

THE MIDTERM WILL COVER CHAPTERS 1-6.

Chapter 1 – Exploring Data

- Quantitative data
- Histograms, stemplots, boxplots, dotplots
- Qualitative (Categorical) data
- Two way tables, bar graphs, pie charts
- Describe distributions – CCSSO
- Outliers – $1.5 \cdot \text{IQR}$ rule

Chapter 3 – Regression

- Least Squares Regression Line
- correlation coefficient (r)
- coefficient of determination (r^2)
- Find these values from formulas/calculator
- Interpretations/Definitions
- slope, intercept, r , r^2 , residual, residual plot, scatterplot

Chapter 5 – Sampling/Experimentation (Know Vocabulary!)

- Random Digit Table
- Simple Random Sample
- Cluster, Stratified, Multi Stage, Systematic
- Types of Sampling/Bias
- Experiment vs. Observational Study
- Blocking/matched pairs
- Randomization/Control/Replication
- Simulations

Chapter 2 - Density Curves

- Normal Distributions
- Z-scores/percentiles
- empirical rule (68-95-99.7)
- calculator – normalcdf(lower, upper, μ , σ)

Chapter 4 – More on Bivariate Data

- transforming relationships
- modeling nonlinear data (power & exponential)
- lurking variables
- causation, confounding, common response
- categorical relationships
- Simpson's Paradox

Chapter 6 – Probability

- OR vs. AND
- independent
- mutually exclusive (disjoint)
- calculating probabilities from two-way tables, tree diagrams, word problems
- conditional probability

1. An outlier is an observation that

A) is seen more frequently than the other observations in the data set.

B) is seen less frequently than the other observations in the data set.

C) is always smaller than the other observations in the data set.

D) is always larger than the other observations in the data set.

E) is significantly different from the other observations in the data set.

2. Suppose that a Normal model describes the acidity (pH) of rainwater, and that water tested after last week's storm had a z-score of +1.8. This means that the acidity of that rain

A) had a pH 1.8 standard deviations higher than that of average rainwater.

B) varied with a standard deviation of 1.8.

C) had a pH 1.8 higher than average rainfall.

D) had a pH 1.8 times that of average rainwater.

E) had a pH of 1.8.

3. A large company has offices in two locations, one in New Jersey and one in Utah. The mean salary of office assistants in the New Jersey office is \$28,500. The mean salary of office assistants in the Utah office is \$22,500. The New Jersey office has 128 office assistants and the Utah office has 32 office assistants. What is the mean salary paid to the office assistants in this company?

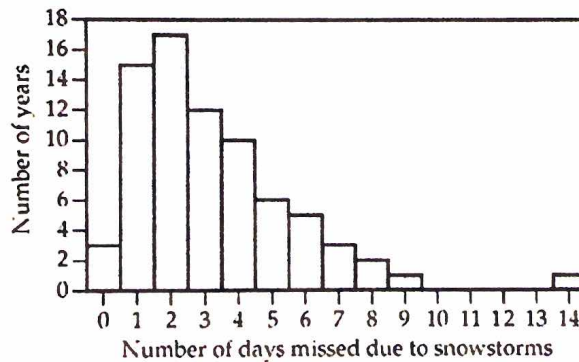
- A) \$22,500 B) \$23,700
 C) \$25,500 **D) \$27,300** E) \$28,500

4. Which statement about bias is true?

- I. Bias results from random variation and will always be present.
 II. Bias results from a sampling method likely to produce samples that do not represent the population.
 III. Bias is usually reduced when sample size is larger.

- A) I only B) III only
C) II only D) I and III only E) II and III only

5. In the northern US, schools are sometimes closed during winter due to severe snowstorms. At the end of the school year, schools have to make up for the days missed. The following graph shows the frequency distribution of the number of days missed due to snowstorms per year, using data collected from the past 75 years.



Which of the following should be used to describe the center of this distribution?

- A) Mean, because it is an unbiased estimator.
B) Median, because the distribution is skewed.
 C) IQR, because it excludes outliers and includes on the middle 50 percent of data.
 D) First quartile, because the distribution is left skewed.
 E) Standard deviation, because it is unaffected by the outliers.

6. Increasing the frequencies in the tails of a distribution will

- A) increase the standard deviation.** B) not affect the standard deviation.
 C) not affect the standard deviation as long as the increases are balanced on each side of the mean.
 D) reduce the standard deviation. E) none of these

7. The probability that a car will skid on a bridge on a rainy day is 0.75. Today the weather station announced that there is a 20 percent chance of rain. What is the probability that it will rain today and that a car will skid on the bridge?

