

STATISTICS 368/501  
INTRODUCTION TO DESIGN  
AND ANALYSIS OF EXPERIMENTS  
Doug Wiens\*  
March 16, 2015

\*© Douglas P. Wiens, Department of Mathematical & Statistical Sciences, Faculty of Science, University of Alberta (2015).

# Contents

- I INTRODUCTION 8**
- 1 Introduction . . . . . 9
  
- II SIMPLE COMPARATIVE EXPERIMENTS 15**
- 2 Basic concepts; sampling distributions . . . 16
- 3 Comparing two means . . . . . 24
- 4 Randomized block designs . . . . . 32

**III SINGLE FACTOR EXPERIMENTS 39**

5 Concepts of ANOVA . . . . . 40

6 Mathematics of ANOVA . . . . . 49

7 Model checking . . . . . 55

8 Inferences; multiple comparisons . . . . . 61

**IV RANDOMIZED BLOCKS, LATIN SQUARES, AND RELATED DESIGNS 67**

9 Randomized complete block designs . . . . . 68

10 RCBD II . . . . . 74

11 RCBD III; Latin squares . . . . . 81

12 Latin and Graeco-Latin Squares . . . . . 88

13 Balanced Incomplete Block Designs . . . . . 95

**V INTRODUCTION TO FACTORIAL DESIGNS 102**

14 Introduction to Factorial Designs . . . . . 103

15 ANOVA; model checking . . . . . 110

16 Test for non-additivity; 3 factor designs . . . 117

**VI THE  $2^k$  FACTORIAL DESIGN 124**

17  $2^2$  factorials . . . . . 125

18  $2^k$  factorials . . . . . 133

**VII BLOCKING AND CONFOUNDING** **145**

19 Blocking & confounding . . . . . 146

20  $2^k$  factorials in 2 blocks . . . . . 151

21  $2^k$  factorials in  $2^p$  blocks . . . . . 158

22 Partial confounding . . . . . 165

**VIII FRACTIONAL FACTORIAL DESIGNS** **172**

23 Fractional factorials - introduction . . . . . 173

24 Fractional factorials - examples . . . . . 180

## **IX EXPERIMENTS WITH RANDOM FACTORS 187**

25 One-factor random effects model . . . . .	188
26 Two-factor random effects model . . . . .	195
27 Two-factor mixed model . . . . .	201

## **X NESTED AND SPLIT-PLOT DESIGNS 210**

28 Two-stage nested designs . . . . .	211
29 Nested and crossed factors . . . . .	217
30 Split-plot designs . . . . .	223

**XI SPECIAL TOPICS 232**

31 Regression . . . . . 233

32 ANOVA via regression . . . . . 241

33 Analysis of covariance . . . . . 250

34 Repeated measures and crossover designs . 258

# **Part I**

## **INTRODUCTION**



## 1. Introduction

Example of a medical experiment (major employers of statisticians are hospitals, pharmaceutical firms, medical research facilities ... )

- Two headache remedies are to be compared, to see which is more effective at alleviating the symptoms.
- Control (Drug 1) and New (Drug 2) - Drug type is the factor whose effect on the outcome is to be studied.
- Outcome variable: rate of blood flow to the brain, one hour after taking the drug.
- Other factors may well affect the outcome - gender of the patient (M or F), dosage level (LO or HI), etc.

- Single factor experiment: here the drug type will be the only factor in our statistical model.
  - Patients suffering from headaches volunteer for the study, and are randomly divided into two groups. One group receives Drug 1, the other receives Drug 2. Within each group the dosage levels are randomly assigned.
  - We hope that the randomization will control for the other factors, in that (on average) there should be approximately equal numbers of men and women in each group, equal numbers at each dosage level, etc.

- A statistical model: Suppose that  $n$  patients receive Drug 1 ( $i = 1$ ) and  $n$  receive Drug 2 ( $i = 2$ ). Define  $y_{ij}$  to be the response of the  $j^{\text{th}}$  subject ( $j = 1, \dots, n$ ) in the  $i^{\text{th}}$  group. A possible model is

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

where the parameters are:

- $\mu$  = 'overall mean' (mean effect of the drugs, irrespective of which one is used), estimated by the overall average

$$\bar{y}_{..} = \frac{1}{2n} \sum_{i=1}^2 \sum_{j=1}^n y_{ij},$$

- $\tau_i$  = ' $i^{\text{th}}$  treatment effect', estimated by  $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$ , where

$$\bar{y}_{i.} = \frac{1}{n} \sum_{j=1}^n y_{ij},$$

and

- $\epsilon_{ij}$  = 'random error' - measurement error, and effects of factors which are not included in the model, but perhaps should be.
  
- We might instead control for the dosage level as well. We could obtain  $4n$  volunteers, split each group into four (randomly) of size  $n$  each and assign these to the four combinations Drug 1/LO dose, Drug 1/HI dose, Drug 2/LO dose, Drug 2/HI dose. This is a  $2^2$  factorial experiment: 2 factors - drug type, dosage - each at 2 levels, for a total of  $2 \times 2 = 2^2$  combinations of levels and factors.
  - Now the statistical model becomes more involved - it contains parameters representing the effects of each of the factors, and possibly interactions between them (e.g., Drug 1 may only be more effective at HI dosages, Drug 2 at LO dosages).

- We could instead split the group into men and women (the 2 blocks) of subjects and run the entire experiment, as described above, within each block. We expect the outcomes to be more alike within a block than not, and so the blocking can reduce the variation of the estimates of the factor effects.
- There are other experimental designs which control for the various factors but require fewer patients - Latin squares, Graeco-Roman squares, repeated measures, etc.
- Designing an experiment properly takes a good deal of thought and planning. In this course we will be doing MUCH MORE than presenting formulas and plugging numbers into them.

- Planning phase of an experiment
  - What is to be measured? (i.e. what is the response variable?)
  - What are the influential factors?
  
- Experimental design phase
  - Control the known sources of variation (as in the factorial experiment described above, or through blocking)
  - Allow estimation of the size of the uncontrolled variation (e.g. by assigning subjects to drug type/dosage level blocks randomly, so as to average out a 'gender' effect).
  
- Statistical analysis phase
  - Make inferences on factors included in the statistical model
  - Suggest more appropriate models

## **Part II**

# **SIMPLE COMPARATIVE EXPERIMENTS**

## 2. Basic concepts; sampling distributions

- Example of an industrial experiment: compare two types of cement mortar, to see which results in a stronger bond. We'll call these 'modified' and 'unmodified'; more details in text. Data collection: 10 samples of the unmodified formulation were prepared, and 10 of the modified were prepared. They are prepared and their 20 bond strengths measured, in random order, by randomly assigned technicians, etc. (A 'completely randomized' design.)
- Data in text and on course web site:

> mod

16.85 16.40 17.21 16.35 16.52

17.04 16.96 17.15 16.59 16.57

> unmod

17.50 17.63 18.25 18.00 17.86

17.75 18.22 17.90 17.96 18.15



- Summary statistics:

```
> summary(mod)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
16.35	16.53	16.72	16.76	17.02	17.21

```
> summary(unmod)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
17.50	17.78	17.93	17.92	18.11	18.25

## Graphical displays:

- `boxplot(mod, unmod)` yields the following boxplots (Q2=median, 'hinges' = approx. Q1, Q3, 'whiskers' go to the most extreme observations within 1.5 IQR of Q2).

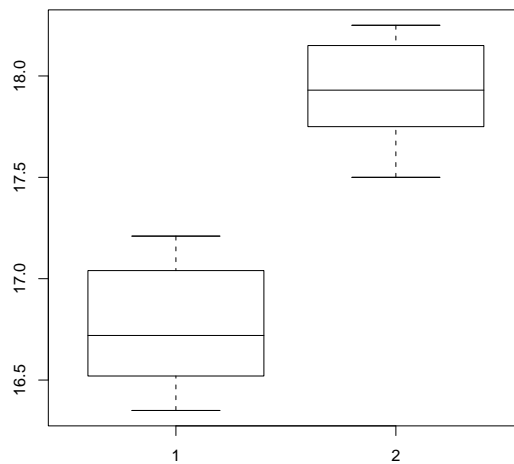


Fig 2.1

## Some basic notions from mathematical statistics

- $X$  a random variable, e.g. strength of a randomly chosen sample of unmodified mortar. (Think about how the randomness might arise.)
- $E[X]$  is the expected value, or mean, of  $X$ ; also written  $\mu_X$  (or just  $\mu$ ). It can be calculated from the probability distribution of  $X$ ; see §2.2 or review STAT 265 for details.
- If  $f(X, Y)$  is a function of r.v.s  $X$  and  $Y$  such as  $X^2$  or  $XY$ , then  $f(X, Y)$  is another r.v.; call it  $Z$  and then  $E[f(X, Y)]$  is defined to be  $E[Z]$ .
- Basic tool is linearity:

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

for constants  $a, b$  and  $c$ .

- Example:

$$\begin{aligned}
 E \left[ (X - \mu_X)^2 \right] &= E \left[ X^2 - 2\mu_X X + \mu_X^2 \right] \\
 &= E \left[ X^2 \right] - 2\mu_X E \left[ X \right] + \mu_X^2 \\
 &= E \left[ X^2 \right] - \mu_X^2.
 \end{aligned}$$

The quantity being studied here is the variance  $\sigma_X^2$ . Its square root  $\sigma_X$  is the standard deviation. The identity established above is sometimes rearranged as

$$E \left[ X^2 \right] = \mu_X^2 + \sigma_X^2.$$

- In Normal ( $N(\mu, \sigma^2)$ ) samples 95% of the values lie within  $1.96\sigma$  of  $\mu$ . Thus the variance of an estimator (which is typically at least approximately normally distributed) is inversely related to the accuracy - an estimator with a small variance will, with high probability, be close to its mean. If this mean is in fact the quantity one is trying to estimate, we say the estimator is unbiased.

- Example: Let  $Y_1, \dots, Y_n$  be a sample of mortar measurements (independent, and all randomly drawn from the same population). If the population mean is  $\mu_Y$  and the variance is  $\sigma_Y^2$ , then the mean and variance of the sample average

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

are (why?)

$$\mu_{\bar{Y}} = \frac{1}{n} \sum_{i=1}^n E[Y_i] = \mu_Y,$$

$$\sigma_{\bar{Y}}^2 = \sum_{i=1}^n \text{VAR} \left[ \frac{Y_i}{n} \right] \stackrel{\text{WHY?}}{=} \sum_{i=1}^n \frac{\text{VAR}[Y_i]}{n^2} = \frac{\sigma_Y^2}{n}.$$

Thus  $\bar{Y}$  is an unbiased estimator of  $\mu_Y$ ; it can be shown that in normal samples, no other unbiased estimator has a smaller variance. But note the lack of robustness.

- Similarly (see derivation in text),

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

is an unbiased estimator of  $\sigma_Y^2$ . The  $n-1$  is the ‘degrees of freedom’; this refers to the fact that only the  $n-1$  differences  $Y_1 - \bar{Y}, \dots, Y_{n-1} - \bar{Y}$  can vary, since  $\sum_{i=1}^n (Y_i - \bar{Y}) = 0$  implies that

$$Y_n - \bar{Y} = - \sum_{i=1}^{n-1} (Y_i - \bar{Y}).$$

- We will very often have to estimate the variance of other estimates, typically the effects of certain ‘treatments’, or other factors, on experimental units (e.g., effect of the modification on the mortar). In all cases the estimate will take the form

$$\hat{\sigma}^2 = \frac{SS}{df},$$

where  $SS$  is a sum of squares of departures from the estimate, and  $df$  is the degrees of freedom.

In many such cases we will be able (in theory, at least) to represent  $SS$  as

$$SS = \sum_{i=1}^{df} Z_i^2,$$

where  $Z_1, \dots, Z_{df}$  are independent  $N(0, \sigma^2)$  r.v.s. This defines a chi-square r.v.:

$$\begin{aligned} \frac{SS}{\sigma^2} &= \sum_{i=1}^{df} \left( \frac{Z_i}{\sigma} \right)^2 \\ &= \text{a sum of } df \text{ squares of independent } N(0, 1)\text{'s} \\ &\sim \chi_{df}^2. \end{aligned}$$

See some densities (Figure 2-6 of the text).

- Note  $E[\chi_{df}^2] = df$  (why?); also  $VAR[\chi_{df}^2] = 2df$ . Thus

$$\hat{\sigma}^2 \sim \sigma^2 \frac{\chi_{df}^2}{df}$$

with  $E[\hat{\sigma}^2] = \sigma^2$ . What is the variance of this estimator? How does it vary with  $df$ ?

### 3. Comparing two means

- Testing whether or not certain treatments have the same effects often involves estimating the variance among the group averages, and comparing this to an estimate of the underlying, ‘residual’ variation. The former is attributed to differences between the treatments, the latter to natural variation that one cannot control. Large values of the ratio of these variances indicates true treatment differences. Mathematically, one computes

$$F_0 = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2},$$

where  $\hat{\sigma}_j^2 \sim \sigma^2 \frac{\chi_{df_j}^2}{df_j}$  for  $j = 1, 2$ , and  $\hat{\sigma}_1^2, \hat{\sigma}_2^2$  are independent r.v.s. Then  $F_0$  is the numerical value of

$$F \sim \frac{\chi_{df_1}^2 / df_1}{\chi_{df_2}^2 / df_2},$$

where the two  $\chi^2$ 's are independent; this is the definition of an  $F$  r.v. on  $(df_1, df_2)$  degrees of freedom:  $F \sim F_{df_1}^{df_2}$ .



- Inferences involving just one mean (or the difference between two means) can be based on the  $t$ -distribution. One can generally reduce the problem to the following. We observe  $Y_1, \dots, Y_n \stackrel{ind}{\sim} N(\mu, \sigma^2)$ . Then

$$\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ independently of } S^2 \sim \sigma^2 \frac{\chi_{n-1}^2}{n-1}.$$

Here we use:

1. R.v.s  $Y_1, \dots, Y_n$  are jointly normally distributed *if and only if every linear combination of them is also normal*; this is the single most important property of the Normal distribution.
2. In Normal samples the estimate of the mean and the estimate of the variation around that mean are independent.

- Thus  $t_0 = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$  is the numerical value of the ratio of independent r.v.s

$$t = \frac{\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}}{S/\sigma} \sim \frac{Z}{\sqrt{\chi_{n-1}^2/(n-1)}}, \text{ where } Z \sim N(0, 1);$$

this is the definition of “Student’s  $t$ ” on  $n - 1$  d.f.

- You should now be able to show that  $t^2$  follows an  $F_{n-1}^1$  distribution.
- Return now to the mortar comparison problem. Let  $\{Y_{ij}\}_{j=1}^{n_i}$  be the samples ( $i = 1$  for modified,  $i = 2$  for unmodified;  $n_1 = n_2 = 10$ ). Assume:

$$Y_{11}, \dots, Y_{1,n_1} \stackrel{ind}{\sim} N(\mu_{mod}, \sigma_1^2),$$

$$Y_{21}, \dots, Y_{2,n_2} \stackrel{ind}{\sim} N(\mu_{unmod}, \sigma_2^2).$$

First assume as well that  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ , say. Then

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_{mod} - \mu_{unmod}, \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right).$$

The 'pooled' estimate of  $\sigma^2$  is

$$\begin{aligned}
 S_p^2 &= \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2} \\
 &\sim \sigma^2 \frac{\chi_{n_1-1}^2 + \chi_{n_2-1}^2}{n_1 + n_2 - 2} \\
 &\sim \sigma^2 \frac{\chi_{n_1+n_2-2}^2}{n_1 + n_2 - 2}
 \end{aligned}$$

(the sum of independent  $\chi^2$ 's is again  $\chi^2$ ; the d.f. are additive). Thus

$$\begin{aligned}
 t &= \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_{mod} - \mu_{unmod})}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\
 &= \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_{mod} - \mu_{unmod})}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \bigg/ \frac{S_p}{\sigma} \\
 &\sim t_{n_1+n_2-2}.
 \end{aligned}$$

To test the hypotheses

$$H_0 : \mu_{mod} = \mu_{unmod}$$

$$H_1 : \mu_{mod} \neq \mu_{unmod}$$

we assume that the null hypothesis is true and

compute

$$t_0 = \frac{\bar{Y}_1 - \bar{Y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

If indeed  $H_0$  is true we know that  $|t_0| \sim |t_{n_1+n_2-2}|$ . If  $H_0$  is false then  $|t_0|$  behaves, on average, like a multiple of  $|\mu_{mod} - \mu_{unmod}| > 0$ . Thus values of  $t_0$  which are large and positive, or large and negative, support the alternate hypothesis.

- The *p-value* is the prob. of observing a  $|t_{n_1+n_2-2}|$  which is as large or larger than  $|t_0|$ :

$$p = P \left( t_{n_1+n_2-2} \geq |t_0| \text{ or } t_{n_1+n_2-2} \leq -|t_0| \right).$$

If  $p$  is sufficiently small, typically  $p < .05$  or  $p < .01$ , we “reject  $H_0$  in favour of  $H_1$ ”.

```
> t.test(mod,unmod,var.equal=T)
Two Sample t-test
data: mod and unmod
t = -9.1094, df = 18, p-value = 3.678e-08
alternative hypothesis: true difference
      in means is not equal to 0
95 percent confidence interval:
 -1.4250734 -0.8909266
sample estimates:
mean of x      mean of y
 16.764        17.922
```

- If  $\sigma_1^2 = \sigma_2^2$  is not a reasonable assumption then

$$\bar{Y}_1 - \bar{Y}_2 \sim N \left( \mu_{mod} - \mu_{unmod}, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right),$$

and  $\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$  is an unbiased estimate of the variance, independent of  $\bar{Y}_1 - \bar{Y}_2$ . However, this variance estimate is no longer  $\sim$  as a multiple of a  $\chi^2$ , so that

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

is not exactly a  $t$  r.v. It is however well approximated by the  $t_\nu$  distribution, (“Welch’s approximation”) with random d.f.

$$\nu = \frac{\left( \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}.$$

```
> t.test(mod,unmod) #Welch's two-sample t-test
```

```
Welch Two Sample t-test
data: mod and unmod
t = -9.1094, df = 17.025, p-value = 5.894e-08
alternative hypothesis: true difference
    in means is not equal to 0
95 percent confidence interval:
 -1.4261741 -0.8898259
sample estimates:
mean of x    mean of y
 16.764      17.922
```

#### 4. Randomized block designs

- **Confidence intervals.** The following treatment seems general enough to cover the cases of practical interest. Any  $t$ -ratio can be written in the form

$$t = \frac{\hat{q} - q}{se(\hat{q})},$$

where  $q$  is a quantity to be estimated (e.g.  $\mu_1 - \mu_2$ ),  $\hat{q}$  is an estimate of  $q$  (e.g.  $\bar{y}_1 - \bar{y}_2$ ) and  $se(\hat{q})$  is an estimate of the standard deviation of  $\hat{q}$  (e.g.  $S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ ). Then if  $\pm t_{\alpha/2}$  are the points on the horizontal axis which each have  $\alpha/2$  of the probability of the  $t$ -distribution lying beyond them, we have

$$\begin{aligned} 1 - \alpha &= P\left(-t_{\alpha/2} \leq t \leq t_{\alpha/2}\right) \\ &= P\left(-t_{\alpha/2} \leq \frac{\hat{q} - q}{se(\hat{q})} \leq t_{\alpha/2}\right) \\ &= P\left(\hat{q} - t_{\alpha/2} \cdot se(\hat{q}) \leq q \leq \hat{q} + t_{\alpha/2} \cdot se(\hat{q})\right), \end{aligned}$$



after a rearrangement. Thus, before we take the sample, we know that the random interval

$$CI = \left[ \hat{q} - t_{\alpha/2} \cdot se(\hat{q}), \hat{q} + t_{\alpha/2} \cdot se(\hat{q}) \right]$$

will, with probability  $1 - \alpha$ , contain the true value of  $q$ . After the sample is taken, and  $\hat{q}, se(\hat{q})$  calculated numerically, we call  $CI$  a  $100(1 - \alpha)\%$  confidence interval.

- In the mortar example, the difference in averages was  $\hat{q} = 16.764 - 17.922 = -1.158$ ; then from the computer output,

$$-9.1094 = t = \frac{-1.158}{se(\hat{q})} \Rightarrow se(\hat{q}) = \frac{-1.158}{9.1094} = .1271.$$

There were 18 d.f., and so for a 95% interval we find  $t_{18, \alpha/2}$  on R:

$$\begin{aligned} &> qt(.975, 18) \\ &[1] 2.100922. \end{aligned}$$

Thus  $CI = -1.158 \pm 2.1009 \cdot .1271 = -1.158 \pm .2670 = [-1.4250, -.8910]$ , in agreement with the computer output.

**Sample size calculations.** The power of a test is the probability of rejecting the null hypothesis when it is false. Suppose that we would like a power of at least .99 when the mortar means differ by  $\delta = \mu_{mod} - \mu_{unmod} = .5$ . Furthermore, suppose that we will reject  $H_0$  if the p-value is  $< .05$ . (We 'use  $\alpha = .05$ '.) We need an estimate of  $\sigma$ , here I'll use  $\sigma = .4$ , which is somewhat larger than  $S_p$  was.

```
> power.t.test(delta = .5, sd = .4,
sig.level = .05,power = .99,type = "two.sample",
alternative = "two.sided")
```

Two-sample t test power calculation

n = 24.52528

delta = 0.5

sd = 0.4

sig.level = 0.05

power = 0.99

alternative = two.sided

NOTE: n is number in \*each\* group

Thus in a future study, if  $\sigma = .4$  is accurate, we will need two equal samples of at least 25 each in order to get the required power. Use `> help(power.t.test)` to get more details on this function.

- **Paired comparisons.** Suppose that the mortar data had arisen in the following way. Ten samples of raw material were taken, and each was split into two. One (randomly chosen) of the two became modified mortar, the other unmodified. This is then a ‘randomized block’ design - the raw materials are blocks, and observations within a block are expected to be more homogeneous (alike) than those in different blocks. Since the blocks (raw materials) might well affect the strength of the mortar, we should compare treatments using the same raw material, i.e. we put

$$d_j = Y_{1j} - Y_{2j} \sim N(\mu_d = \mu_{mod} - \mu_{unmod}, \sigma_d^2)$$

and make inferences about  $\mu_d$  by carrying out a one-sample or paired  $t$ -test.

```
> t.test(mod,unmod,paired=TRUE)
Paired t-test
data: mod and unmod
t = -10.2311, df = 9, p-value = 2.958e-06
alternative hypothesis: true difference
in means is not equal to 0
95 percent confidence interval:
-1.4140405 -0.9019595
sample estimates:
mean of the differences
-1.158
```

- **Testing the normality assumption.** If  $X \sim N(\mu, \sigma^2)$ , then  $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$  and (with  $\Phi(x) = P(Z \leq x)$ ):

$$\begin{aligned} P\left(X \leq \sigma\Phi^{-1}(i/n) + \mu\right) &= P\left(Z \leq \Phi^{-1}(i/n)\right) \\ &= \Phi(\Phi^{-1}(i/n)) \\ &= i/n. \end{aligned}$$

On the other hand, if  $x_{(1)} < x_{(2)} < \dots < x_{(n)}$  are the order statistics of a sample from a population of  $X$ 's, then  $i/n$  is the sample-based estimate of  $P(X \leq x_{(i)})$ . Thus, if the population is indeed Normal, we expect

$$x_{(i)} \approx \sigma\Phi^{-1}(i/n) + \mu,$$

so that a plot of  $x_{(i)}$  against the Normal quantiles  $\Phi^{-1}(i/n)$  should be approximately a straight line, with slope  $\sigma$  and intercept  $\mu$ . In practice  $i/n$  is replaced by something like  $(i - .5)/n$  to avoid  $\Phi^{-1}(i/n) = \infty$  when  $i = n$ .

```
> diff = mod-unmod
> qqnorm(diff)
> qqline(diff)
> cor(sort(diff),qnorm((.5:9.5)/10))
[1] 0.9683504
```

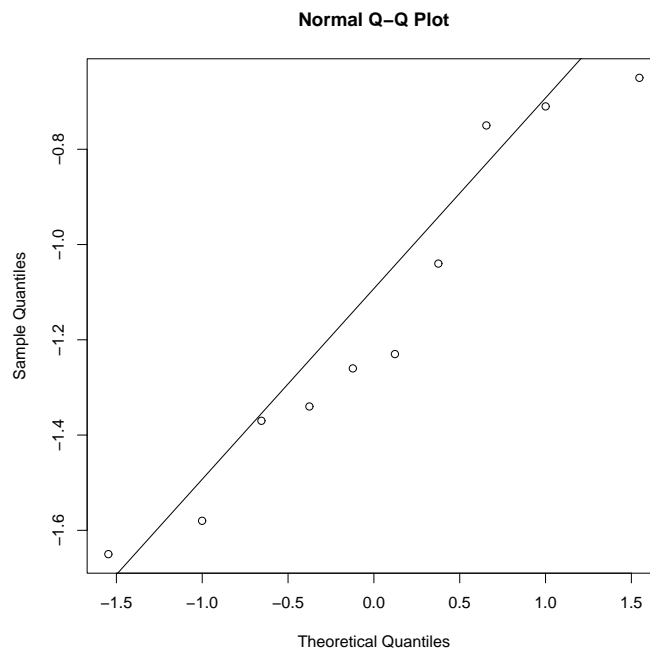


Fig 2.2

Some packages (e.g. MINITAB) will carry out a formal test, rejecting the null hypothesis of Normality if the correlation between  $(x_{(i)}, \Phi^{-1}((i - .5)/n))$  is too low.

## **Part III**

# **SINGLE FACTOR EXPERIMENTS**

## 5. Concepts of ANOVA

- The  $t$ -test studied in the previous chapter does not apply directly when there are more than two means to be compared, but the basic ideas extend in a natural way.
- Example: We are to investigate the formulation of a new synthetic fibre that will be used to make cloth for shirts. The cotton content varies from 10% - 40% by weight (the one factor is cotton content) and the experimenter chooses 5 levels of this factor: 15%, 20%, 25%, 30%, 35%. The response variable is  $Y =$  tensile strength. There are 5 replicates (complete repetitions of the experiment). In a replicate five shirts, each with a different cotton content, are randomly chosen from the five populations of shirts. The 25 tensile strengths are measured, in random order. This is then a **Completely Randomized Design** (CRD). (Why random order?)



- Data on web site and below. The levels 15%, ..., 35% are labelled 1, ..., 5, or  $A, \dots, E$ ; their numerical values are not important for the analysis. Write  $y_{ij}$  for the  $j^{\text{th}}$  observation at level  $i$  ( $i = 1, \dots, a = \#$  of levels). Thus, e.g.,  $y_{23} = 12$ . Totals and averages at level  $i$  are  $y_{i.}$  and  $\bar{y}_{i.} = y_{i.}/n$  ( $n = 5 = \#$  of replicates).

Tensile strength data

	Replicate					Totals	Averages
Level	1	2	3	4	5	$y_{i.}$	$\bar{y}_{i.}$
$A$	7	7	15	11	9	49	9.8
$B$	12	17	12	18	18	77	15.4
$C$	14	18	18	19	19	88	17.6
$D$	19	25	22	19	23	108	21.6
$E$	7	10	11	15	11	54	10.8

$$y_{..} = 376 \quad \bar{y}_{..} = 15.04$$

Measurement order

18	22	2	19	7
17	9	11	12	16
13	1	24	14	3
6	8	15	20	25
4	10	21	23	5

- Questions: Does changing the cotton content (level) change the mean strength? If so, is there a level which results in the maximum mean strength? From the following 'stripchart' we suspect that the answers are 'yes' and 'D'.

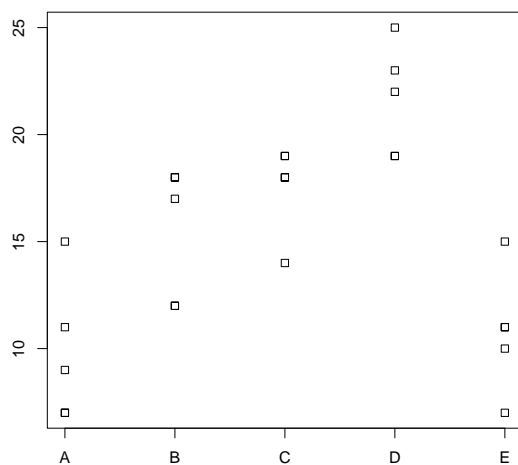


Fig. 3.1

- To answer the first question, we rephrase as: is the variation between the means at the 5 levels large enough, relative to the underlying random variation, that we can conclude that it is not arising purely by chance?
- We carry out an 'Analysis of Variance' (ANOVA); in this example the numerical output (commands on web site) is

```
> g <- lm(strength~content)
```

```
> anova(g)
```

```
Analysis of Variance Table
```

```
Response: strength
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
content	4	475.76	118.94	14.757	9.128e-06
Residuals	20	161.20	8.06		

```
> A <- c(7, 7, 15, 11, 9)
      etc.
> data <- rbind(A, B, C, D, E)
> data
  [,1] [,2] [,3] [,4] [,5]
A     7     7    15    11     9
B    12    17    12    18    18
C    14    18    18    19    19
D    19    25    22    19    23
E     7    10    11    15    11

#The sample mean for each treatment
> apply(data, 1, mean)
  A     B     C     D     E
9.8 15.4 17.6 21.6 10.8

# The overall mean
> mean(data)
[1] 15.04

# Total SS:
> (25-1)*var(strength)
[1] 636.96
```

```
# Draw a scatterplot of 'data' as a data frame
# with treatments as columns
> stripchart(data.frame(t(data)), vertical = TRUE)

