1. Do not open this exam until you are told to do so.

2. There are 20 multiple-choice questions on this exam, each worth the same amount. Please mark your answers clearly on a GRAY Scantron sheet. Multiple marks will be counted wrong.

3. You must mark your Scantron form with
   (a) Your NAME and UIN.
   (b) Your correct SECTION (Thursday 11:10 is 507, 12:45 is 508, 2:20 is 509, and 3:55 is 510).
   (c) This test FORM (A, B, C, or D).
   (d) Your Form letter which is above.

4. You will have only 50 minutes to finish this exam.

5. You may use the following:
   (a) One 8$\frac{1}{2}$ × 11 formula sheet (both sides) of your own making.
   (b) A copy of the CI and HT handout.
   (c) A copy of the Z tables.
   (d) A copy of the categorical data handout.
   (e) A stand-alone calculator, i.e., one that cannot communicate with the internet or anything outside itself.

6. You must put all possessions, besides the materials listed and your scantron, pencil(s) and eraser, along the walls or at the front of the room out of everyone else’s way. This includes cell phones, which must be turned off!

7. If you have questions, please write out what you are thinking on this test so that we can discuss it after your results are returned to you.

8. 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.

9. When you are finished please make sure you have marked your Section and Form and have an answer for every question, then turn in your scantron and show your ID.

10. Good luck!
1. A commonly prescribed drug on the market for relieving nervous tension is believed to be 60% effective. Experimental results with a new drug administered to a random sample of 100 adults who were suffering from nervous tension showed that 72 of them showed improvement. Is there sufficient evidence that the new drug is superior to the one commonly prescribed? Which set of hypotheses is the most appropriate for this situation? Which set of hypotheses is the most appropriate for this situation? Which set of hypotheses is the most appropriate for this situation? Which set of hypotheses is the most appropriate for this situation? Which set of hypotheses is the most appropriate for this situation? Which set of hypotheses is the most appropriate for this situation? Which set of hypotheses is the most appropriate for this situation? Which set of hypotheses is the most appropriate for this situation?

A. \( H_0 : p \leq 0.72 \) vs. \( H_A : p > 0.72 \)
B. \( H_0 : p = 0.72 \) vs. \( H_A : p \neq 0.72 \)
C. \( H_0 : \pi = 0.72 \) vs. \( H_A : \pi \neq 0.72 \)
D. \( H_0 : \pi = 0.6 \) vs. \( H_A : \pi \neq 0.6 \)
E. \( H_0 : \pi \leq 0.6 \) vs. \( H_A : \pi > 0.6 \)

2. Which of the following is true?

A. You should use a larger \( \alpha \), say 0.10, if you want to reduce the chance of a Type II error.
B. The smaller \( \alpha \)-level you use, the more evidence you need to reject the null.
C. When using a confidence interval to decide whether to reject or not in a 2-sample test, the hypothesized value is always 0.
D. All of the above are true.
E. Only two of the above are true.

3. Suppose that a 90% confidence interval for the true proportion of A&M students who watch \textit{Survivor} is (0.22, 0.53) and one for the students is (0.29, 0.42). From this information we can conclude

A. more Aggies watch the show than the u. students.
B. more the u. students watch the show than Aggies.
C. it’s plausible that the true proportions are the same.
D. we have more confidence in the A&M interval since it’s wider.
E. Two of the above are correct.

4. Which of the following best describes a categorical variable?

A. The data can only be described with words, never numbers.
B. You can calculate a median but not a mean for the data.
C. You can not calculate anything from this data; you can only display it in some graphs.
D. Exactly two of the above are correct.
E. None of the above are correct.

5. Using the information below, what is the correct range of the \( p \)-value if I wanted to test \( H_0 : \pi_1 = \pi_2 \) vs. \( H_A : \pi_1 \neq \pi_2 \)?

\[
\begin{align*}
90\% & : (-0.015, 0.215) \\
95\% & : (-0.037, 0.237) \\
99\% & : (-0.080, 0.280)
\end{align*}
\]

A. \( p \)-value > 0.10
B. 0.10 > \( p \)-value > 0.05
C. 0.05 > \( p \)-value > 0.01
D. \( p \)-value < 0.01
E. There is no hypothesized value to compare.

6. Which of the following would be necessary to establish a cause-and-effect relation between two variables?

A. a ‘large’ correlation coefficient, \( r \) close to 1
B. a statistically significant result
C. a strong relation under many different settings
D. for the alleged cause to be plausible
E. All of the above are needed to prove causation.

7. Suppose I wanted to test \( H_0 : \pi_1 = \pi_2 \) vs. \( H_A : \pi_1 \neq \pi_2 \) but I calculated a 95% confidence interval instead. Which of the following would be true?

A. If the 95% confidence interval contained 0, then I would reject \( H_0 \) at the 5 and 1% levels.
B. If the 95% confidence interval contained 0, then I would reject \( H_0 \) at the 5 and 10% levels.
C. If the 95% confidence interval contained 0, then I would fail to reject \( H_0 \) at the 5 and 1% levels.
D. If the 95% confidence interval contained 0, then I would fail to reject \( H_0 \) at the 5 and 10% levels.
E. If the 95% confidence interval contained 0, then I would reject \( H_0 \) at the 5% level only.

