Stat/For/Hort 572 — Midterm I, Spring 2002 — Solutions

1. (a) A curvature in the residual plot indicates that the linear-line model is inadequate. The 6th residual is close to -2 and is a potential outlier. The independence assumptions should be checked via a time-series plot, because data were collected over time and the residuals could be potentially correlated across time. The assumptions of normal distribution and equal variances do not seem to be violated.

   (b) Use the formula $T = \frac{y^* - \hat{y}_{\text{pred}}}{se(\hat{y}_{\text{pred}})}$, where $y^* = 10.8$ and $\hat{y}_{\text{pred}} = 13.03$. Also, since SSError $= 3.50$ and dfError $= 5$, we have MSError $= 3.50/5 = 0.70$ and $se(\hat{y}_{\text{pred}}) = \sqrt{\text{MSError} + se(\hat{y}_{\text{est}})^2} = \sqrt{0.70 + 0.39^2} = 0.9231$. Hence $T = (10.8 - 13.03)/0.9231 = -2.42$ on df $= 5$, with $0.05 <$ p-value < 0.10. Comparing with Bonferroni cutoff $\alpha/8 = 0.05/8$, we do not reject H_0. Hence the observation is not an outlier.

   (c) Use the additional sum of squares principle. Model I is the reduced model with $\text{SSE}_R = 156.39 \times (1 - R^2) = 156.39 \times (1 - 0.951) = 7.66$ on 6 df. Model II is the full model with $\text{SSE}_F = 156.39 \times (1 - R^2) = 156.39 \times (1 - 0.983) = 2.66$ on 5 df. Hence $F = \frac{(SSE_R - SSE_F)/(df_R - df_F)}{SSE_F/df_F} = \frac{(7.66-2.66)/1}{2.66/5} = 9.41$ on (1,5) df. Since $0.01 <$ p-value < 0.05, we reject H_0. The linear-line model is not adequate.

2. (a) The coefficient b_3 represents the intercept difference between cabbage B and C. This test is given by the t-ratio for `w3`. There, $T = 3.60$ on 6 df and the p-value is 0.011, so there is weak evidence against H_0.

   (b) Relative to the full model $Y = b_0 + b_1 w_1 + b_2 w_2 + b_3 w_3 + b_4 w_4 + b_5 w_5 + e$, we need to test $H_0 : (b_2 = b_3 = 0 | b_0, b_1, b_4, b_5)$. We can do this using the additional sum of squares principle and get:

   $$F = \frac{(12.40 - 3.215)/2}{0.536} = 8.57.$$

   Comparing to $F_{2,6}$, we get a p-value between 0.01 and 0.05 and so there is weak evidence against H_0.

   (c) Relative to the full model $Y = b_0 + b_1 w_1 + b_2 w_2 + b_3 w_3 + b_4 w_4 + b_5 w_5 + e$, we need to test $H_0 : (b_4 = b_5 = 0 | b_0, b_1, b_2, b_3)$. We can do this using the additional sum of squares principle and get:

   $$F = \frac{(5.896 + 41.616)/2}{0.536} = 44.33.$$

   Comparing to $F_{2,6}$, we get a p-value less than 0.001 and so there is strong evidence against H_0.

3. (a) The final model is $y = b_0 + b_1 x_1 + e$.

   (b) True. The final model using both forward selection and stepwise procedure is $y = b_0 + b_1 x_1 + e$.