#Arrange responses by column
> strength <- c(data)
> strength
 [1]  7 12 14 19  7  7 17 18 25 10 15 ...
#Set the factor at 5 levels
> d <- rep(c("A", "B", "C", "D", "E"), times=5)
> content <- as.factor(d)
> content
 [1] A B C D E A B C D E A B C D E ...
Levels: A B C D E
# Perform one-way ANOVA
# lm stands for 'linear model'
> g <- lm(strength~content)
> anova(g) # Produces the ANOVA table
```

- Interpretation: The Total Sum of Squares ( $SS_T$ ) measures the total variability around the overall average:

$$SS_T = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = 636.96.$$

How much of this is attributable to differences between the level (treatment) means?

$$\begin{aligned} SS_{Treatments} &= \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})^2 \\ &= n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 = 475.76. \end{aligned}$$

How much is attributable to random variation of the  $y_{ij}$  around these treatment means?

$$SS_E = \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 = 161.20.$$

Degrees of freedom of the SS's: Note that  $SS_T$  is the sum of  $N = an$  squares, but the sum of the unsquared terms is  $\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..}) = 0$ . Thus one of them cannot vary and there are  $N - 1 = 24$  d.f. Similarly  $SS_{Treatments}$  has

$a - 1 = 4$  d.f. The error sum of squares has  $\sum_{i=1}^a (n - 1) = N - a$  d.f., since  $\sum_{j=1}^n (y_{ij} - \bar{y}_{i.}) = 0$  for  $i = 1, \dots, a$ .

- The theoretical ANOVA table is

Source	SS	df	MS	$F_0$
Treatments	$SS_{Tr}$	$a - 1$	$MS_{Tr} = \frac{SS_{Tr}}{a-1}$	$F_0 = \frac{MS_{Tr}}{MS_E}$
Error	$SS_E$	$N - a$	$MS_E = \frac{SS_E}{N-a}$	
Total	$SS_T$	$N - 1$		

It can be shown that if the mean strengths are the same at all levels (i.e. 'under the hypothesis of no treatment effects'), so that  $\bar{y}_{1.}, \dots, \bar{y}_{a.}$  and  $\bar{y}_{..}$  are all estimating the same thing (the 'overall mean'), then (assuming as well that the observations are independently and normally distributed, with common variance  $\sigma^2$ )

$$\frac{SS_{Tr}}{\sigma^2} \sim \chi_{a-1}^2, \text{ ind. of } \frac{SS_E}{\sigma^2} \sim \chi_{N-a}^2,$$

so that

$$\frac{\frac{SS_{Tr}}{\sigma^2} / (a - 1)}{\frac{SS_E}{\sigma^2} / (N - a)} = \frac{MS_{Tr}}{MS_E} = F_0 \sim F_{N-a}^{a-1}.$$

If the mean strengths are not equal we expect larger values of  $MS_{Tr}$  than otherwise, hence larger values of  $F_0$ ; thus large values of  $F_0$  indicate that the hypothesis of no treatment effects should be rejected. The corresponding  $p$ -value is

$$P\left(F_{N-a}^{a-1} > F_0\right),$$

which in our example is  $P\left(F_{20}^4 > 14.757\right) = 9.128e-06$ ; certainly very small! It appears that (at least some of) the treatment means are significantly different.



## 6. Mathematics of ANOVA

It is no accident that  $SS_T = SS_{Tr} + SS_E$ . The following technique is basic; learn it now.

$$\begin{aligned}
 SS_T &= \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 \\
 &= \sum_{i=1}^a \sum_{j=1}^n \left( (y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y}_{..}) \right)^2 \\
 &= \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 + \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})^2 \\
 &\quad + 2 \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..}) (y_{ij} - \bar{y}_{i.}) \\
 &= SS_E + SS_{Tr} + 0,
 \end{aligned}$$

since the last term is

$$\sum_{i=1}^a \left\{ (\bar{y}_{i.} - \bar{y}_{..}) \sum_{j=1}^n (y_{ij} - \bar{y}_{i.}) \right\}$$

and each inner sum is 0 (why?).

- A more methodical and rigorous development of ANOVA starts with a statistical model of the data, and goes on to derive procedures according to some optimality principle. In this case there are two popular models:

Means model:  $y_{ij} = \mu_i + \epsilon_{ij},$

Effects model:  $y_{ij} = \mu + \tau_i + \epsilon_{ij}.$

In both cases we assume that the random errors  $\epsilon_{ij}$  are independently and normally distributed, with common variance  $\sigma^2$ . Thus the means model is equivalent to

$$y_{ij} \stackrel{ind.}{\sim} N(\mu_i, \sigma^2),$$

and we seek estimates of the parameters  $\mu_1, \dots, \mu_a, \sigma^2$ . The hypothesis that all treatments are equally effective becomes ' $\mu_1 = \dots = \mu_a$ '. In the effects model, we interpret the 'overall mean'  $\mu$  as the common effect of the treatments, i.e. as the mean response we would see if all the treatments were equally effective. The 'treatment effects'  $\tau_i$

must then  $= 0$  on average, since a non-zero average would be included in  $\mu$ . Thus in the effects model,

$$y_{ij} \stackrel{ind.}{\sim} N\left(\mu + \tau_i, \sigma^2\right), \text{ and } \sum_{i=1}^a \tau_i = 0.$$

Here we concentrate on the effects model, and derive parameter estimates. A common and in certain senses optimal method of estimation is *Least Squares*, by which we minimize the total squared discrepancy between the observations and their means:

$$\begin{aligned} S(\mu, \boldsymbol{\tau}) &= \sum_{i=1}^a \sum_{j=1}^n \left(y_{ij} - E[y_{ij}]\right)^2 \\ &= \sum_{i=1}^a \sum_{j=1}^n \left(y_{ij} - \mu - \tau_i\right)^2. \end{aligned}$$

We aim to find estimates  $\hat{\mu}, \hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \dots, \hat{\tau}_a)$ , with  $\sum \hat{\tau}_i = 0$ , minimizing  $S(\mu, \boldsymbol{\tau})$ . By calculus we get the minimizers

$$\begin{aligned} \hat{\mu} &= \bar{y}_{..}, \\ \hat{\tau}_i &= \bar{y}_{i.} - \bar{y}_{..} \end{aligned}$$

(Show that  $\sum \hat{\tau}_i = 0$ .) Here is an algebraic proof that these are the minimizers; the technique used is important. Decompose  $S(\mu, \tau)$  as  $S(\mu, \tau) =$

$$\begin{aligned} & \sum_{i=1}^a \sum_{j=1}^n \left[ (y_{ij} - \bar{y}_{i.}) - (\mu - \bar{y}_{..}) - (\tau_i - (\bar{y}_{i.} - \bar{y}_{..})) \right]^2 \\ &= \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 + \sum_{i=1}^a \sum_{j=1}^n (\mu - \bar{y}_{..})^2 \\ & \quad + \sum_{i=1}^a \sum_{j=1}^n (\tau_i - (\bar{y}_{i.} - \bar{y}_{..}))^2 + 3 \text{ cross-products} \\ &= SS_E + N(\mu - \hat{\mu})^2 + n \sum_{i=1}^a (\tau_i - \hat{\tau}_i)^2 + 0, \end{aligned}$$

since, e.g.,

$$\sum_{i=1}^a \sum_{j=1}^n (\mu - \bar{y}_{..}) (\tau_i - \hat{\tau}_i) = n(\mu - \bar{y}_{..}) \sum_{i=1}^a (\tau_i - \hat{\tau}_i) = 0.$$

(Why? Show that the other two cross-products vanish.) Thus  $S(\mu, \tau)$  is the sum of the non-negative terms  $SS_E$ ,  $N(\mu - \hat{\mu})^2$  and  $n \sum_{i=1}^a (\tau_i - \hat{\tau}_i)^2$ . The parameters occur only in the final two, which are clearly minimized by  $\mu = \hat{\mu}$  and  $\tau_i = \hat{\tau}_i$ . The minimum attainable value of  $S(\mu, \tau)$  is  $S(\hat{\mu}, \hat{\tau}) = SS_E$ .

**A basic principle of hypothesis testing.** We see how much the minimum value of  $S$  increases if we assume that the hypothesis is true (this is the SS 'due to the hypothesis') and compare this to its absolute minimum value. Formally:

$$F_0 = \frac{(\min_{H_0} S - \min S) / (\text{change in d.f.})}{\min S / \text{d.f.}} \sim F_{\text{d.f.}}^{\text{change in d.f.}}$$

The  $F$ -distribution holds when the errors are  $\overset{ind.}{\sim} N(0, \sigma^2)$  and when  $H_0$  is true.

As an example, to test  $H_0: \tau_1 = \dots = \tau_a = 0$  vs.  $H_1$ : 'not all = 0' we first assume  $H_0$  is true, so that

$$S = S(\mu, \mathbf{0}) = SS_E + N(\mu - \hat{\mu})^2 + n \sum_{i=1}^a \hat{\tau}_i^2,$$

$$\min_{H_0} S = SS_E + n \sum_{i=1}^a \hat{\tau}_i^2 = SS_E + SS_{Tr} = SS_T,$$

on  $N - 1$  d.f.

Then since  $\min S = SS_E$  on  $N - a$  d.f., we have

$$F_0 = \frac{SS_{Tr}/(a-1)}{SS_E/(N-a)} = \frac{MS_{Tr}}{MS_E} \sim F_{N-a}^{a-1}.$$

- Sometimes  $\min S$  and  $\min_{H_0} S$  are written as  $SS_{Full}$  and  $SS_{Reduced}$ : ‘full’ model containing all terms, ‘reduced’ if  $H_0$  holds.
- It is shown in the text (p. 68) that  $E[MS_E] = \sigma^2$ . Also (asst. 1)  $E[MS_{Tr}] = \sigma^2 + \frac{n \sum_{i=1}^a \tau_i^2}{a-1}$ ; thus we expect  $F_0$  to be near 1 under  $H_0$ , and  $> 1$  otherwise.
- **Unbalanced case.** If the treatment groups have differing sizes  $n_i$  we say the design is unbalanced. In this case we impose the condition  $\sum_{i=1}^a n_i \tau_i = 0$  and find that the LSEs are the same as before, but that now  $SS_{Tr} = \sum_{i=1}^a n_i \hat{\tau}_i^2$ . Everything else remains the same.

## 7. Model checking

- **Model adequacy testing.** Are our assumptions on the random errors satisfied? To answer this we note that the behaviour of the random errors  $\varepsilon_{ij} = y_{ij} - E[y_{ij}]$  should be reflected in the residuals

$$e_{ij} = y_{ij} - \text{'est. of } E[y_{ij}] \text{' } = y_{ij} - \hat{y}_{ij}.$$

Here the 'fitted values'  $\hat{y}_{ij}$  are

$$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) = \bar{y}_{i.}.$$

– Note that, since

$$e_{ij} = y_{ij} - \bar{y}_{i.},$$

we have that  $SS_E = \sum_{i,j} e_{ij}^2$ ; for this reason it is often written  $SS_{Resid}$ .

We look at qq-plots, and plots of residuals versus  $\bar{y}_i$ , or versus time, ... looking for anything that might contradict the assumptions of normality, independence, equality of variances.

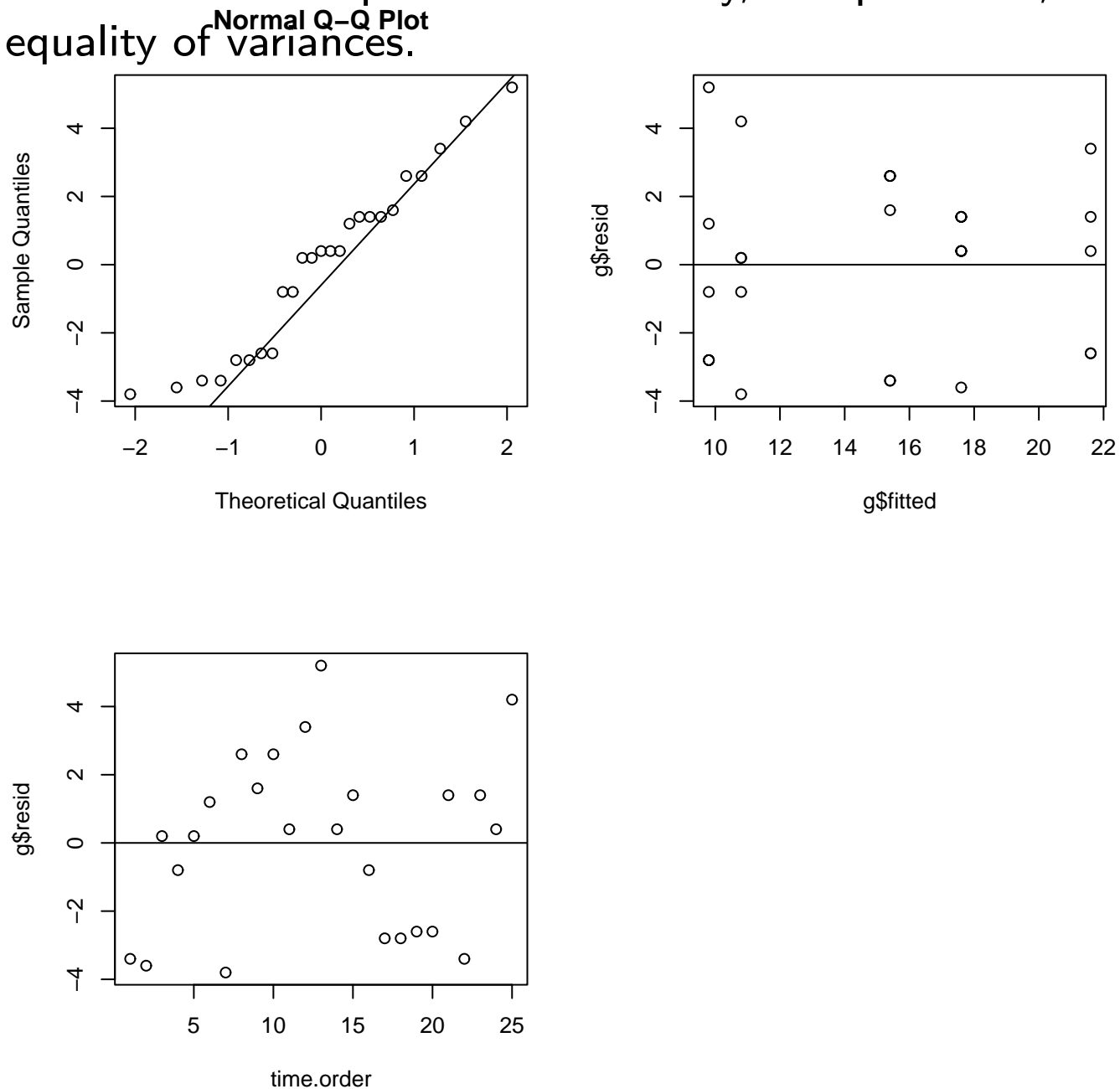


Fig. 3.2



**Formal tests of equality of variances.** We suppose that  $\varepsilon_{ij} \sim (0, \sigma_i^2)$  and test  $H_0: \sigma_1^2 = \dots = \sigma_a^2$ . A test which is optimal if the data are Normal, but can be quite misleading otherwise, is ‘Bartlett’s test’. It is based on the fact that if  $T_1, \dots, T_a$  are any positive numbers, and  $w_1, \dots, w_a$  are any weights summing to 1, then the ratio of the weighted arithmetic mean to the weighted geometric mean is always  $\geq 1$ :

$$\frac{\sum w_i T_i}{\prod T_i^{w_i}} \geq 1,$$

with equality iff  $T_1 = \dots = T_a$ . This is applied with  $T_i = S_i^2$ , the sample variance from the  $i^{\text{th}}$  group, and  $w_i = (n_i - 1) / (N - a)$ . The log of the ratio is computed; it is positive and large values support  $H_1$ .

```
> bartlett.test(g$resid, content)
```

```
  Bartlett test for homogeneity of variances
data:  g$resid and content
Bartlett's K-squared = 0.9331, df = 4,
      p-value = 0.9198
```

A test which is more robust against non-normality is ‘Levene’s test’. First calculate absolute deviations from the group medians  $\tilde{y}_i$ :

$$d_{ij} = |y_{ij} - \tilde{y}_i|.$$

Equality of the variances is indicated by equality of the group means of the absolute deviations, which is tested by the usual F-test. So we compute the  $d_{ij}$  (see the commands on the web site), fit a linear model (called g.d) as before, and do the ANOVA:

```
> anova(g.d)
```

```
Analysis of Variance Table
```

```
Response: abs.deviations
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
content.d	4	4.96	1.24	0.3179	0.8626
Residuals	20	78.00	3.90		

If our assumptions seem to be violated we might:

1. try a transformation of the  $y_{ij}$  (e.g.  $\sqrt{y_{ij}}$ , or  $\log y_{ij}$ ) hoping that the transformed values are 'more normal' - read §3-4.3 for details; or
2. apply a nonparametric procedure. In the latter case we work with the ranks  $R_{ij}$  of the  $y_{ij}$  rather than their numerical values (so this is another kind of transformation - the *rank transformation*). Nonparametric tests are more robust; if they give the same results as the standard tests, we take this as evidence that the assumptions are met.

```
> strength
```

```
 7 12 14 19  7  7 17 18
25 10 15 12 18 22 11 11
18 19 19 15  9 18 19 23 11
```

```
R <- rank(strength)
```

```
2.0  9.5 11.0 20.5  2.0  2.0 14.0 16.5
25.0  5.0 12.5  9.5 16.5 23.0  7.0  7.0
16.5 20.5 20.5 12.5  4.0 16.5 20.5 24.0  7.0
```

Then do an ANOVA on the ranks:

```
> g.rank <- lm(R~content)
> anova(g.rank)
Analysis of Variance Table
Response: R
          Df Sum Sq Mean Sq  F value  Pr(>F)
content    4 1020.70  255.17   19.310 1.212e-06
Residuals 20  264.30   13.21
```

Note that, apart from ties,  $SS_T$  is just  $N - 1$  times the variance of the ranks  $1, \dots, N$ . It is not random and can be calculated without knowing the data. The only randomness is in  $SS_{Tr}$  (since  $SS_E = SS_T - SS_{Tr}$ ) and there is a  $\chi^2$ -approximation to the distribution of  $SS_{Tr}$ . This is used in the Kruskal-Wallis test (see §3-10.1), which typically gives results very similar to the rank transformation followed by the  $F$ -test.

```
> kruskal.test(strength, content)
      Kruskal-Wallis rank sum test
data:  strength and content
Kruskal-Wallis chi-squared = 19.0637, df = 4,
p-value = 0.0007636
```

## 8. Inferences; multiple comparisons

- What have learned about these data so far? The  $F$ -test and the more robust Kruskal-Wallis test both proclaimed the treatment effects to be significant (an indication that non-normality is not a problem), both of the tests of equality of the variances were non-significant, and the qq- and other plots gave no cause for alarm. We can now go on to answer our second question - which of the treatment means ( $\mu_i = \mu + \tau_i$ ) are significantly different? Can we say with any confidence that a particular one is largest? This is a problem of 'multiple comparisons' - comparing several means with each other.
- A confidence interval on one mean:  $\mu_i$  is estimated by  $\bar{y}_i$ , whose variance  $\sigma^2/n_i$  is estimated by  $MS_E/n_i$ . This results in

$$CI = \bar{y}_i \pm t_{\alpha/2, N-a} \sqrt{MS_E/n_i}.$$

Similarly, a confidence interval on one difference  $\mu_i - \mu_j = \tau_i - \tau_j$  is

$$CI = \bar{y}_{i.} - \bar{y}_{j.} \pm t_{\alpha/2, N-a} \sqrt{MS_E \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}.$$

Typically we are interested in contrasts of the treatment means:  $\Gamma = \sum c_i \mu_i$ , where  $\sum c_i = 0$ .

- Examples.

1. Assess the difference between treatments 1 and 2:  $\Gamma = \mu_1 - \mu_2$ ,  $c_1 = 1$ ,  $c_2 = -1$ .

2. Is the average effect of treatments 1 and 2 better than that of treatments 3 and 4? We study  $\Gamma = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$ .

- $\hat{\Gamma} = \sum c_i \hat{\mu}_i = \sum c_i \bar{y}_{i.} \sim N \left( \Gamma, \sigma^2 \sum \frac{c_i^2}{n_i} \right)$ , so that a CI on  $\Gamma$  is

$$\begin{aligned} CI &= \hat{\Gamma} \pm t_{\alpha/2, N-a} \cdot se(\hat{\Gamma}) \\ &= \sum c_i \bar{y}_{i.} \pm t_{\alpha/2, N-a} \sqrt{MS_E \cdot \sum \frac{c_i^2}{n_i}}. \end{aligned}$$

Problems with these methods:

- We might end up constructing many CIs. Before sampling the probability that any one of them will be wrong is  $\alpha$ , so that the probability that at least one will be wrong (the “experimentwise error rate”) is much more than  $\alpha$ .
- We might not know which comparisons we would like to make until after looking at the data. We need a method which accommodates this kind of ‘data-snooping’. One such method is *Scheffé’s method* for comparing all contrasts. A single CI has the property that

$$1 - \alpha = P \left( \left| \frac{\hat{\Gamma} - \Gamma}{se(\hat{\Gamma})} \right| \leq t_{\alpha/2, N-a} \right).$$

In Scheffé’s method we seek a number  $k_\alpha$  so that

$$\begin{aligned} 1 - \alpha &= P \left( \left| \frac{\hat{\Gamma} - \Gamma}{se(\hat{\Gamma})} \right| \leq k_\alpha \text{ for } \underline{\text{all}} \text{ contrasts } \Gamma \right) \\ &= P \left( \max \left| \frac{\hat{\Gamma} - \Gamma}{se(\hat{\Gamma})} \right| \leq k_\alpha \right), \quad (*) \end{aligned}$$

where the max is taken over all contrasts. It turns out that

$$k_\alpha = \sqrt{(a-1) F_{N-a}^{\alpha-1}(\alpha)}.$$

We can then calculate as many CIs as we wish, each of the form

$$CI = \hat{\Gamma} \pm k_\alpha \cdot se(\hat{\Gamma}) \quad (**)$$

and make the statement “With confidence  $1 - \alpha$ , the CIs I have calculated, as well as all others that I could have calculated but didn’t, all contain the true values of the contrasts.”

- Example. For the tensile strength data, using `k <- sqrt(4*qt(.95,4,20))` on R gives  $k_{.05} = 3.386$ . The CIs on all the treatment differences would each have

$$k_{.05} \cdot se(\hat{\Gamma}) = k_{.05} \sqrt{\frac{2 \cdot MS_E}{n}} = 3.386 \sqrt{\frac{2 \cdot 8.06}{5}} = 6.08$$

so that the CIs on  $\mu_i - \mu_j$  (for all  $i, j = 1, \dots, 5$ ) are

$$\bar{y}_i. - \bar{y}_j. \pm 6.08.$$



This compares to  $\bar{y}_{i.} - \bar{y}_{j.} \pm 3.75$  if only one comparison is made, using  $t_{.975,20} = 2.086$ . So the Scheffé intervals are quite wide. On the other hand, using Scheffé's method we can go on to calculate CIs on many other contrasts, without affecting the error rate.

- If we know in advance that we are only interested in the treatment differences, then we can use *Tukey's procedure*, in which in (\*) we maximize only over contrasts of the form  $\mu_i - \mu_j$ ; we seek a number  $q$  such that

$$1 - \alpha = P \left( \max \left| \frac{(\bar{y}_{i.} - \bar{y}_{j.}) - (\mu_i - \mu_j)}{\sqrt{MS_E \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}} \right| \leq \frac{q_\alpha}{\sqrt{2}} \right).$$

Then (\*\*), with  $k_\alpha$  replaced by  $q_\alpha/\sqrt{2}$ , holds for all treatment differences with an experimentwise error rate of  $\alpha$ .

- $qtukey(.95, 5, 20)/\sqrt{2}$  gives  $q_{.05}/\sqrt{2} = 2.992$ , so that the Tukey CIs on  $\mu_i - \mu_j$  (for all  $i, j = 1, \dots, 5$ ) and only these are  $\bar{y}_{i.} - \bar{y}_{j.} \pm 5.37$ . Using this method on the tensile strength data allows us to declare  $\mu_i$  significantly different (at the 5% level) than  $\mu_j$  if  $|\bar{y}_{i.} - \bar{y}_{j.}| > 5.37$ .

Treatment	1	5	2	3	4
Average	9.8	10.8	15.4	17.6	21.6
$\bar{y}_{4.} - \bar{y}_{j.}$	11.8	10.8	6.2	4	

We can declare  $\mu_4$  to be significantly different (and larger) than  $\mu_1$ ,  $\mu_5$  and  $\mu_2$ , but not significantly different than  $\mu_3$ .

- There is another method (*Dunnett's procedure*) available if one of the treatments is a control, and we only want to compare all others with it (i.e. to see if any of them are better than a standard, control treatment). See §3-5.8.

## **Part IV**

# **RANDOMIZED BLOCKS, LATIN SQUARES, AND RELATED DESIGNS**

## 9. Randomized complete block designs

- **Example:** A hardness testing machine presses a pointed rod (the 'tip') into a metal specimen (a 'coupon'), with a known force. The depth of the depression is a measure of the hardness of the specimen. It is feared that, depending on the kind of tip used, the machine might give different readings. The experimenter wants 4 observations on each of the 4 types of tips. Note that the differences in readings might also depend on which type of metal specimen is used, i.e. on the coupons.
- A Completely Randomized design would use 16 coupons, making 1 depression in each. The coupons would be randomly assigned to the tips, hoping that this would average out any differences between the coupons. Here 'coupon type' is a 'nuisance factor' - it may affect the readings, but we aren't very interested in measuring its effect.

It is also controllable, by blocking: we can use 4 coupons (the 'blocks') and apply each of the 4 treatments (the tips) to each coupon. This is preferable to hoping that randomization alone will do the job; it also uses fewer coupons.

- There may be unknown and uncontrollable factors affecting the readings (the eyesight of the operator, ... think of others). Here is where randomization might help - within each block, the treatments are applied in random order. So each block can be viewed as one CR designed experiment. This is a *Randomized Complete Block Design* (RCBD). 'Complete' means that each block contains all of the treatments.
- Common blocking variables: Day of week, person, batch of raw material, ... . A basic idea is that *the responses should be less highly varied within a block than between blocks*.

### Hardness testing design and data.

$y_{ij}$  = machine reading for tip  $i$ , coupon  $j$ ;  
order in parentheses.

	Coupon				
Tip	1	2	3	4	$\bar{y}_i.$
1	9.3 (3)	9.4 (3)	9.6 (2)	10.0 (1)	9.575
2	9.4 (1)	9.3 (4)	9.8 (1)	9.9 (4)	9.600
3	9.2 (4)	9.4 (2)	9.5 (3)	9.7 (3)	9.450
4	9.7 (2)	9.6 (1)	10.0 (4)	10.2 (2)	9.875
$\bar{y}_{.j}$	9.400	9.425	9.725	9.950	$\bar{y}_{..} = 9.625$

- Note that the layout of the data is the same as for a CR design, where the columns would be labelled 'replicates'. But the design is different - if this were a CRD the 'times' would be (1), ..., (16) in random order. Here the randomization is restricted - it is done separately within each block. This will allow us to attribute some of the variation to the blocks ( $SS_{Blocks}$ ), and thus remove it from experimental error ( $SS_E$ ).

```
> par(mfrow=c(1,2))  
> boxplot(y~blocks, xlab="blocks")  
> boxplot(y~treatments, xlab="treatments")
```

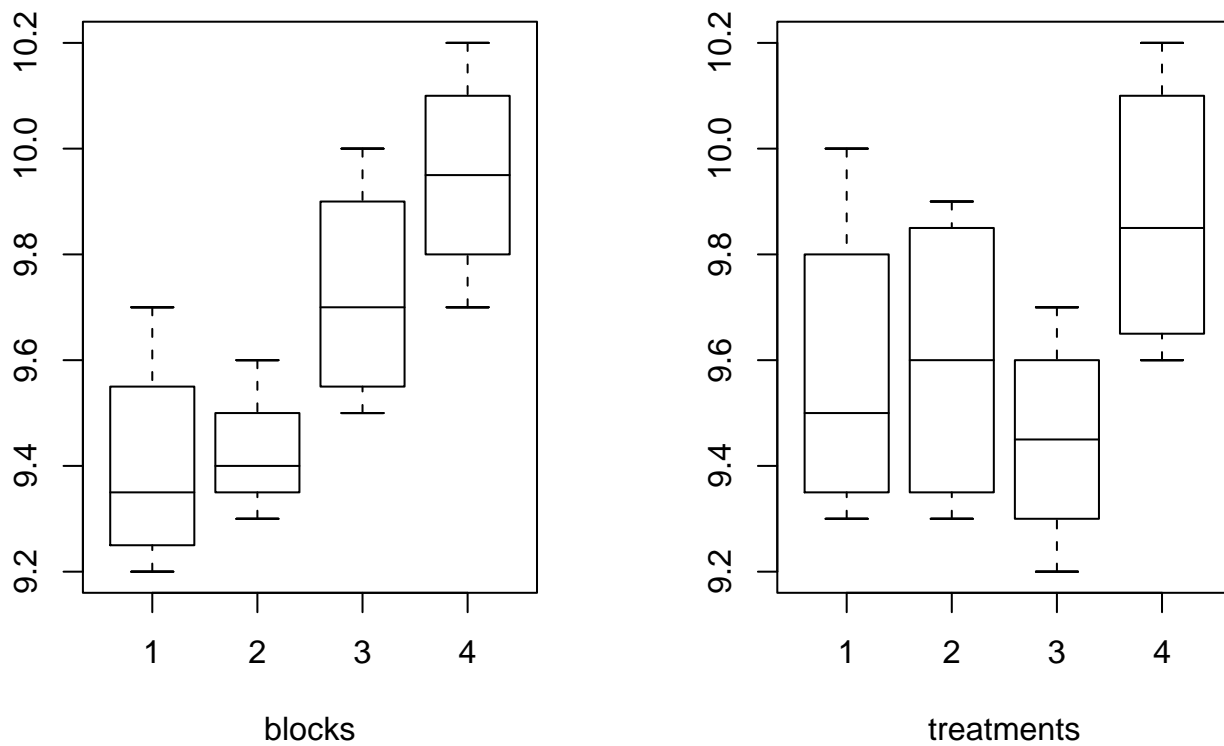


Fig. 4.1

- Effects model:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

$$i = 1, \dots, a = \# \text{ of treatments}$$

$$j = 1, \dots, b = \# \text{ of blocks}$$

$$\tau_i = \text{effect of } i^{\text{th}} \text{ treatment}$$

$$\beta_j = \text{effect of } j^{\text{th}} \text{ block}$$

$$\sum \tau_i = \sum \beta_j = 0.$$

- Assume  $\varepsilon_{ij} \stackrel{ind}{\sim} (0, \sigma^2)$ . Putting  $\mu_{ij} = \mu + \tau_i + \beta_j = E[y_{ij}]$  gives the means model:

$$y_{ij} = \mu_{ij} + \varepsilon_{ij}.$$

- We consider the effects model, and
  - Decompose  $SS_T$  into sums of squares attributable to (i) treatment differences, (ii) blocks, (iii) experimental error,
  - Obtain least squares estimates of  $\mu, \tau_i, \beta_j$ .



Decomposition of  $SS_T$ . Guided by a hunch that the LSEs will turn out to be

$$\hat{\mu} = \bar{y}_{..}, \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}, \hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$$

we write  $SS_T = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2$  as

$$\begin{aligned} & \sum_{i=1}^a \sum_{j=1}^b \left\{ \begin{array}{l} (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) \\ + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) \end{array} \right\}^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 \\ & \quad + \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 + 3 \text{ cross-products} \\ &= b \sum_{i=1}^a \hat{\tau}_i^2 + a \sum_{j=1}^b \hat{\beta}_j^2 + \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2, \\ & \quad \text{since all 3 cross-products vanish (how?)} \\ &= SS_{Tr} + SS_{Blocks} + SS_E. \end{aligned}$$

Degrees of freedom:

$$df(SS_{Tr}) = a - 1, \quad df(SS_{Blocks}) = b - 1,$$

$$df(SS_E) = ab - 1 - (a - 1) - (b - 1) = (a - 1)(b - 1).$$

## 10. RCBD II

The theoretical ANOVA table for a RCBD is

Source	SS	df	MS	F <sub>0</sub>
Treat.	$SS_{Tr}$	$a - 1$	$MS_{Tr} = \frac{SS_{Tr}}{a-1}$	$F_0 = \frac{MS_{Tr}}{MS_E}$
Blocks	$SS_{Blocks}$	$b - 1$	$MS_{Bl} = \frac{SS_{Bl}}{b-1}$	
Error	$SS_E$	$(a-1) \cdot (b-1)$	$MS_E = \frac{SS_E}{df}$	
Total	$SS_T$	$ab - 1$		

The expected mean squares can be derived as in assignment 1, and are

$$E[MS_{Tr}] = \sigma^2 + \frac{b \sum_{i=1}^a \tau_i^2}{a-1},$$

$$E[MS_{Blocks}] = \sigma^2 + \frac{a \sum_{j=1}^b \beta_j^2}{b-1},$$

$$E[MS_E] = \sigma^2.$$

Notice a pattern?

For the hardness data the R output is:

```
> g <- lm(y~treatments + blocks)
> anova(g)
Analysis of Variance Table
Response: y
          Df Sum Sq Mean Sq  F value Pr(>F)
treatments 3  0.38500  0.12833   14.438 0.0008713 ***
blocks      3  0.82500  0.27500   30.938 4.523e-05 ***
Residuals  9  0.08000  0.00889
```

Thus at any level  $\alpha > .00087$ , we would reject the null hypothesis of no treatment effects ( $H_0: \tau_1 = \dots = \tau_a = 0$ ). It also appears that the blocks have a significant effect. A caution here though - the randomization alone ensures that the F-test for treatments is approximately valid even if the errors are not very normal. Because of the randomization restriction, the same is not true for testing the significance of blocks by looking at  $MS_{Bl}/MS_E$ . Thus the  $p$ -value of  $4.523e - 05$  for blocks should be used only as a guide, unless one is sure of the normality.

Suppose that we erroneously analyzed these data as a CRD? What would the ANOVA be? The  $SS_{Blocks}$  must still be accounted for, and if it can't be anywhere else it ends up in experimental error (since  $SS_E = SS_T - SS_{Tr}$  for a CRD):

### Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatments	3	0.38500	0.12833	1.7017	0.2196
Residuals	9	0.08000	0.07542		
	+3	+.82500			
	= 12	= 0.90500			

( $MS_E = .90500/12$ ;  $F_0 = .12833/.07542$ .)

To verify our guess about the LSEs, we show that they minimize

$$\begin{aligned} S(\mu, \boldsymbol{\tau}, \boldsymbol{\beta}) &= \sum_{i=1}^a \sum_{j=1}^b \left\{ y_{ij} - E[y_{ij}] \right\}^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b \left\{ y_{ij} - \mu - \tau_i - \beta_j \right\}^2 \end{aligned}$$

subject to  $\sum \hat{\tau}_i = \sum \hat{\beta}_j = 0$ . Write this as  $S(\mu, \boldsymbol{\tau}, \boldsymbol{\beta}) =$

$$\begin{aligned} &\sum_{i=1}^a \sum_{j=1}^b \left\{ \begin{array}{l} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) \\ - (\mu - \hat{\mu}) - (\tau_i - \hat{\tau}_i) - (\beta_j - \hat{\beta}_j) \end{array} \right\}^2 = \\ &SS_E + ab(\mu - \hat{\mu})^2 + b \sum_{i=1}^a (\tau_i - \hat{\tau}_i)^2 + a \sum_{j=1}^b (\beta_j - \hat{\beta}_j)^2. \end{aligned}$$

The constraints are satisfied, and are used to show that the cross-products vanish. Now it is clear that the minimizers are as claimed, and that the minimum value of  $S(\mu, \boldsymbol{\tau}, \boldsymbol{\beta})$  is  $S(\hat{\mu}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\beta}}) = SS_E$ .

