Autonomous University of the State of Mexico
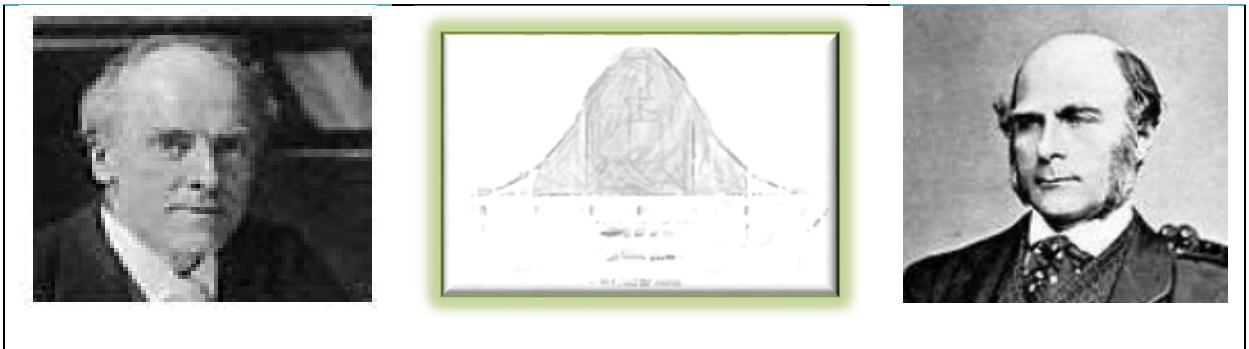
Amecameca University Center

Bachelor's Degree in Nutrition

Biostatistics I

Credits 8

EXERCICES

M en E. S. Narciso Campero Garnica

ÍNDICE

Pag.

PRESENTACIÓN

La Unidad de aprendizaje de Bioestadística I requiere de un compendio de ejercicios que permitan al alumno acceder con mayor facilidad los temas que integran el programa de dicha Unidad y por otro lado, a través del tiempo hemos detectado en la Licenciatura, ya a través del Departamento de Control Escolar, o bien mediante información de la Coordinación de la carrera en reuniones de Academia, o también con la participación del Programa de Tutoría Institucional que los alumnos presentan serias dificultades para la adquisición de las competencias en el idioma Inglés. En una reunión del Claustro de profesores de la licenciatura se sugirió que los docentes apoyáramos incluyendo lecturas, trabajos y/o ejercicios en Inglés.

Es así como surge la presente propuesta de un problemario que auxilie a los alumnos por un lado a practicar una de las habilidades del idioma que es la Lectura de comprensión y por otro lado poner a su disposición materiales que refuercen lo aprendido del contenido de Bioestadística I.

No se exige más a los jóvenes de primer semestre ya que no se pretende multiplicar el grado de dificultad que de por sí les representa el manejo de nuestra temática de Bioestadística, por lo cual se solicitará a los estudiantes que practiquen la lectura de comprensión de los ejercicios sin el auxilio de un traductor aprovechando los cursos que recibieron con anterioridad y que el Primer Curso de Inglés lo tomarán el siguiente semestre.

Las respuestas a los ejercicios, que no sean numéricas, los jóvenes podrán redactarlas en español para su facilidad.

Mis deseos son que este material sirva de apoyo realmente para nuestros alumnos y que tengamos la oportunidad de hacer más amenas nuestras sesiones.

Unidad I: Bioestadística descriptiva

Objetivo: Identificar y aplicar adecuadamente las técnicas descriptivas de la estadística en el ámbito de la salud, para su aplicación en el desarrollo de investigación científica.

Contenido:
       1.1 Recolección y presentación de datos
       1.2 Medidas de tendencia central y de dispersión.
       1.3 Medidas de posición, de simetría y curtosis.

1. The portion of the population that is selected for analysis is called:
    (a) a sample
    (b) a frame
    (c) a parameter
    (d) a statistic

2. A summary measure that is computed from only a sample of the population is called:
    (a) a parameter
    (b) a population
    (c) a discrete variable
    (d) a statistic

3. The height of an individual is an example of a:
    (a) discrete variable
    (b) continuous variable
    (c) categorical variable
    (d) constant

4. The body style of an automobile (sedan, coupe, wagon, etc.) is an example of a:
    (a) discrete variable

(b) continuous variable

(c) categorical variable

(d) constant

5. Statistical inference occurs when you:

    (a) compute descriptive statistics from a sample

    (b) take a complete census of a population

    (c) present a graph of data

    (d) take the results of a sample and draw conclusions about a population

6. The human resources director of a large corporation wants to develop a dental benefits package and decides to select 100 employees from a list of all 5,000 workers in order to study their preferences for the various components of a potential package. All the employees in the corporation constitute the _____.

    (a) sample

    (b) population

    (c) statistic

    (d) parameter

7. The human resources director of a large corporation wants to develop a dental benefits package and decides to select 100 employees from a list of all 5,000 workers in order to study their preferences for the various components of a potential package. The 100 employees who will participate in this study constitute the _____.

    (a) sample

    (b) population

    (c) statistic

    (d) parameter

8. Those methods involving the collection, presentation, and characterization of a set of data in order to properly describe the various features of that set of data are called:

    (a) statistical inference

    (b) the scientific method

    (c) sampling

    (d) descriptive statistics

9. A summary measure that is computed to describe a characteristic of an entire population is called:
   (a) a parameter
   (b) a population
   (c) a discrete variable
   (d) a statistic

10. Which of the following is a discrete variable?
    (a) The favorite flavor of ice cream of students at your local elementary school
    (b) The time it takes for a certain student to walk to your local elementary school
    (c) The distance between the home of a certain student and the local elementary school
    (d) The number of teachers employed at your local elementary school

11. Which of the following is a continuous variable?
    (a) The eye color of children eating at a fast-food chain
    (b) The number of employees of a branch of a fast-food chain
    (c) The temperature at which a hamburger is cooked at a branch of a fast-food chain
    (d) The number of hamburgers sold in a day at a branch of a fast-food chain

12. The possible responses to the question "How long have you been living at your current residence?" are values from a continuous variable.
    (a) True
    (b) False

13. The possible responses to the question "How many times in the past three months have you visited a museum?" are values from a discrete variable.
    (a) True
    (b) False

14. An insurance company evaluates many variables about a person before deciding on an appropriate rate for automobile insurance.
    a) The number of accidents a person has had in the past three years is an example of a _____ variable.
    b) The distance a person drives in a day is an example of a _____ variable.
    c) A person's marital status is an example of a _____ variable.

15. What are some benefits of representing data sets using frequency distributions?

16. What are some benefits of representing data sets using graphs of frequency distributions?

17. What is the difference between class limits and class boundaries?

18. What is the difference between frequency and relative frequency?

True or False? In Exercises 19–28, determine whether the statement is true or false. If it is false, rewrite it as a true statement.

19. The midpoint of a class is the sum of its lower and upper limits.

20. The relative frequency of a class is the sample size divided by the frequency of the class.

21. An ogive is a graph that displays cumulative frequency.

22. Class limits are used to ensure that consecutive bars of a histogram do

23. The median is the measure of central tendency most likely to be affected by an extreme value (an outlier).

24. Every data set must have a mode.

25. Some quantitative data sets do not have a median.

26. The mean is the only measure of central tendency that can be used for data at the nominal level of measurement.

27. Give an example in which the mean of a data set is *not* representative of a typical number in the data set.

28. Give an example in which the median and the mode of a data set are the same.

In Exercises 29–33,

(a) find the mean, median, and mode of the data, if possible. If it is not possible, explain why the measure of central tendency cannot be found.

(b) determine which measure of central tendency best represents the data. Explain your reasoning.

29. **Education**. The education cost per student (in thousands of dollars) from a sample of 10 liberal arts colleges

22 26 19 20 20 18 21 17 19 14

30. **Cholesterol.** The cholesterol level of a sample of 10 female employees

154 216 171 188 229 203 184 173 181 147

31. **Air Quality**. The responses of a sample of 1040 people who were asked if the air quality in their community is better or worse than it was 10 years ago

Better: 346   Worse: 450   Same: 244

32. **Eating Disorders.** The number of weeks it took to reach a target weight for a sample of five patients with eating disorders treated by psychodynamic psychotherapy *(Source: The Journal of Consulting and Clinical Psychology)*

15.0   31.5   10.0   25.5   1.0

33. **Eating Disorders**. The number of weeks it took to reach a target weight for a sample of 14 patients with eating disorders treated by psychodynamic psychotherapy and cognitive behavior techniques *(Source: The Journal of Consulting and Clinical Psychology)*

2.5   20.0   11.0   10.5   17.5   16.5   13.0
15.5   26.5   2.5   27.0   28.5   1.5   5.0

In Exercises 34–36, construct a frequency distribution and a frequency histogram of the data using the indicated number of classes. Describe the shape of the histogram as symmetric, uniform, negatively skewed, positively skewed, or none of these.

