STAT 30100 Exam Jam

Contents

Chapter 1: Statistics, Data, and Statistical Thinking

1. For each of the following situation, state which type of sampling method is used.

   a. In order to assess whether there is a gender gap in their employee's salaries, Equality Inc. randomly selects 750 men and women employees and records their salaries.

   b. An author collected sample data by randomly selecting 15 pages from their book and then counts the number of words on each of those pages in order to approximate the average number per page in the entire book.

   c. An ecologist takes measurements from different locations in a river by stretching a rope across it and collecting samples at every interval of 3 feet.

   d. To learn how its employees felt about higher student fees imposed by the legislature, a university divided its employees into three categories: staff, faculty, and student employees. A random sample was selected from each group and they were telephoned and asked for their opinions.

Solution

a. Simple Random Sample

b. Cluster Sampling

c. Systematic Sampling

d. Stratified Sampling

2. Identify whether the following variables are categorical or quantitative:

   a. Corn yield in bushels
   b. Trunk circumference of a Redwood tree
   c. Writing utensil used on an exam
   d. Brand of phone
   e. Number of ounces of coffee consumed per day
   f. Calories burned during 60 minutes of exercise

Solution

a. Quantitative
b. Quantitative
c. Categorical
d. Categorical
e. Quantitative
f. Quantitative

3. Identify whether the following quantitative variables are discrete or continuous:

   a. Length of a podcast
   b. Number of hits on a website
   c. Typos on a page
   d. Amount of gasoline in a tank
   e. Students in the MAC STAT per hour
   f. Time spent in the MAC STAT

Solution

a. Continuous
b. Discrete
c. Discrete
d. Continuous
e. Discrete
f. Continuous

4. A survey by the National Center for Health Statistics was conducted last year to determine the impact of the internet on Americans' health knowledge. NHIS is a nationally representative survey of the household population of the United States conducted by the Centers for Disease Control and Preventions National Center for Health Statistics. The center used a survey of 7,192 adults aged 18 to 64 to determine the proportion of US adults that have used the internet to look up health information.

   Answer the following questions based on the situation described above.

   a. What is the population of interest?
   b. What is the sample?
   c. What is the variable of interest?
   d. What is the unit or individual?
   e. Identify whether the variable of interest is quantitative or qualitative.

Solution

a. American adults
b. 7,192 adults 18 to 64
c. Proportion of US adults that use the internet to look up health information
d. one adult (aged 18 to 64)
e. Qualitative

5. A pollster used a computer to generate 500 random numbers and then interviews voters corresponding to those numbers. Which type of sampling is this?

   a. Systematic
   b. **Simple random**
   c. Stratified random
   d. Cluster

Chapter 2: Methods for Describing Sets of Data

6. The following is a distribution of quiz scores from a stem-and-leaf plot.

Stem	Leaf
5	2
5	
6	1
6	6 6 7
7	0
7	6 7 8 9 9
8	0 2 2 4 4
8	5 5 5 5 5 8 8 9
9	0 0 1 2 3 3 3 4
9	8 8 9 9
10	0 0

a. How many observations were in the original data set?
b. How do you describe the shape of the data set?
c. What is the range of the data set?

Solution

a. 38 pieces of data
b. Skewed left
c. Range = 100 - 52 = 48

7. Suppose a data set contains five observations, $\{4, 1, 0, 2, 3\}$.

   a. Find $\sum_{i=1}^{5} x_i$.

   b. Find $\sum_{i=1}^{5} (x_i + 1)^2$.

   c. Find the mean, median, and the mode.

   d. Calculate the 90^{th} percentile and interpret the result.

   e. Calculate the standard deviation and interpret the result.
   $\left(\textit{Hint.}\ s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}. \right)$

Solution

a. $\sum_{i=1}^{5} x_i = 4 + 1 + 0 + 2 + 3 = 10$

b. $\sum_{i=1}^{5} (x_i + 1)^2 = (4 + 1)^2 + (1 + 1)^2 + (0 + 1)^2 + (2 + 1)^2 + (3 + 1)^2 = 55$

c. mean $\bar{x} = \frac{\sum x}{n} = \frac{10}{5} = 2$

d. $P_{90} : 0.90 * n = 0.90 * 5 = 4.5 \rightarrow 5^{\text{th}}$ value $\rightarrow 4$

e. $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2} = \sqrt{\frac{1}{5-1} \left((4-2)^2 + (1-2)^2 + (0-2)^2 + (2-2)^2 + (3-2)^2 \right)}$
   $s = \sqrt{\frac{10}{4}} \approx 1.58$
   The sample data deviates from the mean on average by about 1.58

8. 20 STAT 301 students were randomly sampled, and their exam scores recorded. The data is shown below:

Stem	Leaf
1	8
2	
3	
4	
5	1 4
6	6 9
7	6 7 9
8	2 3 5 7 8 8 9
9	1 4 4 8 9

a. Identify the cases (or individuals) for this study.

b. Comment on the shape of the distribution. If the distribution is skewed identify the direction.

c. Compute the five number summary for the data.

d. Calculate the IQR and interpret.

Solution

a. Scores of a STAT 301 exam

b. Skewed to the left with one possible outlier

c.

$$\begin{aligned}
Q_1 &= 0.25 * 20 = 5 \rightarrow \frac{5^{\text{th}} + 6^{\text{th}}}{2} = \frac{69+76}{2} = 72.5 \\
Q_2 &= 0.5 * 20 = 10 \rightarrow \frac{10^{\text{th}} + 11^{\text{th}}}{2} = \frac{83+85}{2} = 84 \\
Q_3 &= 0.75 * 20 = 15 \rightarrow \frac{15^{\text{th}} + 16^{\text{th}}}{2} = \frac{89+91}{2} = 90
\end{aligned}$$

18	69	84	90	99

d. $IQR = Q_3 - Q_1 = 90 - 72.5 = 17.5$

The range of the middle 50% of the data is 21

9. Which of the following graphs are BEST for each situation?

   Pie Chart Histogram Stem-and-Leaf Plot Pareto Graph

   a. Which cheese is most preferred on a burger.
   b. Sleep (in number of hours) data collected from 18 new parents.
   c. Distance traveled to get to class for 1,300 IUPUI students.
   d. Grade distribution before the Final exam.

