STAT303 Secs 509–511  
Spring 2002  
Exam #1  
Form A  

Instructor: Julie Hagen Carroll  

PRINT Your Name: ____________________________  
Thursday Section: __________  
Computer Number: __________  

1. Don’t EVEN open this until you are told to do so.  
2. Be sure to MARK your REGISTERED section number and your test form (A, B, C or D) on the scantron!  
3. Sign your name where indicated on your scantron and write your (Thursday) section number and computer number beside it. Also, PRINT your name at the top of this exam and include your (Thursday) section and computer number. You will get your scantron and exam back.  
4. There are 20 multiple-choice questions on this exam, each worth 5 points. There is partial credit. Please mark your answers clearly on the scantron. Multiple marks will be counted wrong.  
5. You will have 60 minutes to finish this exam.  
6. If you are caught cheating or helping someone to cheat on this exam, you both will receive a grade of zero on the exam. You must work alone.  
7. This exam is worth 100 points, and will constitute 20% of your final grade.  
8. Good luck!
1. An ecologist wants to know the average humidity (at 2 ft. above ground level) on Manitou Island. She knows from an earlier study that 50% of the land is field, 40% is forest and 10% is beach. Based on this information, she decides to get a simple random sample of 20 field plots, 16 forest plots, and 4 beach plots. What sampling method did she use?

A. simple random sample
B. stratified sample
C. cluster sample
D. segmented sample
E. She did not use random samples since she picked exactly 20 field, 16 forest and 4 beach plots.

2. Which of the following is true about the point (50,30) on the bottom edge of the graph?

A. It should be deleted since no one has a dad that’s only 30 inches tall.
B. It is an influential point since it drags the line down to it, i.e., the slope and intercept would change if we deleted the point.
C. Since it’s obviously an outlier, it’s not included and doesn’t influence the line.
D. All of the above are true.
E. Only two of the above are true.

3. How would you describe the correlation between Dads Ht and Moms Ht (without the point above included)?

A. strongly negative
B. moderately negative
C. weak
D. moderately positive
E. strongly positive

4. What would be the approximate height of a dad for a mom 6 ft. 8 in. tall?

A. 80 in. since that’s mom’s height
B. 79 in. using the equation
C. 82 in. since the dads are taller than the moms
D. It’s impossible for a mom to be that tall.
E. Since the mom’s height is outside the range of the data, we shouldn’t use the equation to determine the dad’s height.

5. Suppose I wanted to convert the data above to feet, rather than inches. Which of the following would be true?

A. The slope would stay the same since the x and y values had the same scale change.
B. The intercept would stay the same since the x and y values had the same scale change.
C. The correlation would decrease by the scale change since there are less feet than inches (1 ft. vs. 12 in.).
D. All of the above are true.
E. None of the above are true.

6. Which of the following is/are NOT affected by outliers?

A. the median, ˜x
B. a z-score for any observation
C. the first quartile, Q1
D. None of the above are affected by outliers.
E. Exactly two of the above are NOT affected.

7. Some companies “grade on a bell curve” to compare the performance of their managers and professional workers. This forces the use of some low performance ratings, so that not all workers are graded “above average”. Until the threat of lawsuits forced a change, Ford Motor Company’s “performance management process” assigned 15% A’s and C’s and 70% B’s to the company’s 18,000 managers. Assuming the distribution is approximately bell-shaped (normal) what are the approximate mean and standard deviation if C’s have scores of 25 or less and A’s have scores of 475 or more?

A. x = 200 and s = 175
B. x = 200 and s = 275
C. x = 250 and s = 225
D. x = 200 and s = 225
E. x = 225 and s = 200
8. Suppose we want to know whether vitamin C actually reduces the chance of catching a cold. Why would it be hard to prove?
   
   A. There are too many confounding variables that we cannot control.
   B. This is an observational study, not a controlled experiment.
   C. ‘Human opinion’ could bias the responses.
   D. All of the above are reasons we cannot prove vitamin C reduces the chance of catching a cold.
   E. We don’t need to prove it again. Everyone knows that vitamin C helps reduce the chance of catching a cold.