8. Which of the statements does not contain a statistical blunder?

A. There is a strong negative correlation between a person’s gender and the amount that he or she pays for automobile insurance. Males must pay more for their insurance.
B. Since men are, on average, taller than women, the correlation between men and women’s heights is positive.
C. The mean height of young women is 64 inches, and the correlation between their heights and weights is 0.6 inches.
D. The correlation between height and weight for adult females is about \( r = 1.2 \).
E. All four statements contain blunders.
9. Which best describes the correlation between MDadColl, x, and MMomColl, y?
   A. strongly negative
   B. moderately negative
   C. weak
   D. moderately positive
   E. strongly positive

10. Years of college for males’ moms, MMomColl, is on the y-axis, and years of college for males’ dads, MDadColl, is on the x-axis. If we switched the axes, let x be y and vice versa, which of the following would be true?
   A. The new slope would be 1 over the old slope (the inverse).
   B. The new slope would be the old intercept and the new intercept would be the old slope (they would switch position since x and y switched).
   C. The correlation would be exactly the same.
   D. The new correlation would be 1 over the old correlation.
   E. Two of the above are true.

11. Looking at the original plot (not the switched one), if we add a mom with 3 years of college and pair it with a dad with 5 years of college. What would this do to the correlation, slope and intercept?
   A. Since the point (5,3) would be off of the line, the correlation would decrease, and the slope and intercept would change some.
   B. Since the point (5,3) is about the mean, none of the 3 statistics would change much.
   C. Since the point (3,5) would be off of the line, the correlation would decrease, but the slope and intercept would not change.
   D. Since the point (3,5) would be on of the line, the correlation would decrease, the slope decrease some and intercept would increase some.
   E. Since the point (3,5) would be off of the line, the correlation would decrease, the slope decrease some and intercept would increase some.

12. Suppose the true proportion of Aggies that agree with the verdict in a recent trial is 70%. Joe did a hypothesis test and concluded, based on his sample, that the proportion is significantly greater than 70% at the 0.01 level, then he
   A. committed a Type I error.
   B. committed a Type II error.
   C. made a correct decision.
   D. performed an invalid test.
   E. used the wrong α level.

13. Suppose the p-value for testing $H_0 : \pi_1 = \pi_2$ vs. $H_A : \pi_1 > \pi_2$ is 0.12. Which of the following is true?
   A. There is a 12% probability that $H_0$ is true ($H_A$ is false).
   B. There is a 12% probability of obtaining a difference between $p_1$ and $p_2$ at least as big as this one if $\pi_1$ and $\pi_2$ are really equal.
   C. There is a 12% probability of obtaining a difference between $p_1$ and $p_2$ smaller than this one if $\pi_1$ and $\pi_2$ are really different.
   D. There is a 12% probability of obtaining a difference between $p_1$ and $p_2$ at least as big as this one if $\pi_1$ and $\pi_2$ are really different.
   E. There is a 12% probability of obtaining a difference between $p_1$ and $p_2$ at least as big as this one if $\pi_1$ and $\pi_2$ are really different.

14. In the testing procedure for a two-sided $H_A (\neq)$, we rejected $H_0$ at the $\alpha = 0.05$ level. Sampling from the same population and using the same and set of hypotheses
   A. $H_0$ might not be rejected at the $\alpha = 0.01$ level.
   B. $H_0$ might not be rejected at the $\alpha = 0.10$ level.
   C. the p-value will always be less than $\alpha$.
   D. Two of the above are true.
   E. All of the above are true.
15. Here is a two-way table of the numbers of college students (U.S. citizens only) classified by racial/ethnic group and by whether they attend a public or a private college. The entries are in thousands of students.

<table>
<thead>
<tr>
<th></th>
<th>Public</th>
<th>Private</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Hispanic White</td>
<td>7094</td>
<td>1982</td>
<td>9076</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>831</td>
<td>290</td>
<td>1121</td>
</tr>
<tr>
<td>Hispanic</td>
<td>336</td>
<td>47</td>
<td>383</td>
</tr>
<tr>
<td>Asian</td>
<td>166</td>
<td>32</td>
<td>198</td>
</tr>
<tr>
<td>Native American</td>
<td>68</td>
<td>9</td>
<td>77</td>
</tr>
<tr>
<td>Total</td>
<td>8495</td>
<td>2360</td>
<td>10855</td>
</tr>
</tbody>
</table>

How likely is a public college student to be non-Hispanic black?
A. 831/10855
B. 831/1121
C. 831/8495
D. 1121/10855
E. 8495/10855

16. Using the same table, how likely is a student to be native American and attend a private college?
A. 9/10855
B. 9/77
C. 9/2360
D. 77/10855
E. 2360/10855

17. Suppose the $\chi^2$ test for independence using the previous table had a $p$-value = 0.384. Which of the following would be the correct conclusion?
A. There is no difference in the true proportions of racial/ethnic groups at public and private colleges.
B. There is a difference in the true proportions of racial/ethnic groups at public and private colleges.
C. Private and public colleges are independent.
D. Racial/ethnic groups are independent at both public and private colleges.
E. None of the above are correct conclusions.

18. If the correlation, $r = 0$, then it could be
A. the scatterplot of $x$ and $y$ shows no pattern (just a random scatter of points).
B. $y$ is constant ($s_y = 0$).
C. there is a strong non-linear relationship between $x$ and $y$.
D. All of the above are possible.
E. Only two of the above are possible.

19. An article in the Battalion said “Social drinkers make more money than abstainers”. This is most likely true because
A. drinkers are more upwardly mobile than abstainers.
B. all CEO’s are drinkers.
C. high paying jobs are stressful and cause everyone to drink.
D. there’s a lurking variable to which both drinking and income are related.
E. It’s not true. The article was fabricated. (a lie)

20. If we wanted to show the relationship between drinking (yes or no) and income, which of the following should we use?
A. a two-way table with drinking as the column variable and income as the row variable
B. side-by-side boxplots of income for drinkers and non
C. a scatterplot of income vs. drinks per week
D. two pie charts since drinking is a categorical variable
E. Any of the above would work.