

UNIVERSIDAD AUTONOMA DEL ESTADO DE MEXICO

ESCUELA PREPARATORIA TEXCOCO



MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN PARA DATOS NO AGRUPADOS

MATERIAL DIDACTICO SOLO VISION

ASIGNATURA QUE CORRESPONDE: ESTADISTICA

RESPONSABLE: LIC. EN R.C. ANA LILIA CARRILLO FLORES

SEPTIEMBRE 2015

PRESENTACION

La Estadística está estrechamente relacionada con el método científico, en la recopilación, organización, representación y análisis de datos; tanto para obtener conclusiones como para tomar decisiones con base en los análisis efectuados. Así, en el proceso educativo, la enseñanza de la Estadística contribuye a la formación del estudiante, mediante la construcción y el desarrollo de conocimientos y destrezas intelectuales para realizar un manejo eficiente de datos.

La importancia del estudio de la Estadística radica en su aplicación en todas las áreas del quehacer humano; así, encontramos estadísticas sobre población, deportes, salud, estudios económicos, pruebas de laboratorio, etc. por lo que se pretende que el estudiante de bachillerato logre obtener las bases necesarias propias de la Estadística Descriptiva para que, en un futuro, pueda realizar inferencias estadísticas en la licenciatura de su elección.

PRESENTACION (continuación)

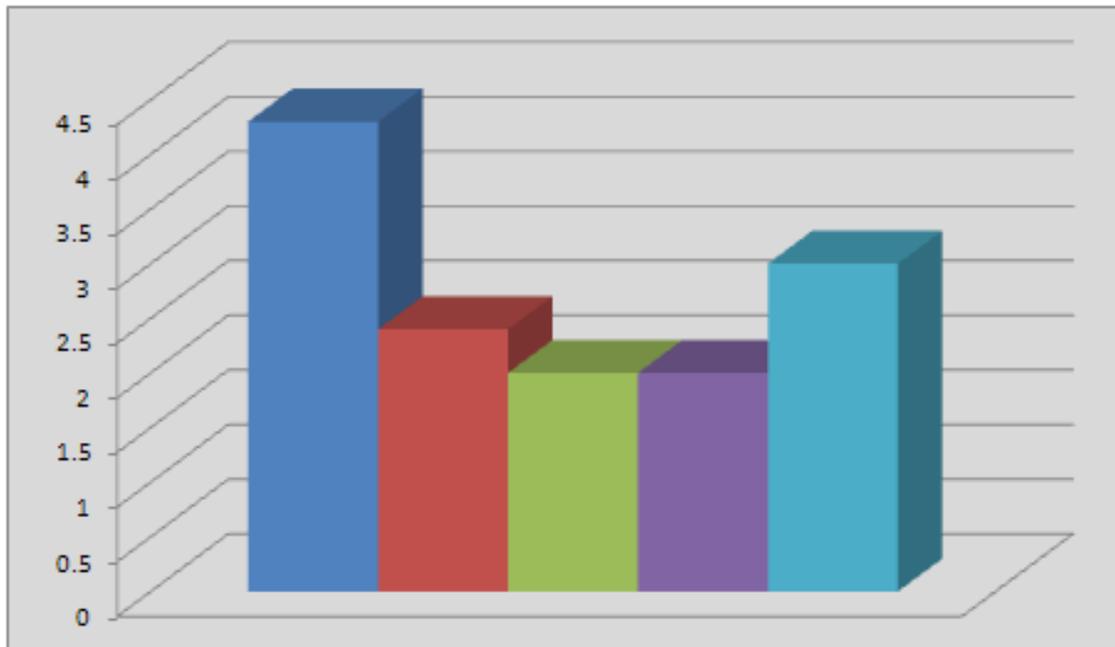
El propósito principal de esta asignatura, es inducir a los estudiantes a abordar diversos problemas reales de actualidad, a través de los principios generales de la Estadística, y motivarlos a comprender su importancia y la manera de utilizarlos en su vida personal y profesional. Los conceptos y herramientas estadísticas tendrán significado para los estudiantes al investigar por sí mismos problemas de la vida cotidiana, al recopilar información de fuentes reales (dependencias públicas y privadas, encuestas ciudadanas, internet, etc.), al formular sus hipótesis, interpretar resultados y tomar decisiones.

