STAT 250 PRACTICE PROBLEMS MIDTERM 1

PECK CHAPTERS 1-4,6

1. Which of the following statistical studies is an observational study?

|  |  |
| --- | --- |
| a. | A group of students is surveyed to determine the percentage who believe in reincarnation. |
| b. | A random sample of students is shown a documentary on the evolution of reincarnation beliefs throughout human history. A second random sample is not shown the documentary. The object of the study is to determine whether viewing the documentary affects belief in reincarnation. |
| c. | To determine whether Vitamin C has any effectiveness at lessening the duration of the common cold, one set of cold sufferers is given no Vitamin C, a second set is given 2 grams per day, and a third set is given 5 grams of Vitamin C per day. |
| d. | All of these studies are observational. |
| e. | None of these studies is observational. |

1. Select which sampling method is described by the following situation.

To analyze the level of employee satisfaction at stores in a certain large city, a grocery store chain selects eight store locations at random and surveys all employees at these locations.

|  |  |
| --- | --- |
| a. | Simple random sampling |
| b. | Stratified random sampling |
| c. | Cluster sampling |
| d. | Systematic sampling |
| e. | Convenience sampling |

1. A researcher wishes to study the relationship between the level of background noise and mental concentration. The treatment (noise level) will have three levels: no noise, low-intensity noise, and high-intensity noise. the subjects are to be divided into three groups, and each group is to receive one of the treatments. He has available to him a set of 60 female volunteers and a set of 90 male volunteers. What experimental design strategy would help him eliminate the introduction of gender as a confounding variable?

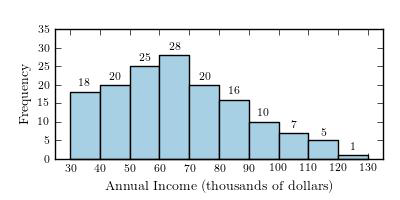
|  |  |
| --- | --- |
| a. | stratified sampling |
| b. | replication |
| c. | blocking |
| d. | systematic sampling |
| e. | double-blind trials |

1. **Patients in a nursing home were given tests of mental agility. For a year, half of the patients were randomly assigned to spend 3 hours a week working on puzzles. After a year, all patients were re-evaluated.. If the puzzle-solving patients showed a better change in mental ability than the non puzzle-solving patients, can we say that puzzle-solving causes better mental ability scores for nursing home patients?** 
   1. **No, we cannot determine causality from an observational study**
   2. **Yes, this was a well-designed experiment**
   3. **No, because the patients were not blind to the treatment**
   4. **Yes, because measurement bias is not present**
2. A text-in poll of opinion on a political issue is conducted during a news show. Which type of bias are the results MOST likely to suffer from?
   1. Selection
   2. Response
   3. Measurement
   4. Nonresponse
3. 60 volunteers for an experiment are to be assigned to two treatment groups. Which of the following is an appropriate way to assign the volunteers to the two groups?
4. Put all the names in a hat. The first 30 drawn out are assigned to group 1
5. List the volunteers. Beginning at the top and proceeding until a group reaches 30, flip a coin for each volunteer and assign them to group 1 if the coin lands on heads and group 2 if it lands on tails. The remaining volunteers get assigned to the group that did not fill first.
   1. I only
   2. II only
   3. I and II
   4. Neither I nor II
6. Which of the following variables from a data set about students is a discrete numeric variable?
   1. Distance of residence to campus
   2. Major Code
   3. Gender
   4. Number of courses enrolled in
7. Which of the following types of graphical display would be appropriate for a sample of student heights?
8. Pie chart
9. Scatterplot
10. Stemplot
11. All of the above since we are dealing with numbers.
12. Which of the following stem and leaf plots corresponds a data distribution that is positively (right) skewed?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| I. |  | II. |  | III. |  |

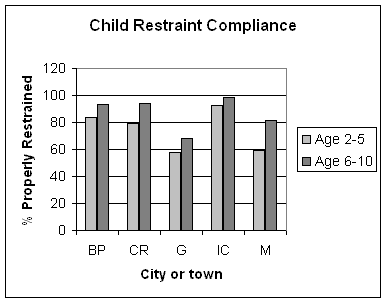
|  |  |
| --- | --- |
| a. | I only |
| b. | II only |
| c. | III only |
| d. | I and III only |
| e. | None of the distributions is positively skewed. |

1. The annual household incomes, in thousands of dollars, for 150 households in a small community are summarized in the histogram below.