The 'change in SS' principle used to test for treatments effects can be written as

$$F_0 = \frac{(SS_{Reduced} - SS_{Full}) / \nabla df}{SS_{Full} / df (SS_{Full})}$$

where:

1. the 'full' model is  $E[y_{ij}] = \mu + \tau_i + \beta_j$ , with  $SS_{Full} = \min S(\mu, \boldsymbol{\tau}, \boldsymbol{\beta}) = SS_E$ ,
2. the 'reduced' model assumes that the null hypothesis is true, so that  $E[y_{ij}] = \mu + \beta_j$ , with

$$\begin{aligned} SS_{Reduced} &= \min S(\mu, \mathbf{0}, \boldsymbol{\beta}) = SS_E + b \sum_{i=1}^a \hat{\tau}_i^2 \\ &= SS_E + SS_{Tr}; \end{aligned}$$

thus  $F_0 = MS_{Tr} / MS_E \sim F_{(a-1)(b-1)}^{a-1}$  when the null hypothesis is true (and the assumptions hold).

NOTE: If there are only 2 treatments ( $a = 2$ ), then  $F_0$  reduces to  $F_0 = t_0^2 = \left\{ \bar{d} / (s_d / \sqrt{b}) \right\}^2$  for  $d_j = y_{1j} - y_{2j}$ ; i.e. to the square of the paired sample t statistic.

Check these assumptions. (i) qqplot of residuals  $e_{ij} = y_{ij} - \hat{y}_{ij}$ , where the fitted values are  $\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j$ . If `g <- lm(y~treatments + blocks)`, then these are `g$residuals` and `g$fitted.values`. (ii) residuals vs. treatment labels, block labels, fitted values.

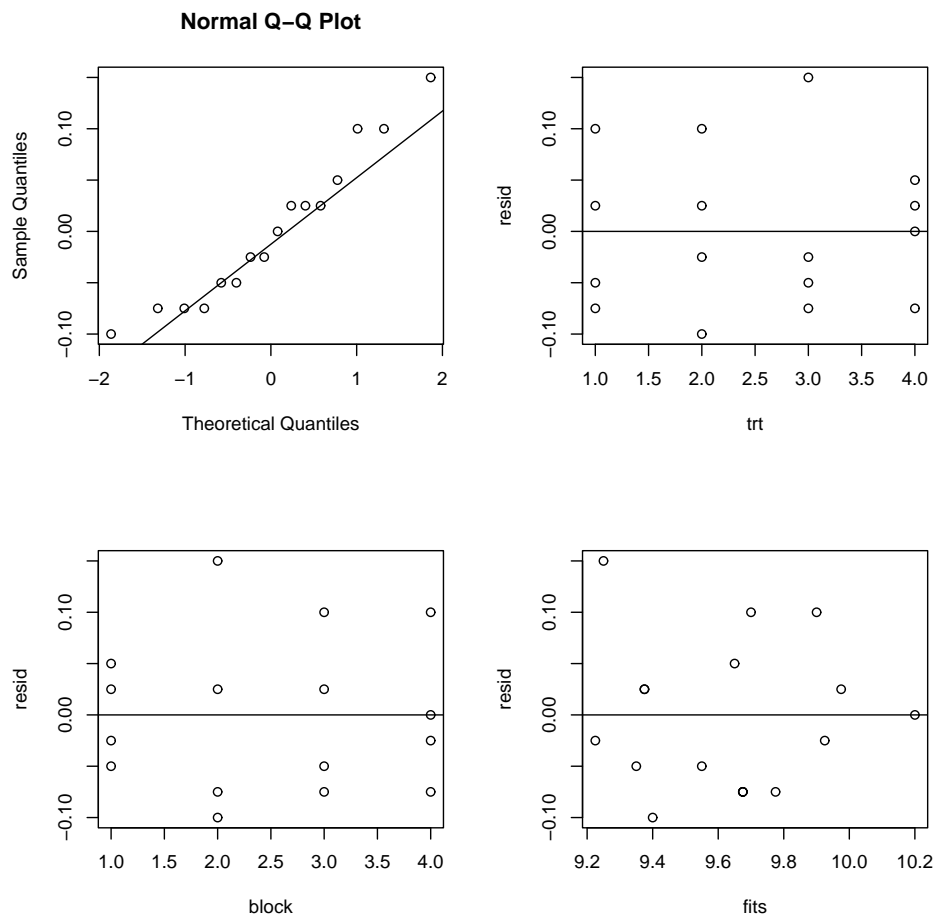


Fig. 4.2

Does the error variance depend on the treatment, or on the block? Apparently not:

```
> bartlett.test(y,treatments)
Bartlett test for homogeneity of variances
data: y and treatments
Bartlett's K-squared = 0.4477, df = 3,
      p-value = 0.9302
```

```
> bartlett.test(y,blocks)
Bartlett test for homogeneity of variances
data: y and blocks
Bartlett's K-squared = 0.9463, df = 3,
      p-value = 0.8142
```

The normality-based tests can be justified here since we have little evidence of non-normality. It's a good idea to run nonparametric tests too, to reassure ourselves that we reach the same conclusions without assuming normality.



## 11. RCBD III; Latin squares

**Levene's test for equal variances in each block.** The blocks correspond to the 4 columns of the data, so we first compute the medians of the columns, subtract then from each column and take absolute values:  $d_{ij} = |y_{ij} - \text{med}\{y_{1j}, \dots, y_{aj}\}|$ . Then do an anova to see if  $E[d_{ij}]$  is the same in each block.

```
data.matrix <- as.matrix(data)
abs.diff <- matrix(nrow=4, ncol=4)
for(i in 1:4) {for (j in 1:4) abs.diff[i,j] <-
  abs(data.matrix[i,j] - median(data.matrix[ ,j]))}
g.blocks <- lm(c(t(abs.diff))~blocks)
anova(g.blocks)
```

Analysis of Variance Table

Response: c(t(abs.diff))

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
blocks	3	0.022500	0.007500	0.5806	0.6389
Residuals	12	0.155000	0.012917		

Using treatments rather than blocks gives  $p = .698$  (you should check this).

The analogue of the Kruskal-Wallis test, for a RCBD, is 'Friedman's test'. The observations are replaced by their ranks *within each block*, and the usual ANOVA is run. This method does not take account of the fact that the denominator of the  $F$  is not random (as before); the function `friedman.test` will do so. The differences are generally slight.

```
> friedman.test(y,treatments,blocks)
Friedman rank sum test
data: y, treatments and blocks
Friedman chi-squared = 8.8462, df = 3,
p-value = 0.03141
```

So - the assumptions seem to be met, and at least some of the differences in the treatment means, i.e. in the mean readings  $\mu_i. = \mu + \tau_i$ , are significant - the readings of the hardness testing device depend on which tip is being used. This is bad news for the engineers. Is there any one tip responsible for the differences? We should look at all of the differences  $\hat{\mu}_i. - \hat{\mu}_j. = \bar{y}_i. - \bar{y}_j.$  to see which are significant.

- Method 1: Fisher's LSD ("Least Significant Difference"). A  $100(1 - \alpha)\%$  confidence interval on one difference  $\mu_i. - \mu_j.$  is

$$\begin{aligned} & \bar{y}_i. - \bar{y}_j. \pm t_{\alpha/2, df(MSE)} \sqrt{MSE \left( \frac{1}{b} + \frac{1}{b} \right)} \\ = & \bar{y}_i. - \bar{y}_j. \pm t_{\alpha/2, 9} \sqrt{.00889 \left( \frac{2}{4} \right)}. \end{aligned}$$

With  $\alpha = .05$ , the 95% interval is  $\bar{y}_i. - \bar{y}_j. \pm .151$ . Converting this to a hypothesis test, we see that the hypothesis of equality is rejected if

$$|\bar{y}_i. - \bar{y}_j.| > LSD = .151.$$

Since

$$\begin{aligned}
 |\bar{y}_{1.} - \bar{y}_{2.}| &= .025 < .151, \\
 |\bar{y}_{1.} - \bar{y}_{3.}| &= .125 < .151, \\
 |\bar{y}_{1.} - \bar{y}_{4.}| &= .300 > .151, * \\
 |\bar{y}_{2.} - \bar{y}_{3.}| &= .150 < .151, \\
 |\bar{y}_{2.} - \bar{y}_{4.}| &= .275 > .151, * \\
 |\bar{y}_{3.} - \bar{y}_{4.}| &= .425 > .151, *
 \end{aligned}$$

we conclude that tips 1,2 and 3 produce identical hardness readings but that tip 4 gives significantly different (and higher) readings. In making these statements our experimentwise error rate is  $< 6 \times .05 = .3$ , so our overall confidence is  $> 70\%$ .

- Method 2: Tukey's procedure replaces  $t_{\alpha/2,9}$  with  $\frac{q_{\alpha}}{\sqrt{2}} = \text{qtukey}(.95, 4, 9) / \text{sqrt}(2) = 3.1218$  to get

$$\frac{q_{\alpha}}{\sqrt{2}} \cdot se(\bar{y}_{i.} - \bar{y}_{j.}) = 3.1218 \sqrt{.00889 \left(\frac{2}{4}\right)} = .208.$$

The same conclusions are drawn, with an experimentwise error rate of only .05.

## Latin Squares

Same (hardness) example. Suppose that the ‘operator’ of the testing machine was also thought to be a factor. We suppose that there are  $p = 4$  operators,  $p = 4$  coupons, and  $p = 4$  tips. The first two are nuisance factors, the last is the ‘treatment’. We can carry out the experiment, and estimate everything we need to, in only  $p^2 = 16$  runs (as before), if we use a *Latin Square Design*. Here each tip is used exactly once on each coupon, and exactly once by each operator. Represent the treatments by the Latin letters A, B, C, D and consider the Latin square:

Coupon	Operator			
	$k = 1$	$k = 2$	$k = 3$	$k = 4$
$i = 1$	A	D	B	C
$i = 2$	B	A	C	D
$i = 3$	C	B	D	A
$i = 4$	D	C	A	B

Each letter appears exactly once in each row and in each column. There are many ways to construct a Latin square, and the randomization enters into things by randomly choosing one of them. Suppose the data were:

Coupon	Operator			
	$k = 1$	$k = 2$	$k = 3$	$k = 4$
$i = 1$	$A = 9.3$	$B = 9.3$	$C = 9.5$	$D = 10.2$
$i = 2$	$B = 9.4$	$A = 9.4$	$D = 10.0$	$C = 9.7$
$i = 3$	$C = 9.2$	$D = 9.6$	$A = 9.6$	$B = 9.9$
$i = 4$	$D = 9.7$	$C = 9.4$	$B = 9.8$	$A = 10.0$

We use an effects model

$$y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k, \quad i, j, k = 1, \dots, p$$

where:

(i)  $y_{ijk}$  is the observation in row  $i$ , column  $k$ , using treatment  $j$ . (So  $y_{243} = 10.0$ ,  $y_{223}$  does not exist - only  $p^2$  of them do.)

(ii)  $\alpha_i, \tau_j, \beta_k$  are the row, treatment and column effects, all summing to zero.

Note that the model is additive in that there is no interaction effect: any treatment has the *same* effect regardless of the levels of the other factors.

## 12. Latin and Graeco-Latin Squares

For the Latin square design the LSEs are

$$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...}, \quad \hat{\tau}_j = \bar{y}_{.j.} - \bar{y}_{...}, \quad \hat{\beta}_k = \bar{y}_{..k} - \bar{y}_{...}$$

as usual, with (as usual)

$$SS_{Rows} = p \sum_{i=1}^p \hat{\alpha}_i^2, \quad SS_{Tr} = p \sum_{j=1}^p \hat{\tau}_j^2, \quad SS_{Col} = p \sum_{k=1}^p \hat{\beta}_k^2$$

$$SS_E = SS_T - SS_{Rows} - SS_{Tr} - SS_{Col}.$$

The d.f. of these are  $p - 1, p - 1, p - 1$  and  $p^2 - 1 - 3(p - 1) = (p - 2)(p - 1)$ .

Theoretical ANOVA table:

Source	SS	df	MS	F <sub>0</sub>
Treat.	$SS_{Tr}$	$p - 1$	$MS_{Tr} = \frac{SS_{Tr}}{p-1}$	$F_0 = \frac{MS_{Tr}}{MS_E}$
Rows	$SS_{Row}$	$p - 1$	$MS_{Row} = \frac{SS_{Row}}{p-1}$	
Col.	$SS_{Col}$	$p - 1$	$MS_{Col} = \frac{SS_{Col}}{p-1}$	
Error	$SS_E$	$(p-1) \cdot (p-2)$	$MS_E = \frac{SS_E}{df}$	
Total	$SS_T$	$p^2 - 1$		



The value of  $F_0$  is of course compared to  $F_{(p-1)(p-2)}^{p-1}$ :  
 p-value for the hypothesis of equal treatments effects  
 is  $P\left(F_{(p-1)(p-2)}^{p-1} > F_0\right)$ .

```
> y <- c(9.3, 9.4, 9.2, 9.7, ... etc.)
> operators <- as.factor(rep(1:4, each=4))
> coupons <- as.factor(rep(1:4, times=4))
> tips <- as.factor(c("A", "B", "C", "D", etc.))
> data <- data.frame(y, operators, coupons, tips)
> data
```

	y	operators	coupons	tips
1	9.3	1	1	A
2	9.4	1	2	B
3	9.2	1	3	C
4	9.7	1	4	D
5	9.3	2	1	B
6	9.4	2	2	A
7	9.6	2	3	D
8	9.4	2	4	C
9	9.5	3	1	C
				etc.

```
> g <- lm(y~tips + operators + coupons)
```

```
> anova(g)
```

```
Analysis of Variance Table
```

```
Response: y
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
tips	3	0.38500	0.12833	38.5	0.0002585	***
operators	3	0.82500	0.27500	82.5	2.875e-05	***
coupons	3	0.06000	0.02000	6.0	0.0307958	*
Residuals	6	0.02000	0.00333			

```
---
```

From the ANOVA table the means are significantly different

```
treatment.means <- vector(length=4)
```

```
for (j in c("A","B","C","D")) treatment.means[j]
  <- mean(y[tips==j])
```

```
> treatment.means
```

```
[1] 9.600 9.625 9.700 9.575
```

```
# These are the averages for tips A, B, C, D
```

```
> qtukey(.95,4,6)*sqrt(.0033/4)
```

```
[1] 0.1406154
```

- Tukey's procedure says that  $\mu_{.k.}$  and  $\mu_{.l.}$  are significantly different ( $\alpha = .05$ ) if

$$|\bar{y}_{.k.} - \bar{y}_{.l.}| > \text{qtukey}(.95, 4, 6) \sqrt{\frac{MS_E}{p}} = .141,$$

so again we conclude that tip 4 gives significantly different readings. Tips 2 and 3 seem significantly different as well. Now the coupons don't seem to affect the readings, although it appears that the operators do.

- The usual model checks should be done - qq-plot of residuals (to check normality), plots of residuals against row labels, column labels, treatment labels (to check for constant variances). Bartlett's test can be carried out if normality is assured. Levene's test for variance equality in the treatment groups would involve computing the absolute deviations from the medians in each treatment group and doing a one-way ANOVA on them:

```
y.d <- y
for(j in c("A","B","C","D")) y.d[tips==j]<-
  abs(y[tips==j]-median(y[tips==j]))
> g.d <- lm(y.d ~tips)
> anova(g.d)
```

### Analysis of Variance Table

Response: y.d

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
tips	3	0.022500	0.007500	0.4865	0.698
Residuals	12	0.185000	0.015417		

The Latin square notion extends to *Graeco-Latin squares*. Suppose that we had one more factor - day of the week, at four levels  $\alpha$  (Monday),  $\beta$  (Tuesday)  $\gamma$  (Wednesday)  $\delta$  (Thursday), of importance if the whole experiment took 4 days to complete. Superimpose a  $4 \times 4$  Latin squares consisting of these Greek letters, in such a way that each (Latin,Greek) combination of letters occurs exactly once:

Coupon	Operator			
	$k = 1$	$k = 2$	$k = 3$	$k = 4$
$i = 1$	$A\alpha = 9.3$	$B\beta = 9.3$	$C\gamma = 9.5$	$D\delta = 10.2$
$i = 2$	$B\delta = 9.4$	$A\gamma = 9.4$	$D\beta = 10.0$	$C\alpha = 9.7$
$i = 3$	$C\beta = 9.2$	$D\alpha = 9.6$	$A\delta = 9.6$	$B\gamma = 9.9$
$i = 4$	$D\gamma = 9.7$	$C\delta = 9.4$	$B\alpha = 9.8$	$A\beta = 10.0$

The additive model has terms for all four factors - coupons, operators, days and tips. Each is estimated by the sample average for that level of that factor, minus the overall average. For example the LSE of the effect of Tuesday is

$$\hat{\theta}_2 = \frac{9.2 + 9.3 + 10.0 + 10.0}{4} - \bar{y} \dots$$

and

$$SS_{Days} = p \sum_{l=1}^p \hat{\theta}_l^2.$$

Each factor uses  $(p - 1)$  d.f., so that  $SS_E$  is on only  $p^2 - 1 - 4(p - 1) = (p - 3)(p - 1)$  d.f.

```
> days <- as.factor(c(1,4,2,3, 2,3,1,4,
                      3,2,4,1, 4,1,3,2))
> h <- lm(y~tips + operators + coupons + days)
> anova(h)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
tips	3	0.38500	0.12833	25.6667	0.012188	*
operators	3	0.82500	0.27500	55.0000	0.004029	**
coupons	3	0.06000	0.02000	4.0000	0.142378	
days	3	0.00500	0.00167	0.3333	0.804499	
Residuals	3	0.01500	0.00500			

### 13. Balanced Incomplete Block Designs

In the RCBD, C='Complete' means that each block contains each treatment. E.g. each coupon is subjected to each of the 4 tips. Suppose that a coupon is only large enough that 3 tips can be used. Then the blocks would be 'incomplete'. One way to run the experiment is to randomly assign 3 tips to each block, perhaps requiring that each tip appears 3 times in total. There is a more efficient way. An incomplete block design is 'balanced' if *any two treatments appear in the same block an equal number of times*. This is then a *Balanced Incomplete Block Design*.

Hardness testing  
BIBD design and data.

	Coupon					
Tip	1	2	3	4	$y_{i.}$	$Q_i$
1	9.3	9.4	—	10.0	28.7	— .1000
2	—	9.3	9.8	9.9	29.0	— .1667
3	9.2	9.4	9.5	—	28.1	— .4333
4	9.7	—	10.0	10.2	29.9	.7000
$y_{.j}$	28.2	28.1	29.3	30.1		

Notation:

$a = \#$  of treatments,  $b = \#$  of blocks.  $a = 4, b = 4.$

$k = \#$  of treatments per block.  $k = 3.$

$r = \#$  of times each treatment appears in the entire experiment.  $r = 3.$

$N = ar = bk = \#$  of observations.  $N = 12.$

$\lambda = \#$  of times each pair of treatments appears together.  $\lambda = 2.$  It can be shown that

$$\lambda = \frac{r(k-1)}{a-1}.$$

Since these parameters have to be integers, BIBD designs don't exist for all choices of  $a, b, k, r, \lambda.$  There are tables available of the ones that do exist.



Model is as for a RCBD:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

with both sums of effects vanishing. As usual, the total sum of squares  $SS_T = \sum_{i,j} (y_{ij} - \bar{y}_{..})^2$  on  $N-1$  d.f. and the SS for Blocks is  $SS_{Blocks} = k \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2$  on  $b-1$  d.f. The treatment SS depends on the 'adjusted total for the  $i^{th}$  treatment'

$$Q_i = y_{i.} - \frac{1}{k} \sum_{j=1}^b n_{ij} y_{.j}$$

where  $n_{ij} = 1$  if treatment  $i$  appears in block  $j$  and  $= 0$  otherwise. So  $\sum_{j=1}^b n_{ij} y_{.j}$  is the total of the block totals, counting only those blocks that contain treatment  $i$ :

$$Q_1 = 28.7 - \frac{1}{3} (28.2 + 28.1 + 30.1) = -.1000,$$

$$Q_2 = 29.0 - \frac{1}{3} (28.1 + 29.3 + 30.1) = -.1667,$$

$$Q_3 = 28.1 - \frac{1}{3} (28.2 + 28.1 + 29.3) = -.4333,$$

$$Q_4 = 29.9 - \frac{1}{3} (28.2 + 29.3 + 30.1) = .7000.$$

(As a check, it is always the case that  $\sum Q_i = 0$ .)

Why so complicated? Due to the incompleteness of the blocks,  $\bar{y}_{i.} - \bar{y}_{..}$  is no longer an unbiased estimate of  $\tau_i$ . For instance in our example

$$\begin{aligned} E[y_{1.}] &= E[y_{11} + y_{12} + y_{14}] \\ &= 3\mu + 3\tau_1 + \beta_1 + \beta_2 + \beta_4; \\ E[y_{..}] &= 12\mu + r(\tau_1 + \tau_2 + \tau_3 + \tau_4) + k \sum_{j=1}^4 \beta_j \\ &= 12\mu. \end{aligned}$$

Then

$$\begin{aligned} E[\bar{y}_{1.} - \bar{y}_{..}] &= \frac{3\mu + 3\tau_1 + \beta_1 + \beta_2 + \beta_4}{3} - \frac{12\mu}{12} \\ &= \tau_1 + \frac{\beta_1 + \beta_2 + \beta_4}{3}. \end{aligned}$$

The block totals must be brought in, in order to adjust for this bias. It turns out that the LSEs of the treatment effects are

$$\hat{\tau}_i = \frac{kQ_i}{\lambda a},$$

and that these are unbiased. (In asst. 2 you will show this for the hardness experiment, when  $i = 1$ .)

The 'Sum of Squares of Treatments, adjusted for Blocks' is

$$SS_{Tr(Bl)} = \frac{k}{\lambda a} \sum_{i=1}^a Q_i^2,$$

on  $a - 1$  d.f. In our case

$$SS_{Tr(Bl)} = \frac{3}{2 \cdot 4} \times .7155 = .2683.$$

The idea is that we first estimate the block effects and then see how much of the remaining variation is attributable to treatments. Doing it in the other order results in something quite different.

Correct analysis:

```
> g <- lm(y ~ coupons + tips)
```

```
> anova(g)
```

```
Analysis of Variance Table
```

```
Response: y
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
coupons	3	0.90917	0.30306	29.3280	0.001339	**
tips	3	0.26833	0.08944	8.6559	0.020067	*
Residuals	5	0.05167	0.01033			

```
---
```

Incorrect analysis:

```
> h <- lm(y ~ tips + coupons)
```

```
> anova(h)
```

```
Analysis of Variance Table
```

```
Response: y
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
tips	3	0.56250	0.18750	18.145	0.004054	**
coupons	3	0.61500	0.20500	19.839	0.003311	**
Residuals	5	0.05167	0.01033			

```
---
```

To make inferences we use the fact that the  $\hat{\tau}_i$  are independent and equally varied, with

$$VAR[\hat{\tau}_i] = \frac{k\sigma^2}{\lambda\alpha},$$

so that

$$se(\hat{\tau}_i - \hat{\tau}_j) = \sqrt{\frac{2k}{\lambda\alpha} \cdot MS_E}.$$

In our example this is .0880, so that single confidence intervals are

$$\hat{\tau}_i - \hat{\tau}_j \pm t_{\alpha/2,5} \cdot .0880$$

and simultaneous Tukey-type intervals replace  $t_{\alpha/2,5}$  by  $qtukey(1 - \alpha, 4, 5) / \sqrt{2}$ .

## **Part V**

# **INTRODUCTION TO FACTORIAL DESIGNS**

## 14. Introduction to Factorial Designs

- We study the effects of two or more factors, each at several levels. A **Factorial Design** has observations at all combinations of these levels.
- Example. Batteries are of two types ('1' and '2';  $a = 2$  levels of Factor A) and their lifetimes may depend on the temperature at which they are used (LO = '1', HI='2';  $b = 2$  levels of factor B).  $n = 4$  observations are made at each of the  $2^2$  ( $= \text{factors}^{\text{levels}}$ ) combinations of levels. The  $nab = 16$  observations are made in random order, so this is also a CRD.

Type (A)	Temperature level (B)	
	1 (LO)	2 (HI)
1	130, 155, 74, 180	20, 70, 82, 58
2	150, 188, 159, 126	25, 70, 58, 45

We see the effect of changing one factor, while leaving the other fixed, by plotting the means  $\bar{y}_{ij}$  at the 4 combinations.

```

y <- c(130,...,58,150,...,45)
type <- as.factor(rep(1:2,each=8))
temp <- as.factor(rep(1:2, each=4, times=2))
data <- data.frame(y,type,temp)
interaction.plot(type,temp,y)
interaction.plot(temp,type,y)

```

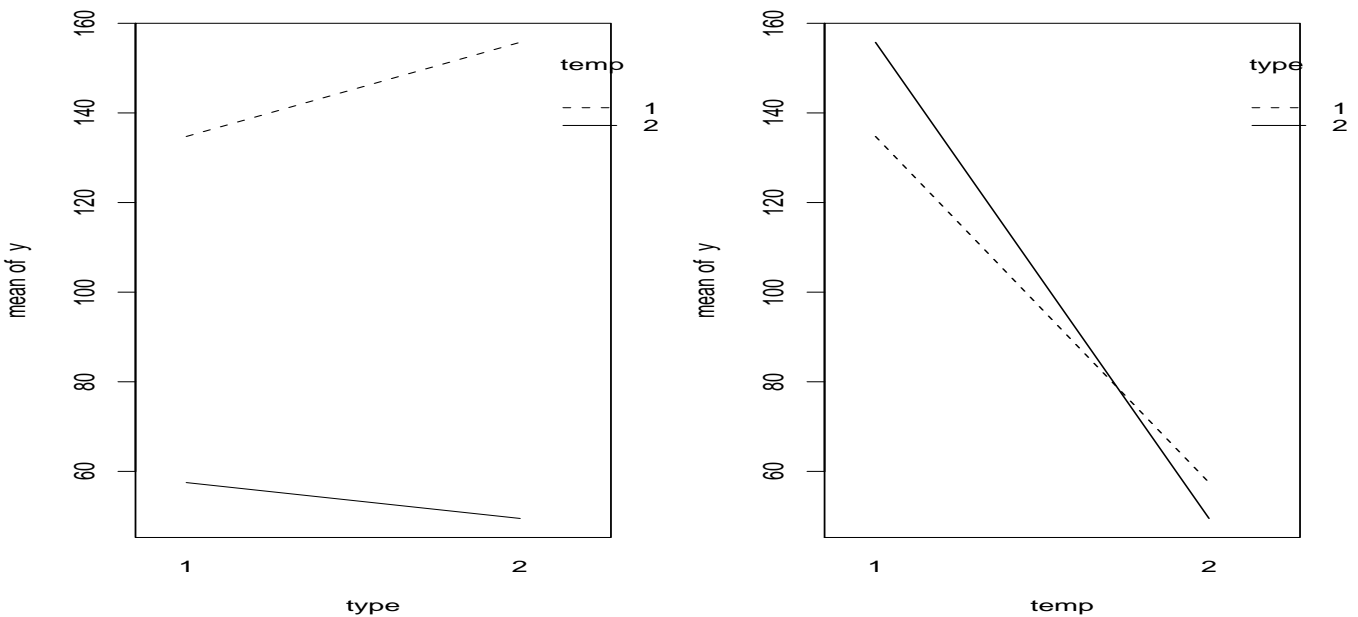


Fig. 5.1. Some interaction shown.

The average lifetimes at the 4 combinations of levels are



Type (A)	Temperature level (B)	
	1 (LO)	2 (HI)
1	134.75	57.5
2	155.75	49.5

The 'main effect of A' is the change in response caused by changing the level of A. Here it is estimated by the difference in the average responses at those levels of Factor A:

$$A = \frac{155.75 + 49.5}{2} - \frac{134.75 + 57.5}{2} = 6.5.$$

Similarly

$$B = \frac{57.5 + 49.5}{2} - \frac{134.75 + 155.75}{2} = -91.75.$$

Because of the interactions, these main effects are misleading. At the low level of B, the effect of A is  $155.75 - 134.75 = 21.00$ . At the high level, it is  $49.5 - 57.5 = -8.0$ . The 'interaction effect' is measured by the average difference between these two:

$$AB = \frac{-8.0 - (21.00)}{2} = -14.5.$$

We will extend these ideas to more complex situations.

- Same example, but now 3 levels of each factor ( $3^2$  factorial);  $n = 4$  observations at each of the  $ab = 9$  combinations of levels. So  $N = nab = 36$ . Data in text (Table 5.1) and on web site.

The effects model, including terms for interaction, is that the  $k^{\text{th}}$  observation at level  $i$  of A,  $j$  of B is

$$\begin{aligned}
 y_{ijk} &= \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \\
 i &= 1, \dots, a = \text{levels of Factor A} \\
 j &= 1, \dots, b = \text{levels of Factor B} \\
 k &= 1, \dots, n.
 \end{aligned}$$

Constraints  $\sum_i \tau_i = 0$  (average effect of levels of A is 0),  $\sum_j \beta_j = 0$  (average effect of levels of B is 0), and average interactions  $\sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0$ .

Reasonable estimates of these effects, obeying these constraints, are  $\hat{\mu} = \bar{y}_{...}$  and

$$\begin{aligned}
 \hat{\tau}_i &= \bar{y}_{i..} - \bar{y}_{...}, \\
 \hat{\beta}_j &= \bar{y}_{.j.} - \bar{y}_{...}, \\
 (\widehat{\tau\beta})_{ij} &= (\bar{y}_{ij.} - \bar{y}_{...}) - \hat{\tau}_i - \hat{\beta}_j \\
 &= \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}.
 \end{aligned}$$

These are shown to be the LSEs in the usual way:

$$\begin{aligned}
1. \text{ Decompose } SS_T \text{ as } SS_T &= \sum_{i,j,k} (y_{ijk} - \bar{y}_{...})^2 = \\
&= \sum_{i,j,k} \left( \hat{\tau}_i + \hat{\beta}_j + (\widehat{\tau\beta})_{ij} + (y_{ijk} - \bar{y}_{ij.}) \right)^2 \\
&= nb \sum_i \hat{\tau}_i^2 + na \sum_j \hat{\beta}_j^2 + n \sum_{i,j} (\widehat{\tau\beta})_{ij}^2 \\
&\quad + \sum_{i,j,k} (y_{ijk} - \bar{y}_{ij.})^2 + 6 \text{ zeros (how?)} \\
&= SS_A + SS_B + SS_{AB} + SS_E,
\end{aligned}$$

on  $a - 1$ ,  $b - 1$ ,  $(a - 1)(b - 1)$  and  $ab(n - 1)$  d.f.

2. Minimize

$$\begin{aligned}
S(\mu, \tau, \beta, (\tau\beta)) &= \sum_{i,j,k} (y_{ijk} - E[y_{ijk}])^2 \\
&= \sum_{i,j,k} (y_{ijk} - \mu - \tau_i - \beta_j - (\tau\beta)_{ij})^2
\end{aligned}$$

by writing it as  $S(\mu, \tau, \beta, (\tau\beta)) =$

$$\begin{aligned} & \sum_{i,j,k} \left( \begin{array}{l} [y_{ijk} - \bar{y}_{ij.}] - [\mu - \hat{\mu}] - [\tau_i - \hat{\tau}_i] \\ - [\beta_j - \hat{\beta}_j] - [(\tau\beta)_{ij} - (\widehat{\tau\beta})_{ij}] \end{array} \right)^2 \\ = & SS_E + N[\mu - \hat{\mu}]^2 + nb \sum_i [\tau_i - \hat{\tau}_i]^2 \\ & + na \sum_j [\beta_j - \hat{\beta}_j]^2 + n \sum_{i,j} [(\tau\beta)_{ij} - (\widehat{\tau\beta})_{ij}]^2. \end{aligned}$$

Thus

$$S(\mu, \tau, \beta, (\tau\beta)) \geq SS_E = S(\hat{\mu}, \hat{\tau}, \hat{\beta}, (\widehat{\tau\beta})),$$

so that  $\hat{\mu}, \hat{\tau}, \hat{\beta}, (\widehat{\tau\beta})$  are the minimizers and the minimum value is  $SS_E$ .