34. **Hospitalization**
Number of classes: 6
Data set: The number of days 20 patients remained hospitalized

6   9   7   14   4   5   6   8   4   11
10   6   8   6   5   7   6   6   3   11

35. *Hospital Beds*

Number of classes: 5

Data set: The number of beds in a sample of 24 hospitals

149	167	162	127	130	180	160	167	221	145	137
194										
207	150	254	262	244	297	137	204	166	174	180
151										

36. *Height of Males*

Number of classes: 5

Data set: The heights (to the nearest inch) of 30 males

67	76	69	68	72	68	65	63	75	69
66	72	67	66	69	73	64	62	71	73
68	72	71	65	69	66	74	72	68	69

37. *Coffee Content* During a quality assurance check, the actual coffee content (in ounces) of six jars of instant coffee was recorded as 6.03, 5.59, 6.40, 6.00, 5.99, and 6.02.

(a) Find the mean and the median of the coffee content.
(b) The third value was incorrectly measured and is actually 6.04. Find the mean and median of the coffee content again.
(c) Which measure of central tendency, the mean or the median, was affected more by the data entry error?

38. *U.S. Exports* The following data are the U.S. exports (in billions of dollars) to 19 countries for a recent year. *(Source: U.S. Department of Commerce)*

Canada	160.8	Japan	51.4
Mexico	97.5	United Kingdom	33.3
Germany	26.6	South Korea	22.6
Taiwan	18.4	Singapore	16.2
Netherlands	18.3	France	19.0
China	22.1	Brazil	12.4
Australia	13.1	Belgium	13.3
Malaysia	10.3	Italy	10.1

| Switzerland | 7.8 | Thailand | 4.9 |
| Saudi Arabia | 4.8 | | |

(a) Find the mean and median.

(b) Find the mean and median without the U.S. exports to Canada.

(c) Which measure of central tendency, the mean or the median, was affected more by the elimination of the Canadian export data?

In Exercises 39-42, find the range, mean, variance, and standard deviation of the population data set.

| 39. | 11 | 10 | 8 | 4 | 6 | 7 | 11 | 6 | 11 | 7 | 20 | 13 |

| 40. | 13 | 23 | 15 | 13 | 18 | 13 | 15 | 14 | 20 | 20 | 18 | 17 |

| 41. | 15 | 8 | 12 | 5 | 19 | 14 | 8 | 6 | 13 | 11 | 23 | 12 |

| 42. | 24 | 26 | 27 | 23 | 9 | 14 | 8 | 8 | 26 | 15 | 15 | 27 |

43. Find the range of the data set.   Key: 2 | 3 = 23

```
2 | 3 3
3 | 0 0 2 3 3 8
4 | 0 1 2 3 3 8
5 | 0 1 1 9
6 | 1 2 9 9
7 | 5 9
8 | 4 8
9 | 0 2 5 6
```

44. Explain how to find the range of a data set. What is an advantage of using the range as a measure of variation? What is a disadvantage?

45. Explain how to find the deviation of an entry in a data set. What is the sum of all the deviations in any data set?

46. Why is the standard deviation used more frequently than the variance? (*Hint:* Consider the units of the variance.)

47. Explain the relationship between variance and standard deviation. Can either of these measures be negative? Explain. Find a data set for which

$$n = 5, \quad \overline{X} = 7, \text{ and } \quad s = 0.$$

48. The number of gallons of water consumed per day by a small village are listed. Make a frequency distribution (using five classes) for the data set. Then approximate the population mean and the population standard deviation of the data set.

167	180	192	173	145	151	174	175	178	163
175	178	160	195	224	244	146	167	180	163
162	146	177	163	149	188	192	173	145	151

49. The following table shows the heights (in inches) and weights (in pounds) of the members of a basketball team. Find the coefficient of variation for each data set. What can you conclude?

HEIGHT	WEIGHTS
72	180
74	168
68	225
76	201
74	189
69	192
72	197
79	162
70	174
69	171
77	185
73	210

50. In the following list, post a D for the situations in which statistical techniques are used for the purpose of description and an I for those in which the techniques are used for the purpose of inference.

_____ (a) The price movements of 50 issues of stock are analysed to determine whether stocks in general have gone up or down during a certain period of time.

_____ (b) A statistical table is constructed for the purpose of presenting the passenger-miles flown by various commercial airlines in the United States.

_____ (c) The average of a group of test scores is computed so that each score in the group can be classified as being either above or below average.

_____ (d) Several manufacturing firms in a particular industry are surveyed for the purpose of estimating industrywide investment in capital equipment.

51. No matter how few elements are included in a statistical population, however, a sample taken from that population (can/cannot) be larger than the population itself.

52. Thus any descriptive measurement of a population is considered to be a (statistics/parameter), and a descriptive measurement of a sample is a sample _____.

53. The word "statistics" has at least three distinct meanings, depending on the context in which it is used. It may refer to:
(i) the procedure of statistical analysis
(ii) descriptive measures of a sample
(iii) the individual measurements, or elements, that make up either a sample or a population.

(a) When one becomes "an accident statistics" by being included in some count of accident frequency, the term is used in the sense of definition ____.
(b) According to the definitions in a course of study called "Business Statistics" the term "statistics" is usually used in the sense of definition ____.
(c) According to the definitions when such sample statistics as the proportion of a sample in favour of a proposal and the average age of those in the sample are determined, the term "statistics" is being used in the sense of definition ____.

54. The two major applications of the tools of statistical analysis are directed toward the purposes of statistical _____ and statistical _____.

55. When all the elements in a statistical population are measured, the process is referred to as "taking a _____". If only a portion of the elements included in a statistical population is measured, the process is called _____.

56. Which of the following measures of variability is not dependent on the exact value of each observation?
(a) range
(b) variance

(c) standard deviation

(d) coefficient of variation

57. A measure of dispersion which is insensitive to extreme values in the data set is the:
    (a) Quartile deviation
    (b) Standard deviation
    (c) Average absolute deviation
    (d) All of the above

58. An absolute measure of dispersion which expresses variation in the same units as the original data is the:
    (a) Standard deviation
    (b) Coefficient of variation
    (c) Variance
    (d) All of the above

59. How does the computation of a sample variance differ from the computation of a population variance?
    (a) μ is replaced by $\overline{X}$
    (b) N is replaced by n −1
    (c) N is replaced by n
    (d) a and c but not b
    (e) a and b but not c

60. Why isn't an average computed from a group frequency distribution exactly the same as that computed from the original raw data used to construct the distribution?

61. For which type of distribution (positively skewed, negatively skewed, or symmetric) is:
    (a) The mean less than the median?
    (b) The mode less than the mean?
    (c) The median less than the mode?

62. The following scores represent the final examination grade for an elementary statistics course:

23	60	79	32	57	74	52	70	82	36	80	77
81	95	41	65	92	85	55	76	52	10	64	75
78	25	80	98	81	67	41	71	83	54	64	72

| 88 | 62 | 74 | 43 | 60 | 78 | 89 | 76 | 84 | 48 | 84 | 90 |
| 15 | 79 | 34 | 67 | 17 | 82 | 69 | 74 | 63 | 80 | 85 | 61 |

Using 10 class intervals with the lowest starting at 9:

(a) Set up a frequency distribution.

(b) Construct a cumulative frequency distribution.

(c) Construct a frequency histogram.

(d) Construct a smoothed cumulative frequency polygon.

(e) Estimate the number of people who made a score of at least 60 but less than 75.

(f) Discuss the skewness of the distribution.

(g) Construct de box-plot and interpret it.

63. Classify the following random variables as discrete or continuous.

(a) The number of automobile accidents each year in Hong Kong.

(b) The length of time to do problem 1 above.

(c) The amount of milk produced yearly by a particular cow.

(d) The number of eggs laid each month by 1 hen.

(e) Numbers of shares sold each year in the stock market.

(f) The weight of grain in kg produced per acre.

64. In four attempts it took a person 48, 55, 51 and 50 minutes to do a certain job.

(a) Find the mean, the range, and the standard deviation of these four sample values.

(b) Subtract 50 minutes from each of the times, recalculate the mean, the range, and the standard deviation, and compare the results with those obtained in part (a).

(c) Add 10 minutes to each of the times, recalculate the mean, the range, and the standard deviation, and compare the results with those obtained in part (a).

(d) Multiply each of the sample values by 2, recalculate the mean, the range, and the standard deviation, and compare the results with those obtained in part (a).

