

Math 222: Exam 2
Spring – 2019
04/12/2019
50 Minutes

Name: _____

Write your name on the appropriate line on the exam cover sheet. This exam contains 9 pages (including this cover page) and 4 questions. Check that you have every page of the exam. Answer the questions in the spaces provided on the question sheets. Be sure to answer every part of each question and show all your work. If you run out of room for an answer, continue on the back of the page — being sure to indicate the problem number.

| Question | Points | Score |
|----------|--------|-------|
| 1 | 12 | |
| 2 | 16 | |
| 3 | 20 | |
| 4 | 32 | |
| Total: | 80 | |

1. (12 points) Mark each of the following statements as True (T) or False (F).
- (a) _____ A χ^2 -analysis is used to analyze categorical variables.
 - (b) _____ If a variable used in a regression is not significant, then the value of its coefficient must be 'approximately' 0.
 - (c) _____ If the coefficient for a variable in a regression is approximately 0, then the variable is insignificant for the model.
 - (d) _____ The degrees of freedom for a χ^2 -test with 8 rows and 6 columns is 48.
 - (e) _____ In performing an F-test for a regression which examined 5 variables and 31 people, if the F-statistic was 3.72, then using the F-table the p -value would be 0.025.
 - (f) _____ For an F-test with 2 degrees of freedom in the numerator and 366 degrees of freedom in the denominator, if the F-statistic was 3.72, then the p -value is 0.050.
 - (g) _____ For a regression with an F-test having 6 degrees of freedom in the numerator and 37 degrees of freedom in the denominator, 43 observations must have been used in making the regression.
 - (h) _____ For a χ^2 -goodness of fit test, there is only one row in the expected value table.
 - (i) _____ For a χ^2 -goodness of fit, the degrees of freedom is the number of columns.
 - (j) _____ The null hypothesis for an F-test with two variables is $H_0 : \beta_1 = \beta_2 = 0$ while the alternative hypothesis is $H_a : \beta_1 = \beta_2 \neq 0$.

2. A guidance counselor at a high school is trying to determine if there is a relationship between gender and whether a graduating student ends up choosing a STEM or a non-STEM major upon graduating and entering college. They collect data which is summarized in the table below.

| | STEM | Non-STEM | Total |
|--------|------|----------|-------|
| Female | 91 | 75 | 166 |
| Male | 70 | 84 | 154 |
| Total | 161 | 159 | 320 |

- (a) (8 points) Complete the expected value table (on the left) and the χ^2 -contribution table (on the right).

| | STEM | Non-STEM |
|--------|------|----------|
| Female | | |
| Male | | |

| | STEM | Non-STEM |
|--------|------|----------|
| Female | | |
| Male | | |

- (b) (6 points) Perform a χ^2 -test for this analysis by giving the null and alternative hypothesis (in the problem context), the degrees of freedom, test statistic, p -value, and stating your conclusion. Use $\alpha = 0.05$.

$$\left\{ \begin{array}{l} H_0 : \\ \\ \\ \\ \\ H_a : \end{array} \right.$$

degrees of freedom: _____

test statistic: _____

p -value: _____

Conclusion:

- (c) (2 points) What are the assumptions for a χ^2 -test? Does this χ^2 -test meet these assumptions?

3. Certain isotopes of sodium are used in cardiovascular traces. A researcher has an unknown isotope of sodium. Starting with 1 g of the substance, they record the natural log of how much remains over a period of approximately 36 hours. Using algebraic manipulation, the slope of the linear model can be used to determine the half-life, and thus the isotope. The data for their linear fit is found below.

Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|------------|----|--------|---------|---------|---------|
| Regression | 1 | 6.5721 | 6.57207 | 1663.86 | 0.000 |
| Time | 1 | 6.5721 | 6.57207 | 1663.86 | 0.000 |
| Error | 27 | 0.1066 | 0.00395 | | |
| Total | 28 | 6.6787 | | | |

Model Summary

| S | R-sq | R-sq (adj) | R-sq (pred) |
|-----------|--------|------------|-------------|
| 0.0628482 | 98.40% | 98.34% | 98.11% |

Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|----------|----------|---------|---------|---------|------|
| Constant | 0.0178 | 0.0254 | 0.70 | 0.488 | |
| Time | -0.04742 | 0.00116 | -40.79 | 0.000 | 1.00 |

- (a) (1 point) How many data points were used in constructing the model?
- (b) (1 point) What percent of variability in the response variable 'Amount' is explained by this model?
- (c) (2 points) What is value of the correlation coefficient?
- (d) (2 points) What is the equation of the linear regression for this data?

-
- (e) (5 points) Construct a 95% confidence interval for β_0 .
- (f) (3 points) Show that $\sum(x_i - \bar{x})^2 \approx 2935.4$.
- (g) (6 points) Construct a 96% confidence interval for the natural log of the amount of sodium left after 12 hours. [The average amount of hours used in the model was 19.4]

4. An engineer is trying to predict the milage per gallon (MPG) of a new car they are designing. Using data collected from other cars, they create a predictive model based on the number of cylinders, the horsepower, and the weight of a car (all variables which can be controlled during construction). The model is (partially) summarized below.¹

Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|------------|-------|---------|---------|---------|---------|
| Regression | _____ | _____ | 5618.52 | _____ | _____ |
| Cylinders | _____ | 30.4 | 30.41 | 1.69 | 0.194 |
| Horsepower | _____ | 242.7 | _____ | 13.52 | 0.000 |
| Weight | _____ | 1209.1 | 1209.05 | 67.37 | 0.000 |
| Error | _____ | 6963.4 | 17.95 | | |
| Total | 391 | 23819.0 | | | |

Model Summary

| S | R-sq | R-sq (adj) | R-sq (pred) |
|-------|--------|------------|-------------|
| _____ | _____% | 70.54% | 70.20% |

Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|------------|-----------|----------|---------|---------|------|
| Constant | _____ | 0.796 | 57.46 | 0.000 | |
| Cylinders | -0.389 | _____ | -1.30 | 0.194 | 5.66 |
| Horsepower | -0.0427 | 0.0116 | _____ | 0.000 | 4.36 |
| Weight | -0.005272 | 0.000642 | -8.21 | _____ | 6.49 |

The regression equation is

$$\text{MGP} = 45.737 - 0.389 \text{ Cylinders} - 0.0427 \text{ Horsepower} - 0.005272 \text{ Weight}$$

¹Dua, D. and Graff, C. (2019) UCI Machine Learning Repository. <http://archive.ics.edu/ml>. Irvine, CA. University of California School of Information and Computer Science.
Quinlan, R. (1993). Combining Instance-Based and Model-Based Learning. In Proceedings on the Tenth International Conference of Machine Learning, 236-243, University of Massachusetts, Amherst. Morgan Kaufmann.

-
- (a) (15 points) Fill in the missing table entries for this model.
- (b) (2 points) For this model, what is the value of $\sum (y_i - \bar{y})^2$?
- (c) (2 points) For this model, what is the value of $\sum (y_i - \hat{y}_i)^2$?
- (d) (6 points) Construct a 95% confidence interval for β_2 .
- (e) (3 points) Of the variables used, which are significant predictors for this regression model and which are not. Explain your answer.
- (f) (4 points) Is this regression model significant? To answer this, perform an F -test for this regression model. Be sure to state the null and alternative hypotheses, the degrees of freedom of the numerator and denominator, F -value, p -value, and your conclusions at the 10% significance level.

BONUS. (8 points) Below is a partial ANOVA table for a linear regression model.

| Source | DF | Adj SS | Adj MS | F-Value |
|------------|-----|---------|--------|---------|
| Regression | ___ | 7362.7 | _____ | 16.43 |
| Error | ___ | 10753.8 | _____ | |
| Total | 50 | 18116.5 | | |

Complete the table above. For credit, you must show all your computations in the space below.