Chapter 10

Analysis of experiments

All of the analysis techniques described in this chapter can be applied to observational studies, provided the structure of the data is the same. For example, if there are two categorical explanatory variables (factors), the standard analysis is a two-way ANOVA, whether the data arise from an observational study or a designed experiment. But the interpretation of a designed experiment allows more possibilities, as it can give more credence to causation.

In this lab, we also look at *interaction* between explanatory variables, which more often arises in the context of a designed experiment.

10.1 Model fitting review

We have seen a number of models (and will see more in this chapter) of the form:

$$\begin{array}{ccccc} y & = & M(\ ) & + & e \\ \uparrow & & \uparrow & & \uparrow \\ \text{observations} & & \text{model} & & \text{error} \\ \text{(response variable)} & & \text{(explanatory variables)} & & \end{array}$$

where the errors e are assumed to be independent, and normally distributed with constant variance. The model, $M(\ )$, depends on explanatory variables, which may be factors (categorical variables) or covariates (numerical variables).

one sample: $M(\ ) = \mu$.

straight line regression: $M(\ ) = \beta_0 + \beta_1 x$

multiple linear regression: $M(\ ) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$ ($x_1, \ldots, x_k$ covariates)

k **independent samples:** $M(\ ) = \mu + \alpha_i$ (α_i = level of factor A)
"one-way analysis of variance"

k **matched samples:** $M(\ ) = \mu + \alpha_i + \beta_j$ (α_i = level of factor A, β_j = level of factor B)
"two-way analysis of variance" *(Note the different use of β in this model.)*

10.2 Completely randomised designs

The data from a completely randomised design can be analysed by a one-way analysis of variance, which reduces to an independent samples t-test when there are only two treatments. The ANOVA enables us to answer the following questions:

- How much variation is due to differences *between* the treatments?
- How much variation is due to differences *within* each set of observations for the same treatments?
- In formal hypothesis testing terms: how likely is it that we would observe the differences we have observed, if there are no true differences between the treatments?

Example: effect of sugars on pea growth in tissue culture.

An experiment was conducted to examine the effect of different sugars on length (in mm) of pea sections grown in tissue culture with auxin present. Four types of sugar were allocated to 40 pea sections in a completely randomised design. The results were as follows (the data are stored in `sugars.csv`):

sugar											mean
control	8.6	7.6	8.0	8.6	7.4	8.1	7.6	7.6	8.7	7.8	8.00
glucose	6.5	6.6	6.8	6.7	7.1	6.8	6.8	6.5	6.7	7.0	6.75
fructose	6.6	7.0	6.4	6.6	6.5	6.4	7.0	6.8	6.5	6.6	6.64
sucrose	7.1	7.5	7.4	7.2	7.3	7.1	7.4	7.4	7.1	7.6	7.31

An appropriate statistical model for these data is

$$\begin{aligned} \text{response} &= \text{average length for sugar } i + \text{error} \\ y_{ij} &= \mu_i + e_{ij} \end{aligned}$$

$j = 1, \ldots, 10$ represents the jth replicate of sugar $i = 1, \ldots, 4$, and the errors e_{ij} are independent $\mathrm{N}(0, \sigma)$ random variables.

The null hypothesis H_0 is tested against the alternative hypothesis H_1, where

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \ \ (= \mu)$.

$H_1 : H_0$ is not true.

From the table, there appear to be substantial differences between the sugar means. This is confirmed in the ANOVA:

