ST 314 Final Exam

SPRING 2023

Name: Ya Zou

Student ID Number: 934-417-686

Instructions

- You will have exactly **80 minutes** to complete the exam.
  - Do not open the exam until you are told to do so.
  - You must stop working when Erin announces the exam is over.
  - If you finish early, please hand your exam to Erin or one of the TAs.
- There are 17 questions on this exam.
- You may use any hard copy/paper notes during the exam.
- You MAY NOT use any digital notes, a calculator, computer, tablet, or phone during the exam.
- You are NOT expected to work out calculations on the exam.
  - For example, if an answer to a question is found by doing $10 + \frac{85}{\sqrt{150}}$, you may leave it as the expression previously written. Do not spend time trying to work out the simplified answer to a complicated expression.
- For multiple choice questions, please fill in the bubble corresponding to the answer you wish to select.
  - For example, to select answer A, mark:

● A
○ B
○ C
○ D

1. (1 point) A manufacturing company that produces electronic devices at numerous different facilities wants to optimize the production process by identifying the factors that significantly influence the quality of the final product. Quality is measured in the defect rate of devices produced at each facility. In addition to defect rate, the company has information on temperature during manufacturing, humidity during manufacturing, speed of the assembly line, and duration of the manufacturing process. The company would like to use all four of these variables to model defect rate.

   Which statistical procedure may be appropriate in this scenario?

   - ○ Simple linear regression
   - ○ One-sample z-test for proportions
   - ● Multiple linear regression
   - ✗ Single-factor ANOVA

2. (1 point) The Oregon Department of Transportation (ODOT) wants to examine the safety of different-sized cars when involved in a crash. ODOT collects a random sample of 30 cars in each of the following categories: compact, midsize, and full-size. Each car in the study is put through a safety test to assess the pressure applied to a dummy driver's head during a crash.

   Which test is appropriate to see whether the average pressure applied to the driver's head during a crash test is different between car types?

   - ✗ Simple linear regression
   - ○ One-sample z-test for proportions
   - ○ Multiple linear regression
   - ● Single-factor ANOVA

Use the following information for questions 3-6

A construction company wants to compare the strength of two different types of concrete mixtures for a particular project. They are considering using either a conventional concrete mixture or a new experimental concrete mixture. To evaluate the strength, the construction company takes 35 random samples of concrete specimens from each mixture type. They perform compression tests on the specimens to measure their strength in pounds per square inch (psi).

Sampled information from the 70 concrete specimens is displayed in the table below.

Group	Sample Mean	Sample Standard Deviation	Sample Size
Conventional Mixture	3206 psi	401 psi	35
Experimental Mixture	3744 psi	355 psi	35

3. (1 point) State the null hypothesis for testing whether the average strength of the experimental mixture is greater than the average strength of the conventional mixture.

- ○ $H_0: \bar{x}_{\text{experimental}} - \bar{x}_{\text{conventional}} < 0$
- ○ $H_0: \bar{x}_{\text{experimental}} \neq \bar{x}_{\text{conventional}}$
- ○ $H_0: \mu_{\text{experimental}} - \mu_{\text{conventional}} > 0$
- ◉ $H_0: \mu_{\text{experimental}} - \mu_{\text{conventional}} = 0$

4. (1 point) State the alternative hypothesis for testing whether the average strength of the experimental mixture is greater than the average strength of the conventional mixture.

- ○ $H_A: \bar{x}_{\text{experimental}} - \bar{x}_{\text{conventional}} > 0$
- ○ $H_A: \bar{x}_{\text{experimentall}} \neq \bar{x}_{\text{conventional}}$
- ◉ $H_A: \mu_{\text{experimental}} - \mu_{\text{conventional}} > 0$
- ○ $H_A: \mu_{\text{experimental}} - \mu_{\text{conventional}} = 0$

5. (1 point) Calculate the point estimate for the difference in average strength between the experimental and conventional mixtures.

You do not need to compute the final answer. Your answer can be a mathematical expression. Please write your answer ***in the box below.***

3744 − 3206

6. (1 point) Which of the following expressions correctly represents the test statistic for this hypothesis test?

