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

2. Be sure to mark your CORRECT section number and your test form on the scantron!

3. Sign your name where indicated on your scantron. You may keep this exam.

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

7. 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 AND NOT discuss this exam with anyone until AFTER the grades are posted.

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

A. Bias (or lack thereof) refers to the center of the distribution of a sample statistic.
B. The distribution of the sample statistic is bell-shaped.
C. We can make the spread of the distribution of the sample statistic as small as we want by using a large sample.
D. All of the above are true.
E. Only two of the above are true.

2. Referring to the normal quantile plot (the graph under the histogram, not beside it), if the mean is 13,000 and the standard deviation is 3500, then we could say

A. about 95% of the observations will fall between 6,000 and 20,000.
B. about 95% of the observations will fall between 9,500 and 16,500.
C. all of the observations will fall between 2,500 and 23,500.
D. nothing since we don’t know if the data is normal or not.
E. nothing since the data isn’t normal.

3. Suppose you have bivariate data, \((x_i, y_i)\). If you changed the scale of \(x\), i.e., you multiplied the \(x\) variable by a number, which of the following statistics would be affected?

A. the slope, \(b_1\) of the regression line
B. the intercept, \(b_0\), of the regression line
C. the correlation coefficient of \(x\) and \(y\)
D. all of the above
E. exactly two of the above

or 3. Assuming the \(x\) and \(y\) are linearly related, which of the following will be true if we multiply the \(y\)'s by 10?

A. \(\bar{y}_{\text{new}} = 10\bar{y}_{\text{old}}\) and \(s_{y(\text{new})} = 10s_{y(\text{old})}\)
B. \(b_0(\text{new}) = 10b_0(\text{old})\) and \(b_1(\text{new}) = 10b_1(\text{old})\)
C. \(r_{\text{new}} = 10r_{\text{old}}\)
D. all of the above
E. exactly two of the above

4. What does the coefficient of determination, \(R^2\), represent?

A. the strength of the relationship between any two variables
B. the strength of the relationship between any two numerical variables
C. the strength of the linear relationship between any two variables
D. the strength of the linear relationship between any two numerical variables
E. the proportion of variation that is explained by the regression line

5. Suppose at the end of the semester I calculate the \(z\)-scores for each of your exams using the mean and standard deviation for each, not the overall. Which one should I drop if I’m dropping your lowest exam score?

A. \(z\)-scores are not useful here since your lowest exam score may not be your lowest \(z\)-score.
B. the lowest \(z\)-score
C. the one furthest from 0, above or below, since it is most likely the outlier
D. the one from the exam with the highest average since it is most likely to be your worst
E. the one from the exam with the lowest average since it is most likely to be your worst

or 5. What does it mean to get a \(z\)-score of +3.0 on this exam?

A. It means that you missed 3 questions on the exam.
B. It means that you got 3 times as many questions on this exam correct as the average student.
C. It means that your grade was 3 standard deviations above the average on this exam.
D. It means that your grade on this exam was in the upper 3% of the class.
E. It means that your grade was 3 points higher than the average on this exam.
6. A market research company wishes to find out which of two internet search engines the population of students at a university prefers to use: Google or MSN Search. A random sample of students is selected, and each one is asked to search for a certain subject using Google and then MSN, or vice versa. The order of the two searches was determined at random. They then indicate which internet search engine they prefer. What type of study is this?

A. an observational study
B. an experiment, but not a double-blind experiment
C. a blind experiment
D. a matched-pairs experiment
E. a multistage sample design

7. What are the approximate five number summaries for the histogram? Each bar represents a range of $25,000 and the numbers below mean the bars with that particular lower endpoint.

A. 0,100000,200000,300000,400000
B. 0,50000,100000,20000,400000
C. 0,25000,50000,75000,100000
D. 0,25000,50000,75000,400000
E. 0,12500,25000,50000,40000

or 7. If the five number summaries for a distribution is $(1.72, 6.06, 6.98, 7.08, 8.95)$, outside of what two numbers will the outliers fall?

A. 5.04 and 8.1
B. 5.96 and 8.0
C. 5.45 and 8.51
D. $-9.125$ and 19.795
E. 4.53 and 8.61

8. One potential advantage of a stem-and-leaf plot over a histogram is that

A. it is smaller (takes up less space).
B. it is easier to see a pattern.
C. it contains more (class) intervals.
D. it uses the class midpoints as the stems.
E. it shows the original data.

or 8. Stemplots

A. are only used for small datasets.
B. display the data values.
C. are the only graphs (of those we discussed in class) from which you can determine the median, $\bar{x}$.
D. All of the above are true statements.
E. Only two of the first 3 answers are true statements.

9. Suppose the grade distribution is normal with a mean of 75 and a standard deviation of 10, $X \sim N(75, 10^2)$. Approximately what grade would you have to make to be in the top 10% of the class?

