STAT 420: Exam #1

Spring 2015, Sections: DU, DG

Solutions

Name:____________________________________

NetID:____________________________________

Be sure to show all your work; your partial credit might depend on it.

No credit will be given without supporting work.

Place your answers in the space provided.

You may use a calculator and one 8.5 x 11 inch sheet with notes.

Page	1	2	3	4	5	6	7	8	Total
Earned	2								

Academic Integrity
The University statement on your obligation to maintain academic integrity is:

If you engage in an act of academic dishonesty, you become liable to severe disciplinary action. Such acts include cheating; falsification or invention of information or citation in an academic endeavor; helping or attempting to help others commit academic infractions; plagiarism; offering bribes, favors, or threats; academic interference; computer related infractions; and failure to comply with research regulations.

Rule 33 of the Code of Policies and Regulations Applying to All Students gives complete details of rules governing academic integrity for all students. You are responsible for knowing and abiding by these rules.

[Problem 1] Consider the following dataset:

x_1	4	4	6	6	8	8	10	10	2	2
x_2	20	14	24	22	29	25	28	30	13	15
y	40	28	278	314	228	240	226	370	126	90

Then consider the model,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,$$

where $\epsilon_i \sim N(0, \sigma^2)$.

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 10 & 60 & 220 \\ 60 & 440 & 1480 \\ 220 & 1480 & 5200 \end{bmatrix} \qquad \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1} = \begin{bmatrix} 3.05 & 0.425 & -0.25 \\ 0.425 & 0.1125 & -0.05 \\ -0.25 & -0.05 & 0.025 \end{bmatrix}$$

$$\boldsymbol{X}^T\boldsymbol{Y} = \begin{bmatrix} 1940 \\ 13960 \\ 47800 \end{bmatrix} \qquad \hat{\boldsymbol{\beta}} = \begin{bmatrix} -100 \\ 5 \\ 12 \end{bmatrix} \qquad \sum (y_i - \hat{y}_i)^2 = 50400 \qquad \sum (y_i - \bar{y})^2 = 123440$$

a) (13) Perform the significance of the regression test at a 5% level of significance.

$H_0 : \beta_1 = \beta_2 = 0$
H_1 : at least one of β_1, β_2 not 0.
$n = 10$.
$p = 3 = (\#\text{ of } \beta\text{s}) = (\#\text{ of columns of } \boldsymbol{X})$

Source	SS	df	MS	F
Regression	73040	$p - 1 = 2$	36520	5.0722222
Error	50400	$n - p = 7$	7200	
Total	123440	$n - 1 = 9$		

Critical Value: $F_{0.05}(2, 7) = 4.74$
Rejection Region: Reject if $F > F_{0.05}(2, 7)$.
Decision: **Reject** H_0 at $\alpha = 0.05$.

b) (8) Test $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$ at a 1% level of significance.

$\hat{Var}(\hat{\beta}_2) = C_{22} \cdot s^2 = 0.025 \cdot 7200 = 180$

$$T = \frac{\hat{\beta}_2 - \beta_{20}}{\sqrt{\hat{Var}(\hat{\beta}_2)}} = \frac{12}{\sqrt{180}} = 0.8944272.$$

Degrees of Freedom: $n - p = 7$
Critical Value: $\pm t_{0.005} = \pm 3.499$
Rejection Region: Reject if $T > t_{0.005}$ or $T < -t_{0.005}$
Decision: **Do NOT Reject** H_0 at $\alpha = 0.01$.

c) (6) Construct a 90% confidence interval for β_1.

$\hat{Var}(\hat{\beta}_1) = C_{11} \cdot s^2 = 0.1125 \cdot 7200 = 810$
Degrees of Freedom: $n - p = 7$
$t_{0.05} = 1.895$

$$5 \pm 1.895 \cdot \sqrt{810} \qquad \mathbf{5 \pm 53.933} \qquad [\mathbf{-48.933}, \mathbf{58.933}]$$

[Problem 2] (8) Andy's Awesome Autos sells used automobiles. Assuming the price of cars X (in thousands of dollars) and the number of miles on the odometer Y (in thousands of miles) follow a joint bivariate normal distribution with parameters

$$\mu_X = 14, \qquad \sigma_X = 4, \qquad \mu_Y = 85, \qquad \sigma_Y = 20, \qquad \rho = -0.60$$

Suppose that Jillian buys a car at Andy's Awesome Autos with 65000 miles on the odometer. What is the probability that she pays more than 19600 for the car?