Solution

a. Pie
b. Stem-and-Leaf
c. Histogram
d. Pareto

10. Many firms use on-the-job training to teach their employees computer programming. Suppose you work in the personnel department of a firm that just finished training a group of its employees to program, and you have been requested to review the performance of one of the trainees on the final test that was given to all trainees. The mean of the test scores is 78. Additional information indicated that the median of the test scores was found to be about 90. What type of distribution most likely describes the shape of the test scores?

    a. Symmetric.
    b. Skewed to the right.
    c. $\boxed{\textbf{Skewed to the left.}}$
    d. Unable to determine from the given information.

Chapter 3: Probability

11. Four hundred accidents that occurred on a Saturday night were analyzed. The number of vehicles involved and whether alcohol played a role in the accident were recorded. The results are shown below.

Did Alcohol Play a Role?	Number of Vehicles Involved			Total
	1	2	≥ 3	
Yes	53	100	17	170
No	24	176	30	230
Total	77	276	47	400

a. What is the probability that the number of vehicles involved in an accident is exactly two?

b. What is the probability that alcohol played a role in an accident?

c. What is the probability that the number of vehicles involved in an accident was greater than or equal to three when alcohol played a role?

d. Given that an accident involved three or more vehicles, what is the probability that alcohol was involved?

Solution

a. $\dfrac{276}{400} = 0.69$

b. $\dfrac{170}{400} = 0.425$

c. $P(\geq 3|\text{Yes}) = \dfrac{17}{170} = 0.10$

d. $P(\text{Yes}|\geq 3) = \dfrac{17}{47} \approx 0.362$

12. The outcome of an experiment is the number of resulting heads when a nickel and dime are flipped simultaneously. What is the sample space for this experiment?

    a. $\boxed{\{0, 1, 2\}}$
    b. $\{HH, HT, TT\}$
    c. $\{HH, HT, TH, TT\}$
    d. {nickel, dime}

13. The table below shows the political affiliation of 1000 randomly selected American voters and their positions on the school of choice program.

	Poltical Party		
Position	Democrat	Republican	Other
Favor	260	120	240
Oppose	40	240	100

Let the events D =Democratic voter, R =Republican voter, T =Other voter, F indicate that the voter favors the school of choice program, and O indicate that the voter opposes the school of choice program.

a. What is the probability that a randomly selected voter favors the school of choice program?
b. What is the probability that a randomly selected voter supports neither Democrats or Republicans?
c. Given that a randomly selected voter supports the “Other” party, what is the probability that they oppose the school of choice program?
d. Are the events D and O independent? Show your work.

Solution

a. $\frac{620}{1000} = 0.62$
b. $\frac{340}{1000} = 0.34$
c. $P(\text{Oppose}|\text{Other}) = \frac{100}{340} = 0.29$
d. $P(A \cap B) = P(A)P(B) \rightarrow \text{Independent}$
   $P(D \cap O) = \frac{40}{1000} = 0.04$
   $P(D)P(O) = \frac{300}{1000}\frac{380}{1000} = 0.114$

   $0.4 \neq 0.114 \rightarrow \text{Not Independent}$

Chapter 4: Random Variables

14. Consider the given discrete probability distribution.

x	1	2	3	4	5
$p[x]$	0.1	0.2	0.2	0.3	0.2

a. Find $E[X]$ and interpret the result.
b. Calculate $P(2 \leq X \leq 4)$.
c. Construct a graph for $p[X]$.
d. Calculate $Var[X]$.
Hint: $Var[X] = E[X^2] - (E[X])^2$

Solution

a. $E[X] = \sum xp[x] = 1(0.1) + 2(0.2) + 3(0.2) + 4(0.3) + 5(0.2) = 3.3$
The average or expected value that X takes is 3.3
b. $P(2 \leq X \leq 4) = P[X = 2] + P[X = 3] + P[X = 4] = 0.2 + 0.2 + 0.3 = 0.7$
c. stuff and things
d.

$$\begin{aligned} E[X^2] &= 1^2(0.1) + 2^2(0.2) + 3^2(0.2) + 4^2(0.3) + 5^2(0.2) = 12.5 \\ Var[X] &= 12.5 - 3.3^2 = 1.61 \end{aligned}$$

15. Consider the given discrete probability distribution.

x	3	5	7	9
$p[x]$	0.24	0.4	0.26	?

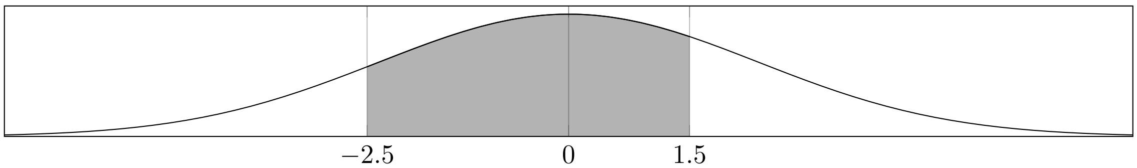
Find the probability that X exceeds 3.

