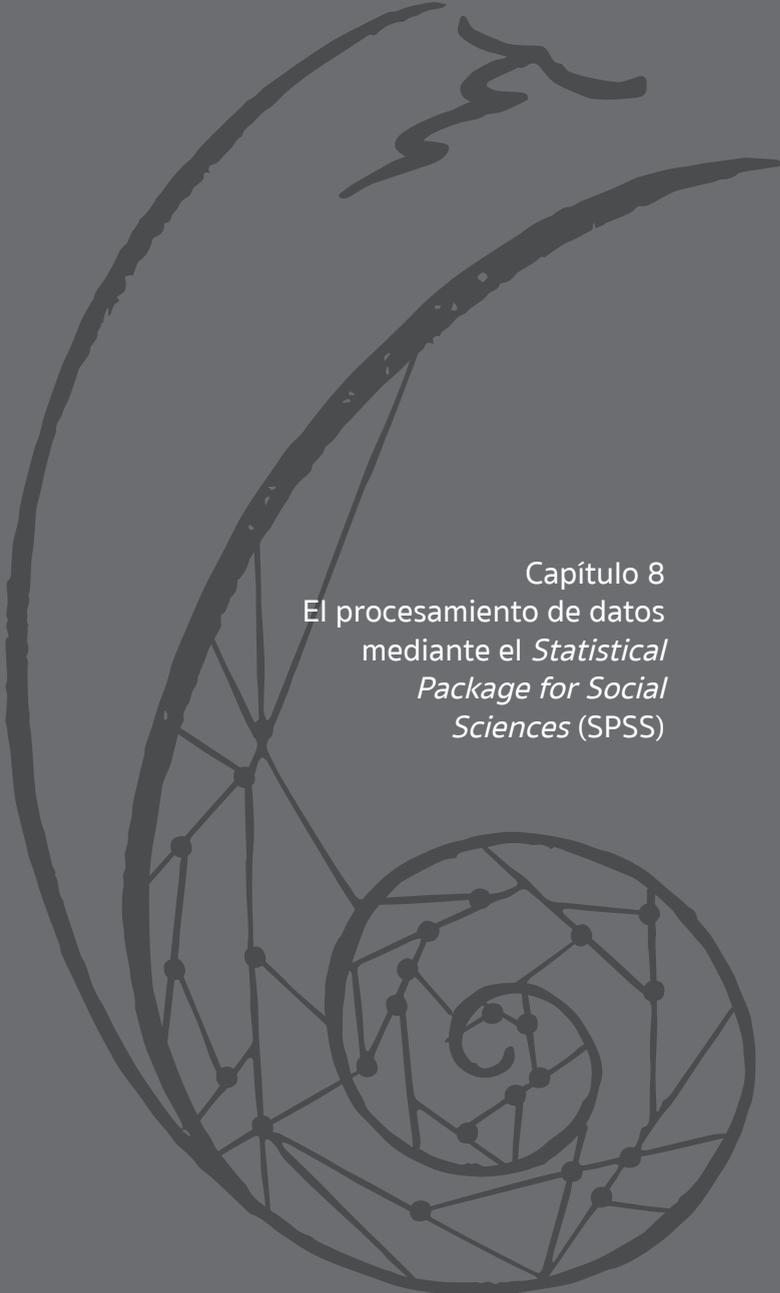
Menor a quince, inaceptable, mayor a 15, aceptable. En este ejemplo, todos los reactivos se encuentran en un rango de aceptables (26.92, 42.30), por encima del 15%, todos son confiables.

Como una etapa complementaria de estos dos cálculos, hay que conjuntar los resultados obtenidos para los grados de dificultad y de discriminación. En ocasiones sucede que el reactivo satisface un criterio pero no el otro, caso en el que también es mejor desecharlo. También se presentan casos en los que el reactivo claramente no cumple con ninguno de los dos criterios. En el ejemplo presentado, todos los reactivos son confiables.

Debido a que existe la posibilidad de rechazar reactivos, siempre hay que hacer las pruebas de este tipo de instrumentos con un número de reactivos superior al que se ha planeado para la aplicación posterior.

Como puede apreciarse, este procedimiento de análisis de los reactivos de un instrumento de medición de opción múltiple es laborioso. Sin embargo, sus resultados garantizan que, al aplicarlo, se mide lo que deseamos en forma confiable.





Capítulo 8
El procesamiento de datos
mediante el *Statistical
Package for Social
Sciences* (SPSS)



Este capítulo constituye una guía general para el manejo del SPSS como una de las posibles opciones para el proceso de datos en forma electrónica, a través de los así denominados *paquetes estadísticos*.

Se introducen someramente las opciones que hay en este momento en el mercado así como comentarios relacionados con el uso de este tipo de paquetes. El segundo y tercer incisos exponen el trabajo previo de diseño que hay que llevar a cabo antes de la captura de los datos en cualquier paquete estadístico, la forma de asignarles códigos para que se manejen en forma numérica y las plantillas que van a crearse.

Se indica las plantillas básicas del SPSS, una para variables y otra para datos. Los últimos cuatro incisos indican el procesamiento general de datos en el SPSS y casos específicos como la manera de procesar datos de varios grupos; de calcular el valor de una variable, como el puntaje total dado por la suma de los puntajes de varios reactivos; la transformación del valor de una variable en rangos; y la elaboración de gráficas.

8.1. Paquetes estadísticos

Para el procesamiento estadístico de datos, en el mercado existen varios *Paquetes estadísticos*. Un paquete estadístico es un software diseñado para este tipo de proceso que permite alimentar datos, en general voluminosos, llevar a cabo operaciones específicas de este tipo de procesos y producir información en diversas formas, como cifras, o bien como gráficas.

Los paquetes considerados como los más usados son el SAS (*Statistical Analysis System*), el R y el SPSS (*Statistical Package for Social Scien-*

ces). Datacamp (2014) analiza estos tres paquetes mediante un cuadro comparativo en el que presenta las ventajas y desventajas de cada uno desde diversos puntos de vista. Enseguida se presenta un extracto de este cuadro.

Cuadro 8.1. Comparativo de paquetes estadísticos SAS, R y SPSS

Paquete	SAS	R	SPSS
Versión actual	9.4, Julio/2013	3.1.0, Abril/2014	22, Agosto/2013
Historia Autores	Jim Goodnight Jim Bar	Ross Ihaka Robert Gentleman	Norman H. Nie Dale H. Bent H. Adlai Tex Hull
Año de creación	1972	1995	1968
Institución	Universidad del Estado de Carolina del Norte	Universidad de Auckland, Nueva Zelandia y Fundación R	Adquirido por IBM en 2009
Propósito y usabilidad	Posición de liderazgo en el mercado de análisis	Usado en academia e investigación por mu- cho tiempo	Es una buena herramienta para los que no saben es- tadística
	Fuerte capacidad de ma- nejo de datos Actualizaciones de soft- ware bien probadas Solución cara	Es una fuente abierta, contraparte del SAS Desarrolla capacida- des gráficas a través de otros softwares: ggplot2, googleVis y rCharts Las técnicas se desa- rollan y ponen a dis- posición con mucha rapidez, gracias a la gran comunidad que le da soporte	Tiene aplicaciones en mu- chos campos pero posee un papel de liderazgo en ciencias sociales Como el SAS tiene un pre- cio considerable
Facilidad de aprendi- zaje	Es fácil obtener una comprensión básica pero para manejarlo bien hay que trabajar en muchos aspectos Hay muchos tutoriales Pueden obtenerse certi- ficaciones de SAS	Tiene fama de ser difí- cil de aprender. En vez de establecer un análisis completo de golpe, los usuarios necesitan aprender a usar datos interactiva- mente La comunidad de so- porte de R está crean- do tutoriales de alta calidad	El SPSS es con gran ventaja el más fácil de aprender de los tres Una de sus ventajas más grandes es su similitud con Excell con el que la mayoría está familiarizada

Fuente: adaptado de Datacamp (2014).

Al analizar la información del cuadro, los elementos que resaltan en cuanto a las características de los paquetes estadísticos son:

- Costo
Hay paquetes para los que hay que pagar una licencia, en general con un costo alto y los que son gratuitos. SAS y SPSS requieren una licencia, R no tiene costo.
- Aprendizaje
Los hay de fácil o difícil aprendizaje. El más fácil de aprender es el SPSS y no hay necesidad de saber mucho de estadística. El R es difícil de aprender y exige preparación adicional. El SAS necesita el manejo de más elementos.
- Contenido
Para dar servicio integral a todas las necesidades de proceso de los datos, los paquetes tienen dos modalidades, o son muy completos en cuanto al manejo de procesos estadísticos o contemplan trabajo adicional. En este sentido, el que da mejor servicio es el SPSS, porque abarca procesos de datos para los dos tipos de estadística, descriptiva e inferencial, en forma holística y porque fue creado para dar servicio primordialmente al área de las ciencias sociales. El R requiere software adicional. El SPSS es una buena herramienta para los que no saben estadística (Datacamp, 2014).

Adicionalmente, a pesar de no ser considerado como uno de los paquetes más usados, también existe el PSPP, creado por la Free Software Foundation (2014), como una alternativa similar al SPSS, sin costo. Hernández y Cuevas opinan que “PSPP por su parte, ha ganado terreno paulatinamente a SPSS en aplicaciones generales” (2013, pp. 170).

Todos estos paquetes estadísticos, al ser herramientas creadas para su manejo a través de computadoras, poseen una estructura similar para el manejo de las entidades, los archivos y los comandos para el procesamiento de datos. En este libro se presenta el uso del SPSS por ser el paquete más completo desarrollado para el proceso de datos en ciencias sociales. Si bien, por ejemplo, el R se ha definido para su aplicación en materia académica y de investigación, al hablar de estudios en materia de lingüística aplicada, particularmente de estudios en los salones de clase de lengua, cuyo contenido contempla procesos de enseñanza y aprendizaje, las entidades manejadas pertenecen más a procesos del área de las ciencias sociales.

El aprendizaje del manejo del SPSS facilita el de otros paquetes de la misma naturaleza, como el PSPP antes mencionado.

Cualquiera que sea el paquete estadístico elegido, debe prepararse el procesamiento de los datos. Es necesario definir la hoja de códigos (ver inciso 8.2) y la distribución de las columnas en la plantilla que conforma el archivo. Generalmente, las entidades manejadas se ubican en las columnas y los renglones corresponden a los sujetos o informantes.

A continuación se encuentran los pasos necesarios para la captura y procesamiento de los datos levantados: definición de códigos y creación de la plantilla de datos, trabajo que debe desarrollarse previamente, sea cual sea el paquete estadístico elegido; y carga de datos en archivo electrónico y procesamiento de datos en archivo electrónico, mediante el SPSS.

8.2. Definición de códigos

En la mayoría de los casos, los valores de las variables se establecen en términos numéricos. Los datos que identifican a cada sujeto, como su nombre y dirección, son considerados alfabéticos o alfanuméricos. Los demás datos, se manejan como numéricos. Para definir los códigos de cada dato a capturar, se establecen los posibles valores recabados. Por ejemplo: para el sexo del sujeto encuestado, puede establecerse que el 1 significa masculino y que el 2 significa femenino. La definición de códigos de cada dato se registra en la llamada Hoja de códigos, en la cual se indica los valores y denominaciones de las posibilidades de presentación de cada dato.

Existen dos posibilidades para la asignación de códigos: los de los datos cuantitativos y los de los datos cualitativos. Si los datos son cuantitativos, las opciones de respuesta se conocen de antemano, por lo que pueden tipificarse fácilmente. Si los datos son cualitativos, los códigos se asignan al momento de la captura en el que se establecen las categorías de las respuestas.

En el cuadro 8.2 se muestra la hoja de codificación de datos de la investigación *Aptitud y rendimiento en el aprendizaje de la lengua inglesa como lengua extranjera* (Uribe, 2010), en la que se incluyó la información relativa a la relación entre los datos y las columnas de las variables en el SPSS. La primera columna se designa para un número consecutivo de sujeto. En el margen izquierdo está este número. No obstante, en ocasiones se requiere la separación de los datos o su agrupamiento.

Es necesario tener la primera columna como respaldo para reordenar el archivo en la forma en la que se capturó originalmente. En este caso, hay cuatro zonas de captura y cálculo en la plantilla SPSS. Datos nominales de los sujetos, los resultados de la medición con el instrumento para Rendimiento, los resultados de la medición con el instrumento para la aptitud para una lengua extranjera (MLAT) y las columnas para los procesos posteriores a la captura.

**Cuadro 8.2. Ejemplo de una hoja de codificación
 para la captura de datos en el SPSS**

Anexo M

Hoja de codificación para la captura de datos de la prueba MLAT en el SPSS (*Statistical Package for Social Science*)

Este anexo forma parte del procedimiento de análisis de datos presentado en el inciso 5.10. La numeración corresponde a cada una de las columnas manejadas en el archivo. Los datos más importantes son:

- A. Datos nominales de los sujetos, columnas: 5, 6, 7, 16, 17, 19, 20, 21, 23, 25, 26, 27 y 28.
 - B. Resultados de las pruebas de rendimiento, columnas: 10, 11, 12, 13, 14, 15 y 42
 - C. Resultados de la prueba MLAT, columnas: 30, 32, 34, 36, 38 y 40
 - D. Columnas para facilitar el procesamiento de datos, columnas: 1, 2, 3, 4, 22, 29, 31, 33, 35, 37, 39, 41, 43 y 44
1. Número consecutivo de sujetos
 2. Fecha de presentación de la prueba
 3. Grupo al que se le aplicó
 4. Nivel de lengua en el que se aplicó
 5. Número de cuenta
 6. Nombre
 7. Columna para los que presentaron el QPT: 1, Si; 2, No
 8. Columna sin uso
 9. Columna para examen PET, FIRST o CAE: 1, PET; 2, FIRST; 3, CAE; 4, Sin prueba; 5, francés
 10. Columnas 10 a 15, para resultados del examen PET, FIRST o CAE
 - Columna 10. Speaking
 - Columna 11. Reading
 - Columna 12. Use of English
 - Columna 13. Listening
 - Columna 14. Writing
 - Columna 15. Puntaje total PET, FIRST o CAE
 16. Sexo. 1, hombre; 2, mujer
 17. Generación, 2003, 2004, 2005, 2006
 18. Categorías de la edad: 1=18-21, 2=22-25, 3=26-29, 4=30-33, 5=34-39
 19. Edad
 20. Docencia o traducción como opción terminal. 0, sin respuesta; 1, docencia; 2, traducción
 21. Inglés o francés como opción terminal. 0, sin respuesta, 1, inglés; 2, francés
 22. Categorías por edad de inicio de aprendizaje: 1=0-4, 2=5-9, 3=10-14, 4=15-19, 5=20-24, 6=25-29
 23. Edad de inicio de aprendizaje de inglés, años
 24. Meses
 25. Vivido en país de lengua inglesa. Cero, sin respuesta; 1, sí; 2, no
 26. Número de años
 27. Número de meses
 28. Hablan inglés en la casa. Cero, sin respuesta; 1, sí; 2, no
 29. Categorías para los puntajes MLAT-1, 1=0-10.75; 2=10.76-21.50; 3=21.60-32.25 y 4=32.26-43.00
 30. Puntaje MLAT, 1
 31. Categorías para los puntajes MLAT-2, 1=0-7.5; 2=7.6-15.0; 3=15.1-22.5 y 4=22.6-30.0
 32. Puntaje MLAT, 2
 33. Categorías para los puntajes MLAT-3, 1=0-12.5; 2=12.6-25.0; 3=25.1-37.5 y 4=37.6-50.0
 34. Puntaje MLAT, 3
 35. Categorías para los puntajes MLAT-4, 1=0-11.25; 2=11.26-22.5; 3=22.6-33.75 y 4=33.26-45.00
 36. Puntaje MLAT, 4
 37. Categorías para los puntajes MLAT-5, 1=0-6.0; 2=6.1-12.0; 3=12.1-18.0 y 4=18.1-24.0
 38. Puntaje MLAT, 5
 39. Categorías para los puntajes totales, 1=0-48.0; 2=48.1-96.0; 3=96.1-144.0 y 4=144.1-192.00
 40. Puntaje total
 41. Categoría de examen de inglés: 1, 0-5.9; 2, 6.0-7.5; 3, 7.6-8.9; 4, 9.0-10.0
 42. Calificación del inglés en semestre de inicio
 43. Sólo alumnos con PET, FIRST y CAE = 1; 2= sin prueba, 3= francés
 44. Sólo alumnos con PET y FIRST = 1, Otros = 2

Fuente: Uribe (2010, p. 228).

El cuadro muestra tanto los códigos asignados para la captura de cada dato como el número de columna en la que se captura en el SPSS. Por ejemplo: en la columna 21 se captura el dato que indica si el sujeto tiene como opción terminal el inglés o el francés. Como en todo reactivo, se considera la opción en la que el sujeto en estudio no emite ninguna respuesta, codificada como 0, sin respuesta. A la opción terminal inglés, se le asigna uno y a la de francés, dos.