(e) In general, what effect does (1) adding a constant to each sample value, and (2) multiplying each sample value by a positive constant, have on the mean, the range, and the standard deviation of a sample?

65. Find the mean, median and mode for the set of numbers

(a) 3, 5, 2, 6, 5, 9, 5, 2, 8, 6;

(b) 51.6, 48.7, 50.3, 49.5, 48.9.

66. The lengths of a large shipment of chromium strips have a mean of 0.44 m and standard deviation of 0.001 m. At least what percentage of these lengths must lie between
    (a) 0.438 and 0.442 m?
    (b) 0.436 and 0.444 m?
    (c) 0.430 and 0.450 m?

67. The 1971 populations and growth rates for various regions are given below. Find the growth rate for the world as a whole Region Population (millions) Annual Growth Rate (%).

Europe	470	0.8
USSR	240	1.1
N. America	230	1.3
Oceania	20	2.1
Asia	2,100	2.3
Africa	350	2.6
S. America	290	2.9

68. Suppose that the annual income of the residents of a certain country has a mean of $48,000 and a median of $34,000. What is the shape of the distribution?

69. Explain what is meant by the term *population*.

70. Explain what is meant by the term *sample*.

71. Explain how a sample differs from a population.

72. Explain what is meant by the term *sample data*.

73. Explain what a *parameter* is.

74. Explain what a *statistic* is.

75. Give an example of a population and two different characteristics that may be of interest.

76. Describe the difference between *descriptive statistics* and *inferential statistics*. Illustrate with an example.

77. Identify each of the following data sets as either a population or a sample:
    a. The grade point averages (GPAs) of all students at a college.
    b. The GPAs of a randomly selected group of students on a college campus.
    c. The ages of the nine Supreme Court Justices of the United States on January 1, 1842.
    d. The gender of every second customer who enters a movie theater.
    e. The lengths of Atlantic croakers caught on a fishing trip to the beach.

78. Identify the following measures as either quantitative or qualitative:
    a. The 30 high-temperature readings of the last 30 days.
    b. The scores of 40 students on an English test.
    c. The blood types of 120 teachers in a middle school.
    d. The last four digits of social security numbers of all students in a class.
    e. The numbers on the jerseys of 53 football players on a team.

79. Identify the following measures as either quantitative or qualitative:
    a. The genders of the first 40 newborns in a hospital one year.
    b. The natural hair color of 20 randomly selected fashion models.
    c. The ages of 20 randomly selected fashion models.
    d. The fuel economy in miles per gallon of 20 new cars purchased last month.
    e. The political affiliation of 500 randomly selected voters.

80. A researcher wishes to estimate the average amount spent per person by visitors to a theme park. He takes a random sample of forty visitors and obtains an average of $28 per person.
    a. What is the population of interest?
    b. What is the parameter of interest?
    c. Based on this sample, do we know the average amount spent per person by visitors to the park? Explain fully.

81. A researcher wishes to estimate the average weight of newborns in South America in the last five years. He takes a random sample of 235 newborns and obtains an average of 3.27 kilograms.
    a. What is the population of interest?
    b. What is the parameter of interest?
    c. Based on this sample, do we know the average weight of newborns in South America? Explain fully.

82. A researcher wishes to estimate the proportion of all adults who own a cell phone. He takes a random sample of 1,572 adults; 1,298 of them own a cell phone, hence 1298∕1572 ≈ .83 or about 83% own a cell phone.

a. What is the population of interest?

b. What is the parameter of interest?

c. What is the statistic involved?

d. Based on this sample, do we know the proportion of all adults who own a cell phone? Explain fully.

83. A sociologist wishes to estimate the proportion of all adults in a certain region who have never married. In a random sample of 1,320 adults, 145 have never married, hence 145∕1320 ≈ .11 or about 11% have never married.

a. What is the population of interest?

b. What is the parameter of interest?

c. What is the statistic involved?

d. Based on this sample, do we know the proportion of all adults who have never married? Explain fully.

84. a. What must be true of a sample if it is to give a reliable estimate of the value of a particular population parameter?

b. What must be true of a sample if it is to give *certain* knowledge of the value of a particular population parameter?

85. Which of the following are part(s) of statistics? Select all that apply.
a) numerical calculations
b) graphs
c) interpretations and decisions based on the numbers and graphs

86. Higher rates of ice cream consumption and drowning for a city correspond. This leads people to believe that eating ice cream can somehow put you at risk of drowning. Can you think of another interpretation?

a)   Yes
b)   No

87. You hear in a commercial that 80% of children prefer to eat a certain kind of cereal for breakfast. What do you conclude?
a) This cereal is superior to all others...at least according to kids.
b) You need to know more about where these data came from before making any conclusions.

c) 20% of kids prefer to eat Trix.

88. What should you take into consideration when evaluating statistical claims? Check all that apply.
    a) The statistics presented
    b) The sources of the statistical findings
    c) The procedures used to generate the claims

89. Two basic divisions of statistics are
    a) inferential and descriptive.
    b) population and sample.
    c) sampling and scaling.
    d) mean and median.

90. Check all that apply. Descriptive statistics
    a) allow random assignment to experimental conditions.
    b) use data from a sample to answer questions about a population.
    c) summarize and describe data.
    d) allow you to generalize beyond the data at hand.

Unidad II: Probabilidades y Teorema de Bayes.

Objetivo: Identificar, calcular e interpretar adecuadamente las medidas de probabilidad y el Teorema de Bayes en las ciencias de la salud para distinguir su aplicación en la recolección e interpretación de datos.

Contenido:

2.1 Cálculo de la probabilidad individual.

2.2 Cálculo de la probabilidad compuesta: probabilidad conjunta y probabilidad de unión.

2.3 Cálculo de la probabilidad condicional.

2.4 Teorema de Bayes: sensibilidad, especificidad, falsos positivos, falsos negativos, valor predictivo positivo y valor predictivo negativo.

1. A child orders breakfast at a restaurant. The restaurant has two choices of drinks: milk and orange juice. The restaurant also has three choices of meat: sausage, ham, and bacon. Suppose the child orders one drink and one type of meat.

    a) Give the sample space that shows all the different outcomes for what the child might order.
    b) How many different outcomes are possible?

2. The lunch at The Diner has a choice of ham, turkey, or roast beef on rye or white bread with coffee or milk. Draw a tree diagram that illustrates what a person might have for lunch if they pick only one meat, one bread, and one drink.

3. Find the value for each expression below. Show your work by hand and use your calculator to verify your results.
    a) 5!
    b) $_6P_3$
    c) $_7C_5$
    d) $(5 - 2)!$
    e) $4! - 2!$

4. A homeless shelter has decided to purchase all new kitchen appliances. They need one oven, one refrigerator, and one dishwasher. The appliance store has 7 brands of ovens, 6 brands of refrigerators, and 5 brands of dishwashers. In how many brand arrangements can they purchase their appliances?

5. An ice cream shop has 8 different flavors of ice cream available. How many 2-scoop cones can be made if you are allowed to have the same flavor for both scoops?

6. An ice cream shop has 8 different flavors of ice cream available. How many 2-scoop cones can be made if you decide not to have the same flavor for both scoops?

7. For $7.99, a restaurant will sell you their lunch special. The special is either a hamburger or chicken sandwich, onion rings or fries, and soda or coffee.
    a) Make a tree diagram showing the different ways a customer may order the lunch special.
    b) How many outcomes are there? Use the Fundamental Counting Principle to justify your answer.

8. A couple is planning to have 3 children. Consider the different results that might occur in terms of gender. For example one outcome might be Boy, Boy, Girl (BBG).

    a) Using the Fundamental Counting Principle, calculate the number of different outcomes for the children in this family.
    b) Build a tree diagram that shows the different orders of children the couple might have.

c) Construct the sample space that shows all the different orders of children the couple might have.

9. A woman has three skirts, five shirts, and four hats. How many different outfits can she wear if she picks one skirt, one shirt, and one hat for her outfit?

10. How many different five-digit ZIP codes are possible if the digits can be repeated?

11. How many different five-digit ZIP codes are possible if the digits cannot be repeated?

12. In how many ways can a baseball manager arrange a batting order of nine players?

13. A store manager wishes to display six different brands of laundry soap by lining them up in a row on a shelf. In how many ways can this be done?

14. There are 8 different statistics books, 6 different geometry books, and 3 different trigonometry books being considered for next year. In how many ways can a textbook committee select one of each book?

15. At a film festival, there are eight different films that will be shown. In how many different orders can these films be shown?

16. The call letters of a radio station must have four letters. The first letter must be a K or a W. How many different call letter combinations are possible if letters may not be repeated?

17. The call letters of a radio station must have four letters. The first letter must be a K or a W. How many different call letter combinations are possible if letters may be repeated?

