Stat/For/Hort 572 — Midterm I, Spring 2005 — Solutions

1. (a) H_0 : no LOF versus H_A : LOF. From the R `anova(lake.lm, lake.aov)` output, $f = 1.7966$ on df $= (7, 6)$. The p-value is 0.2463 and we do not reject H_0 at 5% level. There is no evidence of a lack of fit of the SLR model.

   (b) Since $y^* = 58, \hat{y}_{\text{pred}} = 46.75, s.e.(\hat{y}_{\text{pred}}) = \sqrt{7.06 + 0.824^2} = 2.78$, we have $t = \frac{58-46.75}{2.78} = 4.05$ on df $= 12$. The p-value is less than $2 \times 0.001 = 0.002$, which is smaller than the comparisonwise significance level $\alpha/n = 0.05/15 = 0.003$. Thus we reject H_0 at 5% level and there is evidence that the second observation is an outlier.

   (c) The second observation is not an influential observation, because it is in the middle of the data range and the fitted intercepts and slopes are similar whether or not the observation is in the data set. One could also compute the leverage value of this observation and it is quite small.

2. (a) Step 1: Select x_3, as the SLR `y~x3` has the largest $R^2 = 0.4524$. Keep x_3, as the t-value is -4.359 which has a magnitude greater than 2.

   (b) Step 2: Select x_2 in addition to x_3, as the MLR `y~x2+x3` has the largest $R^2 = 0.5664$. Keep x_2, as the t-value is 2.405 which is greater than 2. Also check x_3 and continue to keep x_3, as the t-value is -3.719 which has a magnitude greater than 2.

   (c) Step 3: Do not select x_1, because the the t-value is 0.267 which is less than 2. Stop and the final model by the stepwise selection is the MLR `y~x2+x3`.

3. (a) The intercept b_0 is the expected clarity of lakes in ecoregion A. A 90% confidence interval for b_0 is $\hat{b}_0 \pm t_{0.05,12} \times s.e.(\hat{b}_0)$, which is $40.4 \pm 1.782 \times 1.92$ or 40.4 ± 3.42.

   (b) $H_0 : [b_1 = b_2 = 0 | b_0]$. From the `summary(ecolake.lm)` output, $f = 61.67$ on df $= (2, 12)$. The p-value is less than 0.001 and reject H_0 at 5% level. There is strong evidence that the expected lake clarity is not the same in all three ecoregions.

   (c) Since the clarity difference between A and C (i.e. b_1) is assumed to be 0, the test of interest is $H_0 : [b_2 = 0 | b_0]$. To find SS$(b_2|b_0)$, consider the fitting order of x_2, x_1 with

Source	df	SS
x_2	1	SS$(b_2\|b_0)$
x_1	1	SS$(b_1\|b_0, b_2)$
Error	12	SSE $= 221.20$

   Since $f = \frac{SS(b_1|b_0,b_2)/1}{221.20/12} = t^2 = (3.609)^2$, $SS(b_1|b_0, b_2) = 240.048$. Also SS$(b_2|b_0)$ + SS$(b_1|b_0, b_2)$ = SS$(b_1|b_0)$ + SS$(b_2|b_0, b_1) = 1293.63 + 980.10 = 2273.73$. Thus SS$(b_2|b_0) = 2273.73 - 240.048 = 2033.68$ on df $= 1$. Further SSE of the model `y~w2` is SS$(b_1|b_0, b_2)$ + SSE $= 240.048 + 221.20 = 461.248$ on df $=$ 13. Thus to test $H_0 : [b_2 = 0 | b_0]$, the observed $f = \frac{2033.68/1}{461.248/13} = 57.32$ on df $= (1, 13)$. The p-value is less than 0.001 and we reject H_0 at 5%. There is very strong evidence against H_0.

Grade Distribution

```
  100:5
90-99:17
80-89:28
70-79:17          mean = 80, median = 82
60-69:14
  <60:8
```