

1. Which of the following graphical displays would you use to portray the distribution of a **categorical** variable?
 - (a) box plot
 - (b) bar chart
 - (c) histogram
 - (d) stem plot

2. True or False. The **standard normal distribution** has mean equal to 0.
 - (a) True
 - (b) False

3. A large data set consisting of stock market rate of returns (measured in percentages) has mean 5.3 percent and standard deviation 3.6 percent. A histogram of the percentages closely resembles a normal distribution. The **Empirical Rule** says that approximately 95 percent of the rate of returns should fall between which two values?
 - (a) 3.6 and 7.0 percent
 - (b) 1.7 and 8.9 percent
 - (c) -1.9 and 12.5 percent
 - (d) -5.5 and 16.1 percent

4. On this exam, suppose that the mean score is 75 and that the standard deviation is 8. You make a 79. What is your **standardized score**?
 - (a) 0.3
 - (b) 0.7
 - (c) 0.5
 - (d) -0.5

5. Which of the following statements is **false**?
 - (a) The value of the sample mean is not sensitive to outliers, but the median's value is.
 - (b) Standardized values are unitless quantities.
 - (c) The value of the correlation r can be sensitive to outliers.
 - (d) The sample standard deviation measures how spread out observations are about their mean.

6. If a density curve is **skewed to the left** (the low side), then
- (a) the mean will be less than the median.
 - (b) the mean will be greater than the median.
 - (c) the mean and median will be equal.
7. Which graphical display would you use to portray the relationship between **two quantitative variables** measured on the same individuals?
- (a) bar graph
 - (b) pie chart
 - (c) line plot
 - (d) scatterplot
8. Which of the following variables is **categorical**?
- (a) the amount of fertilizer (in pounds) applied in an agricultural experiment
 - (b) starting salaries for undergraduate engineering majors
 - (c) the mean daily CD4 cell count (number of cells) for a group of HIV patients
 - (d) marital status of USC professors
9. Which statistic is **not** part of the Five Number Summary?
- (a) median
 - (b) maximum
 - (c) minimum
 - (d) mean
10. Which of the following statements concerning the sample standard deviation s is **false**?
- (a) The larger the standard deviation, the more “spread out” the distribution of the data.
 - (b) The value of s can never be less than zero.
 - (c) It is measured in original units of the data.
 - (d) Its value is not heavily affected by outliers.

11. In a recent *University of South Carolina Times* press release, President Sorensen said that by the year 2010, he would like a freshmen class boasting SAT scores of 1100 at the 25th percentile and 1300 at the 75th percentile. If this goal is attained, what will the **interquartile range** be for the distribution of freshman SAT scores?

- (a) 100
- (b) 200
- (c) 400
- (d) 1200

12. True or False. The **area** under any density curve is equal to 1.

- (a) True
- (b) False

13. Which of the following values is closest to the **75th percentile** for the standard normal distribution?

- (a) 0.7
- (b) 1.6
- (c) 2.0
- (d) -2.0

14. A stem and leaf display is shown below. The stem is the tens place and the leaf is the units place.

0		3	6				
1		0	2	2	6	8	
2		1	2	5	6		
3		1					

What is the **median** of this distribution?

- (a) 18
- (b) 17
- (c) 15
- (d) 12

15. Refer to the data set in Question 14. What is the **range** of this distribution?

- (a) 23
- (b) 28
- (c) 13
- (d) 18

16. These data represent the number of accidents (per month) at a busy intersection in Jackson, Mississippi, for $n = 20$ consecutive months.

6 4 3 4 6 6 4 6 0 7 6 6 4 6 3 7 6 6 5 5

I have computed that $\sum x_i = 100$. What is the **sample mean**?

- (a) 4
- (b) 5
- (c) 6
- (d) 10

17. A survey records many variables of interest to the researchers conducting the survey. Which of the following variables is **quantitative**?

- (a) county of residence
- (b) age of respondent
- (c) political party affiliation of the household head
- (d) gender of the household head

18. Quiz scores for a large undergraduate class follow a normal distribution with mean 20 and standard deviation 5. Your score is 18. What is your **standardized score**?

- (a) 0.34
- (b) 0.66
- (c) 0.4
- (d) -0.4

19. A unimodal data distribution that is **not** symmetric is said to be

- (a) confounded.
- (b) continuous.
- (c) unbiased.
- (d) skewed.

20. When describing relationships with the correlation, often times people are careless. Which one of the following statements does **not** contain a mistake?