18. How many different four-digit ID tags can be made if repeats are allowed?

19. How many different four-digit ID tags can be made if it must start with a 7 and no repeats are allowed?

20. In how many different ways can the Harry Potter series of books (7 books total) be arranged in a row on a shelf?

21. A coin is tossed 8 times. How many different outcomes are there for this series of 8 flips?

22. A pizza restaurant offers 6 different toppings for their pizzas. How many different pizzas are possible?

23. Use a tree diagram to find all possible outcomes for the result of a series of coin flips if the coin is flipped two times. Write a list of the possible results when complete.

24. The Super-Cool Ice Cream Shoppe sells sundaes, cones, or ice cream bars. You will pick either butterscotch or chocolate and you may choose to have it with nuts or without nuts.

    a) Draw a tree diagram to illustrate the different types of ice cream treats that you could order.
    b) How could you find the number of outcomes using the Fundamental Counting Principle?
    c) How many different outcomes are possible?

25. A quiz has four true/false questions on it. Use a tree diagram to show all the different possible answer keys.

26. A box contains a $1 bill, a $5 bill, and a $10 bill. Two bills are selected one after the other without replacing the first bill. Draw a tree diagram to show all possible amounts of money that may be drawn.

27. The Eagles and Hawks play each other in a hockey tournament. The first team to win two games is the champion. Use a tree diagram to show all different possible outcomes for the tournament.

28. Suppose you are going to order an ice cream cone with two different flavored scoops. You are going to take a picture of your ice cream cone for use in the school newspaper. The ice cream shop has 5 flavors to choose from; chocolate, vanilla, orange, strawberry, and mint. How many different ice cream cone photos are possible?

29. Suppose that there are 12 employees in an office. The boss needs to select 4 of the employees to go on a business trip to California. In how many ways can she do this?

30. In how many ways can a student select five questions to answer from an exam containing nine questions?

31. In how many ways can a student select five questions to answer from an exam containing nine questions if the student is required to answer the first and the last question?

32. The general manager of a fast-food restaurant chain must select 6 restaurants from 11 for a promotional program. In how many different possible ways can this selection be done?

33. There are 7 women and 5 men in a department. In how many ways can a committee of 4 people be selected?

34. In how many ways can committee of 3 people be chosen if there are 8 men and 4 women available for selection and we require that two men and one woman be on the committee?

35. In your own words, state how you can tell the difference between a combination and permutation problem.

36. Your closet contains 10 different styles of shoes. In how many ways can you pick out five different styles of shoes for the school week if you don't care which day of the week you wear each style?

37. There are 7 women and 5 men in a department. Four people are needed for a committee.
    a) In how many ways can a committee of 4 people be selected?
    b) In how many ways can this committee be selected if there must be exactly 2 men and 2 women on the committee?
    c) In how many ways can this committee be selected if there must be at least 2 women on the committee?

38. Flakes-R-Us cereal comes in two types, Sugar Sweet and Touch O' Honey. If a researcher has ten boxes of each type, how many ways can she select two boxes of each for a quality control test?

39. A bag contains some jelly beans. There are a total of 6 red jelly beans, 4 green jelly beans, 2 black jelly beans, 5 yellow jelly beans, and 3 orange jelly beans in the bag. Suppose one jelly bean is drawn from the bag.
    a) Find P(purple).
    b) Find P(yellow).
    c) Find P(˘red)

40. If we randomly pick two television sets in succession from a shipment of 240 television sets of which 15 are defective, what is the probability that they will both be defective?

41. A poll of 500 people determines that 382 like ice cream and 362 like cake. How many people like both if each of them likes at least one of the two? (Hint: Use $P(A \cap B) = P(A) + P(B) - P(A \cup B)$).

42. The Mathematics Department of the University of Louisville consists of 8 professors, 6 associate professors, 13 assistant professors. In how many of all possible samples of size 4, chosen without replacement, will every type of professor be represented?

43. A pair of dice consisting of a six-sided die and a four-sided die is rolled and the sum is determined. Let A be the event that a sum of 5 is rolled and let B be the event that a sum of 5 or a sum of 9 is rolled. Find:

 (a) P(A),

(b) P(B), and

(c) P(A ∩ B).

44. A faculty leader was meeting two students in Paris, one arriving by train from Amsterdam and the other arriving from Brussels at approximately the same time. Let A and B be the events that the trains are on time, respectively. If P(A) = 0.93, P(B) = 0.89 and P(A ∩ B) = 0.87, then find the probability that at least one train is on time.

45. Bill, George, and Ross, in order, roll a die. The first one to roll an even number wins and the game is ended. What is the probability that Bill will win the game?

46. Suppose a box contains 4 blue, 5 white, 6 red and 7 green balls. In how many of all possible samples of size 5, chosen without replacement, will every color be represented?

47. A box contains five green balls, three black balls, and seven red balls. Two balls are selected at random without replacement from the box. What is the probability that both balls are the same color?

48. Find the sample space of the random experiment which consists of tossing a coin until the first head is obtained. Is this sample space discrete?

49. In a particular softball league each team consists of 5 women and 5 men. In determining a batting order for 10 players, a woman must bat first, and successive batters must be of opposite sex. How many different batting orders are possible for a team?

50.- Find de sample space of the random experiment wuich consists of tossing a coin infinitely many times. Is the sample space discrete?

51. An urn contains 3 red balls, 2 green balls and 1 yellow ball. Three balls are selected at random and without replacement from the urn. What is the probability that at least 1 color is not drawn?

52. A box contains four $10 bills, six $5 bills and two $1 bills. Two bills are taken at random from the box without replacement. What is the probability that both bills will be of the same denomination?

53. Mr. Flowers plants 10 rose bushes in a row. Eight of the bushes are white and two are red, and he plants them in random order. What is the probability that he will consecutively plant seven or more white bushes?

54. **U.S. Population, Female** The following table shows the number of females in the United States in 2000, broken down by age.4 Numbers are in millions.

Age	0-18	18-25	25-35	35-45	45-55	55-65	65-75	75 and over
Number	35.2	13.3	19.8	22.7	19.2	12.6	10.1	10.5

Construct the associated probability distribution (with probabilities rounded to four decimal places) and use the distribution to compute the following.
(a) $P(18 \leq X \leq 55)$
(b) $P(X \leq 45)$
(c) $P(X \leq 45)$

55. **U.S. Population, Male** The following table shows the number of males in the United States in 2000, broken down by age.5 Numbers are in millions.

Age	0-18	18-25	25-35	35-45	45-55	55-65	65-75	75 and over
Number	37.1	13.9	20.1	22.4	18.5	11.6	8.3	6.18

Construct the associated probability distribution (with probabilities rounded to four decimal places) and use the distribution to compute the following.
(a) $P(25 \leq X \leq 65)$
(b) $P(X \leq 18)$
(c) $P(X \geq 15)$

56. John forgot to study for his history quiz, so he will guess on each question. The quiz has 5 true-false questions and 5 multiple-choice questions (with 4 choices each). He will guess an answer for each question. In how many possible ways might John answer all of the questions?

57. What is the probability that John will get all of the questions correct?

58. You might be interested in finding out a patient's probability of having liver disease if they are an alcoholic. "Being an alcoholic" is the **test** (kind of like a litmus test) for liver disease.

- **A** could mean the event "Patient has liver disease." Past data tells you that 10% of patients entering your clinic have liver disease. P(A) = 0.10.
- **B** could mean the litmus test that "Patient is an alcoholic." Five percent of the clinic's patients are alcoholics. P(B) = 0.05.
- You might also know that among those patients diagnosed with liver disease, 7% are alcoholics. This is your **B|A:** the probability that a patient is alcoholic, given that they have liver disease, is 7%.

59. **Given the following statistics, what is the probability that a woman has cancer if she has a positive mammogram result?**
- One percent of women over 50 have breast cancer.
- Ninety percent of women who have breast cancer test positive on mammograms.
- Eight percent of women will have false positives.

60. Approximately 1% of women aged 40-50 have breast cancer. A woman with breast cancer has a 90% chance of a positive test from a mammogram, while a woman without has a 10% chance of a false positive result. What is the probability a woman has breast cancer given that she just had a positive test?

61.- Assuming women and men exist in equal number, and assuming that 5% of the men are colour blind and that 0,25% of the women are colour blind, evaluate the probability that a person drawn at random is colour blind. Then evaluate the probability that, having drawn a colour-blind person, this is a male.

62. A drug test (random variable T) has 1% false positives (i.e., 1% of those not taking drugs show positive in the test), and 5% false negatives (i.e., 5% of those taking drugs test negative). Suppose that 2% of those tested are taking drugs. Determine the probability that somebody who tests positive is actually taking drugs (random variable D).