8.3. Creación de la plantilla de datos para el SPSS

Antes de capturar los datos, con el paquete estadístico elegido, es necesario definir la plantilla de captura de los mismos. Esta plantilla, en principio, se crea a partir de la Hoja de códigos definida previamente. Las secciones de columnas de la plantilla más comunes son:

- Número consecutivo de los sujetos o informantes capturados
- Datos nominales de los sujetos
- Datos levantados, organizados por entidades o indicadores del objeto de estudio
- Resultados de pruebas aplicadas o de otros instrumentos de levantamiento de datos
- Columnas para facilitar el procesamiento de datos

Para el SPSS, se carga el programa. Aparece una pantalla que pregunta *¿Qué desea hacer?*

Hay varias respuestas. Las que más se usan son:

- Abrir un origen de datos existentes
- Introducir los datos

Para crear el archivo, se elige la opción y se acepta la elección.

- Introducir datos
- Aceptar

El SPSS contiene dos plantillas, para la definición del tipo de datos manejados y para su captura. Como en el caso de Excel, tiene cejas en la parte inferior de la plantilla; en este caso, dos, para indicar si se está en la página de definición de las características de los datos (ceja inferior derecha, denominada Vista de variables) o en la página de su captura (ceja inferior izquierda, llamada Vista de datos). El PSPP sigue la misma definición. El siguiente esquema muestra la ubicación de las cejas en la plantilla.

Esquema 8.1. Indicadores del tipo de plantilla

17			
18			
19			▼
	◀		▶
Vista de datos	Vista de variables		

En un inicio, el manejo de estas dos plantillas, datos y variables, puede desconcertar, porque en la sección de definición de las variables o valores posibles de cada dato, se registran las características de cada variable asignando un renglón a cada una de ellas y en las columnas se indican sus características. En cambio, en la sección de datos, las filas corresponden a cada uno de los sujetos que participaron en el levantamiento de datos y las columnas corresponden a los valores obtenidos para cada una de las variables.

Para crear la plantilla de definición de las variables, se elige la ceja de Vista de variables. El siguiente esquema muestra la plantilla para las variables, con los elementos necesarios para definir las.

Esquema 8.2. Plantilla de definición de las características de las variables

											X
Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Unidades Ventana Ayuda											
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1											
2											
3											

En cada renglón o fila se declara una variable. A continuación se mencionan las características que deben declararse en esta plantilla de las variables.

Cuadro 8.3. Características de los datos cargados en la plantilla del SPSS

Característica	Tipos de características
Nombre de la variable	<i>Nombre de la variable</i>
Tipo	Numérica
	Una coma
	Puntos
	Notación científica
	Fecha
	Dólar
	Moneda personalizada
	Cadena
Ancho	Indica el número de dígitos de la columna
Decimales	Indica si el número se maneja con decimales y con cuántos
Etiqueta	Indica el nombre utilizado para reportar cifras de esa variable
Valores	Los valores son los establecidos en la hoja de codificación
Perdidos	Existen tres posibilidades. En general, se maneja que no hay perdidos
Columnas	Indica el número de columnas que ocupa el dato, coincide con el ancho de la columna
Alineación	Izquierda
	Centrada
	Derecha, para los números
Nivel de medición	Nominal
	Ordinal
	Escala
Rol	Entrada, en general se elige esta opción Objetivo Ambos Ninguna Partición Segmentación

Las posibilidades para declarar el tipo de variable son las ocho mencionadas en el cuadro 8.3; las más utilizadas son las de dato numérico y

cadena (alfabética o alfanumérica). El ancho de la columna depende del número de dígitos manejado, enteros y decimales más uno. Para cargar los valores que puede asumir la variable, se teclea sobre la columna de valores en la celda de la variable que se define.

El visor pide el valor de la variable y la etiqueta (el nombre) que se asigna a ese valor. Por ejemplo:

- Se inserta 1 en el cuadro para Valor
- Se inserta la leyenda "Lee diariamente" en el cuadro para Etiqueta
- Añadir
 Se repite el proceso en concordancia con el número de valores que puede asumir la variable.
- Se inserta 2 en el cuadro para Valor
- Se inserta la leyenda "Lee los fines de semana" en el cuadro para Etiqueta
- Añadir
- Se inserta 3 en el cuadro para Valor
- Se inserta la leyenda "No lee" en el cuadro para Etiqueta
- Añadir

Corresponde una columna de la plantilla de variables a cada característica diferente, componente, dimensión o indicador. Para la captura de los datos se utiliza la ceja Vista de datos. El siguiente esquema muestra la plantilla correspondiente.

Esquema 8.3. Plantilla para la captura de los datos de cada sujeto

												X
Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Unidades Ventana Ayuda												
Visible 0 de 0 variables												
	Var											
1												
2												

Los datos completos de cada sujeto se capturan en una sola fila. La primera corresponde al primer sujeto capturado y así sucesivamente. En esta plantilla las columnas son encabezadas por los nombres asignados a las variables en la plantilla de su definición.

8.4. Carga de los datos en el archivo electrónico

Como se mencionó en el inciso 8.2, para cargar la plantilla básica con los datos levantados, es necesario crear, al inicio de la plantilla de variables, una variable denominada secuencia, de tal forma que, al cargar los datos de cada sujeto, esa primera columna también registre un número consecutivo que parte del uno, para control posterior del manejo de agrupamientos de los datos que se plantearán más adelante.

Se crearán las columnas que sean necesarias para obtener cifras totales de los procesos de los datos. Por ejemplo, si se trabaja con una escala compuesta por tres subescalas, se declaran cuatro columnas de puntajes totales, al final de las columnas definidas para cada reactivo, tres para las subescalas y una para el puntaje total.

En el transcurso de la carga de los datos o bien al finalizarla, el archivo se guarda como en cualquier otro programa de computadora.

Los comandos más utilizados son Guardar y Guardar como.

Existe una ventana denominada el Visor. En ella, el SPSS da los resultados de los procesos que se le piden.

8.5. Procesamiento de los datos en el archivo electrónico

El SPSS cuenta con una serie de comandos que permiten el manejo de los datos y cálculos estadísticos con ellos. El siguiente esquema muestra los comandos generales.

Esquema 8.4. Comandos utilizados en el SPSS

												X
Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Unidades Ventana Ayuda												
Visible 0 de 0 variables												
	Var											
1												
2												
3												
4												

En el esquema 8.4, se aprecian los comandos: Archivo, Edición, Ver, Datos, Transformar, Analizar, Marketing directo, Gráficos, Utilidades, Ventana y Ayuda. Entre estos comandos, los más utilizados son: Archivo, Edición, Datos, Transformar, Analizar y Gráficos. Solamente se presenta el uso de los comandos y subcomandos más frecuentes.

El comando Archivo se utiliza en la misma forma que para otros programas, con las funciones principales: Nuevo, Abrir, Cerrar, Guardar, Guardar como, Presentación preliminar, Imprimir y Salir.

El comando de Edición tiene los subcomandos: Deshacer, Rehacer, Cortar, Copiar, Pegar, Insertar variable e Insertar caso (caso se refiere a la información de otro sujeto).

En el comando Datos, hay la opción de Segmentar archivos, con las opciones para analizar todos los datos u organizar los resultados por grupos (según un parámetro de agrupación, el proceso se encuentra en el inciso 8.7. Proceso de datos de varios grupos) y Ordenar datos en orden Ascendente o Descendente.

En el de Transformar, se encuentra el de Calcular variable, caso en el que hay que alimentar el nombre de la variable que recibe el cálculo y la fórmula para calcularla (consultar el inciso 8.8).

También está Recodificar en distintas variables, para clasificar el valor de la variable a través de rangos de valores (consultar el inciso 8.9).

El siguiente comando, Analizar, es el que permite el cálculo de elementos estadísticos. Los principales comandos de los que consta son: Estadísticos descriptivos, Comparar medias, Correlaciones, Regresión lineal, Escala y Pruebas no paramétricas.

Los estadísticos descriptivos más usados son: Frecuencias y Descriptivos. En el comando de Frecuencias, están estadísticas, gráficas y formato. En las estadísticas están: Medidas de Tendencia central, Dispersión y Distribución. Las medidas de tendencia central son: Media, Mediana, Moda y Suma; las de dispersión: Desviación estándar, Varianza, Mínimo, Máximo y Rango; en Distribución, está Asimetría y Curtosis.

En las gráficas están Tipo de gráfica: Ninguna, Barras, Sectores e Histogramas, con opción a graficar la curva normal en el histograma y Valores de la gráfica: en Frecuencias y en Porcentaje.

En Formato, los datos pueden ordenarse en forma Ascendente o Descendente.

En el comando de Descriptivos, están Tendencia central, Media, Suma; Dispersión, Desviación estándar, Varianza, Rango, Mínimo, Máximo; Distribución, Curtosis y Asimetría; y Presentación, Lista de variables y Medias ascendentes y descendentes.

El comando Comparar medias permite: Comparar medias, prueba t con una muestra, prueba t con muestras independientes y prueba t con muestras relacionadas y ANOVA de un factor.

Para Correlaciones, se tiene Correlaciones bivariadas, Pearson, Tau-b de Kendall y Spearman, con prueba de significancia unilateral o bilateral, con opción de marcar las correlaciones significativas, con opción para marcar Medias y Desviaciones.

Para regresión lineal, tiene una pantalla o visor que permite declarar las variables del proceso. Para Escala, se cuenta con análisis de fiabilidad, con prueba Alfa y algunas estadísticas. En Pruebas no paramétricas, abarca para Una muestra, Muestras independientes y Muestras relacionadas.

El contenido del comando Analizar se resume como sigue:

Esquema 8.4. Comandos utilizados en el SPSS

Subcomandos de Analizar	Subcomandos	Subcomandos	Subcomandos
Estadísticos descriptivos	Frecuencias	Estadísticas	Medidas tendencia central: <ul style="list-style-type: none"> • Media • Mediana • Moda • Suma
			Dispersión: <ul style="list-style-type: none"> • Desviación estándar • Varianza • Mínimo • Máximo • Rango
			Distribución: <ul style="list-style-type: none"> • Asimetría • Curtosis
	Gráficas	Formato	Tipo de gráfica: <ul style="list-style-type: none"> • Ninguna • Barras • Sectores • Histogramas (Opción de trazar la curva normal) • Valores de la gráfica • Frecuencias • Porcentajes
			Valores: <ul style="list-style-type: none"> • Ascendentes • Descendentes
			Tendencia central <ul style="list-style-type: none"> • Media • Suma
	Descriptivos	Descriptivos	Dispersión <ul style="list-style-type: none"> • Desviación estándar • Varianza • Rango • Mínimo • Máximo
			Distribución <ul style="list-style-type: none"> • Curtosis • Asimetría
			Presentación <ul style="list-style-type: none"> • Lista de variables • Medias ascendentes • Medias descendentes

Comparar medias	Comparar medias prueba t ANOVA	<ul style="list-style-type: none"> • Una muestra • Muestras independientes • Muestras relacionadas • de un factor 	
Correlaciones	Bivariadas	<ul style="list-style-type: none"> • Pearson • Tau-b de Kendall • Spearman 	<ul style="list-style-type: none"> • Con prueba de significación unilateral o bilateral • Con opción de marcar correlaciones significativas • Con opción para marcar medias y desviaciones
Regresión lineal			
Escala	<ul style="list-style-type: none"> • Análisis de confiabilidad (fiabilidad) 	<ul style="list-style-type: none"> • Alfa • Otras opciones 	
Pruebas no paramétricas	<ul style="list-style-type: none"> • Una muestra • Muestras independientes • Muestras relacionadas 		

Cada vez que se efectúa una operación con un comando o con la combinación de varios, los resultados se presentan en una pantalla propia para ello. Cuando hay necesidad de señalar las variables en proceso, en general, se transfieren de la columna de la izquierda del visor o pantalla hacia la derecha, en ventanas de operaciones específicas.

8.6. Pruebas en general

La mayoría de las pruebas de hipótesis que se utilizan hacen uso de la comparación de medias.

Para comparar medias, se siguen los siguientes pasos:

1. Analizar, Comparar medias
2. Paso de las variables a las ventanas centrales
3. Variables dependientes y variable independiente
4. Aceptar

Se obtiene: la media y el número de sujetos, por el total y por los valores que puede asumir la variable o variables dependientes.

8.7. Procesamiento de datos de varios grupos

Dependiendo del tipo de estadística empleada, los datos pueden analizarse en formas diversas. En cualquiera de estas formas, cuando se trabaja con más de un grupo, en el SPSS puede darse una indicación para que los cálculos se hagan separando cada grupo. En el caso del manejo de más de un grupo, la forma de indicar la separación de los cálculos por grupo es la siguiente.

- Datos, segmentar archivo
- En la pantalla, hay tres alternativas: analizar todos los casos sin crear grupos, comparar los grupos y organizar los datos por grupos
La primera opción es la opción estándar, sin necesidad de elegirla.

Para la segunda, comparar grupos.

- Se marca la opción. Se activa ordenar archivo según variables de agrupación
- Se pasa al centro la variable que identifica los grupos
- Aceptar

Para la tercera opción, organizar los datos por grupos

- Se marca la opción. Se activa ordenar archivo según variables de agrupación
- Se pasa al centro la variable que identifica los grupos
- Aceptar

Como se vio en el inciso 8.4, en el comando llamado Transformar, se encuentra el de Calcular variable, caso en el que hay que alimentar el nombre de la variable que recibe el cálculo y la fórmula para calcularla y Recodificar en distintas variables, para clasificar el valor de la variable a través de rangos de valores. Los incisos 8.8 y 8.9 presentan el manejo de estos dos casos.

8.8. Calcular el valor de una variable

Es un proceso muy utilizado, por ejemplo cuando se quiere indicar que el puntaje total de una escala o de una subescala está constituido por la suma de las respuestas dadas a todos los reactivos o a cierto número de ellos, respectivamente.

■ Transformar

Se abre una ventana con el siguiente contenido. Del lado izquierdo, hay una ventanita para indicar el nombre de la variable que recibirá el resultado de sumar los contenidos de cada reactivo (cada reactivo está definido por una variable en la plantilla de variables). Esta variable puede definirse de antemano en la plantilla de variables o bien darle el nombre para que el SPSS la defina. Se recomienda definirla previamente. En la parte inferior está la lista de las variables definidas para el caso particular.

En el cuadro ubicado en la parte superior, a la derecha, dice "expresión numérica", ahí es donde se define el cálculo, por ejemplo, si se tuviera una escala con dos reactivos y se quisiera conocer el puntaje total en los dos, es decir, para calcular la expresión:

$$\text{Puntaje-total} = \text{Reactivo-uno} + \text{Reactivo-dos.}$$

En el cuadro de la izquierda se escribiría: Puntaje-total. A continuación:

- Se señala el nombre de la variable que corresponde al primer reactivo
- Se tecléa sobre la flecha hacia la derecha
- El nombre de la variable va a la ventana superior de la derecha

Cuadro 8.5. Dato de la expresión numérica

Expresión numérica:

Reactivo-uno

Debajo de la ventana están los signos de las operaciones que pueden hacerse: + - * / **; < ≤ = & ~; > ≥ ~ = | () y los números del 0 al 9 y la leyenda Eliminar. Las expresiones para los cálculos se construyen con estos símbolos y números.

- Se tecléa el signo +
- Se señala el nombre de la variable que corresponde al segundo reactivo
- Se tecléa sobre la flecha hacia la derecha
- El nombre de la variable va a la ventana superior de la derecha

Cuadro 8.6. Segundo dato de la expresión numérica

Expresión numérica:

Reactivo-uno + reactivo-dos

- Aceptar

Pregunta si se cambia el valor de la variable dada para recibir el resultado para clasificar el valor de la variable a través de rangos de valores.

- Cambiar

Después de este proceso, la columna definida para recibir la suma de los dos reactivos la contiene.

8.9. Transformar variable por rangos de sus valores

Como se mencionó, este proceso se usa para clasificar el valor de la variable a través de rangos de valores. Por ejemplo, para la edad de los sujetos participantes.

El proceso se realiza en varias etapas. En la primera, hay que ubicarse en la plantilla de definición de las variables, como lo muestra el esquema 8.5, para obtener los valores máximo y mínimo de la variable en proceso. En este caso, la edad.