a. 0.27
b. $\boxed{\mathbf{0.76}}$
c. 0.73
d. 0.49

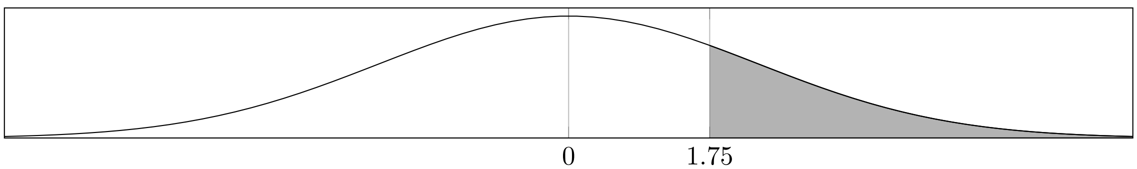
Chapter 4: Normal Distribution

16. Shade the appropriate area for each problem and then use the standard normal distribution to find the probability.

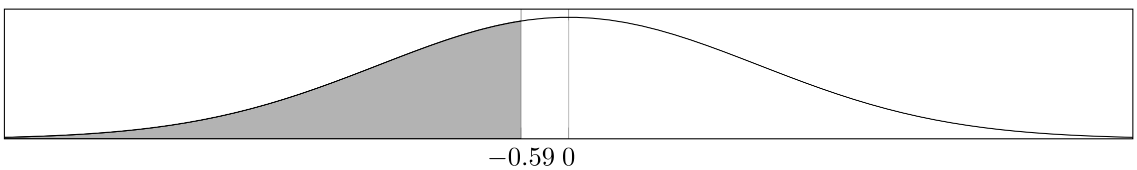
    a. $P(-2.5 < Z < 1.5)$.



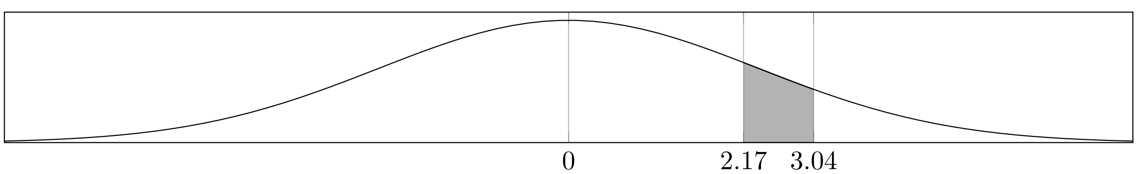
    b. $P(Z > 1.75)$.



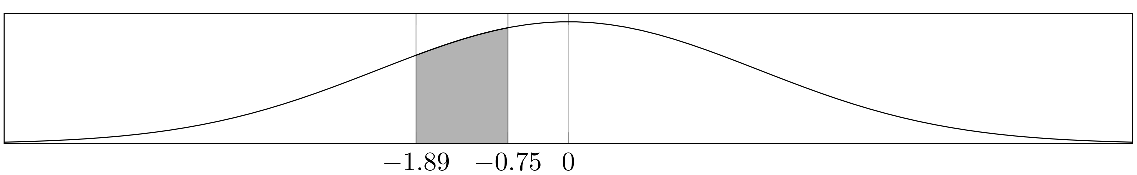
    c. $P(Z < -0.59)$.



    d. $P(2.17 < Z < 3.04)$



    e. $P(-1.89 < Z < -0.75)$.



17. What geometric shape is used to represent areas for a normal distribution?

    a. Rectangle.

    b. Curve.

    c. $\boxed{\textbf{Bell Curve.}}$

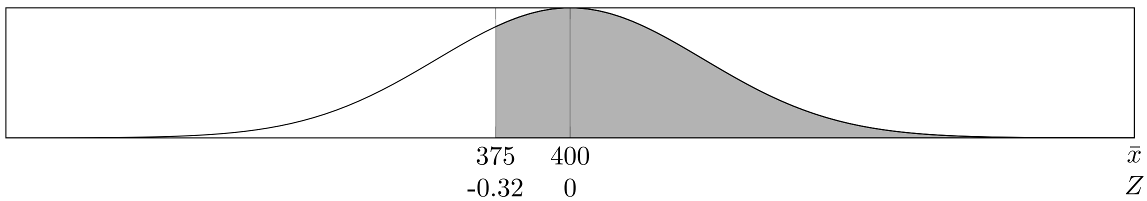
    d. Triangle.

18. Use the standard normal distribution to find $P(Z < -2.33 \text{ or } Z > 2.33)$.

    a. 0.9809
    b. $\boxed{\mathbf{0.0198}}$
    c. 0.7888
    d. 0.0606

19. The distribution of the amount of money undergraduate students spend on books for a term is normally distributed with a mean $\mu = \$400$ and a standard deviation of $\sigma = \$78$. If a student is selected at random what is the approximate probability that this student spends more than \$375 on books?

   a. Label the variable.
   b. Formulate the problem.
   c. Shade the appropriate area under the curve based on part (b).



   d. Calculate the Z-score.
   e. Show your Z-score using the Z-scale on the above graph.
   f. Use a Z-table to find the final answer.

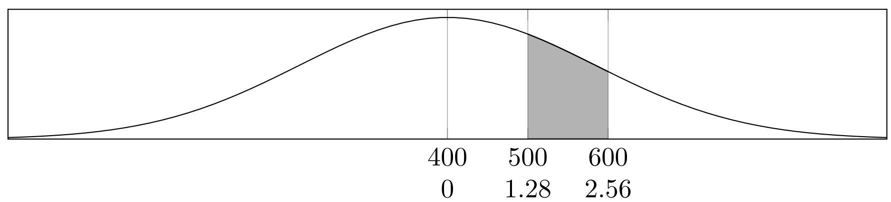
Solution

   a. Let x be the amount of money undergrads spend on books.
   b. $P(x > 375)$
   d. $Z = \dfrac{x - \mu}{\sigma} \dfrac{375 - 400}{78} = -0.32$
   f. $P(x > 375) = P(Z > -0.32) = 0.6255$

20. The distribution of the amount of money undergraduate students spend on books for a term is normally distributed with a mean $\mu = \$400$ and a standard deviation of $\sigma = \$78$. If a student is selected at random, what is the approximate probability that this student spends between \$500 and \$600 on books?

   a. Formulate the problem.

   b. Shade the appropriate area under the curve based on part (a).



   c. Calculate the Z-score.

   d. Show your Z-score using the Z-scale on the above graph.

   e. Use a Z-table to find the final answer.