Unidad III: Muestreo

1. Which of the following is an example of *nonprobabilistic sampling*?
   a) simple random sampling
   b) stratified simple random sampling
   c) cluster sampling
   d) judgment sampling
   e) None of the above answers is correct

2. Which of the following sampling methods does *not* lead to *probability samples*?
   a) stratified sampling
   b) cluster sampling
   c) systematic sampling
   d) convenience sampling
   e) None of the above answers is correct

3. *Stratified random sampling* is a method of selecting a sample in which:
   a) the sample is first divided into strata, and then random samples are taken from each stratum
   b) various strata are selected from the sample
   c) the population is first divided into strata, and then random samples are drawn from each stratum
   d) None of the above answers is correct

4. Each student at a school has a student identification number. Counselors have a computer generate 50 random identification numbers, and the students associated with those numbers are asked to take a survey. What type of sample is this?

5. An airline company wants to survey its customers one day, so they randomly select 5 flights that day and survey every passenger on those flights. What type of sample is this?

6. Come up with one sentence summarizing the key difference between stratified and cluster sampling as if you were teaching this concept to a student who is just starting off in a statistics course.

7. Suppose you are going to be conducting a study on Duke students, asking for their opinion on an issue of interest to you (could be related to the university, or a wider societal issue). First, formulate your research question. Then, describe how you would carry out the sampling of students using the following methods:
   a. simple random sampling
   b. stratified sampling
   c. cluster sampling

8. Think about what attributes of the Duke student population make sense to stratify vs. cluster on with respect to your research question.

9. The following data refers to the Kapas yield of 96 plants.

82	102	88	93	97	38	103	92	102	62	54
63	72	64	68	59	69	73	65	46	79	55
87	84	29	52	28	36	37	53	49	51	91
30	37	56	66	42	37	35	97	32	35	12
89	99	54	72	26	67	18	27	60	72	82
33	42	52	82	14	22	57	73	63	61	76
63	92	40	58	62	61	43	25	42	36	50
17	30	75	87	47	56	76	36	35	44	80
56	51	111	73	93	58	49	89			

Select a sample of 25 plants by using simple random sampling method. Also calculate the mean of the 25 samples and verify whether the mean is equal to the mean of the 96 plants.

10. Suppose we have a *population of 25 elements*:

01	02	03	04	05	06	07	08	09	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25					

Suppose the first row of the table of random number is

88547 09896 95436 79115 08303 01041 20030 63754 08459 28364

Please use *systematic sampling* to obtain *a sample of 5* elements and then find *the sample mean*.

11. Let's say, 100 (N_h) students of a school having 1000 (N) students were asked questions about their favorite subject. It's a fact that the students of the 8th grade will have different subject preferences than the students of the 9th grade. For the survey to deliver precise results, the ideal manner is to divide each grade into various strata.

Here's a table of the number of students in each grade:

Grade	Number of students (N)
5	150
6	250
7	300
8	200
9	100

Calculate the sample of each grade using the stratified random sampling formula:

12. You hypothesize that students in fraternities/sororities on your campus are more extroverted (on an introversion/extroversion scale) than other students. To answer your research question, you need to obtain a sample of "Greek" students and a sample of "independent" students. For each of these two populations of students, how would you obtain simple random samples?

13. For the hypothesis in question 12, describe how you would obtain samples using stratified random sampling, with stratification based on the sex of the student. Assume that 60% of the Greeks are women.

14. For the hypothesis in question 12, describe how you would obtain samples using convenience sampling.

15. For the hypothesis in question 12, describe how you would obtain samples using quota sampling based on the sex of the student. Again, assume that 60% of the Greeks are women.

16. For the hypothesis in question 12, describe a poor method of sampling that you believe would result in systematic sampling error and thus bias the research findings. Explain why you believe the sample would be biased

17. Willy runs a small company with 10 employees. He decides to pick a simple random sample of 3 employees to go on a business trip. He numbers them 0- 9 and uses the random digit table printed below to select the sample. Which employees are in the sample?

    10056 42211 65491

18. What is the difference between a cluster sample and a stratified random sample?

19. When would you use non-random sampling?

20. A store prints a request on each receipt asking customers to fill out a satisfaction survey online if they are willing. What type of sample is this?

21. A politician asks his neighbors their opinions about a controversial issue. What type of sample is this?

22. Each law firm in one state registers its phone number with the state court system. An employee of the state court system uses a computer to select 500 random registered phone numbers, and the law firms associated with those numbers will be selected for an audit. What type of sample is this?

23. A large company surveys 100 employees by taking random samples of 10 managers and 90 non-managerial employees. What type of sample is this?

24. A restaurant leaves comment cards on all of its tables to learn about customer satisfaction. Each comment card is a brief survey that customers can fill out if they choose to. What type of sample is this?

25. A truck manufacturer selects 3 trucks at random from each of 6 models for safety testing. What type of sample is this?

26. A support hotline asks callers to stay on the line after they have completed their call to respond to a short survey. What type of sample is this?

27. A principal orders t-shirts and wants to check some of them to make sure they were printed properly. She randomly selects 2 of the 10 boxes of shirts and checks every shirt in those 2 boxes. What type of sample is this?

28. A quality control worker at a factory selects the first 10 items she sees as her sample for the day. What type of sample is this?

29. A teacher puts students' names in a hat and chooses without looking to get a sample of 3 students. What type of sample is this?

For each of the following questions (30 – 34), suggest an appropriate sampling design and sample size.

30. A public health official wants to estimate the number of babies who are being born infected with HIV.

31. The City Manager wants to know, by next Monday, the extent of pothole damage in the city caused by the latest El Niño storms.

32. The Chief County Librarian wants to know what the patrons think of the county's branch libraries, and whether they would be willing to support a tax increase for the libraries.

33. Why do some residents in nursing homes get broken hips when they fall, while other residents fall but do not break their hips, and other residents do not fall at all?

34. Do older Hispanic women who live in neighborhoods with higher proportions of Hispanic residents get fewer preventive health care checkups than older Hispanic women who live in neighborhoods with lower proportions of Hispanic residents?

35. Suppose your school has 500 students and you need to conduct a short survey on the quality of the food served in the cafeteria. You decide that a sample of 10 students should be sufficient for your purposes. In order to get your sample, you assign a number from 1 to 500 to each student in your school. To select the sample, you use a table of randomly generated numbers. What type of sample is this?

Unidad IV: Bioestadística inferencial: Distribuciones de variables discretas y continuas.

1. Patient noncompliance is one of many potential soureces of bias in medical studies. Consider a study where patients are asked to take 2 tablets of a certain medication in the morning, and two tabletes at bed time. Suppose however, that the patients do not always fully cumply and take both tablets at both times; it can also occur that only 1 tablet, or even none, are taken at either of these times.

   a) Explicitly construct the sample space S of all posible daily outcomes for a randomly selected patient.

   b) Explicitly list of outcomes in the evento that a patient takes *al least* one Tablet at both times, and calculate the probability, <u>assuming that the outcomes are eccually likely.</u>

   c) Construct a probability table and corresponding probability histogram for the random variable X = "the daily total numbers of tablets taken by a random patient."

   d) Calculate the daily mean number of tablets taken.

   e) Suppose that the outcomes are not equally likely, but vary as follows:

# tablets	AM probability	PM probability
0	0.1	0.2
1	0.3	0.3
2	0.6	0.5

   Rework parts b) – d) using these probabilities. Assume independence between AM and PM.

Wheel 1		Wheel 2		Wheel 3	
Outcome	Probability	Outcome	Probability	Outcome	Probability
A		A		A	
B		B		B	
C		C		C	
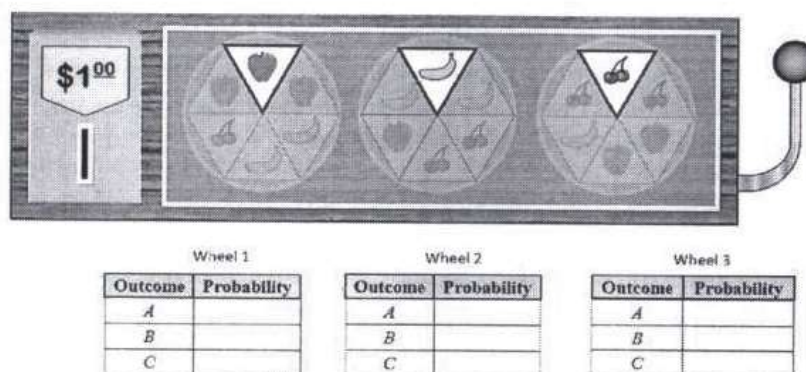
2. A casino slot machine consists of three wheels, each with images of three types of fruit: apples, bananas, and cherries. When a player pulls the handle, the wheels spin independently of one another, until each one stops at a random image displayed in its window, as shown above. Thus, the

sample space S of possible outcomes consists of the 27 orderer triples shown below, where events A = "Apple", B = "Banana"", and C = " Cherries".