Esquema 8.5. Descriptivos de la variable edad

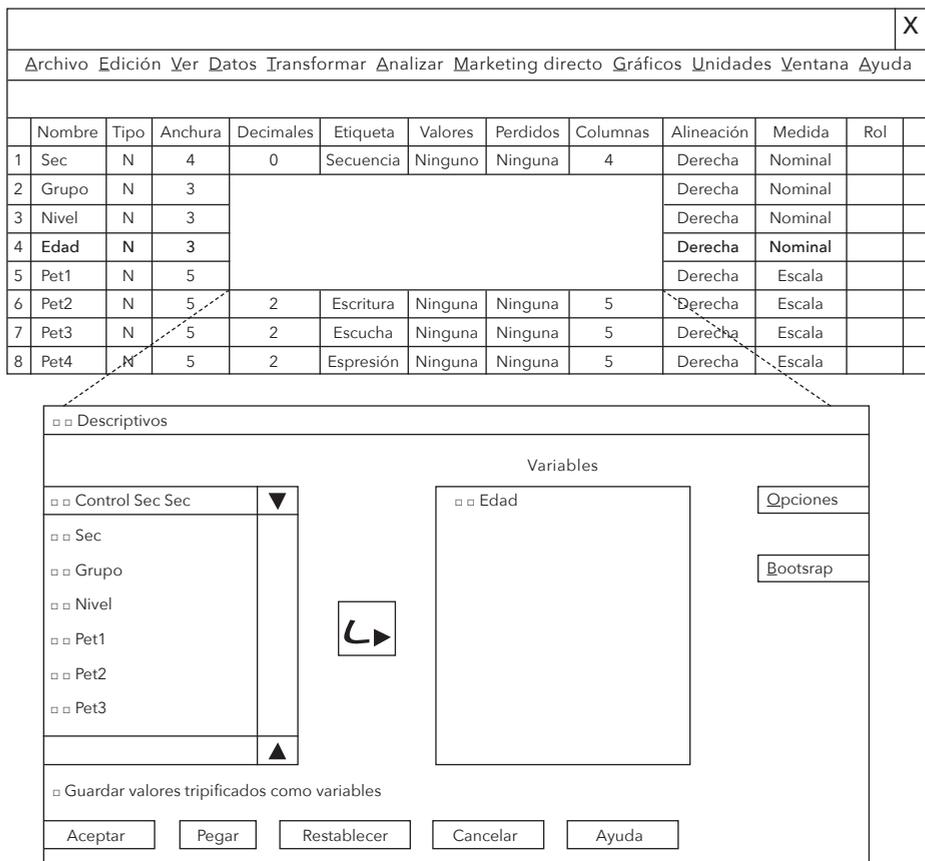
												X
Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Unidades Ventana Ayuda												
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol	
1	Sec	N	4	0	Secuencia	Ninguno	Ninguna	4	Derecha	Nominal		
2	Grupo	N	3	0	Grupo	Ninguno	Ninguna	3	Derecha	Nominal		
3	Nivel	N	3	0	Nivel	1. semanal	Ninguna	3	Derecha	Nominal		
4	Edad	N	3	0	Edad	Ninguno	Ninguna	3	Derecha	Nominal		
5	Pet1	N	5	2	Lectura	Ninguno	Ninguna	5	Derecha	Escala		
6	Pet2	N	5	2	Escritura	Ninguno	Ninguna	5	Derecha	Escala		
7	Pet3	N	5	2	Escucha	Ninguno	Ninguna	5	Derecha	Escala		
8	Pet4	N	5	2	Expresión	Ninguno	Ninguna	5	Derecha	Escala		

Cuando se está en la plantilla, se obtienen los valores mínimo y máximo de la edad, con los siguientes comandos:

Analizar, Estadísticos descriptivos, Descriptivos

Sobre la pantalla de definición de las variables se abre una ventana según se muestra en el esquema 8.6. A la izquierda está la lista de variables, ahí se marca la variable edad y se tecléa sobre la flecha para que el nombre de la variable vaya al cuadro del centro.

Esquema 8.6. Ventana para trabajar descriptivos de la variable específica (edad)



A continuación se tecléa lo siguiente:

- Opciones

En la ventana que se abre se marca

- Mínimo y máximo
- Continuar
- Aceptar

Presenta el siguiente cuadro como resultado.

Cuadro 8.7. Valores máximo y mínimo de la variable
Estadísticos descriptivos

	N	Mínimo	Máximo
Edad	162	18	39
N válido (según lista)	162		

Estos valores máximo y mínimo sirven para establecer los rangos de clasificación de la edad.

Se resta 18 de 39 ($39-18=21$). Supongamos que queremos tres rangos de edad para clasificar a los sujetos. Se divide 21 entre tres, lo que da un total de 7. A partir de aquí, los rangos quedan como:

18 a 24
25 a 31 y
32 a 39

En la segunda etapa, se crea una variable para que reciba los valores de la edad clasificados en rangos de edad. Se asigna un nombre a cada rango. Por ejemplo:

18 a 24 Edad más baja
25 a 31 Edad intermedia y
32 a 39 Edad más alta

- Transformar
- Recodificar en distintas variables

En la pantalla que presenta, igualmente, a la derecha está la lista de variables. Se pasa al centro la variable que quiere desglosarse.

Cuadro 8.8. Variable para desglosarse en rangos

Variable numérica: ---> Variable de resultado:

edad -->

El cuadro de la derecha permite dar un nombre a la variable que contendrá los rangos y una etiqueta para fines de identificación de resultados. Por ejemplo: raedad, Rangos de edad.

- Cambiar

El cuadro indicará

Cuadro 8.9. Nombres de las variables originales y por rangos

Variable numérica: ---> Variable de resultado:

edad --> raedad

En la parte inferior se teclea

- Valores antiguos y nuevos

Abre una ventana con valores antiguos a la izquierda y valores nuevos a la derecha. Los rangos establecidos se manejan en el lado izquierdo.

- El primer rango utiliza: Rango, INFERIOR hasta valor. En este caso, 24, con valor en la derecha de uno.
- Añadir

En el cuadro central del lado derecho aparece:

Cuadro 8.10. Rango inicial definido

Antiguo: ---> Nuevo:

Lowest thru --> 1

El rango intermedio, en este caso sólo uno o los que hubiera, se declaran usando:

- Rango
- 25
- hasta
- 31
- Valor a derecha de dos
- Añadir

La pantalla se modifica a:

Cuadro 8.11. Rangos inicial e intermedio definidos

Antiguo: ---> Nuevo:

Lowest thru 24 --> 1
25 thri 31 --> 2

Para el último rango se utiliza:

- Rango, valor hasta SUPERIOR
- 39
- Valor a la derecha de tres
- Añadir

La pantalla se presenta como:

Cuadro 8.12. Todos los rangos definidos

Antiguo: ---> Nuevo:

Lowest thru 24 --> 1
25 thru 31 --> 2
39 thru Highest --> 3

- Continuar

Regresa a la pantalla anterior

- Aceptar

Aparece la pantalla de resultados, con los siguientes cuadros:

```
DESCRIPTIVES VARIABLES=edad  
/STATISTICS=MIN MAX.
```

Descriptivos

[Conjunto_de_datos1]F:\Datos cognitiva y afectiva\20121110pruebas mlat rendimto y ansiedad 162.sav

Cuadro 8.13. Valores máximo y mínimo

Estadísticos descriptivos

	N	Mínimo	Máximo
Edad	162	18	39
N válido (según lista)	162		

```
RECODE edad (Lowest thru 24=1) (25 thru 31=2) (39 thru Highest=3) INTO raedad.
```

```
VARIABLE LABELS raedad 'Rangos de edad'.
```

```
EXECUTE.
```

Los valores establecidos como rangos quedan en la variable denominada como *raedad*.

El siguiente paso para cerrar el proceso es definir los valores y las etiquetas en la variable creada para contener los rangos; en este caso, *raedad*.

En la plantilla para definir variables hay que definir los valores 1, 2 y 3, para las etiquetas: edad más baja, edad intermedia y edad más alta.

Enseguida, se hace el proceso para conocer la situación de la variable *raedad*, ubicándose en la plantilla de captura de datos.

Analizar, Estadísticos descriptivos, Frecuencias

Aparece la pantalla:

Esquema 8.7. Comportamiento de la variable con rangos de edad

	Sec	Grupo	Nivel	Edad	Pet1	Pet2	Pet3	Pet4	Puntaje T	Redad		
1	1	10	7	24	6.80	8.80	6.90	6.00	7.12	1		
2	2	5	5						6.50	3		
3	3	2	3						8.60	2		
4	4	7	4						9.20	3		
5	5	4	4						8.20	3		
6	6	2	4	26	7.20	7.60	7.66	6.79	7.31	2		
7	7	5	3	27	8.56	8.68	9.45	9.34	9.00	2		
8	8	4	4	39	9.12	8.45	7.34	8.43	8.83	3		

Frecuencias

Control Sec Sec

Sec

Grupo

Nivel

Edad

Pet1

Pet2

Mostrar tablas de frecuencias

Variables

Rangos de edad

Estadísticos

Gráficos

Formato

Bootstrap

Aceptar Pegar Restablecer Cancelar Ayuda

Se teclean los siguientes comandos:

- Estadísticos
- Media
- Continuar
- Aceptar

Los resultados se presentan como sigue:

Frecuencias

[Conjunto_de_datos1]F:\Datos cognitiva y afectiva\20121110pruebas mlat rendimto y ansiedad 162.sav

Cuadro 8.14. Número de casos válidos y valor de la media

Estadísticos		
Rangos de edad		
N	Válidos	154
	Perdidos	8
Media		1.1364

Cuadro 8.15. Frecuencia y porcentaje de los tres rangos

		Rangos de edad			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Edad más baja	134	82.7	87.0	87.0
	Edad intermedia	19	11.7	12.3	99.4
	Edad más alta	1	.6	.6	100.0
	Total	154	95.1	100.00	
Perdidos	Sistema	8	4.9		
Total		162	100.0		

Este último cuadro muestra los sujetos en cada uno de los rangos definidos, tanto en cifra como en porcentaje. Adicionalmente están los sujetos que no declararon su edad, considerados como Perdidos. El resultado muestra que el 82.7% de los sujetos tienen entre 18 y 24 años de edad.

8.10. Gráficas

En el inciso 4.5 se vieron los tipos de gráficas que pueden presentarse. Este inciso indica la forma de obtenerlas con el SPSS. En la misma línea de trabajo de otros procesos, aquí no se revisan todos los comandos sino los más usados. La descripción de los pasos se basa en Bakieva *et al.* (2010), con la versión 15 del SPSS y en la versión 18 del propio SPSS.

Existen dos procedimientos para hacer gráficas.

- Iniciando con el comando Analizar o
- Directamente con el comando Gráficos

Para la primera opción dentro del comando Analizar, se teclea lo siguiente.

SPSS

- Analizar
- Estadísticos descriptivos
- Frecuencias
- Gráfica

Aparece la pantalla que tiene a la izquierda todas las variables y el cuadro del centro para pasar las variables en proceso, ya visto en el esquema 8.6.

Para la segunda opción, a través del comando Gráficos, se hace directamente.

SPSS

- Gráficos
- Opción a, Generador de gráficos
- Opción b, Cuadro de gráficos antiguos
- Aceptar

Opción a. En la pantalla aparecen tres cuadros superiores y uno inferior. El superior izquierdo corresponde a la lista de variables que pueden seleccionarse. El superior derecho es una gran ventana, denominado por Bakieva *et al.* (2010) como el lienzo, podría denominarse la pantalla de la derecha, contiene dos sugerencias:

1. Arrastre hasta aquí un gráfico de la galería para utilizarlo como punto de partida y

2. Pulse en la ficha elementos básicos para crear un gráfico mediante elementos

Debajo del superior izquierdo hay un cuadro que indica información sobre categorías.

En la parte inferior de la pantalla, hay tres secciones. De izquierda a derecha, la primera es denominada **Galería** y contiene cinco indicadores: elementos básicos, Grupos/ID de puntos, Títulos/notas al pie, propiedades del elemento y opciones.

A continuación hay un cuadro denominado Elija entre, Este cuadro contiene una lista de los tipos de gráficas antes analizados. Finalmente, en la parte inferior derecha, hay imágenes de diversos tipos de gráficas.

Bakieva *et al.* (2010) consideran que el método más fácil para genera una gráfica es mediante el uso de la Galería.

Una vez seleccionada la Galería, los pasos que sugieren son los siguientes.

- En la lista de gráficas, se elige la que se desea

Al señalarla, la ventana inferior permite elegir el tipo de gráfica deseada: favoritos, de barras, de líneas, de áreas, de sectores, de dispersión, los histogramas, de máximos y mínimos, de caja y ejes dobles.

El ejemplo muestra la opción de Barras.

- Cada una ofrece varios tipos
- Se selecciona el subtipo de gráfica que se desea
- Se lleva la imagen de la gráfica a la pantalla superior derecha (al lienzo) o se pulsa dos veces sobre la imagen.

Si había una gráfica en el lienzo, se sustituye por la seleccionada

- Se llevan las variables tomándolas de la lista en la sección superior izquierda y se ubican en el eje y, en la zona de colocación de agrupamiento, cuando el texto que aparece en la zona tenga color azul.

Puede suceder que ya esté ahí un elemento estadístico que se desea utilizar, en cuyo caso no hay que llevar ninguna variable a esa zona; el texto aparece en negro.

- Se selecciona el tipo de gráfica y el subtipo en éste
- Se lleva la imagen del gráfico a la pantalla superior derecha o se tecllea dos veces sobre la imagen

La elección de una gráfica de la lista lleva implícita la consideración de un conjunto de ejes y elementos gráficos. Aparece la pantalla de propiedades del elemento.

Los pasos y recomendaciones dados por Bakieva *et al.* (2010) son:

- Arrastrar variables desde la lista Variables y colocarlas en las zonas de colocación del eje y, si está disponible, en la zona de colocación de agrupamiento.
- Si una zona de colocación del eje ya muestra un estadístico que se desea utilizar, no tendrá que arrastrarse ninguna variable a la zona de colocación. Sólo deberá añadirse una variable a la zona cuando el texto de la zona sea azul.
- Si el texto es negro, la zona ya contiene una variable o un estadístico

Bakieva *et al.* (2010) recomiendan cuidar que se haya definido el nivel de medición de las variables.

- El nivel de medición de las variables es importante
- El generador de gráficas establece la configuración por defecto según el nivel de medición durante la generación de una gráfica
- Además, la gráfica resultante también puede tener un aspecto distinto para los diferentes niveles de medición
- Puede cambiarse temporalmente el nivel de medición de una variable pulsando con el botón derecho del ratón en la variable y eligiendo una opción
- Si es necesario, pueden cambiarse las propiedades de los elementos de la gráfica, a través de las propiedades del elemento. Por ejemplo, puede elegirse la media (suma, porcentaje, recuento... etc.) como el elemento estadístico para los sectores (o barras) de la gráfica
- Si desea transponer la gráfica (por ejemplo, para que las barras sean horizontales), se pulsa en la pestaña Elementos básicos y, a continuación, se pulsa en Transponer
- Si necesita agregar más variables a la gráfica (por ejemplo, para la agrupación o la adición de paneles) se pulsa en la pestaña Grupos/ID de puntos y se selecciona una o más opciones
- Se arrastran las variables categóricas a las nuevas zonas de colocación que aparecen la pantalla derecha
- Aceptar

Si desea cambiar una gráfica por otra, por ejemplo una gráfica de sectores por una gráfica de barras, en el Menú Editor de gráficos se selecciona:

- Cambiar gráfico y seleccionar un gráfico desde la lista de opciones
- Pueden cambiarse las propiedades de las gráficas, a través de la elección de la media, suma o porcentaje, como el elemento estadístico para los sectores o barras de la gráfica.
- La barra de elementos básicos permite hacer estos cambios.
- Aceptar

En el visor de resultado aparece la gráfica solicitada. Para la opción b, Cuadro de gráficos antiguos:

- Cuadro de gráficos antiguos

Da enseguida los tipos de gráficas que pueden construirse que ya se han enunciado y algunas otras, barras, barras 3D, líneas, áreas, sectores, máximos y mínimos, diagramas de caja, barras de error, pirámides de población, dispersión/puntos e histograma.

- Al elegir la gráfica, cada una de ellas tiene sus propios parámetros de construcción

Por ejemplo, la gráfica de barras indica que puede ser: simple, de una sola variable compuesta o apilada, para dos variables.

- Definir
- Hay una pantalla adicional que define los elementos de la gráfica, las variables, los ejes y los títulos.

Bibliografía

- Alford, A. (2011). *Prueba de bondad de ajuste Kolmogorov-Smirnov*. Recuperado de <http://pruebadbondaddeajuste.blogspot.mx/> [30/09/2014].
- Anguita, M. (2000). *Práctica del curso de programación de Applet's; uso de los recursos de internet para el aprendizaje de Estadística descriptiva- 1º Bachillerato*. Recuperado de <http://thales.cica.es/rd/Recursos/rd99/ed99-0122-04/ed99-0122-04.html> [11/09/2014].
- Anónimo (2012). "Prueba t de Student para dos muestras relacionadas", en Conexionismo. Recuperado de http://www.conexionismo.com/leer_articulo.php?ref=prueba_t_de_student_para_dos_muestras_relacionadas-902ybsd7 [04/08/2014].
- Alaminos, A. y J. Castejón (2006). *Elaboración, análisis e interpretación de encuestas, cuestionarios y escalas de opinión*. Vicerrectorado de calidad y administración europea, Instituto de ciencias de la educación. Serie Docencia Universitaria-EEES. Universidad de Alicante: Marfil. Recuperado de <http://rua.ua.es/dspace/bitstream/10045/20331/1/Elaboraci%c3%b3n%2c%20an%c3%a1lisis%20e%20interpretaci%c3%b3n.pdf> [12/06/2014].
- Álvarez, M. (s/f). *Introducción a la estadística descriptiva; apuntes para el módulo 2*. México: Universidad Nacional Autónoma de México.
- Álvarez, T. (2005). *Didáctica del texto en la formación del profesorado*. Madrid: Síntesis.
- Bakieva, M. et al. (2010). *Gráficos en SPSS. Material elaborado en la Convocatoria de innovación de 2010 del Vicerectorat de Convergència Europea I Qualitat de la Universitat de València*. Recuperado de http://www.uv.es/innovamide/spss/SPSS/SPSS_0303a.pdf [04/07/2014].
- Benemérita Universidad Autónoma de Puebla (2014). *Doctorado en Ciencias del Lenguaje*. Recuperado de <http://estudios.universia.net/mexico/estudio/buap-doctorado-ciencias-del-lenguaje> [16/06/2014].
- Benzécri, J. (1973). *L'Analyse de données. Tome I. La Taxinomie*. Paris: Dunod, Escuela Francesa de Análisis de Datos, en R. J. Rodríguez. ¿Cómo analizar cuantitativamente datos cualitativos?; aná-

lisis de datos cualitativos. Recuperado de <http://www.gestiopolis.com/como-analizar-cuantitativamente-datos-cualitativos/> [27/10/2015].