What proportion of households have annual incomes of $90,000 or more?

|  |  |
| --- | --- |
| a. | 0.85 |
| b. | 0.67 |
| c. | 0.08 |
| d. | 0.15 |
| e. | 0.11 |

1. A statewide survey released by the University of Iowa Injury Prevention Research Center was designed to assess compliance with Iowa seatbelt law. The display below shows the results for 5 Iowa cities. What type of display is this?
   1. Box plot
   2. Comparative histogram
   3. Stem plot
   4. Comparative bar chart
2. A study was conducted to investigate whether music training has an influence on the ability to memorize. Plots of the data sets are shown below. Which of the following is FALSE?

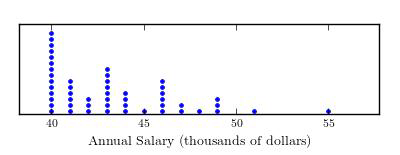


* 1. The distribution for the no music group is right-skewed.
  2. The mean for the music group is 9.
  3. The center of the data is lower for the no music group than for the music group.
  4. The interquartile range is larger for the music group than for the no music group.
  5. 25% of people in the music group scored at or below 6.

1. An athlete completed an 800-m race in 140 seconds. The distribution of 800-m race times

followed a bell curve, with mean 150 seconds and standard deviation 6. The same athlete also competed in a swim, finishing in 10.25 minutes. The distribution of swim times also followed a bell curve, with mean 12 minutes and standard deviation 1.5 minutes. In which event does the athlete have a better standing relative to the other competitors in the event?

1. The 800-m race because that has the higher z-score.
2. The swim because that has the higher z-score
3. The 800-m race because that has the lower z-score
4. The swim because that has the lower z-score
5. The athlete has the same standing in both events.
6. Scores follow a normal distribution with a mean of 100 and a standard deviation of 15. According the empirical rule, 95% of observations fall
   1. Below 100
   2. Between 85 and 115
   3. Between 70 and 130
   4. Above 70
7. If a distribution is shown to be positively skewed, then we can conclude:
8. The mean and median of that distribution are the same
9. The mean is greater than the median of that distribution.
10. The mean is smaller than the median of that distribution.
11. Cannot be determined
12. Which of the following is true about the coefficient of determination?
13. Its values range from -1 to 1.
14. It is the fraction (percent) of the variation accounted for by the linear relationship.
15. The lower it is the better the predictive power.
16. It is equal to the square root of the correlation coefficient.
17. The dot plot below shows the annual salaries, in thousands of dollars, of a sample of non-management staff who work for a small banking company.



Which of the following statistics would be the most reasonable estimate of the typical annual salary?

|  |  |
| --- | --- |
| a. | mean |
| b. | mode |
| c. | median |
| d. | interquartile range |
| e. | Range |

1. Data on ** the weight of a pickup truck (pounds) and ** distance (in feet) required for a truck traveling 40 miles per hour to come to a complete stop for 30 trucks was used to fit the least squares regression line . Which of the following statements is a correct interpretation of the value 0.05 in the equation of the regression line?

|  |  |
| --- | --- |
| a. | On average, the truck weight goes up 0.05 pound for each additional foot required to stop the truck. |
| b. | On average, the stopping distance is 0.05 foot when the truck weight is 0. |
| c. | The correlation coefficient for this data set is 0.05 |
| d. | On average, the stopping distance goes up 0.05 foot for each 1-pound increase in truck weight. |
| e. | Approximately 5% of the variation in the stopping distances can be explained by the linear relationship between stopping distance and truck weight. |

1. Is the explanatory variable cardiovascular fitness (as measured by time to exhaustion running on a treadmill (min)) related to an athlete’s performance in a 20-km ski race (min)?

The equation for the regression line is .

