Stat 512 Spring 2009 Midterm Exam 1

Instructor: Bowei Xi

Name:

Department:

Solution

To receive full credit:
Show all work that lead to the final answer on the attached pages

Problem 1 (8): One regression model is $Y = \beta_0 + \beta_1 \times X + \epsilon$. Fit the regression model and we have $b_0 = 2$ and $b_1 = 5$. Answer the following questions.

1. (1) Given $X = 10$, what is the mean fitted value $\hat{Y}$?
2. (1) For the data point $X = 20$ and $Y = 107$, what is the residual?
3. (1.5) 26 data points are used to fit the regression model. ($n = 26$). The error sum of squares (SSE) is 97. What is the estimate for σ^2?
4. (1.5) The SAS output regarding the slope is:

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
X	1	5	0.37	??	<.0001

   What is the value of t-statistic?
5. (1.5) Without actually constructing a 95% confidence interval, can you tell whether the 95% confidence interval for the slope contains 0 or not? Explain.
6. (1.5) There is another predictor variable $W = 2X + 1$. We fit another regression model $Y = \beta_0^* + \beta_1^* \times W + \epsilon^*$. What are the fitted intercept and slope b_0^* and b_1^*?

(1) $\hat{Y} = 2 + 5 \times 10 = 52$

(2) $\hat{Y} = 5 \times 20 + 2 = 102 \quad e = 107 - 102 = 5$

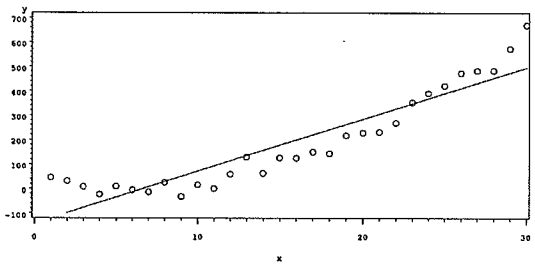
(3) $\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{97}{24} = 4.04$

(4) $t\text{-stat} = \frac{5}{0.37} = 13.51$

(5) No. Since p-value < 0.05, we reject the null $H_0: \beta_1 = 0$. Hence the 95% confidence interval for β_1 doesn't contain 0.

(6) $\hat{Y} = 2 + 5X = 2.5 \times (2X+1) - 0.5 = -0.5 + 2.5W$

$\hat{\beta}_0^* = b_0^* = -0.5$, $b_1^* = 2.5$



Problem 2 (8): We have the following ANOVA table from the SAS output.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	209859	209859	2520.3	<.0001
Error	56	4633	83.27		
Corrected Total	57	214522			

1. (3) Fill in the missing values on the ANOVA table.
2. (1) Write down the hypotheses H_0 and H_a. What is your conclusion based on ANOVA?
3. (1) What is the sample size n?
4. (1) What is R^2? And what is the estimate for σ^2?
5. (1) The Normal QQplot of the residuals is the above. Is there anything wrong? Explain.
6. (1) The Box-Cox transformation suggests $\lambda = -1$. What is the suggested transformation of Y?

(1) above.

(2) $H_0: \beta_1 = 0$ $H_A: \beta_1 \neq 0$. Since p-value < 0.0001, we reject H_0.

(3) $n = 58$

(4) $R^2 = \frac{209859}{214522} = 97.83\%$

$\hat{\sigma}^2 = MSE = 83.27$

(5). The residuals are not from a Normal distribution.
We may need to transform Y or use a different model.

(6) $Y^{-1} = \frac{1}{Y}$.

Problem 3 (9): We have the following partial SAS output:

Root MSE	21	R-Square	0.8733
Dependent Mean	300	Adj R-Sq	0.8987
Coeff Var	2.6		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	Pr > |t|
Intercept	1	-26	7	<.0001
X	1	2156	98	<.0001

1. (1.5) Write down the fitted regression line equation.
2. (1) What is the estimate for σ^2?
3. (1) What is the Pearson correlation coefficient between Y and X?
4. (2) Sample size $n = 21$. The critical value (the t-distribution percentile) is $t_{19}(0.975) = 2.093$. Compute the 95% confidence interval for the slope.
5. (2) The $R^2 = 0.8733$. What is the value of the F-statistic on ANOVA table?
6. (1.5) Given $X = 1$, there are two intervals for Y, $[2120, 2140]$ and $[2110, 2150]$. Which is the confidence interval and which is the prediction interval? Explain.

(1) $\hat{Y} = -26 + 2156X$

(2) $\hat{\sigma}^2 = (\text{Root MSE})^2 = 21^2 = 441$

(3) Pearson correlation coefficient $\rho = \sqrt{R\text{-Square}} = \sqrt{0.8733} = 0.9345$

(4) $2156 \pm 2.093 \times 98 = [1950.9, 2361.1]$

(5) $F = \dfrac{R^2/(p-1)}{(1-R^2)/(n-p)} = 130.96 \quad (p=2)$

(6) Prediction interval is $[2110, 2150]$, wider than confidence interval.