_____ (1976). *L'Analyse de données. Tome II. L'Analyse des correspondances*. Paris: Dunod.

Bernia, J. (1979). *Psicología experimental I*. Valencia: Nau Libres.

Brown, J. (2004). "Research Methods for Applied Linguistics: Scope, Characteristics and Standards", en A. Davies y C. Elder (eds.), *Handbook of Applied Linguistics*. Oxford: Blackwell Publishing Ltd., 475-500.

Brown, S. (2014). *Measures of Shape: Skewness and Kurtosis*. Recuperado de <http://www.tc3.edu/instruct/sbrown/stat/shape.htm> [14/04/2014].

Carter, R. y D. Nunan (eds.) (2001). "The Cambridge Guide to Teaching English to Speakers of Other Languages", Cambridge: Cambridge University Press, en N. Schmitt y M. Celce-Murcia. *An Overview of Applied Linguistics*, en N. Schmitt. *An Introduction to applied linguistics*. Londres: Arnold.

Campbell, D. y J. Stanley (1963). "Experimental and quasi-experimental design for research", en A. Alaminos y J. Castejón. *Elaboración, análisis e interpretación de encuestas, cuestionarios y escalas de opinión*. Vicerrectorado de calidad y administración europea, Instituto de ciencias de la educación, Universidad de Alicante: Marfil.

Celorrio, A. (2002). "Pruebas de hipótesis no paramétricas de Kolmogorov-Smirnov", en *Monografías.com*. Recuperado de <http://www.monografias.com/trabajos11/docima/docima.shtml> [29/09/2014].

Cervantes, V. (2005). "Interpretaciones del coeficiente Alpha de Cronbach", en *Avances de medición*, 3, 9-38.

Chernobilsky, L. (2007). "El uso de la computadora como auxiliar en el análisis de datos cualitativos", en Irene Vasilachis de Gialdino (coord.). *Estrategias de investigación cualitativa*. Madrid: Gedisa, 239-272.

Collado, J. (1978). *Fundamentos de Lingüística general*. Madrid: Gredos.

_____ (1986). *Fundamentos de Lingüística aplicada*. Madrid: Gredos.

- Cook, T. y D. Campbell (1979). *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNall.
- Datacamp (2014). *Statistical Language Wars: The Infograph*. Recuperado de <http://datacamp.wpengine.com/wp-content/uploads/2014/05/infograph.png> [09/10/2014].
- Davies, A. (1999). *An Introduction to Applied Linguistics: From Practice to Theory*. Edimburgo: Edinburgh University Press.
- _____ (2007). *An Introduction to Applied Linguistics: From Practice to Theory*. Cambridge: Edinburgh University Press.
- Davies, A. y C. Elder (eds.) (2004). *The Handbook of Applied Linguistics*. Oxford: Blackwell Publishing Ltd.
- Decarlo, L. (1997). "On the meaning and use of kurtosis", en *Psychological Methods*, 2 (3), 292-307. Recuperado de <http://www.columbia.edu/~ld208/psymeth97.pdf> [14/04/2014].
- Dehouve, D. (2013). Las funciones rituales de los altos personajes mexicas. *Estudios de cultura náhuatl*, 45, enero-junio. Recuperado de http://www.scielo.org.mx/scielo.php?pid=S0071-16752013000100003&script=sci_arttext [20/11/2015].
- Ducrot, O. y T. Todorov (2006). *Diccionario enciclopédico de las ciencias del lenguaje*. 24ª ed., México: Siglo XXI.
- El osio de los santos (s/f). *Tablas de distribución F*. Recuperado de <http://www.elosiodelosantos.com/sergiman/div/tablaf.html> [02/09/2014].
- Escuela Nacional de Antropología e Historia (2014). Programa del Doctorado en Ciencias del Lenguaje. Recuperado de http://hispanismo.cervantes.es/Departamentos_ficha.asp?DOCN=2160 [16/06/2014].
- Florit (2013). *Fonética, fonología, morfología, principios de contrastividad*. Programa académico. Recuperado de http://www.lenguas.unc.edu.ar/carrerasdegrado/espanol/primero/Fonetica_Fonologia_Morfologia_espanolas_y_principios_de_contrastividad.pdf [14/10/2015].

- Free Software Foundation (2014). "GNU PSPP", en *Operating system*. Recuperado de <http://www.gnu.org/software/pspp/> [09/10/2014].
- Gardner, R. (1993). "A student's contribution to second language learning. Part II: affective variables", en *Language Teaching*, 26, 1-11.
- Gardner, R. y W. Lambert (1972). *Attitudes and Motivation in Second Language Learning*. Massachusetts: Newbury House Publishers.
- Gardner, R. y P. McIntyre (1993). "A student's contribution to second language learning. Part I: cognitive variables", en *Language Teaching*, 26, 1-11.
- Gómez, J. (2015). *Coeficiente Alfa de Cronbach; validación del instrumento de recolección de datos*. [Power Point]. Recuperado de http://www.google.com.mx/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&sqi=2&ved=0CCMQFjAB&url=http%3A%2F%2Fficad.milaulas.com%2Fpluginfile.php%2F202%2Fmod_folder%2Fcontent%2F0%2FJose%2520Manuel%2520Gomez.ppt%3Fforcedownload%3D1&ei=xlsWVO_RF-vP8QHs3oCIC-g&usg=AFQjCNEFq1-Nyfyi5w17vaZzx1Nx1dtiw&sig2=iIPBO3il-TwgklnbJa0J66g&bvm=bv.75097201,d.b2U [14/09/2014].
- Graphpad Software (2014). *GraphPad Prism*. Recuperado de <http://www.graphpad.com/> [25/08/2014].
- _____ (2014). *Interpreting results: Skewness and Kurtosis*. Recuperado de http://www.graphpad.com/guides/prism/6/statistics/index.htm?stat_skewness_and_kurtosis.htm [14/04/2014].
- Guillén, A. (2006). *Procesamiento de datos; paquetes estadísticos*. Recuperado de <http://www.revistasden.org/files/3-CAP%203.pdf> [10/10/2014].
- Hernández, R. et al. (2006). *Metodología de la Investigación*. México: Mc Graw Hill.
- Hernández, S. y J. Cuevas (2013). "Programas informáticos de uso libre y su aplicación en la enseñanza de la estadística", en *Revista Investigación Operacional*, 34 (2), 166-174. Recuperado de <http://rev-inv-ope.univ-paris1.fr/files/34213/34213-08.pdf> [13/10/2014].

- Hohenthal, A. (s/f). "Measurement Techniques. What is a language attitude?", en *The Literature and Culture of the Indian Subcontinent (South Asia) in the Postcolonial Web*. Recuperado de <http://www.postcolonialweb.org/india/hohenthal/6.1.html> [23/11/2015].
- Holland, J. y J. Miller (1991). "Learning and adaptive economic behavior", en *American Economic Review*, 81 (2), 365-370. Recuperado de <http://www.iula.upf.edu/doctores.htm> [09/06/2014].
- Juárez, F. et al. (2002). *Apuntes de estadística inferencial*. México: Instituto Nacional de Psiquiatría Ramón de la Fuente.
- Kaplan, R. (ed.) (2002). *The Oxford Handbook of Applied Linguistics*. Oxford: Oxford University Press.
- Laerd Statistics (s/f). *Multiple Regression Analysis using SPSS Statistics. Take the Tour Plans & Pricing SIGN UP*. Recuperado de <https://statistics.laerd.com/spss-tutorials/multiple-regression-using-spss-statistics.php> [25/08/2014].
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. London: Routledge.
- Lenarduzzi, R. (2002). *Nuevos enfoques y disciplinas en la ciencias del lenguaje y su proyección aplicada en campo didáctico: los elementos extraoracionales*. Testi specialistici e nuovi saperi nelle lingue iberiche, Messina, A. Lippolis (ed.), 151-164. Recuperado de http://cvc.cervantes.es/literatura/aispi/pdf/16/16_149.pdf [31/05/2014].
- Likert, R. (1932). *A technique for the measurement of attitude*, *Archives of Psychology*, 140. [Traducción al castellano en G. Summers (1976). *Medición de actitudes*]. México: Trillas.
- Lincoln, S. y E. Guba (1995). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- López, L. (2011). "El salón de clase de Nicolás Paris", en *Cultura. El Espectador*. Recuperado de <http://www.mineducacion.gov.co/observatorio/1722/article-273482.html> [04/06/2014].
- Madsen, H. (1983). *Techniques in Testing*. Oxford: Oxford American English.

- Manzano, A. (2011). *Prueba de Levene de ad de Varianzas*. Recuperado de <http://es.scribd.com/doc/61087146/Prueba-de-Levene-de-ad-de-Varianzas> [14/09/2014].
- Marin, C. (1975). *Manual de investigación en psicología social*. México: Trillas.
- Marketing Directo (2004). *SPSS, breve historia de un producto que se ha convertido en una marca*. [Nota de prensa 19/04/2004]. Recuperado de <http://www.marketingdirecto.com/actualidad/marketing/spss-breve-historia-de-un-producto-que-se-ha-convertido-en-una-marca/> [09/10/2014].
- Martínez, A. (2005). "Chi-Cuadrado. Pruebas estadísticas de SPSS", en *SPSS Free*. Recuperado de <http://www.spssfree.com/spss/tablas24.html> [16/09/2014].
- Medina, N. (2012). *El diario de campo: valiosas contribuciones a la investigación cualitativa*. Recuperado de <http://dcprecisionesmetodologicas.blogspot.mx/2012/07/normal-0-21-false-false-false-es-x-none.html> [03/06/2013].
- Mendoza, F. (2009). *El cambio en el salón de clases; un proceso permanente*. Recuperado de <http://felixmendoza60-laesquinadesaber.blogspot.mx/2009/08/el-cambio-en-el-salon-de-clases-un.html> [04/06/2014].
- Merriam-Webster (2014). *Dictionary and Encyclopaedia Britannica Company*. Recuperado de <http://www.merriam-webster.com/dictionary/linear%20regression> [07/09/2014].
- Miles y Huberman (1994). *Qualitative Data Analysis*. London: SAGE Publications.
- Molina, E. (2008). Edgar Morin y el pensamiento complejo 01, entrevista. [archivo de video]. Recuperado de: https://www.youtube.com/watch?v=t_sNKVcnfZQ [04/06/2014].
- Morales, J. (1981). "Metodología y Teoría de la Psicología", en A. Alaminos y J. Castejón. *Elaboración, análisis e interpretación de encuestas, cuestionarios y escalas de opinión*. Vicerrectorado de calidad y administración europea, Instituto de ciencias de la educación, Universidad de Alicante: Marfil.

- Morosini, E. (2012). *Medición escalar de sujeto mediante escalas sumativas. Escalas de actitud tipo Likert*. [Power Point]. Recuperado de <http://escaladeactitud.blogspot.mx/> [17/03/2014].
- Murphy, B. (2009). *Grammaticality Judgments as Linguistic Evidence Alternatives to Judgements. ESSLLI, Bordeaux*. Recuperado de <http://clic.cimec.unitn.it/brian/esslli/alternativeMethods.pdf> [22/05/2014].
- Nadelsticher, A. (1983). *Técnicas para la construcción de cuestionarios de actitudes y opción múltiple*. México: Instituto nacional de ciencias penales.
- NAV(2016). *Regresión y correlación. Fórmulas básicas en la regresión lineal simple*. Universidad Centroamericana José Simeón Cañas. Recuperado de http://www.uca.edu.sv/matematica/upload_w/file/REGRESION%20SIMPLE%20Y%20MULTIPLE.pdf [24/09/2014].
- Nist/Sematech (2012). *Engineering Statistics Handbook*. Recuperado de <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm> [14/09/2014].
- Nunan, D. (1999). *Research Methods in Language Learning*. Cambridge: Cambridge University Press.
- Payrató, Ll. (1998). "El Supermercado lingüístico", en *De Profesión, lingüista. Panorama de la Lingüística Aplicada (SIC)*. Barcelona: Ariel. Recuperado de <http://ced.cele.unam.mx/linguistica/framesu2.htm> [20/05/2014].
- Pennycook, A. (2001). *Critical Applied Linguistics; a Critical Introduction*. N.J.: L. Erlbaum.
- Profe de lengua (s/f). *Relaciones entre la lingüística y otras ciencias*. Recuperado de http://www.profedelengua.es/Relaciones_linguistica-ciencias.pdf [31/05/2014].
- Quezada, A. y E. Canessa (2008). "La complejidad de los procesos educativos en el aula de clases", en *Educación en Revista*, 32, 103-119. Chile: Universidad Adolfo Ibáñez. Recuperado de <http://ojs.c3sl.ufpr.br/ojs2/index.php/educar/article/view/13922/9376> [04/06/2014].

- Rada, G. (2007). *Intervalos de confianza*. Recuperado de <http://escuela.med.puc.cl/Recursos/recepidem/EPIANAL9.HTM> [04/07/2014].
- Raymundo, R. (2014a). *Pruebas paramétricas. Psicología para estudiantes*. Recuperado de http://www.ray-design.com.mx/psicoparaest/index.php?option=com_content&view=category&id=52&Itemid=61 [04/06/2014].
- _____ (2014b). *Prueba t de Student para datos no relacionados (muestras independientes). Psicología para estudiantes*. Recuperado de http://www.ray-design.com.mx/psicoparaest/index.php?option=com_content&view=article&id=233:t-student-dn-r&catid=52:pruebaspara&Itemid=61 [25/09/2014].
- _____ (2014c). *Prueba t de Student para datos relacionados (muestras dependientes). Psicología para estudiantes*. Recuperado de http://www.ray-design.com.mx/psicoparaest/index.php?option=com_content&view=article&id=232:t-student-dr&catid=52:pruebaspara&Itemid=61 [25/09/2014].
- _____ (2014d). *Prueba t de Student-Welch para dos muestras independientes con varianzas no homogéneas. Psicología para estudiantes*. Recuperado de http://www.ray-design.com.mx/psicoparaest/index.php?option=com_content&view=article&id=234:t-student-welch&catid=52:pruebaspara&Itemid=61 [25/09/2014].
- Reyes, J. (2003). "¿Cómo se ha desarrollado la Lingüística Aplicada?", en *Introducción a la lingüística aplicada*. Recuperado de <http://ced.cele.unam.mx/linguistica/framesu2.htm> [20/05/2014].
- Rivera, R. (2011). *Las Variables. PP, Slideshare*. Recuperado de <http://es.slideshare.net/profesorrene/las-variables-en-una-investigacion-cientifica> [28/05/2014].
- Rodríguez, G. et al. (1999). *Metodología de la investigación cualitativa*. Málaga: Aljibe.
- Rodríguez, R. (2003). *¿Cómo analizar cuantitativamente datos cualitativos?; análisis de datos cualitativos*. Recuperado de <http://www.gestiopolis.com/como-analizar-cuantitativamente-datos-cualitativos/> [27/10/2015].

- Salamanca, A. y C. Martín-Crespo (2007). "El diseño en la investigación cualitativa", en *Revista NURE Investigación*, 26. Recuperado de http://www.nureinvestigacion.es/ficheros_administrador/f_metodologica/fmetodologica_26.pdf [11/06/2014].
- Salcedo, R. (2002). "Experiencias docentes, calidad y cambio escolar. Investigación e innovación en el aula", en *Biblioteca Luis Ángel Arango*. Recuperado de <http://www.banrepultural.org/blaavirtual/educacion/expedocen/expedocen8a.htm> [07/11/2014].
- Salgado, A. (2007). "Investigación cualitativa: diseños, evaluación del rigor metodológico y retos", en *Liberabit*, 13, 71-78. Recuperado de <http://www.scielo.org.pe/pdf/liber/v13n13/a09v13n13.pdf> [07/09/2014].
- Salkind, N. (2008). *Statistics for people who (think they) hate statistics*. Thousand Oaks: Sage Publications.
- Schmitt, N. (2000). *An Introduction to Applied Linguistics*. Londres: Arnold.
- Schmitt, N. y M. Celce-Murcia (2000). "An overview of Applied Linguistics", en *An Introduction to Applied Linguistics*. Londres: Arnold, 1-18.
- Statistics Solutions (2013). *ANOVA (Analysis of Variance)*. Recuperado de <https://www.statisticssolutions.com/manova-analysis-anova/> [07/09/2014].
- Suárez, M. (2011a). "Coeficiente de correlación de Karl Pearson", en *Monografías.com*. Recuperado de <http://www.monografias.com/trabajos85/coeficiente-correlacion-karl-pearson/coeficiente-correlacion-karl-pearson.shtml> [20/09/2014].
- _____ (2011b). "Coeficiente de correlación por rangos de Spearman", en *Monografías.com*. Recuperado de <http://www.monografias.com/trabajos85/coeficiente-correlacion-rangos-spearman/coeficiente-correlacion-rangos-spearman.shtml> [20/09/2014].
- _____ (2011c). "Medidas de forma: asimetría y curtosis", en *Monografías.com*. Recuperado de <http://www.monografias.com/trabajos87/medidas-forma-asimetria-curtosis/medidas-forma-asimetria-curtosis.shtml#curtosisoa> [12/09/2014].