3. Under the hypothesis of no interactions, the quantity to be minimized is  $S(\mu, \tau, \beta, 0)$ , with minimum value

$$\begin{aligned} SS_{Reduced} &= SS_E + n \sum_{i,j} (\widehat{\tau\beta})_{ij}^2 \\ &= SS_E + SS_{AB}. \end{aligned}$$

Thus the appropriate F-ratio is

$$F_0 = \frac{MS_{AB}}{MS_E} \sim F_{(a-1)(b-1), ab(n-1)}.$$

If the interactions are not significant then it makes sense to ask about the significance of the levels of the factors, using  $MS_A/MS_E$ , etc. The expected values of the mean squares turn out to be what one would expect:  $E[MS_E] = \sigma^2$  and

$$\begin{aligned} E[MS_{AB}] &= \sigma^2 + \frac{n \sum_{i,j} (\tau\beta)_{ij}^2}{(a-1)(b-1)}, \\ E[MS_A] &= \sigma^2 + \frac{nb \sum_i \tau_i^2}{a-1}, \\ E[MS_B] &= \sigma^2 + \frac{na \sum_j \beta_j^2}{b-1}. \end{aligned}$$

## 15. ANOVA; model checking

**Summary.** Theoretical ANOVA table for a two factor factorial experiment with  $n$  observations per cell:

Source	SS	df	MS	$F_0$
A	$SS_A$	$a - 1$	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B	$SS_B$	$b - 1$	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
AB	$SS_{AB}$	$df(AB)$	$MS_{AB} = \frac{SS_{AB}}{df(AB)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
Error	$SS_E$	$df(Err)$	$MS_E = \frac{SS_E}{df(Err)}$	
Total	$SS_T$	$abn - 1$		

$$df(AB) = (a - 1)(b - 1), \quad df(Err) = ab(n - 1)$$

$$SS_A = nb \sum_i \hat{\tau}_i^2, \quad SS_B = na \sum_j \hat{\beta}_j^2,$$

$$SS_{AB} = n \sum_{i,j} (\widehat{\tau\beta})_{ij}^2, \quad SS_E = \sum_{i,j,k} (y_{ijk} - \bar{y}_{ij.})^2.$$

$$\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}, \quad \hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...},$$

$$(\widehat{\tau\beta})_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}.$$

In the battery experiment the various averages are

```
> means <- matrix(nrow=3,ncol=3)
> for(i in 1:3) {for (j in 1:3) means[i,j] <-
  mean(y[type==i & temp == c(15,70,125)[j]])}
> means
```

	[Temp=15]	[Temp=70]	[Temp=125]
[Type1]	134.75	57.25	57.5
[Type2]	155.75	119.75	49.5
[Type3]	144.00	145.75	85.5

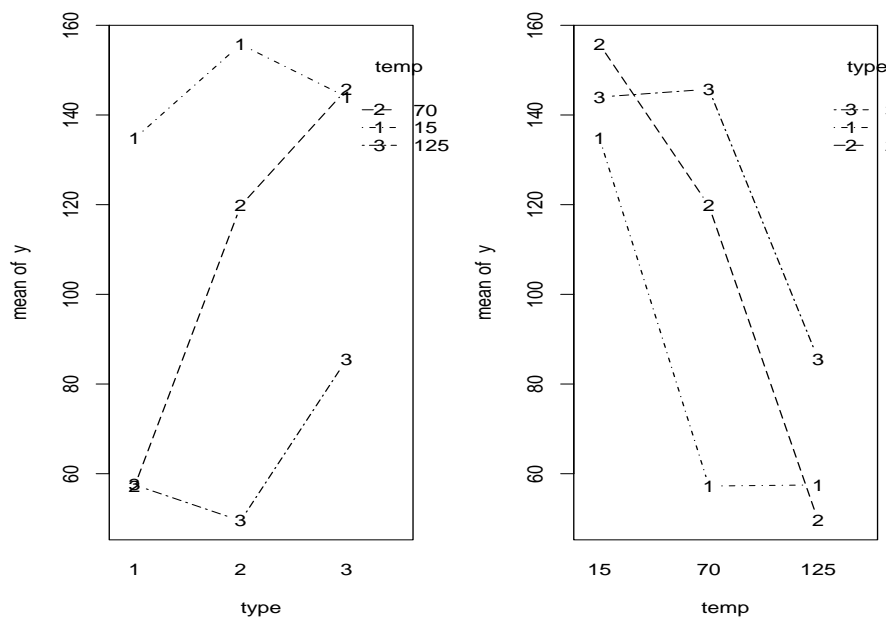


Fig. 5.2

```
> g <- lm(y~type+temp+type*temp)
> anova(g)
```

### Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
type	2	10684	5342	7.9114	0.001976
temp	2	39119	19559	28.9677	1.909e-07
type:temp	4	9614	2403	3.5595	0.018611
Residuals	27	18231	675		

As suspected, the interaction effects are quite significant. There is no battery type which is ‘best’ at all temperatures. If interactions were NOT significant one could compare the  $\mu + \tau_i$  by seeing which of the differences  $\hat{\mu} + \hat{\tau}_i = \bar{y}_{i..}$  were significantly different from each other (using  $se(\bar{y}_{i..} - \bar{y}_{k..}) = \sqrt{\frac{2MS_E}{nb}}$ ).



As it is, we can only make comparisons at fixed levels of the other factor. For instance when  $temp = 70$ , ( $j = 2$ ) we can compare the means

$$\mu_{i2} = \mu + \tau_i + \beta_2 + (\tau\beta)_{i2},$$

with estimates  $\bar{y}_{i2}$ . (each an average of  $n$  observations). The 95% Tukey CIs on  $\mu_{i2} - \mu_{k2}$  are

$$\begin{aligned} & \bar{y}_{i2} - \bar{y}_{k2} \pm \text{qtukey}(.95, 3, 27) \sqrt{\frac{MS_E}{n}} \\ & = \bar{y}_{i2} - \bar{y}_{k2} \pm 45.55. \end{aligned}$$

Since  $\bar{y}_{12} = 57.25$ ,  $\bar{y}_{22} = 119.75$  and  $\bar{y}_{32} = 145.75$  we conclude that  $\mu_{12}$  is significantly less than  $\mu_{22}$  and  $\mu_{32}$ , but that these two are not significantly different from each other.

**Model checking.** First suppose that we incorrectly ignored interactions, and fit an additive model:

```
> h <- lm(y~type+temp)
```

```
> anova(h)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
type	2	10684	5342	5.9472	0.006515
temp	2	39119	19559	21.7759	1.239e-06
Residuals	31	27845	898		

The error would show up in the residual plot against the fitted values.

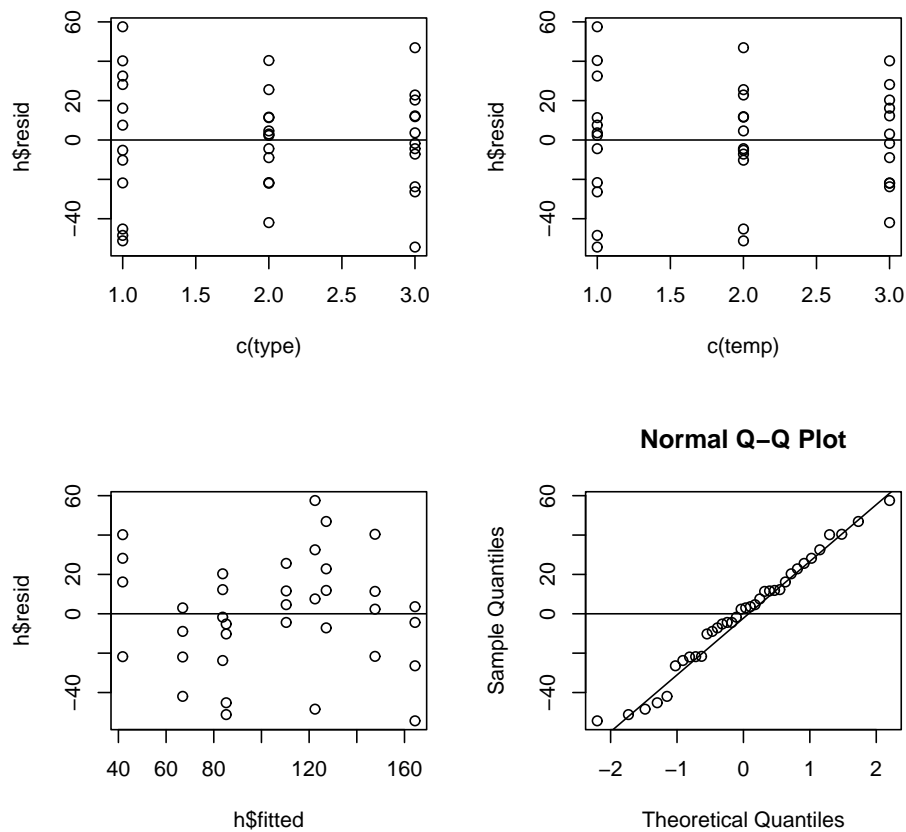


Fig 5.3.

Note that within each of the 9 groups the residuals tend to be of the same sign, with the signs alternating as we move from group to group.

The residuals from the correct fit look better:

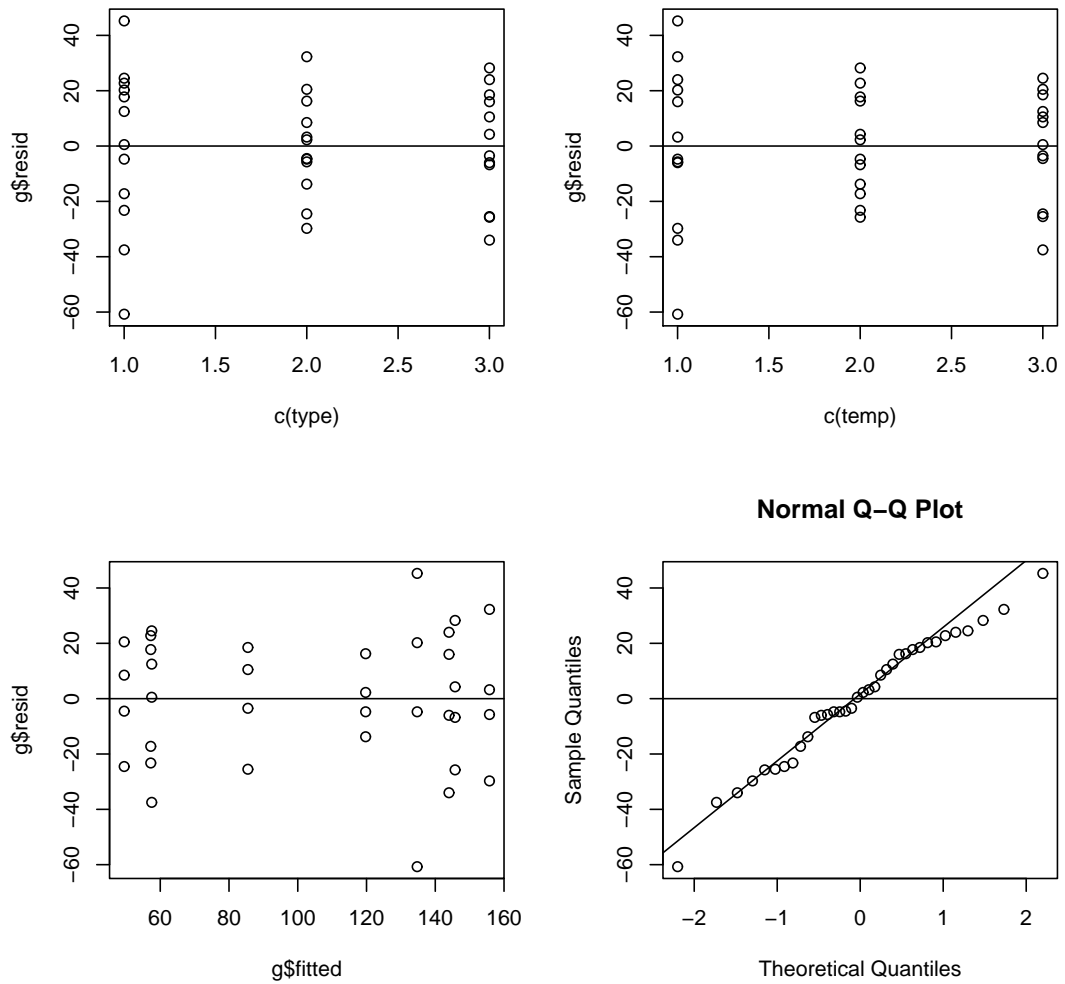


Fig 5.4.

## 16. Test for non-additivity; 3 factor designs

Sometimes we can make only one observation per cell ( $n = 1$ ). Then all  $y_{ij1} - \bar{y}_{ij.} = 0$ , so  $SS_E = 0$  on  $ab(n - 1) = 0$  d.f. The interaction SS, which for  $n = 1$  is

$$\sum_{i,j} \left( y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..} \right)^2, \quad (*)$$

is what we should be using to estimate experimental error. There is still however a way to test for interactions, if we assume that they take a simple form:

$$(\tau\beta)_{ij} = \gamma\tau_i\beta_j.$$

We carry out ‘Tukey’s one d.f. test for interaction’, which is an application of the usual ‘reduction in SS’ hypothesis testing principle. Our ‘full’ model is

$$y_{ij} = \mu + \tau_i + \beta_j + \gamma\tau_i\beta_j + \varepsilon_{ij}.$$

Under the null hypothesis  $H_0: \gamma = 0$  of no interactions, the ‘reduced’ model is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

in which the minimum SS (i.e.  $SS_{Red}$ ) is (\*) above. One computes

$$F_0 = \frac{SS_{Red} - SS_{Full}}{MS_E(Full)} \sim F_{(a-1)(b-1)-1}^1.$$

The difference

$$SS_N = SS_{Red} - SS_{Full}$$

is called the 'SS for non-additivity', and uses 1 d.f. to estimate the one parameter  $\gamma$ . The ANOVA becomes

Source	SS	df	MS
A	$SS_A$	$a - 1$	$MS_A = \frac{SS_A}{a-1}$
B	$SS_B$	$b - 1$	$MS_B = \frac{SS_B}{b-1}$
N	$SS_N$	1	$MS_N = \frac{SS_N}{1}$
Error	$SS_E$	$(a-1)(b-1)$	$MS_E = \frac{SS_E}{df(Err)}$
Total	$SS_T$	$ab - 1$	

The error SS is  $SS_{Full}$ . To obtain it one has to minimize

$$\sum_{i,j} \left( y_{ij} - \left[ \mu + \tau_i + \beta_j + \gamma\tau_i\beta_j \right] \right)^2.$$

After a calculation it turns out that

$$SS_N = \frac{ab \left\{ \sum_{i,j} y_{ij} \bar{y}_{i.} \bar{y}_{.j} - \bar{y}_{..} (SS_A + SS_B + ab\bar{y}_{..}^2) \right\}^2}{SS_A \cdot SS_B}.$$

Then  $SS_E$  is obtained by subtraction:  $SS_E = SS_{Red} - SS_N$ .

An R function to calculate this, and carry out the F-test, is at “R commands for Tukey’s 1 df test” on the course web site.

**Example.** For the experiment at Example 5.2 of the text there are  $a = 3$  levels of temperature and  $b = 5$  of pressure; response is  $Y =$  impurities in a chemical product.

```
> h <- tukey.1df(y,temp,press)
      SS      df MS      F0      p
A  23.333  2  11.667  42.949 1e-04
B  11.6    4   2.9    10.676 0.0042
N   0.099  1   0.099   0.363 0.566
Err  1.901  7   0.272
Tot  36.933 14
```

**A 3 factor example.** Softdrink bottlers must maintain targets for fill heights, and any variation is a cause for concern. The deviation from the target ( $Y$ ) is affected by %carbonation ( $A$ ), pressure in the filler ( $B$ ), line speed ( $C$ ). These are set at  $a = 3, b = 2, c = 2$  levels respectively, with  $n = 2$  observations at each combination ( $N = nabc = 24$  runs, in *random order*).

	y	carbon	press	speed
1	-3	10	25	200
2	-1	10	25	200
3	0	12	25	200
4	1	12	25	200
5	5	14	25	200
6	4	14	25	200
		.....		
19	1	10	30	250
20	1	10	30	250
21	6	12	30	250
22	5	12	30	250
23	10	14	30	250
24	11	14	30	250



- > plot.design(data)
- > interaction.plot(carbon,press,y)
- > interaction.plot(carbon,speed,y)
- > interaction.plot(press,speed,y)

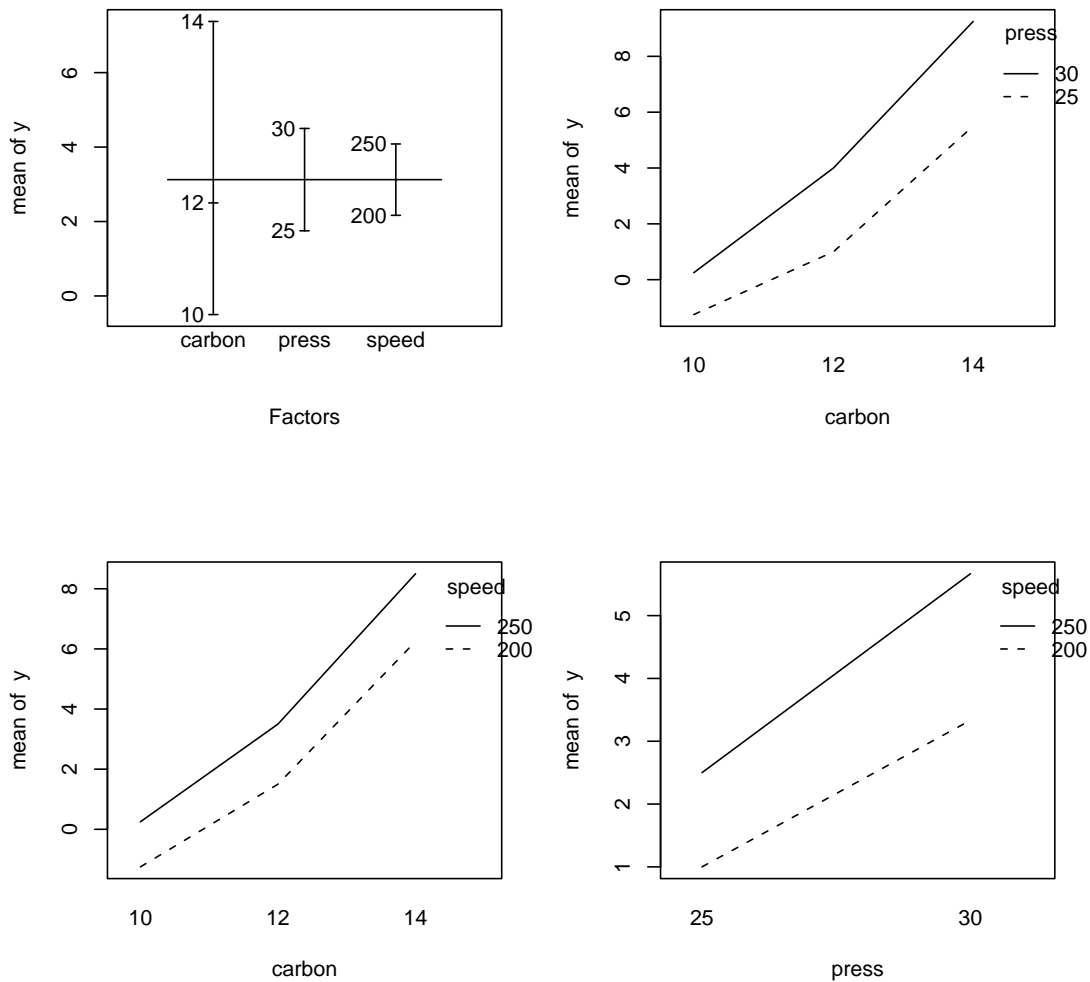


Fig. 5.5.

Full 3 factor model:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl}.$$

```
> g <- lm(y ~ carbon + press + speed + carbon*press
+ carbon*speed + press*speed + carbon*press*speed)
> anova(g)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
C	2	252.750	126.375	178.4118	1.186e-09
P	1	45.375	45.375	64.0588	3.742e-06
S	1	22.042	22.042	31.1176	0.0001202
C:P	2	5.250	2.625	3.7059	0.0558081
C:S	2	0.583	0.292	0.4118	0.6714939
P:S	1	1.042	1.042	1.4706	0.2485867
C:P:S	2	1.083	0.542	0.7647	0.4868711
Resid	12	8.500	0.708		

It seems that interactions are largely absent, and that all three main effects are significant. In particular, the low level of pressure results in smaller mean deviations from the target. A CI on  $\beta_2 - \beta_1 = E[\bar{y}_{.2.} - \bar{y}_{.1.}]$  is ( $\alpha = .05$ )

$$\begin{aligned} & \bar{y}_{.2.} - \bar{y}_{.1.} \pm t_{\alpha/2, 12} \sqrt{MS_E \left( \frac{1}{12} + \frac{1}{12} \right)} \\ &= 1.75 - 4.5 \pm 2.1788 \sqrt{\frac{.708}{6}} \\ &= -2.75 \pm .75 \end{aligned}$$

or  $[-3.5, -2]$ .

## Part VI

# THE $2^k$ FACTORIAL DESIGN

17.  $2^2$  factorials

- We'll start with a basic  $2^2$  design, where it is easy to see what is going on. Also, these are very widely used in industrial experiments.
- Two factors (A and B), each at 2 levels - low ('-') and high ('+'). # of replicates =  $n$ .
- Example - investigate yield ( $y$ ) of a chemical process when the concentration of a reactant (the primary substance producing the yield) - factor A - and amount of a catalyst (to speed up the reaction) - factor B - are changed. E.g. nickel is used as a 'catalyst', or a carrier of hydrogen in the hydrogenation of oils (the reactants) for use in the manufacture of margarine.

Factor		$n = 3$ replicates			Total	Label
A	B	I	II	III		
-	-	28	25	27	80 =	(1)
+	-	36	32	32	100 =	$a$
-	+	18	19	23	60 =	$b$
+	+	31	30	29	90 =	$ab$

- Notation

(1) = sum of obs'ns at low levels of both factors,

$a$  = sum of obs'ns with A high and B low,

$b$  = sum of obs'ns with B high and A low,

$ab$  = sum of obs'ns with both high.

- Effects model. Use a more suggestive notation:

$$y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ijk} \quad (i, j = 1, 2, k = 1, \dots, n)$$

- E.g.  $A_1$  = main effect of low level of A,  $A_2$  = main effect of high level of A. But since  $A_1 + A_2 = 0$ , we have  $A_1 = -A_2$ .

- We define the 'main effect of Factor A' to be

$$A = A_2 - A_1.$$

- What is the LSE of  $A$ ? Since  $A$  is the effect of changing factor  $A$  from high to low, we expect

$$\begin{aligned}\hat{A} &= \text{average } y \text{ at high } A - \text{average } y \text{ at low } A \\ &= \frac{a + ab}{2n} - \frac{(1) + b}{2n} \\ &= \frac{a + ab - (1) - b}{2n}.\end{aligned}$$

This is the LSE.

Reason: We know that the LSE of  $A_2$  is

$$\hat{A}_2 = \text{average } y \text{ at high } A - \text{overall average } y,$$

and that of  $A_1$  is

$$\hat{A}_1 = \text{average } y \text{ at low } A - \text{overall average } y,$$

so that

$$\begin{aligned}\hat{A} &= \hat{A}_2 - \hat{A}_1 \\ &= \text{average } y \text{ at high } A - \text{average } y \text{ at low } A.\end{aligned}$$

- Often the 'hats' are omitted (as in the text). Similarly,

$$\begin{aligned}
 B &= \frac{b + ab - a - (1)}{2n} \\
 AB &= \text{difference between effect of A at high B,} \\
 &\quad \text{and effect of A at low B} \\
 &= \frac{ab - b}{2n} - \frac{a - (1)}{2n} \\
 &= \frac{ab - b - a + (1)}{2n}.
 \end{aligned}$$

With  $(1) = 80$ ,  $a = 100$ ,  $b = 60$ ,  $ab = 90$  we find

$$\begin{aligned}
 A &= 8.33, \\
 B &= -5.0 \\
 AB &= 1.67.
 \end{aligned}$$

- It appears that increasing the level of A results in an increase in yield; that the opposite is true of B, and that there isn't much interaction effect. To confirm this we would do an ANOVA.



```
> A <- c(-1, 1, -1,1)
> B <- c(-1, -1, 1, 1)
> I <- c(28, 36, 18, 31)
> II <- c(25, 32, 19, 30)
> III <- c(27, 32, 23, 29)
>
> data <- data.frame(A, B, I, II, III)
> data
  A  B  I  II  III
1 -1 -1 28 25  27
2  1 -1 26 32  32
3 -1  1 18 19  23
4  1  1 31 30  29

# compute sums for each combination
> sums <- apply(data[,3:5], 1, sum)
> names(sums) <- c("(1)", "(a)", "(b)", "(ab)")
> sums
(1)  (a)  (b) (ab)
 80 100  60  90
```

```

# Interaction plots
> ybar <- sums/3
> par(mfrow=c(1,2))
> interaction.plot(A, B, ybar)
> interaction.plot(B, A, ybar)

# Build ANOVA table

> y <- c(I, II, III)
> factorA <- as.factor(rep(A,3))
> factorB <- as.factor(rep(B,3))
> g <- lm(y ~factorA + factorB + factorA*factorB)
> anova(g)

```

### Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factorA	1	208.333	208.333	53.1915	8.444e-05
factorB	1	75.000	75.000	19.1489	0.002362
AB	1	8.333	8.333	2.1277	0.182776
Residuals	8	31.333	3.917		

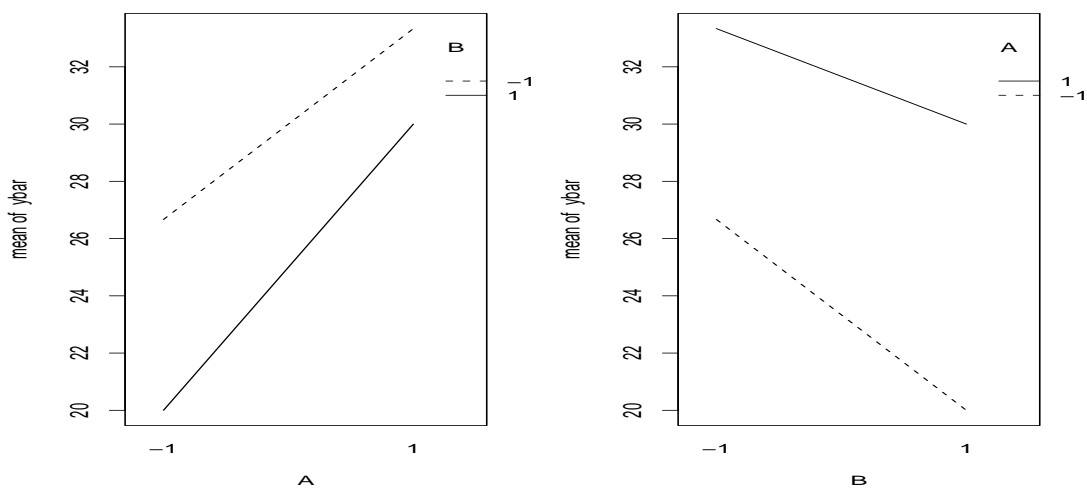


Fig. 6.1

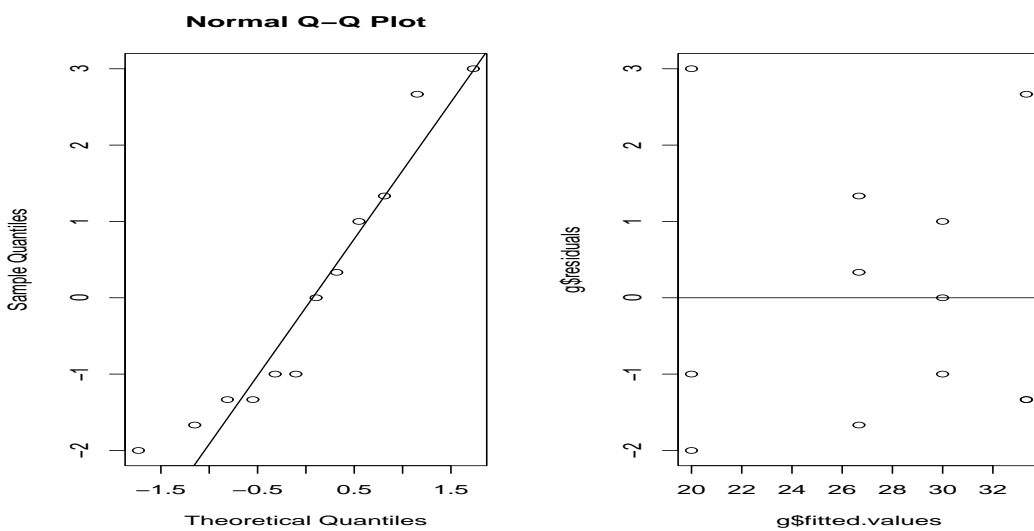


Fig. 6.2

**Contrasts.** The estimates of the effects have used only the terms  $ab, a, b$  and  $(1)$ , each of which is the sum of  $n = 3$  independent terms. Then

$$\begin{aligned} A &= \frac{ab + a - b - (1)}{2n} = \frac{C_A}{2n}, \\ B &= \frac{ab - a + b - (1)}{2n} = \frac{C_B}{2n}, \\ AB &= \frac{ab - a - b + (1)}{2n} = \frac{C_{AB}}{2n}, \end{aligned}$$

where  $C_A, C_B, C_{AB}$  are orthogonal contrasts (why?) in  $ab, a, b$  and  $(1)$ . In our previous notation, the SS for Factor A (we might have written it as  $bn \sum \hat{A}_i^2$ ) is

$$SS_A = 2n (\hat{A}_1^2 + \hat{A}_2^2) = 4n \hat{A}_2^2 = nA^2 = \frac{C_A^2}{4n},$$

and similarly

$$SS_B = \frac{C_B^2}{4n}, \quad SS_{AB} = \frac{C_{AB}^2}{4n}.$$

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}.$$

In this way  $SS_A = [90 + 100 - 60 - 80]^2 / 12 = 208.33$ .

18.  $2^k$  factorials

- All of this generalizes to the  $2^k$  factorial, in which  $k$  factors are investigated, each at two levels. To easily write down the estimates of the effects, and the contrasts, we start with a table of  $\pm$  signs, done here for  $k = 3$ . Label the rows (1), then the product of  $a$  with (1). Then all products of  $b$  with the terms which are already there:  $b \times (1) = b, b \times a = ab$ . Then all products of  $c$  with the terms which are already there. (This is the 'standard' order.) Now put in the signs. Start with  $2^k = 8$  +'s under the I, then alternate -'s and +'s, then in groups of 2, finally (under C) in groups of 4 ( $= 2^{k-1}$ ). Then write in the products under the interaction terms.

	Effect							
	I	A	B	C	AB	AC	BC	ABC
(1)	+	-	-	-	+	+	+	-
<i>a</i>	+	+	-	-				+
<i>b</i>	+	-	+	-				+
<i>ab</i>	+	+	+	-	+	-	-	-
<i>c</i>	+	-	-	+				+
<i>ac</i>	+	+	-	+				-
<i>bc</i>	+	-	+	+				-
<i>abc</i>	+	+	+	+				+

- Interpretation: Assign the appropriate signs to the combinations (1), ..., *abc*. Effect estimates are

$$A = \frac{[(1) + b + c + bc]}{4n} + \frac{[a + ab + ac + abc]}{4n},$$

*etc.*

$$ABC = \frac{[a + b + c + abc] - [(1) + ab + ac + bc]}{4n},$$

all with  $2^{k-1}n$  in the denominator.

- These are all of the form  $C / (2^{k-1}n)$  for a contrast in the sums  $(1), \dots, abc$ ; the corresponding  $SS$  is  $C^2 / (2^k n)$ . For example

$$SS_{ABC} = \frac{\left\{ \begin{array}{l} [a + b + c + abc] \\ - [(1) + ab + ac + bc] \end{array} \right\}^2}{8n}.$$

- The sums of squares are all on 1 d.f. (including  $SS_I$ , which uses the 1 d.f. usually subtracted from  $N = 2^k n$  for the estimation of the overall mean  $\mu$ ), so that  $SS_E$ , obtained by subtraction, is on  $N - 2^k = 2^k (n - 1)$  d.f. Then, e.g., the F-ratio to test the effect of factor A is

$$F_0 = \frac{MS_A}{MS_E},$$

where  $MS_A = SS_A$  and  $MS_E = SS_E / df(SS_E)$ .

