MSA 8190 Statistical Foundations

Fall 2016

Assignment 8

Issued: November 8, 2016

Due: November 15, 2016

Problem 1

Let x be the compressive strength and y be the intrinsic permeability of various concrete mixes and cures. Summary quantities are $n = 15$, $\sum_{i=1}^{n} y_i = 572$, $\sum_{i=1}^{n} y_i^2 = 23,530$, $\sum_{i=1}^{n} x_i = 43$, $\sum_{i=1}^{n} x_i^2 = 157.42$, and $\sum_{i=1}^{n} x_i y_i = 1697.80$. Assume that the two variables are related according to the simple linear regression model.

(a) Calculate the least squares estimates of the slope and intercept.

(b) Use the equation of the fitted line to predict what permeability would be observed when the compressive strength is $x = 4.3$.

(c) Give a point estimate of the mean permeability when compressive strength is $x = 3.7$.

(d) Suppose that the observed value of permeability at $x = 3.7$ is $y = 46.1$. Calculate the value of the corresponding residual.

(e) Test for significance of regression using $\alpha = 0.05$. Find the P-value for this test. Can you conclude that the model specifies a useful linear relationship between these two variables?

(f) Estimate σ^2 and the standard deviation of $\hat{\beta}_1$.

(g) What is the standard error of the intercept in this model?

Problem 2

Using data in Problem 1, find 95% confidence interval on each of the following:

(a) Slope

(b) Intercept

(c) Mean permeability when $x = 2.5$

(d) Find a 95% prediction interval on permeability when $x = 2.5$. Explain why this interval is wider than the interval in part (c).

Problem 3

Data concerning the performance of the 28 National Football League teams in 1976 are shown below. Excel file can be found in D2L. It is suspected that the number of games won (y) is related to the number of yards gained rushing by an opponent (x).

Teams	Games Won (y)	Yards Rushing by Opponent (x)	Teams	Games Won (y)	Yards Rushing by Opponent (x)
Washington	10	2205	Detroit	6	1901
Minnesota	11	2096	Green Bay	5	2288
New England	11	1847	Houston	5	2072
Oakland	13	1903	Kansas City	5	2861
Pittsburgh	10	1457	Miami	6	2411
Baltimore	11	1848	New Orleans	4	2289
Los Angeles	10	1564	New York Giants	3	2203
Dallas	11	1821	New York Jets	3	2592
Atlanta	4	2577	Philadelphia	4	2053
Buffalo	2	2476	St. Louis	10	1979
Chicago	7	1984	San Diego	6	2048
Cincinnati	10	1917	San Francisco	8	1786
Cleveland	9	1761	Seattle	2	2876
Denver	9	1709	Tampa Bay	0	2560

(a) Calculate the least squares estimates of the slope and intercept. What is the estimate of σ^2? Graph the regression model.

(b) Find an estimate of the mean number of games won if the opponents can be limited to 1800 yards rushing.

(c) What change in the expected number of games won is associated with a decrease of 100 yards rushing by an opponent?

(d) To increase by 1 the mean number of games won, how much decrease in rushing yards must be generated by the defense?

(e) Given that $x = 1917$ yards (Cincinnati), find the fitted value of y and the corresponding residual.

(f) Test for significance of regression using $\alpha = 0.01$. Find the P-value for this test. What conclusions can you draw?

(g) Estimate the standard errors of the slope and intercept.

(h) Test (using $\alpha = 0.01$) $H_0 : \beta_1 = -0.01$ versus $H_1 : \beta_1 \neq -0.01$. Would you agree with the statement that this is a test of the claim that if you can decrease the opponent's rushing yardage by 100 yards the team will win one more game?

Problem 4
Using data in Problem 3, find 95% confidence interval on each of the following:

(a) Slope

(b) Intercept

(c) Mean number of games won when opponents rushing yardage is limited to $x = 1800$.

(d) Find a 95% prediction interval on the number of games won when opponents rushing yards is 1800.

Problem 5
For data in Problem 3:

(a) Calculate R^2 for this model and provide a practical interpretation of this quantity.

(b) Prepare a normal probability plot of the residuals from the least squares model. Does the normality assumption seem to be satisfied?

(c) Plot the residuals versus y and against x. Interpret these graphs.

Problem 6
Show that in a simple linear regression model the point $(\bar{x}, \bar{y})$ lies exactly on the least squares regression line.

Problem 7
Suppose we wish to fit a regression model for which the true regression line passes through the point $(0, 0)$. The appropriate model is $Y = \beta x + \epsilon$. Assume that we have n pairs of data $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. Find the least squares estimate of the model.

Problem 8
Let $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ be the estimated intercept, slope and variance of linear regression model for n pairs of data $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. Suppose that each value of x_i is multiplied by a positive constant a, and each value of y_i is multiplied by another positive constant b. Let β_0^*, β_1^*, and σ^{*2} be the estimated intercept, slope and variance of linear regression model for n pairs of data $(ax_1, by_1), (ax_2, by_2), \cdots, (ax_n, by_n)$.

(a) How the new estimators β_0^*, β_1^*, and σ^{*2} are related to the old estimators $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$, respectively?

(b) Show that the t-statistic for testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ is unchanged in value.

Problem 9

The 1976 team performance statistics for the teams in the National Football League is given in the Excel file.

(a) Fit a multiple regression model relating the number of games won to the teams' passing yardage (x_2), the percent rushing plays (x_7), and the opponents' yards rushing (x_8).

(b) Estimate σ^2.

(c) What are the standard errors of the regression coefficients?

(d) Use the model to predict the number of games won when $x_2 = 2000$ yards, $x_7 = 60\%$, and $x_8 = 1800$.

(e) Test for significance of regression using $\alpha = 0.05$. What is the P-value for this test?

(f) Conduct the t-test for each regression coefficient β_2, β_7, and β_8. Using $\alpha = 0.05$, what conclusions can you draw about the variables in this model?

(g) Find the amount by which the regressor x_8 (opponents'yards rushing) increases the regression sum of squares.

(h) Use the results from part (e) above and to conduct an F-test for $H_0 : \beta_8 = 0$ versus $H_1 : \beta_8 \neq 0$ using $\alpha = 0.05$. What is the P-value for this test? What conclusions can you draw?

(i) Find a 95% confidence interval on β_8.

(j) What is the estimated standard error of $\hat{\mu}_{Y|\mathbf{x}_0}$ when $x_2 = 2000$ yards, $x_7 = 60\%$, and $x_8 = 1800$ yards?

(k) Find a 95% confidence interval on the mean number of games won when $x_2 = 2000$ yards, $x_7 = 60\%$, and $x_8 = 1800$.