La materia está dividida en cuatro módulos:

- I. Conceptualización de la estadística y distribución de frecuencias.
- II. Medidas de tendencia central y medidas de posición.
- III. Medidas de dispersión
- IV. Regresión lineal

En esta presentación se abordan los temas de Medidas de Tendencia Central (incluida en el módulo I) y las Medidas de Dispersión (incluidas en el módulo II) para datos no agrupados

Medidas de Tendencia Central para datos no agrupados

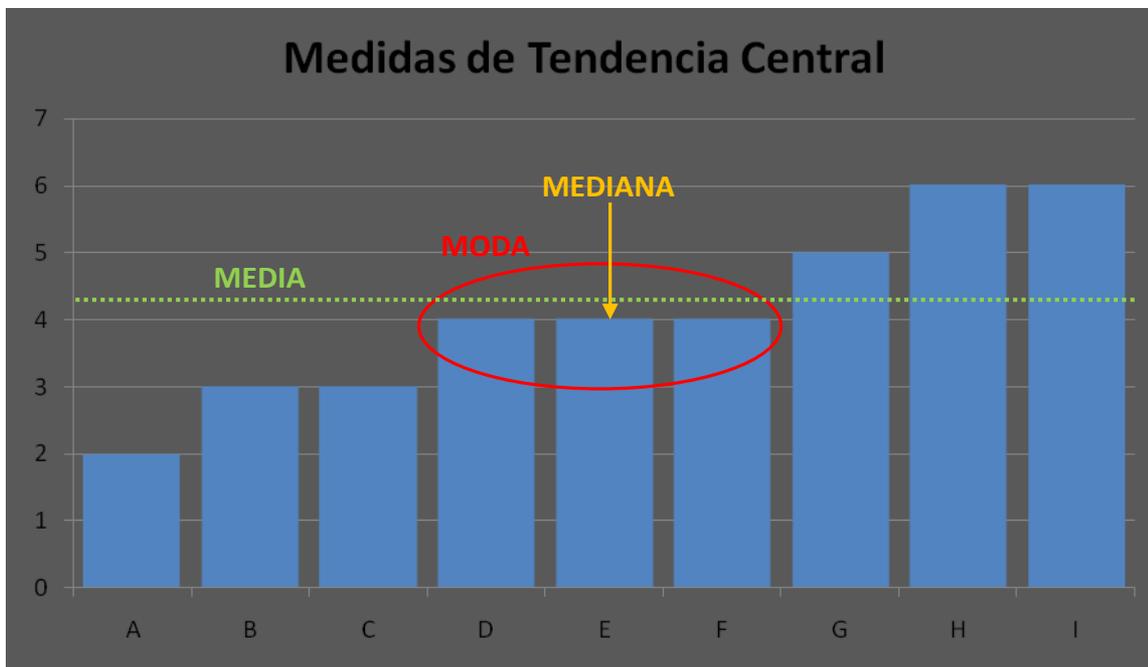


Objetivos

- Al finalizar la clase el alumno comprenderá los conceptos de media, mediana y moda para datos no agrupados, mismos que utilizará para resolver problemas prácticos.

¿Qué son las medidas de tendencia central?

- Son indicadores estadísticos que muestran hacia que valor (o valores) se agrupan los datos. Deben su nombre a que normalmente se encuentran en la posición central de una distribución.



