Stat 8311, Fall 2006, Barley data

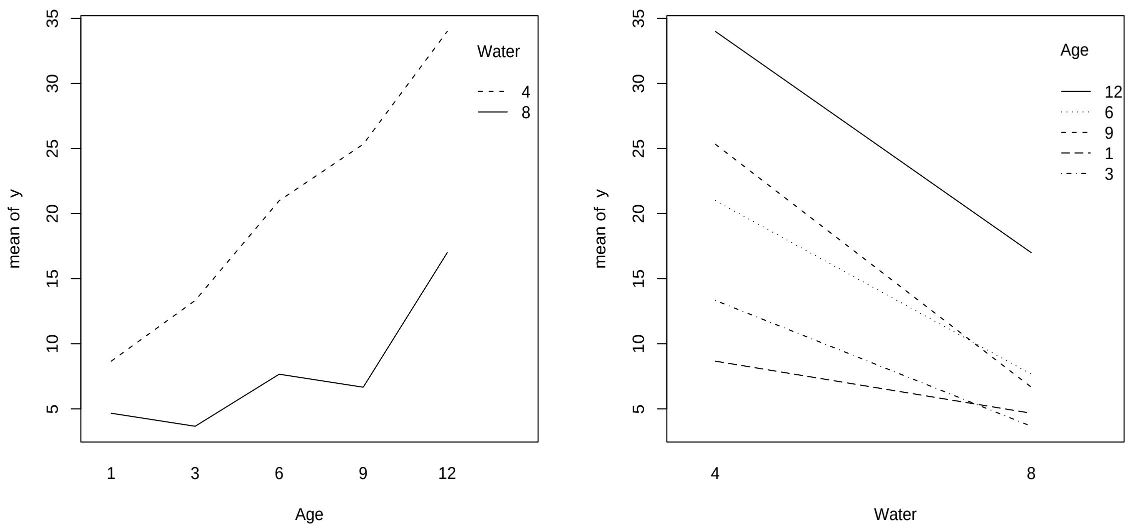
This handout uses the Barley sprouting data from Oehlert, p. 166. We first fit the usual way:

```
> barley <- data.frame(Age <- factor(rep(c(1, 3, 6, 9, 12),
+     c(6, 6, 6, 6, 6))), Water <- factor(rep(rep(c(4, 8),
+     c(3, 3)), 5)), Rep = factor(rep(c("R1", "R2", "R3"),
+     10)), y = c(11, 9, 6, 8, 3, 3, 7, 16, 17, 1, 7, 3, 9,
+     19, 35, 5, 9, 9, 13, 35, 28, 1, 10, 9, 20, 37, 45, 11,
+     15, 25))
> xtabs(y ~ Rep + with(barley, Age:Water), data = barley)
```

```
    with(barley, Age:Water)
Rep  1:4 1:8 3:4 3:8 6:4 6:8 9:4 9:8 12:4 12:8
  R1  11   8   7   1   9   5  13   1   20   11
  R2   9   3  16   7  19   9  35  10   37   15
  R3   6   3  17   3  35   9  28   9   45   25
```

```
> attach(barley)
> interaction.plot(Age, Water, y)
```

```
> interaction.plot(Water, Age, y)
```



Here are some summaries of the data:

```
> mean(y)
```

```
[1] 14.2
```

```
> tapply(y, Age, mean)
```

```
        1         3         6         9        12
 6.666667  8.500000 14.333333 16.000000 25.500000
```

```
> tapply(y, list(Water, Age), mean)
```

```
         1         3         6         9 12
4 8.666667 13.333333 21.000000 25.333333 34
8 4.666667  3.666667  7.666667  6.666667 17
```

```
> tapply(y, Water, mean)
```

```
        4         8
20.466667  7.933333
```

The option `contrasts` controls how R selects the basis for various subspaces. The default is the "drop first level" parameterization, which does not provide a basis for the row space orthogonal to the column space.

```
> getOption("contrasts")
```

```
        unordered           ordered
"contr.treatment"      "contr.poly"
```

```
> contrasts(barley$Age)
```

```
   3 6 9 12
1  0 0 0  0
3  1 0 0  0
6  0 1 0  0
9  0 0 1  0
12 0 0 0  1
```

```
> summary(m1 <- lm(y ~ 1 + Age + Water + Water:Age, data = barley))
```

```
Call:
lm(formula = y ~ 1 + Age + Water + Water:Age, data = barley)

Residuals:
     Min       1Q   Median       3Q      Max
-14.0000  -2.6667   0.8333   3.2500  14.0000

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)     8.667      4.462   1.942  0.06632
Age3            4.667      6.310   0.740  0.46819
Age6           12.333      6.310   1.954  0.06478
Age9           16.667      6.310   2.641  0.01567
Age12          25.333      6.310   4.014  0.00068
Water8         -4.000      6.310  -0.634  0.53335
Age3:Water8    -5.667      8.924  -0.635  0.53265
Age6:Water8    -9.333      8.924  -1.046  0.30812
Age9:Water8   -14.667      8.924  -1.643  0.11592
Age12:Water8  -13.000      8.924  -1.457  0.16072

Residual standard error: 7.729 on 20 degrees of freedom
Multiple R-Squared: 0.6939,        Adjusted R-squared: 0.5561
F-statistic: 5.037 on 9 and 20 DF,  p-value: 0.001269
```