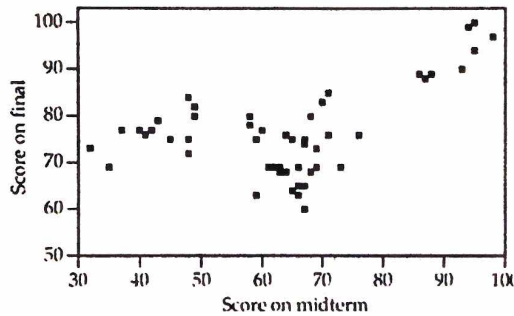
- A) 0.0300 B) 0.0375
 C) 0.1500 D) 0.3000 E) 0.9500

8. Randomizing

- I. can protect you against factors you know are in the data.
- II. can help protect you against factors that you aren't even aware of.
- III. makes it impossible for us to draw inferences about the population when we see only a sample.

- A) none of these B) III only
 C) I, II, and III D) II and III only E) I and II only

9. Extra study sessions were offered to students after the Midterm to help improve their understanding of statistics. Student scores on the Midterm and the Final Exam were recorded. The following scatterplot shows the Final test scores against the Midterm test scores.



Which of the following statements correctly interprets the scatterplot?

- A) All students have shown significant improvement in the Final Exam scores as a result of the extra study sessions.
 B) The extra study sessions were of no help. Each student's Final Exam score was about the same as his or her score on the Midterm.
 C) The extra study sessions further confused the students. All student scores decreased from Midterm to Final Exam.
 D) Students who scored below 55 on the Midterm showed considerable improvement on the Final Exam; those who scored between 55 and 80 on the Midterm showed minimal improvement on the Final Exam; and those who scored above 80 on the Midterm showed almost no improvement on the Final Exam.
 E) Students who scored below 55 on the Midterm showed minimal improvement on the Final Exam; those who scored between 55 and 80 on the Midterm showed moderate improvement on the Final Exam; and those who scored above 80 on the Midterm showed considerable improvement on the Final Exam.

10. A residuals plot is helpful because

- I. it will help us see whether our model is appropriate.
- II. it might show a pattern in the data that was hard to see in the original scatterplot.
- III. it will clearly identify influential points.

- A) I only B) II only
 C) I, II, and III D) I and III only E) I and II only

11. All but one of the statements below contain a mistake. Which one could be true?

A) The correlation between height and weight is 0.568 inches per pound.

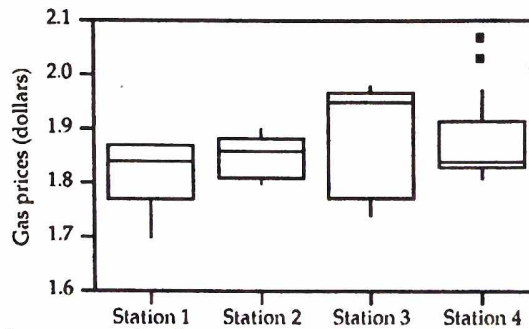
B) If the correlation between blood alcohol level and reaction time is 0.73, then the correlation between reaction time and blood alcohol level is -0.73.

C) The correlation between the breed of a dog and its weight is 0.435.

D) The correlation between gender and age is -0.171.

E) The correlation between weight and length of foot is 0.488.

12. A resident of Auto Town was interested in finding the cheapest gas prices at nearby gas stations. On randomly selected days over a period of one month, he recorded the gas prices (in dollars) at four gas stations near his house. The box plots of gas prices are as follows:



Which station has more consistent gas prices?

A) Station 1

B) Station 2

C) Station 3

D) Station 4

E) Station 4 without the outliers

13. The observed time (in minutes) it takes a runner to complete a marathon are approximately normally distributed. The z-score for his running time this week was -2. Which one of the following statements is a correct interpretation of his z-score?

A) This week his time was two minutes lower than his time last week.

B) This week his time was two minutes lower than his best time ever.

C) This week his time was two minutes lower than his average time.

D) This week his time was two standard deviations lower than his average time.

E) This week his time was two standard deviations lower than his time last week.

14. A numerical measure of the direction and strength of a linear association is its

A) association.

B) slope.

C) predicted value.

D) intercept.

E) correlation.

* 15. With observational data, as opposed to data from a designed experiment, there is no way to be sure that a(n) _____ is not the cause of any apparent association.

