STAT303 Secs 508–510
Fall 2003
Exam #1
Form A
Instructor: Julie Hagen Carroll
September 23, 2003

1. Don’t EVEN open this until you are told to do so.

2. 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.

3. You will have 60 minutes to finish this exam.

4. 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.

5. This exam is worth 100 points, and will constitute 20% of your final grade.

6. Good luck!
1. In a tire-treadwear study one brand of tires was tested at different temperatures. What is the explanatory variable??

A. brand of tire
B. temperature
C. both brand and temperature
D. tire-treadwear
E. both tire-treadwear and temperature

2. Which of the following best describes the distribution above?

A. skewed left, \( \bar{x} \) is between 3 and 5, and \( \mu \) is slightly larger
B. skewed right, \( \bar{x} \) is between 3 and 5, and \( \mu \) is slightly larger
C. skewed left, \( \bar{x} \) is between 3 and 5, and \( \mu \) is slightly smaller
D. skewed right, \( \bar{x} \) is between 3 and 5, and \( \mu \) is slightly smaller
E. skewed left, \( \mu \) is between 3 and 5, and \( \bar{x} \) is slightly larger

3. Which of the following would also be correct about the distribution in the histogram of the last problem?

A. A boxplot of the data would have the box near the bottom with a longer tail (whisker) stretching upwards.
B. It’s unlikely that there would be outliers indicated in the boxplot.
C. The median and standard deviation would be the best summary numbers to use for this dataset.
D. All of the above are true.
E. Exactly two of the above are true.

4. Assuming that the distribution of grades on this exam is approximately normal, what is the minimum z-score would you have to have to be in the top 25% of the class?

A. 0.7500
B. 0.250
C. 0.6750
D. 2.50
E. It’s impossible to tell without knowing the mean and standard deviation.

5. Confounding is:

A. when we establish that changes in the explanatory variable cause changes in the response variable.
B. when the changes in both, the response and explanatory variables, are caused by a lurking variable.
C. mixing of influences between the variables.
D. just another fancy term for regression.
E. when someone uses statistics to confuse the reader.

6. Regarding the Empirical Rule, which of the following is FALSE?

A. Approximately 95% of the data fall within 3 standard deviations of the mean.
B. Approximately 99.7% of the data fall within 4 standard deviations from the median.
C. Approximately 68% of the data fall within 1 standard deviation from the mean.
D. Two of the above are false.
E. All of the above are false.

7. There is a negative statistical association between heart attacks and amount of hair. Guys with lots of hair are less likely to have heart attacks than bald guys (less hair, more heart attacks). But of course getting older is the real cause. Old guys have less hair and more heart attacks than young guys. What idea does this illustrate?

A. Women don’t go bald.
B. Heart attacks cause baldness.
C. The negative association between heart attacks and amount of hair is a coincidence.
D. Baldness causes heart attacks.
E. Both variables result from a common cause.
8. The scatterplot above shows the relationship between how many hours of sleep you get and how well you do on the exam. Assuming the regression equation is \( \text{grade} = -12.5 + 12.5 \times \text{sleep} \), which of the following is TRUE?

A. If you don’t sleep at all, you’ll make a −12.5 on the test.
B. If you’d but sleep another hour, you’d raise your score a full letter grade.
C. If you sleep 9 hours, you’ll make a 100.
D. All of the above are true.
E. None of the above are true.

9. Suppose we added a sick person to the data above, one who slept 12 hours but still didn’t pass (less than 60 on the test). What would this do?

A. Nothing since they would be an outlier.
B. The correlation would decrease since they’d not fit the line.
C. The correlation would decrease since the slope would then be negative.
D. They’d be an influential point, and the line would probably change direction.
E. They’d be an influential point, so the correlation would increase.

10. Still talking about the scatterplot, which of the following is/are TRUE?

A. Since the correlation is strong, we can accurately predict how well you’re going to do by looking at how long you slept.
B. Although the correlation is strong, we know that doesn’t have anything to do with how well you do on the test.
C. The correlation is strong, but there is likely some other influence involved that we don’t have data on.
D. The correlation is only moderate since it’s obvious that how much sleep you get cannot determine how well you do on an exam.
E. The correlation is only moderate because our prediction of how well you’ll do on the exam is only an estimate.

11. Ok, one more about sleep and grades. What would happen if we recorded sleep in minutes rather than hours?

A. Nothing, everything would be the same.
B. The prediction would be more accurate since minutes are more exact than hours.
C. The correlation would be stronger since minutes are more exact than hours.
D. The slope would and the intercept would decrease by a factor of 60.
E. The slope would decrease by a factor of 60.

12. If the weights of students at A&M are normally distributed with a mean of 160 and a standard deviation of 25, what is the probability that a randomly selected student will weigh more than 150?

A. 0.6554
B. 0.9536
C. 0.4
D. 0.3446
E. 0.25

13. If we survey the class and ask each person for their Zip Code, what would be the best way to graphically display this data?

A. Histogram
B. Stem plot
C. Scatter plot
D. Pie chart
E. None of the above
14. Suppose we have a fairly normal looking distribution of probabilities of getting a date to the next Aggie football game where the average is 40% (on average you get a date 40% of the time) with a standard deviation of 4.9%. So \( P \sim N(0.4, 0.049^2) \). What is the 95th percentile of this distribution?

A. 82.89%
B. 1.645
C. \( \pm 2 \text{ sd} \)
D. 48%
E. 0.495

15. In a study of the relationship between gender and the progress of students in doctoral programs, a sample of 1024 students is obtained. From the table below, which of the following are TRUE?

<table>
<thead>
<tr>
<th></th>
<th>men</th>
<th>women</th>
<th>totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>completed</td>
<td>423</td>
<td>98</td>
<td>521</td>
</tr>
<tr>
<td>still</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>enrolled</td>
<td>134</td>
<td>33</td>
<td>167</td>
</tr>
<tr>
<td>dropped</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>out</td>
<td>238</td>
<td>98</td>
<td>336</td>
</tr>
<tr>
<td>totals</td>
<td>795</td>
<td>229</td>
<td>1024</td>
</tr>
</tbody>
</table>

A. 66% of the people in the study are men.
B. Of the women, 14.4% are still enrolled.
C. Of those that have completed a doctoral degree, 53.2% are men.
D. Only A and B are true.
E. Only A and C are true.

16. How likely are we to see a male drop-out according to the study above?

A. more than 70% of the time
B. less than 25% of the time
C. about 30% of the time
D. almost 80% of the time
E. about a third of the time

17. If the correlation coefficient \( r \) is \(-0.92\) for variables A and B, then

A. there is no linear relationship between A and B.
B. as the values of A increase the values of B tend to decrease.
C. there would be a stronger correlation between A and B if \( r = 0.92 \).
D. the correlation between B and A would be \( r = 0.92 \).
E. None of the above are true.

18. The time taken to prepare the envelopes to mail a weekly report to all executives in a company has a normal distribution with mean 35 minutes and a standard deviation of 2 minutes. 10% of the time, the mailing preparation takes more than

A. 32.4 minutes.
B. 37.6 minutes.
C. 35.2 minutes, the mean plus 10% of 2.
D. 38.5 minutes, the mean plus 10%.
E. None of the above are correct.

19. An outlier

A. is an observation that is very different from the rest of the data set.
B. greatly affects the value of the sample mean and sample standard deviation.
C. greatly affects the sample median.
D. Only A and B are true.
E. All of the above - except D.

20. What’s the chance a standard normal random variable is between -2.1 and 1.37?

A. 0.73
B. 0.7673
C. 0.8968
D. 0.9326
E. 0.2327

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