- (a) "The correlation between yield (in kg) and amount of fertilizer (sodium nitrate, in kg) was close to 0. There is virtually no linear relationship between the yield and the amount of sodium nitrate applied."
- (b) "For our experiment, we computed the correlation to be $r = 0.78$ pounds."
- (c) "The correlation between gender and promotion status was significantly higher than zero. This suggests that males and females are promoted at different rates."
- (d) "The correlation between age and number of failures was much stronger for the experimental formulation ($r = 0.21$) than it was for the control formulation ($r = -0.96$)."

21. True or False. We use population parameters to estimate sample statistics.

- (a) True
- (b) False

22. What is another word/phrase for **average**?

- (a) variance
- (b) mean
- (c) median
- (d) interquartile range

23. You have data which give the monthly sales, recorded in dollars, and regions (A or B) for 3 consecutive years (so that there are 36 observations per region). You are not interested in the longitudinal behavior in the sales; rather, you are interested in **comparing** the two regions. Which graphical display would be best?

- (a) pie charts for each region
- (b) a time plot for all 72 observations
- (c) a histogram for all 72 observations
- (d) side-by-side box plots

24. True or False. For any data set, the **sum of the deviations** from the mean; i.e.,

$$\sum (x_i - \bar{x}),$$

is always equal to 0.

- (a) True
- (b) False

25. If the **sample mean** of a data set is 15 and the **sample variance** is 9, what percent of the data would you expect to fall between 6 and 24, assuming that the data distribution is fairly symmetric? Note that I'm giving you the sample variance here; not the sample standard deviation.

- (a) 68 percent
- (b) 81.5 percent
- (c) 95 percent
- (d) 99.7 percent

26. In a regression setting, the **vertical distance** from a data point to the regression line is called a(n)

- (a) residual
- (b) lurking variable measure
- (c) correlation
- (d) outlier

27. We have computed the probability of an event to be 0.0003. Which of the following statements would **not** be correct?

- (a) If the event did occur, we would consider it to be unusual.
- (b) The event will never occur.
- (c) The event is unlikely to occur.
- (d) We would expect the event to occur about 0.03 percent of the time.

28. True or False. The standard deviation is measured in **original units** of the data.

- (a) True
- (b) False

29. Consider the following **straight-line regression model**, obtained from an SRS of first-year students at a small school in Menomonie, WI.

$$\text{GPA} = -2.22 + 0.045\text{IQ}$$

The r^2 statistic from the data was computed to be 0.812 (or 81.2 as a percent). What is the interpretation of this r^2 statistic?

- (a) Since the r^2 is pretty high for this regression, the model must be a good model for the data.
- (b) About 81 percent of the data values fall on the regression line.
- (c) About 81 percent of the variation in the GPA data is explained by the IQ scores.
- (d) We are 81 percent confident that our predictions will include the mean GPA.

30. A statistic is said to be an **unbiased estimator** of a parameter if

- (a) the standard deviation of the statistic is equal to the parameter.
- (b) its sampling distribution is approximately normal.
- (c) the statistic can be computed in repeated sampling.
- (d) its sampling distribution has mean equal to the parameter.

31. As a recent home-buyer, I have become interested in the **longitudinal** behavior of mortgage rates since 2000. I have monthly data on these rates from January 2000 until March 2006 (a total of 75 months). To examine these data graphically, with the hope of studying the pattern in these rates over this 75 month period, which graphical display would be best?

- (a) boxplot
- (b) time plot
- (c) scatterplot
- (d) stem plot

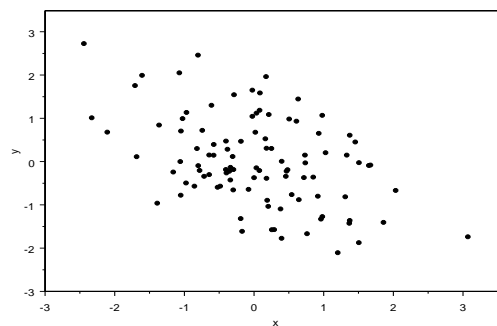
32. Which of the following values is closest to the **5th percentile** for the standard normal distribution?

- (a) 2.0
- (b) -2.0
- (c) -1.6
- (d) 1.6

33. Which statistic does **not** provide a measure of variability (or spread)?

- (a) median
- (b) range
- (c) interquartile range
- (d) variance

34. In the scatterplot below, which value is closest to the **correlation**?



- (a) $r = 0.01$
- (b) $r = 0.22$
- (c) $r = -0.99$
- (d) $r = -0.50$