The p-value for the hypothesis of no effect is

$$P\left(F_{2^k(n-1)}^1 > F_0\right).$$

- Suppose  $n = 1$ , so that no d.f. are available for the estimation of  $\sigma^2$ . In the  $2^2$  there was Tukey's test for non-additivity, which relied on the assumption that the interactions were of a certain mathematically simple but statistically dubious form (even more so for  $k > 2$ ). A more common remedy is to not even try to estimate certain effects - usually higher order interactions - and use the d.f. released in this way to estimate error.
- A graphical way of identifying the important effects which must be in the model, and those which can be dropped to facilitate error estimation, is a normal probability plot of the absolute values of the effect estimates - a 'half-normal' plot. Those effects which deviate significantly from the qqline tend to be the important ones.
- Example. Data in Table 6-10. A chemical product is produced using two levels each of temperature (A), pressure (B), concentration of formaldehyde (C) and rate (D) at which the product is stirred. Response (Y) is the 'filtration rate'.



	A	B	C	D	y
1	-1	-1	-1	-1	45
2	1	-1	-1	-1	71
3	-1	1	-1	-1	48
4	1	1	-1	-1	65
5	-1	-1	1	-1	68
6	1	-1	1	-1	60
7	-1	1	1	-1	80
8	1	1	1	-1	65
9	-1	-1	-1	1	43
10	1	-1	-1	1	100
11	-1	1	-1	1	45
12	1	1	-1	1	104
13	-1	-1	1	1	75
14	1	-1	1	1	86
15	-1	1	1	1	70
16	1	1	1	1	96

```

> g <- lm(y ~ (A+B+C+D)^4)
# Easier than, but equivalent to,
# y ~ A + B + ... + B*C*D + A*B*C*D
> anova(g)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	1870.56	1870.56		
B	1	39.06	39.06		
C	1	390.06	390.06		
D	1	855.56	855.56		
A:B	1	0.06	0.06		
A:C	1	1314.06	1314.06		
A:D	1	1105.56	1105.56		
B:C	1	22.56	22.56		
B:D	1	0.56	0.56		
C:D	1	5.06	5.06		
A:B:C	1	14.06	14.06		
A:B:D	1	68.06	68.06		
A:C:D	1	10.56	10.56		
B:C:D	1	27.56	27.56		
A:B:C:D	1	7.56	7.56		
Residuals	0	0.00			

```
> g$effects
```

# Note that these are twice as large in absolute value as those in the text, and the signs sometimes differ. This is because of R's definition of 'effect', and makes no difference for comparing their absolute values.

```
(Intercept)          A1          B1          C1
      -280.25       -43.25       -6.25       19.75
                etc.
      A1:B1:D1      A1:C1:D1      B1:C1:D1  A1:B1:C1:D1
      -8.25         3.25         5.25         2.75
```

```
> effects <- abs(g$effects[-1])
```

```
> qq <- qqnorm(effects, type="n")
```

```
> text(qq$x, qq$y, labels = names(effects))
```

"n" means no plotting - I only wanted the names to appear

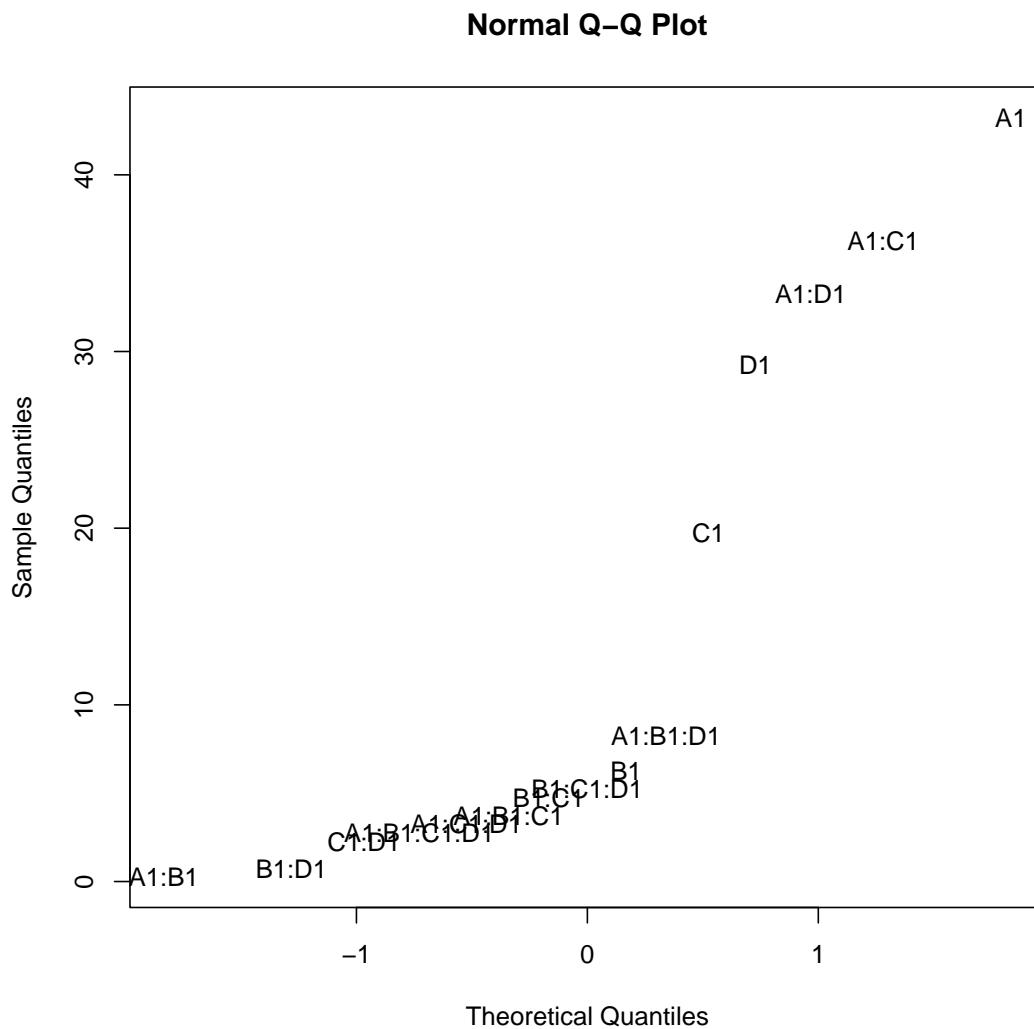


Fig. 6.3

The significant terms seems to be A, C, D and the interactions AC, AD. So let's just drop B and fit all terms not involving B.

```
> h <- lm(y ~ (A+C+D)^3)
```

```
> anova(h)
```

### Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
A	1	1870.56	1870.56	83.3677	1.667e-05	***
C	1	390.06	390.06	17.3844	0.0031244	**
D	1	855.56	855.56	38.1309	0.0002666	***
A:C	1	1314.06	1314.06	58.5655	6.001e-05	***
A:D	1	1105.56	1105.56	49.2730	0.0001105	***
C:D	1	5.06	5.06	0.2256	0.6474830	
A:C:D	1	10.56	10.56	0.4708	0.5120321	
Residuals	8	179.50	22.44			

Note that this is now a  $2^3$  factorial with  $n = 2$  replicates.

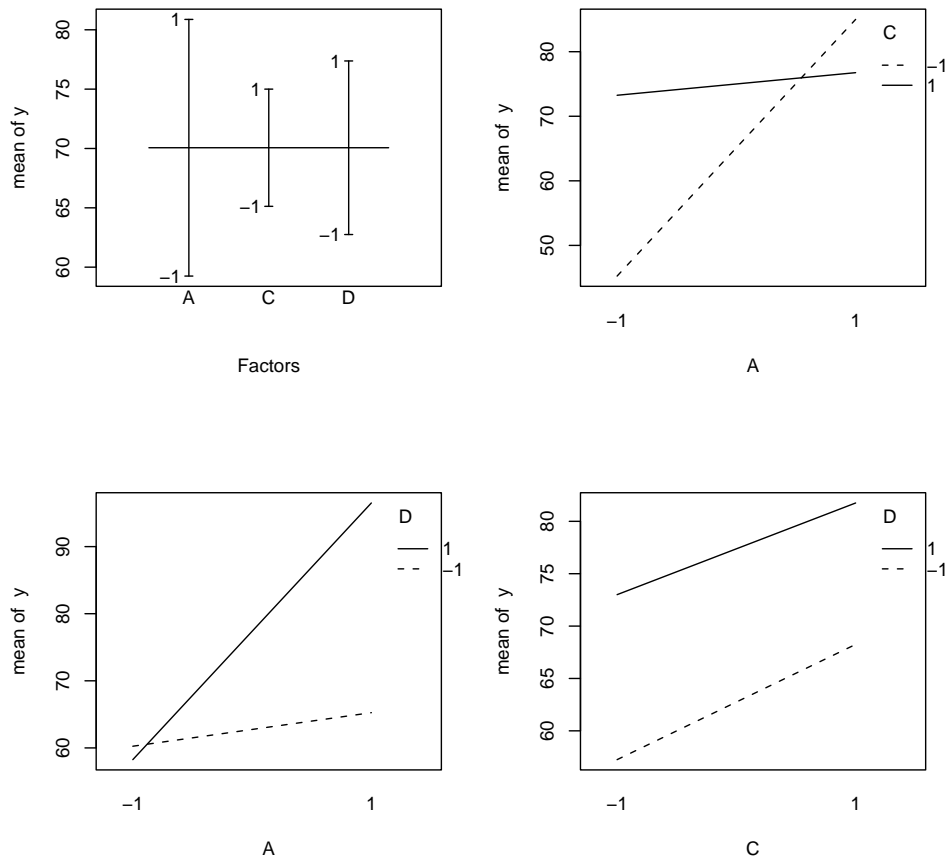


Fig. 6.4

Although the main effects plot indicates that C high is best, the interaction plots show that the best settings are A high, C low and D high.

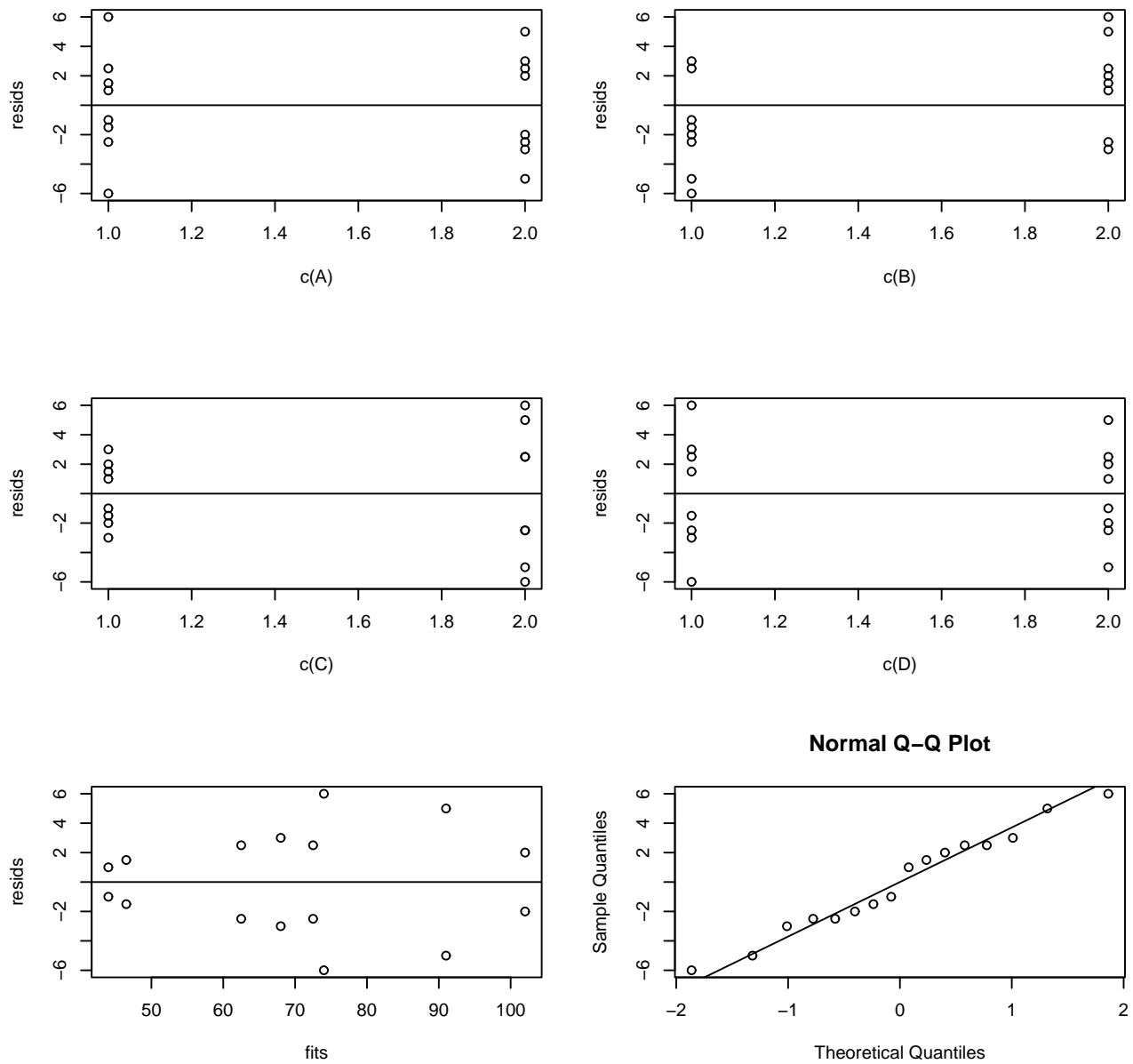


Fig. 6.5. Residual plots. Note that we plot against omitted factors (B), and anything else which is available (e.g. run order).





## **Part VII**

# **BLOCKING AND CONFOUNDING**

## 19. Blocking & confounding

- Importance of blocking to control nuisance factors  
- day of week, batch of raw material, etc.
- **Complete Blocks.** This is the easy case. Suppose we run a  $2^2$  factorial experiment, with all 4 runs made on each of 3 days. So there are 3 replicates (= blocks), 12 observations. There is 1 d.f. for each of I, A, B, AB, leaving 8 d.f. Of these, 2 are used for blocks and the remaining 6 for  $SS_E$ . The LSEs of the block effects are

$\widehat{Bl}_i =$  average of 4 obs'ns in block  $i$  - overall average

and

$$SS_{Blocks} = \sum_{\text{all obs'ns}} (\widehat{Bl}_i)^2 = 4 \sum_{i=1}^3 (\widehat{Bl}_i)^2.$$

Note the randomization used here - it is only within each block. If we could run the blocks in random order, for instance if they were batches of raw material, then we would also do so.

- Example - chemical experiment from lecture 17, with 'Replicates' re-labelled as 'Blocks'.

Factor		Block			Total	Label
<i>A</i>	<i>B</i>	<i>I</i>	<i>II</i>	<i>III</i>		
–	–	28	25	27	80	(1)
+	–	36	32	32	100	<i>a</i>
–	+	18	19	23	60	<i>b</i>
+	+	31	30	29	90	<i>ab</i>
averages		28.25	26.5	27.75	27.5	
$\widehat{Bl}$		.75	–1	.25		

$$SS_{Blocks} = 4 \left( (.75)^2 + (-1)^2 + (.25)^2 \right) = 6.5.$$

- Check on R.

- **Incomplete blocks.** Consider again a  $2^2$  factorial, in which only 2 runs can be made in each of 2 days (the blocks). Which 2 runs? Consider

Block 1: (1),  $ab$

Block 2:  $a, b$ .

What is the LSE of the block effect? Think of the blocks as being at a 'high' level - Block 1 - and a 'low' level - Block 2. Then the estimate is

$$\begin{aligned}
 Bl &= \text{average at high level} - \text{average at low level} \\
 &= \frac{ab + (1) - a - b}{2} \\
 &= \frac{[ab - a] - [b - (1)]}{2} \\
 &= \text{effect of B when A is high} \\
 &\quad - \text{effect of B when A is low} \\
 &= AB.
 \end{aligned}$$

We say that AB is confounded with blocks.

- The confounding can be seen from the table of  $\pm$  signs. All runs in which  $AB = +$  are in block 1, all with  $AB = -$  are in block 2.

	Effect				Block
	I	A	B	AB	
(1)	+	-	-	+	1
<i>a</i>	+	+	-	-	2
<i>b</i>	+	-	+	-	2
<i>ab</i>	+	+	+	+	1

- This is fine if we feel that interactions are not an issue, and don't want to estimate them. But what effect is confounded with blocks in this design?

	Effect				Block
	I	A	B	AB	
(1)	+	-	-	+	1
<i>a</i>	+	+	-	-	2
<i>b</i>	+	-	+	-	1
<i>ab</i>	+	+	+	+	2

- As in the last example, we can choose which effect is to be confounded with the two blocks. For instance in a  $2^3$  design, with 4 runs in each of 2 blocks, we confound ABC with blocks in the following way:

	Effect								Blocks
	I	A	B	C	AB	AC	BC	ABC	
(1)	+	-	-	-	+	+	+	-	1
<i>a</i>	+	+	-	-				+	2
<i>b</i>	+	-	+	-				+	2
<i>ab</i>	+	+	+	-	+	-	-	-	1
<i>c</i>	+	-	-	+				+	2
<i>ac</i>	+	+	-	+				-	1
<i>bc</i>	+	-	+	+				-	1
<i>abc</i>	+	+	+	+				+	2

- Can you guess at a way to do this over 4 days with 2 runs per day?

20.  $2^k$  factorials in 2 blocks

- For running a  $2^k$  factorial in 2, 4, 8, 16, ... blocks, there are useful algebraic methods to decide on a confounding scheme. For 2 blocks, start with a 'defining contrast'

$$L = \alpha_1 x_1 + \cdots + \alpha_k x_k$$

where

$$x_i = \begin{cases} 1, & \text{if factor } i \text{ is high,} \\ 0, & \text{if factor } i \text{ is low,} \end{cases}$$

and  $\alpha_i$  is the exponent of factor  $i$  in the effect to be confounded. E.g.  $2^3$  factorial with  $ABC = A^1 B^1 C^1$  confounded with blocks has  $\alpha_1 = \alpha_2 = \alpha_3 = 1$  and  $L = x_1 + x_2 + x_3$ . If  $AB = A^1 B^1 C^0$  is to be confounded, then  $\alpha_1 = \alpha_2 = 1, \alpha_3 = 0$ ,  $L = x_1 + x_2$ . Now evaluate  $L$  at all treatment combinations, using 'arithmetic mod 2':

$x \pmod{2} =$  remainder when  $x$  is divided by 2.

- All those with  $L = 0 \pmod{2}$  go in one block, those with  $L = 1 \pmod{2}$  in another.

With ABC confounded,  $L = x_1 + x_2 + x_3 \pmod{2}$ :

$$(1) : L = 0 + 0 + 0 = 0 \pmod{2}$$

$$a : L = 1 + 0 + 0 = 1 \pmod{2}$$

$$b : L = 0 + 1 + 0 = 1 \pmod{2}$$

$$ab : L = 1 + 1 + 0 = 0 \pmod{2}$$

$$c : L = 0 + 0 + 1 = 1 \pmod{2}$$

$$ac : L = 1 + 0 + 1 = 0 \pmod{2}$$

$$bc : L = 0 + 1 + 1 = 0 \pmod{2}$$

$$abc : L = 1 + 1 + 1 = 1 \pmod{2}$$

Thus

Block 1 : (1),  $ab$ ,  $ac$ ,  $bc$

Block 2 :  $a$ ,  $b$ ,  $c$ ,  $abc$ ,

in agreement with what we got using the  $\pm$  signs.

- For 4 ( $= 2^p$ ) blocks we would pick two ( $= p$ ) effects to be confounded, and get two contrasts  $L_1$  and  $L_2$ . The combinations  $(L_1, L_2) = (0, 0), (0, 1), (1, 0), (1, 1) \pmod{2}$  define the 4 blocks. More on this later.



- Example. Run the  $2^4$  factorial for the filtration rate experiment in 2 blocks, corresponding to different batches of formaldehyde. Confound the ABCD interaction with blocks, so that  $L = x_1 + x_2 + x_3 + x_4 \pmod{2}$  will be 0 if  $\{x_1, x_2, x_3, x_4\}$  contains 0, 2 or 4 ones: (1),  $ab, ac, ad, bc, bd, cd, abcd$  and 1 otherwise:

Block 1: (1),  $ab, ac, ad, bc, bd, cd, abcd,$

Block 2:  $a, b, c, d, abc, abd, acd, bcd.$

The data have been modified by subtracting 20 from all Block 1 observations, to simulate a situation where the first batch of formaldehyde is inferior.

	A	B	C	D	ABCD	y	Block
1	-1	-1	-1	-1	1	25	1
2	1	-1	-1	-1	-1	71	2
3	-1	1	-1	-1	-1	48	2
4	1	1	-1	-1	1	45	1
5	-1	-1	1	-1	-1	68	2
6	1	-1	1	-1	1	40	1
7	-1	1	1	-1	1	60	1
8	1	1	1	-1	-1	65	2
9	-1	-1	-1	1	-1	43	2
10	1	-1	-1	1	1	80	1
11	-1	1	-1	1	1	25	1
12	1	1	-1	1	-1	104	2
13	-1	-1	1	1	1	55	1
14	1	-1	1	1	-1	86	2
15	-1	1	1	1	-1	70	2
16	1	1	1	1	1	76	1

Note one block has  $ABCD = 1$ , the other  $ABCD = -1$ .

**In all output, 'ABCD' is to be interpreted as 'Blocks'**

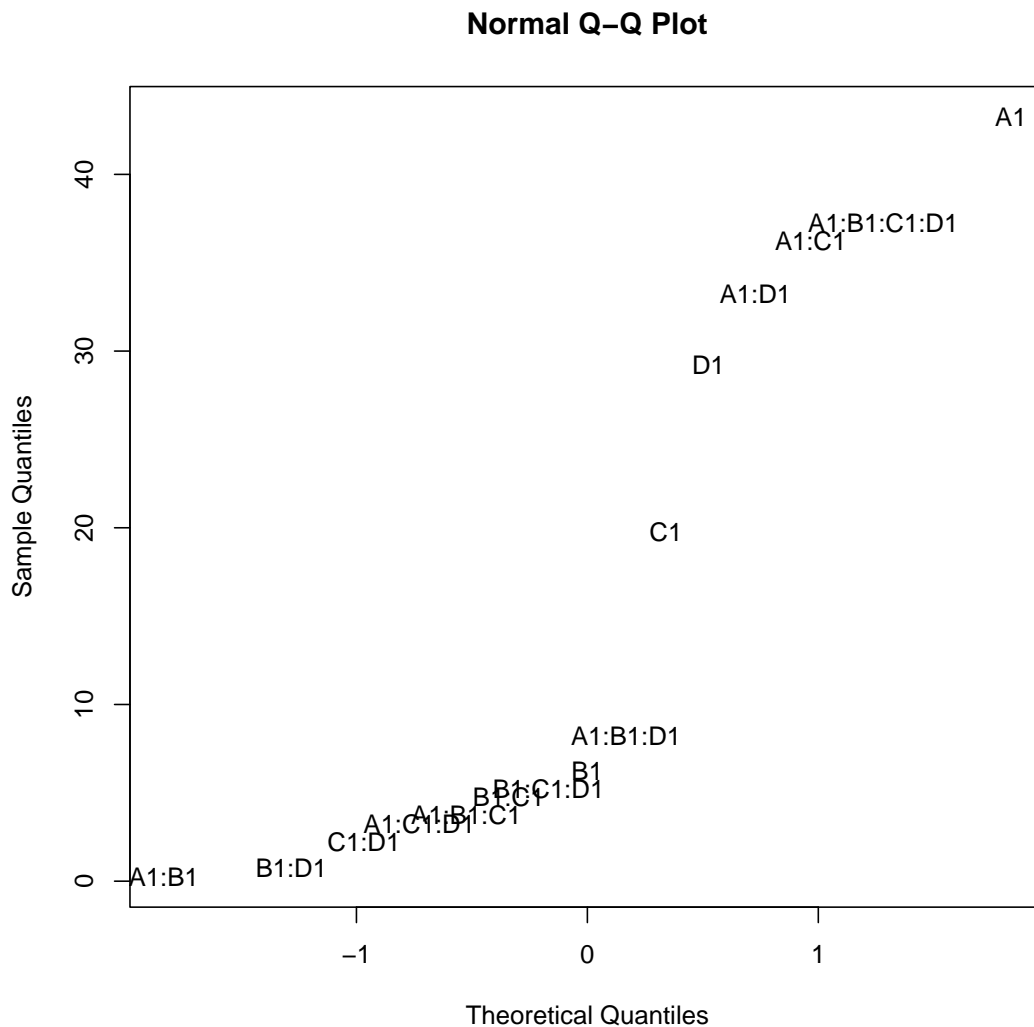
```
> g <- lm(y ~ (A+B+C+D)^4)
```

```
> anova(g)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	1870.56	1870.56		
B	1	39.06	39.06		
C	1	390.06	390.06		
D	1	855.56	855.56		
A:B	1	0.06	0.06		
A:C	1	1314.06	1314.06		
A:D	1	1105.56	1105.56		
B:C	1	22.56	22.56		
B:D	1	0.56	0.56		
C:D	1	5.06	5.06		
A:B:C	1	14.06	14.06		
A:B:D	1	68.06	68.06		
A:C:D	1	10.56	10.56		
B:C:D	1	27.56	27.56		
A:B:C:D	1	1387.56	1387.56		
Residuals	0	0.00			



**Fig. 7.1.** Half normal plot. Significant effects are  $A$ ,  $C$ ,  $D$ ,  $AC$ ,  $AD$  and Blocks ( $= ABCD$ ). We can run the ANOVA again, estimating only these effects.

```
> Blocks <- ABCD
> h <- lm(y ~A + C + D + A*C + A*D + Blocks)
> anova(h)
```

### Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
A	1	1870.56	1870.56	89.757	5.600e-06	***
C	1	390.06	390.06	18.717	0.0019155	**
D	1	855.56	855.56	41.053	0.0001242	***
Blocks	1	1387.56	1387.56	66.581	1.889e-05	***
A:C	1	1314.06	1314.06	63.054	2.349e-05	***
A:D	1	1105.56	1105.56	53.049	4.646e-05	***
Residuals	9	187.56	20.84			

## 21. $2^k$ factorials in $2^p$ blocks

Suppose we needed four batches of formaldehyde, and could do only 4 runs per batch. This is then a  $2^4$  factorial in  $2^2$  blocks.

- Some more algebra: If two effects are confounded with blocks, then so is their product, which is defined by 'multiplication mod 2':  $A^0 = A^2 = 1$ ,  $A^1 = A$  E.g.  $AB * BC = AB^2C = AC$ .
- Pick two effects to be confounded with blocks:  $ABC$  and  $ACD$ . Then also  $ABC * ACD = BD$  is confounded. We wouldn't pick  $ABC$  and  $ABCD$ , since  $ABC * ABCD = D$ .

- For the choices  $ABC$  and  $ACD$  we have

$$L_1 = 1 \cdot x_1 + 1 \cdot x_2 + 1 \cdot x_3 + 0 \cdot x_4 = x_1 + x_2 + x_3,$$

$$L_2 = 1 \cdot x_1 + 1 \cdot x_2 + 1 \cdot x_3 + 1 \cdot x_4 = x_1 + x_3 + x_4,$$

with

	(1)	a	b	ab	c	ac	bc	abc
$L_1$	0	1	1	0	1	0	0	1
$L_2$	0	1	0	1	1	0	1	0
Block	I	IV	II	III	IV	I	III	II

	d	ad	bd	abd	cd	acd	bcd	abcd
$L_1$	0	1	1	0	1	0	0	1
$L_2$	1	0	1	0	0	1	0	1
Block	III	II	IV	I	II	III	I	IV

Block I : (1),  $ac$ ,  $abd$ ,  $bcd$

Block II :  $b$ ,  $abc$ ,  $ad$ ,  $cd$

Block III :  $ab$ ,  $bc$ ,  $d$ ,  $acd$

Block IV :  $a$ ,  $c$ ,  $bd$ ,  $abcd$

	A	B	C	D	blocks	ABC	ACD	BD	y
1	-1	-1	-1	-1	1	-1	-1	1	25
2	1	-1	-1	-1	4	1	1	1	71
3	-1	1	-1	-1	2	1	-1	-1	48
4	1	1	-1	-1	3	-1	1	-1	45
5	-1	-1	1	-1	4	1	1	1	68
6	1	-1	1	-1	1	-1	-1	1	40
7	-1	1	1	-1	3	-1	1	-1	60
8	1	1	1	-1	2	1	-1	-1	65
9	-1	-1	-1	1	3	-1	1	-1	43
10	1	-1	-1	1	2	1	-1	-1	80
11	-1	1	-1	1	4	1	1	1	25
12	1	1	-1	1	1	-1	-1	1	14
13	-1	-1	1	1	2	1	-1	-1	55
14	1	-1	1	1	3	-1	1	-1	86
15	-1	1	1	1	1	-1	-1	1	20
16	1	1	1	1	4	1	1	1	76



```
> g <- lm(y ~blocks + A + B + C + D + A*B + A*C +
          A*D + B*C + C*D + A*B*D + B*C*D + A*B*C*D)
```

```
> anova(g)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
blocks	3	3787.7	1262.6		
A	1	1105.6	1105.6		
B	1	826.6	826.6		
C	1	885.1	885.1		
D	1	33.1	33.1		
A:B	1	95.1	95.1		
A:C	1	1.6	1.6		
A:D	1	540.6	540.6		
B:C	1	217.6	217.6		
C:D	1	60.1	60.1		
A:B:D	1	3.1	3.1		
B:C:D	1	22.6	22.6		
A:B:C:D	1	5.1	5.1		
Residuals	0	0.0			

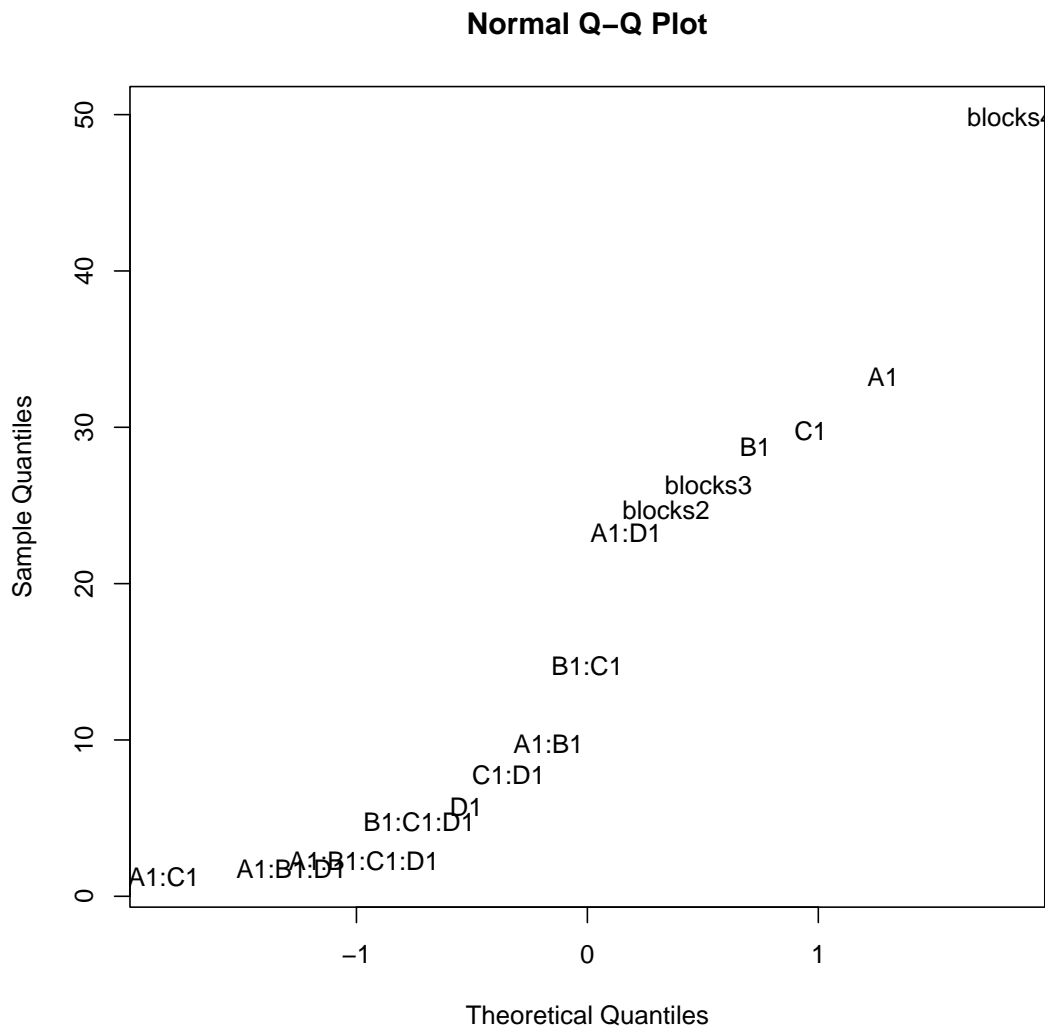


Fig. 7.2. Half normal plot for  $2^4$  factorial in  $2^2$  blocks. It looks like we can drop the main effect of 'D' if we keep some of its interactions.

R will, by default, estimate a main effect if an interaction is in the model. To fit blocks, A, B, C, AB, AD, BC, CD *but not D*, we can add the SS and df for D to those for Error.

