STAT 512 Midterm 1 (Total 70 Points) – Spring 2014

Name: Solution

Section#3 --- 8:30am; **Section#2 --- 9:30am**

- **Exam time: 8:00-10:00pm.**
- **Must show all work to get credits.**

1. (8 points) Short answer questions.

(a) For a model with four parameters, suppose we want to make simultaneous inference to guarantee that the joint coverage of the four confidence intervals is at least 90%, what confidence level you should use to construct each interval using the Bonferroni correction method?

$\frac{0.1}{4} = 0.025$ or $1 - \frac{0.1}{4} = 0.975$

(b) The correlation between two variables Y and X is -0.6. With a simple linear regression model, what percent of the variation in Y can be explained by X?

$r = -0.6$

$R^2 = (-0.6)^2 = 0.36$

(c) Under alternative hypothesis, what is the distribution of the t test statistic? What are the parameters for this distribution?

(1) non-central t distribution

(1) degree of freedom & non-centrality parameter

(d) In simple linear regression, what would you do if the residuals are badly behaved?

transform Y

2. (27 points) A researcher wants to fit a simple linear regression model to a set of data with 27 observations. The estimates he obtained are: $b_0=5$ (with standard error 3), $b_1=1.5$ (with standard error 0.8), and SSE=100.

(2-1). Write down the model and corresponding assumptions.

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2) \qquad i=1,2,\cdots 27$

(2-2). Find the predicted value of the response variable when the explanatory variable is 2.

$\hat{Y} = 5 + 1.5 \times 2 = 8$

(2-3). What is the estimate of the standard deviation of the error in the model?

$\hat{\sigma} = S = \sqrt{MSE} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{100}{27-2}} = \sqrt{4} = 2$

(2-4). Give an estimate of the change in the response variable if the explanatory variable increase by 5.

$\Delta Y = 1.5 \times 5 = 7.5$ $\uparrow$

orignal: X^* $\qquad Y^* = b_0 + b_1 X^*$

new: $X^* + 5$ $\qquad Y^{**} = b_0 + b_1 (X^* + 5)$

$= b_0 + b_1 X^* + 5b_1$

$\Delta Y = Y^{**} - Y^* = 5b_1$

(2-5). Give a 95% confidence interval for your estimate in (2-4).

$7.5 \pm t_{25, 0.05} \cdot S(5b_1)$

$= 7.5 \pm 2.06 \times 5 \times (0.8)$

$= 7.5 \pm 8.24$

$= (-0.74, 15.74)$

(2-6). What is the residual corresponding to the data point with x=1 and y=6?

$$r = y - \hat{y}$$
$$= 6 - (5 + 1.5)$$
$$= -0.5$$

(2-7). Suppose the confidence interval for the mean response at x=2 is (5, 11), what is the prediction interval at x=2?

$$8 \pm t_{25,(0.05)} \times S(\hat{y}) = 8 \pm 2.06 \times S(\hat{y}) = (5, 11)$$

$$MSE = \frac{SSE}{n-2} = \frac{100}{25}$$

$$\Rightarrow 2.06 \cdot S(\hat{y}) = 3 \quad \Rightarrow S^2(\hat{y}) = \left(\frac{3}{2.06}\right)^2 = 2.1208$$

Predicted interval: $8 \pm 2.06 \times \sqrt{[S^2(\hat{y}) + s^2]} = 8 \pm 2.06 \sqrt{[2.1208 + 4]}$

$$= 8 \pm 2.06 \times \sqrt{6.1208} = 8 \pm 2.06 \times 2.474$$
$$= 8 \pm 5.0964 = (2.9036, 13.0964)$$

(2-8). Test the hypothesis: Ho: β_1=2 vs. Ha: $\beta_1 \neq 2$ at α=0.05.

$H_0: \beta_1 = 2 \qquad H_a: \beta_1 \neq 2$

$$TS = \frac{1.5 - 2}{0.8} = \frac{-0.5}{0.8} = -0.6250 < t_{25, 0.05}$$

$$t_{27-2, (0.05)} = 2.06$$

Fail to reject H_0

(2-9). Describe the procedure to construct a confidence band for the regression line.

$$\underline{\hat{Y}_h} \pm W \cdot s(\hat{Y}_h) \qquad W = \sqrt{2 \cdot F_{2, n-2}(1-\alpha)} \qquad \cdot\, \underline{s(\hat{Y}_h)}$$

$$= \sqrt{2 F_{2,15}(0.95)}$$

3. (35 points) An experiment was conducted to determine the effect of temperature (x1), humidity (x2), and rate (x3) on the viscosity of a polymer (y). We fit the data (17 observations in total) using a multiple linear regression model including both liner and quadratic terms.