61.4% of the observed variation in ski time can be explained by the regression of ski time on run time to exhaustion.

What is the correlation between ski time and time exhaustion time?

1. .614
2. .888
3. -.233
4. -.783
5. A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

 = 10.9 + 0.23x

One student in the sample was 73 inches tall with a foot length of 29 cm. What is the predicted foot length for this student?

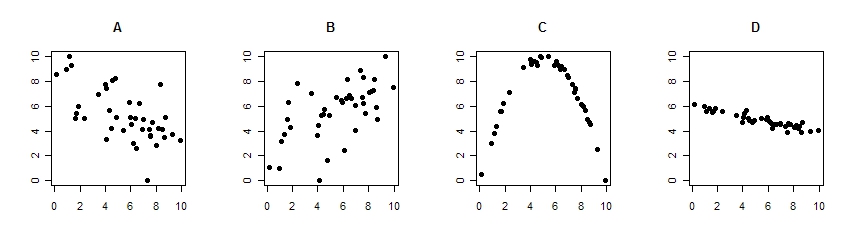
A. 17.57 cm

B. 27.69 cm

C. 29 cm

D. 33 cm

1. A residual plot is used to
   1. Determine how strongly the response and explanatory variable are related
   2. Determine whether linear regression is an appropriate model for the relationship between the response and explanatory variable
   3. Determine whether there is a causal relationship between the explanatory and response variable
   4. Determine the direction of the relationship between explanatory and response variables



1. Order the correlations from largest to smallest in absolute value.
   1. A, B, C, D
   2. D, A, B, C
   3. D, A, C, B
   4. C, D, B, A
2. Based on 1988 census data for the 50 States in the United States, the correlation between the number of churches per state and the number of violent crimes per state was 0.85. What can we conclude?
   * 1. The presence of a lot of churches in a state causes the number of violent crimes in the state to increase.
     2. The correlation is spurious because of the confounding variable of population size: both number of churches and number of violent crimes are related to the state’s population size.
     3. Since the data comes from a census, or nearly complete enumeration of the United States, there must be a causal relationship between the number of churches and the number of violent crimes.
     4. The relationship is not causal because only correlations of +1 or –1 show causal relationships.
3. A researcher investigating the association between two variables collected some data and was surprised when he calculated the correlation. He had expected to find a fairly strong association, yet the correlation was near 0. Discouraged, he didn’t bother making a scatterplot. Why might a scatterplot still reveal a strong association as the researcher expected?

A. There is a non-linear relationship between the variables

B. The data is influenced by outliers

C. Both A and B may be contributing factors.

D. None of the above, a correlation of 0 means that the variables have no association.

1. An electronics store sells a particular brand of computer notebook. Let X be the number of computer notebooks sold in a day. The probability distribution for X is below

X 0 1 2 3 4

Probability 0.1 0.45 0.25 \_\_\_\_ 0.15

1. What is the probability that three computer notebooks are sold in a day?
2. Find the probability that no more than two computer notebooks are sold in a day.
3. Find the mean for the above distribution.
4. Hens usually begin laying eggs when they are about 6 months old. Young hens produce eggs with an average weight of 51 grams and a standard deviation of 1.2 grams.
5. What proportion of eggs are between 49 and 52 grams?
6. Supermarkets like eggs to be of consistent weight, therefore, farmers reject the lowest and highest 7% of eggs. What are the minimum and maximum weights of eggs under this restriction?
7. A drug is effective in 60% of patients. Consider giving the drug to 10 patients and let X be the number of patients for whom the drug is effective.
   1. Find the mean and standard deviation of X.
   2. What is the probability the drug is effective for at most one patient?
   3. What is the probability the drug is effective for 7 or more patients?
8. A drug is effective in 60% of patients. Consider giving the drug to 100 patients and let X be the number of these patients for whom the drug is effective.
   1. Find the mean and standard deviation of X.
   2. What is the probability the drug is effective for 70 or more patients? Compute the probability using the normal approximation to the binomial distribution. Explain why it is appropriate to use the normal approximation in this case.
   3. Is your answer in this problem bigger or smaller than your answer in 3c? Explain why.