- ○ $z = \frac{3744-3206}{\sqrt{\frac{355(1-355)}{70}}}$
- ● $t = \frac{3744-3206}{\sqrt{\frac{355^2}{35}+\frac{401^2}{35}}}$
- ○ $t = \frac{3744-3206}{\sqrt{70}}$
- ○ $F = \frac{3744-3206}{355-401}$

Use the following information for questions 7-9

Interested in assessing the efficacy of different teaching styles on test performance, researchers randomly divide 200 students in an introductory biology course into 5 groups of 40 students. Each group is taught the same material using a different teaching method, and their mean test scores are compared using an analysis of variance.

7. (1 point) State the null hypothesis used to test whether there is a difference in average test scores between the five teaching methods.

- ● The average test scores for all five teaching methods are the same.
- ○ The average test score for each teaching method is 0.
- ○ All of the average test scores for the five teaching methods differ from one another.
- ○ At least two teaching methods differ in average test scores.

8. (1 point) Identify the null distribution of the test statistic used in this single factor ANOVA F test.

- ○ *t* distribution with 199 degrees of freedom.
- ○ *t* distribution with 39 degrees of freedom.
- ● *F* distribution with 4 and 195 degrees of freedom.
- ○ *F* distribution with 39 and 199 degrees of freedom.

F = 3.59 0.0076.

9. (1 point) The test statistic for this single factor ANOVA F test is 3.59 and the p-value is 0.0076. Using this information, identify the correct conclusion for the hypothesis test.

- ○ There is convincing evidence to suggest that the average test scores for all five teaching methods differ from one another.
- ● There is convincing evidence to suggest that at least one of the teaching methods yields an average test score that differs from the average test score of at least one other teaching method.
- ○ There is no evidence to suggest that the average test scores for all five teaching methods differ from one another.
- ○ There is no evidence to suggest that at least one of the teaching methods yields an average test score that differs from the average test score of at least one other teaching method.

Use the following information for questions 10-11

A rocket motor is manufactured by bonding together two types of propellants, an igniter and a sustainer. A random sample of 26 specimens is used to investigate the relationship between the shear strength of the bond in psi and the age of the propellant in weeks. The estimated linear regression model is $\hat{y} = 2609.2 - 42.3x$ where y represents the shear strength of the bond in psi and x represents the age of the propellant in weeks.

10. (1 point) Interpret the estimate for the **slope** coefficient.

- ○ When the age of the propellant is 0 weeks, the predicted shear strength is 2609.2 psi.
- ○ When the age of the propellant is 0 weeks, the predicted shear strength is -42.3 psi.
- ○ For every one week increase in propellant age, shear strength is predicted to increase by 2609.2 psi.
- ● For every one week increase in propellant age, shear strength is predicted to decrease by 42.3 psi.

11. (2 points) Calculate the residual for a propellant that is 15 weeks old with a shear strength of 1968 psi.
You do not need to compute the final answer. Your answer can be a mathematical expression. Please write your answer ***in the box below.***

$e_i = y_i - \hat{y}_i = 1968\,psi - (2609.2 - 42.3(15))$

Use the following information for questions 12-15

A security company that develops a new antivirus software claims that their software can successfully detect and remove a specific type of malware with an accuracy rate of at least 95%. To validate this claim, they conduct a study in which they randomly select a sample of 500 computers and install the malware on those systems. These 500 infected systems are used to test the antivirus software. From the sample of 500 computers, they find that 472 are successful in detecting and removing the specific type of malware. That is, from the sample, the proportion of successful uses of the software was 0.944.

12. (2 points) Write the null and alternative hypotheses needed to test whether the proportion of successful uses of this antivirus software is significantly different than the value claimed by the company.

H_0: $p = 0.944$ H_A: $p \neq 0.944$

p

$0.538 > 0.05$

13. (1 point) The p-value for the hypothesis test performed to test whether the proportion of successful uses of this antivirus software is significantly different than the value claimed by the company is 0.538. What information does this p-value provide?

- ○ The p-value gives the probability of observing a sample proportion of successful uses as or more extreme than what was observed (i.e. less than 0.944 or greater than 0.956), when the true proportion of all successful uses of the software is 0.95.
- ○ The p-value gives the probability of observing a sample proportion of successful uses as or more extreme than what was observed (i.e. less than 0.538), when the true proportion of all successful uses of the software is 0.944.
- ○ The p-value states that there is a 53.8% chance that, from a random sample of 500 uses of the antivirus software, the proportion of successful uses is less than or equal 0.95 when the true proportion of successful uses is 0.944.
- ◉ The p-value states that there is a 95% chance that from a random sample of 500 uses of the antivirus software, the proportion of successful uses is less than or equal 0.538 when the true proportion of successful uses is 0.95.