- _____ (2012a). "Prueba de hipótesis con la razón F de Fisher", en *Monografías.com*. Recuperado de <http://www.monografias.com/trabajos91/prueba-hipotesis-f-fisher-empleando-excel-y-winstats/prueba-hipotesis-f-fisher-empleando-excel-y-winstats.shtml> [04/09/2014].
- _____ (2012b). "Interaprendizaje de probabilidades y estadística inferencial con Excel, Winstats y Graph", en *Monografías.com*. Recuperado de <http://www.monografias.com/trabajos91/prueba-hipotesis-f-fisher-empleando-excel-y-winstats/prueba-hipotesis-f-fisher-empleando-excel-y-winstats.shtml> [04/10/2014].
- Taylor, S. y R. Bogdan (1990). "Introducción a los métodos cualitativos de investigación", en A. Salgado. *Investigación cualitativa: diseños, evaluación del rigor metodológico y retos*. Recuperado de <http://www.scielo.org.pe/pdf/liber/v13n13/a09v13n13.pdf> [07/09/2014].
- The American Heritage New Dictionary of Cultural Literacy (2005). Published by Houghton Mifflin Company. Recuperado de <http://dictionary.reference.com/browse/statistical-significance> [07/09/2014].
- Universidad de Córdoba / Departamento de Ciencias del Lenguaje (2014). *Programa del doctorado en ciencias del lenguaje*. Recuperado de http://www.uco.es/ciencias_lenguaje/ [16/06/2014].
- Universidad de Coruña (2014). *Tablas estadísticas*. Recuperado de <http://www.udc.gal/dep/mate/estadistica2/documentos-pdf/dmtablas.pdf> [27/08/2014].
- Universidad de Málaga (s/f). *Bioestadística: métodos y aplicaciones, versión electrónica*. Recuperado de <http://www.bioestadistica.uma.es/libro/node22.htm> [11/09/2014].
- Universidad Nacional de Córdoba (2014). *Programa del Doctorado en ciencias del lenguaje*. Recuperado de http://www.lenguas.unc.edu.ar/posgrado/posgrado_doctorado.html [16/06/2014].
- Universidad Pompeu Fabra (2014). *Programa del Doctorado en traducción y ciencias del lenguaje*. Recuperado de <http://www.upf.edu/doc-torats/es/programes/comunicacio/presentacio/> [16/06/2014].

- Universidad Veracruzana (2014a). *Programa del Doctorado en Estudios del lenguaje y lingüística aplicada*. Recuperado de <http://www.uv.mx/idiomas/posgrados/dcl/> [16/06/2014].
- _____ (2014b). *Programa del Doctorado en ciencias el lenguaje*. Recuperado de <http://www.uv.mx/idiomas/posgrados/dcl/> [16/06/2014].
- Uribe, C. (2002). *Relación del estilo de enseñanza del profesor y las actitudes de los alumnos hacia la lengua inglesa: un estudio en la Facultad de Contaduría y Administración de la UAEM*, tesis de Maestría. México: Universidad Nacional Autónoma de México.
- Uribe, C. (2010). *Aptitud y rendimiento en el aprendizaje de la lengua inglesa como lengua extranjera*, tesis Doctoral. México: Universidad Nacional Autónoma de México.
- Valdéz, I. (s/f). *Tabla 2. Valores de χ^2 de la distribución Ji-cuadrada*. Recuperado de http://www.dcb.unam.mx/profesores/irene/Notas/tablas/Ji_cuadrada.pdf [14/10/2014].
- Van Lier, L. (1988). *The classroom and the language learner: Ethnography and Second Language Classroom Research*. London: Longman.
- Vasilachis de Gialdino, I. (coord.) (2007). *Estrategias de investigación cualitativa*. Madrid: Gedisa.
- Vila, A. et al. (2014). *Correlación lineal y análisis de regresión*. Recuperado de <http://es.scribd.com/doc/216720470/Re-Gresi-on-Lineal> [15/04/2014].
- Villavicencio, L. (2008). *Las variables. Definición con ejemplo*. Recuperado de <http://www.monografias.com/trabajos64/variables-definicion-ejemplo/variables-definicion-ejemplo2.shtml> [29/02/2014].
- Visauta, B. (2007). *Análisis estadístico con SPSS 14; estadística básica*. 3ª ed., Madrid: McGrawHill.
- Vitutor (2010). *Cuartiles, deciles y percentiles*. Recuperado de http://www.vitutor.net/2/11/cuartiles_percentiles.html [11/09/2014].

WiŚniewski, K. (2007). *Applied Linguistics*. Recuperado de <http://www.tlumaczenia-angielski.info/linguistics/applied-linguistics.htm> [06/05/2014].

Word Reference (2014). *Language Dictionaries. Diccionario de la lengua española*. Recuperado de <http://www.wordreference.com/definicion/coeficiente> [10/04/2014].



Anexos

Anexo A

Ciencias del lenguaje y sus componentes, de acuerdo con la propia concepción de Álvarez (2005), presentadas en nueve cuadros, uno para cada una de ellas.

Para la Psicolingüística, se tiene el siguiente cuadro.

Cuadro A.1. Componentes de la Psicolingüística

Ciencia del lenguaje	Componentes	Subcomponentes
Psicolingüística	Conducta lingüística y procesos psicológicos	
	Relación Lenguaje-pensamiento	
	Actividad discursiva	Representaciones del discurso en la memoria Análisis estructural
	Procesos cognitivos	Comprensión de textos Producción de textos Macroestructuras semánticas Operaciones mentales y Macroreglas Marcos cognitivos y coherencia del discurso Superestructuras del contenido semántico
	Memoria	Memoria de corto plazo, almacén de información fonológica, morfológica y sintáctica Memoria de largo plazo, información semántica Almacenamiento estructurado, operaciones de conocer y recordar para la comprensión textual

Fuente: Álvarez (2005).

En el cuadro A.2., se muestran los componentes considerados para la Semiótica.

Cuadro A.2. Componentes de la Semiótica

Ciencia del lenguaje	Componentes	Subcomponentes
Semiótica	Sistemas de signos para comunicarse	
	Unidades y relaciones	
	Articuladora de investigaciones y la lingüística como ciencia social	
	Consideración del signo	Significado, semántica Secuencia, sintaxis Uso y efectos en destinatarios, pragmática
	Discursos	Análisis narrativo Significado del texto
	Fundamentación dialéctica de la lengua	¿Cómo hablar en un debate? ¿Cómo llamar la atención de los interlocutores?
	Dar sentido a lo que se dice	

Fuente: Álvarez (2005).

El tercer cuadro muestra la Retórica.

Cuadro A.3. Componentes de la Retórica

Ciencia del lenguaje	Componentes	Subcomponentes
Retórica	Hablar bien para persuadir	
	Entre la dialéctica y a gramática, Aristóteles	
	Discurso oral, emisor, receptor y contenido del mensaje	
	Pragmática	Elementos del acto Comunicativo Discurso Enunciados Uso del lenguaje, Lingüística del texto
	Análisis crítico	
	Análisis totalizador del texto	
	Errores argumentativos	Argumentos confiables

Fuente: Álvarez (2005).

El cuarto cuadro muestra la Pragmática.

Cuadro A.4. Componentes de la Pragmática

Ciencia del lenguaje	Componentes	Subcomponentes
Pragmática	Explicación de la interpretación en habla real	
	El lenguaje como actividad	
	El lenguaje generado	
	Sistema multifuncional generado por la intención del hablante	
	Relaciones entre estructuras de texto y contexto	
	Condiciones y reglas para los actos de habla	
	Discurso	Construcción colectiva Realización interactiva Contribución a la interacción social Reglas para el de los elementos lingüísticos
	Lingüística interaccionista	Teorías pragmáticas Teoría de los actos de habla
	Lenguaje y comunicación	Comunicación humana Contexto y comprensión Conocimiento mutuo Principios de cooperación
	Locutor	Marca en el enunciado Inscripción en el mensaje Relación con el mensaje
Alumno	Interprete lo que se dice verbalmente Aprenda a declarar algo Se inmiscuya en lo que declara	

Fuente: Álvarez (2005).

El quinto cuadro presenta la Enunciación.

Cuadro A.5. Componentes de la Enunciación

Ciencia del lenguaje	Componentes	Subcomponentes	Subsubcomponentes	
Enunciación	Puesta en funcionamiento de una lengua			
	Acto individual, consideración del hablante en la realización del discurso			
	Conversión individual de la lengua en enunciado discursivo			
	Hablante, manifestaciones y lengua			
	Huellas de la presencia del lector	Deixis		
		Marcas de modalización		
		Conectores		
		Habla y sistema		
		Hablante		
		Relaciones enunciado y protagonista del discurso		
Relaciones enunciado y situación comunicativa				
Procedimientos para el emisor		Dejar su importancia en el enunciado Inscribirse en él Relacionarse con el mensaje		

Fuente: Álvarez (2005).



En el sexto cuadro está la Sociolingüística.

Cuadro A.6. Componentes de la Sociolingüística

Ciencia del lenguaje	Componentes	Subcomponentes
Sociolingüística	Lenguaje y sociedad	
	Entre la lingüística y la antropología y la sociología	
	Lenguaje como producto social	
	Lenguaje como forma de conducta social	
	Vínculo entre la forma del discurso y la conducta social	
	Vínculo entre el lenguaje y otros sistemas simbólicos	
	Vínculo entre los miembros reales de la comunidad de hablantes	
	Lengua y contexto sociocultural	
	Objeto de estudio	Habla en su aspecto cualitativo Bilingüismo Lenguas en contacto Diglosia Desigualdad lingüística y social Competencia comunicativa
	Registros de lengua	

Fuente: Álvarez (2005).

El séptimo cuadro se refiere a la Etnografía de la comunicación.

Cuadro A.7. Componentes de la Etnografía de la comunicación

Ciencia del lenguaje	Componentes	Subcomponentes	Subsubcomponentes	
Etnografía de la Comunicación	Hecho comunicativo o de habla			
	Intercambio verbal			
	Enfoques	Antropología cultural		
		Sociología		
		Lingüística		
		Relación lengua y cultura		
		Relación con los sistemas locales de conducta y conocimiento		
		Competencia comunicativa condicionada social y culturalmente		
		Interacción	Estrategias verbales del hablante cortés Tiempo para la adquisición de los sistemas fonológico y sintáctico	
		Cooperación conversacional	Conocimiento de las convenciones lingüísticas y comunicativas Gramática Contextualización	
Hecho comunicativo				

Fuente: Álvarez (2005).

El cuadro A.8 contiene lo relativo al Análisis del discurso.

Cuadro A.8. Componentes del Análisis del discurso

Ciencia del lenguaje	Componentes	Subcomponentes
Análisis del discurso	Organizaciones transoracionales del lenguaje	Diálogo Conversación Texto escrito
	Atención multidisciplinaria al estudio del discurso	
	Bases en: antropología, lingüística, semiótica, poética, psicología, sociología y comunicación y comunicación de masas	
	Análisis de lengua en su uso comunicativo	
	Uso del lenguaje en contextos sociales	Diálogo entre hablantes
	Aspectos del lenguaje	
	Problemas en la lingüística del discurso	
	Dimensión pragmática	
	Análisis crítico del discurso	

Fuente: Álvarez (2005).

Por último, en el cuadro A.9, está la Lingüística textual.

Cuadro A.9. Componentes de la Lingüística textual

Ciencia del lenguaje	Componentes
Lingüística textual	<p>Trascender la oración como unidad completa de significado Atender al discurso/texto y su contexto de producción</p> <p>Observación del lenguaje como instrumento de comunicación Unidad textual del lenguaje dentro del uso real Unidades completas de significado, acontecimientos comunicativos Coherencia y cohesión textuales</p> <p>Búsqueda y análisis de regularidades lingüísticas y textuales Hablar y escribir más y mejor</p>

Fuente: Álvarez (2005).

Anexo B

Ciencias del lenguaje a través de instituciones que ofrecen estudios de Doctorado con esa denominación.

Cuadro B.1. Ciencias del Lenguaje a través de instituciones que ofrecen estudios de Doctorado con esa denominación

Universidad y Facultad	Posgrado	Áreas	Cursos
Benemérita Universidad Autónoma de Puebla, México Instituto de Ciencias sociales y humanidades	Doctorado en Ciencias del Lenguaje	<ul style="list-style-type: none"> • Lengua y literatura: lenguas autóctonas • Problemas relacionados con el manejo del lenguaje 	<ul style="list-style-type: none"> • Ciencias del lenguaje I y II • Adquisición y desarrollo del Lenguaje • Descripción de lenguas • Pragmática de la comunicación • Organización del discurso • Psicolingüística • Neurolingüística • Hermenéutica y fenomenología • Constructivismo y ciencias cognitivas • Lingüística cognitiva. Interacción socio verbal (discurso, cultura y sociedad) • Discursos e instituciones sociales • Adquisición de segunda lengua • Sociología del lenguaje y políticas lingüísticas • Filosofía del lenguaje • Estética de la recepción y discurso literario
ENAH Escuela Nacional de Antropología e Historia, México	Doctorado en Ciencias del Lenguaje	Seminarios y grupos de estudio relacionados con el español	<ul style="list-style-type: none"> • Enfoque funcional y tipológico • Español frente a lenguas románicas e indígenas de México
ENAH Escuela Nacional de Antropología e Historia, México	Doctorado en Ciencias del Lenguaje	<ul style="list-style-type: none"> • Análisis gramatical del español: sincronía y diacronía • Sociolingüística interaccional 	<ul style="list-style-type: none"> • Fenómenos de variación y cambio lingüístico del español colonial mexicano • Interacción verbal bilingüe
Universidad Veracruzana, México	Doctorado en Estudios del Lenguaje y Lingüística Aplicada Doctorado en Ciencias del Lenguaje	<ul style="list-style-type: none"> • Diversidad lingüística • Enseñanza de lenguas • Políticas del lenguaje de las lenguas y culturas nacionales y extranjeras 	<ul style="list-style-type: none"> • Investigación etnográfica • Estudios del lenguaje • Divulgación del conocimiento
Universidad Pompeu Fabra, Barcelona, España	Doctorado en Traducción y Ciencias del Lenguaje	<ul style="list-style-type: none"> • Adquisición del lenguaje • Análisis del discurso • Aprendizaje de idiomas • Estudios de traducción • Lexicología y Lexicografía • Lingüística computacional e ingeniería del lenguaje • Lingüística formal y descriptiva y variación • Recepción y traducción literarias • Terminología y gestión del conocimiento • Traducción especializada 	<ul style="list-style-type: none"> • No los reporta



<p>Universidad Pompeu Fabra, Barcelona, España</p> <p>IULA Instituto Universitario de Lingüística Aplicada</p>	<p>Doctorado en Ciencias del Lenguaje y LA</p>	<ul style="list-style-type: none"> • Aspectos aplicados de la teoría lingüística y diseño de aplicaciones específicas • Temas transversales de comunicación y discurso que impliquen el lenguaje ... 	<ul style="list-style-type: none"> • Lingüística teórica • Lingüística aplicada • Lingüística computacional e ingeniería lingüística • Lexicografía • Terminología • Variación lingüística • Discurso general y especializado • Gestión del conocimiento • Lingüística forense
<p>Universidad de Córdoba, España</p> <p>Facultad de Filosofía y letras</p> <p>Departamento de ciencias del lenguaje</p>	<ul style="list-style-type: none"> • Ciencias del Lenguaje • Docencia e investigación 	<ul style="list-style-type: none"> • Didáctica de la lengua y la literatura • Filología italiana • Lengua española • Lingüística general • Teoría de la literatura y literatura comparada <p>La docencia y la investigación se desarrollan indistintamente en las Facultades de:</p> <ul style="list-style-type: none"> • Filosofía y letras • Ciencias de la educación • Ciencias del trabajo • Derecho y Ciencias Económicas y Empresariales 	<ul style="list-style-type: none"> • No lo reporta
<p>Universidad Nacional de Córdoba, Argentina</p> <p>Facultad de Lenguas</p>	<p>Doctorado en Ciencias de Lenguaje</p>	<ul style="list-style-type: none"> • Culturas y literaturas comparadas • Lingüística aplicada • Traductología 	<ul style="list-style-type: none"> • Estructurados • No estructurados

Anexo D

Tabla para la prueba estadística Kolmogorov-Smirnov. Para esta prueba se tomó la tabla que presenta Alford (2011).

Las instrucciones para su uso indican que se consulta utilizando el número de sujetos o informantes y el nivel de significancia deseado.