A) causal variable B) confounding variable

C) outlier

D) lurking variable

E) regression variable

16. The probability that Ted enrolls in an English class is $\frac{1}{3}$. If he does enroll in an English class, the probability that he enrolls in a mathematics class is $\frac{1}{5}$. If he does not enroll in an English class, the probability that he enrolls in a mathematics class is $\frac{3}{5}$. If we learn that Ted enrolled in a mathematics class, what is the probability that he enrolled in an English class?

A) $\frac{1}{15}$

B) $\frac{1}{7}$

C) $\frac{1}{4}$

D) $\frac{1}{3}$

E) $\frac{1}{2}$

* 17. The linear regression model contains

I. two easily-estimated parameters.

II. a meaningful measure of how well the model fits the data.

III. the ability to predict new values.

IV. a self-check to help us avoid silly mistakes.

A) I and III only

B) I, II, and III only

C) all of these

D) none of these

E) II only

18. A school committee member is lobbying for an increase in the gasoline tax to support the county school system. The local newspaper conducted a survey of county residents to assess their support for such an increase. What is the population of interest here?

A) all school-aged children

B) all county residents

C) all county residents with school-aged children

D) all county residents with children in the county school system

E) all county school system teachers

19. A manufacturer of ready-bake cake mixes is interested in designing a study to test the effects of four different temperature levels ($300^\circ, 325^\circ, 350^\circ, 375^\circ$ F $^\circ$), two different types of pans (glass and metal), and three different types of ovens (gas, electric, and microwave) on the texture of its cakes. How many different treatment combinations are to be used in this study?

A) 3

B) 9

C) 18

D) 20

E) 24

20. A regression analysis of students' AP Statistics scores vs. the number of hours they spent doing homework found $R^2 = 0.32$. Which of these is true?

I. 32% of student test scores can be correctly predicted with this model.

II. Homework accounts for 32% of your grade in AP Stats.

III. There's a 32% chance that you'll get the score this model predicts for you.

A) all of these

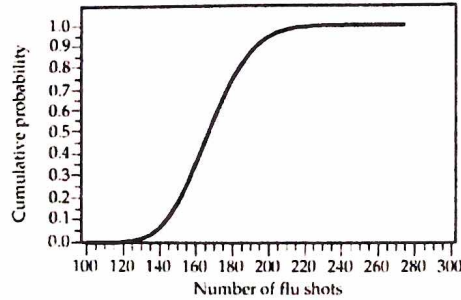
B) none of these

C) II only

D) III only

E) I and II

21. During flu season, a city medical center needs to keep a large supply of flu shots. A nurse's aide compiles data on the number of flu shots given per day in the past few years during flu season. A cumulative probability chart of the collected data is as follows:



How many flu shots should the center store every day to meet the demand on 95 percent of the days?

- A) At most 190
- B) At most 140
- C) Exactly 170
- D) At least 150
- E) At least 200

22. The relationship between the selling price of a car (in \$1,000) and its age (in years) is estimated from a random sample of cars of a specific model. The relation is given by the formula $\text{selling price} = 15.9 - 0.983(\text{age})$. Which of the following can we conclude from this equation?

- A) For every year the car gets older, the selling price goes down by approximately 9.83 percent.
- B) A new car costs on the average \$9,830.
- C) For every year the car gets older, the selling price drops by approximately \$1,590.
- D) A new car costs \$16,883.
- E) For every year the car gets older, the selling price drops by approximately \$983.

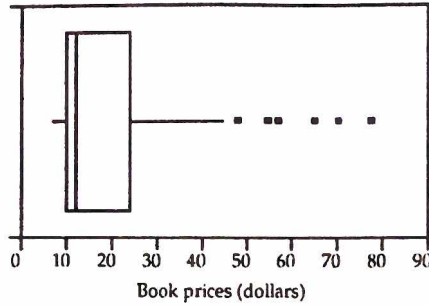
23. After a frost warning was issued, the owner of a large orange grove asked his workers to spray all his trees with water. The water was supposed to freeze and form a protective covering of ice around the orange blossom. Nevertheless, the owner suspected that some trees suffered considerable damage due to the frost. To estimate the proportion of trees that suffered more than 50 percent damage due to the frost, he took a random sample of 100 trees from his grove. What constitutes an observation (measurement) in this experiment?

