Chapter 8

Experimental design

1 Studies in crop variation

R. A. Fisher created a lot of statistical theory (which is still heavily used) while working at Rothamsted agricultural research station. In particular, Fisher developed a method of designing and analyzing complex experiemnts. (For more about the history see a presentation by Roger Payne.)

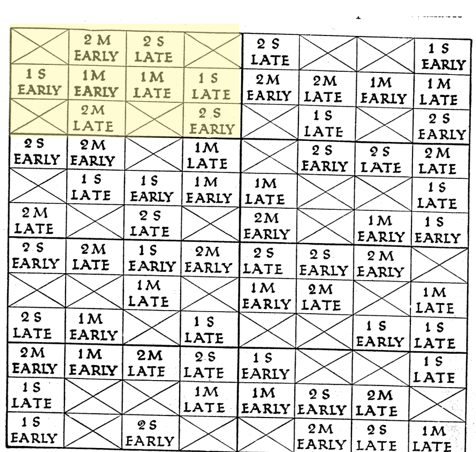
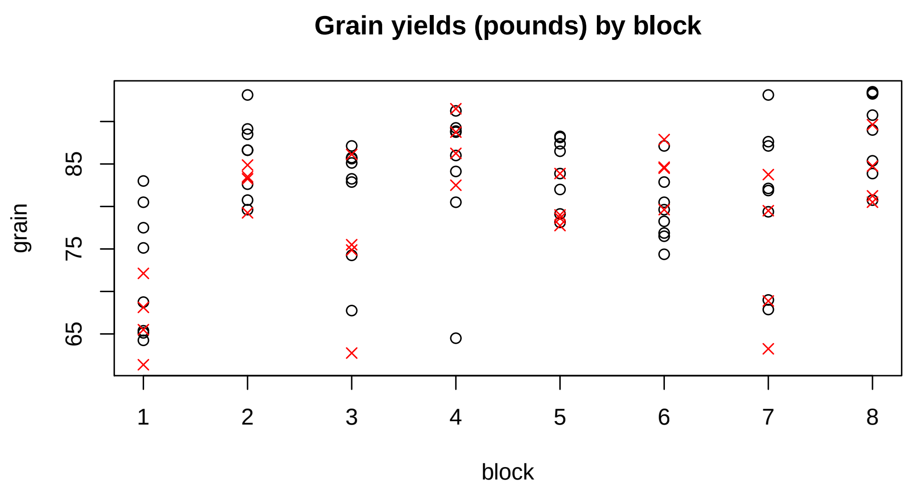


Fig. 1. A complex experiment with winter oats. (Reproduced from the *Journal of the Ministry of Agriculture* by permission of the Controller of H.M. Stationery Office.)

In a important early paper, Eden and Fisher (1927) described a way to compare the effects of various fertilizer treatments on the yield of grain (Grey Winter oats). They had two different nitrogen fertilizers (M = muriate of ammonia, S = sulphate of ammonia), applied in three different amounts (0, 1, or 2 cwt/acre), at two different stages of crop growth (E= early, L= late). They assigned the "treatments" to 96 plots of size 1/40 acre, arranged in 8 blocks of 12 plots each. Within each block, they assigned "treatments" to plots in a random order: 4 plots with no treatment (that is, amount = 0), and each of the eight possible combinations of $\{M, S\}$, $\{1, 2\}$, and $\{E, L\}$ appearing once.

The random allocation was intended to offset differences in fertility between different plots within each block, which were known to exist. (You can see these differences by looking at just the yields for the no-treatment plots, the red X's in the next plot.)



##	I	II	III	IV	V	VI	VII	VIII	total
## none	61.38	79.25	75.50	91.50	78.62	84.62	68.88	81.25	621.00
## none	65.50	83.50	74.88	86.25	79.00	84.50	79.50	80.50	633.62
## none	68.12	83.25	62.75	88.75	83.88	87.88	63.25	89.62	627.50
## none	72.12	84.88	86.12	82.50	77.75	79.62	83.75	84.75	651.50
## 1ME	77.50	80.75	85.12	80.50	88.25	76.88	69.00	90.75	648.75
## 1ML	80.50	93.12	67.75	88.88	88.12	79.62	67.88	80.75	646.62
## 1SE	65.38	89.12	85.75	86.00	86.50	76.50	79.38	93.50	662.12
## 1SL	75.12	86.62	85.62	89.25	87.38	87.12	87.62	93.25	692.00
## 2ME	83.00	86.62	83.25	64.50	82.00	82.88	82.12	85.38	649.75
## 2ML	64.25	79.62	87.12	88.75	79.12	74.38	87.12	89.00	649.38
## 2SE	68.75	88.50	82.88	84.12	83.88	78.25	81.88	83.88	652.12
## 2SL	65.12	82.62	74.25	91.25	78.12	80.50	93.12	93.38	658.38
## total	846.75	1017.88	951.00	1022.25	992.62	972.75	943.50	1046.00	7792.75

