

NOTE to students: This is intended to give you an idea of the type questions the instructor asks and the approximate length of the exam. It does NOT indicate the exact questions or the topics covered. Students should refer to the information posted on Moodle to determine the coverage of the material.

Instructions:

- Read each question carefully.
- The exam is open book
 - **Only course materials (e.g. lecture notes and videos, quizzes, practice problems) may be used on the exam**
 - **Using the internet to search for answers or post questions from the exam is a violation of academic integrity**
- **You may not work with any other students on this exam; the work submitted must be your own**
- Multiple-choice/True-false questions are worth 3 points each; points for short answer questions are indicated in parentheses at the start of the question.
- For written responses, points are awarded for answers that are complete and show complete statistical reasoning. Points are not awarded based on the length of the answer, so **do NOT add unnecessary or irrelevant information to your answer!**
- Some questions will allow you to type or upload an answer; you may choose one method, you do not need to do both.
 - If you choose to upload your answer, you can scan or photograph a handwritten document (make sure it is legible!) or you can typeset your answer using LaTeX or the Word equation editor (or a similar software)
- Once you start the exam, you will have **2 hours** to complete it
 - Any saved work will be automatically submitted when the timer runs out; if you finish before the 2 hours is up, you should submit your work (like you would with a quiz)
 - Any saved work will be automatically submitted when the exam window closes, even if the 2 hour window is not up

Honor Pledge:

I certify that I have not received or given unauthorized aid in taking this exam.

Signed: _____

1. **True False** The Chi-square distribution is generally skewed to the right.
2. **True False** When the decision of a hypothesis test is “do not reject H_0 ”, a Type II Error was made.
3. **True False** The appropriate degrees of freedom for a multiplier for a confidence interval based on paired data is $n_1 + n_2 - 2$.
4. When is it appropriate to use the pooled version of the 2-sample t-test or confidence interval?
 - a. When the sample sizes are large.
 - b. When the sample sizes are equal.
 - c. When the population variances are large.
 - d. When the population variances are equal.
5. A researcher conducted a study to see if an energy education program for 6th grade children would reduce the amount of electricity their families used. The study found that the values were significantly lower (p-value = 0.0003) with an average yearly savings of \$0.37 per household. This would be a case in which the results are
 - a. Both statistically and practically significant.
 - b. Practically significant but not statistically significant.
 - c. Statistically significant, but not practically significant.
 - d. Neither statistically significant nor practically significant.
6. A biologist is collecting data to test the hypothesis of $H_0: p_1 - p_2 = 0$ vs. $H_A: p_1 - p_2 > 0$ (a one-sided test) based on a random sample of 50 units (25 from each population). Another biologist suggests he use $H_0: p_1 - p_2 \neq 0$ (a two-sided test) and makes the following statement: “If we would reject the null hypothesis using a one-sided test, then we would also reject it using a two-sided test on the same data.” Is he correct?
 - a. Yes, if we reject using a 1-sided test we would also reject using a 2-sided test.
 - b. Yes, the alternative does not make a difference if we will reject or not.
 - c. No, we may or may not reject the null hypothesis if we are using a two sided alternative.
 - d. No, we would never reject the null hypothesis.
7. Consider the previous problem. Which null distribution would be appropriate for the biologist to use to find the p-value?
 - a. Standard normal distribution
 - b. t-distribution with 24 degrees of freedom
 - c. t-distribution with 48 degrees of freedom
 - d. It is impossible to determine without knowing which alternative hypothesis the biologist picked.
8. Which of the following conditions are required for a hypothesis test for the difference in population proportions to be valid?
 - a. The sample sizes must be sufficiently large ($n_1 \geq 30$ and $n_2 \geq 30$).
 - b. The sample sizes must be sufficiently large ($n_1\hat{p}_1, n_1(1 - \hat{p}_1), n_2\hat{p}_2, n_2(1 - \hat{p}_2)$ all at least 10).
 - c. The populations must be normally distributed.
 - d. The populations must be random.

9. Statistical *power* is defined as
- The probability of finding an effect if one exists.
 - The chance of not making a type 1 error.
 - The chance of not making a type 2 error.
 - Both a. and c. but not b.
10. The best way (within the researcher's control) to increase the power of a test is to
- Increase the sample size.
 - Decrease the significance level.
 - Increase the size of the effect you are looking for.
 - Both a. and c. but not b.