Here is the upper 5×5 submatrix of the covariance matrix of the estimated coefficients. The grand mean is not independent of the row effects

```
> round(vcov(m1)[1:5, 1:5])

            (Intercept) Age3 Age6 Age9 Age12
(Intercept)          20  -20  -20  -20   -20
Age3                -20   40   20   20    20
Age6                -20   20   40   20    20
Age9                -20   20   20   40    20
Age12               -20   20   20   20    40
```

We can change to the "zero-sum" parameterization, or indeed, *any* parameterization, by changing value of the `contrasts` argument to `options`.

```
> opt <- options(contrasts = c("contr.sum", "contr.poly"))
> contrasts(barley$Age)

   [,1] [,2] [,3] [,4]
1     1    0    0    0
3     0    1    0    0
6     0    0    1    0
9     0    0    0    1
12   -1   -1   -1   -1

> summary(m2 <- update(m1))

Call:
lm(formula = y ~ 1 + Age + Water + Water:Age, data = barley)

Residuals:
     Min       1Q   Median       3Q      Max
-14.0000  -2.6667   0.8333   3.2500  14.0000

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.2000     1.4111  10.063 2.84e-09
Age1         -7.5333     2.8221  -2.669 0.014731
Age2         -5.7000     2.8221  -2.020 0.057011
Age3          0.1333     2.8221   0.047 0.962786
Age4          1.8000     2.8221   0.638 0.530829
Water1        6.2667     1.4111   4.441 0.000251
Age1:Water1  -4.2667     2.8221  -1.512 0.146211
Age2:Water1  -1.4333     2.8221  -0.508 0.617086
Age3:Water1   0.4000     2.8221   0.142 0.888706
Age4:Water1   3.0667     2.8221   1.087 0.290116

Residual standard error: 7.729 on 20 degrees of freedom
Multiple R-Squared: 0.6939,      Adjusted R-squared: 0.5561
F-statistic: 5.037 on 9 and 20 DF,  p-value: 0.001269
```

```
> round(vcov(m2)[1:5, 1:5])

            (Intercept) Age1 Age2 Age3 Age4
(Intercept)           2    0    0    0    0
Age1                  0    8   -2   -2   -2
Age2                  0   -2    8   -2   -2
Age3                  0   -2   -2    8   -2
Age4                  0   -2   -2   -2    8
```

Since Age is numeric, would could use an orthogonal polynomial parameterization. First we return to the original options, and then define Page with the "right" orthogonal polynomials.

```
> options(opt)
> barley$Page <- C(barley$Age, contr.poly(5, c(1, 3, 6, 9,
+     12)), 4)
> contrasts(barley$Page)

            .L          .Q          .C          ^4
1  -0.58578768  0.5127027 -0.3888972  0.2067551
3  -0.36048473 -0.1356181  0.5799242 -0.5615571
6  -0.02253030 -0.5650755  0.1615028  0.6738686
9   0.31542414 -0.3429044 -0.6368217 -0.4211679
12  0.65337857  0.5308954  0.2842919  0.1021013

> summary(o1 <- lm(y ~ Page * Water, data = barley))$coef

                 Estimate Std. Error     t value     Pr(>|t|)
(Intercept)    20.4666667   1.995551 10.25615016 2.064265e-09
Page.L         19.8491903   4.462187  4.44831002 2.467812e-04
Page.Q          0.1321266   4.462187  0.02961028 9.766714e-01
Page.C          1.2865479   4.462187  0.28832229 7.760669e-01
Page^4          1.2575477   4.462187  0.28182318 7.809746e-01
Water8        -12.5333333   2.822135 -4.44108256 2.509756e-04
Page.L:Water8 -10.8671125   6.310485 -1.72207244 1.004872e-01
Page.Q:Water8   4.1701656   6.310485  0.66083123 5.162611e-01
Page.C:Water8   0.8506597   6.310485  0.13480101 8.941166e-01
Page^4:Water8   1.7425288   6.310485  0.27613231 7.852797e-01

> anova(m1)

Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
Age        4 1321.13  330.28  5.5293 0.003645
Water      1 1178.13 1178.13 19.7232 0.000251
Age:Water  4  208.87   52.22  0.8742 0.496726
Residuals 20 1194.67   59.73

> anova(m2)
```

```
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
Age        4 1321.13  330.28  5.5293 0.003645
Water      1 1178.13 1178.13 19.7232 0.000251
Age:Water  4  208.87   52.22  0.8742 0.496726
Residuals 20 1194.67   59.73
```

```
> anova(o1)
```

```
Analysis of Variance Table

Response: y
           Df  Sum Sq Mean Sq F value   Pr(>F)
Page        4 1321.13  330.28  5.5293 0.003645
Water       1 1178.13 1178.13 19.7232 0.000251
Page:Water  4  208.87   52.22  0.8742 0.496726
Residuals  20 1194.67   59.73
```

