(SOLUTIONS)

STATISTICS 3340, Midterm Exam, October 30, 2025

Name: ______________________________ **Banner #:** B00__________________

Materials permitted: calculator

Check that your paper has 4 pages. Show your work in the space provided.

Exam is out of 40 points

1. A regression analysis is to be carried out for the model $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$. There are eight observations, with $\boldsymbol{y}' = (4, 3, 2, 1, -1, -2, -3, -4)$ and

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}$$

(7) (a) Find the least squares estimator of $\boldsymbol{\beta}$.

$$X'X = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 8 \end{pmatrix} \quad 2$$

$$(X'X)^{-1} = \begin{pmatrix} 1/8 & 0 & 0 \\ 0 & 1/8 & 0 \\ 0 & 0 & 1/8 \end{pmatrix} \quad 3$$

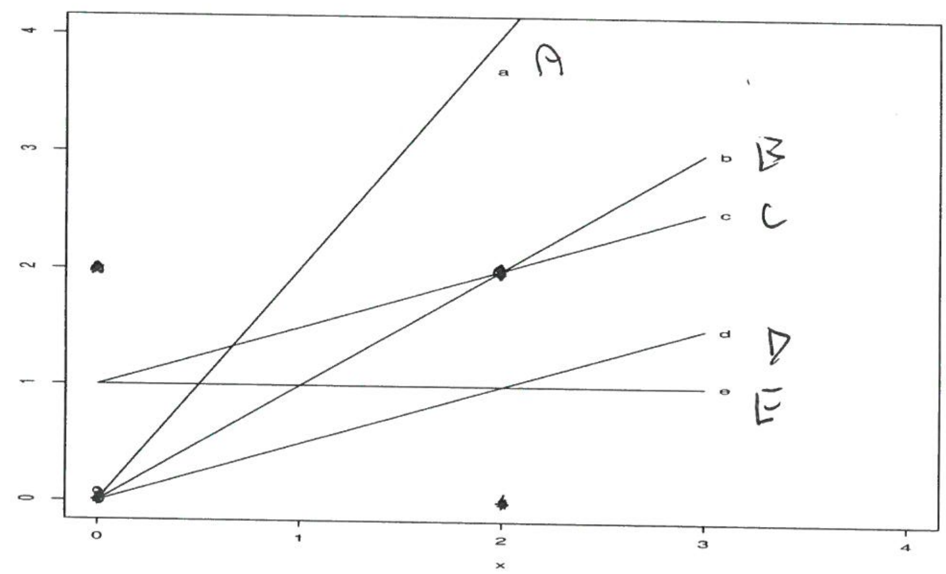
$$X'y = \begin{pmatrix} 0 \\ 20 \\ 4 \end{pmatrix} \quad 1$$

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} 0 \\ 20/8 \\ 4/8 \end{pmatrix} = \begin{pmatrix} 0 \\ 2\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \quad 2$$

7

2. The plot below shows four data points (0,0), (0,2), (2,0), (2,2), represented by *'s. Also shown are five straight lines, labelled a, b, c, d and e.

line	SSE
A	24
B	8
C	6
D	6
E	4



(2) (a) What is the total sum of squares ($\sum_{i=1}^{n}(y_i - \bar{y})^2$) for these data?

$\bar{y} = (0+0+2+2)/4 = 1$ 1

$\sum (y_i - \bar{y})^2 = 4(1^2) = 4$

(3) (b) Which line best represents the least squares line?

E

(2) (c) What is the residual (or error) sum of squares for the line you chose in part (b)?

4 (or entry in table above)

3. The regression model with unknown intercept β_0 and fixed slope 1 is given by

$$y_i = \beta_0 + x_i + \epsilon_i$$

(3) Using calculus, derive the least squares estimator of the intercept β_0 for this model. (i.e. Differentiate the error sum of squares $\sum_{i=1,n}(y_i - \beta_0 - x_i)^2$ with respect the the unknown parameter, and solve the appropriate equation.)