$(A A A), (A A B), (A A C), (A B A), (A B B), (A B C), (A C A), (A C B), (A C C)$

$(B A A), (B A B), (B A C), (B B A), (B B B), (B B C), (B C A), (B C B), (B C C)$

$(C A A), (C A B), (C A C), (C B A), (C B B), (C B C), (C C A), (C C B), (C C C)$

a) Complete the individual tables above, and use these to construct the probability table (including the outcomes) for the discrete random variable X = "# Apples" that are displayed when the handle is pulled.

X	Outcomes	Probability f(x)

b) Sketch the corresponding probability histogram of X. Label all relevant features.
c) Calculate the mean μ and variance σ^2 of X.
d) Similar to X = "# Apples", define random variables Y = "# Bananas" and Z = "# Cherries" displayed in one play. The player wins if all three displayed images are of the same fruit. Using these variables, calculate the probability of a win.
e) Suppose it costs one dollar to play this game once. The result is that either the player loses the dollar, or if the player wins, the slot machine pays out ten dollars in coins. If the player continues to play this game indefinitely, should he/she expect to win money, lose money, or neither, in the long run? If win or lose money, how much per play? Show all the work.

3. When the health department tested private wells in a county for two impurities commonly found in drinking water, it found that 20% of the wells had neither impurity, 40% had impurity A, and 50% had impurity B. (Obviously, some had both impurities). If a well is randomly chosen from those in the county, find the probability distribution for Y, the number of impurities found in the well.

4. In a family of 11 children, what is the probability that there will be more boys than girls?

5. Seven out of every 100 people in the United States have Type O negative blood. If 6 people are chosen at random, what is the probability that exactly 2 of them have Type O negative blood?

6. Nine out of every 100 people in the United States have Type B positive blood. If 8 people are chosen at random, what is the probability that exactly 3 of them have Type B positive blood?

7. In a doctor's office, 80% of the patients are adults. If 15 patients are scheduled for an appointment on a given day, what is the expected number of adults? What is the standard deviation?

8. One in every 6 coffee drinkers at a restaurant prefers coffee with cream, sugar, or both. If 4 coffee drinkers are chosen at random, what is the probability that at least 1 coffee drinker prefers coffee with cream, sugar, or both?

9. A multiple-choice examination has 15 questions, each with five possible answers, only one of which is correct. Suppose that one of the students who takes the examination answers each of the questions with and independent random guess. What is the probability that he answers at least ten questions correctly?

10. Tay-Sachs disease is a genetic disorder that is usually fatal in young children. If both parents are carriers of the disease, the probability that the offspring will develop the disease is approximately 0.25. Suppose that a husband and wife are both carriers and that they have three children. If the outcomes of the three pregnancies are mutually independent, what are the probabilities of the following events?
    a) All three children develop Tay-Sachs.
    b) Only one child develops Tay-Sachs.
    c) The third child develops Tay-Sachs, given that the first two did not.

11. Suppose that 80% of adults with allergies report symptomatic relief with a specific medication. If the medication is given to 10 new patients with allergies, what is the probability that it is effective in exactly seven?

12. Of the volunteers donating blood in a clinic, 80% have the Rhesus (Rh) factor present in their blood.
    a) If five volunteers are randomly selected, what is the probability that al least one does not have the Rh factor)
    b) If five volunteers are randomly selected, what is the probability that at most four have the Rh factor?
    c) What is the smallest number of volunteers who must be selected if we want to be at least 90% certain that we obtain at least five donors with the Rh factor?

13. The likelihood that a patient with a heart attack dies of the attack is 0.04 (i.e., 4 of 100 die of the attack). Suppose we have 5 patients who suffer a heart attack, what is the probability that all will survive?

14. Suppose the occurrence of a rare disease in a certain population is known to follow a Poisson distribution, with an average of $\lambda = 2.3$ cases per year. In a typical year, what is the probability that...
    a) no cases occur?
    b) exactly one case occurs?
    c) exactly two cases occur?
    d) three or more cases occur?
    e) Answer (a) – (d) for a typical two-year period. (Assume independence from year to year.)

15. Suppose a fast food restaurant can expect two customers every 3 minutes, on average. What is the probability that four or fewer patrons will enter the restaurant in a 9 minutes period?

16. Consider a simple emergency room example where 2 patients arrive, on average, every 10 minutes. Consecutive arrivals are statistically independent. Letting N(t) represent the number of such arrival in t minutes, evaluate the following probabilities, using the Poisson distribution.
    a) $P[N(60)] = 10$

b) P [N(60)] ≤ 10
c) P [N(60)] < 10
d) P [N(60)] > 10
e) P [N(60)] ≥ 10

17. Let Y denote a random variable that has a Poisson distribution with mean $\lambda = 2$. Find
   a) $P(Y = 4)$.
   b) $P(Y \geq 4)$
   c) $P(Y < 4)$
   d) $P(Y \geq 4 / Y \geq 2)$.

18. The random variable Y has a Poisson distribution and is such that $p(0) = p(1)$. What is $p(2)$?

19. The probability that a mouse inoculated with a serum will contract a certain disease is 0.1. Using the Poisson approximation, find the probability that al most 3 of 30 inoculated mice will contract the disease.

20. Imagine that a certain disease occurs in a large population in such a way that the probability of a randomly selected individual having the disease remains constant al $\pi = 0.008$, independent of any other randomly selected individual having the disease. Suppose now that a sample of n = 500 individuals is to be randomly selected from this population. Define the discrete random variable X = "the number of diseased individuals," capable of assuming any value in the set 0, 1, 2, ..., 500 for this sample.
   a) Calculate the probability distribution function $p(x) = P(X = x)$ –"the probability that the number of diseased individuals equals x"- for x = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. Do this computations two ways. First, using the Binomial Distribution and second, using the Poisson Distribution, and arrange these values into a probability table. (For the sake of comparison, record al least five decimal places.)

x	Binomial	Poisson
0		
1		
2		
3		
4		
5		
5		
6		
7		
8		
9		
10		
etc	etc	etc

   b) Using either the Binomial or Poisson Distribution, what is the mean member of diseased individuals to be expected in the sample, and what is its probability? How does this probability compare with the probabilities of other numbers of diseased individuals?

33

c) Suppose that, after sampling n = 500 individuals, you find that X = 10 of them actually have this disease, before performing any formal statistical test, what assumptions -if any- might you suspect have been violated in this scenario?

21. If scores are normally distributed with a mean of 30 and a standard deviation of 5, what percent of the scores is: (a) greater than 30? (b) greater than 37? (c) between 28 and 34?

22. a) What are the mean and standard deviation of the standard normal distribution?
    b) What would be the mean and standard deviation of a distribution created by multiplying the standard normal distribution by 10 and then adding 50?

23. The normal distribution is defined by two parameters. What are they?

24. What proportion of a normal distribution is within one standard deviation of the mean?

25. What proportion is more than 1.8 standard deviations from the mean?

26. What proportion is between 1 and 1.5 standard deviations above the mean?

27. A test is normally distributed with a mean of 40 and a standard deviation of 7.
    a) What score would be needed to be in the 85th percentile?
    b) What score would be needed to be in the 22nd percentile?

28. Assume a normal distribution with a mean of 90 and a standard deviation of 7. What limits would include the middle 65% of the cases.

29. The height for boys at age 0 is normally distributed with mean 49.99cm and standard deviation 2.66cm. Find the percentiles given in the next table.

Percentile	3	5	10	25	50	75	90	95	97
Height [cm]									

30. The following data gives the length-at-birth for girls:

Percentile	3	5	10	25	50	75	90	95	97
Height [cm]	45.09	45.58	46.34	47.68	49.29	51.02	52.70	53.77	54.50

Assume that the height-at-birth is normally distributed. Find the mean and the standard deviation.

31. Family income ~ N ($25,000, $10,000). If the poverty level is $10,000, what percentage of the population lives in poverty?

32. A new tax law is expected to benefit "middle income" families, those with incomes between $20,000 and $30,000. If Family income ~ N ($25000, $10000), what percentage of the population will benefit from the law?

33. Suppose that weights of bags of potato chips coming from a factory follow a normal distribution with mean 12.8 ounces and standard deviation .6 ounces. If the manufacturer wants to keep the mean at 12.8 ounces but adjust the standard deviation so that only 1% of the bags weigh less than 12 ounces, how small does he/she need to make that standard deviation?