Solution

a. Let x be the amount of money undergrads spend on books.

   $P(x > 375)$

c. $Z = \dfrac{x - \mu}{\sigma} = \dfrac{375 - 400}{78} = -0.32$

e. $P(x > 375) = P(Z > -0.32) = 0.6255$

Chapter 4: Sampling Distributions

21. The Central Limit Theorem is considered powerful in statistics because

    a. **it works for any population distribution provided the sample size is sufficiently large.**

    b. it works for any sample size provided the population is normal.

    c. it works for any population distribution provided the population mean is known.

    d. it works for any sample provided the population distribution is known.

22. Suppose a random sample of $n = 64$ measurements is selected from a population with mean $\mu = 65$ and standard deviation $\sigma = 12$. Find the Z-score corresponding to a value of $\overline{x} = 68$.

    a. 0.25

    b. -0.25

    c. $\boxed{\mathbf{2}}$

    $$Z = \frac{x - \mu}{\sigma} = \frac{68 - 65}{12/\sqrt{64}} = 2$$

    d. -2

23. The daily revenue at a university snack bar has been recorded for the past five years. Records indicate that the mean daily revenue is \$3500 and the standard deviation is \$550. The distribution is skewed to the right due to several high volume days. Suppose that 100 days are randomly selected and the average daily revenue computed. Which of the following describes the sampling distribution of the sample mean?

   a. **Normally distributed with a mean of \$3500 and a standard deviation of \$55.**
   b. Normally distributed with a mean of \$3500 and a standard deviation of \$550.
   c. Normally distributed with a mean of \$350 and a standard deviation of \$55.
   d. Skewed to the right with a mean of \$3500 and a standard deviation of \$550.

24. The number of cars running a red light in a day at a given intersection is given by a distribution with a mean of 2.7 cars and a standard deviation of 4. The number of cars running the red light was observed on 100 randomly chosen days and the mean number of cars calculated.

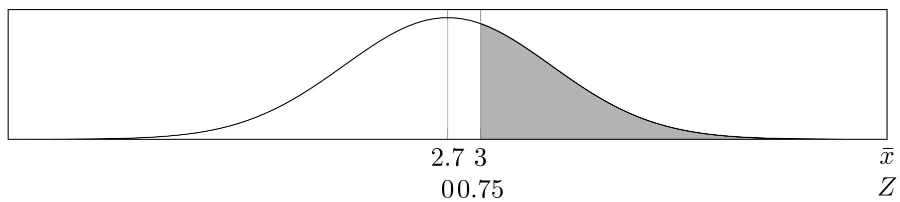
    a. What is the mean of the sampling distribution of the sample mean of the number of cars running the red light?

    b. What is the standard deviation of the sampling distribution of the sample mean of the number of cars running the red light?

    c. What is the distribution of the sampling distribution of the sample mean of the number of cars running the red light?

    d. Find the probability that the sample mean of the number of cars running the red light is greater than three by completing the following steps.

       1. Shade the appropriate area under the curve based on part (d).



       2. Calculate the Z-score.
       3. Show your Z-score value by using the Z-scale on the above graph.
       4. Use a Z-table to find the final answer.

Solution

a. The same as the population mean, 2.7

b. $s = \dfrac{\sigma}{\sqrt{100}} = \dfrac{4}{\sqrt{100}} = 0.4$

c. Normally distributed ($n > 30$)

d. $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{3.5 - 2.7}{0.4} = 0.75$
   $P(\bar{x} > 3) = P(Z > 0.75) = 0.2266$

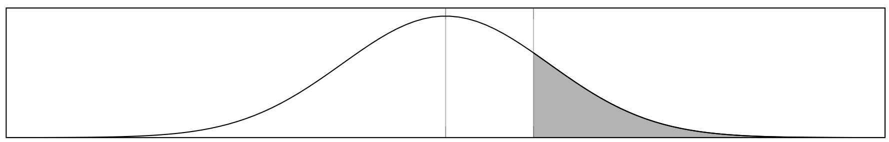
25. The weight of corn chips dispensed into a 17 ounce bag has been identified as possessing a normal distribution with a mean of 17.5 ounces and a standard deviation of 0.5. Suppose that 100 bags of chips are selected at random, find the probability that the mean weight of the bags exceeds 17.6 ounces.

    a. What is the mean of the sampling distribution of the sample mean ounces?

    b. What is the standard deviation of the sampling distribution of the sample mean ounces?

    c. Find $P(\bar{x} > 17.6)$.

    1. Shade the appropriate area under the curve based on part (b).



    2. Calculate the Z-score.
    3. Show your Z-score using the Z-scale on the above graph.
    4. Use a Z-table to find the final answer.

