THE UNIVERSITY OF MINNESOTA

Statistics 5401/8401 October 5, 2005

Assignment Sheet No. 4

Reading

Week of October 3 - 7: J&W Sec. 6.3

Week of October 10-14: J&W, 6.4-6.9, review 7.1-7.6

Note: Sec. 7.1-7.6 is a summary of univariate (single response) multiple regression analysis. I assume you are familiar from it from Stat 5302 or equivalent.

Written Assignment (due in class Friday, October 14)

1. Below is a listing of data set `catches` in file `plankton.txt`. It is available through the class web page or directly at URL

`http://www.stat.umn.edu/~kb/classes/5401/files/data/plankton.txt`

```
catches         25    10   LABELS
) Data on plankton catches in a marine ecology study
) 25 successive hauls were made, each using two nets, one at
) 29 meters and one at 31 meters.  For each net in each haul,
) estimates were made of the "catch" (number caught) for 5
) varieties A, B, C, D, and E.
) Cols  1-5: log10(catch) for varieties A-E at 29 meters
) Cols 6-10: log10(catch) for varieties A-E at 31 meters
) Note this data is artificial but is based on real data.
)         29 Meters                    31 Meters
)   A     B     C     D     E     A     B     C     D     E
)"%lf %lf %lf %lf %lf %lf %lf %lf %lf %lf"
(10f6.3)
 2.321 1.537 3.332 2.945 1.907 2.339 1.783 3.093 2.316 1.744
 2.835 2.316 2.745 3.075 1.994 2.862 1.966 2.383 2.991 1.808
 2.748 2.007 2.929 3.107 1.838 2.578 1.605 2.352 3.054 1.435
 2.981 2.142 2.956 3.124 1.762 2.832 1.518 2.179 2.478 1.593
 2.130 1.386 2.586 2.876 1.670 2.186 1.478 2.428 3.096 1.419
 2.867 1.874 3.276 2.843 1.860 2.642 1.717 2.822 2.433 1.589
 2.492 1.896 2.352 3.065 1.949 2.560 1.792 1.881 2.757 1.265
 2.493 1.829 3.086 3.143 1.933 2.387 1.901 2.822 1.896 1.634
 2.573 1.885 2.863 3.290 1.663 2.388 1.421 2.575 2.410 1.293
 2.848 2.108 3.887 3.239 2.042 2.803 2.025 3.105 2.693 1.758
 3.443 2.130 3.888 2.922 1.927 3.460 1.897 3.252 2.665 2.084
 2.856 2.157 2.194 3.332 1.912 2.199 1.769 1.648 3.051 1.010
 2.986 2.469 2.778 3.502 1.728 2.364 2.042 2.205 2.839 1.256
 2.898 2.095 3.276 2.953 1.965 3.122 1.999 3.074 2.555 1.784
 2.548 2.027 2.648 3.447 2.120 2.378 2.032 2.030 2.639 1.426
 2.751 1.861 3.125 2.924 1.853 2.974 1.543 2.615 2.426 1.623
 3.106 1.787 3.522 3.362 1.920 2.902 1.568 3.013 2.969 1.529
 3.134 2.021 3.242 3.262 2.146 2.585 1.858 2.617 2.694 1.515
 2.812 2.455 3.261 3.069 1.877 2.646 2.117 2.613 2.685 1.502
 2.888 2.022 3.257 3.181 1.840 2.521 1.785 2.472 2.812 1.653
 2.924 2.373 2.873 3.479 1.935 2.356 1.770 1.946 2.988 1.142
 2.723 1.846 3.249 2.731 2.002 2.897 1.795 2.861 2.302 2.019
 2.513 1.848 2.934 3.037 1.747 2.565 1.676 2.379 2.825 1.580
 2.496 1.268 2.956 2.718 1.911 2.547 1.487 2.835 2.558 1.796
 2.808 1.699 3.224 2.995 1.722 2.576 1.828 2.745 2.722 1.618
```

The data matrix contains data on the catches of 5 varieties of plankton recorded by a marine research vessel. It has column labels `A_29`, `B_29`, `C_29`, `D_29`, `E_29`, `A_31`, `B_31`, `C_31`, `D_31` and `E_31` identifying the depth and variaty.