Use the following for questions 11 and 12: A large corporation manufactures laundry detergent in boxes that are advertised to be 42 ounces. The quality control manager is concerned that the boxes are being filled with more than 42 ounces of detergent (this would cost the company extra money). The manager takes a random sample of 200 boxes from the most recent production lot. The resulting sample mean was 42.7 ounces and the sample standard deviation was 7.3 ounces.

11. What is the alternative hypothesis the manager would like to test?
- $H_a: \mu = 42$
 - $H_a: \mu = 42.7$
 - $H_a: \mu > 42$
 - $H_a: \mu < 42.7$
12. Calculate the value of the test statistic for this test.
- 1.356
 - 0.186
 - 1.356
 - 82.7

Use the following for questions 13 and 14: A researcher obtained a random sample of 120 drivers from across the U.S. and recorded for each driver their seat belt usage and the region where they live. The researcher wishes to use the resulting data (shown in the table below) to answer the question: Is there is an association between seat belt usage and region?

	REGION				Total
	Midwest	Northeast	South	West	
SEAT BELT No	7	5	12	7	31
USAGE Yes	34	15	8	32	89
Total	41	20	20	39	120

13. Which procedure would be best to address the researcher's question?
- Chi-square test of homogeneity
 - Chi-square test of independence
 - Simple linear regression analysis
 - Paired samples t-test
14. If there is no association, how many drivers from the South would we expect to wear a seatbelt?
- 14.83
 - 0.067
 - 5.17
 - 8

15. A manufacturer claims 50% of candies they produce are brown and that candy pieces are randomly placed into bags. Sam buys a large family size bag of these candies and Kerry buys a small fun size bag. Which bag is more likely to have more than 70% brown candies?
- Sam's, because a larger bag is more likely to have a larger proportion of brown candies.
 - Kerry's, because there is more variability in proportions of colors among smaller samples.
 - Both have the same chance because the bags they buy are both random samples of candy pieces.

Short answer: Give one answer in the space provided.

16. (6 points) One analogy often used to describe statistical hypothesis testing is that of a criminal trial. Under this analogy, the hypotheses are: H_0 : the defendant is innocent vs. H_A : the defendant is guilty. Describe a type 1 error in the context of this analogy.

17. Does sugar content for cereal differ by the supermarket shelf that the cereal is on? Data was collected and used to create the partial ANOVA table. Use it to answer the questions that follow.

	DF	SS	MS	F	p-value
Shelf	2		122.1		0.0012
Error					
Total	76	1486.7			

- (1 point) How many shelves were considered in this study?
- (1 point) How many subjects (types of cereal) were considered in this study?
- (4 points) Fill in the blanks in the ANOVA Table above. Show your work.

Problem 17 continued:

- d. (3 points) Draw a well-labeled picture of the p-value provided in the ANOVA table.

- e. (6 points) Would it be appropriate to conduct a multiple comparison's procedure for this data? Answer Yes or No and explain your reasoning. If it is appropriate, use the output at the right to do so; explain your reasoning.

shelf Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
2 – 3	3.0913	0.4002	5.7823
2 – 1	4.7690	1.7070	7.8311
3 – 2	-3.0913	-5.7823	-0.4002
3 – 1	1.6778	-1.0554	4.4110
1 – 2	-4.7690	-7.8311	-1.7070
1 – 3	-1.6778	-4.4110	1.0554

18. (14 points) A large university is planning to build a new parking garage on campus that will be paid for by a larger student parking fee. The student newspaper wanted to determine if the majority of students favor the new garage. They email a random sample of 100 students and find that 57 of them are in favor of the new garage. Conduct an appropriate test of hypothesis using a 5% significance level. Be sure to include all necessary steps starting with Step 2 (to save time, you may skip Step 1). Note: the p-value is 0.081.

19. (20 points) A survey of undergraduate students at a large east coast university examined many questions regarding student's health and lifestyles. The survey recorded information on 575 students. The survey recorded the student's biological sex and the number of hours of exercise they perform each week. The resulting answers were used to create the following table. Does this data indicate that there is a difference in the average amount of exercise based on biological sex (Male – Female)? Consider both the results of the hypothesis test and the confidence interval to address this question. Use a 5% significance level and a 95% confidence level. Again, to save time, you may skip Step 1 of the test. Note: the p-value is 0.0002 and the multiplier is 1.96.

Sex	Mean	St .Dev	n
Male	6.5	5.8	235
Female	4.8	4.3	340