Solution

a. The same mean as the population, 17.5

b. $\dfrac{\sigma}{\sqrt{n}} = \dfrac{0.5}{\sqrt{100}} = 0.05$

c. $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{17.6 - 17.5}{0.05} = 2$

$P(\bar{x} > 17.6) = P(Z > 2) = 0.0228$

Chapter 5: Estimation with Confidence Intervals

26. What is $Z_{\alpha/2}$ when $\alpha = 0.01$?

    a. 1.96
    b. $\boxed{\textbf{2.575}}$
    c. 1.645
    d. 2.33

27. Find the value of t_0 such that $P(-t_0 \leq t \leq t_0) = 0.95$ where $df = 15$.

    a. $\boxed{\mathbf{2.131}}$
    b. 1.753
    c. 2.602
    d. 2.947

28. Suppose a large labor union wishes to estimate the mean number of hours per month a union member is absent from work. The union decides to sample 358 of its members at random and monitor the working time of each of them for 1 month. At the end of the month, the total number of hours absent from work is recorded for each employee. Which of the following should be used to estimate the parameter of interest for this problem?

    a. A small sample confidence interval for μ.

    b. **A large sample confidence interval for μ.**

    c. A large sample confidence interval for p.

    d. A small sample confidence interval for p.

29. The director of a hospital wishes to estimate the mean number of people who are admitted to the emergency room during a 24 hour period. The director randomly selects 81 different 24 hour periods and determines the number of admissions for each. For the sample, $\overline{x} = 16.6$ and $s^2 = 25$. Estimate the mean number of admissions per 24 hour period with a 95% confidence interval.

    a. Label the parameter.
    b. Verify conditions.
    c. Interval calculations.
    d. Interpret your results in the context of the problem.

Solution

a. Let x be the number of people admitted to an emergency room in a 24 hour period.
b. $n \geq 30$, Random Sample $\rightarrow$ Approximately Normal
c. $\bar{x} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}}$
   $16.6 \pm 1.96 \frac{5}{\sqrt{81}}$
   $(15.51, 17.64)$
d. We are 95% confident that the true population mean is contained in the interval $(15.51, 17.64)$

30. A random sample of 50 employees of a large company were asked if they participated in the company's stock purchase plan. The answers are shown below.

yes	no	no	yes	no	no	yes	yes	no	no
no	yes	yes	yes	no	yes	no	no	yes	yes
no	yes	yes	no	yes	yes	no	yes	yes	yes
yes	no	no	yes	yes	yes	yes	yes	no	yes
no	yes	yes	no	yes	yes	yes	yes	yes	yes

Use a 90% confidence interval to estimate the proportion of employees who participate in the company's stock purchase plan by completing the following steps.

a. Label the parameter.
b. Verify the conditions.
c. Interval calculations.
d. Interpret your result in the context of the problem.

Solution

a. Let p be the proportion of people that participate in the company's stock purchase program $\hat{p} = \frac{32}{50} = 0.64$
b. Random Sample, $n\hat{p} \geq 15 \rightarrow 50 * .64 \geq 15$, $n(1-\hat{p}) \geq 15 \rightarrow 50 * 0.36 \geq 15$
c. $\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
$0.66 \pm 1.645\sqrt{\frac{0.64(1-0.64)}{50}}$
$(0.5483, 0.7717)$
d. We are 90% confident that the true proportion falls in the interval $(0.5498, 0.7702)$

31. We intend to estimate the average driving time of Chicago commuters. From a previous study, we believe that the average time is 42 minutes with a standard deviation of 6 minutes. We want our 99% confidence interval to have a margin of error of no more than ± 2 minutes. What is the smallest sample size that we should consider?

Solution

$$n = \left(\frac{Z_{\alpha/2}\sigma}{MOE}\right)^2$$
$$n = \left(\frac{2.575 * 6}{2}\right)^2$$
$$n = 59.67 \rightarrow 60$$

32. What is the best explanation of the use of the term "95% confidence?"

    a. We can never be 100% confidence in statistics, we can only be 95% confident.
    b. The sample statistic will fall in 95% of the confidence intervals we construct.
    c. **The population parameter will fall in 95% of the confidence intervals we construct.**
    d. We are confident that 95% of all samples will give us the same intervals.

33. What is the confidence level of the following confidence interval for μ?

$$\overline{x} \pm 0.95 \cdot \frac{\sigma}{\sqrt{n}}$$

a. 80%
b. 95%
c. $\boxed{\textbf{66\%}}$
d. 5%

34. A university dean is interested in determining the proportion of students who receive some sort of financial aid. Rather than examine the records of all students, the dean randomly selects 200 students and finds that 118 of them are receiving financial aid. If the dean wanted to estimate the proportion of all students receiving financial aid to within 1% with 98% confidence, how many students would need to be sampled?

    a. $\boxed{\text{n=13133}}$
    b. $n = 3177$
    c. $n = 5637$
    d. $n = 132$

35. How much money does the average professional football fan spend on food at a single football game? That question was posed to ten randomly selected football fans. The sample results provided a sample mean and standard deviation of \$16.00 and \$3.35, respectively. Use this information to construct a 99% confidence interval for the mean.

    a. Label the parameter.
    b. Verify the conditions.
    c. Interval calculations.
    d. Interpret the result from part (c).

Solution

a. Let μ be the population average of how much football fans spend on food at a single game.
b. The population standard deviation σ is unknown, Random Sample, $n < 30 \to t$
c. $\bar{x} \pm t_{\alpha/2}\dfrac{s}{\sqrt{n}}$
   $16 \pm 3.250\dfrac{3.35}{\sqrt{10}}$
   $(12.577, 19.443)$
d. We are 99% confident that the population mean amount of money spent by a football fan at a single game is between \$13.27 and \$18.73.

Chapter 6: Test of Hypotheses

36. The owner of Get Away Travel has recently surveyed a random sample of 381 customers to determine whether mean age of the agency's customers is over 25. The appropriate hypotheses are $H_0 : \mu = 25$, $H_A : \mu > 25$. If he concludes that the mean age is over 25 when it is not, then he makes a ______ error. If he concludes that the mean age is not over 25 when it is, then he makes a ______ error.

    a. $\boxed{\textbf{Type 1, Type 2.}}$
    b. Type 2, Type 2.
    c. Type 1, Type 1.
    d. Type 2, Type 1.