Remark. E&F Table I gave the grain yields in eighths of a pound:

```
##          I    II   III   IV    V    VI   VII  VIII total
## none    491  634  604  732  629  677  551  650  4968
## none    524  668  599  690  632  676  636  644  5069
## none    545  666  502  710  671  703  506  717  5020
## none    577  679  689  660  622  637  670  678  5212
## 1ME     620  646  681  644  706  615  552  726  5190
## 1ML     644  745  542  711  705  637  543  646  5173
## 1SE     523  713  686  688  692  612  635  748  5297
## 1SL     601  693  685  714  699  697  701  746  5536
## 2ME     664  693  666  516  656  663  657  683  5198
## 2ML     514  637  697  710  633  595  697  712  5195
## 2SE     550  708  663  673  671  626  655  671  5217
## 2SL     521  661  594  730  625  644  745  747  5267
## total 6774 8143 7608 8178 7941 7782 7548 8368 62342
```

If you look at the paper, be aware that some tabulations are for pounds and some are for eighths of a pound.

The four untreated plots within each block give a way of estimating the variability within blocks:

```
## lm(formula = grain ~ -1 + block, data = EFdata, subset = notreat)
##             Estimate Std. Error t value Pr(>|t|)
## blockI        66.781      2.838  23.531        0
## blockII       82.719      2.838  29.147        0
## blockIII      74.812      2.838  26.361        0
## blockIV       87.250      2.838  30.743        0
## blockV        79.812      2.838  28.123        0
## blockVI       84.156      2.838  29.653        0
## blockVII      73.844      2.838  26.020        0
## blockVIII     84.031      2.838  29.609        0
## Estimate of sigma =  5.68 from  24 degrees of freedom
```

E&F also estimated σ using the residuals from an additive fit (not the way they put it):

```
out.bt <- lm(grain ~  block + treat,EFdata)
sighat <- sqrt(sum(out.bt$res^2)/out.bt$df)
round(sighat,3) # on 80 degrees of freedom

## [1] 6.405
```

Using an F-test, E&F decided that the two estimates of σ were not significantly different. They then declared that "the value derived from the whole 80 degrees of freedom may be used with confidence".

The analysis of variance table suggests that overall effect of the treatments is only at the noise level:

```
anova(out.bt)

## Analysis of Variance Table
##
## Response: grain
##           Df Sum Sq Mean Sq F value    Pr(>F)
## block      7 2286.4  326.63  7.9620 2.617e-07 ***
## treat      8  387.0   48.38  1.1792     0.322
## Residuals 80 3281.9   41.02
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

However, E&F also pointed out that the design of the experiment allows for more detailed comparisons. For example, there were 32 plots for each of the fertilizer treatments. The average yields provide a broad comparison of the three levels:

```
mean.fert <- tapply(EFdata$grain,EFdata$fert,mean)
round(mean.fert,3)

##      M   none      S
## 83.270 79.176 81.078
```

The difference 2.19 between the means for the M and S fertilizers is down at the level of the (estimated) standard error for a difference of two such averages: $\widehat{\sigma}\sqrt{2/32} = 1.6$.

E&F commented that the only significant differences appeared to be in the amount of fertilizer used:

```
round(tapply(EFdata$grain,EFdata$amount,mean),2)

##     d  none     s
## 81.55 79.18 82.80
```

> **Remark.** These numbers are different from those in the first row of Table VI (E&F$_{560}$). My numbers are about 1.04 times bigger. Maybe that is the conversion factor for pounds to bushels, although I have my doubts. Probably I have made a silly mistake somewhere. The effects differences don't look very significant to me.

It is possible to carry out formal t-tests without so much manual labor.

2 Decomposition of treatment effects

The E&F design involves a few complications that I'll avoid by first discussing a simpler data set from Box et al. (1978, Section 10.1). The data involve three factors: temperature (at 160 or 180 degrees Celsius), concentration (at 20% or 40%), and catalyst (A or B), with the yield (in grams) as the response in a pilot study. (BHH devoted quite a few pages to the example.)