(3-1). Write down the model and corresponding assumptions.

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \beta_4 X_{i,1}^2 + \beta_5 X_{i,2}^2 + \beta_6 X_{i,3}^2 + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$

$i = 1, 2, \cdots 17$

(3-2). Complete the following ANOVA table from SAS output.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	[6]	[43743]	[7290.5]	[18.5132]	[<0.01]
Error	[10]	[3938]	[393.8]		
Corrected Total	[16]	47681			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS	95% Confidence Limits	
Intercept	1	754.08139	1477.13021	0.51	0.6208	2012400	102.63165	-2537.16981	4045.33260
humidity	1	-1.77614	5.54349	-0.32	0.7553	18731	40.42676	-14.12781	10.57553
temp	1	-19.54188	38.36517	-0.51	0.6215	21175	102.17404	-105.02481	65.94106
rate	1	20.50195	62.49984	0.33	0.7497	3470.55480	42.37549	-118.75637	159.76026
hum2	1	0.02614	0.07597	0.34	0.7379	132.37219	46.62205	-0.14314	0.19542
tem2	1	0.14926	0.21504	0.69	0.5034	223.23157	189.73312	-0.32987	0.62839
rat2	1	-0.50181	3.02266	-0.17	0.8715	10.85399	10.85399	-7.23673	6.23310

(3-3). What is the R-square and adjusted R-square?

$$R^2 = \frac{43743}{47681} = 0.9174$$

$$\text{adj } R^2 = 1 - \frac{3938/10}{47681/16} = 1 - 0.1321 = 0.8679$$

(3-4). Perform the test for model significance. Please state your hypotheses, decision rule, and conclusion.

H_0: $\beta_j = 0 \quad \forall j = 1, \dots 6$

H_a: at least one $\beta_j \neq 0$

P-value < 0.01

Reject if $F > F_{6,10}(0.05) = 3.22$

(3-5). The distribution of one of the explanatory variables seems to be skewed (i.e., not normal). Is this a concern? Why?

No $\because$ no assumption on distribution of X

(3-6). Write the estimated regression equation for the fitted multiple regression model.

$$\hat{y} = 754.08 - 1.77 \cdot hum - 19.54 * temp + 20 \cdot rate + 0.02 \cdot hum^2 + 0.15 \cdot tem^2 - 0.50 \, rate^2$$

(3-7). Test whether the quadratic terms are significant or not (one test). Give the null and alternative hypotheses, p value, and your conclusion.

H_0: $\beta_4 = \beta_5 = \beta_6 = 0$

H_a: at least one $\beta_j \neq 0$ $\quad j = 4, 5, 6$

$$TS = \frac{\left(SSM(F) - SSM(R)\right) / \left[df(F) - df(R)\right]}{SSE(F) / df_E(F)}$$

$$= \frac{\left[43743 - (18731 + 21175 + 3471)\right] / 3}{393.8}$$

$$= \frac{(43743 - 43377)/3}{393.8} = \frac{366/3}{393.8} = \frac{122}{393.8} = 0.3098$$

$F_{3,10}(0.05) = 3.71$

Fail to reject H_0

(3-8) If temperate=80, humidity=30, and rate=10, what is the average viscosity under this condition?

$$\hat{y} = 754.08 - 1.78 \times 30 - 19.54 \times 80 + 20.5 \times 10 + 0.03 \times 30^2 + 0.15 \times 80^2 - 0.50 \times 10^2$$

$$= 279.48$$

(3-9) Give the confidence interval for the average viscosity in (3-~~7~~8).

(3-10). Based on the SAS output of Analysis of Variance and Parameter Estimates, is there any contradiction you observe? Please interpret it.

Model is significat but none of the predictors is significant

due to multi-collinearity

(3-11) Suggest the remedies for the contradiction in (3-10).

Centering the predictors

5 (3-12) Describe the procedure you will use if you were asked to fit the model.