Stat/For/Hort 572 – Midterm II, Spring 2004 — Brief Solutions for Problem 1

First we assess the effect of habitat types on the abundance using a one-way ANOVA approach. From the following output, we conclude that there is strong evidence of a habitat effect ($p < 0.0001$). Further pairwise comparison using the LSD suggests that the habit types are different from one another, with the exception that the clay disturbed and the sandy undisturbed are not significantly different. The residual plot reveals a possible fan shape but otherwise the 4 model assumptions seem to be satisfied.

```
Dependent Variable: abund
                                       Sum of
Source                     DF         Squares    Mean Square   F Value   Pr > F
Model                       3     19643.35000     6547.78333     50.53   <.0001
Error                      16      2073.20000      129.57500
Corrected Total            19     21716.55000

                           t Tests (LSD) for abund
                    Alpha                            0.05
                    Error Degrees of Freedom           16
                    Error Mean Square             129.575
                    Critical Value of t           2.11991
                    Least Significant Difference   15.262

               t Grouping          Mean      N    type

                        A       252.400      5    1

                        B       208.400      5    2
                        B
                        B       206.000      5    3

                        C       163.800      5    4
```

Next we relate the abundance to the environmental variables using MLR. The full model has two explanatory variables, namely, soil moisture ($p < 0.001$) and soil temperature ($p = 0.13$). Notice potential multicollinearity due to a fairly high negative correlation between soil moisture and soil temperature ($r = -0.77$).

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  85.0241    37.6017   2.261   0.0372 *
moist         7.9440     0.6961  11.412 2.16e-09 ***
temp         -1.3888     0.8679  -1.600   0.1280

Residual standard error: 7.201 on 17 degrees of freedom
Multiple R-Squared: 0.9594,    Adjusted R-squared: 0.9546
```

By a backward elimination, the best model has only soil moisture as the explanatory variable. The residual plot does not reveal any serious departure from the 4 model assumptions.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.6799     9.5891   2.782   0.0123 *
moist         8.8063     0.4594  19.168 2.00e-13 ***

Residual standard error: 7.506 on 18 degrees of freedom
Multiple R-Squared: 0.9533,    Adjusted R-squared: 0.9507
```

Finally we study the effect of soil moisture and soil temperature on the abundance for different habitat types using MLR. We create new variables `i1,i2,i3` to represent the interaction between `moist` and `w1,w2,w3`, and new variables `ii1,ii2,ii3` to represent the interaction between `temp` and `w1,w2,w3`. For model comparisons, we fit several models assuming equal intercepts and/or equal slopes, in addition to the full model. Using the principle of additional sum of squares, it appears that the intercepts and slopes for the models `abund`∼`moist+temp` are different for different habitat types ($p = 0.0063$) and the difference is likely due to the differences in the slopes ($p = 0.029$) rather than the differences in the intercepts ($p = 0.66$). For further understanding of the data, one could compare the regression in a pairwise fashion or for any specific habitat types of particular interest.

```
> abund.lm <- lm(abund~moist+temp+w1+w2+w3+i1+i2+i3+ii1+ii2+ii3,data=abund);
> summary(abund.lm);

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 138.2308    37.1840   3.717  0.00589 **
moist         8.5385     0.9412   9.072 1.75e-05 ***
temp         -3.8846     1.2801  -3.035  0.01620 *
w1          -46.7762    99.3745  -0.471  0.65042
w2          -89.8478    79.2290  -1.134  0.28962
w3            1.3216    48.5233   0.027  0.97894
i1           -1.9930     3.1321  -0.636  0.54235
i2           -2.2406     1.5044  -1.489  0.17473
i3           -8.2147     1.7827  -4.608  0.00174 **
ii1           3.1573     1.7545   1.800  0.10962
ii2           4.8846     2.1104   2.315  0.04933 *
ii3           4.4751     1.6322   2.742  0.02538 *

# test for equal intercepts and equal slopes
> abund.lm1 <- lm(abund~moist+temp,data=abund);
> anova(abund.lm1, abund.lm);

Model 1: abund ~ moist + temp
Model 2: abund ~ moist + temp + w1 + w2 + w3 + i1 + i2 + i3 + ii1 + ii2 + ii3
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1     17 881.41
2      8 101.34  9    780.08 6.8426 0.006278 **

# test for equal slopes, given that the intercepts are different
> abund.lm2 <- lm(abund~moist+temp+w1+w2+w3,data=abund);
> anova(abund.lm2, abund.lm);

Model 1: abund ~ moist + temp + w1 + w2 + w3
Model 2: abund ~ moist + temp + w1 + w2 + w3 + i1 + i2 + i3 + ii1 + ii2 + ii3
  Res.Df    RSS Df Sum of Sq     F  Pr(>F)
1     14 435.90
2      8 101.34  6    334.56 4.402 0.02910 *

# test for equal intercepts, given that the slopes are different
> abund.lm3 <- lm(abund~moist+temp+i1+i2+i3+ii1+ii2+ii3,data=abund);
> anova(abund.lm3, abund.lm);

Model 1: abund ~ moist + temp + i1 + i2 + i3 + ii1 + ii2 + ii3
Model 2: abund ~ moist + temp + w1 + w2 + w3 + i1 + i2 + i3 + ii1 + ii2 + ii3
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     11 122.421
2      8 101.336  3    21.085 0.5549 0.6593
```