Stat/For/Hort 572 – Midterm II, Spring 2005 — Brief Solutions for Problem 2

The study is a completely randomized design with subsampling. The colonies are the samples and the two repeats are the subsamples. Arrange the data in the file `bark.dat`.

```
"environment" "id" "distance"
1 1 44
1 1 52
1 2 42
1 2 66
1 3 53
1 3 42
1 4 55
1 4 57
2 1 46
2 1 37
2 2 55
2 2 44
2 3 34
2 3 18
2 4 51
2 4 40
3 1 30
3 1 39
3 2 38
3 2 36
3 3 39
3 3 38
3 4 31
3 4 10
```

We analyze the data using the following SAS commands.

```
data bark;
    infile "bark.dat" missover firstobs=2;
    input env id y;

proc glm data=bark;
    class env id;
    model y=env id(env);
    output out=newbark p=pred r=resid;
    test h=env e=id(env);
    means env / lsd e=id(env);
    contrast 'rural vs urban+bound' env 2 -1 -1 / e=id(env);
    contrast 'urban vs rural+bound' env -1 -1 2 / e=id(env);
    run;

proc plot data=newbark;
    plot resid*pred/vref=0;
    run;
```

The (edited) SAS output looks like the following. The residual plot at the sampling level (not shown) looks fine when checking for the usual model assumptions, even though the residual plot at the subsampling level (not shown) shows a curious pattern.

```
Dependent Variable: y

Source                     DF        Squares    Mean Square   F Value   Pr > F
Model                      11    2568.458333     233.496212      3.00   0.0361
Error                      12     935.500000      77.958333
Corrected Total            23    3503.958333

Source                     DF      Type I SS    Mean Square   F Value   Pr > F
env                         2    1416.333333     708.166667      9.08   0.0040
id(env)                     9    1152.125000     128.013889      1.64   0.2085

Source                     DF    Type III SS    Mean Square   F Value   Pr > F
env                         2    1416.333333     708.166667      9.08   0.0040
id(env)                     9    1152.125000     128.013889      1.64   0.2085

    Tests of Hypotheses Using the Type III MS for id(env) as an Error Term
Source                     DF    Type III SS    Mean Square   F Value   Pr > F
env                         2    1416.333333     708.166667      5.53   0.0271

                          t Tests (LSD) for y
```

```
NOTE: This test controls the Type I comparisonwise error rate, not the
                      experimentwise error rate.
                  Alpha                              0.05
                  Error Degrees of Freedom              9
                  Error Mean Square             128.0139
                  Critical Value of t            2.26216
                  Least Significant Difference    12.797

                          t Tests (LSD) for y
      Means with the same letter are not significantly different.
              t Grouping          Mean      N    env

                     A        51.375      8    1
                     A
                  B  A        40.625      8    2
                  B
                  B           32.625      8    3

   Tests of Hypotheses Using the Type III MS for id(env) as an Error Term
Contrast                  DF    Contrast SS    Mean Square   F Value   Pr > F
rural vs urban+bound       1    1160.333333    1160.333333      9.06   0.0147
urban vs rural+bound       1     954.083333     954.083333      7.45   0.0232
```

(a) Based on the SAS printout, we conclude that there is moderate evidence of an environment effect (p = 0.027).

(b) The LSD suggests that there is a difference between the rural environment and the urban environment, but there is no difference between the rural environment and the boundary, nor between the urban environment and the boundary.

(c) By the two contrasts, we conclude that there is moderate evidence that there is a difference between the rural area and the urban/urban-rural boundary area (p = 0.015) and that there is moderate evidence that there is a difference between the urban area and the urban/urban-rural boundary area (p = 0.023).

(d) The 95% confidence interval of the within colony variation is

$$\frac{\text{SSE}}{\chi^2_{12,0.975}} < \sigma^2 < \frac{\text{SSE}}{\chi^2_{12,0.025}},$$

where SSE = 935.5, $\chi^2_{12,0.025} = 4.40$, and $\chi^2_{12,0.975} = 23.34$. Hence the CI is (40.08, 212.61).

(e) It would be helpful to repeat the experiment more than twice in each colony, because the estimated within-colony variance component is 77.96 whereas the estimated among-colony variance component is

$$\frac{128.01 - 77.96}{2} = 25.03$$

which is smaller.