Ma 322: Biostatistics
Homework Assignment 9

Prof. Wickerhauser

Read Chapter 15, "ANOVA and Regression," pages 263–287 of our text.

NOTE: Machine-readable data for the problems below is in

`http://www.math.wustl.edu/~victor/classes/ma322/hw09data.txt`.

Cut and paste from that document into a text file, or into an `R` variable by use of the `scan()` function.

1. The following fake data mimics a study of amino acids in six imaginary species of millipedes:

   Alanine concentration in millipede hæmolymph (mg/100 ml)

Species 1	Species 2	Species 3	Species 4	Species 5	Species 6
21.5	14.5	16.0	14.8	12.1	14.4
19.6	17.4	20.3	15.6	11.4	14.7
20.9	15.0	18.5	13.5	12.7	13.8
22.8	17.8	19.3	16.4	14.5	12.0

   (a) Test, at the $\alpha = 0.05$ significance level, the hypothesis H_0: There is no difference in mean alanine concentration among the species. Use one-factor ANOVA.

   (b) Test, at the $\alpha = 0.05$ significance level, the hypothesis H_0: There is no difference in mean alanine concentration between species A and B. Use pairwise t-tests for every pair A, B.

   (c) Test, at the $\alpha = 0.05$ significance level, the hypothesis H_0: There is no difference in mean alanine concentration between species A and B. Use Tukey's HSD test for every pair A, B.

2. Test for all factor and interaction effects in the following 3×2 fixed-effects analysis of variance with equal replication:

```
         Response to Factors A and B
        a1            a2            a3
    ----------    ----------    ----------
     b1    b2      b1    b2      b1    b2
    ----  ----    ----  ----    ----  ----
    34.1  35.6    38.6  40.3    41.0  42.1
    36.9  36.3    39.1  41.3    41.4  42.7
    33.2  34.7    41.3  42.7    43.0  43.1
    35.1  35.8    41.4  41.9    43.4  44.8
    34.8  36.0    40.7  40.8    42.2  44.5
```

3. Test for all factor and interaction effects in the following $4 \times 3 \times 2$ fixed-effects analysis of variance, where a_i is the level of factor A, b_i is the level of factor B, and c_i is the level of factor C.

Response to Factors A, B and C

	a1			a2			a3			a4		
	b1	b2	b3	b1	b2	b3	b1	b2	b3	b1	b2	b3
c1:												
	4.1	4.6	3.7	4.9	5.2	4.7	5.0	6.1	5.5	3.9	4.4	3.7
	4.3	4.9	3.9	4.6	5.6	4.7	5.4	6.2	5.9	3.3	4.3	3.9
	4.5	4.2	4.1	5.3	5.8	5.0	5.7	6.5	5.6	3.4	4.7	4.0
	3.8	4.5	4.5	5.0	5.4	4.5	5.3	5.7	5.0	3.7	4.1	4.4
c2:												
	4.8	5.6	5.0	4.9	5.9	5.0	6.0	6.0	6.1	4.1	4.9	4.3
	4.5	5.8	5.2	5.5	5.3	5.4	5.7	6.3	5.3	3.9	4.7	4.1
	5.0	5.4	4.6	5.5	5.5	4.7	5.5	5.7	5.5	4.3	4.9	3.8
	4.6	6.1	4.9	5.3	5.7	5.1	5.7	5.9	5.8	4.0	5.3	4.7

4. Given the following data:

Y	X_1	X_2	X_3	X_4
51.4	0.2	17.8	24.6	18.9
72.0	1.9	29.4	20.7	8.0
53.2	0.2	17.0	18.5	22.6
83.2	10.7	30.2	10.6	7.1
57.4	6.8	15.3	8.9	27.3
66.5	10.6	17.6	11.1	20.8
98.3	9.6	35.6	10.6	5.6
74.8	6.3	28.2	8.8	13.1
92.2	10.8	34.7	11.9	5.9
97.9	9.6	35.8	10.8	5.5
88.1	10.5	29.6	11.7	7.8
94.8	20.5	26.3	6.7	10.0
62.8	0.4	22.3	26.5	14.3
58.4	6.6	15.7	8.7	26.3
81.6	2.3	37.9	20.0	0.5

(a) Fit the multiple regression $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$ to the data, computing the sample partial regression coefficients and Y intercept.

(b) Test $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ by ANOVA at the 0.05 level.

(c) Compute the standard error of each partial regression coefficient and test $H_0 : \beta_i = 0$, at the $\alpha = 0.05$ level, individually for each $i = 1, 2, 3, 4$.

(d) Calculate the standard error of estimate and the coefficient of determination.

(e) What is the predicted mean population value $\hat{Y}$ at $X_1 = 5.4$, $X_2 = 20.3$, $X_3 = 18.7$, $X_4 = 11.2$?

(f) What is the 95% confidence interval for $\hat{Y}$ in part (e)?

5. Perform a stepwise regression analysis of the data in Problem 4.

6. Analyze the five variables in Problem 4 as a multiple correlation.

   (a) Compute the simple correlation coefficient for each pair of variables.

   (b) Compute the multiple correlation coefficient R for each variable in terms of the other 4, and test $H_0 : R = 0$ at the 0.05 level in each case.

   (c) Compute the partial correlation coefficients for the five variables.

7. Each of five research papers was read by each of six reviewers. Each reviewer then marked the quality of the five papers as follows:

	Paper				
Reviewer	1	2	3	4	5
A	5	4	3	1	2
B	4	5	3	2	1
C	5	4	1	2	3
D	5	3	2	4	1
E	4	5	2	3	1
F	5	4	1	3	2

   (a) Calculate the Kendall coefficient of concordance.

   (b) Test, at the $\alpha = 0.01$ significance level, whether the rankings by the six reviewers are in agreement.