34. Heights of adult women in the United States are normally distributed with a population mean of $\mu = 63.5$ inches and a population standard deviation of $\sigma = 2.5$. A medical re- searcher is planning to select a large random sample of adult women to participate in a future study. What is the standard value, or z-value, for an adult woman who has a height of 68.5 inches?

35. Suppose we know that the birth weight of babies is Normally distributed with mean 3500g and standard deviation 500g. What is the probability that a baby is born that weighs less than 3100g?

36. The length of human pregnancies from conception to birth approximates a normal distribution with a mean of 266 days and a standard deviation of 16 days.
    a) What proportion of all pregnancies will last between 240 and 270 days (roughly between 8 and 9 months)?
    b) What length of time marks the shortest 70% of all pregnancies?

37. Suppose that we are told that the heights of adult males in a particular region of the world are normally distributed with mean of 70 inches and standard deviation of 2 inches.
    a) Approximately what proportion of adult males are taller than 73 inches?
    b) What proportion of adult males are between 72 and 73 inches?
    c) What height corresponds to the point where 20% of all adult males are greater than this height?
    d) What height corresponds to the point where 20% of all adult males are less than this height?

38. The mean weight of 500 college students is 70 kg and the standard deviation is 3 kg. Assuming that the weight is normally distributed, determine how many students weigh:
    a) Between 60 kg and 75 kg.
    b) More than 90 kg.
    c) Less than 64 kg.
    d) Exactly 64 kg.
    e) 64 kg or less.

39. A bottling company uses a machine to fill the bottles with olive oil. The bottles are designed to contain 475 milliliters (ml). In fact, the contents vary according to a normal distribution with a mean of 473 ml and standard deviation of 3 ml.
    a) What is the probability that one bottle is less than 470 ml?
    b) What is probability that the mean of six bottles is more than 475 ml?

40. Suppose that the weight of navel oranges is normally distributed with mean $\mu = 8$ ounces, and standard deviation $\sigma = 1.5$ ounces. We can write $X \sim N(8, 1.5)$. Answer the following questions:
    a) What proportion of oranges weigh more than 11.5 ounces? (or if you randomly select a navel orange, what is the probability that it weighs more than 11.5 ounces?).
    b) What proportion of oranges weigh less than 8.7 ounces?
    c) What proportion of oranges weigh less than 5 ounces?

d) What proportion of oranges weigh more than 4.9 ounces?

e) What proportion of oranges weigh between 6.2 and 7 ounces?

f) What proportion of oranges weigh between 10.3 and 14 ounces?

g) What proportion of oranges weigh between 6.8 and 8.9 ounces?

h) Find the 80th percentile of the distribution of X. This question can also be asked as follows: Find the value of X below which you find the lightest 80% of all the oranges.

i) Find the 5th percentile of the distribution of X

j) Find the interquartile range of the distribution of X.

41. The lengths of the sardines received by a certain cannery is normally distributed with mean 4.62 inches and a standard deviation 0.23 inch. What percentage of all these sardines is between 4.35 and 4.85 inches long?

42. A baker knows that the daily demand for apple pies is a random variable which follows the normal distribution with mean 43.3 pies and standard deviation 4.6 pies. Find the demand which has probability 5% of being exceeded.

43. Suppose that the height of UCLA female students has normal distribution with mean 62 inches and standard deviation 8 inches.

a) Find the height below which is the shortest 30% of the female students.

b) Find the height above which is the tallest 5% of the female students.

MENSA is an organization whose members possess IQs in the top 2% of the population.

44. If IQs are normally distributed, with mean 100 and a standard deviation of 16, what is the minimum IQ required for admission to MENSA?

45. If three individuals are chosen at random from the general population what is the probability that all three satisfy the minimum requirement for MENSA?

46. Suppose that the height (X) in inches, of a 25-year-old man is a normal random variable with mean $\mu =$ 70 inches. If $P(X > 79) = 0.025$ what is the standard deviation of this random normal variable?

47. Suppose that the weight (X) in pounds, of a 40-year-old man is a normal random variable with standard deviation $\sigma = 20$ pounds. If 5% of this population is heavier than 214 pounds what is the mean μ of this distribution?

At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.1 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of the bottle goes below 35.8 oz. or above 36.2 oz., then the bottle will be declared out of control.

48. If the process is in control, meaning $\mu = 36$ oz. and $\sigma = 0.1$ oz., find the probability that a bottle will be declared out of control.

49. In the situation of (48), find the probability that the number of bottles found out of control in an eight-hour day (16 inspections) will be zero.

50. In the situation of (48), find the probability that the number of bottles found out of control in an eight-hour day (16 inspections) will be exactly one.

51. If the process shifts so that $\mu = 37$ oz and $\sigma = 0.4$ oz, find the probability that a bottle will be declared out of control.

The chickens of the Ornithes farm are processed when they are 20 weeks old. The distribution of their weights is normal with mean 3.8 lb, and standard deviation 0.6 lb. The farm has created three categories for these chickens according to their weight: petite (weight less than 3.5 lb), standard (weight between 3.5 lb and 4.9 lb), and big (weight above 4.9 lb).

52. What proportion of these chickens will be in each category? Show these proportions on the normal distribution graph.

53. Find the 60th percentile of the distribution of the weight. In other words find c such that $P(X < c) = 0.60$.

54. Suppose that 5 chickens are selected at random. What is the probability that 3 of them will be petite?

55. Suppose that the height (X) in inches, of a 25-year-old man is a normal random variable with mean $\mu = 70$ inches. If $P(X > 79) = 0.025$ what is the standard deviation of this random normal variable?

56. Suppose that the weight (X) in pounds, of a 40-year-old man is a normal random variable with standard deviation $\sigma = 20$ pounds. If 5% of this population weigh less than 160 pounds what is the mean μ of this distribution?

57. A bag of cookies is underweight if it weighs less than 500 grams. The filling process dispenses cookies with weight that follows the normal distribution with mean 510 grams and standard deviation 4 grams.
   a) What is the probability that a randomly selected bag is underweight?
   b) If you randomly select 5 bags, what is the probability that exactly 2 of them will be underweight?

58. The weight X of water melons is normally distributed with mean $\mu = 10$ pounds and standard deviation $\sigma = 2$ pounds. Find c such that $P(X > c) = 0.60$.

59. The diameters of apples from the Milo Farm follow the normal distribution with mean 3 inches and standard deviation 0.3 inch. Apples can be size-sorted by being made to roll over a mesh screens. First the apples are rolled over a screen with mesh size 2.5 inches. This separates out all the apples with diameters < 2.5 inches. Second, the remaining apples are rolled over a screen with mash size 3.2 inches. Find the proportion of apples with diameters:
   a) < 2.5 inches,
   b) between 2.5 and 3.2 inches, and
   c) greater than 3.2 inches.

60. 95% of students at school are between 1,1m and 1,7m tall. Assuming this data is normally distributed can you calculate the mean and standard deviation?

61. 95% of students at school weigh between 62 kg and 90 kg. Assuming this data is normally distributed, what are the mean and standard deviation?

62. The Fresha Tea Company pack tea in bags marked as 250 g. A large number of packs of tea were weighed and the mean and standard deviation were calculated as 255 g and 2.5 g respectively. Assuming this data is normally distributed, what percentage of packs are underweight?

63. The mean June midday temperature in Desertville is 36°C and the standard deviation is 3°C. Assuming this data is normally distributed, how many days in June would you expect the midday temperature to be between 39°C and 42°C?

64. The heights of male adults are Normally distributed with mean 1.7 m and standard deviation 0.2 m. In a population of 400 male adults, how many would you expect to have a height between 1.4 m and 1.6 m?

65. The ages of the population of a town are Normally distributed with mean 43 and standard deviation 14. The town has a population of 5,000. How many would you expect to be aged between 22 and 57?

66. The weight of men aged 18 to 74, in pounds, are Normally distributed with $\mu = 173$ and $\sigma = 30$.
    a) Find the percentage of men with weights between 190 lbs and 225 lbs.
    b) Among 400 men aged 18 to 74, how many are expected to weigh between 190 lbs lbs and 225 lbs?

67. The heights of male freshmen entering a large state university are normally distributed with a mean of 68 inches ($\mu = 68$ inches). It is know that about 2% of the freshmen are taller than 72 inches.
    a) Sketch a normal curve that represents the distribution of heights and clearly indicate its mean.
    b) What is the standard deviation of distribution of heights?
    c) About what percent of the freshmen males at this college have a height of 66 inches or less? Sketch a normal curve and clearly indicate the corresponding shaded region that illustrates your solution.