Given $y = 65$, we know that X is normally distributed with mean and variance

$$\mu^* = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y) = 14 - -0.6\frac{4}{20}(65 - 85) = 16.4$$
$$\sigma_*^2 = \sigma_X^2(1 - \rho^2) = 4^2(1 - -0.6^2) = 10.24$$

$$P(X > 85|Y = 65) = P\left(Z > \frac{19.6 - 16.4}{\sqrt{10.24}}\right) = P(Z > 1) = \mathbf{0.1587}$$

[Problem 3] Andy's Awesome Autos sells used automobiles. For $n = 10$ cars for sale, he has data,

- y = price of car (in thousands of dollars)
- x = age of car (in years)

x	2	1	5	2	3	4	3	5	1	4
y	19	27	8	9	18	14	24	14	21	6

$$\sum x_i = 30 \quad \sum y_i = 160 \quad \sum x_i^2 = 110 \quad \sum y_i^2 = 3004 \quad \sum x_i y_i = 420$$

$$\sum (x_i - \bar{x})^2 = 20 \quad \sum (y_i - \bar{y})^2 = 444 \quad \sum (x_i - \bar{x})(y_i - \bar{y}) = -60$$

a) (8) Find the equation of the least-squares regression line.

$$\hat{\beta}_1 = \frac{SXY}{SXX} = \frac{-60}{20} = -3, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 16 - -3(3) = \mathbf{25}.$$

So, the least squares regression line is given by:

$$\hat{y} = 25 + -3x.$$

b) (11) Perform the significance of the regression test at a 5% level of significance.

$H_0 : \beta_1 = 0$
$n = 10.$
$p = 2 = (\#\text{ of } \beta\text{s})$
$SSR = \beta_1^2 \cdot SXX = (-3)^2 \cdot 20 = 180$

Source	SS	df	MS	F
Regression	180	$p-1=1$	180	5.4545455
Error	264	$n-p=8$	33	
Total	444	$n-1=9$		

Critical Value: $F_{0.05}(1,8) = 5.32$
Rejection Region: Reject if $F > F_{0.05}(1,8)$.
Decision: **Reject** H_0 at $\alpha = 0.05$.

c) (3) Interpret β_0 in the context of the problem.

The average price of a new (age zero) car.

d) (3) Interpret β_1 in the context of the problem.

The change (decrease) in average price of a car per each additional year of age.

e) (7) Construct a 90% confidence interval for the average price of a car that is 3.5 years old.

Degrees of freedom: $n-2=8$
$t_{\alpha/2}=t_{0.05}=1.86$
$s_e^2=MSE=33$
$s_e=5.745$
$\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x=25+(-3)\cdot 3.5=14.5$

Confidence interval for $\mu_{y|x}$:

$$\hat{y}\pm t_{\alpha/2}\cdot s_e\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{SXX}}=14.5\pm 1.86\cdot 5.745\sqrt{\frac{1}{10}+\frac{(3.5-3)^2}{20}}$$

$$14.5\pm 3.584 \qquad [\mathbf{10.916},\mathbf{18.084}]$$

f) (7) Construct a 90% prediction interval for the price of a car that is 3.5 years old.

Degrees of freedom: $n-2=8$
$t_{\alpha/2}=t_{0.05}=1.86$
$s_e^2=MSE=33$
$s_e=5.745$
$\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x=25+(-3)\cdot 3.5=14.5$

Prediction interval for new y :

$$\hat{y}\pm t_{\alpha/2}\cdot s_e\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{SXX}}=14.5\pm 1.86\cdot 5.745\sqrt{1+\frac{1}{10}+\frac{(3.5-3)^2}{20}}$$

$$14.5\pm 11.271 \qquad [\mathbf{3.229},\mathbf{25.771}]$$

[Problem 4] (14) Suppose the model,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \epsilon_i,$$

was fit to $n = 40$ data points.

```
sum( lm( y ~ 1 )$residuals^2 )
[1] 400

sum( lm( y ~ x1 + x3 + x5 )$residuals^2 )
[1] 280

sum( lm( y ~ x2 + x4 + x6 + x7 )$residuals^2 )
[1] 300

sum( lm( y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 )$residuals^2 )
[1] 240
```

Test $H_0 : \beta_1 = \beta_3 = \beta_5 = 0$ at a 10% level of significance.

Full Model: $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \epsilon_i$ $p = 8$.
$SSE_{Full} = 240$.

Null Model: $Y_i = \beta_0 + \beta_2 x_{i1} + \beta_4 x_{i3} + \beta_6 x_{i5} + \beta_7 x_{i6} + \epsilon_i$ $q = 5$.
$SSE_{Null} = 300$

Source	SS	df	MS	F
Difference	60	3	20	2.6666667
Full	240	32	7.5	
Null	300	35		

Critical Value: $F_{0.1}(3, 32) = 2.26$
Rejection Region: Reject if $F > F_{0.1}(3, 32)$.
Decision: **Reject** H_0 at $\alpha = 0.1$.