Tabla D.1. Tabla para la prueba estadística Kolmogorov-Smirnov

N	Nivel de significancia				
	0.2	0.1	0.05	0.02	0.01
1	0.900	0.950	0.975	0.990	0.995
2	0.684	0.776	0.842	0.900	0.929
3	0.565	0.636	0.780	0.785	0.829
4	0.493	0.565	0.624	0.689	0.734
5	0.447	0.509	0.563	0.627	0.669
6	0.410	0.468	0.519	0.577	0.617
7	0.381	0.436	0.483	0.538	0.576
8	0.358	0.410	0.454	0.507	0.542
9	0.339	0.387	0.430	0.480	0.513
10	0.323	0.369	0.409	0.457	0.489
11	0.308	0.352	0.391	0.437	0.468
12	0.296	0.338	0.375	0.419	0.449
13	0.285	0.325	0.365	0.404	0.432
14	0.275	0.314	0.349	0.390	0.418
15	0.266	0.304	0.338	0.377	0.404
16	0.258	0.295	0.327	0.366	0.392
17	0.250	0.286	0.318	0.355	0.381
18	0.244	0.279	0.309	0.346	0.371
19	0.237	0.271	0.301	0.337	0.361
20	0.232	0.265	0.294	0.329	0.352
21	0.226	0.259	0.287	0.321	0.344
22	0.221	0.253	0.281	0.314	0.337
23	0.216	0.247	0.275	0.307	0.330
24	0.212	0.242	0.269	0.301	0.323
25	0.208	0.238	0.264	0.295	0.317
26	0.204	0.233	0.259	0.290	0.311
27	0.200	0.239	0.254	0.284	0.305
28	0.197	0.235	0.250	0.279	0.300
29	0.193	0.221	0.246	0.275	0.295
30	0.190	0.218	0.242	0.270	0.290
31	0.187	0.214	0.238	0.266	0.285
32	0.184	0.211	0.234	0.262	0.281
33	0.182	0.208	0.231	0.258	0.277
34	0.179	0.205	0.227	0.254	0.273
35	0.177	0.202	0.224	0.251	0.269
36	0.174	0.199	0.221	0.247	0.265
37	0.172	0.196	0.218	0.244	0.262
38	0.170	0.194	0.215	0.241	0.258
39	0.168	0.191	0.213	0.238	0.255
40	0.165	0.189	0.210	0.235	0.252
	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.66}{\sqrt{N}}$

Anexo E

Tabla para la prueba pruebas estadística t. Esta tabla fue tomada de las Tablas estadísticas publicadas por la Universidad de la Coruña, sin fecha de publicación. Fue simplificada para considerar solamente los valores que corresponden a los niveles de significancia .05 y .01.

Tabla E.1. Distribución de la prueba t

N	0.050	0.01	N	0.050	0.01
1	6.3137	31.821	17	1.7396	2.5669
2	2.9200	6.9645	18	1.7341	2.5524
3	2.3534	4.5407	19	1.7291	2.5395
4	2.1318	3.7469	20	1.7247	2.5280
5	2.0150	3.3649	21	1.7207	2.5176
6	1.9432	3.1427	22	1.7171	2.5083
7	1.8946	2.9979	23	1.7139	2.4999
8	1.8595	2.8965	24	1.7109	2.4922
9	1.8331	2.8214	25	1.7081	2.4851
10	1.8125	2.7638	26	1.7056	2.4786
11	1.7959	2.7181	27	1.7033	2.4727
12	1.7823	2.6810	28	1.7011	2.4671
13	1.7709	2.6503	29	1.6991	2.4620
14	1.7613	2.6245	30	1.6973	2.4573
15	1.7531	2.6025	∞	1.6449	2.3264
16	1.7459	2.5835	-	-	

La primera y cuarta columnas de la tabla corresponden a los grados de libertad utilizados y, la segunda y tercera, y quinta y sexta, a los valores de t para .05 y .01 de significancia.

Anexo F

Tabla para la prueba estadística F. Para esta prueba se tomaron dos tablas, para los niveles de significancia .05 y .01, del sitio de El osio de los santos (s/f). Las instrucciones para su uso indican que: "Los grados de libertad del numerador se encuentran en la primera fila, en tanto que los grados de libertad del denominador se encuentran en la primera columna, cada tabla corresponde a un valor diferente de alfa o 'p'".

Tabla F.1. Valores de F para nivel de significancia .05

Alfa =	Grados de libertad del numerador													
	1	2	3	4	5	6	7	8	9	10	20	60	100	10000
1	161.45	199.5	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	248.02	252.2	253.04	254.3
2	18.513	19	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396	19.446	19.479	19.486	19.496
3	10.128	9.5521	9.2766	9.1172	9.0134	8.9407	8.8867	8.8452	8.8123	8.7855	8.6602	8.572	8.5539	8.5267
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.041	5.9988	5.9644	5.8025	5.6878	5.664	5.6284
5	6.6079	5.7861	5.4094	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.5581	4.4314	4.4051	4.3654
6	5.9874	5.1432	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.099	4.06	3.8742	3.7398	3.7117	3.6693
7	5.5915	4.7374	4.3468	4.1203	3.9715	3.866	3.7871	3.7257	3.6767	3.6365	3.4445	3.3043	3.2749	3.2302
8	5.3176	4.459	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.1503	3.0053	2.9747	2.9281
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	2.9365	2.7872	2.7556	2.7072
10	4.9646	4.1028	3.7083	3.478	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.774	2.6211	2.5884	2.5384
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.948	2.8962	2.8536	2.6464	2.4901	2.4566	2.405
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.5436	2.3842	2.3498	2.2967
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.671	2.4589	2.2966	2.2614	2.207
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.3879	2.2229	2.187	2.1313
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.3275	2.1601	2.1234	2.0664
16	4.494	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.2756	2.1058	2.0685	2.0102
17	4.4513	3.5915	3.1968	2.9647	2.81	2.6987	2.6143	2.548	2.4943	2.4499	2.2304	2.0584	2.0204	1.961
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.1906	2.0166	1.978	1.9175
19	4.3808	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.1555	1.9795	1.9403	1.8787
20	4.3513	3.4928	3.0984	2.8661	2.7109	2.599	2.514	2.4471	2.3928	2.3479	2.1242	1.9464	1.9066	1.8438
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3661	2.321	2.096	1.9165	1.8761	1.8124
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.0707	1.8894	1.8486	1.7838
23	4.2793	3.4221	3.028	2.7955	2.64	2.5277	2.4422	2.3748	2.3201	2.2747	2.0476	1.8648	1.8234	1.7577
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.0267	1.8424	1.8005	1.7338
25	4.2417	3.3852	2.9912	2.7587	2.603	2.4904	2.4047	2.3371	2.2821	2.2365	2.0075	1.8217	1.7794	1.7117
26	4.2252	3.369	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	1.9898	1.8027	1.7599	1.6913
27	4.21	3.3541	2.9603	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	1.9736	1.7851	1.7419	1.6724
28	4.196	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.236	2.19	1.9586	1.7689	1.7251	1.6548
29	4.183	3.3277	2.934	2.7014	2.5454	2.4324	2.3463	2.2782	2.2229	2.1768	1.9446	1.7537	1.7096	1.6384
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	1.9317	1.7396	1.695	1.623
40	4.0847	3.2317	2.8387	2.606	2.4495	2.3359	2.249	2.1802	2.124	2.0773	1.8389	1.6373	1.5892	1.5098
50	4.0343	3.1826	2.79	2.5572	2.4004	2.2864	2.1992	2.1299	2.0733	2.0261	1.7841	1.5757	1.5249	1.4392
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.097	2.0401	1.9926	1.748	1.5343	1.4814	1.3903
70	3.9778	3.1277	2.7355	2.5027	2.3456	2.2312	2.1435	2.0737	2.0166	1.9689	1.7223	1.5046	1.4498	1.354
80	3.9604	3.1108	2.7188	2.4859	2.3287	2.2142	2.1263	2.0564	1.9991	1.9512	1.7032	1.4821	1.4259	1.3259
90	3.9469	3.0977	2.7058	2.4729	2.3157	2.2011	2.1131	2.043	1.9856	1.9376	1.6883	1.4645	1.407	1.3032
100	3.9362	3.0873	2.6955	2.4626	2.3053	2.1906	2.1025	2.0323	1.9748	1.9267	1.6764	1.4504	1.3917	1.2845
200	3.8884	3.0411	2.6498	2.4168	2.2592	2.1441	2.0556	1.9849	1.9269	1.8783	1.6233	1.3856	1.3206	1.1903
300	3.8726	3.0258	2.6347	2.4017	2.2441	2.1288	2.0402	1.9693	1.9112	1.8623	1.6057	1.3634	1.2958	1.1521
400	3.8648	3.0183	2.6272	2.3943	2.2366	2.1212	2.0325	1.9616	1.9033	1.8544	1.5969	1.3522	1.2831	1.1303
500	3.8601	3.0138	2.6227	2.3898	2.232	2.1167	2.0279	1.9569	1.8986	1.8496	1.5916	1.3455	1.2753	1.1159
600	3.857	3.0107	2.6198	2.3868	2.229	2.1137	2.0248	1.9538	1.8955	1.8465	1.5881	1.341	1.2701	1.1055
700	3.8548	3.0086	2.6176	2.3847	2.2269	2.1115	2.0226	1.9516	1.8932	1.8442	1.5856	1.3377	1.2664	1.0976
800	3.8531	3.007	2.616	2.3831	2.2253	2.1099	2.021	1.95	1.8916	1.8425	1.5837	1.3353	1.2635	1.0912
900	3.8518	3.0057	2.6148	2.3818	2.224	2.1086	2.0197	1.9487	1.8903	1.8412	1.5822	1.3334	1.2613	1.0861
1000	3.8508	3.0047	2.6138	2.3808	2.2231	2.1076	2.0187	1.9476	1.8892	1.8402	1.5811	1.3318	1.2596	1.0818
1500	3.8477	3.0017	2.6108	2.3779	2.2201	2.1046	2.0157	1.9446	1.8861	1.837	1.5775	1.3273	1.2542	1.0675
2000	3.8461	3.0002	2.6094	2.3764	2.2186	2.1031	2.0142	1.943	1.8846	1.8354	1.5758	1.325	1.2516	1.0593
10000	3.8424	2.9966	2.6058	2.3728	2.215	2.0995	2.0105	1.9393	1.8808	1.8316	1.5716	1.3194	1.2451	1.0334

Enseguida se encuentra la tabla para un valor del nivel de significancia de .01.

Tabla F.2. Valores de F para nivel de significancia .01

Alfa =	0.01	Grados de libertad del numerador												
	1	2	3	4	5	6	7	8	9	10	20	60	100	10000
1	4052.2	4999.3	5403.5	5624.3	5764	5859	5928.3	5981	6022.4	6055.9	6208.7	6313	6333.9	6365.6
2	98.502	99	99.164	99.251	99.302	99.331	99.357	99.375	99.39	99.397	99.448	99.484	99.491	99.499
3	34.116	30.816	29.457	28.71	28.237	27.911	27.671	27.489	27.345	27.228	26.69	26.316	26.241	26.126
4	21.198	18	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.019	13.652	13.577	13.464
5	16.258	13.274	12.06	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.5527	9.202	9.13	9.0215
6	13.745	10.925	9.7796	9.1484	8.7459	8.466	8.26	8.1017	7.976	7.8742	7.3958	7.0568	6.9867	6.8811
7	12.246	9.5465	8.4513	7.8467	7.4604	7.1914	6.9929	6.8401	6.7188	6.6201	6.1555	5.8236	5.7546	5.6506
8	11.259	8.6491	7.591	7.0061	6.6318	6.3707	6.1776	6.0288	5.9106	5.8143	5.3591	5.0316	4.9633	4.8599
9	10.562	8.0215	6.992	6.4221	6.0569	5.8018	5.6128	5.4671	5.3511	5.2565	4.808	4.4831	4.415	4.3116
10	10.044	7.5595	6.5523	5.9944	5.6364	5.3858	5.2001	5.0567	4.9424	4.8491	4.4054	4.0819	4.0137	3.91
11	9.6461	7.2057	6.2167	5.6683	5.316	5.0692	4.886	4.7445	4.6315	4.5393	4.099	3.7761	3.7077	3.6035
12	9.3303	6.9266	5.9525	5.4119	5.0644	4.8205	4.6395	4.4994	4.3875	4.2961	3.8584	3.5355	3.4668	3.3619
13	9.0738	6.7009	5.7394	5.2053	4.8616	4.6203	4.441	4.3021	4.1911	4.1003	3.6646	3.3413	3.2723	3.1665
14	8.8617	6.5149	5.5639	5.0354	4.695	4.4558	4.2779	4.14	4.0297	3.9394	3.5052	3.1813	3.1118	3.0051
15	8.6832	6.3588	5.417	4.8932	4.5556	4.3183	4.1416	4.0044	3.8948	3.8049	3.3719	3.0471	2.9772	2.8695
16	8.5309	6.2263	5.2922	4.7726	4.4374	4.2016	4.0259	3.8896	3.7804	3.6909	3.2587	2.933	2.8627	2.7539
17	8.3998	6.1121	5.185	4.6689	4.336	4.1015	3.9267	3.7909	3.6823	3.5931	3.1615	2.8348	2.7639	2.6542
18	8.2855	6.0129	5.0919	4.579	4.2479	4.0146	3.8406	3.7054	3.5971	3.5081	3.0771	2.7493	2.6779	2.5671
19	8.185	5.9259	5.0103	4.5002	4.1708	3.9386	3.7653	3.6305	3.5225	3.4338	3.0031	2.6742	2.6023	2.4905
20	8.096	5.849	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567	3.3682	2.9377	2.6077	2.5353	2.4224
21	8.0166	5.7804	4.874	4.3688	4.0421	3.8117	3.6396	3.5056	3.3982	3.3098	2.8795	2.5484	2.4755	2.3615
22	7.9453	5.719	4.8166	4.3134	3.988	3.7583	3.5866	3.453	3.3458	3.2576	2.8274	2.4951	2.4218	2.3067
23	7.8811	5.6637	4.7648	4.2635	3.9392	3.7102	3.539	3.4057	3.2986	3.2106	2.7805	2.4471	2.3732	2.2571
24	7.8229	5.6136	4.7181	4.2185	3.8951	3.6667	3.4959	3.3629	3.256	3.1681	2.738	2.4035	2.3291	2.2119
25	7.7698	5.568	4.6755	4.1774	3.855	3.6272	3.4568	3.3239	3.2172	3.1294	2.6993	2.3637	2.2888	2.1706
26	7.7213	5.5263	4.6365	4.14	3.8183	3.5911	3.421	3.2884	3.1818	3.0941	2.664	2.3273	2.2519	2.1327
27	7.6767	5.4881	4.6009	4.1056	3.7847	3.558	3.3882	3.2558	3.1494	3.0618	2.6316	2.2938	2.218	2.0978
28	7.6357	5.4529	4.5681	4.074	3.7539	3.5276	3.3581	3.2259	3.1195	3.032	2.6018	2.2629	2.1867	2.0655
29	7.5977	5.4205	4.5378	4.0449	3.7254	3.4995	3.3303	3.1982	3.092	3.0045	2.5742	2.2344	2.1577	2.0355
30	7.5624	5.3903	4.5097	4.0179	3.699	3.4735	3.3045	3.1726	3.0665	2.9791	2.5487	2.2079	2.1307	2.0075
40	7.3142	5.1785	4.3126	3.8283	3.5138	3.291	3.1238	2.993	2.8876	2.8005	2.3689	2.0194	1.9383	1.8061
50	7.1706	5.0566	4.1994	3.7195	3.4077	3.1864	3.0202	2.89	2.785	2.6981	2.2652	1.909	1.8248	1.6847
60	7.0771	4.9774	4.1259	3.6491	3.3389	3.1187	2.953	2.8233	2.7185	2.6318	2.1978	1.8363	1.7493	1.6023
70	7.0114	4.9218	4.0744	3.5997	3.2907	3.0712	2.906	2.7765	2.6719	2.5852	2.1504	1.7846	1.6954	1.5422
80	6.9626	4.8807	4.0363	3.5631	3.2551	3.0361	2.8713	2.742	2.6374	2.5508	2.1153	1.7459	1.6548	1.496
90	6.9251	4.8491	4.0069	3.535	3.2276	3.0091	2.8445	2.7154	2.6109	2.5243	2.0882	1.7158	1.6231	1.4593
100	6.8953	4.8239	3.9837	3.5127	3.2059	2.9877	2.8233	2.6943	2.5898	2.5033	2.0666	1.6918	1.5977	1.4292
200	6.7633	4.7128	3.881	3.4143	3.11	2.8933	2.7298	2.6012	2.4971	2.4106	1.9713	1.5833	1.4811	1.2812
300	6.7201	4.6766	3.8475	3.3822	3.0787	2.8625	2.6993	2.5709	2.4668	2.3804	1.9401	1.5468	1.441	1.2228
400	6.6987	4.6586	3.8309	3.3664	3.0632	2.8472	2.6842	2.5559	2.4518	2.3654	1.9245	1.5285	1.4207	1.19
500	6.6858	4.6479	3.821	3.3569	3.054	2.8381	2.6751	2.5469	2.4429	2.3565	1.9152	1.5174	1.4084	1.1684
600	6.6773	4.6407	3.8144	3.3506	3.0478	2.8321	2.6691	2.5409	2.4369	2.3505	1.9091	1.5101	1.4001	1.1529
700	6.6713	4.6356	3.8097	3.346	3.0434	2.8278	2.6648	2.5367	2.4327	2.3463	1.9047	1.5048	1.3942	1.1411
800	6.6667	4.6318	3.8062	3.3427	3.0402	2.8245	2.6617	2.5335	2.4295	2.3431	1.9013	1.5008	1.3897	1.1318
900	6.6631	4.6288	3.8034	3.3401	3.0376	2.822	2.6592	2.531	2.427	2.3406	1.8988	1.4978	1.3863	1.1242
1000	6.6603	4.6264	3.8012	3.338	3.0356	2.82	2.6572	2.529	2.425	2.3386	1.8967	1.4953	1.3835	1.1178
1500	6.6518	4.6193	3.7947	3.3317	3.0294	2.814	2.6512	2.5231	2.4191	2.3327	1.8906	1.4879	1.3751	1.0969
2000	6.6476	4.6158	3.7914	3.3286	3.0264	2.811	2.6482	2.5201	2.4162	2.3298	1.8875	1.4842	1.3708	1.085
10000	6.6374	4.6073	3.7836	3.321	3.0191	2.8038	2.6411	2.513	2.4091	2.3227	1.8802	1.4752	1.3606	1.0476