```
> h <- lm(y ~blocks + A + B + C +
          B*C + A*B + A*D + C*D)
```

```
> anova(h)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
blocks	3	3787.7	1262.6	156.5969	0.0001333	***
A	1	1105.6	1105.6	137.1240	0.0003042	***
B	1	826.6	826.6	102.5194	0.0005356	***
C	1	885.1	885.1	109.7752	0.0004690	***
D	1	33.1	33.1	4.1008	0.1128484	
B:C	1	217.6	217.6	26.9845	0.0065401	**
A:B	1	95.1	95.1	11.7907	0.0264444	*
A:D	1	540.6	540.6	67.0465	0.0012117	**
C:D	1	60.1	60.1	7.4496	0.0524755	.
Residuals	4	32.3	8.1			

This would change  $MS_E$  to  $(32.3 + 33.1) / 5 = 13.08$  on 5 d.f. - not a helpful step (since  $MS_D$  was larger than  $MS_E$ ).

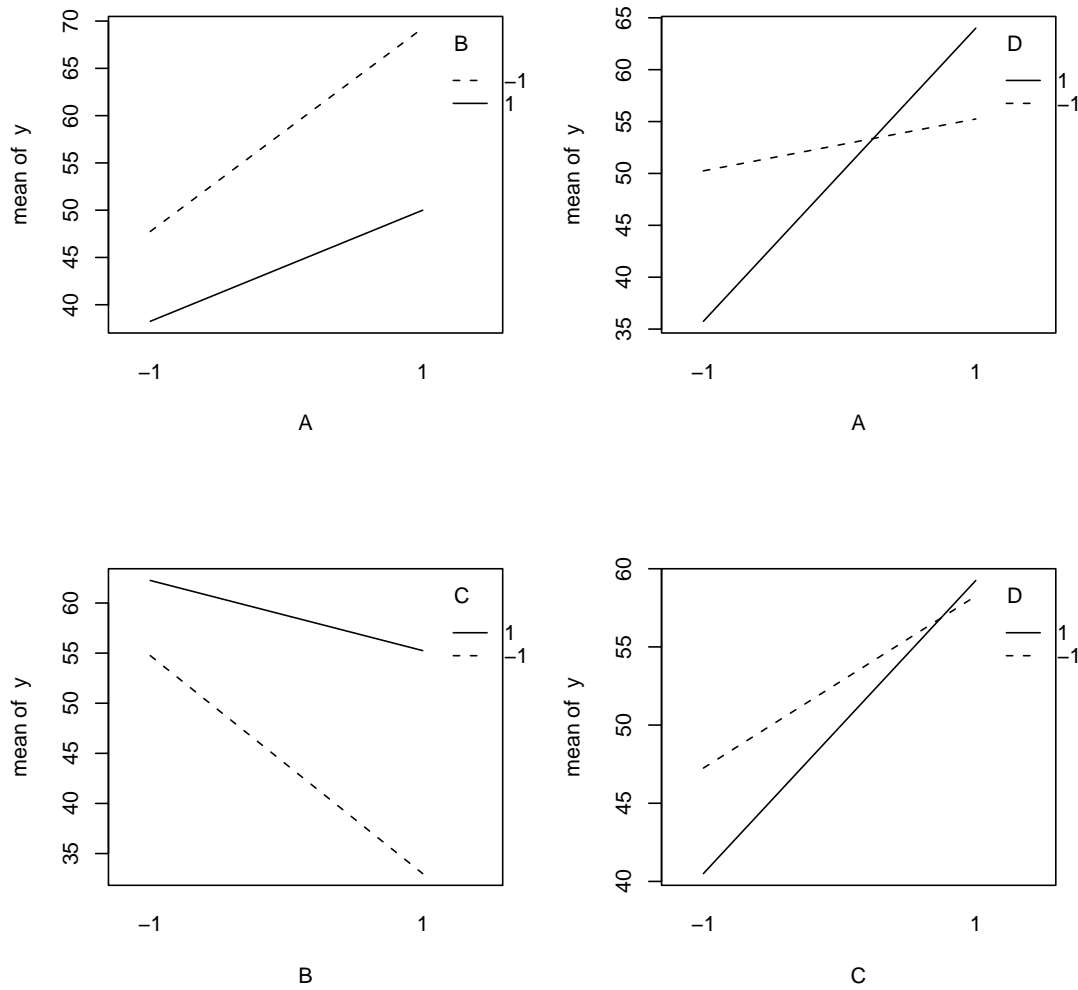


Fig. 7.3. Interaction plots. The best combination seems to be A, C, D high, B low.

## 22. Partial confounding

- To get an estimate of error, we have to either drop certain effects from the model, or replicate the design. If we replicate, we can either:
  - Confound the same effects with blocks in each replication - ‘complete’ confounding, or
  - Confound different effects with each replication - ‘partial’ confounding.
- Partial confounding is often better, since we then get estimates of effects from the replications in which they are *not* confounded.

- Example 7-3 from text. Two replicates of a  $2^3$  factorial are to be run, in 2 block each.
  - Replicate 1: Confound ABC with blocks. So  $L = x_1 + x_2 + x_3 = 0$  for (1),  $ab, ac, bc$  and  $= 1$  for  $a, b, c, abc$ .
  - Replicate 2: Confound AB with blocks. So  $L = x_1 + x_2 = 0$  for (1),  $ab, abc, c$  and  $= 1$  for  $a, b, ac, bc$ .

Rep 1		Rep2	
Block 1	Block 2	Block 3	Block 4
(1) = 550	$a = 669$	(1) = 604	$a = 650$
$ab = 642$	$b = 633$	$c = 1052$	$b = 601$
$ac = 749$	$c = 1037$	$ab = 635$	$ac = 868$
$bc = 1075$	$abc = 729$	$abc = 860$	$bc = 1063$

	A	B	C	Rep	Block	ABC	AB	y
1	-1	-1	-1	I	1	-1	1	550
2	1	1	-1	I	1	-1	1	642
3	1	-1	1	I	1	-1	-1	749
4	-1	1	1	I	1	-1	-1	1075
5	1	-1	-1	I	2	1	-1	669
6	-1	1	-1	I	2	1	-1	633
7	-1	-1	1	I	2	1	1	1037
8	1	1	1	I	2	1	1	729
9	-1	-1	-1	II	3	-1	1	604
10	-1	-1	1	II	3	1	1	1052
11	1	1	-1	II	3	-1	1	635
12	1	1	1	II	3	1	1	860
13	1	-1	-1	II	4	1	-1	650
14	-1	1	-1	II	4	1	-1	601
15	1	-1	1	II	4	-1	-1	868
16	-1	1	1	II	4	-1	-1	1063

When the levels of one factor (Blocks) make sense only within the levels of another factor (Replicates) we say that the first is 'nested' within the second. A way to indicate this in R is as:

```
> h <- lm(y ~Rep + Block%in%Rep + A + B + C
          + A*B + A*C + B*C + A*B*C)
```

```
> anova(h)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Rep	1	3875	3875	1.5191	0.272551	
A	1	41311	41311	16.1941	0.010079	*
B	1	218	218	0.0853	0.781987	
C	1	374850	374850	146.9446	6.75e-05	***
Rep:Block	2	458	229	0.0898	0.915560	
A:B	1	3528	3528	1.3830	0.292529	
A:C	1	94403	94403	37.0066	0.001736	**
B:C	1	18	18	0.0071	0.936205	
A:B:C	1	6	6	0.0024	0.962816	
Residuals	5	12755	2551			

Through the partial confounding we are able to estimate all interactions. It looks like only A, C, and AC are significant.



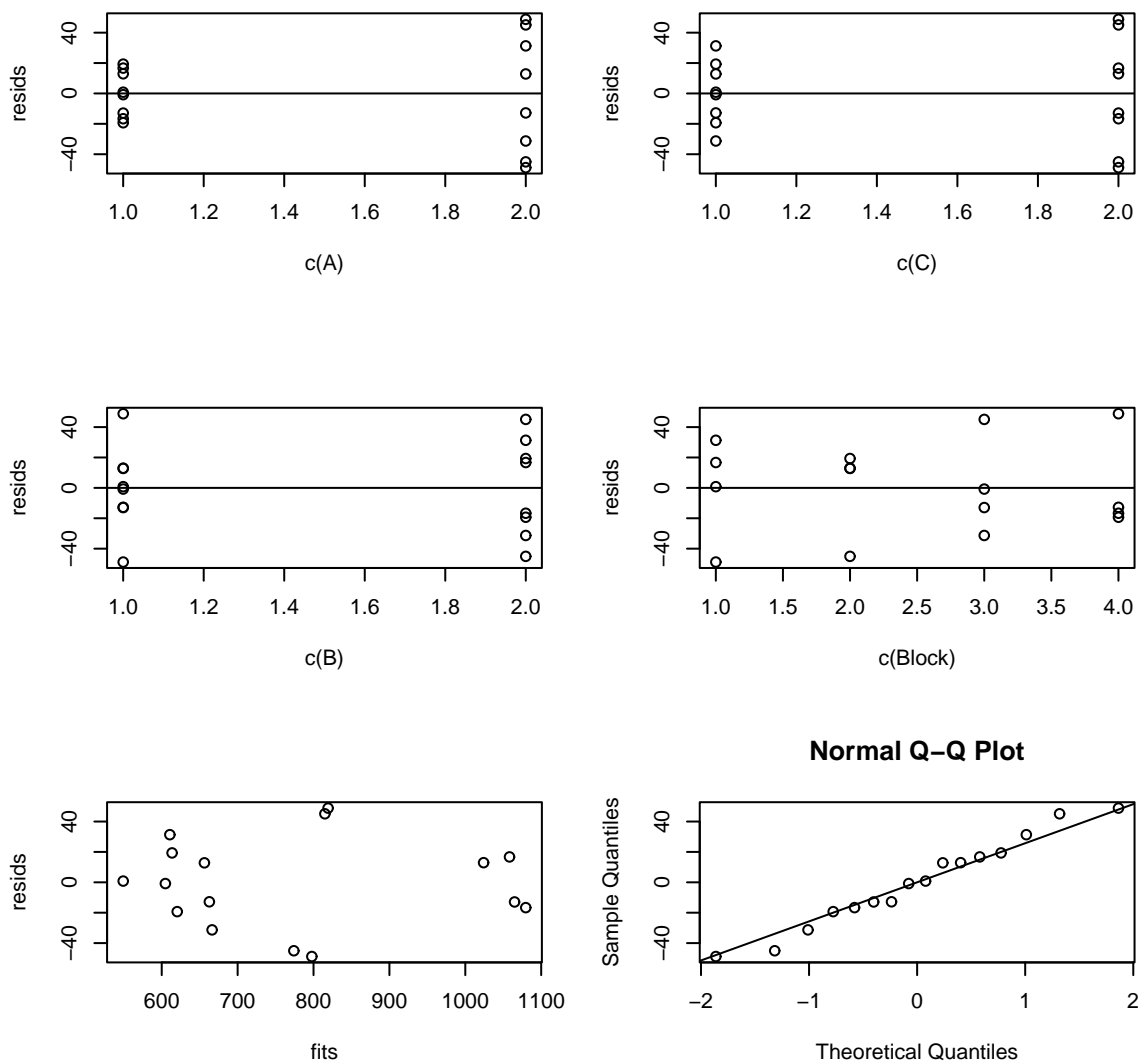


Fig. 7.4. Residuals for fit to A, C, and AC only.

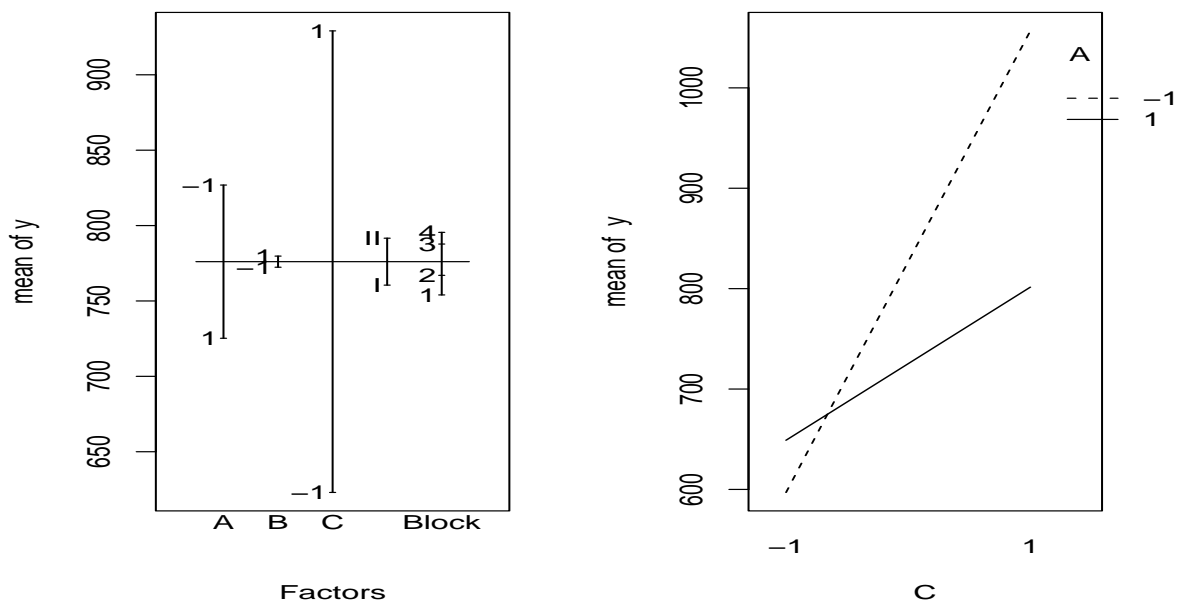


Fig. 7.5. Design and interactions.

- How is  $SS_{Blocks(Rep)}$  computed? One way is to compute  $SS_{ABC}$  in Rep I, where this effect is confounded with blocks, and similarly  $SS_{AB}$  in

Rep II, and add them:

$$SS_{ABC.in.I} = \frac{\left[ \begin{array}{l} -(550 + \dots + 1075) \\ + (669 + \dots + 729) \end{array} \right]^2}{8}$$

$$= 338,$$

$$SS_{AB.in.II} = \dots = 120.125,$$

$$SS_{Blocks(Rep)} = 338 + 120.125 = 458.125,$$

in agreement with the ANOVA output. See the programme on the course web site to see how to do this calculation very easily.

- Another method goes back to general principles. We calculate a SS for blocks within each replicate (since blocks make sense only within the replicates):

$$SS_{Blocks(Rep)} = 4 \sum_{i=1,2} \sum_{j=1,2} (\bar{y}_{ij.} - \bar{y}_{i..})^2 = 458.125.$$

Here  $\bar{y}_{ij.}$  is the average in block  $j$  of replicate  $i$ , and  $\bar{y}_{i..}$  is the overall average **of that replicate**, which is the only one in which that block makes sense. See the R calculation.

## **Part VIII**

# **FRACTIONAL FACTORIAL DESIGNS**

## 23. Fractional factorials - introduction

- Consider a  $2^5$  factorial. Even without replicates, there are  $2^5 = 32$  obs'ns required to estimate the effects - 5 main effects, 10 two factor interactions, 10 three factor interactions, 5 four factor interactions and 1 five factor interaction. If three (or more) factor interactions are not of interest then only 15 effects are left so that (including 1 d.f. for  $\mu$ ) perhaps only half as many obs'ns are needed.
- A  $2^{k-1}$  design, or 'one-half fraction of the  $2^k$  design', is one in which only half of the  $2^k$  treatment combinations are observed.
- Example.  $2^3$  factorial. Recall that we could run this in two blocks, with ABC confounded with blocks, as follows:

	Effect								Blocks
	I	A	B	C	AB	AC	BC	ABC	
(1)	+	-	-	-	+	+	+	-	1
<i>a</i>	+	+	-	-			+	+	2
<i>b</i>	+	-	+	-			-	+	2
<i>ab</i>	+	+	+	-	-	-	-	-	1
<i>c</i>	+	-	-	+			-	+	2
<i>ac</i>	+	+	-	+			-	-	1
<i>bc</i>	+	-	+	+			+	-	1
<i>abc</i>	+	+	+	+			+	+	2

If we run only block 2, then the design uses  $\{a, b, c, abc\}$ . These are those for which  $ABC = +$ ; since also  $I = +$  we say that the 'defining relation' for the design is

$$I = ABC,$$

and we refer to the 'word'  $ABC$  as the 'generator' of the design.

If we used only those combinations with  $A = +$ , then  $A = I$  would be the defining relation and  $A$  the generating word.

- Our one-half fraction is

	Effect							
	I	A	B	C	AB	AC	BC	ABC
<i>a</i>	+	+	-	-	-	-	+	+
<i>b</i>	+	-	+	-	-	+	-	+
<i>c</i>	+	-	-	+	+	-	-	+
<i>abc</i>	+	+	+	+	+	+	+	+

and the estimates of the effects are obtained by applying the  $\pm$  signs appropriately. We use  $[\cdot]$ 's to distinguish these from the full factorial estimates.

$$\begin{aligned}
 [A] &= \frac{a - b - c + abc}{2} = [BC], \\
 [B] &= \frac{-a + b - c + abc}{2} = [AC], \\
 [C] &= ?? = [??].
 \end{aligned}$$

We say that these pairs of effects are 'aliases'.

- Note that in the full factorial,

$$A + BC = \frac{a - b - c + abc}{2},$$

so that  $[A]$  and  $[BC]$  are each estimating the same things as  $A + BC$ . We write

$$[A] \rightarrow A + BC,$$

$$[B] \rightarrow B + AC,$$

etc.

These relations can also be obtained by doing multiplication (mod 2) on the defining relation:

$$I = ABC \Rightarrow A = A^2BC = BC,$$

$$B = AB^2C = AC,$$

etc.

- The one-half fraction with defining relation  $ABC = I$  is called the 'principal fraction'. The other half, in which  $ABC = -$ , is called the 'alternate' or 'complementary' fraction, and has defining relation

$$I = -ABC.$$



Complementary fraction								
Effect								
	I	A	B	C	AB	AC	BC	ABC
(1)	+	-	-	-	+	+	+	-
<i>ab</i>	+	+	+	-	+	-	-	-
<i>ac</i>	+	+	-	+	-	-	-	-
<i>bc</i>	+	-	+	+	-	+	+	-

- The estimates from the complementary fraction are

$$[A]' = \frac{-(1) + ab + ac - bc}{2} = -[BC]',$$

$$[B]' = \frac{-(1) + ab - ac + bc}{2} = -[AC]',$$

$$[C]' = ?? = [??].$$

In the full factorial,

$$A - BC = \frac{-(1) + ab + ac - bc}{2},$$

so that

$$[A]' \rightarrow A - BC,$$

$$[B]' \rightarrow B - AC,$$

etc.

If we were to run one fraction, and then decide to run the other, we could view the entire experiment as a full  $2^3$  factorial run in two blocks, with  $ABC$  confounded with blocks. We could combine our estimates to get estimates of all main effects and two factor interactions:

$$\frac{[A] + [A]'}{2} = \frac{a - b - c + abc}{4} + \frac{-(1) + ab + ac - bc}{4} = A,$$

$$\frac{[A] - [A]'}{2} = \dots = BC,$$

etc.

- We probably wouldn't want to run one half of a  $2^3$  factorial unless we were confident that the two factor interactions were not significant.

- We call this half of a  $2^3$  factorial a Resolution III design, since there are 3 letters in the generating word. We write it as  $2_{III}^{3-1}$ . A half of a  $2^4$  factorial with defining relationship  $I = ABC$  would also be resolution III, and would have  $D = ABCD$  but  $A = BC$ , etc. If the defining relationship were  $I = ABCD$  (Resolution IV), then  $A = BCD$ ,  $AB = CD$ , etc. - main effects are aliased only with 3 factor interactions, 2 factor interactions with each other. Obviously, the best we can do by running half of a  $2^k$  design is Resolution K, with  $I = AB \cdots K$ .
- Another way to view the resolution is as the (minimum) total length of any aliased words (where  $I$  has length 0). If it is small then two short words could be aliased. The larger the resolution, the better.

## 24. Fractional factorials - examples

Recall the filtration rate example started in Lecture 18. A chemical product is produced using two levels each of temperature (A), pressure (B), concentration of formaldehyde (C) and rate (D) at which the product is stirred. Response (Y) is the 'filtration rate'. Suppose we decide to investigate the effects of these factors by running half of a  $2^4$  factorial. We attain Resolution IV (the best possible) with the defining relation

$$I = ABCD.$$

Note that:

- (i) The defining relationship implies that  $D = ABC$ .
- (ii) The principal (or complementary) half of a  $2^k$  factorial is a full  $2^{k-1}$  factorial for  $k-1$  of the factors.

Thus we can get the design by writing down a full  $2^3$  factorial for A, B and C, and computing the signs for D from  $D = ABC$ . The resulting  $2_{IV}^{4-1}$  design, with simulated data, is

	Effect					
	<i>I</i>	<i>A</i>	<i>B</i>	<i>C</i>	$D = ABC$	<i>y</i>
(1)	+	-	-	-	-	45
<i>a</i>	+	+	-	-	+	100
<i>b</i>	+	-	+	-	+	45
<i>ab</i>	+	+	+	-	-	65
<i>c</i>	+	-	-	+	+	75
<i>ac</i>	+	+	-	+	-	60
<i>bc</i>	+	-	+	+	-	80
<i>abc</i>	+	+	+	+	+	96

The aliasing relationships are

$$A = BCD, B = ACD, C = ABD, D = ABC, \\ AB = CD, AC = BD, AD = BC.$$

Thus

$$[A] \rightarrow A + BCD, \\ [AB] \rightarrow AB + CD, \\ \text{etc.}$$

These estimates can be computed as

$$[A] = \frac{a + ab + ac + abc}{4} - \frac{(1) + b + c + bc}{4},$$

etc.

For the analysis, first (try to) fit the full  $2^4$  model:

```
> g <- lm(y ~ (A+B+C+D)^4)
```

```
> anova(g)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	722.0	722.0		
B	1	4.5	4.5		
C	1	392.0	392.0		
D	1	544.5	544.5		
A:B	1	2.0	2.0		
A:C	1	684.5	684.5		
A:D	1	722.0	722.0		
Residuals	0	0.0			

Only one member of each aliased pair is exhibited; by default it is the shortest word in the pair. From the ANOVA it looks like only A, C and D have significant main effects. Of course these could be caused by significant effects of BCD, ABD or ABC, but ... The half normal plot backs up these conclusions.

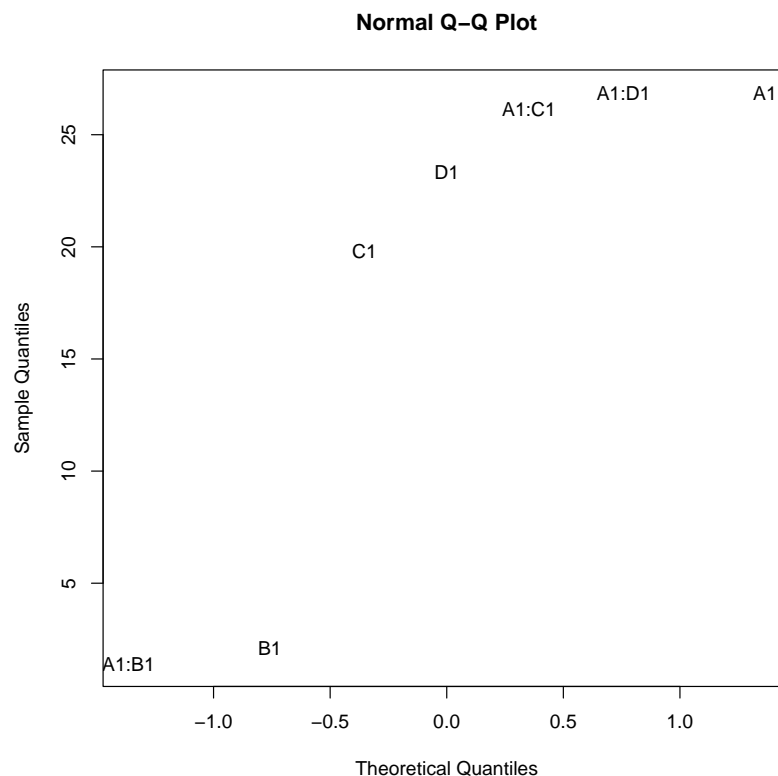


Fig. 8.1. Half normal plot for  $2_{IV}^{4-1}$  design in filtration example.

The half normal plot also points to  $AD (= BC)$  and  $AC (= BD)$  as significant. Which ones? Since  $B$  is not significant we wouldn't expect  $BC$  or  $BD$  to be significant either. We conclude that the factors of interest are  $A, C, D$  and the interactions  $AC, AD$ . If we drop  $B$  from the model then the table of  $\pm$  signs becomes

	Effect							
	$I$	$A$	$C$	$D$	$AC$	$AD$	$CD$	$y$
(1)	+	-	-	-	+	+	+	45
$a$	+	+	-	+	-	+	-	100
$b$	+	-	-	+	+	-	-	45
$ab$	+	+	-	-	-	-	+	65
$c$	+	-	+	+	-	-	+	75
$ac$	+	+	+	-	+	-	-	60
$bc$	+	-	+	-	-	+	-	80
$abc$	+	+	+	+	+	+	+	96

and we can estimate all main effects and two factor interactions without any being aliased.



```
> h <- lm(y ~ (A+C+D)^2)
```

```
> anova(h)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
A	1	722.0	722.0	160.4444	0.05016	.
C	1	392.0	392.0	87.1111	0.06795	.
D	1	544.5	544.5	121.0000	0.05772	.
A:C	1	684.5	684.5	152.1111	0.05151	.
A:D	1	722.0	722.0	160.4444	0.05016	.
C:D	1	2.0	2.0	0.4444	0.62567	
Residuals	1	4.5	4.5			

If the insignificant  $SS_{CD}$  were combined with  $SS_E$ , we would have  $MS_E = 3.25$  on 2 d.f., and all  $F_0$ 's would be  $4.5/3.25 = 1.38$  times as large. They would become

$$F_0 = (222.15, 120.61, 167.54, 210.62, 222.15)$$

respectively, with corresponding p-values

$$P(F_2^1 > F_0) = (0.004, 0.008, 0.006, 0.005, 0.004).$$

All this generalizes. Basic idea: To run a one-quarter fraction of a  $2^6$  factorial we would choose two defining relationships, say

$$\begin{aligned} I &= ABCE, I = BCDF; \text{ implying} \\ I &= ABCE * BCDF = ADEF \text{ and} \\ E &= ABC, F = BCD. \end{aligned} \quad (*)$$

We construct the design by writing down all 16 rows of  $\pm$  signs for a full  $2^4$  factorial in factors A,B,C and D. Then the signs for E and F are computed from (\*).

This is a Resolution IV design:  $2_{IV}^{6-2}$ , and all two factor interactions are aliased with other two (or more) factor interactions. For instance

$$\begin{aligned} AB &= CE = ACDF \\ EF &= ABCF = BCDE. \end{aligned}$$

## **Part IX**

# **EXPERIMENTS WITH RANDOM FACTORS**

## 25. One-factor random effects model

- Up to now we have considered only fixed factors - fixed levels of temperature, pressure, etc. Then our inferences are confined only to those levels.
  
- Often factor levels are chosen at random from a larger population of potential levels, and we wish to make inferences about the entire population of levels.
  - Example: A drug company has its products manufactured in a large number of locations, and suspects that the purity of the product might vary from one location to another. Three locations are randomly chosen, and several samples of product from each are selected and tested for purity.
  
  - Example: When pulp is made into paper, it is bleached to enhance the brightness. The type of bleaching chemical is of interest. Four

chemicals are chosen from a large population of potential bleaching agents, and each is applied to five batches of pulp. One wants to know, initially, if there is a difference in the brightness resulting from the chemical types.

- In each of these examples we could use the model

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (*)$$

with  $i$  running over the  $a$  factors (the chosen locations, or the chosen chemicals) and  $j$  running over the  $n$  items in each sample from the factor. The difference now is that **the  $\tau_i$  are random variables** - if the factors are randomly chosen then their effects are random.

- Assume  $\tau_i, \varepsilon_{ij}$  are independent and normally distributed, with means = 0 and variances  $\sigma_\tau^2$  and  $\sigma^2$  respectively. These are called **variance components** and (\*) a **random effects model**. Then

$$\left( \text{var} [y_{ij}] = \right) \sigma_y^2 = \sigma_\tau^2 + \sigma_\varepsilon^2.$$

- The hypothesis of interest becomes  $H_0: \sigma_\tau^2 = 0$ , to be tested against  $H_1: \sigma_\tau^2 > 0$ . (Why?)
- The decomposition

$$SS_T = SS_{Treatments} + SS_E$$

still holds, with  $SS_{Treatments} = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2$ .

We still have

$$E[MS_E] = \sigma_\varepsilon^2, \quad df(MS_E) = N - a.$$

( $N = an$ ) and  $df(MS_{Tr}) = a - 1$ . However (see derivation in text)

$$E[MS_{Tr}] = \sigma_\varepsilon^2 + n\sigma_\tau^2.$$

- Each  $MS$  is  $\sim E[MS] \cdot \chi_{df}^2/df$ , and they are independent, so that

$$F_0 = \frac{MS_{Treatments}}{MS_E} \sim \frac{\sigma_\varepsilon^2 + n\sigma_\tau^2}{\sigma_\varepsilon^2} F_{N-a}^{a-1}.$$

This is  $\sim F_{N-a}^{a-1}$  when  $H_0$  is true, but tends to be larger than an  $F_{N-a}^{a-1}$  when  $H_0$  is false.

Thus  $H_0$  is rejected for large  $F_0$ , and the p-value is

$$p = P\left(F_{N-a}^{a-1} > F_0\right).$$

Since  $\sigma^2 = E[MS_E]$  and  $\sigma_\tau^2 = (E[MS_{Tr}] - E[MS_E])/n$ , common estimators of these are

$$\hat{\sigma}_\varepsilon^2 = MS_E, \quad \hat{\sigma}_\tau^2 = \frac{MS_{Tr} - MS_E}{n}.$$

With unequal group sizes  $n_1, \dots, n_a$  we replace  $n$  in this last equation by

$$n_0 = \bar{n} + \frac{\text{var}(n_1, \dots, n_a)}{a\bar{n}}.$$

- See discussion in text re negative estimates  $\hat{\sigma}_\tau^2$ .

- **Example:** Fabric is woven on a large number of looms, all of which should yield fabric of the same strength. To test this 4 looms are chosen at random, and four fabrics woven from each. Their strengths are measured. Completely Randomized design.

```
loom1    98    97    99    96
loom2    91    90    93    92
loom3    96    95    97    95
loom4    95    96    99    98
```

```
> g <- lm(y~looms)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
looms	3	89.188	29.729	15.681	0.0001878	***
Residuals	12	22.750	1.896			

$$\hat{\sigma}_T^2 = (29.729 - 1.896)/4 = 6.96,$$

$$\hat{\sigma}_y^2 = \hat{\sigma}_T^2 + \hat{\sigma}_\varepsilon^2 = 8.85.$$

We estimate that  $6.96/8.85 = 78\%$  of the variability in the  $y$ 's is attributable to variation between looms.