Each row contains data for a single haul, 25 in all. In each haul, two fine mesh nets were simultaneously dragged one at 29 meters and the other at 31 meters depth.

Separate counts of each plankton variety caught on a haul were made for each depth. This resulted in 10 numbers for each haul, 5 each from 29 and 31 meters. The values of the counts of A, B, C, D and E for depth 29 are in columns 1-5; similar counts for depth 31 m are in columns 6-10.

You should view this as multivariate data with n = 25 and p = 10. Each row corresponds to one case. You may assume the data consist of a random sample of size 25 from a multivariate normal population.

Since each number is measuring the same thing at a particular depth and for a particular variety, it is *repeated* measures data, with p = 10 variables and two *within-subject* factors, variety and depth.

(a) Is there statistical evidence at the 5% level that the expected catches of plankton were the same for all 10 variety-depth combinations? That is, test the null hypothesis that all p = 10 means (5 varieties at two depths) were the same.

Hint: Think of this as a one-sample problem with derived data consisting of any 9 independent comparisons among the 10 variety – depth combinations, say the comparisons of column 1 (variety A at depth 29) with the remaining 9 columns (variety B-E at depth 29 and variety A-E at depth 31). You can compute these by as the new data matrix $\mathbf{Y} = \mathbf{X}\,\mathbf{C}_1'$ where $\mathbf{C}_1 = [\mathbf{1}_9, -\mathbf{I}_9]$. Or you can compute it as `Y <- X[,1] - X[,-1]`, if `X` is the data matrix. Alternatively, you can directly compute the estimated means and variance matrix of these contrasts by $\bar{\mathbf{y}} = \mathbf{C}_1\bar{\mathbf{x}}$ and $\mathbf{S}_y = \mathbf{C}_1\mathbf{S}_x\mathbf{C}_1'$ and compute T^2 from them.

One way to compute $\mathbf{C}_1$ in MacAnova is

```
Cmd> ones9 <- rep(1,9); i9 <- dmat(9,1)
Cmd> c1 <- hconcat(ones9,-i9)
```

(b) Test the null hypothesis (use $\alpha = .05$) that, for each variety, the expected catches of plankton were the same at both depths? Do this two ways, using a Hotelling's T^2 statistic and Bonferronized t-tests. Identify all varieties, if any, which differed between depths.

Hint: Think of this as a multivariate paired comparison with 25 independent set of two 5-dimensional vectors, one at each depth coming from the first 5 and last 5 columns, respectively. Alternatively, you can approach it as a one-sample problem with derived data $\mathbf{X}\,\mathbf{C}_2'$, where $\mathbf{C}_2 = [\mathbf{I}_5, -\mathbf{I}_5]$.

One way to compute $\mathbf{C}_2$ in MacAnova is

```
Cmd> i5 <- dmat(5,1)
Cmd> c2 <- hconcat(i5, -i5)
```

(c) Test the null hypothesis that the between variety differences in expected catch were the same at each depth. This can be viewed as a test of no-interaction between variety and depth.

Hint: The appropriate test is essentially a paired Hotelling's T^2 on any four independent species differences, for example, A – B, A – C, A – D, and A – E or A – B, B – C, C – D, D – E. This can be transformed into a one sample problem by considering the difference between the contrasts for depth 29 and the contrasts for depth 31. These can be obtainedfrom $\mathbf{X}\,\mathbf{C}_3'$, where

$$\mathbf{C}_3 = \begin{bmatrix}\mathbf{1}_4 & -\mathbf{I}_4 & -\mathbf{1}_4 & \mathbf{I}_4\end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The rows of $\mathbf{C}_3$ define contrasts.

One way to create $\mathbf{C}_3$ in MacAnova is the following:

```
Cmd> ones4 <- rep(1,4); i4 <- dmat(4,1)
Cmd> c3 <- hconcat(ones4,-i4,-ones4, i4)
```

2. J&W Ex. 6.16

3. J&W Ex. 6.27