9. Which of the datasets represented by the boxplots above has the largest standard deviation, assuming there are the same number of observations in each?
   
   A. 1
   B. 2
   C. 3
   D. 4
   E. It can’t be determined from boxplots since they don’t show the standard deviation.

10. It is known that the price of houses sold in College Station has a right-skewed distribution. About what percentage of the prices of houses sold are between the first and third quartiles, \( Q_1 \) and \( Q_3 \)?
   
   A. It can’t be determined since the data is not normal.
   B. 50%
   C. 68%
   D. 65%
   E. 99.7%

11. What are the ‘5 Number Summary’ and shape of the data in the output above?
   
   A. 18,19.5,22,26.5,29; it’s uniform
   B. 18,20,21,22,29; it’s skewed right
   C. 18,20,21,22,29; it’s skewed left
   D. 2.44,43.90,92.68,96.34,100.00; it’s uniform
   E. 18,19,20,21,29; it’s skewed right

12. Which of the following is true about histograms, boxplots and stemplots?
   
   A. A histogram is the best graph to use since it doesn’t lose any information.
   B. A stemplot is the only one that shows the individual observations.
   C. Boxplots are the only graphs that can compare different datasets.
   D. Boxplots are the only graphs that give us the value of the median.
   E. All of the above are true.

13. If I told you your \( z \)-score on this exam was a 0.25, what would this mean?
   
   A. Your score was the 25th percentile of the class.
   B. You got 25% of the answers correct.
   C. You made a 75 on the exam.
   D. You did better than average.
   E. You did worse than average.
14. Using the table above how likely is a student to be a College of Geosciences (GEOS) senior (U4)?

A. 2/11
B. 2/26
C. 2/78
D. 26/78
E. 11/78

15. Using the table above, which of the following statements would be valid?

A. More than half of A&M’s students are juniors (U3).
B. A&M has twice as many Liberal Arts (LARTS) majors as Ag and Life Sciences (AGLS).
C. You can’t be in Stat303 if you are a freshman (U1).
D. All of the above are valid.
E. None of the above are valid.

16. So I have the test scores by section and gender. If I randomly pick one of the sections as my sample to determine the average, this would be

A. a simple random sample
B. a stratified sample
C. a cluster sample
D. a segmented sample
E. a biased sample

17. When can you correctly apply the 68-95-99.7% Rule?

A. when the data is bell-shaped
B. when the data is symmetric
C. when 68% of the data is within 1 standard deviation of the mean
D. All of the above are true.
E. Only 2 of the above are true.

18. It's been shown that heights follow a normal distribution. If the average height of men is $\mu = 70$ in. with a standard deviation, $\sigma = 5$, about what proportion of men would be between 65 and 80 inches tall?

A. 81.5%
B. 13.5%
C. 68%
D. 95%
E. 47.5%

19. When can we use a linear equation to predict the response variable?

A. when the explanatory and response variables are related
B. when we use simple linear regression to get the equation
C. when explanatory variable value is within the range of our dataset
D. All of the above are valid.
E. None of the above are sufficient.

20. During the stockmarket boom of the 1990s, initial public offerings (IPOs) of the stock of new companies often produced enormous gains for people who bought the stocks when they first became available. At least that’s what legend says. A study of all 4567 companies that went public in the years 1990 to 2000 (excluding very small IPOs) found that on the average their stock prices had either risen 111% or declined 31% by the end of the 2000. One of these numbers is the mean change in price and the other is the median change. Which is which and why?

A. These calculations must be wrong since one is positive and the other is negative.
B. The mean is 111% and the median is 31% since the data is obviously skewed to the left (more did well).
C. The mean is 111% and the median is 31% since the data is obviously skewed to the right (more failed).
D. The median is 111% and the mean is 31% since the data is obviously skewed to the left (more did well).
E. The median is 111% and the mean is 31% since the data is obviously skewed to the right (more did failed).

1B,2E,3C,4E,5A,6E,7C,8D,9A,10B,11B