Stat 8311 – Unbalanced data from Searle, Table 4.2

```
> searle <- data.frame(soil = rep(c("s1", "s2"), c(7, 8)),
+     var = c("v1", "v2", "v3")[c(1, 1, 1, 2, 2, 3, 3, 1, 1,
+         1, 1, 2, 3, 3, 3)], y = c(6, 10, 11, 13, 15, 14,
+         22, 12, 15, 19, 18, 31, 18, 9, 12))
```

Here is listing of the data file, and counts of the number of observations in each cell:

```
> searle

   soil var  y
1    s1  v1  6
2    s1  v1 10
3    s1  v1 11
4    s1  v2 13
5    s1  v2 15
6    s1  v3 14
7    s1  v3 22
8    s2  v1 12
9    s2  v1 15
10   s2  v1 19
11   s2  v1 18
12   s2  v2 31
13   s2  v3 18
14   s2  v3  9
15   s2  v3 12
```

```
> xtabs(~soil + var, data = searle)

    var
soil v1 v2 v3
  s1  3  2  2
  s2  4  1  3
```

Next, the observed cell means:

```
> xtabs(y ~ soil + var, data = searle)/xtabs(~soil + var, data = searle)

    var
soil v1 v2 v3
  s1  9 14 18
  s2 16 31 13
```

The function `Anova` in the `library(car)` computes Type II and Type III anova:

```
> opt <- options(contrasts = c("contr.SAS", "contr.poly"))
> coef(m1 <- aov(y ~ soil * var, data = searle))

 (Intercept)       soils1        varv1        varv2 soils1:varv1
          13            5            3           18          -12
soils1:varv2
         -22
```

```
> print(anova(m1), digits = 5)

Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
soil       1  52.500  52.500  3.9375 0.078512
var        2 124.734  62.367  4.6775 0.040475
soil:var   2 222.766 111.383  8.3537 0.008888
Residuals  9 120.000  13.333

> library(car)
> print(Anova(m1, type = "II"), digits = 5)

Anova Table (Type II tests)

Response: y
           Sum Sq Df F value   Pr(>F)
soil       83.901  1  6.2926 0.033393
var       124.734  2  4.6775 0.040475
soil:var  222.766  2  8.3537 0.008888
Residuals 120.000  9

> print(Anova(m1, type = "III"), digits = 6)

Anova Table (Type III tests)

Response: y
             Sum Sq Df  F value     Pr(>F)
(Intercept) 507.000  1 38.02500 0.00016541
soil         30.000  1  2.25000 0.16785066
var         247.500  2  9.28125 0.00649617
soil:var    222.766  2  8.35372 0.00888845
Residuals   120.000  9
```

We can also see that order of fitting matters by comparing projections on subspaces, and estimated marginal means, depending on the order specified:

```
> proj(m1)[, ]

   (Intercept)  soil        var  soil:var   Residuals
1           15 -2.00 -2.7659574 -1.234043 -3.0000e+00
2           15 -2.00 -2.7659574 -1.234043  1.0000e+00
3           15 -2.00 -2.7659574 -1.234043  2.0000e+00
4           15 -2.00  5.0531915 -4.053191 -1.0000e-00
5           15 -2.00  5.0531915 -4.053191  1.0000e-00
6           15 -2.00 -0.9042553  5.904255 -4.0000e+00
7           15 -2.00 -0.9042553  5.904255  4.0000e+00
8           15  1.75 -1.6755319  0.925532 -4.0000e+00
9           15  1.75 -1.6755319  0.925532 -1.0000e-00
```

```
10          15  1.75 -1.6755319  0.925532  3.0000e+00
11          15  1.75 -1.6755319  0.925532  2.0000e+00
12          15  1.75  6.1436170  8.106383 -4.1281e-16
13          15  1.75  0.1861702 -3.936170  5.0000e+00
14          15  1.75  0.1861702 -3.936170 -4.0000e+00
15          15  1.75  0.1861702 -3.936170 -1.0000e+00
```

```
> model.tables(m1, type = "mean")
```

```
Tables of means
Grand mean

15

 soil
    s1    s2
    13 16.75
rep  7  8.00

 var
       v1    v2    v3
    12.86 20.42 14.75
rep  7.00  3.00  5.00

 soil:var
     var
soil  v1 v2 v3
  s1   9 14 18
  rep  3  2  2
  s2  16 31 13
  rep  4  1  3
```

```
> proj(m2 <- update(m1, ~var * soil))[, ]
```

```
   (Intercept)           var      soil  var:soil     Residuals
1           15 -2.000000e+00 -2.765957 -1.234043 -3.000000e+00
2           15 -2.000000e+00 -2.765957 -1.234043  1.000000e+00
3           15 -2.000000e+00 -2.765957 -1.234043  2.000000e+00
4           15  4.666667e+00 -1.613475 -4.053191 -1.000000e+00
5           15  4.666667e+00 -1.613475 -4.053191  1.000000e+00
6           15  8.881784e-16 -2.904255  5.904255 -4.000000e+00
7           15  8.881784e-16 -2.904255  5.904255  4.000000e+00
8           15 -2.000000e+00  2.074468  0.925532 -4.000000e+00
9           15 -2.000000e+00  2.074468  0.925532 -1.000000e-00
10          15 -2.000000e+00  2.074468  0.925532  3.000000e+00
11          15 -2.000000e+00  2.074468  0.925532  2.000000e+00
12          15  4.666667e+00  3.226950  8.106383 -1.761829e-16
13          15  8.881784e-16  1.936170 -3.936170  5.000000e+00
14          15  8.881784e-16  1.936170 -3.936170 -4.000000e+00
15          15  8.881784e-16  1.936170 -3.936170 -1.000000e+00
```

```
> model.tables(m2, type = "mean")

Tables of means
Grand mean

15

 var
    v1    v2 v3
    13 19.67 15
rep  7  3.00  5

 soil
       s1    s2
    12.52 17.17
rep  7.00  8.00

 var:soil
     soil
var   s1 s2
  v1   9 16
  rep  3  4
  v2  14 31
  rep  2  1
  v3  18 13
  rep  2  3
```