```
##   Temp Conc Cat yield
## 1  160   20   A    60
## 2  180   20   A    72
## 3  160   40   A    54
## 4  180   40   A    68
## 5  160   20   B    52
## 6  180   20   B    83
## 7  160   40   B    45
## 8  180   40   B    80
```

As usual, the factors can be represented by dummy variables:

```
T2 <- bhh$Temp == "180" ; T1 <- bhh$Temp == "160"
C2 <- bhh$Conc == "40" ; C1 <- bhh$Conc == "20"
K2 <- bhh$Cat == "B" ; K1 <- bhh$Cat == "A"
```

or coded as variables taking the values ± 1:

```
B <- data.frame(int=1, t=T2-T1, c=C2-C1, k=K2-K1)
B$tc <- B$t * B$c ; B$tk <- B$t * B$k ; B$ck <- B$c * B$k
B$tck <- B$t * B$c * B$k
B <- as.matrix(B)
print(B)

##      int  t  c  k tc tk ck tck
## [1,]   1 -1 -1 -1  1  1  1  -1
## [2,]   1  1 -1 -1 -1 -1  1   1
## [3,]   1 -1  1 -1 -1  1 -1   1
## [4,]   1  1  1 -1  1 -1 -1  -1
## [5,]   1 -1 -1  1  1 -1 -1   1
## [6,]   1  1 -1  1 -1  1 -1  -1
```

```
## [7,]    1 -1   1   1 -1 -1   1   -1
## [8,]    1  1   1   1  1  1   1    1
```

You will see in a moment why I created the matrix `B`. Observe that its columns are orthogonal, each with squared length equal to 8:

```
# Matrix(t(B) %*% B) # remove the comment char to see the matrix
```

Why are they orthogonal?

Let me drop the $B\$$ prefix for a while. First note that

$$\langle int, t\rangle = \langle \mathbb{1}_8, T2 - T1\rangle = 4 - 4 = 0.$$

This equality relects the fact that **Temp** appears the same number of times at each of its levels. Similarly

$$\begin{aligned}\langle t, c\rangle &= \langle T2 - T1, C2 - C1\rangle \\ &= sum(T2 * C2 - T1 * C2 - T1 * C1 + T1 * C1) = 2 - 2 - 2 + 2 = 0.\end{aligned}$$

Again the orthogonality comes from balance in the design. The interactions re more interesting.

$$\begin{aligned}\langle t, ck\rangle &= sum\left((T2 - T1) * (C2 - C1) * (K2 - K1)\right) \\ &= sum\left(T2 * C2 * K2 - T2 * C2 * K1 + \cdots - T1 * C1 * K1\right) \\ &= 1 - 1 + \cdots - 1 = 0.\end{aligned}$$

More balance. Finally,

$$\begin{aligned}\langle tk, tck\rangle &= sum(t * k * t * c * k) \\ &= sum(t^2 * c * k^2) = sum(1 * c * 1) = 0.\end{aligned}$$

And so on.

If we divide each of the columns of B by $\sqrt{8}$ we are left with an orthonormal basis for $\mathbb{R}^8$. The coefficients obtained from

```
print( bhh$yield %*% B / 8 )

##         int    t    c    k   tc tk ck  tck
## [1,] 64.25 11.5 -2.5 0.75 0.75  5  0 0.25
```

give a representation of `yield` in this basis. Moreover, $t/\sqrt{8}$ is a unit vector in span$(T1, T2)$ that is orthogonal to $\mathbb{1}$. The least squares fit

```
lm(yield ~ Temp + Conc + Cat, bhh)
```

represents the component of the `yield` vector in the four-dimensional subspace span$(\mathbb{1}, T1, T2, C1, C2, K1, K2)$ in the new basis:

```
## lm(formula = yield ~ Temp + Conc + Cat, data = bhh)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    64.25      2.531  25.385    0.000
## Temp1          11.50      2.531   4.544    0.010
## Conc1          -2.50      2.531  -0.988    0.379
## Cat1            0.75      2.531   0.296    0.782
## Estimate of sigma =  7.16 from  4 degrees of freedom
```

Note the coefficients.

Table 5.4. Calculated Effects and Standard Errors for 2^3 Factorial: Pilot Plant Example

Average	Effect with Standard Error
Main effects	
Temperature, T	23.0 ± 1.4
Concentration, C	-5.0 ± 1.4
Catalyst, K	1.5 ± 1.4
Two-factor interactions	
$T \times C$	1.5 ± 1.4
$T \times K$	10.0 ± 1.4
$C \times K$	0.0 ± 1.4
Three-factor interaction	
$T \times C \times K$	0.5 ± 1.4

The rescaled inner product $\langle bhh\$yield, B\$t\rangle/4$ also has the interpretation

$$\text{mean of yields at } 180^\circ - \text{mean of yields at } 160^\circ,$$

which estimates the "main temperature effect", a difference between the average (over the the other factors) effect of temperature 180° and the average effect of temperature 160°. Similarly, $\langle bhh\$yield, B\$tk\rangle/4$ equals some multiple of the difference between the average temperature effects at the two levels of catalyst, a measure of the interaction between temperature and catalyst.

Remark. I can never keep track of how many averages are involved in these interactions. I much prefer to think of interactions as estimates of departures from additivity under some parametrization of the model. Compare with the estimates of main effects and interactions in the table copied from BHH.

