Stat/For/Hort 572 — Midterm I, Spring 2006 — Solutions

1. (a) Consider the four model assumptions: correct model, independence, equal variance, and normal distribution. The linear line relationship appears inadequate and the equal variance assumption may not be satisfied.

   (b) H_0 : no LOF versus H_A : LOF of the SLR model. From the R output, the observed $f = 9.0348$. Compare to F distribution with df = (12, 14), the p-value is less than 0.001. Thus reject H_0 at 5% level and there is very strong evidence of a lack of fit of the SLR model.

   (c) From the R output, for $x^* = 18$, $\hat{y}_{\text{est}} = 6.76, s.e.(\hat{y}_{\text{est}}) = 0.7084$. Since $t_{0.025,26} = 2.056$, the 95% confidence interval is $\hat{y}_{\text{est}} \pm t_{0.025,26} \times s.e.(\hat{y}_{\text{est}})$, which is 6.76 ± 1.46 or $[5.30, 8.22]$.

   (d) H_0 : no LOF versus H_A : LOF of the quadratic regression model. Since SS Pure Error is 18.601 on 14 df and SS Error is 126.64 on 25 df, SS LOF is $126.64 - 18.601 = 108.039$ on $25 - 14 = 11$ df. By the additional sum of squares principle, the observed $f = \frac{108.039/11}{18.60/14} = 7.39$. Compare to F distribution with df = (11, 14), the p-value is less than 0.001. Thus reject H_0 at 5% level and there is very strong evidence of a lack of fit of the quadratic regression model.

2. (a) The parameter b_3 is the slope difference between the two regression lines corresponding to farms A and B respectively. From the R output, use either the observed $t = -0.608$ on 6 df or $f = 0.3694$ on (1, 6) df. The p-value is 0.5656. Do not reject H_0 at 5% level and there is no evidence of a nonzero b_3.

   (b) The parameter b_0 is the intercept of the regression line for farm A, which represents the expected weight gain for a zero level of diet supplement. Since $\hat{b}_0 = 1.888, s.e.(\hat{b}_0) = 0.3299, t_{0.05,6} = 1.943$, the 90% confidence interval is $\hat{b}_0 \pm t_{0.05,6} \times s.e.(\hat{b}_0)$, which is 1.89 ± 0.64 or $[1.25, 2.53]$.

   (c) The test of interest is $H_0 : [b_2 = b_3 = 0|b_0, b_1]$ versus H_A: not H_0. The additional sum of squares is $10.04 + 0.0106 = 10.0506$ on 2 df and the SSE of the full model is 0.1718 on 6 df. By the additional sum of squares principle, the observed $f = \frac{10.0506/2}{0.1718/6} = 175.51$. Compare to F distribution with df = (2, 6), the p-value is less than 0.001. Thus reject H_0 at 5% level and there is very strong evidence that the two regression lines are not equal.

   (d) The full model is $y = b_0 + b_1 w_1 + b_2 w_2 + e$ which has SSE = $0.0106 + 0.1718 = 0.1824$ on 7 df. The reduced model is $y = b_0 + b_1 w_1 + e$ with additional sum of squares 10.04 on 1 df. By the additional sum of squares principle, the observed $f = \frac{10.04/1}{0.1824/7} = 385.31$. Compare to F distribution with df = (1, 7), the p-value is less than 0.001. Thus reject H_0 at 5% level and there is very strong evidence of a nonzero b_2. That is, although there is no evidence of slope difference, there is strong evidence of intercept difference between the two regression lines for farms A and B.

3. (a) From the R output, the correlation between y and each individual x is the highest for x_3. This implies that the third model $y = b_0 + b_1 x_3 + e$ has the largest R^2 and thus the smallest SSE. The df of SSE is the same for all three models. Thus the third model has the smallest MSE.

   (b) H_0: the third observation is not an outlier versus H_A: not H_0. Because of the way x_4 is coded, use the observed t-value 2.983 on 7 df. The p-value is 0.02043 for one comparison and thus the experiment-wise p-value is $12 \times 0.02043 = 0.2452$. Do not reject H_0 at 5% level and there is no evidence that the third observation is an outlier.

   (c) According to the full model fit, the t-value for b_1 (1.885) is the smallest and is less than 2. Thus eliminate x_1 is the first step. Now fit the model with x_2 and x_3. Since the smallest t-value is 16.905 and is more than 2, stop. The model selected by backward elimination is

$$y = b_0 + b_2 x_2 + b_3 x_3 + e$$

Grade Distribution

```
  100:2
90-99:15
80-89:18
70-79:17         mean = 78, median = 80
60-69:3
  <60:10
```