Often, one would like to see the individual contrasts in the anova table, and the only way I know to do this in R is to add individual polynomial terms. I'll just do the linear here. Creating a numeric variable from a factor requires some ridiculous R code:

```
> barley$nAge <- as.numeric(as.character(barley$Age))

> anova(o2 <- update(o1, ~nAge + .))
```

```
Analysis of Variance Table

Response: y
           Df  Sum Sq Mean Sq F value    Pr(>F)
nAge        1 1246.86 1246.86 20.8738 0.0001864
Page        3   74.27   24.76  0.4145 0.7444701
Water       1 1178.13 1178.13 19.7232 0.0002510
Page:Water  4  208.87   52.22  0.8742 0.4967256
Residuals  20 1194.67   59.73
```

```
> anova(o3 <- update(o1, ~nAge + Page + Water + nAge:Water +
+     Page:Water))
```

```
Analysis of Variance Table

Response: y
           Df  Sum Sq Mean Sq F value    Pr(>F)
nAge        1 1246.86 1246.86 20.8738 0.0001864
Page        3   74.27   24.76  0.4145 0.7444701
Water       1 1178.13 1178.13 19.7232 0.0002510
nAge:Water  1  177.14  177.14  2.9655 0.1004872
Page:Water  3   31.73   10.58  0.1770 0.9106910
Residuals  20 1194.67   59.73
```

All of the above suggest that the *Age* effect can be effectively captured by a linear effect only, possibly with a common intercept.

```
> summary(update(o1, y ~ nAge * Water))

Call:
lm(formula = y ~ nAge + Water + nAge:Water, data = barley)

Residuals:
     Min       1Q   Median       3Q      Max
-13.7276  -2.7843   0.3283   2.5926  14.9805

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.6032     3.3867   1.950   0.0621
nAge          2.2360     0.4600   4.861 4.85e-05
Water8       -4.9433     4.7895  -1.032   0.3115
nAge:Water8  -1.2242     0.6506  -1.882   0.0711

Residual standard error: 7.073 on 26 degrees of freedom
Multiple R-Squared: 0.6667,       Adjusted R-squared: 0.6283
F-statistic: 17.34 on 3 and 26 DF,  p-value: 2.175e-06

> summary(update(o1, y ~ nAge:Water))

Call:
lm(formula = y ~ nAge:Water, data = barley)

Residuals:
     Min       1Q   Median       3Q      Max
-14.3569  -2.9594   0.1248   3.0153  15.7558

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.1316     2.3976   1.723   0.0963
nAge:Water4   2.5188     0.3700   6.807  2.6e-07
nAge:Water8   0.7291     0.3700   1.971   0.0591

Residual standard error: 7.081 on 27 degrees of freedom
Multiple R-Squared: 0.6531,       Adjusted R-squared: 0.6274
F-statistic: 25.41 on 2 and 27 DF,  p-value: 6.21e-07
```

Unbalanced

We have made extensive use of the equal number of replications per cell because in that case the orthogonalization we used of main effects, rows given main effects, columns given rows and main effects, interactions given rows and columns, would be the same regardless of the order. With unbalance, *order matters*.

```
> barley$sel <- rep(TRUE, 30)
> barley$sel[c(1, 2, 18, 25)] <- FALSE
```

```
> m2 <- update(m1, subset = sel, data = barley)
> o2 <- update(o1, subset = sel, data = barley)
> anova(m2)

Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
Age        4 1326.60  331.65  5.9937 0.0038190
Water      1 1290.86 1290.86 23.3289 0.0001847
Age:Water  4  299.70   74.93  1.3541 0.2934098
Residuals 16  885.33   55.33

> anova(o2)

Analysis of Variance Table

Response: y
           Df  Sum Sq Mean Sq F value    Pr(>F)
Page        4 1326.60  331.65  5.9937 0.0038190
Water       1 1290.86 1290.86 23.3289 0.0001847
Page:Water  4  299.70   74.93  1.3541 0.2934098
Residuals  16  885.33   55.33

> drop1(m2, scope = ~Age * Water)

Single term deletions

Model:
y ~ 1 + Age + Water + Water:Age
          Df Sum of Sq     RSS     AIC
<none>                  885.33  111.72
Age        4   1238.92 2124.25  126.48
Water      1      1.33  886.67  109.76
Age:Water  4    299.70 1185.04  111.31

> drop1(o2, scope = ~Page * Water)

Single term deletions

Model:
y ~ Page * Water
           Df Sum of Sq     RSS     AIC
<none>                   885.33  111.72
Page        4   1238.92 2124.25  126.48
Water       1   1056.64 1941.97  130.15
Page:Water  4    299.70 1185.04  111.31
```