- A) the proportion of trees that suffered more than 50 percent damage due to frost
- B) the number of trees affected by the frost
- C) for each sampled tree, whether it suffered more than 50 percent damage or at most 50 percent damage
- D) for each sampled tree, whether it was sprayed with water or not sprayed with water
- E) the number of trees sampled from the grove

25. A small kiosk at the Atlanta airport carries souvenirs in the price range of \$3.99 to \$29.99, with a mean price of \$14.75. The airport authorities decide to increase the rent charged for a kiosk by five percent. To make up for the increased rent, the kiosk owner decides to increase the prices of all items by 50 cents. As a result, which of the following will happen?

- A) The mean price and the range of prices will increase by 50 cents.
- B) The mean price will remain the same, but the range of prices will increase by 50 cents.
- C) The mean price and the standard deviation of prices will increase by 50 cents.
- D) The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.
- E) The mean price and the standard deviation will remain the same.

26. The following boxplot summarizes the prices of books at a bookstore.



Which of the following is true about the prices of the books at this store?

- A) This store carries more books priced above the mean than below the mean.
- B) This store carries more books priced below the mean than above the mean.
- C) This store carries about the same number of books at different prices in the entire price range.
- D) The mean price of books at this store is the same as the median price of books.
- E) The mean price of books at this store is lower than the median price of books.

27. An experiment was designed to test the effects of three different types of paint on the durability of wooden toys. Because boys and girls tend to play differently with toys, a randomly selected group of children was divided into two groups by sex.

Which of the following statements about this experiment is true?

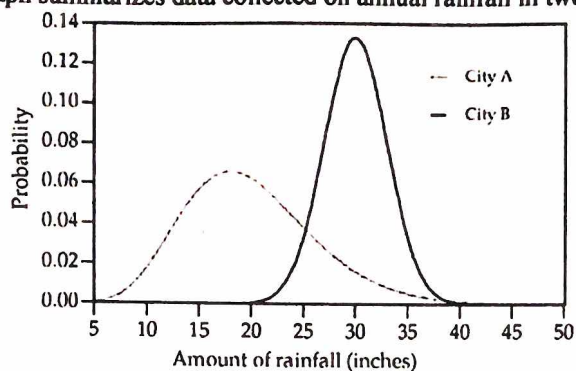
- A) There are three types of paint and two sex groups, giving a total number of six treatment combinations in this experiment.
- B) This experiment should be blocked by paint type.
- C) This experiment should be blocked by sex.
- D) This is a completely randomized experimental design.
- E) This is a matched-pairs experimental design in which one boy and one girl are matched by age to form a pair, then they are assigned to a group.

~~28~~

29. Data were collected on two variables, X and Y , and a least squares regression line was fitted to the data. The estimated equation for this data is $\hat{y} = -2.29 + 1.70x$. One point has $x = 5, y = 6$. What is the residual for this point?

- A) 7.91
- B) 6.21
- C) 0.21
- D) -0.21
- E) -2.91

30. The following graph summarizes data collected on annual rainfall in two cities for the past 150 years.



Which of the following conclusions can be made from this graph?

- A) The cities have different mean annual rainfalls, but the range of their annual rainfalls is approximately the same.
- B) On average, City B gets more rain than City A, but it has a smaller range of annual rainfall.
- C) On average, City B gets more rain than City A, and it has a larger range of annual rainfall.
- D) On average, City A gets more rain than City B, but it has a smaller range of annual rainfall.
- E) On average, City A gets more rain than City B, and it has a larger range of average rainfall.

31. On a boxplot, an observation is classified as an outlier if

- A) it is smaller than $\bar{X} - 3S$ or larger than $\bar{X} + 3S$
- B) it is larger than $\bar{X} - 3S$ and smaller than $\bar{X} + 3S$
- C) it is smaller than $Q1$ or larger than $Q3$
- D) it is smaller than $Q1 - 1.5(IQR)$ or larger than $Q3 + 1.5(IQR)$
- E) it is larger than $Q1 - 1.5(IQR)$ and smaller than $Q3 + 1.5(IQR)$

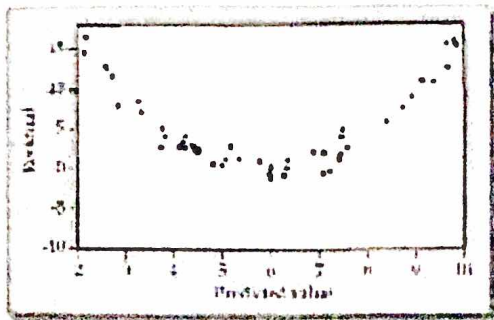
32. Which of the following statements about any two events A and B is true?

