Stat/For/Hort 572 – Midterm II, Spring 2005 — Brief Solutions for Problem 1

First we assess the relationship between toxicity and the treatment (i.e., chemotherapy and dose rate). We create a new variable `dcint` to represent the interaction between `dose` and `chemo`. For model comparisons, we evaluate the regression of toxicity on dose rate for the patients who received chemotherapy versus those who did not receive chemotherapy. From the following output, we conclude that there is strong evidence that the slopes of the regression are different for the two groups of patients ($p = 0.00167$) but there is no evidence that the intercepts are different ($p = 0.51$). Note that we could also fit a reduced model with only the term `dose` and use the principle of sum of squares to conclude that the there is strong evidence that the intercepts and the slopes are not all equal ($f = 131.47$ on $df = (2, 36)$, $p < 0.0001$). The residual plot does not reveal any obvious violation of the model assumptions (i.e., correct model, equal variance, and normal distribution).

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.68704   13.46001   1.537  0.13305
dose         -0.03804    0.51121  -0.074  0.94110
chemo       -12.62091   19.11855  -0.660  0.51337
dcint         2.51736    0.74089   3.398  0.00167 **
---

Residual standard error: 10.19 on 36 degrees of freedom
Multiple R-Squared: 0.8803,     Adjusted R-squared: 0.8703
F-statistic: 88.26 on 3 and 36 DF,  p-value: < 2.2e-16
```

Next we assess the relationship between toxicity and the patient information (i.e., age and smoking history). We create a new variable `asint` to represent the interaction between `age` and `smoke`. For model comparisons, we evaluate the regression of toxicity on age for the patients who are smokers versus those who are non-smokers. From the following output, there is no evidence that the slopes of the regression are different for the two groups of patients ($p = 0.49$) and there is no evidence that the intercepts are different ($p = 0.49$). Note that we could also fit a reduced model with only the term `age` and use the principle of sum of squares to conclude that there is no evidence that the intercepts and the slopes are different ($f = 0.24$ on $df = (2, 36)$, $p = 0.79$). The residual plot does not reveal any obvious violation of the model assumptions (i.e., correct model, equal variance, and normal distribution).

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.3830    35.7140  -0.179    0.859
age           1.2334     0.8298   1.486    0.146
smoke        36.4452    52.7564   0.691    0.494
asint        -0.8865     1.2744  -0.696    0.491
---

Residual standard error: 28.51 on 36 degrees of freedom
Multiple R-Squared: 0.06225,    Adjusted R-squared: -0.0159
F-statistic: 0.7966 on 3 and 36 DF,  p-value: 0.5039
```

Finally we study the relationship between toxicity and all the independent variables. We create new variables `doseint1, doseint2, doseint3` to represent the interaction between `dose` and `w1, w2, w3`, and new variables `ageint1, ageint2, ageint3` to represent the interaction between `age` and `w1, w2, w3`. For model comparisons, we fit several models assuming equal intercepts and/or equal slopes, in addition to the full models. Using the principle of additional sum of squares, we conclude that there is strong evidence of differences among the intercepts and the slopes ($p < 0.0001$). Furthermore there is no evidence of difference among the intercepts ($p = 0.59$) but there is moderate evidence of difference among the slopes ($p = 0.024$). For further understanding of the data, one could assess the difference among the slopes for drug doses and that for age separately. The conclusion is similar to (a) and (b) such that there is evidence that the toxicity is related to dose rate and the relationship is different for the patients who received chemotherapy versus those who did not receive chemotherapy. There is no obvious evidence of relationship between toxicity and patient's age and smoking history.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.70389   29.42991   0.432   0.6693
dose         -0.61130    0.77833  -0.785   0.4388
age           0.55078    0.62317   0.884   0.3843
w1          -25.81164   34.78511  -0.742   0.4642
w2          -15.66020   35.05533  -0.447   0.6585
w3          -48.31944   37.63491  -1.284   0.2097
doseint1      0.31921    0.97656   0.327   0.7462
doseint2      2.64278    1.00070   2.641   0.0134 *
doseint3      2.95080    1.21297   2.433   0.0216 *
ageint1       0.43398    0.73599   0.590   0.5601
ageint2       0.03573    0.68969   0.052   0.9591
ageint3       0.59521    0.96883   0.614   0.5439
---

Residual standard error: 8.951 on 28 degrees of freedom
Multiple R-Squared: 0.9281,     Adjusted R-squared: 0.8999
F-statistic: 32.86 on 11 and 28 DF,  p-value: 4.379e-13

# test for equal intercepts and equal slopes
Model 1: y ~ dose + age
Model 2: y ~ dose + age + w1 + w2 + w3 + doseint1 + doseint2 + doseint3 +
    ageint1 + ageint2 + ageint3
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     37 29631.1
2     28  2243.3  9   27387.9 37.984 2.323e-13 ***
---

# test for equal intercepts, given that the slopes are different
Model 1: y ~ dose + age + doseint1 + doseint2 + doseint3 + ageint1 + ageint2 +
    ageint3
Model 2: y ~ dose + age + w1 + w2 + w3 + doseint1 + doseint2 + doseint3 +
    ageint1 + ageint2 + ageint3
  Res.Df    RSS Df Sum of Sq     F Pr(>F)
1     31 2398.8
2     28 2243.2  3     155.5 0.647 0.5914

# test for equal slopes, given that the intercepts are different
Model 1: y ~ dose + age + w1 + w2 + w3
Model 2: y ~ dose + age + w1 + w2 + w3 + doseint1 + doseint2 + doseint3 +
    ageint1 + ageint2 + ageint3
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     34 3649.9
2     28 2243.3  6    1406.6 2.9262 0.02414 *
---
```