37. A 90% confidence interval for p is given as $(0.59, 0.81)$. How large was the sample used to construct this interval?

Solution

$$\begin{aligned}
\hat{p} &= \frac{0.59 + 0.81}{2} = 0.70 \\
90\% \rightarrow Z_{\alpha/2} &= 1.645 \\
MOE &= 0.81 - 0.70 = 0.11 \\
n &= \left(\frac{Z_{\alpha/2}}{MOE}\right)^2 \hat{p}(1-\hat{p}) \\
n &= \left(\frac{1.645}{0.11}\right)^2 (0.70)(0.30) \\
n &= 46.86 \rightarrow 47
\end{aligned}$$

38. A survey was conducted of 258 drivers to assess their attitudes toward speeding. The data recorded included the response to "Are you sure you can resist your friends' persuasion to drive faster?" The average recorded was 4.98 with a standard deviation of 1.62. Suppose it is known that the true mean response for students is $\mu = 4.7$. Test whether the student response is higher than claimed. Use $\alpha = 0.01$.

    a. Hypothesis:
        1. Label the parameter.
        2. Null Hypothesis.
        3. Alternative hypothesis.
    b. Test Statistic:
        1. Write the name of the test you will be using.
        2. Verify the required conditions for your proposed test.
    c. Level of significance.
    d. Rejection region.
    e. Calculation of test statistic.
    f. Decision.
    g. Conclusion based on context of the problem.

Solution

a. Let x be the mean response for students who answer the question "Are you sure can resist your friend's persuasion to drive faster?"

   $H_0 : \mu = 4.7$
   $H_1 : \mu > 4.7$

b. A random sample of $n = 258 > 30 \rightarrow Z$ test

c. $\alpha = 0.01$

d. $RR = \{Z > 2.33\}$

e. $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{4.98 - 4.7}{1.62/\sqrt{258}} = 2.776$

f. $2.776 > 2.33 \rightarrow$ Reject H_0

g. Based on the $\alpha = 0.01$ level of significance, there is sufficient evidence to say that the student response is higher than claimed.

39. A study of 15 adults assessed whether adults have dental anxiety when they go to the dentist. On a score from 0 = none to 25 (extreme), the average recorded was 10.7 with a standard deviation of 3.6. Conduct a hypothesis test to assess whether the anxiety score is different from $\mu = 11$. Use $\alpha = 0.05$.

   a. Hypothesis:
      1. Label the parameter.
      2. Null Hypothesis.
      3. Alternative hypothesis.
   b. Test Statistic:
      1. Write the name of the test you will be using.
      2. Verify the required conditions for your proposed test.
   c. Level of significance.
   d. Rejection region.
   e. Calculation of test statistic.
   f. Calculate the p-value.
   g. Make a decision using RR and then using the p-value.
   h. Conclusion based on context of the problem.

Solution

a. Let x be the mean anxiety score of adults when they go to the dentist.

   $H_0 : \mu = 11$
   $H_1 : \mu \neq 11$

b. Random sample of $n = 15 < 30 \rightarrow t$ test

c. $\alpha = 0.05$

d. $RR = \{t < -2.145 \text{or} t > 2.145\}$

e. $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{10.7 - 11}{3.6/\sqrt{15}} = -0.323$

f. p-value ¿ 0.100

g. The test statistic is not in the rejection region and the p-value is greater than α so we cannot reject the null hypothesis.

h. There is insufficient evidence to suggest that at the 95% level of confidence that that the mean anxiety score differs from 11.

40. A national organization has been working with utilities throughout the nation to find sites for large wind turbines that generate electricity. Wind speeds must average more than 22 miles per hour (mph) for a site to be acceptable. Recently, the organization conducted wind speed tests at a particular site. Based on a sample of $n = 33$ wind speed recordings (taken at random intervals), the wind speed at the site averaged $\overline{x} = 22.8$ mph, with a standard deviation of $s = 4.3$ mph. To determine whether the site meets the organization's requirements consider the test $H_0: \ \mu = 22$ versus $H_a: \ \mu > 22$, where μ is the true mean wind speed at the site and $\alpha = 0.01$. Suppose the value of the test statistic was computed to be 1.07. State the conclusion.

    a. **At $\alpha = 0.01$ there is insufficient evidence to conclude that the true mean wind speed at the side exceeds 22 mph.**

    b. At $\alpha = 0.01$ there is sufficient evidence to conclude the true mean wind speed exceeds 22 mph.

    c. At $\alpha = 0.01$ there is sufficient evidence to conclude the true mean wind speed at the site is equal to 22 mph.

    d. None.

41. What is the p-value for a test where $H_0: \mu = 3$ and $H_a: \mu \neq 3$ with a test statistic $Z = 1.3774$?

   a. $\boxed{\textbf{0.1676}}$
   b. 0.0838
   c. 0.9162
   d. 0.8324

42. Given $H_0: \mu = 25$, $H_a: \mu \neq 25$ and p-value=0.034. To reject H_0 what α do you need?

   a. 1%
   b. 2%
   c. 3%
   d. $\boxed{\textbf{4\%}}$

43. The article "Theaters Losing Out to Living Rooms" (San Luis Obispo Tribune, June 17, 2005) states that movie attendance declined in 2005. The Associated Press found that 730 of 1000 randomly selected adult Americans preferred to watch movies at home rather than at a movie theater. Is there convincing evidence that more than half of adult Americans prefer to watch movies at home? Test the relevant hypotheses using $\alpha = 0.05$.

   a. Hypothesis.
      1. Label the parameter.
      2. Null hypothesis.
      3. Alternative hypothesis.
   b. Test statistic.
      1. Name the test you are going to use.
      2. Verify the required conditions for your proposed test.
   c. Calculation of the proposed test.
   d. Calculation of the p-value.
   e. Level of significance.
   f. Decision and conclusion based on the context of the problem.