Media Aritmética

- La media aritmética es el valor resultante que se obtiene al dividir la sumatoria de un conjunto de datos sobre el número total de datos. Solo es aplicable para el tratamiento de datos cuantitativos.
- Se le llama también promedio o, simplemente, media.
- Se obtiene mediante la fórmula:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Ejemplo de cálculo de media aritmética

- Calcular la media para 5 datos obtenidos de personas a las que se le preguntó la talla de calzado. Los datos son: 4, 6, 3, 7 y 5.

Aplicando la fórmula, tenemos:

$$\bar{x} = \frac{4 + 6 + 3 + 7 + 5}{5} = \frac{25}{5} = 5$$

La media que representa a las cinco personas es 5.

Ventajas Vs Desventajas en el uso de la media

Ventajas	Desventajas
Cálculo sencillo, intervienen todos los datos	Una misma media puede representar a dos poblaciones muy distintas
Su valor es único	Se ve muy afectada por valores extremos
Se usa para comparar poblaciones	
Se emplea en cálculos estadísticos posteriores (Desviación media, varianza, covarianza, etc)	

Mediana

- La mediana es un valor de la variable que deja por debajo de sí a la mitad de los datos, una vez que estos están ordenados de menor a mayor.
- La mediana del número de hijos de un conjunto de trece familias, cuyos respectivos hijos son: 3, 4, 2, 3, 2, 1, 1, 2, 1, 1, 2, 1 y 1, una vez ordenados los datos:

$\underbrace{1, 1, 1, 1, 1, 1}_{\text{Mitad inferior}}, \quad \underbrace{2}_{\text{Mediana}}, \quad \underbrace{2, 2, 2, 3, 3, 4}_{\text{Mitad superior}}$

Mediana (continuación)

- En caso de un número par de datos, la mediana no correspondería a ningún valor de la variable, por lo que se conviene en tomar como mediana el valor intermedio entre los dos valores centrales.
- Por ejemplo, en el caso de doce datos como los anteriores:

$\underbrace{1, 1, 1, 1, 1}_{\text{Valores inferiores}} \quad \underbrace{1, 2}_{\text{Valores intermedios}} \quad \underbrace{2, 2, 3, 3, 4}_{\text{Valores superiores}}$

$$1,5 = \frac{1 + 2}{2}$$

Ventajas Vs Desventajas en el uso de la mediana

Ventajas	Desventajas
No se ve alterada por valores extremos	Su resultado puede ser de otro tipo de variable
No se ve afectada por la dispersión	

Moda

- La moda es el dato más repetido, el valor de la variable con mayor frecuencia absoluta. En cierto sentido la definición matemática corresponde con la locución "estar de moda", esto es, ser lo que más se lleva.
- Su cálculo es extremadamente sencillo, pues sólo necesita un recuento.

Ejemplo de obtención de moda

- El número de personas en distintos vehículos en una carretera: 5, 7, 4, 6, 9, 5, 6, 1, 5, 3, 7.
Encontrar la moda.
- El número que más se repite es 5, entonces la moda es 5.

Ventajas Vs Desventajas en el uso de la mediana

Ventajas	Desventajas
Cálculo sencillo	Puede no haber moda en una distribución
Interpretación muy clara	No siempre se sitúa hacia el centro de la distribución
Puede calcularse para variables cualitativas	Puede haber más de una moda
Es estable a valores extremos	