A. more than 1
B. more than 85
C. more than 75
D. more than 90
E. between 75 and 85

10. What’s the first thing you should always do when given a dataset to analyze?

A. calculate the mean
B. plot the data
C. find out if the data is categorical or numerical
D. find out where the data came from
E. find out how much you’ll be paid to analyze the data

11. Which of the following is not a major principle of experimental design?

A. comparative experimentation
B. replication
C. randomization
D. segmentation
E. All of the above are principles of experimental design.

or 11. Which of the following is not a concept used in experimental design?

A. comparative experimentation
B. replication
C. randomization
D. blinding
E. All of the above can be used in experimental design.
12. A researcher reports that, on average, the participants in his study lost 10.4 pounds after two months on his new diet. A friend of yours comments that she tried the diet for two months and lost no weight, so the report must be a fraud. Which of the following statements is correct?

A. Your friend must not have followed the diet correctly since she didn’t lose any weight.
B. The report must not be correct since your friend didn’t lose any weight.
C. We would have to add your friend’s results to the study for the report to be correct.
D. The report only states the average and doesn’t give the individual results, so it could still be correct.
E. Never believe what you read in the paper. It’s always biased.

<table>
<thead>
<tr>
<th>College</th>
<th>AGLS</th>
<th>EDUC</th>
<th>ENGI</th>
<th>LART</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>lcity</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>metro</td>
<td>4</td>
<td>21</td>
<td>1</td>
<td>5</td>
<td>31</td>
</tr>
<tr>
<td>outsid</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>scity</td>
<td>3</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>town</td>
<td>5</td>
<td>14</td>
<td>1</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>57</td>
<td>4</td>
<td>10</td>
<td>86</td>
</tr>
</tbody>
</table>

13. How likely are you to be from a metroplex if you are in the College of Agriculture?

A. 4/86  
B. 4/31  
C. 4/15  
D. 15/86 * 31/86  
E. 4/15 * 4/31

or

13. How likely are you to be in the College of Liberal Arts if you are from a metroplex?

A. 5/10  
B. 5/31  
C. 5/86  
D. 10/86 * 31/86  
E. 5/10 * 5/31

14. How likely are you to be in the College of Education and come from a town?

A. 14/57  
B. 14/23  
C. 14/86  
D. 14/57 * 14/23  
E. 14/(57 * 23)

15. Referring to the boxplots which represent the sale price for houses in neighborhoods C, D, E, F and G, which of the following is/are true?

A. The houses in neighborhood D are more expensive than all of the other houses.
B. The sale price of houses in neighborhood C are more variable than in neighborhood D.
C. Most houses in neighborhood E sold for more than $60,000.
D. All of the above are true.
E. Only two of the above are true.

16. A standardized test designed to measure math anxiety has a mean of 100 and a standard deviation of 10 in a population of first year college students. If we assume the shape of the distribution is normal, which of the following observations would be considered an outlier, i.e., not a probable part of this distribution? (Hint: use one of the Rules for summarizing data sets.)

A. 140  
B. 135  
C. 100  
D. exactly two of the above  
E. none of the above would be an outlier since normal distributions don’t have outliers

17. Which of the following best describes random sampling?

A. all of the observations from the population are equally spread out in the sample  
B. all possible samples of size n are equally likely  
C. all possible samples of size n are normally distributed  
D. all of the observations from the population are normally distributed  
E. all possible samples of size n are equally spread out
17. Referring to the histogram, which of the following is/are true? Each bar represents a range of $25,000 and the numbers below mean the bars with that particular lower endpoint.

A. The median is between $25,000 and $50,000 and the mean is larger since the data is skewed to the right.
B. The median is between $50,000 and $75,000 and the mean is larger since the data is skewed to the right.
C. The median is between $50,000 and $75,000 and the mean is smaller since the data is skewed to the right.
D. The median is between $25,000 and $50,000 and the mean is larger since the data is skewed to the left.
E. The median is between $25,000 and $50,000 and the mean is smaller since the data is skewed to the left.

18. Which of the following is/are true?

A. It’s impossible to get an accurate or complete census.
B. Matched-pairs design can only be used on twins.
C. Control, as a basic principle of experimental design, means we measure all possible variables so that there are no lurking variables.
D. All of the above are true.
E. None of the above are true.

19. Referring to the scatterplot (the one beside the histogram), how would you describe the correlation between appraised land value and sale price?

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

20. What does the 3.92 in the regression equation mean?

A. For every dollar increase in appraised land value, the sale price goes up $1.
B. For every dollar increase in appraised land value, the sale price goes up $3.92.
C. The average sale price is 3.92 times the average appraised land value.
D. The average sale price equals 3.92 times the average appraised land value, since plugging in the average appraised land value gives you the average sale price.
E. None of the above describe the 3.92 correctly.