- Can we put this estimate of 78% in a confidence interval? Put  $ICC = \sigma_\tau^2 / (\sigma_\tau^2 + \sigma_\varepsilon^2)$ , the **intra-class correlation coefficient**. Then  $\widehat{ICC} = .78$  and we want a CI on  $ICC$ .
- Let  $l_{N-a}^{a-1}(\alpha/2), u_{N-a}^{a-1}(\alpha/2)$  be the lower and upper  $\alpha/2$ -points in the  $F_{N-a}^{a-1}$  distribution, so that

$$\begin{aligned}
 1 - \alpha &= P\left(l_{N-a}^{a-1}(\alpha/2) \leq F_{N-a}^{a-1} \leq u_{N-a}^{a-1}(\alpha/2)\right) \\
 &= P\left(\begin{array}{l} l_{N-a}^{a-1}(\alpha/2) \leq F_0 \cdot \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + n\sigma_\tau^2} \\ \leq u_{N-a}^{a-1}(\alpha/2) \end{array}\right) \\
 &= P\left(\begin{array}{l} L = \frac{1}{n} \left[ \frac{F_0}{u_{N-a}^{a-1}(\alpha/2)} - 1 \right] \leq \frac{\sigma_\tau^2}{\sigma_\varepsilon^2} \\ \leq \frac{1}{n} \left[ \frac{F_0}{l_{N-a}^{a-1}(\alpha/2)} - 1 \right] = U \end{array}\right) (**) \\
 &= P\left(\frac{L}{1+L} \leq ICC \leq \frac{U}{1+U}\right).
 \end{aligned}$$

Thus a  $100(1 - \alpha)\%$  CI is  $\left[\frac{L}{1+L}, \frac{U}{1+U}\right]$  with  $L, U$  defined in (\*\*).

- In the looms example:

```

> u <- qf(.975,3,12); l <- qf(.025,3,12)
> F0 <- 15.681
> L <- (F0/u-1)/4; U <- (F0/l-1)/4
> L/(1+L); U/(1+U)
[1] 0.3850669
[1] 0.9824416

```

95% confidence interval on  $ICC$  is [.385, .982].

Go through the same process to get a CI on  $\sigma_\varepsilon^2$ . This is much easier, and is based on the fact that

$$\frac{(N - a) MS_E}{\sigma_\varepsilon^2} \sim \chi_{N-a}^2.$$

Check that the 95% confidence interval is [.975, 5.16].

## 26. Two-factor random effects model

- Here we'll extend the two-factor factorial design to the case of random factors. We'll consider two random factors  $A$  and  $B$ , with the interaction effects random as well.
- Example: Parts used in a manufacturing process are measured with a certain gauge. Variability in the readings can arise from the parts being measured ( $\sigma_{\tau}^2$ ), from the operators doing the measuring ( $\sigma_{\beta}^2$ ), from the interaction between these two ( $\sigma_{\tau\beta}^2$ ), and from the gauge itself. Twenty parts are randomly chosen, and each is measured twice by each of 3 operators chosen from a large population of operators. All 120 measurements are made in a random order, so this is a  $a \times b = 20 \times 3$  factorial with  $n = 2$  replicates. Data in Table 13-3 of the text, R commands on web site.

- Model:

$$\begin{aligned}
 y_{ijk} &= \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \\
 i &= 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n, \\
 \tau_i &\sim N(0, \sigma_\tau^2), \quad \beta_j \sim N(0, \sigma_\beta^2), \\
 (\tau\beta)_{ij} &\sim N(0, \sigma_{\tau\beta}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2),
 \end{aligned}$$

with all r.v.s independent. Thus

$$\sigma_y^2 = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma_\varepsilon^2$$

and the 'gauge variation' is  $\sigma_\varepsilon^2 + \sigma_\beta^2$  - the variation attributable to the instrument or the person operating it.

- There is no change in the ANOVA, the formation of the mean squares, or the d.f. One still computes  $MS_A$  on  $a - 1$  d.f.,  $MS_B$  on  $b - 1$  d.f.,  $MS_{AB}$  on  $(a - 1)(b - 1)$  d.f., and  $MS_E$  on  $ab(n - 1)$  d.f. However, the relevant F-ratios are not necessarily what they were in the fixed factor case. One must start by determining the expected values of the mean squares. There are

rules for doing this (see §13-5 for details) which you will eventually learn if you find work in which you are designing experiments regularly. In this  $a \times b$  factorial these expected values are

$$\begin{aligned} E[MS_A] &= \sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 + bn\sigma_\tau^2, \\ E[MS_B] &= \sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 + an\sigma_\beta^2, \\ E[MS_{AB}] &= \sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2, \\ E[MS_E] &= \sigma_\varepsilon^2. \end{aligned}$$

- Suppose we test the no-interaction hypothesis  $H_0: \sigma_{\tau\beta}^2 = 0$  vs.  $H_1: \sigma_{\tau\beta}^2 > 0$ . Under  $H_0$ ,  $MS_{AB}$  is an unbiased estimate of  $\sigma_\varepsilon^2$ , which is also the expected value of  $MS_E$ . This suggests using

$$F_0 = \frac{MS_{AB}}{MS_E} \sim \frac{\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2}{\sigma_\varepsilon^2} F_{(a-1)(b-1), (n-1)ab},$$

since this is  $\sim F_{(a-1)(b-1), (n-1)ab}$  under  $H_0$  and tends to be larger than this under  $H_1$ .

- Instead, test  $H_0: \sigma_\tau^2 = 0$  vs.  $H_1: \sigma_\tau^2 > 0$ . Under  $H_0$ ,  $MS_A$  is an unbiased estimate of  $\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2$ , so we should compare  $MS_A$  to  $MS_{AB}$ :

$$F_0 = \frac{MS_A}{MS_{AB}} \sim \frac{\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 + bn\sigma_\tau^2}{\sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2} F_{(a-1)(b-1)}^{a-1}$$

and this is  $\sim F_{(a-1)(b-1)}^{a-1}$  under  $H_0$  and tends to be larger than this under  $H_1$ .

- Similarly with  $\sigma_\beta^2$ .
- Note the implication: when you do the ANOVA on R, there will be F-values and p-values printed out in the 'A' and 'B' rows. **These will be for a fixed effects model**, i.e. will be for  $F_0 = MS_A/MS_E$  and  $F_0 = MS_B/MS_E$ , hence can't all be used in this analysis.

- The variance components can be estimated by equating the mean squares to their expected values and solving the resulting equations. Thus

$$\begin{aligned}\hat{\sigma}_\varepsilon^2 &= MS_E, \\ \hat{\sigma}_{\tau\beta}^2 &= \frac{MS_{AB} - MS_E}{n}, \\ \hat{\sigma}_\beta^2 &= \frac{MS_B - MS_{AB}}{an}, \\ \hat{\sigma}_\tau^2 &= \frac{MS_A - MS_{AB}}{bn}.\end{aligned}$$

In the Example, this leads to

$$\hat{\sigma}_{\tau\beta}^2 = -0.14.$$

Of course  $\sigma_{\tau\beta}^2 \geq 0$ , so one way to deal with this is to assume that, indeed,  $\sigma_{\tau\beta}^2 = 0$  (the p-value for this hypothesis was .862). If so, then  $\tau\beta \sim N(0, \sigma_{\tau\beta}^2 = 0)$ : the interactions are all = 0 with probability 1. This implies the **reduced model**

$$y_{ijk} = \mu + \tau_i + \beta_j + \varepsilon_{ijk},$$

which we fit in the usual way. In it,

$$\begin{aligned}E[MS_A] &= \sigma_\varepsilon^2 + bn\sigma_\tau^2, \\ E[MS_B] &= \sigma_\varepsilon^2 + an\sigma_\beta^2, \\ E[MS_E] &= \sigma_\varepsilon^2,\end{aligned}$$

and hypotheses are tested by comparing  $MS_A$  or  $MS_B$  to  $MS_E$  (go back to the  $E[MS]$ 's to check this).

- In this reduced model, the mean squares are

$$\begin{aligned}MS_A &= 62.391, \\MS_B &= 1.308, \\MS_E &= .88,\end{aligned}$$

so that

$$\begin{aligned}\hat{\sigma}_\tau^2 &= \frac{MS_A - MS_E}{bn} = 10.25, \\ \hat{\sigma}_\beta^2 &= \frac{MS_B - MS_E}{an} = .011, \\ \hat{\sigma}_\varepsilon^2 &= .88\end{aligned}$$

and the variability of the gauge (arising from operator variability and random error) is estimated by

$$\hat{\sigma}_{gauge}^2 = \hat{\sigma}_\varepsilon^2 + \hat{\sigma}_\beta^2 = .891.$$

This is much smaller than  $\hat{\sigma}_\tau^2$  (estimating the variability between the parts being measured), so that the gauge should easily distinguish between one part and another (why?).



## 27. Two-factor mixed model

- In the previous measurement systems experiment it was assumed that the 3 operators were chosen from a large pool of possible operators, about which we want to make inferences. Suppose instead that the manufacturer employs only 3 operators, and wishes to make inferences only about these three. Then factor B - 'operator' - is fixed, while factor A - 'part' - is still random.
- This is a 'mixed' model - some factors fixed, others random. To avoid confusion with what is in the text, let's re-define the factors so that  $A =$  operators is fixed and  $B =$  parts is random. The usual ('restricted') model used is

$$\begin{aligned}
 y_{ijk} &= \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \\
 i &= 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n, \\
 \beta_j &\sim N(0, \sigma_\beta^2), \\
 (\tau\beta)_{ij} &\sim N\left(0, \left(\frac{a-1}{a}\right) \cdot \sigma_{\tau\beta}^2\right), \quad \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2), \\
 \sum_{i=1}^a \tau_i &= 0, \quad \sum_{i=1}^a (\tau\beta)_{ij} = 0.
 \end{aligned}$$

- The  $\beta_j$  and  $\varepsilon_{ijk}$  are assumed to be independent of each other, and of the  $(\tau\beta)_{ij}$ . However, since the  $(\tau\beta)_{ij}$  are required to obey the restriction  $\sum_{i=1}^a (\tau\beta)_{ij} = 0$ , they cannot be independent of each other. In fact it can be shown that for each  $j$ , any two of them have covariance

$$\text{cov} \left[ (\tau\beta)_{ij}, (\tau\beta)_{i'j} \right] = -\frac{1}{a} \sigma_{\tau\beta}^2,$$

and hence

$$\text{corr} \left[ (\tau\beta)_{ij}, (\tau\beta)_{i'j} \right] = -\frac{1}{a-1}.$$

- Defining  $\sigma_{\tau\beta}^2$  in such a way that  $\text{var} \left[ (\tau\beta)_{ij} \right] = \left( \frac{a-1}{a} \right) \cdot \sigma_{\tau\beta}^2$  is done merely so that many other expressions become cleaner.

- Major points:
  - The fixed factor obeys the same constraint as in the model with all factors fixed
  - The (main) random factor obeys the same assumptions as in the model with all factors random
  - The interactions are restricted by the sum condition.
  
- There is an ‘unrestricted’ formulation of the model as well, which drops the sum condition, so that the  $(\tau\beta)_{ij}$  are all independent of each other. This is a matter of some controversy.

- The Mean Squares are computed in the usual ways. As in the model with all factors random, to form appropriate F-ratios we have to look at their expected values. Using the rules in §13.5 for Expected Mean Squares, or by a direct calculation, one shows that they are

$$\begin{aligned}
 E[MS_A] &= \sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 + \frac{bn \sum_{i=1}^a \tau_i^2}{a-1}, \\
 E[MS_B] &= \sigma_\varepsilon^2 + an\sigma_\beta^2, \\
 E[MS_{AB}] &= \sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2, \\
 E[MS_E] &= \sigma_\varepsilon^2.
 \end{aligned}$$

- Test  $H_0$ : all  $\tau_i = 0$  using  $F_0 = ???$
- Test  $H_0$ :  $\sigma_\beta^2 = 0$  using  $F_0 = ???$
- Test  $H_0$ :  $\sigma_{\tau\beta}^2 = 0$  using  $F_0 = ???$

- Estimate the fixed effects?

$$\hat{\mu} = \bar{y}_{...},$$

$$\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}.$$

- Estimate the variance components? Using the usual analysis of variance method we equate the mean squares to their expectations and solve for the components to get

$$\hat{\sigma}_{\beta}^2 = \frac{MS_B - MS_E}{an},$$

$$\hat{\sigma}_{\tau\beta}^2 = \frac{MS_{AB} - MS_E}{n},$$

$$\hat{\sigma}_{\varepsilon}^2 = MS_E.$$

- Confidence interval on a difference of treatment means, i.e. on  $\mu_i - \mu_{i'} = \tau_i - \tau_{i'}$ ? As always, it is

$$(\bar{y}_{i..} - \bar{y}_{i'..}) \pm t_{\alpha/2, df} \cdot se(\bar{y}_{i..} - \bar{y}_{i'..}).$$

But what now is the standard error? df? It can be shown (derivation below) that

$$\begin{aligned} \text{var} [\bar{y}_{i..} - \bar{y}_{i'..}] &= 2 \cdot \frac{E[MS_{AB}]}{nb} \\ &= 2 \cdot \frac{E[\text{denominator MS in test for A}]}{\# \text{ of observations used in each } \bar{y}}. \end{aligned}$$

The appropriate CI is thus ( $df = df(MS_{AB})$ )

$$(\bar{y}_{i..} - \bar{y}_{i'..}) \pm t_{\alpha/2, (a-1)(b-1)} \cdot \sqrt{\frac{2}{nb} MS_{AB}}.$$

For multiple comparisons using Tukey's method, replace  $t_{\alpha/2, (a-1)(b-1)}$  by

$$qtukey(1 - \alpha, a, (a - 1)(b - 1)) / \sqrt{2}.$$

Derivation:

$$\begin{aligned}
 \bar{y}_{i..} - \bar{y}_{i'..} &= \frac{1}{nb} \sum_{j,k} y_{ijk} - \frac{1}{nb} \sum_{j,k} y_{i'jk} \\
 &= \frac{1}{nb} \sum_{j,k} \left\{ \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \right\} \\
 &\quad - \frac{1}{nb} \sum_{j,k} \left\{ \mu + \tau_{i'} + \beta_j + (\tau\beta)_{i'j} + \varepsilon_{i'jk} \right\} \\
 &= \tau_i - \tau_{i'} + \frac{1}{b} \sum_j \left\{ (\tau\beta)_{ij} - (\tau\beta)_{i'j} \right\} \\
 &\quad + \frac{1}{nb} \sum_{j,k} \varepsilon_{ijk} - \frac{1}{nb} \sum_{j,k} \varepsilon_{i'jk}.
 \end{aligned}$$

Thus (why?)  $\text{var} [\bar{y}_{i..} - \bar{y}_{i'..}] =$

$$\begin{aligned}
 &\frac{1}{b^2} \sum_j \text{var} \left[ (\tau\beta)_{ij} - (\tau\beta)_{i'j} \right] + \frac{\sigma_\varepsilon^2}{nb} + \frac{\sigma_\varepsilon^2}{nb} \\
 &= \frac{1}{b^2} \sum_j \left\{ 2 \left( \frac{a-1}{a} \right) \cdot \sigma_{\tau\beta}^2 - 2 \left( -\frac{1}{a} \sigma_{\tau\beta}^2 \right) \right\} + \frac{2\sigma_\varepsilon^2}{nb} \\
 &= \frac{2\sigma_{\tau\beta}^2}{b} + \frac{2\sigma_\varepsilon^2}{nb} = \frac{2}{nb} \left( \sigma_\varepsilon^2 + n\sigma_{\tau\beta}^2 \right) = \frac{2}{nb} E[MS_{AB}].
 \end{aligned}$$

- Example: Measurement systems (again).

```
> g <- lm(y~(operator + part)^2)
```

```
Response: y
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
operator	2	2.62	1.31	1.3193	0.2750	
part	19	1185.43	62.39	62.9151	<2e-16	***
operator:part	38	27.05	0.71	0.7178	0.8614	
Residuals	60	59.50	0.99			

```
> MSA <- 1.31
```

```
> MSB <- 62.39
```

```
> MSAB <- .71
```

```
> MSE <- .99
```

```
> a <- 3
```

```
> b <- 20
```

```
> n <- 2
```



From the R programme on web site

F-value for A is 1.845070; p-value is 0.1718915  
 F-value for B is 63.0202; p-value is 0  
 F-value for AB is 0.7171717; p-value is 0.8620944  
 Estimate of  $\sigma^2(\beta) = 10.23333$   
 Estimate of  $\sigma^2(\tau.\beta) = -0.14$

```
# Fit the reduced model
> h <- lm(y ~ operator + part)
> anova(h)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
operator	2	2.62	1.31	1.4814	0.2324
part	19	1185.43	62.39	70.6447	<2e-16 ***
Residuals	98	86.55	0.88		

Estimate of  $\sigma^2(\beta) = 10.25167$

# **Part X**

## **NESTED AND SPLIT-PLOT DESIGNS**

## 28. Two-stage nested designs

- Recall Lecture 22 - partial confounding - where there were two replicates of an experiment. In each there were two blocks, but in one pair of blocks ABC was confounded, in the other set AB was confounded. Thus 'Block1' and 'Block2' meant different things within Replicate 1 than within Replicate 2 - the blocks were 'nested within replicates'.
- Another example - the surface finish of metal parts made on four machines is being studied. Different operators are used on each machine. Each machine is run by three different operators, and two specimens from each operator are tested.

Operators nested within machines											
Machine 1			Machine 2			Machine 3			Machine 4		
1	2	3	1	2	3	1	2	3	1	2	3
79	94	46	92	85	76	88	53	46	36	40	62
62	74	57	99	79	68	75	56	57	53	56	46

- Here ‘Operator 1’ makes sense only within the context of the machine on which this operator works - it refers to something different within Machine 1 than within Machine 2, etc.. When the levels of factor B (operators) make sense only within the levels of factor A, we say that B is ‘nested within’ A, and that this is a ‘nested design’.
- Model and ANOVA. The effects model is

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k},$$

$$k = 1, \dots, n, \quad j = 1, \dots, b, \quad i = 1, \dots, a.$$

‘Interaction’ makes no sense here. The SS and df can be decomposed (how?) as

$$\sum_{i,j,k} (y_{ijk} - \bar{y}_{...})^2 = bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$$

$$+ n \sum_{i,j} (\bar{y}_{ij.} - \bar{y}_{i..})^2 + \sum_{i,j,k} (y_{ijk} - \bar{y}_{ij.})^2,$$

$$SS_T = SS_A + SS_{B(A)} + SS_E,$$

$$abn - 1 = (a - 1) + a(b - 1) + ab(n - 1).$$

```
> g <- lm(y~machine + operator %in% machine)
> anova(g)
```

### Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
machine	3	3617.7	1205.9	14.2709	0.000291
machine:operator	8	2817.7	352.2	4.1681	0.013408
Residuals	12	1014.0	84.5		

- The F's and p-values in the preceding ANOVA are for the fixed effects model (when would this example have both factors fixed?). Both F's have  $MS_E$  in their denominators. We conclude that variation between the machines is very significant, and that within one or more machines, variation between operators is quite significant.
  
- In this 'two-stage' nested design we might have
  - both factors fixed (with  $\sum_i \tau_i = 0$ ,  $\sum_j \beta_{j(i)} = 0$  for each  $i$ ),
  - A fixed and B random ( $\sum_i \tau_i = 0$ , each  $\beta_{j(i)} \sim N(0, \sigma_\beta^2)$ ), or
  - both random ( $\tau_i \sim N(0, \sigma_\tau^2)$  and each  $\beta_{j(i)} \sim N(0, \sigma_\beta^2)$ ).

The appropriate F-ratios are determined by the expected mean squares in each case.

$E(MS)$	A fixed B fixed	A fixed B random	A random B random
$MS_A$	$\sigma^2 + \frac{bn \sum_i \tau_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + \frac{bn \sum_i \tau_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2 + bn\sigma_\tau^2$
$MS_{B(A)}$	$\sigma^2 + \frac{n \sum_{i,j} \beta_{j(i)}^2}{a(b-1)}$	$\sigma^2 + n\sigma_\beta^2$	$\sigma^2 + n\sigma_\beta^2$
$MS_E$	$\sigma^2$	$\sigma^2$	$\sigma^2$

- A fixed, B random

- F to test for effect of A is  $F_0 = ?$
- F to test for effect of B(A) is  $F_0 = ?$
- $\hat{\sigma}_\beta^2 = ?$

- A random, B random

- F to test for effect of A is  $F_0 = ?$
- F to test for effect of B(A) is  $F_0 = ?$
- $\hat{\sigma}_\tau^2 = ?$   $\hat{\sigma}_\beta^2 = ?$

If the machines were randomly chosen:

```
> cat("F to test effect of machines is",  
1205.9/352.2, "and p-value is",  
1-pf(1205.9/352.2, 3, 8), "\n")  
F to test effect of machines is 3.423907  
and p-value is 0.07279175
```

If machines are random so are operator effects (since the operators are operating randomly chosen machines):

```
> cat("Estimate of operator within machines  
variance is", (352.2-84.5)/2, "\n")  
Estimate of operator within machines  
variance is 133.85
```



## 29. Nested and crossed factors

- Extending the analysis of nested designs to the case where there are three factors A, B, C with B nested in A and C in B is straightforward.
  - In R: `lm(y ~ A + B%inA% + C%in%B)`
  - Consult or derive the expected mean squares in order to form appropriate F-ratios, and estimates of variance components.
- A design might have some factorial factors and some nested factors. Again, the analysis uses the same basic principles as above.

- Example: Printed circuit boards (used in electronic equipment - stereos, TVs etc.) have electronic components inserted on them by hand. There are three types of equipment (the 'fixtures') and 2 workplace layouts to be investigated. These factors are crossed (i.e. factorials), and fixed. In layout 1, four operators are (randomly) chosen to insert the components (2 replicates for each fixture). In layout 2, which is in a different location, this is done with a different 4 operators. *So operators is a random factor nested within locations.* A fixture/operator interaction (in each location) makes sense here. Response variable is  $y =$  time to assemble.

	Layout 1				Layout 2			
Oper:	1	2	3	4	1	2	3	4
Fix. 1	22	23	28	25	26	27	28	24
	24	24	29	23	28	25	25	23
Fix. 2	30	29	30	27	29	30	24	28
	27	28	32	25	28	27	23	30
Fix. 3	25	24	27	26	27	26	24	28
	21	22	25	23	25	24	27	27

- Effects model:

Obs'n using fixture  $i$ , layout  $j$ ,  
operator  $k$ , replicate  $l$

$$= y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_{k(j)} + (\tau\beta)_{ij} + (\tau\gamma)_{ik(j)} + \varepsilon_{(ijk)l},$$

$$i \leq a = 3, j \leq b = 2, k \leq c = 4, l \leq n = 2.$$

```
> g <- lm(y~(fixture + layout)^2 +
(operator + fixture*operator)%in%layout)
> anova(g)
```

Analysis of Variance Table

Response: time

	Df	Sum Sq	Mean Sq	F value
A - fixture	2	82.792	41.396	17.7411
B - layout	1	4.083	4.083	1.7500
AB - fix:lay	2	19.042	9.521	4.0804
C(B)-lay:oper	6	71.917	11.986	5.1369
AC(B) - fix:lay:oper	12	65.833	5.486	2.3512
Residuals	24	56.000	2.333	

- Expected mean squares for this model:

$$E[MS_A] = \sigma^2 + 2\sigma_{\tau\gamma}^2 + 8 \sum_i \tau_i^2,$$

$$E[MS_B] = \sigma^2 + 6\sigma_{\gamma}^2 + 24 \sum_j \beta_j^2,$$

$$E[MS_{AB}] = \sigma^2 + 2\sigma_{\tau\gamma}^2 + 4 \sum_{i,j} (\tau\beta)_{ij}^2,$$

$$E[MS_{C(B)}] = \sigma^2 + 6\sigma_{\gamma}^2,$$

$$E[MS_{AC(B)}] = \sigma^2 + 2\sigma_{\tau\gamma}^2,$$

$$E[MS_E] = \sigma^2.$$

- What are the F-ratios?

R programme on web site gives:

	F value	p-value
A - fixture	7.546	.0076
B - layout	0.341	.5807
AB - fix:lay	1.736	.2178
C(B)-lay:oper	5.138	.0016
AC(B) - fix:lay:oper	2.351	.0360

- How are the variance components estimated?

R programme on web site gives:

$$\begin{aligned}\hat{\sigma}_{\tau\gamma}^2 &= 1.577 \\ \hat{\sigma}_{\gamma}^2 &= 1.609\end{aligned}$$

- Tukey confidence intervals on differences of the  $\tau_i$ .  
In general, for mixed models these are  $(\bar{y}_{i\dots} - \bar{y}_{i'\dots}) \pm \frac{q_{\alpha}}{\sqrt{2}} \cdot \sqrt{\frac{2}{r}MS}$ , where  $MS$  is the mean square in the denominator of the F to test the effect, and  $r$  is the number of observations used in each treatment mean. In this case we have ( $\alpha = .05$ )

$$\begin{aligned}\frac{q_{\alpha}}{\sqrt{2}} \cdot \sqrt{\frac{2}{r}MS} &= \frac{qtukey(.95, 3, 12)}{\sqrt{2}} \cdot \sqrt{\frac{2}{16}MS_{AC(B)}} \\ &= 2.21.\end{aligned}$$

The three fixture means are  
25.25 , 27.9375 , 25.0625.

- Conclusions:
  - Fixtures are significantly different; fixtures 1 and 3 result in smaller mean assembly times than fixture 2
  - Operators differ significantly within at least one of the layouts
  - The fixture  $\times$  operator interaction is significant within at least one of the layouts (so some operators are quicker than others, using the same fixtures)
  - Layout does not have a significant effect on assembly time
  
- Recommendations:
  - Use only fixtures 1 and 3
  - Retrain the slower operators

## 30. Split-plot designs

- **Split-plot designs.** Example: In an agricultural experiment,  $a = 3$  fields are planted, with a different crop in each field (the ‘whole plots’; ‘crop’ is the ‘whole plot treatment’). Each field is then divided into  $b = 4$  ‘subplots’, or ‘split-plots’, and each subplot is treated with a different fertilizer (the ‘subplot treatment’). Then the whole experiment is replicated  $r = 3$  times. Sounds like an  $a \times b = 3 \times 4$  factorial run replicated in 3 blocks, and it would be if all 12 combinations were applied in a random order in each block. But they aren’t - *the randomization is only within each whole plot.*
  - The hard to change factor (crops) is the whole plot treatment, and the main factor of interest (fertilizer) is the subplot treatment. Note that the whole plot treatments (crops) are confounded with the whole plots (fields) - if there is a systematic difference between the fields (soil quality?) it will show up as a difference between the crops.

- Another example: In making pulp into paper,  $a = 3$  methods are used to prepare the pulp. Then the preparation is baked at one of  $b = 4$  temperatures. Response is  $y =$  strength of paper. Not a  $3 \times 4$  factorial unless 12 different batches of pulp are prepared - this is not feasible economically. Instead, three batches of pulp are prepared, each using one of the three methods. Each batch (whole plot) is split into 4 smaller samples (subplots) and each is baked at one of the 4 temperatures (subplot treatments). The whole process is replicated  $r = 3$  times. *Randomization is only within the methods, not within the replicates.*

	Rep 1			Rep 2			Rep 3		
Method:	1	2	3	1	2	3	1	2	3
Temp:									
200°	30	34	29	28	31	31	31	35	32
225°	35	41	26	32	36	30	37	40	34
250°	37	38	33	40	42	32	41	39	39
275°	36	42	36	41	40	40	40	44	45



- The hard to change factor (preparation method) is the “whole plot treatment”, and the main factor of interest (baking temperature) is the “subplot treatment”. Again the whole plot treatments are confounded with the whole plots - if there is a systematic difference between the batches it will show up as a difference between the preparation methods.
- If this experiment is replicated  $r$  times, then we view the replicates as blocks. The simplest model, and one that can be easily fitted on R, includes effects of {replicates + replicates  $\times$  whole plot interactions}, and {wholeplot and subplot treatments and their interaction}:

$$y_{ijk} = \mu + \underbrace{\tau_i + (\tau\alpha)_{ij}}_{\text{replicate effects}} + \underbrace{\alpha_j + \beta_k + (\alpha\beta)_{jk}}_{\text{treatment effects}} + \varepsilon_{ijk}$$

$$i = 1, \dots, r, \quad j = 1, \dots, a, \quad k = 1, \dots, b.$$

Note my  $\alpha$  is Montgomery's  $\beta$ ; my  $\beta$  is Montgomery's  $\gamma$ .

- If we call these effects  $R$  (replicates, with effects  $\tau_i$ ),  $A$  (whole plot treatments, with effects  $\alpha_j$ ),  $B$  (subplot treatments, with effects  $\beta_k$ ) then the sums of squares are computed by fitting a linear model  $\text{lm}(y \sim (R + RA) + (A + B)^2)$ .

Source	SS	df
$\tau$ : Replicates (blocks)	$SS_R$	$r - 1$
$(\tau\alpha)$ : RA inter'n	$SS_{RA}$	$(r - 1)(a - 1)$
$\alpha$ : Whole plot treatment	$SS_A$	$a - 1$
$\beta$ : Subplot treatment	$SS_B$	$b - 1$
$(\alpha\beta)$ : AB inter'n	$SS_{AB}$	$(a - 1)(b - 1)$
$\varepsilon$ : Subplot error	$SS_E$	$a(r - 1)(b - 1)$

The RA interaction can be viewed as an additional error term, arising because the whole plot treatment (preparation method) is replicated between blocks. The other -  $SS_E$  - arises because the subplot treatment (temperature) is replicated within blocks.