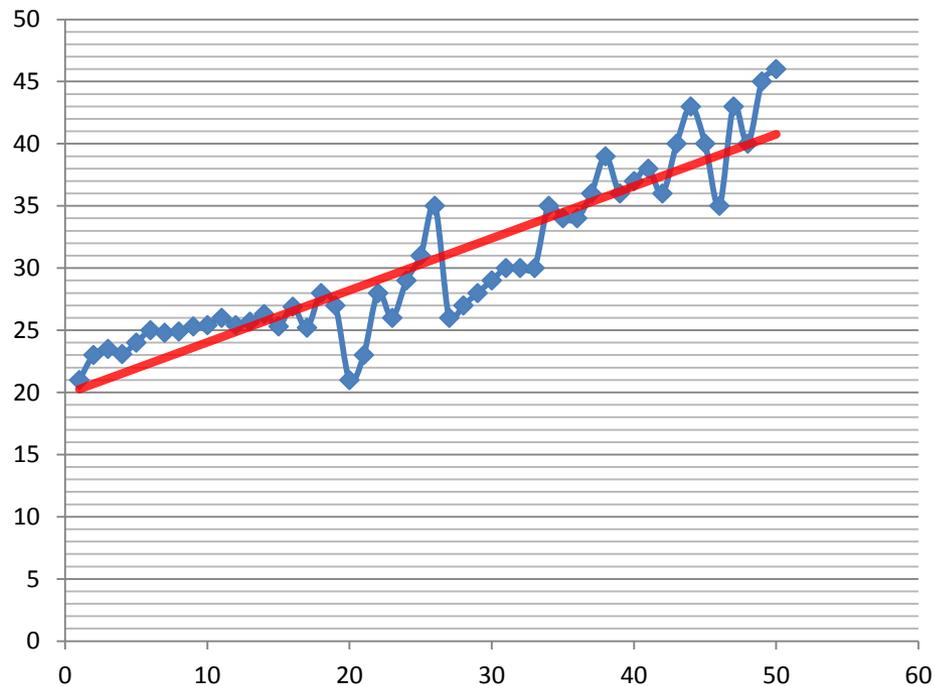
Verificación de la comprensión

- Actividad explicada por el profesor

Conclusiones

- Las medidas de tendencia central son valores a los que se puede recurrir si se quiere tener una referencia de una serie de datos.
- No se debe cometer el error de tomar una decisión únicamente con una medida de tendencia central, es ampliamente recomendable acompañar ésta con una medida de dispersión.

Medidas de Dispersión para datos no agrupados



Objetivos

- Al finalizar la clase el alumno comprenderá los conceptos de Rango, Desviación Media, Varianza, Desviación Estándar y Coeficiente de Variación para datos no agrupados, y en conjunto con las medidas de tendencia central, los aplicará para resolver problemas prácticos.

¿Qué son las medidas de dispersión?

No solo basta con determinar las medidas de tendencia central para comprender el comportamiento de una serie de datos, es importante además, conocer que tan alejados están esos datos respecto a ese punto de concentración.

Las medidas de dispersión nos indican la distancia promedio de los datos respecto a las medidas de tendencia central. Así podremos diferenciar dos conjuntos de datos que poseen iguales medias, siendo los datos de uno más dispersos del otro.

Rango

- Es la diferencia entre los valores extremos: el valor máximo observado menos el valor mínimo observado.

$$\text{Rango} = \text{Valor máximo} - \text{Valor mínimo}$$

Ejemplo de cálculo de Rango

- A una reunión acuden personas con las siguientes edades: 25, 18, 31, 35, 24, 51, 38, 44, 26, 45, 53, 28, 37 y 40
 - Valor máximo: 53
 - Valor mínimo: 18
- $\text{Rango} = 53 - 18 = 35$
- Lo cual quiere decir que entre la persona de mayor edad y la de menor edad existe un rango de 35 años.

Ventajas Vs Desventajas en el uso del rango

Ventajas	Desventajas
Cálculo sencillo	No toma en cuenta todas las observaciones
Resulta fácil percibir el resultado	Depende sensiblemente del número de datos
	Es muy general, tan solo da una idea de cuán amplia es la variación entre puntajes extremos

Desviación media

- Es la media de los valores absolutos de las distancias entre la media y los diferentes datos.

