PhD Qualifying Exam: Applied Statistics, STA-682 (80 Points)

August 2024

General Instructions

1. There are 4 problems with subparts in this exam.

2. Write only on one side of your answer page. Begin each problem on a new sheet with the problem number noted.

3. You must show all your work and justifications correctly and completely to receive full credit. Partial credit may be given for partially correct solutions.

4. For each problem/subpart, hand in only the answer that you want to be graded. If necessary, cross out any answer(s) you do not want graded. Crossed out work will be ignored. Failure to follow this requirement may result in a zero score for that problem.

5. When finished, please collate all pages according to the problem numbers and then number the pages clearly.

6. If a theorem is applied, you must clearly state the theorem.

7. Any new notations must be defined before you use them.

Problem 1 (22 points).

Consider the following Multiple Linear Regression Model (1) $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ...\beta_k X_{ki} + \epsilon_i$, where i=1 to n and $\epsilon_i \sim N(0, \sigma^2 I)$.

a. (4 points) Express this multiple linear regression model (1) in matrix form. Define the matrices and their dimensions clearly.

b. (4 points) Derive the Least Squares estimate of the parameter matrix used in part (a).

c. (3 points) What important assumption must be made about [X] in order to estimate these parameters? What are the consequences of violating this assumption?

d. (5 points) Derive the Maximum Likelihood estimate of the parameter matrix used in part (a).

e. (3 points) Compare the results of parts (c) and (d). What are the relationships between the LSE and MLE estimates?

f. (3 points) Is it necessary that the response $[Y] = [Y_1, Y_2 ... Y_n]'$ have independent elements?

Problem 2 (20 points).

Given the model $Y = XB + \epsilon$, where $\epsilon \sim N(0, \sigma^2 I)$, Y_{nx1}, $X_{nx(k+1)}$, $\beta_{(k+1)x1}$, ϵ_{nx1}, and $p = k+1$, prove the following:

a. (4 points) $b = \hat{\beta} = (X'X)^{-1}X'Y \sim N(\beta, \sigma^2(X'X)^{-1})$

b. (4 points) Use the Quadratic Form and Normal Variables to show $\frac{(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)}{\sigma^2} \sim \chi^2_p$

c. (4 points) Use the Quadratic Form and Normal Variables to show $\hat{\beta}$ and s^2 are independent

d. (4 points) Use the Quadratic Form for Normal Variables to show $\frac{(n-p)s^2}{\sigma^2} \sim \chi^2_{(n-p)}$

e. (4 points) $\overline{Y}$ *and* $\sum_{i=1}^{n} (y_i - \overline{y})^2$ are independent.

Problem 3 (16 points).

For Model (1) given in Problem 1 ($Y = X\beta + \epsilon$):

a. (3 points) Give the Least Squares estimates of the parameters (β, σ^2) in matrix notation.

b. (3 points) Define *estimable* of a linear function of β

c. (8 points) State and prove the Gauss-Markov Theorem

d. (2 points) Interpret β_3 for Model (1).

Problem 4 (22 points).

Suppose Consumer Reports (CR) wants to examine the toner costs of three different brands of laser printers (Brand A, Brand B, Brand C). CR obtains a random sample of three printers from each brand and prints 10,000 identical pages for each printer. Yearly toner costs are provided in the table below.

Cost	Brand_A	Brand_B	Brand_C
	500	600	400
	600	700	500
	700	800	600
sample mean	600	700	500
sample standard deviation	100	100	100

A One-Way Analysis of Variance (ANOVA) model is considered. Assume that these samples were taken from populations with a common variance σ^2 and taken from normal populations.

a.(4 points) Express this ANOVA model in matrix form, where the Design Matrix is full column rank. State the model in general form.

b.(4 points) Determine the least-squares estimate of the parameters. Interpret these estimates.

c.(6 points) Construct the ANOVA table for this model. Specify the sources of variation, sums of squares, and degrees of freedom. Show all calculations.

d.(8 points) We wish to test the Null Hypothesis H_o: The three types of printers have equal mean toner costs.

1. Express the Null Hypothesis as a general linear hypothesis in matrix form.
2. Determine the Test Statistic for the Null Hypothesis.
3. Define the probability distribution that the Test Statistic is based on, including the distribution parameters.
4. Explain how you would test the Null Hypothesis for a decision of Rejection or a Failure to Reject.