Anexo G

Tabla G.1. Tabla de valores de X de la distribución Chi-cuadrada

$$1 - \alpha = P(X \leq x_{\alpha, v})$$

v	1 - α														
	0.001	0.002	0.0025	0.005	0.01	0.02	0.025	0.05	0.1	0.2	0.25	0.3	0.4	0.5	
1	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.004	0.016	0.064	0.102	0.148	0.275	0.455	
2	0.002	0.004	0.005	0.010	0.020	0.040	0.051	0.103	0.211	0.446	0.575	0.713	1.022	1.386	
3	0.024	0.039	0.045	0.072	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424	1.869	2.366	
4	0.091	0.129	0.145	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195	2.753	3.357	
5	0.210	0.280	0.307	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000	3.656	4.351	
6	0.381	0.486	0.527	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828	4.570	5.348	
7	0.599	0.741	0.794	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671	5.493	6.346	
8	0.857	1.038	1.104	1.344	1.647	2.320	2.180	2.733	3.490	4.594	5.071	5.527	6.423	7.344	
9	1.152	1.370	1.450	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393	7.357	8.343	
10	1.479	1.734	1.827	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267	8.295	9.342	
11	1.834	2.126	2.232	2.603	3.053	3.609	3.816	4.575	5.578	6.989	7.584	8.148	9.237	10.341	
12	2.214	2.543	2.661	3.074	3.571	4.178	4.404	5.226	6.304	7.807	8.438	9.034	10.182	11.340	
13	2.617	2.982	3.112	3.565	4.107	4.765	5.009	5.892	7.041	8.634	9.299	9.926	11.129	12.340	
14	3.041	3.440	3.582	4.075	4.660	5.368	5.629	6.571	7.790	9.467	10.165	10.821	12.078	13.339	
15	3.483	3.916	4.070	4.601	5.229	5.985	6.262	7.261	8.547	10.307	11.037	11.721	13.030	14.339	
16	3.942	4.408	4.573	5.142	5.812	6.614	6.908	7.962	9.312	11.152	11.912	12.624	13.983	15.338	
17	4.416	4.915	5.092	5.697	6.408	7.255	7.564	8.672	10.085	12.002	12.792	13.531	14.937	16.338	
18	4.905	5.436	5.623	6.265	7.015	7.906	8.231	9.390	10.865	12.857	13.675	14.440	15.893	17.338	
19	5.407	5.970	6.167	6.844	7.633	8.567	8.907	10.117	11.651	13.716	14.562	15.352	16.850	18.338	
20	5.921	6.514	6.723	7.434	8.260	9.237	9.591	10.851	12.443	14.578	15.452	16.266	17.809	19.337	
21	6.447	7.070	7.289	8.034	8.897	9.915	10.283	11.591	13.240	15.445	16.344	17.182	18.768	20.337	
22	6.983	7.636	7.865	8.643	9.542	10.600	10.982	12.338	14.041	16.314	17.240	18.101	19.729	21.337	
23	7.529	8.211	8.450	9.260	10.196	11.293	11.689	13.091	14.848	17.187	18.137	19.021	20.690	22.337	
24	8.085	8.796	9.044	9.886	10.856	11.992	12.401	13.848	15.659	18.062	19.037	19.943	21.652	23.337	
25	8.649	9.388	9.646	10.520	11.524	12.697	13.120	14.611	16.473	18.940	19.939	20.867	22.616	24.337	
26	9.222	9.989	10.256	11.160	12.198	13.409	13.844	15.379	17.292	19.820	20.843	21.792	23.579	25.336	
27	9.803	10.597	10.873	11.808	12.878	14.125	14.573	16.151	18.114	20.703	21.749	22.719	24.544	26.336	
28	10.391	11.212	11.497	12.461	13.565	14.847	15.308	16.928	18.939	21.588	22.657	23.647	25.509	27.336	
29	10.986	11.833	12.128	13.121	14.256	15.574	16.047	17.708	19.768	22.475	23.567	24.577	26.475	28.336	
30	11.588	12.461	12.765	13.787	14.953	16.306	16.791	18.493	20.599	23.364	24.478	25.508	27.442	29.336	
35	14.688	15.686	16.032	17.192	18.509	20.027	20.569	22.465	24.797	27.836	29.054	30.178	32.282	34.336	
40	17.917	19.032	19.417	20.707	22.164	23.838	24.433	26.509	29.051	32.345	33.660	34.872	37.134	39.335	
45	21.251	22.477	22.899	24.311	25.901	27.720	28.366	30.612	33.350	36.884	38.291	39.585	41.995	44.335	
50	24.674	26.006	26.464	27.991	29.707	31.664	32.357	34.764	37.689	41.449	42.942	44.313	46.864	49.335	
55	28.173	29.605	30.097	31.735	33.571	35.659	36.398	38.958	42.060	46.036	47.610	49.055	51.739	54.335	
60	31.738	33.267	33.791	35.534	37.485	39.699	40.482	43.188	46.459	50.641	52.294	53.809	56.620	59.335	
65	35.362	36.983	37.538	39.383	41.444	43.779	44.603	47.450	50.883	55.262	56.990	58.573	61.506	64.335	
70	39.036	40.747	41.332	43.275	45.442	47.893	48.758	51.739	55.329	59.898	61.698	63.346	66.396	69.334	
75	42.757	44.554	45.169	47.206	49.475	52.039	52.942	56.054	59.795	64.547	66.417	68.127	71.290	74.334	
80	46.520	48.400	49.043	51.172	53.540	56.213	57.153	60.391	64.278	69.207	71.145	72.915	76.188	79.334	
85	50.320	52.282	52.952	55.170	57.634	60.412	61.389	64.749	68.777	73.878	75.881	77.710	81.089	84.334	
90	54.156	56.196	56.892	59.196	61.754	64.635	65.647	69.126	73.291	78.558	80.625	82.511	85.993	89.334	
95	58.022	60.140	60.861	63.250	65.898	68.879	69.925	73.520	77.818	83.248	85.376	87.317	90.899	94.334	
100	61.918	64.110	64.857	67.328	70.065	73.142	74.222	77.929	82.358	87.945	90.133	92.129	95.808	99.334	

v	1 - α												
	0.6	0.7	0.75	0.8	0.9	0.95	0.975	0.98	0.99	0.995	0.9975	0.998	0.999
1	0.708	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.140	9.549	10.827
2	1.833	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	11.983	12.429	13.815
3	2.946	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.320	14.796	16.266
4	4.045	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.424	16.923	18.466
5	5.132	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	18.385	18.908	20.515
6	6.211	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.249	20.791	22.457
7	7.283	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.040	22.601	24.321
8	8.351	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	23.774	24.352	26.124
9	9.414	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	25.463	26.056	27.877
10	10.473	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.112	27.721	29.588
11	11.530	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	28.729	29.354	31.264
12	12.584	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.318	30.957	32.909
13	13.636	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688	29.819	31.883	32.536	34.527
14	14.685	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	33.426	34.091	36.124
15	15.733	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	34.949	35.627	37.698
16	16.780	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	36.456	37.146	39.252
17	17.824	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.718	37.946	38.648	40.791
18	18.868	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	39.422	40.136	42.312
19	19.910	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	40.885	41.610	43.819
20	20.951	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	42.336	43.072	45.314
21	21.992	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	43.775	44.522	46.796
22	23.031	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	45.204	45.961	48.268
23	24.069	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	46.623	47.392	49.728
24	25.106	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	48.034	48.811	51.179
25	26.143	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	49.435	50.223	52.619
26	27.179	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642	48.290	50.829	51.627	54.051
27	28.214	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963	49.645	52.215	53.022	55.475
28	29.249	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.994	53.594	54.411	56.892
29	30.283	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.335	54.966	55.792	58.301
30	31.316	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	56.332	57.167	59.702
35	36.475	38.859	40.223	41.778	46.059	49.802	53.203	54.244	57.342	60.275	63.076	63.955	66.619
40	41.622	44.165	45.616	47.269	51.805	55.758	59.342	60.436	63.691	66.766	69.699	70.617	73.403
45	46.761	49.452	50.985	52.729	57.505	61.656	65.410	66.555	69.957	73.166	76.223	77.179	80.078
50	51.892	54.723	56.334	58.164	63.167	67.505	71.420	72.613	76.154	79.490	82.664	83.656	86.660
55	57.016	59.980	61.665	63.577	68.796	73.311	77.380	78.619	82.292	85.749	89.034	90.061	93.167
60	62.135	65.226	66.981	68.972	74.397	79.082	83.298	84.580	88.379	91.952	95.344	96.403	99.608
65	67.249	70.462	72.285	74.351	79.973	84.821	89.177	90.501	94.422	98.105	101.600	102.691	105.988
70	72.358	75.689	77.577	79.715	85.527	90.531	95.023	96.387	100.425	104.215	107.808	108.930	112.317
75	77.464	80.908	82.858	85.066	91.061	96.217	100.839	102.243	106.393	110.285	113.974	115.124	118.599
80	82.566	86.120	88.130	90.405	96.578	101.879	106.629	108.069	112.329	116.321	120.102	121.280	124.839
85	87.665	91.325	93.394	95.734	102.079	107.522	112.393	113.871	118.236	122.324	126.194	127.401	131.043
90	92.761	96.524	98.650	101.054	107.565	113.145	118.136	119.648	124.116	128.299	132.255	133.489	137.208
95	97.855	101.717	103.899	106.364	113.038	118.752	123.858	125.405	129.973	134.247	138.288	139.546	143.343
100	102.946	106.906	109.141	111.667	118.498	124.342	129.561	131.142	135.807	140.170	144.292	145.576	149.449

Fuente: Valdéz (s/f).



Anexo H

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO FACULTAD DE LENGUAS

Fórmulas para operaciones y pruebas estadísticas

Cuadro H.1. Relación entre operaciones y pruebas estadísticas con los incisos

Operaciones y pruebas estadísticas	Inciso
1. Cuantiles, percentiles, cuartiles y deciles	3.2.2.1
2. Media	3.2.2.2
3. Mediana	3.2.2.2
4. Moda	3.2.2.2
5. Rango	3.2.2.3
6. Desviación estándar	3.2.2.3
7. Varianza	3.2.2.3
8. Curtosis	3.2.2.4
9. Asimetría	3.2.2.4
10. Prueba de normalidad. Kolmogorov-Smirnov	4.1.4
11. Prueba de Levene	4.1.5
12. Alfa de Cronbach	4.2.1
13. Correlación	4.2.2
14. Determinación	4.2.3
15. Regresión lineal	4.2.4
16. Regresión múltiple	4.2.5
17. T de Student	5.1

1. Cuantiles, cuartiles, deciles y percentiles

Anguita (2000) señala que:

Sabemos que la mediana divide a los datos en dos partes iguales, también tiene interés estudiar otros parámetros, llamados **cuantiles**, que dividen los datos de la distribución en función de otras cantidades. Los más importantes son los cuartiles, quintiles, deciles y percentiles.

1.1. Cuartiles

Los cuartiles "Son tres valores que dividen la serie de datos en cuatro partes iguales. Se representan por Q_1 (cuartil primero), Q_2 (cuartil segundo) y

Q_3 (cuartil tercero)" (Anguita, 2000). Vitutor (2010) establece que " Q_1 , Q_2 y Q_3 determinan los valores correspondientes al 25%, al 50% y al 75% de los datos" y que " Q_2 coincide con la mediana". Para calcular los cuartiles, presenta el siguiente procedimiento.

- Se ordenan los datos de mayor a menor
- Se busca el lugar que ocupa cada cuartil, mediante la fórmula:

$$\frac{k \times N}{4}, \text{ con } k = 1, 2, 3$$

Donde:

- N = número de sujetos o informantes.

El número de sujetos puede ser par o impar, lo cual establece una diferencia en el cálculo. Si el número de sujetos o informantes es impar, el lugar y el valor de los tres cuartiles son directos. Si se trata de un número par, los valores de los cuartiles se obtienen mediante un promedio de los dos valores que corresponden a cada lugar.

1.2. Deciles

Los deciles "Son nueve valores que dividen la serie en 10 partes iguales: D_1, D_2, \dots, D_9 " (Anguita, 2000). Vitutor (2010) agrega que "los deciles dan los valores correspondientes al 10%, al 20%,... y al 90% de los datos" y que "el quinto decil coincide con la mediana". Para este caso, la fórmula para conocer el lugar que ocupa el decil es:

$$\frac{k \times N}{10}, \text{ con } k = 1, 2, \dots, 9$$

1.3. Percentiles

Los percentiles "Son 99 valores que dividen la serie de datos en 100 partes iguales: P_1, P_2, \dots, P_{99} " (Anguita, 2000). Vitutor (2010) complementa que "los percentiles dan los valores correspondientes al 1%, al 2%,... y al 90%

de los datos" y que "el percentil 50 coincide con la mediana. Para calcular el lugar que ocupa cada percentil se utiliza la fórmula:

$$\frac{k \times N}{100}, \text{ con } k = 1, 2, \dots, 99$$

2. Media

La media se calcula dividiendo la suma total de las repuestas de cada uno de los sujetos o informantes entre el número total de ellos.

$$\bar{X} = \frac{R_1 + R_2 + \dots + R_n}{N}$$

Donde:

- R_i indica la respuesta i o el valor de la variable i y N , el número total de sujetos o informantes.

3. Mediana

La mediana es un cuantil que divide la serie de datos en dos partes iguales y cuyo valor corresponde al 50% de los datos.

4. Moda

En una serie de datos, la moda es el valor más frecuente.

5. Rango

Para calcular el rango de los valores de las repuestas, éstas se clasifican en orden ascendente, para conocer sus valores máximo y mínimo. El rango significa los valores que puede asumir una variable, partiendo del mínimo y sólo hasta el máximo. Por ejemplo:

En el caso de una escala de 30 reactivos, con repuestas posibles: 1, 2, 3, 4, 5, el valor mínimo a obtener es de 30 reactivos por 1, la respuesta más baja; es decir, un valor de 30. El valor máximo es de 30 reactivos por 5, la respuesta más alta; un valor de 150. Por lo tanto el rango del

puntaje total de la escala para cada sujeto varía entre 30 y 150. El rango es R (30-150).

6. Desviación estándar

Para calcular la desviación estándar, primero se calcula la varianza. La desviación estándar es la raíz cuadrada de la varianza.

7. Varianza

La Universidad de Málaga (s/f) define la varianza como la media de las diferencias cuadráticas de n puntuaciones con respecto a su media aritmética, es decir

$$S^2 = \frac{1}{n} \sum_{i=1}^n (\text{respuesta } R_i - \text{Media})^2$$

Es decir, la varianza es un promedio de la suma del cuadrado de las desviaciones de cada respuesta con respecto a la media. También proporciona la siguiente fórmula equivalente:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (R_i^2) - \text{Media}^2$$

Por ejemplo: las observaciones o respuestas de un levantamiento de datos son: 4, 4, 5, 6, 6.

- Primeramente se calcula el valor de la media: la suma de las observaciones, $4 + 4 + 5 + 6 + 6 = 25$, dividida entre 5, el número de observaciones; esto es $25 \text{ entre } 5 = 5$. La media tiene un valor de 5.
- Con la primera fórmula, la varianza se calcula como sigue.

$$S^2 = \frac{1}{5} ((4-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (6-5)^2)$$

$$S^2 = \frac{1}{5} ((-1)^2 + (-1)^2 + (0)^2 + (1)^2 + (1)^2)$$

$$S^2 = \frac{1}{5} (1 + 1 + 0 + 1 + 1)$$

$$S^2 = \frac{1}{5} (4) = \frac{4}{5} = .8$$

- Con la segunda fórmula, la varianza se calcula como sigue.

$$S^2 = \frac{1}{5} (4^2 + 4^2 + 5^2 + 6^2 + 6^2) - 5^2$$

$$S^2 = \frac{1}{5} (16 + 16 + 25 + 36 + 36) - 25$$

$$S^2 = \frac{1}{5} (129) - 25 = \frac{129}{5} - 25 = 25.8 - 25 = .8$$

8. Curtosis

Suárez (2011a) maneja dos posibilidades para el cálculo de la curtosis, a la cual denomina como medida de Fisher, para datos sin agrupar y para datos agrupados, con las fórmulas siguientes:

- Datos sin agrupar

$$\alpha = \frac{\sum (x_i - \bar{x})^4}{n\sigma^4}$$

- Datos agrupados

$$\alpha = \frac{\sum f(x_i - \bar{x})^4}{n\sigma^4}$$

Donde:

x_i = cada uno de los valores de las repuestas

n = número de datos, observaciones o puntajes

\bar{x} = media aritmética;

σ^4 = cuádruplo de la desviación estándar poblacional;

f = frecuencia absoluta;

Si el valor de la curtosis es menor de 3, la distribución de datos es platocúrtica. Si es igual a 3, es normal o mesocúrtica y si es mayor a 3, es leptocúrtica (Suárez, 2011).

