Supplemental Handouts for Simple Linear Regression (Ch. 7)

page 1  Ecological correlations (Freedman, Pisani, Purves, *Statistics*)

page 2,3  Regression effect/fallacy (FPP) + practice problems

page 4,5  Plots of residuals vs. fitted values (Wiesberg, *Applied Linear Regression*)

page 6-8  Notes on diagnostics (Guenther Walther, STAT201, Stanford)

Outline, 1/13/04

1. Simple Linear Regression (SLR) estimation
2. Assumptions
3. Parameter estimation in SLR
4. Inferential tools: Illustrated using height/nematode example

\[
\begin{array}{cccc}
\text{Value} & \text{Std. Error} & t \text{ value} & \text{Pr(>|t|)} \\
\text{(Intercept) } & 10.3264 & 0.6890 & 14.9876 & 0.0000 \\
\text{Number.of.Nematodes } & -0.5738 & 0.1228 & -4.6740 & 0.0004 \\
\end{array}
\]

Residual standard error: 1.933 on 14 degrees of freedom
Multiple R-Squared: 0.6094
F-statistic: 21.85 on 1 and 14 df, the p-value is 0.0003584

Questions to answer for nematode/plant height problem:
(a) Is there evidence of a linear association between nematode level and plant height?
(b) Does nematode level cause decreased plant heights? What info is needed to infer causality?
(c) Are decreased nematode levels associated with increased plant growth?
(d) What is the mean plant height when 0 nematodes are present? Estimate this mean and give a 95% confidence interval.
(e) Give a 95% CI for the mean height when there are 5000 nematodes present. Hint: centering trick
Solution: subtract 5 (x 1000) from each of the x’s, and re-run the regression. In Splus, the formula is \( Y \overset{I}{(X - 5)} \). \( I() \) notation is used for transformations of x.

\[
\begin{array}{cccc}
\text{Value} & \text{Std. Error} & t \text{ value} & \text{Pr(>|t|)} \\
\text{(Intercept) } & 7.4575 & 0.4987 & 14.9551 & 0.0000 \\
\text{Center.Nematodes } & -0.5738 & 0.1228 & -4.6740 & 0.0004 \\
\end{array}
\]

Residual standard error: 1.933 on 14 degrees of freedom
Multiple R-Squared: 0.6094
F-statistic: 21.85 on 1 and 14 df, the p-value is 0.0003584
95% CI for mean plant height when 5000 nematodes are present: 7.458 +/- t(0.975,14) (0.499)

5. Checking assumptions by looking at residuals
(a) Assumptions checked: mean zero errors, constant variance of errors
(b) Diagnose: nonlinearity, nonconstant variance, presence of outliers
(c) Reading: 5.5.2, p. 131; Display 5.6, p. 133; 8.3.2, p. 214; handouts

6. QQ normal plot of residuals, Reading: p. 224-225
7. Plotting residuals vs. time order (8.6.3), vs. spatial location