68. 1960-62 Public Health Service Health Examination Survey 6,672 Americans 18-79 years old. The woman's heights were approximately normal with 63" and standard deviation 2.5". What percentage of women were over 68" tall?

69. The ages of employees in a certain workplace are normally distributed. It is known that 80% of the workers are under 65 years old, and 67% are under 55 years old. What percentage of the workers are under 45 years old? (Hint: First find μ and σ by calculating the z-scores.)

Respuestas a los ejercicios seleccionados.

1. a
2. d
3. b
5. d
7. a
11. c
13. a
19. (F) El punto medio o marca de clase de un intervalo es la suma de los lìmites superior e inferior del intervalo, dividida entre 2.
23. (F)
29. mean = 19.6; median = 19.5; mode = 19 and 20
31. No se puede calcular la media, mediana = Some; moda = Worse
37. a) Media = 6.005, Mediana = 6.01
    b) Media = 5.945, Mediana = 6.01
    c) Media
41. Rango = 18; Media = 12.166; Varianza = 27.424; Desviaciòn Estàndar = 5.2368
43. 73
47. La desviación estándar es la raíz cuadrada positiva de la varianza, no pueden ser negativas porque representan las variaciones de un conjunto de datos respecto a la media.
    El conjunto de datos es: 7,7,7,7,7.
53. a) ii
    b) i
    c) iii
59. e
61. a) negativa
    b) positiva
    c) negativa
67. Poblaciòn total: 3700 millones. Crecimiento anual: 2%
71. Una muestra es un subconjunto de una población.
73. Un parámetro es un valor representativo (medida de resumen) de una población y se simboliza por una letra griega, sean μ = media, $\P$ = proporción, σ^2 = varianza, σ = desviación estándar.
79) a) cualitativa
    b) cuantitativa
    c) cuantitativa
    d) cualitativa
83) a) Todos los adultos de una cierta región.
    b) $\P$ = proporción de adultos de la población que nunca se han casado.
    c) p = proporción de adultos de la muestra que nunca se han casado.
    d) No se puede conocer con exactitud el valor del parámetro pero se puede estimar con cierto nivel de confianza.
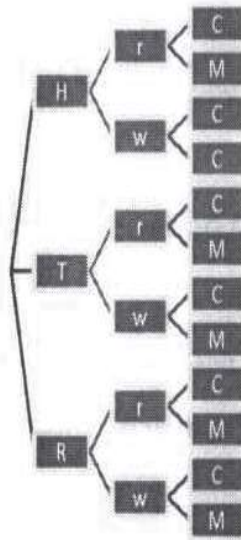
89. Inferencial y descriptiva

ANWERS UNIT 2

1. M = Milk; O = Orange juice; S = Sausage; H = Ham; B = Bacon

    a) (M,S), (M,H), (M,B), (O,S), (O,H), (O,B)

    b) 6

2. H = ham; T = Turkey; R = Roast beaf; r = rye bread; w = white bread; C = Coffee; M = Milk
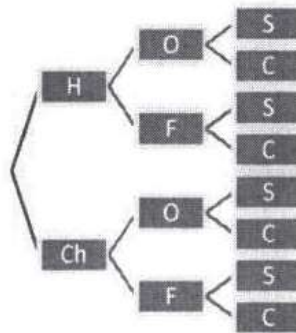

3.

    a) 120

    b) 120

    c) 21

    d) 6

e) 22

7. H = hamburger; Ch = Chiken sandwich; O = Onion rings; F = fries; S = Soda; C = Coffee
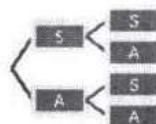

a)
b) 8

11. 120

13. 6! = 720

17. 35152

19. 504

23.

(S,S), (S,A), (A,S), (A,A)

29. 495

31. 210

37.

   a) 495
   b) 210
   c) 285

41. 244

43.

   a) $P(A) = 1/6$
   b) $P(B) = 1/4$
   c) $P(A \cap B) = 1/6$

47. 0.324

51. 0.7

53. 0.07

59. $P(E/+) = 0.102$

61.

   a) $P(D) = 0.02625$
   b) $P(H/D) = 0.95$

ANWERS UNIT 3

1. d

2. d
3. c

5. Cluster

11. 15, 25, 30, 20, 10

13.

17. 1, 0, 5

19. Cuando se selecciona una muestra de la población por el hecho de que sea accesible, con la consecuente imposibilidad de hacer afirmaciones generales con rigor estadístico sobre la población.

23. Aleatorio estratificado proporcional

29. Aleatorio simple

31. Aleatorio por conglomerados. Para el tamaño de la muestra se identifican los conglomerados, se extrae una muestra aleatoria de los mismos y se revisan todos los elementos de cada conglomerado seleccionado.

37.
  a) 6.7%
  b) 9.2%
  c) 71.6 in
  d) 68.4 in

41. 72.1%

43.
  a) 57.8 in
  b) 75.2 in

47. 181.1 pounds
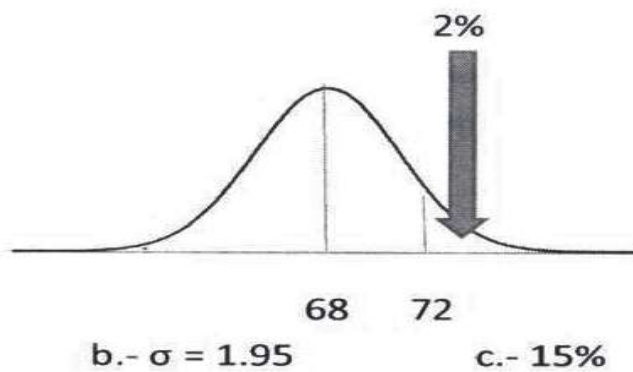
53. 3.95 lb

59.
  a) 4.75%
  b) 70.11%
  c) 24.14%

61. $\mu = 76$ kg; $\sigma = 7.14$ kg

67.

2%

68   72

b.- $\sigma = 1.95$ c.- 15%

Bibliografía

http://www.mypolyuweb.hk/machanck/lectnotes/c1_des.pdf

https://www.ahschools.us/cms/lib08/MN01909485/Centricity/Domain/4529/Anoka_Hennepin_Probability_and_Statistics-4%202013%20NEW.pdf

http://danida.vnu.edu.vn/cpis/files/Books/Fundamentals%20of%20Statistics%20I%20-%20Lecture%20Notes.pdf

https://www.pearsonhighered.com/assets/samplechapter/0/1/3/1/0131467573.pdf

http://esminfo.prenhall.com/samplechps/larson/pdfs/ch02.pdf

http://college.cengage.com/mathematics/larson/calculus_applied_life/1e/resources/app_c/1021662_App_C7.pdf

http://www.mathsisfun.com/data/standard-normal-distribution.html

https://onlinecourses.science.psu.edu/stat500/sites/onlinecourses.science.psu.edu.stat500/files/lesson06/SamplingDistributionExamples_solution.pdf

https://www3.nd.edu/~rwilliam/stats1/x21.pdf

http://www.stat.ucla.edu/~nchristo/introeconometrics/introecon_normal_dist.pdf

http://www.math.louisville.edu/~pksaho01/teaching/Math662TB-09S.pdf

https://www.saylor.org/site/textbooks/Introductory%20Statistics.pdf

http://www.fao.org/ag/humannutrition/32428-0613f516cb07eade922c8c19b4d0452c0.pdf

http://pages.stat.wisc.edu/~ifischer/Intro_Stat/Lecture_Notes/4_-_Classical_Probability_Distributions/4.4_-_Problems.pdf

https://1pdf.net/probability-exercises-anasayfa_58efb332f6065d8e55126d26

http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Probability/BS704_Probability7.html#headingtaglink_3

www.statisticshowto.com/bayes-theorem-problems/math.cornell.edu/~mec/2008

2009/TianyiZheng/Bayes.html

http://home.deib.polimi.it/capone/fcn/Ex-Probability.pdf

https://www.ini.rub.de/PEOPLE/.../Bayes-SolutionsPublic.pdf

https://courses.smp.uq.edu.au/MATH3104/.../bayes_solutions.pdf

https://www2.stat.duke.edu/courses/.../app_1.1_Sampling.pdf

http://web.thu.edu.tw/wenwei/www/Courses/statistics/ex9_3.1.htm

https://www.questionpro.com/blog/stratified-random-sampling/

http://uca.edu/psychology/files/2013/08/Ch7-Sampling-Techniques.pdf

https://www.khanacademy.org/math/ap-statistics/gathering-data-ap/sampling-methods/e/simple-random-samples

https://web.csulb.edu/~msaintg/ppa696/696sampl.htm

https://www.statcan.gc.ca/edu/power-pouvoir/ch13/prob/5214899-eng.htm