Solution

a. Let p be the proportion of Americans that prefer to watch movies at home rather than at the theater

   $H_0 : p = 0.50$
   $H_1 : p > 0.50$

b. A random sample where $n\hat{p} = 500 > 15$ and $n(1-\hat{p}) = 500 > 15$, thus we can use a Z test

c. $Z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}} = \dfrac{0.73 - 0.5}{\sqrt{\dfrac{(0.5).5}{1000}}} = \dfrac{0.23}{0.0158} = 14.5465$

d. p-value is essentially 0 as 14 is far off the chart

e. $\alpha = 0.05$

f. $p-value < 0.05 \rightarrow$ Reject the null hypothesis, which means that the true proportion of Americans who prefer to watch movies at home is greater than 0.5

Chapter 7: Inferences Based on Two Samples

44. Samples of 200 are taken from two independent populations, and give the following results:

$$\bar{x}_1 = 5275 \qquad \bar{x}_2 = 5200$$
$$s_1 = 150 \qquad s_2 = 200$$

a. Use a 95% CI to estimate the difference between the population means $(\mu_1 - \mu_2)$

b. Test the null hypothesis $\mu_1 - \mu_2 = 0$ against $\mu_1 - \mu_2 \neq 0$

State the hypothesis
Test Statistic
Rejection Region
Decision/Conclusion

Solution

a.

$$\bar{x}_1 - \bar{x}_2 \pm Z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
$$5275 - 5200 \pm 1.96\sqrt{\frac{150^2}{200} + \frac{200^2}{200}}$$
$$75 \pm 1.96\sqrt{312.5}$$
$$(40.35, 109.65)$$

b.

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$
$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$
$$Z = \frac{75}{\sqrt{312.5}} = 4.24 RR = \{Z < -1.96 \text{or} Z > 1.96\} 4.24 > 1.96 \rightarrow Reject H_0$$

There is sufficient evidence to conclude at the $\alpha = 0.05$ level of significance to say that the two populations have different means.

45. Using on-line homework programs is an efficient way to deliver and grade the assignments. It is claimed that the students' grades are similar whether they do the homework from a textbook versus on-line. A study was conducted to test this claim, 35 students completed their homework from the textbook and 40 students used the on-line software. Their grades were recorded and is shown below:

$$\begin{array}{ll} n_1 = 35 & n_2 = 40 \\ \bar{x}_1 = 89.6 & \bar{x}_2 = 91.2 \\ s_1 = 4.75 & s_2 = 3.62 \end{array}$$

Test the claim at $\alpha = 0.05$.

a. Hypothesis.
   1. Label the parameter.
   2. Null hypothesis.
   3. Alternative hypothesis.

b. Test statistic.
   1. Name the test you are going to use.
   2. Verify the required conditions for your proposed test.

c. Calculation of the proposed test.

d. Calculation of the p-value.

e. Level of significance.

f. Decision and conclusion based on the context of the problem.

Solution

a. Let $\mu_1 - \mu_2$ be the population difference between homework grades from the group of students who use textbooks versus those that used the on-line software.

   $H_0 : \mu_1 - \mu_2 = 0$
   $H_1 : \mu_1 - \mu_2 \neq 0$

b. Both independent and random samples are large $(n > 30) \rightarrow Z$ test

c. $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{89.6 - 91.2}{\sqrt{\dfrac{4.75^2}{35} + \dfrac{3.62^2}{40}}} = \dfrac{-1.6}{0.9860} = -1.62$

d. $p - value = P(Z < -1.62 \cup Z > 1.62) = 2(0.0537) = 0.1074$

e. $\alpha = 0.05$

f. $p - value = 0.1074 > 0.05 \rightarrow$ Fail to Reject. So, there is insufficient evidence to state that the population means differ from each other.

46. A summer weight loss camp claims to revolutionize a child's nutrition education and their physical activity. We want to test their claim based on the campers' BMI. We randomly selected 76 campers and measured their BMI at the beginning of camp and at the end. The data is shown below:

	Mean	s
Start	34.9	6.9
End	31.6	6.2
Difference	3.3	1.5

a. Use a 90% Confidence Interval to estimate the mean BMI change.

b. Test the hypothesis that the campers have lower BMI after the camp ends (Be sure to note the direction the difference was taken) at $\alpha = 0.10$.

i. Label the parameters

ii. Hypothesis

iii. Calculate the Test Statistic

iv. Identify the level of significance

v. Find the rejection region

vi. Find the p-value

vii. Decision

Based on the Rejection Region

Based on the p-value

viii. Conclusion in context

Solution

a. $\bar{x}_d \pm Z_{\alpha/2}\dfrac{s}{\sqrt{n}}$

$3.3 \pm 1.645\dfrac{1.5}{\sqrt{76}} \rightarrow (3.02, 3.58)$

b. Let μ_d be the population average of the difference in BMI before and after the camp.