- Quite commonly both treatment effects are fixed and only replicates are random (hence so are interactions involving replicates). Then the expected mean squares are:

Model term	MS	E(MS)
$\tau$	$MS_R$	$\sigma_\varepsilon^2 + b\sigma_{\tau\alpha}^2 + ab\sigma_\tau^2$
$\alpha$	$MS_A$	$\sigma_\varepsilon^2 + b\sigma_{\tau\alpha}^2 + \frac{rb \sum \alpha_j^2}{a-1}$
$(\tau\alpha)$	$MS_{RA}$	$\sigma_\varepsilon^2 + b\sigma_{\tau\alpha}^2$
$\beta$	$MS_B$	$\sigma_\varepsilon^2 + \frac{ra \sum \beta_k^2}{b-1}$
$(\alpha\beta)$	$MS_{AB}$	$\sigma_\varepsilon^2 + \frac{r \sum_{j,k} (\alpha\beta)_{jk}^2}{(a-1)(b-1)}$
<i>Error</i>	$MS_E$	$\sigma_\varepsilon^2$

All F-ratios except one have  $MS_E$  in the denominator. The exception is that to test for factor A ( $H_0$ : all  $\alpha_j = 0$ ) we use  $F_0 = MS_A/MS_{RA}$ .

```
g <- lm(y ~ (R/A.pulp) + (A.pulp + B.temp)^2)
# R/A.pulp means "R + R*A.pulp"
```

```
Response: y
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
R	2	77.56	38.78	9.7622	0.001345
A.pulp	2	128.39	64.19	16.1608	9.586e-05
B.temp	3	434.08	144.69	36.4266	7.449e-08
R:A.pulp	4	36.28	9.07	2.2832	0.100284
A.pulp:B.temp	6	75.17	12.53	3.1538	0.027109
Residuals	18	71.50	3.97		

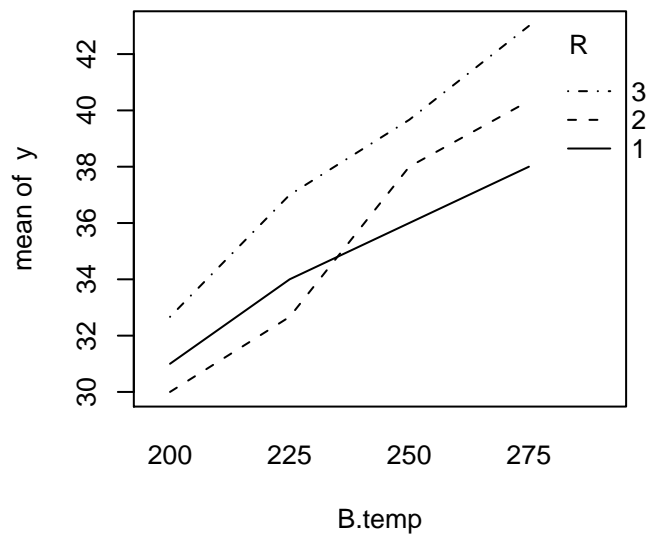
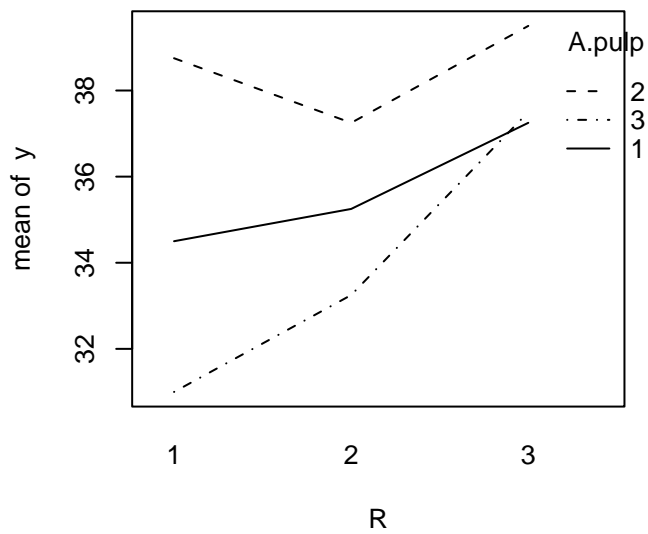
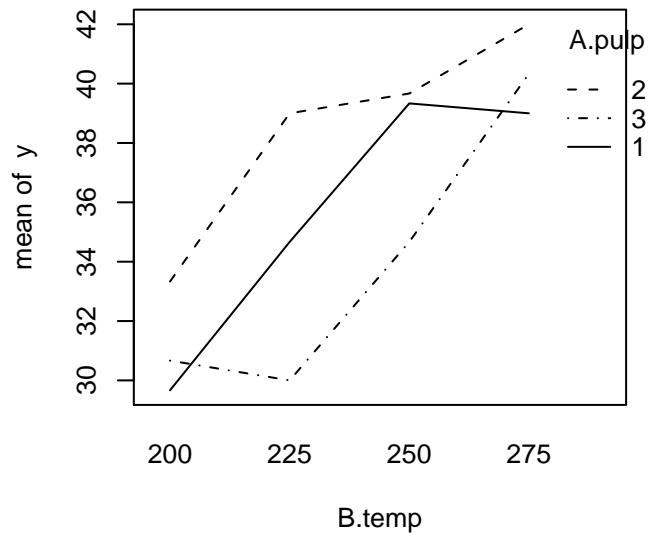
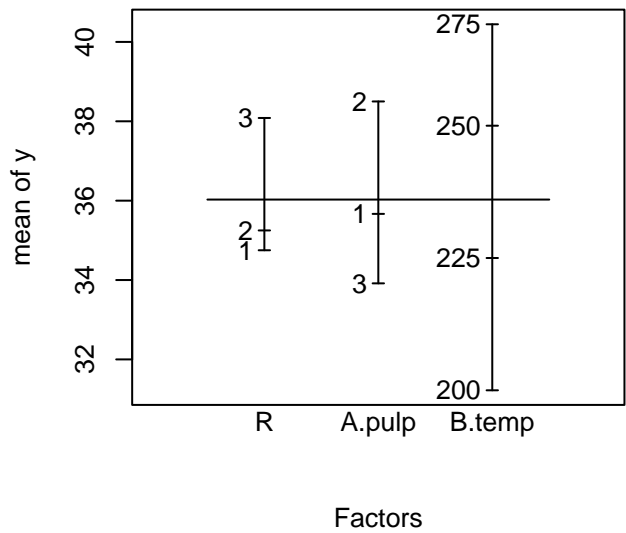
```
---
```

```
> MSA <- 64.19;    dfA <- 2
> MSRA <- 9.07;   dfRA <- 4
```

```
> F.A <- MSA/MSRA
> p.A <- 1-pf(F.A, dfA, dfRA)
```

```
F.A = 7.077178 with p = 0.04854655
```

From the printout, the p-values for B and AB are  $7 \times 10^{-9}$  and .027 respectively. Thus all three effects (A, B, AB) are quite significant.



Design and interaction plots.  $A = 2, B = 275$  looks like the winner.

Here is a way to get everything from one command:

```
> h <- aov(y~(A.pulp+B.temp)^2 + Error(R/A.pulp))
> summary(h)
```

Error: R

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	2	77.556	38.778		

Error: R:A.pulp

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A.pulp	2	128.389	64.194	7.0781	0.04854
Residuals	4	36.278	9.069		

---

Error: Within

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
B.temp	3	434.08	144.69	36.4266	7.449e-08
A.pulp:B.temp	6	75.17	12.53	3.1538	0.02711
Residuals	18	71.50	3.97		

---

The aov command has fitted  $y \sim (A.pulp + B.temp)^2$ , taken the residuals from this model, and fitted R and  $R * A.pulp$  to these residuals to get the correct value of SSE (irrelevant and/or incorrect values XX'd out):

```
k <- lm(y ~ (A.pulp+B.temp)^2) #Gives MSA, MSB, MSAB
anova(k)
```

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A.pulp	2	128.39	64.19	XXX	XXX
B.temp	3	434.08	144.69	XXX	XXX
A.pulp:B.temp	6	75.17	12.53	XXX	XXX
Residuals	XX	XXX	XXX		

```
k2 <- lm(k$resid ~ (R+A.pulp)^2) #Gives MSR, MSRA, SSE
with SSA=0 (since A has already been accounted for)
anova(k2)
```

Response: k\$resid

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
R	2	77.556	38.778	XXX	XXX
A.pulp	2	0	0	0	1.00
R:A.pulp	4	36.278	9.069	XXX	XXX
Residuals	XX	71.500	XXX		

# **Part XI**

# **SPECIAL TOPICS**



## 31. Regression

- **Regression** - a general method to relate one variable ( $y$ , the 'dependent' variable) to one or more 'independent' variables  $x_1, \dots, x_p$ . In the usual models the experimenter chooses the values of the  $x$ 's, and then observes the r.v.  $y$ .
- Model:

$$y = f(x_1, \dots, x_p) + \varepsilon,$$

for some function  $f(x_1, \dots, x_p)$  and random error (possibly measurement error)  $\varepsilon \sim N(0, \sigma^2)$ . Thus the expected value of  $y$ , when it is observed at the values  $x_1, \dots, x_p$ , is

$$E[y|x_1, \dots, x_p] = f(x_1, \dots, x_p).$$

Given observations

$$\left\{ y_i, x_{1i}, \dots, x_{pi} \right\}_{i=1}^n$$

we try to estimate  $f$ ; we can then use this estimate to estimate the mean of  $y$  at any values of the  $x$ 's, observed or not. Or we can predict future values of  $y$ .

- Most common is linear regression, in which  $f$  is a linear function of some (unknown) parameters  $\beta_0, \dots, \beta_p$ :

$$f(x_1, \dots, x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

- Example 1. The time required for a car to stop is related to the initial speed through:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

with  $y =$  time,  $x =$  speed. Interpretation of intercept  $\beta_0$  and slope  $\beta_1$ ? Is it reasonable to take  $\beta_0 = 0$ ?

- Example 2. In Example 1, more realistic might be  $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$  (so  $x_1 = x, x_2 = x^2$ ).

- Example 3. The independent variables need not be continuous - they could act as labels. For instance suppose  $Y$  is the response variable in a CRD experiment, with three treatments. Define  $x_1 = 1$  if treatment 2 is used,  $x_2 = 1$  if treatment 3 is used, both = 0 otherwise. Then

$$E[y|x_1, x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 = \begin{cases} \beta_0, & \text{trt1,} \\ \beta_0 + \beta_1, & \text{trt2,} \\ \beta_0 + \beta_2, & \text{trt3,} \end{cases}$$

so that  $\beta_1$  and  $\beta_2$  represent the differences in treatment means (comparing to treatment 1). All treatments are equally effective if  $\beta_1 = \beta_2 = 0$ .

- Estimation is done by least squares. With  $\beta = (\beta_0, \dots, \beta_p)$  the sum of squares function is (as always)

$$\begin{aligned} S(\beta) &= \sum_{i=1}^n \left( y_i - E[y_i | x_{1i}, \dots, x_{pi}] \right)^2 \\ &= \sum_{i=1}^n \left( y_i - \beta_0 - \beta_1 x_{1i} - \dots - \beta_p x_{pi} \right)^2 \end{aligned}$$

and this is to be minimized. If  $\hat{\beta}_0, \dots, \hat{\beta}_p$  are the minimizers, they must satisfy the “normal equations”: for each  $j = 0, \dots, p$  we must have

$$\begin{aligned} 0 &= \frac{\partial}{\partial \beta_j} S(\boldsymbol{\beta}) \Big|_{\hat{\beta}_0, \dots, \hat{\beta}_p} \\ &= -2 \sum_{i=1}^n \left( y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \dots - \hat{\beta}_p x_{pi} \right) x_{ji}. \end{aligned}$$

(We put  $x_{0i} = 1$  if there is an intercept  $\beta_0$ .) The residuals are

$$\begin{aligned} e_i &= y_i - \hat{y}_i \\ &= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \dots - \hat{\beta}_p x_{pi}, \end{aligned}$$

so that the normal equations can be written

$$\sum_{i=1}^n e_i x_{ji} = 0, \quad j = 0, \dots, p.$$

In particular, if there is an intercept we have  $\bar{e} = 0$  and so

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \dots - \hat{\beta}_p \bar{x}_p.$$

The minimum value of  $S$  is

$$S(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^n e_i^2 = SS_E.$$

- Example: In straight line regression (SLR),  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$  (\*). The other normal equation is

$$\begin{aligned}
 0 &= \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i \\
 &= \sum_{i=1}^n (y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{x})) x_i \text{ from (*)} \\
 \Rightarrow \hat{\beta}_1 &= \frac{\sum_{i=1}^n (y_i - \bar{y}) x_i}{\sum_{i=1}^n (x_i - \bar{x}) x_i} \\
 &= \frac{\sum_{i=1}^n (y_i - \bar{y}) (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \text{ (how?)}.
 \end{aligned}$$

- For  $p \geq 2$  the solution is expressed more simply in matrix terms.
- On R the command `g <- lm(y~x1+x2+. . .+xp)` followed by `summary(g)` or `anova(g)` will do the fitting.

```

> x <- seq(from=1,to=5,length=20)
> y <- 1+2*x+rnorm(20)
> g <- lm(y~x)
> summary(g)

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.6250	0.4644	1.346	0.195
x	1.9966	0.1435	13.913	4.51e-11

---

Residual standard error: 0.7791 on 18 df

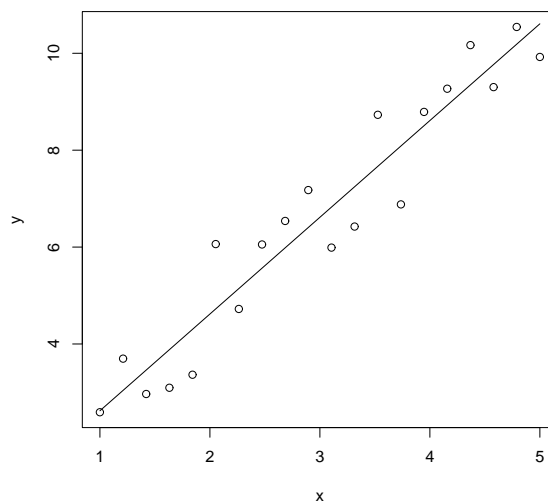
Multiple R-Squared: 0.9149,

Adjusted R-squared: 0.9102

F-statistic: 193.6 on 1 and 18 df,

p-value: 4.509e-11

```
> plot(x,y); lines(x,g$fitted.values)
```



- Testing: After fitting  $E[Y|x_1, \dots, x_p] = \beta_0 + \beta_1x_1 + \dots + \beta_px_p$  we might want to test that some of the  $\beta$ 's are zero, i.e. that the corresponding  $x$ 's do not affect  $E[y]$ . The general procedure is (as always) to fit the reduced model which assumes the null hypothesis that these  $\beta$ 's are zero, and see how much this affects  $SS_E$ . If  $r$  of the  $\beta$ 's are dropped,

$$F_0 = \frac{\{SS_E(\text{Reduced}) - SS_E(\text{Full})\} / r}{MS_E(\text{Full})}$$

and  $F_0 \sim F_{n-p-1}^r$  if  $H_0$  is true. Otherwise it tends to be larger, so the  $p$ -value is  $P(F_{n-p-1}^r > F_0)$ . Note that the df of  $SS_E(\text{Full})$  is  $n - p - 1 = [n - \# \text{ of } \beta\text{'s being estimated}]$ , and this always appears on the printout.

- Special cases:
  - If  $r = 1$  then  $\sqrt{F_0} = t_0 \sim t_{n-p-1}$  is given on the R output, together with the  $p$ -value.
  - If  $r = p$ , and we are testing the hypothesis that  $E[y]$  does not depend on **any** of the  $x$ 's in the study, then  $F_0$  and the  $p$ -value appear on the output.
  
- From the preceding output:
  - $H_0: \beta_0 = 0$  has  $p$ -value  $P(|t_{18}| > 1.346) = .195$ .
  - $H_0: \beta_1 = 0$  has  $p$ -value  $P(|t_{18}| > 13.913) = 4.51 \times 10^{-11}$ .
  - $H_0: \beta_1 = 0$  has  $p$ -value  $P(F_{18}^1 > 193.6) = 4.51 \times 10^{-11}$ .



## 32. ANOVA via regression

- ANOVA can always be performed as an application of regression.
- Example 1: The designs studied so far have been **balanced** - equal numbers of observations in each treatment group, each block, etc. When this balance is absent the regression method can be useful. Recall the RCBD from Lecture 9, but with one observation missing to destroy the balance. Assume that it is “missing at random” (MAR) and not, for instance, because of its unusual  $y$ -value.

Hardness testing design and data.

$y_{ij}$  = machine reading for tip  $i$ , coupon  $j$ .

	Coupon (C)			
Tip (T)	1	2	3	4
1	9.3	9.4	9.6	10.0
2	9.4	9.3	—	9.9
3	9.2	9.4	9.5	9.7
4	9.7	9.6	10.0	10.2

Define independent ‘indicator’ variables

$$x_1 = I(C = 2), x_2 = I(C = 3), x_3 = I(C = 4),$$

$$x_4 = I(T = 2), x_5 = I(T = 3), x_6 = I(T = 4),$$

and fit a regression model

$$E[Y|\mathbf{x}] = \beta_0 + \beta_1 x_1 + \cdots + \beta_6 x_6.$$

		$E[Y \mathbf{x}]$			
		C			
T		1	2	3	4
1		$\beta_0$	$\beta_0 + \beta_1$	$\beta_0 + \beta_2$	$\beta_0 + \beta_3$
2		$\beta_0 + \beta_4$	$\beta_0 + \beta_1 + \beta_4$	$\beta_0 + \beta_2 + \beta_4$	$\beta_0 + \beta_3 + \beta_4$
3		$\beta_0 + \beta_5$	$\beta_0 + \beta_1 + \beta_5$	$\beta_0 + \beta_2 + \beta_5$	$\beta_0 + \beta_3 + \beta_5$
4		$\beta_0 + \beta_6$	$\beta_0 + \beta_1 + \beta_6$	$\beta_0 + \beta_2 + \beta_6$	$\beta_0 + \beta_3 + \beta_6$

From the table of expected values we see that, in each block,

$$\beta_4 = E[y|tip2] - E[y|tip1],$$

$$\beta_5 = E[y|tip3] - E[y|tip1],$$

$$\beta_6 = E[y|tip4] - E[y|tip1].$$



Test  $H_0: \beta_4 = \beta_5 = \beta_6 = 0$  (no treatment effects).

$$\begin{aligned}
 F_0 &= \frac{\{SS_E(\text{Reduced}) - SS_E(\text{Full})\} / r}{MS_E(\text{Full})} \\
 &= \frac{\{.45750 - .06222\} / 3}{.00778} = 16.936, \\
 p &= P(F_8^3 > 16.936) = 0.000796.
 \end{aligned}$$

You should check that this is the  $p$ -value given by `anova(lm(y~as.factor(blocks) + as.factor(tips)))` but not by `anova(lm(y~as.factor(tips) + as.factor(blocks)))` unless the tips  $x$ 's are entered **first** in the regression. A way around this inconvenient feature is to use the regression equation to **estimate** the missing (if MAR) observation - by  $\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_4 = 9.62$  (coefficients obtained from `summary(g)`) - and then do the usual complete data analysis. The only modification required is that  $df(SS_E)$  should be reduced by 1 before the  $p$ -values are computed.

Example 2: One replicate of a  $2^3$  factorial is carried out with factors temperature (T) at levels  $120^\circ C$  and  $160^\circ C$ , Pressure (P) at 40 and 80 pounds and catalyst concentration (C) at levels 15 and 30 grams/litre. There are two nonstandard features of the design. One is that 4 additional observations are made at the centre point:  $T = 140$ ,  $P = 60$ ,  $C = 22.5$ . This is commonly done to investigate a possible quadratic trend. The other is that it was not possible to hold T, P and C at exactly these levels - there was some fluctuation. The actual data, in run order, were

	y	T	P	C
1	32	125	41	14.0
2	46	158	40	15.0
3	57	121	82	15.0
		etc.		
9	50	140	60	22.5
10	44	140	60	22.5
11	53	140	60	22.5
12	56	140	60	22.5

The three factor and TC interactions were not of interest, but a possible quadratic effect was. So we fit the model

$$E[Y|T, P, C] = \beta_0 + \beta_1 T + \beta_2 P + \beta_3 C + \beta_{12} TP + \beta_{23} PC + \beta_{11} T^2 + \beta_{22} P^2 + \beta_{33} C^2.$$

Note that we are using the numerical values of the variables - they are not treated as factors.

```
> g <- lm(y~T + P + C + T*P + P*C
          + I(T^2) + I(C^2)+ I(P^2))
> # See help(formula) for an explanation of I().
> summary(g) # To get the coefficients
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-4.077e+03	7.082e+03	-0.576	0.605
T	6.915e+01	1.213e+02	0.570	0.609
P	-3.550e+01	6.772e+01	-0.524	0.636
C	2.351e+01	4.927e+01	0.477	0.666
I(T^2)	-2.470e-01	4.368e-01	-0.566	0.611
I(C^2)	-3.968e-01	8.938e-01	-0.444	0.687
I(P^2)	2.922e-01	5.398e-01	0.541	0.626

```

T:P          1.398e-02  3.242e-02  0.431  0.696
P:C          -5.815e-02  9.805e-02  -0.593  0.595
> anova(g)   # To get SSE

```

For this model,  $SS_E = 78.75$  on 3 d.f. Individually none of the terms looks significant, but the interpretation of the t-tests is that they give the p-value for that term **if it is entered last**. We test for the significance of all interaction and quadratic terms by fitting the additive model:

```
h <- lm(y~T + P + C)
```

for which  $SS_E = 102.22$  on 8 d.f. The test statistic for the hypothesis that all these terms can be dropped is

$$F_0 = \frac{(102.22 - 78.75)/5}{78.75/3} = .179 < 1$$

with  $p = .953$ . So we accept the reduced model:

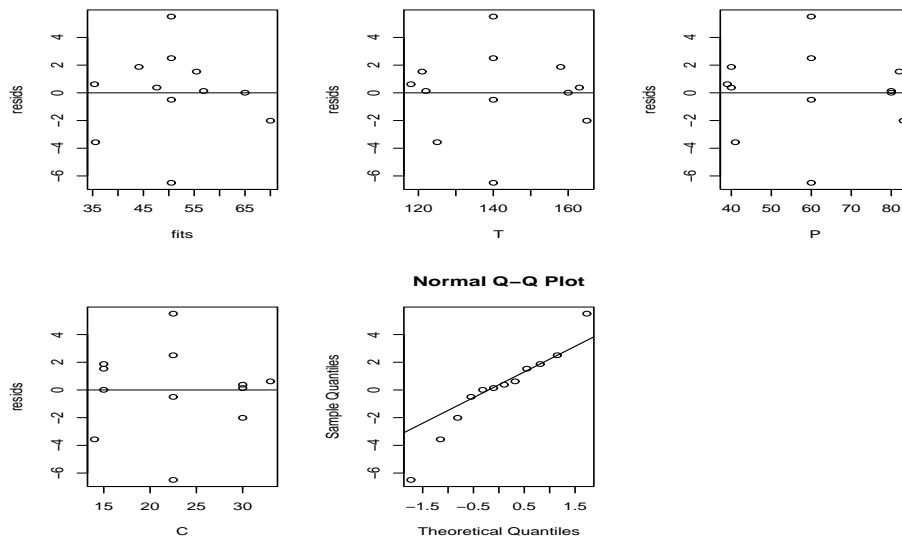
```
> summary(h)
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-21.08256	10.25781	-2.055	0.07390
T	0.27050	0.06265	4.317	0.00256
P	0.50816	0.06127	8.293	3.37e-05
C	0.14300	0.15695	0.911	0.38884

The final regression equation, to predict  $y$  from T, P and C, is

$$\hat{y} = -21.08 + .27T + .51P + .14C.$$



Residual plots





### 33. Analysis of covariance

- Recall that blocking is generally used to **control** for a nuisance factor - e.g. days of week, operator of machine, etc. We can assign the runs of the experiment to particular blocks. Sometimes however there is a nuisance factor that can be observed but not controlled - age of the patients in a medical trial, rainfall in an agricultural experiment, etc. The response ( $y$ ) may be quite closely related to this nuisance factor ( $x$ ) and so the values of  $x$  should play a role in the analysis. We call  $x$  a **covariate**, and the subject **analysis of covariance** (ANCOVA).
- Example. Three machines are used to produce a certain type of fibre. The response variable is  $y =$  strength. However, the thickness ( $x$ ) of the fibre will clearly affect strength. This varies both within and between machines, and can be measured but not controlled.

Machine 1		Machine 2		Machine 3	
$y$	$x$	$y$	$x$	$y$	$x$
36	20	40	22	35	21
41	25	48	28	37	23
39	24	39	22	42	26
42	25	45	30	34	21
49	32	44	28	32	15

The layout of the design is as for a CRD -  $a = 3$  treatments (machines),  $n = 5$  observations made in each treatment group, all  $an$  runs carried out in random order. From the design standpoint the only difference is that the covariate is measured along with the response variable.

- Analysis. Along with the usual terms in the effects model for a single factor CRD, we include a term expressing the departure of the covariate from its overall average, and assume that  $y$  is linearly related to this departure.

If  $y_{ij}$  is the  $j^{\text{th}}$  observation in the  $i^{\text{th}}$  treatment group, the model is

$$y_{ij} = \mu + \tau_i + \beta (x_{ij} - \bar{x}_{..}) + \varepsilon_{ij},$$

$$i = 1, \dots, a, \quad j = 1, \dots, n.$$

Here  $\tau_i$  is the effect of the  $i^{\text{th}}$  treatment, and we assume that  $\sum_i \tau_i = 0$ . We also assume that  $\varepsilon_{ij} \sim N(0, \sigma^2)$  and that the covariate is not affected by the treatments. (E.g. in a drug trial in which drugs are the treatments and  $y$  is a measure of the effectiveness of the drug, we wouldn't use 'time for which the drug was taken' as a covariate - that could obviously depend on which drug was taken).

- Write  $X_{ij}$  for  $x_{ij} - \bar{x}_{..}$ , so that  $E[y_{ij}] = \mu + \tau_i + \beta X_{ij}$  and  $\sum_{i,j} X_{ij} = 0$ . We can view  $\mu + \tau_i$  as the intercept of the regression line relating  $y_{ij}$  to  $X_{ij}$  for fixed  $i$ . Exactly as the LSE of an intercept was derived in Lecture 31, we get that the LSE of  $\mu + \tau_i$  is

$$\bar{y}_{i.} - \hat{\beta} \bar{X}_{i.} = \text{adj } \bar{y}_{i.},$$

the  $i^{th}$  “adjusted treatment mean”. (If there was no covariate then  $\bar{y}_i$  alone would be the estimate.) The unbiased estimate of  $\tau_i$  is

$$\hat{\tau}_i = (adj \bar{y}_i) - \bar{y}_{..}$$

Note that  $\mu + \tau_i = E[y_{ij} | X_{ij} = 0]$ , so that this can also be estimated by  $\hat{y}(X = 0, treatment = i)$ . In fact these estimates are equal:

$$adj \bar{y}_i = \hat{y}(X = 0, treatment = i).$$

Thus a package (virtually any regression package) that analyzes  $\hat{y}$  at arbitrary values of  $X$  will yield the adjusted treatment means and their standard errors.

- The main hypothesis of interest is  $H_0: \tau_1 = \dots = \tau_a = 0$ . This is equivalent to the statement that all  $\mu + \tau_i$  are equal, and so is tested by comparing their estimates - the adjusted treatment means - with each other. On a regression package we can merely enter ‘machines’ last, so that its  $SS$  is automatically adjusted for  $X$ , and then obtain the relevant  $p$ -value from the printout. Remember to enter treatments as a factor.

```
> data
```

```

      y  x machine      X
1  36 20      1 -4.1333333
2  41 25      1  0.8666667
3  39 24      1 -0.1333333
      etc.
10 44 28      2  3.8666667
11 35 21      3 -3.1333333
12 37 23      3 -1.1333333
13 42 26      3  1.8666667
14 34 21      3 -3.1333333
15 32 15      3 -9.1333333

```

```
> g <- lm(y ~X + machine);anova(g)
```

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X	1	305.130	305.130	119.9330	2.96e-07
machine	2	13.284	6.642	2.6106	0.1181
Residuals	11	27.986	2.544		

From this, the  $p$ -value for machines is .1181. We conclude that there is no significant difference between the machines, once their output is adjusted for fibre thickness. (An incorrect 1-way ANOVA, ignoring  $X$ , gives  $p = .04$ .) The 13.284 in the output is referred to as  $SS(\text{machines}|X)$ , the  $SS$  for machines, adjusted for thickness. Entering  $X$  last shows that the variation in fibre thickness accounts for a significant amount of the variation in strength ( $p = 4.264 \times 10^{-6}$ ,  $SS(X|\text{machines}) = 178.014$ ):

```
> g0 <- lm(y ~ machine + X);anova(g0)
```

```
Analysis of Variance Table
```

```
Response: y
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
machine	2	140.400	70.200	27.593	5.170e-05
X	1	178.014	178.014	69.969	4.264e-06
Residuals	11	27.986	2.544		

Of course this  $p$ -value is also that for the hypothesis  $H_0: \beta = 0$ :

```
> summary(g)
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	40.3824	0.7236	55.806	7.55e-15
X	0.9540	0.1140	8.365	4.26e-06
machine2	1.0368	1.0129	1.024	0.328
machine3	-1.5840	1.1071	-1.431	0.180
---				

We can easily get information on the adjusted treatment means on R. After creating a linear model by `g <- lm(· · ·)`, we can estimate the mean response at new values of the variables using, for example, `predict(g, new=data.frame(machine=as.factor(1), X=0), se.fit=T)`. Thus:

```
# Compute the adjusted treatments means
```

```
and their standard errors:
```

```
ybar.adj <- se.of.adjusted.mean <- vector(length=3)
```



```

for(i in 1:3) {
prediction <- predict(g, new=
  data.frame(machine=as.factor(i), X=0), se.fit=T)
ybar.adj[i] <- prediction$fit
se.of.adjusted.mean[i] <- prediction$se.fit
}
tau.hat <- ybar.adj - mean(y)
cat("Adjusted machine means and their
  standard errors are", "\n")
print.matrix(cbind(tau.hat, ybar.adj,
  se.of.adjusted.mean))

```

gives the output

```

Adjusted machine means and their standard errors are
  tau.hat    ybar.adj    se.of.adjusted.mean
  0.1824131  40.38241    0.7236252
  1.2192229  41.41922    0.7444169
 -1.4016360  38.79836    0.7878785

```

Compare with the MINITAB output on p. 583; also contrast the ease of this regression approach with the more algebraic treatment at pp. 576-580.

### 34. Repeated measures and crossover designs

- Repeated measures designs. In some fields (educational experiments, drug trials, etc.) the experimental units are often people, and each of them might receive several treatments. This might be in an attempt to reduce the between-person variability, or it might be because few people are both suitable and willing subjects. Or, the person might receive just one treatment but then be measured repeatedly to record his/her progress.
- Example 1: Three different cardiovascular compounds are to be tested, to see how effective they are at reducing blood pressure. There are  $n$  patients recruited, and each takes each of the  $a = 3$  treatments, in random order and with a sufficient 'washout' period between them that their effects do not depend on previous treatments.

- Analysis: We can treat this as a RCBD, in which the subjects are the  $n$  blocks, and they have a random effect. We model the response of the  $j^{\text{th}}$  patient to the  $i^{\text{th}}$  treatment as

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

$$\sum \tau_i = 0, \beta_j \stackrel{\text{ind.}}{\sim} N(0, \sigma_\beta^2).$$

This is a special case of the mixed model we studied in Lecture 27 - put  $\sigma_{\tau\beta}^2 = 0$  there to get:

$$E[MS_A] = \sigma_\varepsilon^2 + \frac{n \sum_{i=1}^a \tau_i^2}{a-1}, \quad (a-1 \text{ d.f.}),$$

$$E[MS_B] = \sigma_\varepsilon^2 + a\sigma_\beta^2, \quad (b-1 \text{ d.f.}),$$

$$E[MS_E] = \sigma_\varepsilon^2, \quad ((a-1)(b-1) \text{ d.f.}).$$

Then the  $F$  to test  $H_0$ : all  $\tau_i = 0$  is

$$F_0 = \frac{MS_A}{MS_E} \sim F_{(a-1)(b-1)}^{a-1} \text{ under } H_0.$$

We call  $MS_B$  the ‘between patients’ mean square.

- Example 2: Same framework as previously, but suppose that  $n = 3m$  and that each patient receives only one treatment, and is then monitored by having his/her blood pressure measured every 15 minutes for the first hour, and every hour thereafter for the next 6 hours. So there are  $a = 3$  treatment groups, with  $m$  patients in each. We take  $b$  measurements on each patient. There are three controlled sources of variation: treatments (A), times (B), and subjects (C, nested in A). We assume that the first two of these are fixed effects and the third is random. Model: the observation on the  $k^{\text{th}}$  subject in the  $i^{\text{th}}$  treatment group, at the  $j^{\text{th}}$  time, is

$$\begin{aligned}
 y_{ijk} &= \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_{k(i)} + \varepsilon_{ijk}, \\
 i &= 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, m, \\
 \sum_i \tau_i &= \sum_j \beta_j = 0, \quad \sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0, \\
 \gamma_{k(i)} &\overset{\text{ind.}}{\sim} N(0, \sigma_\gamma^2).
 \end{aligned}$$

Recall Lecture 29, where we discussed designs with both factorial and nested factors. This is

such a design. It can be shown that the expected mean squares are:

$$\begin{aligned}
 E[MS_A] &= \sigma_\varepsilon^2 + b\sigma_\gamma^2 + \frac{bm \sum_{i=1}^a \tau_i^2}{a-1}, \\
 E[MS_{C(A)}] &= \sigma_\varepsilon^2 + b\sigma_\gamma^2, \\
 E[MS_B] &= \sigma_\varepsilon^2 + \frac{am \sum_{j=1}^b \beta_j^2}{b-1}, \\
 E[MS_{AB}] &= \sigma_\varepsilon^2 + \frac{m \sum_{i=1}^a \sum_{j=1}^b (\tau\beta)_{ij}^2}{(a-1)(b-1)}, \\
 E[MS_E] &= \sigma_\varepsilon^2.
 \end{aligned}$$

From this it is clear how the appropriate F-ratios are formed.

- Example 3: Here is a design in which all  $3m$  patients receive all  $a = 3$  treatments, but we don't rely on the randomization (of the order in which the treatments are applied to a patient) to control for the effects of changes over time. Label the treatments  $T_1, T_2, T_3$  and make a  $3 \times 3$  Latin square from these letters:

Sequence	Factor B (periods)		
	1	2	3
1	$T_1$	$T_3$	$T_2$
2	$T_2$	$T_1$	$T_3$
3	$T_3$	$T_2$	$T_1$

Choose  $m$  patients at random, and have them take the treatments according to sequence 1. Similarly assign  $m$  patients to each of the other sequences. An appropriate model accounts for variation due to treatments (A), periods (B), sequences (C), patients nested in sequences (D(C)) and an A×B interaction: the  $l^{th}$  patient in sequence  $k$ , when receiving treatment  $i$  in period  $j$ , has response

$$y_{ijkl} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_k + \delta_{l(k)} + \varepsilon_{ijkl}.$$

Only  $\delta_{l(k)}$  is a random factor.

This is called a ‘crossover’ design, since patients cross over to another treatment at the end of each period.