- *Desviación Media (DM) = $\frac{\text{Valor absoluto de la suma de las desviaciones}}{\text{Número total de datos}}$*

$$DM = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

Ejemplo de cálculo de la Desviación Media

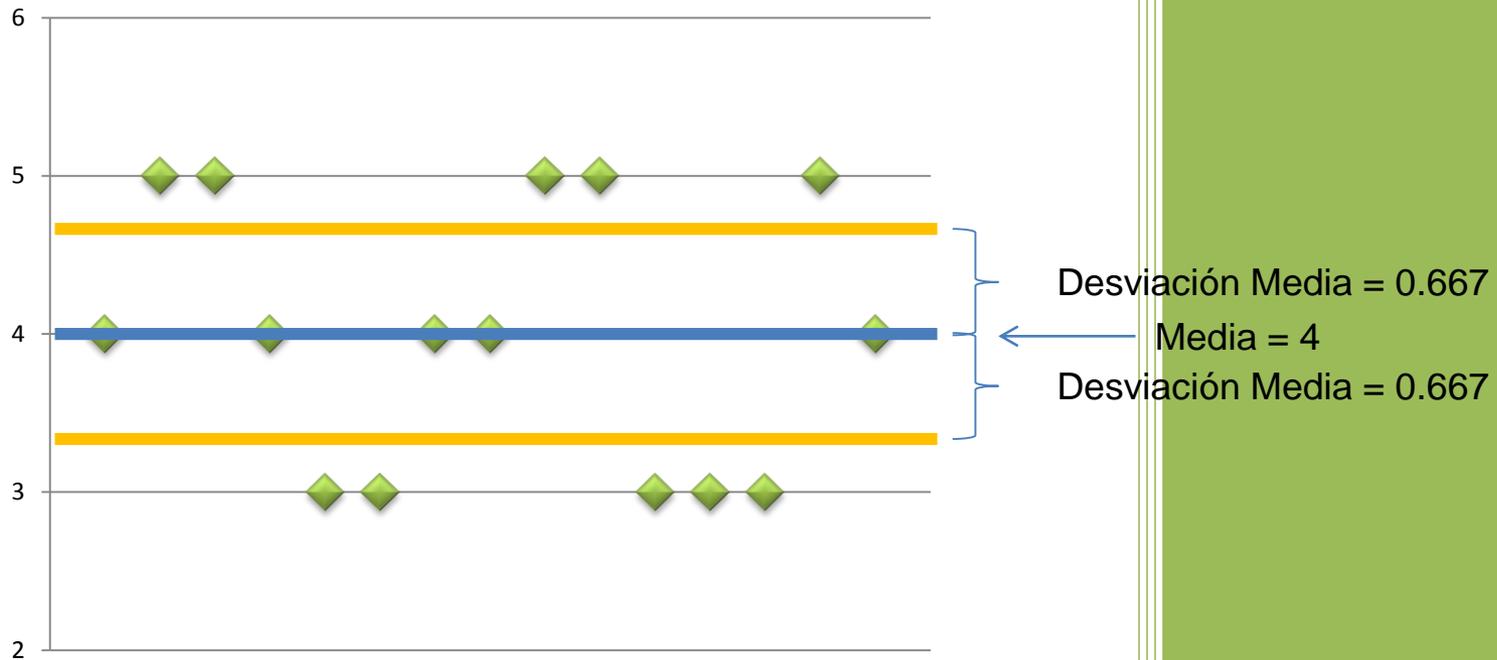
- Obtener la desviación media de los siguientes datos:

3	4	3
4	5	3
4	5	3
5	4	5
5	3	4

- La suma de todos los datos es: 60
- La media es: 4
- La desviación media se obtiene:

$$DM = \frac{|3-4|+|4-4|+|4-4|+|5-4|+|5-4|+|4-4|+|5-4|+|5-4|+|4-4|+|3-4|+|3-4|+|3-4|+|3-4|+|5-4|+|4-4|}{15} = 0.667$$

Interpretación de la DM calculada



Ventajas Vs Desventajas en el uso de la Desviación Media

Ventajas	Desventajas
Cálculo sencillo	Tiene la limitante de estar calculada con respecto a la media
	Es muy general, su uso es limitado

Varianza

- Se basa en las diferencias entre la media aritmética y cada una de las puntuaciones
- Es el promedio de los cuadrados de las distancias de las observaciones a partir de la media (su valor nunca será negativo)