- A) $P(A \cup B)$ implies events A and B are independent.
- B) $P(A \cup B) = 1$ implies events A and B are mutually exclusive.
- C) $P(A \cap B) = 0$ implies events A and B are independent.
- D) $P(A \cap B) = 0$ implies events A and B are mutually exclusive.
- E) $P(A \cap B) = P(A) - P(B)$ implies A and B are equally likely events.

33. A company has 400 employees. Their mean income is \$20,500, and the standard deviation of their incomes is \$3,750. The distribution of incomes is normally distributed. How many of the 400 employees do you expect to have an income of between \$13,000 and \$28,000?

- A) 50
- B) 100
- C) 200
- D) 390
- E) 400

34. Sixty pairs of measurements were taken at random to estimate the relation between variables X and Y . A least squares line was fitted to the collected data. The resulting residual plot is as follows:



Which of the following conclusions is appropriate?

A) A line is an appropriate model to describe the relation between X and Y .

B) A line is not an appropriate model to describe the relation between X and Y .

C) The assumption of The Law of Averages has been violated.

D) The variables X and Y are not related at all.

E) Without knowing what variables X and Y are, there is no way to form an appropriate conclusion about a possible linear connection between them.

35. An insurance agent is successful in selling a life insurance policy to 20 percent of the customers he contacts. He decides to construct a simulation to estimate the mean number of customers he needs to contact before being able to sell a policy. Which of the following schemes should he use to do the simulation?

A) Assign numbers "0, 1" to successfully selling a policy to a customer and numbers "2, 3, 4, 5, 6, 7, 8, 9" to failing to sell a policy to a customer.

B) Assign numbers "0, 1" to successfully selling a policy to a customer and numbers "2, 3, 4" to failing to sell a policy to a customer.

C) Assign number "0" to successfully selling a policy to a customer and number "1" to failing to sell a policy to a customer.

D) Assign numbers "0, 1, 2, 3, 4" to successfully selling a policy to a customer and numbers "5, 6, 7, 8, 9" to failing to sell a policy to a customer.

E) Assign number "20" to successfully selling a policy to a customer and numbers "1, 3, 5, 7, 9, 11, 13, 15, 17, 19" to failing to sell a policy to a customer.

36. A newspaper reporter examined police reports of accidents during the past 12 months to collect data about the speed of a car and its stopping distance. The reporter then constructed a scatterplot and computed a correlation coefficient to show the relation between a car's speed and its stopping distance. This is an example of

A) a double-blind study. B) a single-blind study.

C) a study involving no blinding at all.

D) an observational study.

E) a well-designed experiment.

37. A company with offices in five different countries is interested in estimating the proportion of its employees in favor of banning smoking on the office premises. It is known that the views of people from different countries on issues like smoking tend to vary due to the influence of different local social structures. Which of the following is an appropriate sampling technique to use in this situation, and why?

- A) Stratified sample, because the population is divided into five strata – namely, the five offices in five different countries.
- B) Stratified sample, because people in different countries use different languages and thus will not be able to communicate to each other the results of a simple random sample.
- C) Simple random sample, because the study has not been attempted before
- D) Simple random sample, because this is the only method that gives unbiased results.
- E) Simple random sample, because this is the only method for which inferential techniques are available.

38. In a clinic, 50 patients with sleep disorders are randomly assigned to one of two different groups. Patients in one group are given medication before bedtime. Patients in the other group are given blindfolds and played soft music at bedtime. Each patient is attached to a machine that records breathing patterns. From the patterns, it is possible to determine if the patient is awake or asleep. The data will be used to decide which method is more effective in helping patients with sleep disorders. Which of the following statements is correct in the context of this experiment?

- A) This is a single-blind experiment, because only one group uses blindfolds.
- B) This is a single-blind experiment, because only patients are not doctors use blindfolds.
- C) This is a double-blind experiment, because patients are blindfolded and the doctor does not know which patient receives which treatment.
- D) This experiment cannot be a single-blind experiment, because many patients do not like to be blindfolded.
- E) This experiment cannot be a double-blind experiment, because patients will know which treatment they are receiving, although the examining doctor might not.

39. The relation between studying time (in hours) and grade on a semester exam (0-100) in a random sample of students in math class was found to be $\widehat{grade} = 50.25 + 10.2(hours)$. How will a student's grade be affected is she studies for two hours?

- A) Her grade will go up by 40.05 points.
- B) Her grade will go up by 20.4 points.
- C) Her grade will remain unchanged.
- D) Her grade will go down by 10.2 points.
- E) It cannot be determined from the information given.