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

2. All graphs are on the last page which you may remove.

3. There are 20 multiple-choice questions on this exam, each worth 5 points. There is partial credit. Please mark your answers clearly. Multiple marks will be counted wrong.

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

5. If you have questions, please write out what you are thinking on the back of the page so that we can discuss it after I return it to you.

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 the same as a regular exam (this may differ from section to section).

8. When you are finished please make sure you have marked your CORRECT section (Tuesday 12:45 is 508, 2:20 is 509, and 3:55 is 510) and FORM and 20 answers, then turn in JUST your scantron.

9. Good luck!
1. Which of the following is/are true?

A. The mean of any set of $z$-scores (i.e., calculated from one dataset using its mean and standard deviation) is 0.
B. The IQR for a uniform dataset is approximately 50% of the range.
C. Stemplots are the only numeric graph that shows the actual data values.
D. All of the above are true.
E. Only two of the above are true.

2. Referring to the boxplots on the last page, which one has the 5-Number Summary: 0, 2, 5, 13, 35? Think outliers!

A. A
B. B
C. C
D. D
E. None of them since the only goes to about 32, not 35.

3. Which of the following seems to have the strongest correlation, $r$?

A. your zip code and where you live
B. your SSN and when you were born
C. your area code and what state you live in
D. your height and weight
E. Each of the above has strong correlation but we don’t know exactly how much without the numbers.

4. How likely is a junior, $U3$, to be in the College of AGLS in the survey above?

A. 12/17
B. 12/41
C. 12/78
D. 17/78
E. 41/78

5. Which of the following is/are true?

A. 8/78 is a conditional probability conditioning of being an EDUC major.
B. 12/36 is how likely we are to pick a senior LARTS major.
C. 36/78 is a marginal probability for Class.
D. All of the above are true.
E. None of the above are true.

6. Which of the following is true about outliers?

A. Since they are points that don’t ‘fit’ the rest of the data, they should be omitted in the analysis (calculations).
B. They decrease correlation between two variables, but they won’t affect the $z$-scores for each distribution.
C. They affect location statistics, but not spread statistics.
D. All of the above are true.
E. None of the above are true.

7. Which of the following is true about the histogram on the last page? Each bin is 2 units wide.

A. The median is between 32 and 34 because it is the tallest bin and the mean is somewhat less.
B. The median is between 30 and 32 and the mean is somewhat less.
C. The median is between 28 and 30 and the mean is somewhat more.
D. The median is between 30 and 32 and the mean is somewhat more.
E. The median is between 28 and 30 and the mean is somewhat more.

8. If we changed all the 0 to 2’s in the histogram to 10 to 12’s (i.e., added 10 points to them, but only them), which of the following would be true.

A. The mean would increase and the standard deviation would decrease.
B. The mean would increase by 10 but the standard deviation would not change.
C. The mean and the standard deviation would both increase.
D. The mean and the standard deviation would decrease.
E. The mean and the median would increase.

9. If the correlation between $X$ and $Y$ is $-0.87$, then which of the following is true for the set of data, $(x, y)$?

A. There is a strong linear relationship between $X$ and $Y$.
B. There is a weak linear relationship between $X$ and $Y$ since $r$ is close to the minimum possible value, $-1$.
C. A correlation of 0.56 would be stronger since it is larger than $-0.87$.
D. We would need to plot the data to determine if there is really a strong linear relationship between $X$ and $Y$.
E. More than one of the above is a true statement.
10. There is a new deli in town which sells soups, salads, and sandwiches. There are six different soups, ten different salads and numerous combinations of bread, veggies, cheeses and meats for the sandwiches. To test the quality of the deli, I decided to take a random sample of four of each type of food. I used

A. three separate random samples
B. a stratified sample
C. a cluster sample
D. a multi-stage sample
E. a systematic sample

11. Still at the deli... If on a different day, I took another sample like the one above,

A. I could get exactly the same four soups, salads and sandwiches.
B. I could get four completely different soups, salads and sandwiches.
C. I could get some of the same soups, salads and sandwiches.
D. All of the above are true.
E. Only two of the above are true.

12. And again at the deli... If I only sampled one soup, one salad and one sandwich (come on, that’s a lot of food), which element of a good experiment did I forget that I had before?

A. The experiment is fine. I didn’t omit anything.
B. randomization
C. replication
D. control
E. blinding

13. Suppose you need to convert a distribution, $X$, with mean 50 and standard deviation 5, to a new distribution, $Y = 14 - 8X$. What are the new mean and standard deviation? (Think shift and scale changes.)

A. $\bar{Y} = 50$ and $s_Y = 5$, shift and scale changes don’t affect the new distribution
B. $\bar{Y} = -386$ and $s_Y = -26$
C. $\bar{Y} = -386$ and $s_Y = 26$, standard deviations can’t be negative
D. $\bar{Y} = -386$ and $s_Y = -40$
E. $\bar{Y} = -386$ and $s_Y = 40$

14. 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) with mean 250 and standard deviation 225, what is the approximate minimum score an employee has to make to get an A, be in the upper 15%?

A. 475
B. slightly more than 475
C. slightly less than 475
D. more than 700
E. more than 275

15. Mary’s high school used the 100 point scale with mean 82 and standard deviation 7. Mike’s high school used the 4.0 scale with mean 2.6 and standard deviation 0.3. If Mike’s grade point was 3.1, what is his equivalent grade at Mary’s school?

A. We can’t determine that unless we know the data is normal for both schools.
B. 1.67
C. 94
D. 70
E. 61

16. The correlation for the scatterplot on the last page is closest to

A. 0.9
B. 0.6
C. 0.3
D. −0.6
E. −0.9
17. (Think definitions) The point in the top right of the scatterplot, approximately (40,30) would be considered
   A. an outlier since it doesn’t really fit the line.
   B. an influential point since removing it would change the equation of the line.
   C. an influential point since removing it would decrease the correlation.
   D. an outlier since removing it would decrease the correlation.
   E. an outlier since removing it would increase the correlation.

18. Adding the point (15,15) (it’s above the mean) would
   A. not change any of the bivariate statistics since it’s about on the line.
   B. would decrease the correlation since it’s not at the mean.
   C. would increase the correlation since it’s on the line but not at the mean.
   D. would decrease the correlation since it’s way off the line.
   E. would change the equation of the line but not the correlation.

19. Which of the following is true about the desirable properties of our sample statistics?
   A. As long as we take a random sample, our statistic(s) will be unbiased.
   B. Increasing the sample size, \( n \), affects all three parts of a distribution: the shape, the center and the spread.
   C. Samples of the same size have about the same variability no matter what size their parent population is (assuming that said population is large enough).
   D. All of the above are true.
   E. None of the above are true.

20. Which of the following types of graphs would be the best at displaying exam scores and responses to a question which has answers A,B,C,D,E?
   A. five separate histograms, one for each answer (A,B,C,D,E)
   B. five side-by-side boxplots, one for each answer (A,B,C,D,E)
   C. five bar charts since the responses are categorical
   D. a scatterplot since we have two variables
   E. a two-way table since we have two variables