$SSE = \sum (y_i - \beta_0 - x_i)^2$

$\frac{\partial SSE}{\partial \beta_0} = -2 \sum (y_i - \beta_0 - x_i)$ 1

$\frac{\partial SSE}{\partial \beta_0} = 0 \Rightarrow \sum_{i=1}^{n} (y_i - x_i) = n\beta_0$ 1

$\Rightarrow \hat{\beta}_0 = \bar{y} - \bar{x}$

10

4. The "regression through the origin" model has intercept 0. If there is a single predictor variable with values $(x_1, x_2, \ldots, x_n)$, the model is

$$y_i = \beta x_i + \epsilon_i$$

.

In this case

$$\boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$$

(3) Using this $\boldsymbol{X}$, plug into the general formula $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ to calculate the least squares estimator of β.

$X'X = \sum x_i^2$ 1

$X'y = \sum x_i y_i$ 1

$\hat{\beta} = (X'X)^{-1}X'y = \frac{\sum x_i y_i}{\sum x_i^2}$ 1

5. Four regression models were fit to a dataset containing 102 observations. The model and error sum of squares for each of the models are included in the following table.

Model	SSE
$y = \beta_0 + \epsilon$	208
$y = \beta_0 + \beta_1 x_1 + \epsilon$	200
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$	198
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$	196

(3) (a) When testing the hypothesis $H_0 : \beta_3 = 0$, what is the observed value of F and what are the numerator and denominator degrees of freedom?

Hint: Use the partial F test where $F = \frac{(SSE_{red} - SSE_{full})/r}{SSE_{full}/(n-p)}$.

1

$F = \frac{(198-196)/1}{196/98}$ 1 $= 2/2 = 1$

$n = 102$

$r = 1$

$p = 4$

$n-p = 98$

1

(1, 98)

(3) (b) When testing the hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$, what is the observed value of F and what are the numerator and denominator degrees of freedom?

1

$F = \frac{(208-196)/3}{196/98}$ 1 $= \frac{4}{2} = 2$

$r = 3$

$n-p = 98$

(3, 98)

1

9

6. During early development, the weight y of an animal is assumed to be a linear function of age x.

(4) (a) Write down a single multiple regression equation which allows for completely different linear functions of age for males and females. That is, your multiple regression equation should allow for both different intercepts and different slopes for males and females. Carefully define any variables you need.

$Z = \begin{cases} 1; & \text{males} \\ 0; & \text{females} \end{cases}$ 1

$y = \underbrace{B_0 + B_1 \text{age}}_{1} + B_2 Z + B_3 Z\text{age} + \epsilon$ 1 1

(4) (b) In the following table, fill in the $\boldsymbol{E}(Y)$ which your model specifes for males and for females.

Sex	$\boldsymbol{E}(Y)$
Male	$E(Y) = (B_0 + B_2) + (B_1 + B_3)\text{age}$
Female	$E(Y) = B_0 + B_1 \text{age}$

2 2

(2) (c) Suppose that you want to test that the relationship between mean weight and age has the same intercept, but allowing for different slopes for males and females. State the associated null and alternative hypotheses in terms of the model parameters.

1 $H_0: B_2 = 0$

1 $H_a: B_2 \neq 0$

7. Suppose that X, Y and Z all have mean 0 and variance 1, that $Cov(X, Z) = .5$, $Cov(X, Y) = 0$ and $Cov(Y, Z) = 0$.

(4) What is the covariance between $X + Y$ and $X - 2Z$?

$\text{cov}(x + y, X - 2Z)$

$= \text{cov}(x, x) + \text{cov}(x, -2Z) + \text{cov}(y, x) + \text{cov}(y, -2Z)$

$= V(x) - 2\text{cov}(x, Z) + \text{cov}(y, x) - 2\text{cov}(y, Z)$

$= 1 - 2(.5) + 0 + 0$

$= 0$

14