Linear Regression Models
Statistics GR5205 – Fall 2016

Assignment 4

Reading:

By Tuesday, October 11, read Appendix A and Chapters 1–5 of *Applied Linear Regression Models* by Kutner, Nachtsheim and Neter (KNN).

For Tuesday, October 18, read Chapter 6 and Section 10.1 of KNN.

Homework 4:

The following problems are due in class on Tuesday, October 25.

1. (Problem 3.13 in KNN) Continue with the *Copier maintenance* data.
   (a) What are the alternative conclusions when testing for lack of fit of a linear regression function?
   (b) Conduct the test indicated in part (a) for the *Copier maintenance* data: Clearly state your null and alternative hypotheses; report and interpret the P-value.
   (c) Does the test in part (b) detect other departures from the normal SLR model, such as lack of constant variance or lack of normality in the error terms? Could the results of the test of lack of fit be affected by such departures? Discuss.

2. (Problems 4.3 and 4.7 in KNN) Continue with the *Copier maintenance* data.
   (a) Use the Bonferroni method to obtain simultaneous 95% confidence intervals for β_0 and β_1. Interpret your intervals.
   (b) A consultant has suggested that β_0 should be 0 and β_1 should equal 14. Do the intervals in part (a) support or contradict this claim? Explain.
   (c) Estimate the expected number of minutes spent when there are 3, 5, and 7 copiers to be serviced, respectively. Use interval estimates with simultaneous 95% confidence based on the Bonferroni method. Interpret your results.
   (d) Two service calls for preventive maintenance are scheduled in which the number of copiers to be serviced are 4 and 6, respectively. Use the Bonferroni method to obtain a pair of prediction intervals with joint 95% confidence for the times to be spent on these calls. Interpret your results.

(e) Reproduce the scatterplot of `minutes` versus `copiers`, overlay the least squares line and the Working-Hotelling 95% confidence band for the mean function $E(Y|X = x)$. Interpret the confidence band.

(f) Refer to part (b) above. Does the line $y = 14x$ fall within the 95% confidence band for the true mean function? What is the significance of this?

3. (Problem 4.27 in KNN) Continue with the *SENIC Project* data. Consider the regression of average length of stay on infection risk.

(a) Use the Bonferroni method to obtain simultaneous 95% confidence intervals for β_0 and β_1. Interpret.

(b) A researcher suggested that β_0 should be 7 and β_1 should be 1. Do the data support this expectation? Assess in two ways: (1) by reference to the intervals in part (a); and (2) by reference to a 95% confidence band for the mean function $E(Y|X = x) = \beta_0 + \beta_1 x$. Are your conclusions consistent? Explain.

(c) We wish to simultaneously estimate the expected average stay at hospitals with average infections risks $x = 2, 3, 4, 5$ with 95% family-wise confidence. Use the Bonferroni method to construct the desired intervals. Interpret your results.

4. Return to the *Copier maintenance* data. Find the following quantities by matrix calculations only — no `lm()` allowed!

(a) Vector of estimated regression coefficients $\mathbf{b}$;

(b) vector of fitted values $\hat{\mathbf{Y}}$ and vector of residuals $\mathbf{e}$;

(c) *SSTO* and *SSE*;

(d) estimated variance-covariance matrix of $\mathbf{b}$;

(e) a 95% confidence interval for $\mathrm{E}(Y|X = 6)$, and

(f) a 95% prediction interval for Y_h at $X = x_h = 6$.