9. Asimetría

Suárez (2011), para medir la asimetría, utiliza el coeficiente de Pearson, con la siguiente fórmula:

$$As = \frac{3(\bar{X} - Md)}{s}$$

Donde:

\bar{X} = media aritmética

Md = Mediana

s = desviación típica o estándar

Suárez (2011) señala que el coeficiente de Pearson varía entre -3 y 3. Por lo tanto,

- Si la asimetría es < 0 , la distribución es asimétrica negativa
- Si es igual a cero, es simétrica y
- Si es > 0 , la distribución es asimétrica positiva

10. Prueba de normalidad. Kolmogorov-Smirnov

La información de esta prueba se tomó de Alford (2011). Según este autor, la prueba de Kolmogorov para el análisis de una muestra es considerada como un procedimiento de bondad de ajuste porque permite la medición del grado de concordancia entre la distribución de un conjunto de datos y la distribución teórica específica, en este caso, la normal. Alford (2011) propone la siguiente fórmula.

$$D = \text{m\u00e1ximo de } (F_t - F_o)$$

Donde:

- D = Valor calculado para su b\u00fasqueda en tabla
- F_t = Frecuencia te\u00f3rica
- F_o = Frecuencia observada

Los pasos se\u00f1alados por Alford (2011) para el c\u00e1lculo son los siguientes.

- La hip\u00f3tesis nula afirma que la distribuci\u00f3n de los datos se ajusta a la de una distribuci\u00f3n normal
- Se establecen intervalos de datos
- Se calcula la frecuencia observada de cada uno de los intervalos
- La suma de frecuencias debe ser igual a 100
- Se calcula la frecuencia relativa (frecuencia observada de cada intervalo dividida entre la suma total de la frecuencia observada)
- Se calcula la frecuencia observada relativa acumulada (FORA) y la frecuencia esperada relativa acumulada (FERA)
- Se calculan los valores de D para cada intervalo, con

$$D = (\text{FOR acumulado menos FER acumulado})$$

- Se toma el m\u00e1ximo valor obtenido para D
- Se busca el valor en la tabla correspondiente (ver anexo D), seg\u00fan el n\u00famero de sujetos o informantes (n) y el nivel de significancia deseado
- Si el valor de n es menor o igual a 40, se usa la tabla directamente
- Si el valor es mayor de 40, se usan las f\u00f3rmulas que se ubican en el \u00faltimo rengl\u00f3n de la tabla
- Si la D calculada es menor al valor de la tabla, se acepta la hip\u00f3tesis nula

11. Prueba de Levene

NIST/SEMATECH (2012), en su manual de estad\u00edstica para ingenieros, proporciona el siguiente procedimiento para el c\u00e1lculo de la Prueba de Levene de homogeneidad de varianzas.

La prueba es para examinar si un n\u00famero k de muestras tiene homogeneidad de varianzas. La hip\u00f3tesis nula supone que las varianzas de todos los grupos manejados son iguales. Su planteamiento es el siguiente:

- “Dada una variable Y , con tamaño de muestra N , dividida en k subgrupos, donde N_i es el tamaño de muestra del subgrupo i , la prueba estadística de Levene se define como:” (Nist/Sematech, 2012):

$$W = \frac{N - k}{k - 1} \frac{\sum_{i=1}^k N_i (\bar{Z}_i - \bar{Z}_{...})^2}{\sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - \bar{Z}_i)^2}$$

Donde:

- W = Valor calculado para la prueba
- Y = Variable observada
- Y_{ij} = Valor de la observación j del grupo i
- \bar{Y}_i = Media de las observaciones del grupo i
- N = Tamaño total de las observaciones
- k = Número de grupos
- i = Número de sujeto o informante del grupo i
- N_i = Número de sujetos o informantes del grupo i , tamaño del grupo i
- j = Sujeto o informante específico del grupo i
- Z_i es la media para el grupo i

$$\bar{Z}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Z_{ij}$$

- $Z_{..}$ es la media de todas las observaciones

$$\bar{Z}_{...} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{N_i} Z_{ij}$$

- Z_{ij} , en función de la media $| Y_{ij} - \bar{Y}_i |$

El valor W calculado para la prueba de Levene se compara con el valor de la tabla F (ver anexo F), con $(k-1)$ y $(N-k)$ grados de libertad. Si el valor calculado es menor al valor encontrado en la tabla y el valor del nivel de significancia es mayor de $.05$, se acepta que hay homogeneidad de varianzas.

12. Alfa de Cronbach

Gómez (2015) considera que Alfa de Cronbach es “un índice de consistencia interna que toma valores entre 0 y 1 y que sirve para comprobar si

el instrumento que se está evaluando recopila información defectuosa y por tanto nos llevaría a conclusiones equivocadas o si se tiene un instrumento fiable que hace mediciones estables y consistentes” y que “mide la homogeneidad de las preguntas promediando todas las correlaciones entre todos los ítem, para ver que, efectivamente, se parecen”. Proporciona la siguiente fórmula.

$$\alpha = \frac{K}{K-1} \left[1 - \frac{\sum S_i^2}{\sum T^2} \right]$$

Donde:

- K = Número de reactivos
- S_i^2 = Suma de varianzas de los reactivos
- S_T^2 = Varianza de la suma de los reactivos
- α = Coeficiente Alfa de Cronbach

13. Correlación

El coeficiente de correlación entre dos variables se calcula con dos procedimientos. Para datos que se ajustan a una estadística paramétrica, se utiliza el coeficiente de Pearson y para la no paramétrica, el coeficiente de Spearman.

13.1. Coeficiente de correlación de Pearson

La fórmula para calcular el coeficiente de correlación lineal de Pearson (Suárez, 2011a) es la siguiente:

$$r_{xy} = \frac{\sum_{i=1}^N (x_i)(y_i)}{\sqrt{(\sum_{i=1}^N x_i^2)(\sum_{i=1}^N y_i^2)}}$$

Donde:

- r_{xy} = Coeficiente de correlación entre las variables X y Y
- N = número total de sujetos o informantes en cada grupo, el mismo número
- x_i = $(X_i - \bar{X})$ para cada una de las observaciones de la variable X

- $y_i = (Y_i - \bar{Y})$ para cada una de las observaciones de la variable Y
- $(X_i)(Y_i) = (X_i - \bar{X})(Y_i - \bar{Y})$
- $x_i^2 = (X_i - \bar{X})(X_i - \bar{X})$
- $y_i^2 = (Y_i - \bar{Y})(Y_i - \bar{Y})$

13.2. Coeficiente de correlación de Spearman

Suárez (2011b) señala que "Este coeficiente se emplea cuando una o ambas escalas de medidas de las variables son ordinales, es decir, cuando una o ambas escalas de medida son posiciones. Ejemplo: Orden de llegada en una carrera y peso de los atletas". Proporciona la siguiente fórmula.

$$r_s = 1 - \frac{6 \sum_{i=1}^N d^2}{n(n^2 - 1)}$$

Donde:

- r_s = coeficiente de correlación por rangos de Spearman
- d = diferencia entre los rangos X - Y
- n = número de observaciones o datos

Suárez (2011b) recomienda el siguiente procedimiento:

- Ordenar los datos en rangos
- Asignar un uno al puntaje (s) más alto (s)
- Asignar dos al siguiente y así sucesivamente
- Si se repiten varios puntajes, se calculan las medias aritméticas
- Construir una tabla con los rangos asignados
- Calcular la columna para $d = X - Y$
- Calcular la columna para d^2
- Calcular la suma total de la columna
- Utilizar la fórmula propuesta

14. Determinación

Cuando hay interés por conocer la participación de una variable independiente en el comportamiento de una variable dependiente, el coe-

ficiente de determinación se calcula elevando al cuadrado el valor del coeficiente de correlación.

15. Regresión lineal

Para mostrar la ecuación de la regresión lineal simple, se tomó todo el procedimiento y las fórmulas descritas por NAV (2006), de la Universidad de la Universidad centroamericana José Simeón Cañas, UCA.

“El modelo de regresión es la ecuación que define cómo se relacionan dos variables Y y X” (NAV, 2006).

El modelo de regresión simple es:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Donde:

- Y es la variable dependiente
- X es la variable independiente
- β_0 y β_1 son parámetros del modelo
- ε se denomina error; explica la variabilidad en la Y que no puede explicarse con la relación lineal entre X e Y
- El valor esperado de Y, de notado como E (Y/X) está dado por $\beta_0 + \beta_1 X$

La ecuación de regresión lineal simple implica la estimación de los parámetros β_0 y β_1 , calculados como b_0 y b_1 , a través del método de mínimos cuadrados. La gráfica de la ecuación se denomina línea de regresión lineal estimada.

Para obtener los valores de b_0 y b_1 se utilizan las siguientes ecuaciones simultáneas:

$$\sum_{i=1}^n Y_i = n(b_0) + (\sum_{i=1}^n X_i) b_1$$

$$\sum_{i=1}^n X_i Y_i = (\sum_{i=1}^n X_i)(b_0) + (\sum_{i=1}^n X_i^2) b_1$$

Finalmente, después de varias transformaciones, los valores de b_0 y b_1 , se calculan con las siguientes fórmulas.

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$

16. Regresión múltiple

Para mostrar la ecuación de la regresión múltiple, también se tomó el procedimiento y las fórmulas descritas por NAV (2006), de la Universidad de la Universidad centroamericana José Simeón Cañas, UCA. Se presenta el procedimiento para una variable dependiente Y y varias variables independientes X , desde $i = 1$ hasta p variables independientes.

El análisis de regresión múltiple es el estudio de la forma en que una variable dependiente, y , se relaciona con dos o más variables independientes. En el caso general emplearemos p para representar la cantidad de variables independientes (NAV, 2006).

Las fórmulas que sugiere son las siguientes:

Variables independientes = $X = (X_1, X_2, \dots, X_p)$

Modelo de regresión múltiple $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$

El término del error explica la variabilidad en “ y ” que no puede explicar las p variables independientes. El error es una variable aleatoria distribuida normalmente con media cero y varianza constante, σ^2 , para todos los valores de las X_i . Si consideramos el valor medio de la variable “ y ” dadas las variables independientes $X = (x_1, x_2, \dots, x_p)$, obtenemos la ecuación de regresión lineal (NAV, 2006).

$$\mu_{Y/X} = E(Y/X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

“Utilizando los datos de una muestra de tamaño n y el método de mínimos cuadrados se determina la ecuación de regresión múltiple estimada” (NAV, 2006).

$$\check{Y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

Por último, NAV (2006) señala que “cada coeficiente b_i representa una estimación del cambio en “ y ” que corresponde a un cambio unitario en x_i cuando todas las demás variables independientes se mantienen constantes”.

Adicionalmente, el criterio de mínimos cuadrados es:

$$\min \sum_{i=1}^p (Y_i - \bar{Y}_i)^2 \quad (\text{NAV, 2006})$$

Las pruebas de significancia se llevan a cabo con la prueba F, con la hipótesis nula de que todos los coeficientes de la expresión tienen un valor de cero y la hipótesis de que uno o más de los parámetros es cero (NAV, 2006).

NAV (2006), plantea algunas fórmulas adicionales que se requieren para este proceso.

- Suma de cuadrados debida al error: $SCE = \sum (y_i - \hat{y}_i)^2$
- Suma de cuadrados total: $SCT = \sum (y_i - \bar{y})^2$
- Suma de cuadrados debida a la regresión: $SCR = \sum (\hat{y}_i - \bar{y})^2$
- Relación entre SCT, SCR y SCE: $SCT = SCR + SCE$

La prueba F se calcula con un nivel de significancia de .05, con la fórmula:

$$F_c = \frac{CMR}{CME} = \frac{SCE/p}{SCE/(n-p-1)}$$

“Se rechaza la hipótesis nula si el p-valor de F_c es menor que .05” (NAV, 2006).

17. T de Student

La prueba t de Student se presenta para los casos de dos muestras independientes y de dos muestras relacionadas.

17.1. Prueba t para muestras independientes

Con base en la página de Raymundo (2014c), la fórmula para la prueba t de Student para muestras independientes es la siguiente:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_p \sqrt{\frac{1 + 1}{(N_1)(N_2)}}}$$

Donde:

- t = Valor estadístico de la prueba t de Student
- X_1 = Media del grupo uno
- X_2 = Media del grupo dos
- σ_p = Desviación estándar ponderada de ambos grupos
- N_1 = Tamaño del grupo uno
- N_2 = Tamaño del grupo dos

La fórmula para la desviación estándar ponderada es la siguiente

$$\sigma_p \sqrt{\frac{SC_1 + SC_2}{(N_1 + N_2 - 2)}}$$

Donde:

- σ_p = desviación estándar ponderada
- SC_1 = suma de cuadrados del grupo 1

$$\sum_{i=1}^{N_1} (X_i - \bar{X})^2$$

- SC_2 = suma de cuadrados del grupo 2

$$\sum_{i=1}^{N_2} (X_i - \bar{X})^2$$

- N_1 = tamaño de la muestra 1
- N_2 = tamaño de la muestra 2

Los pasos recomendados por Raymundo (2014c) son:

- Determinar el promedio o media aritmética de cada grupo de población.
- Calcular la varianza e cada grupo, para demostrar la homogeneidad de varianzas (puede usarse la fórmula del inciso 10. Prueba de normalidad. Kolmogorov-Smirnov)

- Calcular la suma de cuadrados de cada grupo, SC
- Calcular la desviación estándar ponderada σ_p , de cada grupo
- Obtener la diferencia absoluta entre los grupos, $(\bar{X}_1 - \bar{X}_2)$
- Aplicar la fórmula y obtener el valor estadístico de t
- Calcular los grados de libertad $gl = N_1 + N_2 - 2$
- Obtener el valor de t en la tabla
- Decidir si se acepta o se rechaza la hipótesis

17.2. Prueba t para muestras relacionadas

Con base en la página de Raymundo (2014b), es un requisito que haya homogeneidad de varianzas y la fórmula para la prueba t de Student para muestras independientes es la siguiente.

$$t = \frac{\bar{d}}{\frac{\sigma_d}{\sqrt{N}}}$$

Donde:

- t = valor estadístico del procedimiento
- \bar{d} = valor promedio o media aritmética de las diferencias entre los momentos antes y después

$$\bar{d} = \frac{\sum_{i=1}^N d}{N}$$

- σ_d = desviación estándar de las diferencias entre los momentos antes y después
- N = tamaño de la muestra

$$\sigma_d = \sqrt{\frac{\sum_{i=1}^N (d - \bar{d})^2}{N - 1}}$$

Raymundo (2014b) recomienda:

- Ordenar los datos en función de los momentos antes y después
- Obtener las diferencias entre ambos

- Calcular la media aritmética de las diferencias
- Calcular la desviación estándar de las diferencias
- Calcular el valor de t por medio de la ecuación
- Calcular los grados de libertad $gl = N - 1$
- Calcular el valor de t calculada con los grados de libertad
- Decidir si se acepta o rechaza la hipótesis

17.3. Prueba t de Student-Welch para dos muestras independientes con varianzas no homogéneas

Raymundo (2014c) plantea que, para la prueba t de Student-Welch, existe una modificación de la ecuación original que corresponde a la t de Student, que es la siguiente:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Donde:

- t = estadístico equivalente a t de Student
- \bar{X}_1 = media aritmética del grupo 1
- \bar{X}_2 = media aritmética del grupo 2
- σ_1^2 = varianza del grupo 1
- σ_2^2 = varianza del grupo 2
- n_1 = tamaño de la muestra del grupo 1
- n_2 = tamaño de la muestra del grupo 2

El cálculo de los grados de libertad se realiza con la fórmula siguiente:

$$gl = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}{\left(\frac{\sigma_1^2}{n_1 - 1}\right)^2 + \left(\frac{\sigma_2^2}{n_2 - 1}\right)^2} - 2$$

Donde:

- σ^2_1 = varianza del grupo 1
- σ^2_2 = varianza del grupo 2
- n_1 = tamaño de la muestra del grupo 1
- n_2 = tamaño de la muestra del grupo 2



El salón de clase de lengua
Procesamiento de datos en Lingüística Aplicada

de Clara Cecilia Uribe Hernández, fue impreso en enero de 2017 en los talleres de Editorial CIGOME, S.A. de C.V., Vialidad Alfredo del Mazo núm. 1524, ex. Hacienda La Magdalena C.P. 50010, Toluca, México. Su edición consta de 300 ejemplares. La edición estuvo a cargo de la Dirección de Difusión y Promoción de la Investigación y los Estudios Avanzados.

Coordinación editorial: Patricia Vega Villavicencio
Diseño de forros e interiores: Juan Manuel García Guerrero
Corrección de estilo: María de los Ángeles García Moreno





El salón de clase de lengua, además de ser espacio para la enseñanza, como agua viva, constituye una fuente de vertientes para la investigación educativa; es crisol donde convergen diversas lenguas, en tanto medio de expresión y materia de aprendizaje; genera la inquietud de analizar su naturaleza y características con relación a la interacción entre personajes y procesos.

Y es apenas el origen; introducirse a la investigación plantea situaciones, cuestionamientos, posibilidades, particularmente sobre el manejo de datos cualitativos, cuantitativos o combinados; como respuesta, este libro conjunta aspectos teóricos someros con elementos precisos, detallados, del procesamiento de datos en lingüística aplicada, con el espíritu de que profesores y estudiantes cuenten con lineamientos fundamentales de soporte para el levantamiento de datos, el uso de herramientas estadísticas y el procesamiento de datos que sustenten el análisis certero para la derivación de resultados.



slEA
Sociedad Latinoamericana de Estadística y Lingüística Aplicada