14. (2 points) The firm takes a random sample of 500 uses of the antivirus software and finds that 472 of them (a proportion of 0.944) are successful in detecting and removing malware. Construct a 99% confidence interval for the true proportion of uses of the software that are successful in detecting and removing malware.

You do not need to do any of the calculations – just set up the construction of the confidence interval. Please write your answer ***in the box below.***

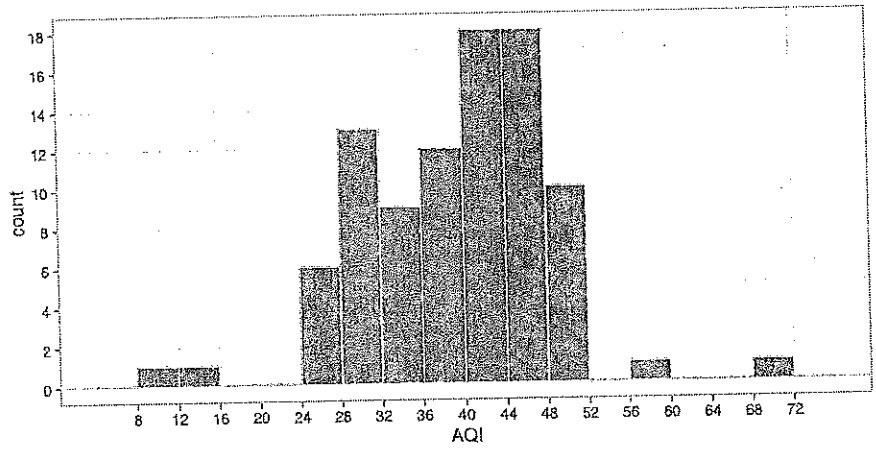
$$\frac{472}{500} \pm 2.574 * \sqrt{\frac{0.944 \cdot (1-0.944)}{500}}$$

Use the following information for questions 15-16

Daily air quality is measured by the air quality index (AQI) reported by the Environmental Protection Agency. Values of this index range from 0 to 300 and a higher value indicates lower air quality.

AQI was reported for 90 randomly sampled days in Durham, NC over the past five years. The histogram below shows the distribution of these values. These 90 days represent a reasonably random and representative sample, and you may assume that the AQI levels for these days are independent.

The average AQI for these 90 days is 39.26, with a standard deviation of 9.13.



15. (1 point) The point estimate for the average AQI in Durham, NC over the last five years is

- 90
- 9.13
- 150
- 39.26

16. (1 point) The sample mean AQI from a sample of 90 days is a random variable that has some standard error associated with it. Based on the sampled data, what is the estimate for the standard error of the sample mean AQI?

- $\frac{64}{\sqrt{90}}$
- $\frac{39.26}{\sqrt{300}}$
- $\frac{9.13}{\sqrt{90}}$
- $\sqrt{\frac{9.13(1-9.13)}{90}}$

17. (1 point) Let y = sales at a fast-food restaurant (in thousands of dollars), x_1 = number of competing restaurants within a 1-mile radius, x_2 = population within a 1-mile radius (in thousands of people), and x_3 be an indicator variable that equals 1 if the restaurant has a drive-up window and 0 otherwise. Suppose that the estimated regression model from a sample of 47 restaurants is $\hat{y} = 10 - 1.2x_1 + 6.8x_2 + 15.3x_3$.

Consider two restaurants. Both have 3 competing restaurants within a 1-mile radius and both have a drive-up window. The difference between these restaurants is that one has population of 2,500 within a 1-mile radius, while the other has a population of 3,500 in a 1-mile radius.

Which of the following statements is true regarding the expected sales when comparing the two restaurants?

- ○ The restaurant with a population of 3,500 within a 1-mile radius is expected to earn \$12,00 less in sales than the restaurant with a population of 2,500 within a 1-mile radius.
- ○ The restaurant with a population of 2,500 within a 1-mile radius is expected to earn \$15,300 less in sales than the restaurant with a population of 3,500 within a 1-mile radius.
- ◉ The restaurant with a population of 3,500 within a 1-mile radius is expected to earn \$6,800 more in sales than the restaurant with a population of 2,500 within a 1-mile radius.
- ○ The restaurant with a population of 2,500 within a 1-mile radius is expected to earn \$6,800 more in sales than the restaurant with a population of 3,500 within a 1-mile radius.

$\hat{y} = 10$.