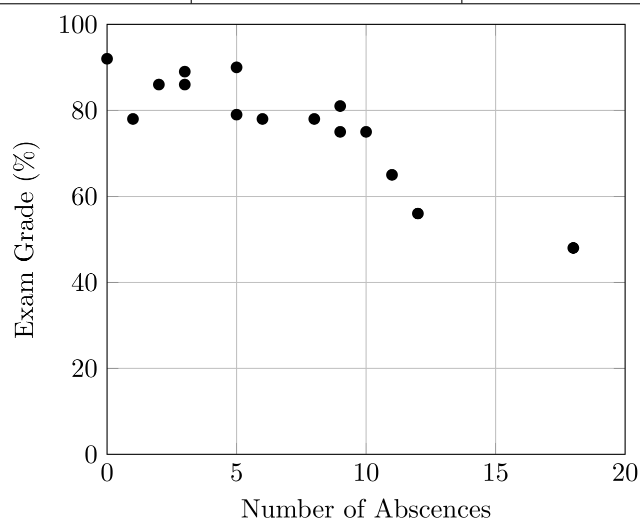
$$H_0 : \mu_d = 0$$
$$H_1 : \mu_1 > 0$$
$$Z = \frac{\bar{x}_d - \mu_d}{s_d/\sqrt{n}} = \frac{3.3 - 0}{1.5/\sqrt{75}}$$
$$Z = 19.05\alpha = 0.10 \rightarrow RR = \{Z > 1.28\}$$
$$p - value = P(Z > 19.05) \approx 0$$

Since $19.05 > 1.28$ and our $p - value < \alpha = 0.10$, we can reject the null hypothesis which implies that there is sufficient evidence to conclude that the mean difference in BMI before and after the camp was significantly greater than 0.

Chapter 9: Simple Linear Regression

47. The data below represents the number of absences and the final grades of fifteen randomly selected students from a statistics class. Describe the trend and strength of the data set.

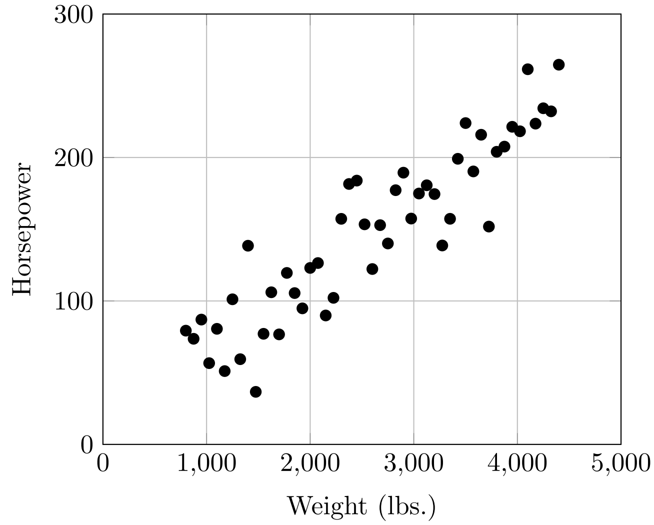
Student Number	Number of Absences	Final Grade (%)
1	5	79
2	6	78
3	2	86
4	12	56
5	9	75
6	5	90
7	8	78
8	18	48
9	0	92
10	1	78
11	9	81
12	3	86
13	10	75
14	3	89
15	11	65



Solution

There is a decently strong negative trend seen above

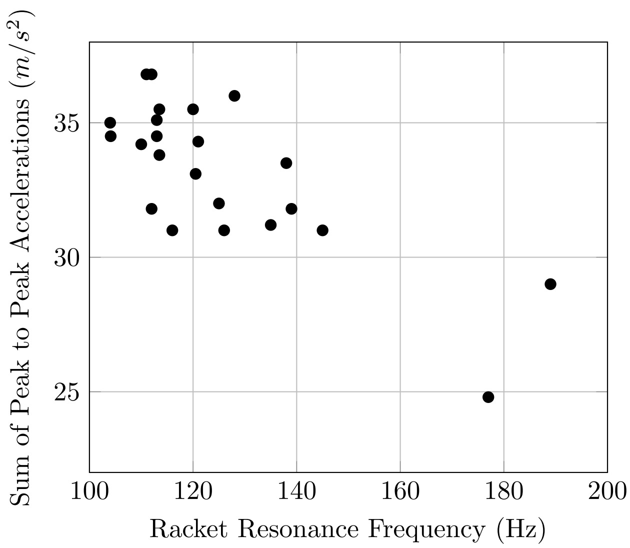
48. Consider the following scatter-plot of the weight of several models of cars (in pounds) versus their horsepower.



A plausible value for the correlation between weight and horsepower is

a. +0.2
b. -0.9
c. $\boxed{\mathbf{+0.8}}$
d. -1.0

49. One factor in the development of tennis elbow, a malady that strikes fear into the hearts of all serious players of that sport, is the impact-induced vibration of the racket-and-arm system at ball contact. It is well known that the likelihood of getting tennis elbow depends on various properties of the racket used. Consider the accompanying scatter plot racket resonance frequency (Hertz) versus the sum of peak-to-peak accelerations (a characteristic of arm vibrations, in meters per second squared) for $n = 23$ different rackets (Transfer of Tennis Racket Vibrations into the Human Forearm, Medicine and Science in Sports Exercise [1992-] 1134-1140).



a. Identify the explanatory and response variables.
b. Describe the overall shape, trend and strength of the scatterplot.
c. Suppose the equation of the least-squares regression line is $y = 40.5 - 0.537x$. Interpret the slope in the context of this problem. Also interpret the intercept.
d. What would you predict the sum of peak-to-peak accelerations to be when the racket resonance frequency is 165 Hz?
e. Given $r = -0.7723$, calculate the coefficient of determination and interpret it in terms of the problem.

Solution

a. The response variable is the sum of peak to peak accelerations (m/s^2) and the explanatory variable that we hope to use to predict the response is the racket resonance frequency (Hz)
b. The general trend is a moderately strong negative line.
c. For every $1Hz$ increase in frequency, the expected sum of peak to peak accelerations should decrease by about $0.537m/s^2$.
   If the resonance was at $0Hz$, then the expected value of the sum of peak to peak accelerations would be $40.5m/s^2$.
d. $\hat{y} = 40.5 - 0.537(165) = -48.105m/s^2$
e. $R^2 = (-0.7723)^2 = 0.5964 \rightarrow$ which implies that 59.64% of the deviations in y is explained by x