Stat 763 Exam 3 April 18, 2025

Name:

1. In the following test scheme

$H_0: A\beta = b$ versus $H_a: A\beta \neq b$
Test statistic $F = \frac{(A\widehat{\beta}-b)'[A(X'X)^{-1}A']^{-1}(A\widehat{\beta}-b)/q}{MSE}$
Reject H_0 if $F > F_\alpha(q, n-p)$

Describe, using your own words, what n, p and q are. (10 points)

n is the number of observations on response y.
p is the number of components in vector β.
q is the number of equations in $A\beta = b$.

2. With observed $y_1, ..., y_n$, write out the formulas for $SSE(\emptyset)$ and $SSE(\beta_0)$. (10 points)

$SSE(\emptyset) = \sum_{i=1}^n y_i^2 = U.SSTO$
$SSE(\beta_0) = \sum_{i=1}^n (y_i - \overline{y})^2 = C.SSTO.$

3. With the same set of observations one obtained tables for two models. Fill in blanks. (20 points)

$y = \beta_0 + \beta_1 x_1 + \epsilon$

Source	SS
Model	58
Error	139
C.Total	197

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$

Source	SS
Model	86
Error	111
C.Total	197

	Type I SS
β_0	3008
β_1	58
β_2	28

For model $y = \beta_0 + \beta_1 x_1 + \epsilon$, $SSE = SSTO - SSI_1 \Longrightarrow SSM = SSI_1 = 58$
For model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, $SSTO = 197$. $SSM = SSI_1 + SSI_2 = 86$.

4. Model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ is reduced by H_0 to $y = \beta_1(x_1 + 2x_2 + 1) + \epsilon$. Find H_0 and write it as $A\beta = b$. (20 points)

Rewrite the reduced model as $y = \beta_1 + \beta_1 x_1 + 2\beta_1 x_2 + \epsilon$.
Clearly it is reduced from the full model by $\beta_0 = \beta_1$ and $\beta_2 = 2\beta_1$.
Thus $H_0: \beta_0 - \beta_1 = 0$ and $2\beta_1 - \beta_2 = 0$.
Or $H_0: A\beta = b$ where $A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -1 \end{pmatrix}$ and $b = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

5. The observations on y and x in three cities, A, B and C are available along with three indicators I_a, I_b, I_c. Based on prior knowledge three polynomial regression models are proposed.

$$\begin{cases} y_a = \beta_{a0} + \beta_{a1}x + \beta_{a2}x^2 + \epsilon \\ y_b = \beta_{a0} + \beta_{b1}x + \beta_{a2}x^2 + \epsilon \\ y_c = \beta_{c0} + \beta_{b1}x + \beta_{a2}x^2 + \epsilon \end{cases}$$

where the first two regression function share the intercepts, the last two regression function share the first order term and all three share the quadratic term. Find the combined model. (20 points)

$y = \beta_0 + \beta_1 x_1 + \beta_2 x^2 + \beta_3 I_c + \beta_4 I_a\, x + \epsilon$ is the combined model. It is equivalent to

$$\begin{cases} y_a = \beta_0 + (\beta_1 + \beta_4)x + \beta_2 x^2 + \epsilon \\ y_b = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon \\ y_c = (\beta_0 + \beta_3) + \beta_1 x + \beta_2 x^2 + \epsilon \end{cases}$$

6. One used this table to test the lack of fit for a regression model.

	DF	MS	F	p-value
Numerator	4	3398	14.80	0.006
Denominator	5	230		

Suppose $n = 11$ and $\sum_i (y_i - \overline{y})^2 = 19883$. Fill out the table and point out the model under the investigation. (20 points)

$SSE = SSLF + SSPE = 4 \times 3398 + 5 \times 230 = 14742.$

Source	DF	SS
Model	1	5141
Error	9	14742
C.Total	10	19883

The model under investigation is $y = \beta_0 + \beta_1 x + \epsilon$.