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Ejemplo de cálculo de la Varianza

- Obtener la varianza de los siguientes datos:

3	4	3
4	5	3
4	5	3
5	4	5
5	3	4

- La suma de todos los datos es: 60
- La media es: 4
- La varianza se obtiene:

$$\sigma^2 = \frac{(3-4)^2 + (4-4)^2 + (4-4)^2 + (5-4)^2 + (5-4)^2 + (4-4)^2 + (5-4)^2 + (5-4)^2 + (4-4)^2 + (3-4)^2 + (3-4)^2 + (3-4)^2 + (3-4)^2 + (3-4)^2 + (5-4)^2 + (4-4)^2}{15} = 0.667$$

Ventajas Vs Desventajas en el uso de la varianza

Ventajas	Desventajas
Considera todos los valores	Al expresarse en el cuadrado de las unidades de los datos, resulta difícil de comprender

Desviación estándar o típica

- Es la medida de dispersión más usada en estadística, tanto para aspectos descriptivos como analíticos. Es la raíz cuadrada del cuadrado de la suma de las desviaciones entre el número total de observaciones, es decir, es la raíz cuadrada de la varianza.

- *Desviación Estándar* (σ) = $\sqrt{\frac{\text{Cuadrado de la suma de las desviaciones}}{\text{Número total de datos}}}$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Ejemplo de cálculo de la Desviación Estándar

- Calcular la desviación estándar para los siguientes datos no agrupados:

3	4	3
4	5	3
4	5	3
5	4	5
5	3	4

- La suma de todos los datos es: 60
- La media es: 4
- La desviación estándar se obtiene:

$$\sigma = \sqrt{\frac{((3-4)^2+(4-4)^2+(4-4)^2+(5-4)^2+(5-4)^2+(4-4)^2+(5-4)^2+(5-4)^2+(4-4)^2+(3-4)^2+(3-4)^2+(3-4)^2+(3-4)^2+(5-4)^2+(4-4)^2)}{15}} = \sqrt{0.667} = 0.816$$

Ventajas Vs Desventajas en el uso de la Desviación Estándar

Ventajas	Desventajas
Considera todos los valores	Es sensible a las unidades de medida
Se expresa en las unidades originales de la variable	

Coeficiente de variación

- El coeficiente de variación indica la importancia de la desviación estándar en relación al promedio aritmético.
- Se expresa en porcentaje

$$\text{Coeficiente de Variación (C.V.)} = \frac{\text{Desviación estándar } (\sigma)}{\text{Media } (\bar{x})} \times 100$$

Ejemplo de cálculo del coeficiente de variación

- Calcular la desviación estándar para los siguientes datos no agrupados:

3	4	3
4	5	3
4	5	3
5	4	5
5	3	4

- La Desviación estándar es: 0.816
- La media es: 4
- El coeficiente de variación se obtiene:

$$C.V. = \left(\frac{0.816}{4} \right) \times 100 = 20.41 \%$$

Ventajas Vs Desventajas en el uso del Coeficiente de Variación

Ventajas	Desventajas
Es útil para comparar la variabilidad entre conjuntos de datos (de diferentes o iguales unidades de medida)	Arrastra las limitaciones que tiene la media como medida de tendencia central
Se expresa en porcentaje, lo que permite la comparación rápida entre diferentes muestras	

CONCLUSIONES

- Las medidas de dispersión son un complemento de análisis para las medidas de tendencia central.
- Al hacer una toma de decisiones, es conveniente contar con ambas para poder analizar y decidir.

BIBLIOGRAFIA CONSULTADA

- LINCOYAN, Portus Govindea (1992) Curso Práctico De Estadística. Mc. Graw Hill
- PORTILLA, Chimal Enrique (1992) Estadística Primer Curso
- García Ferrando, M. (1996) Socioestadística. Introducción a la estadística en sociología