```
## lm(formula = yield ~ Temp + Conc + Cat + Temp:Cat, data = bhh)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    64.25      0.456 140.765    0.000
## Temp1          11.50      0.456  25.195    0.000
```

```
## Conc1          -2.50      0.456  -5.477    0.012
## Cat1            0.75      0.456   1.643    0.199
## Temp1:Cat1      5.00      0.456  10.954    0.002
## Estimate of sigma =  1.29 from  3 degrees of freedom
```

BHH (Section 10.8):

The main effect of a factor should be individually interpreted only if there is no evidence that the factor interacts with otl1er factors. When there is evidence of one or more such interactions, the interacting variables must be considered jointly.

If we keep throwing in interactions we eventually run out of degrees of freedom. We get a perfect fit, which **R** flags as not such a good thing.

```
##
## Call:
## lm(formula = yield ~ Temp * Conc * Cat, data = bhh)
##
## Residuals:
## ALL 8 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       6.425e+01         NA      NA       NA
## Temp1             1.150e+01         NA      NA       NA
## Conc1            -2.500e+00         NA      NA       NA
## Cat1              7.500e-01         NA      NA       NA
## Temp1:Conc1       7.500e-01         NA      NA       NA
## Temp1:Cat1        5.000e+00         NA      NA       NA
## Conc1:Cat1       -1.963e-15         NA      NA       NA
## Temp1:Conc1:Cat1  2.500e-01         NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,Adjusted R-squared:    NaN
## F-statistic:   NaN on 7 and 0 DF,  p-value: NA
```

3 A more complicated factorial design: confounding

Here is one of the data sets that come with **R**:

```
data(npk)
npk

##    block N P K yield
## 1      1 0 1 1  49.5
## 2      1 1 1 0  62.8
## 3      1 0 0 0  46.8
## 4      1 1 0 1  57.0
## 5      2 1 0 0  59.8
## 6      2 1 1 1  58.5
## 7      2 0 0 1  55.5
## 8      2 0 1 0  56.0
## 9      3 0 1 0  62.8
## 10     3 1 1 1  55.8
## 11     3 1 0 0  69.5
## 12     3 0 0 1  55.0
## 13     4 1 0 0  62.0
## 14     4 1 1 1  48.8
## 15     4 0 0 1  45.5
## 16     4 0 1 0  44.2
## 17     5 1 1 0  52.0
## 18     5 0 0 0  51.5
## 19     5 1 0 1  49.8
## 20     5 0 1 1  48.8
## 21     6 1 0 1  57.2
## 22     6 1 1 0  59.0
## 23     6 0 1 1  53.2
## 24     6 0 0 0  56.0
```

We have 6 blocks, each of size 4, and 3 factors, each at 2 levels. Clearly we do not have enough room in each block to make comparisons between all 8 ways of combining the factors. I wonder what will happen if we try to fit all the interactions:

```
anova(lm(yield ~ block + N*P*K,npk))

## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value   Pr(>F)
```

```
## block      5 343.29  68.659  4.4467 0.015939 *
## N          1 189.28 189.282 12.2587 0.004372 **
## P          1   8.40   8.402  0.5441 0.474904
## K          1  95.20  95.202  6.1657 0.028795 *
## N:P        1  21.28  21.282  1.3783 0.263165
## N:K        1  33.13  33.135  2.1460 0.168648
## P:K        1   0.48   0.482  0.0312 0.862752
## Residuals 12 185.29  15.441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To be continued.

References

Box, G. E. P., W. G. Hunter, and J. S. Hunter (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building.* New York: Wiley.

Eden, T. and R. A. Fisher (1927, 10). Studies in crop variation IV: The experimental determination of the value of top dressings with cereals. *The Journal of Agricultural Science 17*(4), 548–562.