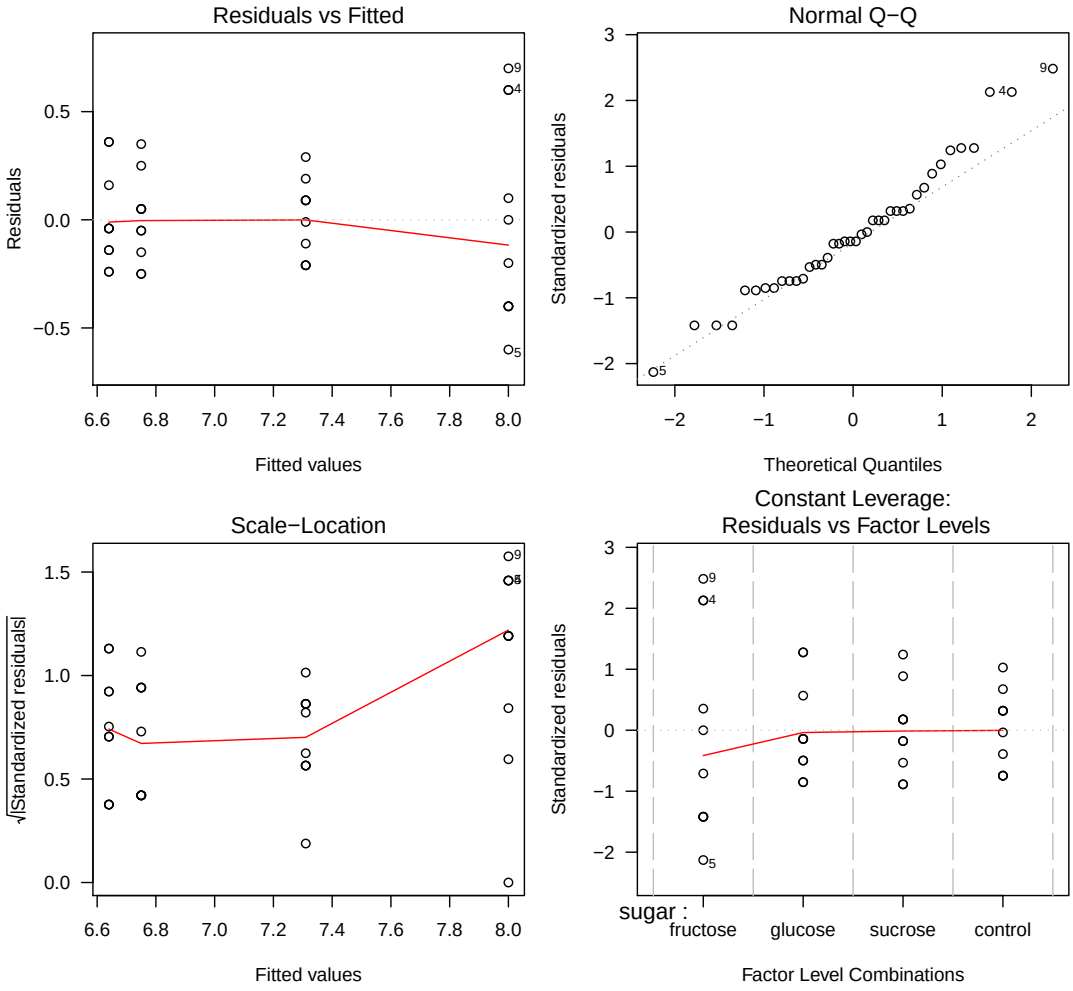
```
> sugars <- read.csv("../data/sugars.csv")
> sugars.lm <- lm(length ~ sugar, data = sugars)
> anova(sugars.lm)

Analysis of Variance Table

Response: length
          Df Sum Sq Mean Sq F value    Pr(>F)
sugar      3 11.657  3.8857  44.016 3.938e-12 ***
Residuals 36  3.178  0.0883
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

It is good to examine diagnostic plots, especially the graph of residuals vs fitted values, which indicates that the assumption of homogeneity of variance is reasonable. Although the largest mean is associated with somewhat greater variability than the others, there is no pattern to the residuals.

```
> par(mfrow = c(2, 2))
> plot(sugars.lm)
```



10.3 Randomised block designs

The data from a randomised block design can be analysed by a two-way ANOVA. The ANOVA enables us to partition the total variation into three sources: blocks, treatments, and error. Typically, the variation due to blocks is of little direct interest; we simply wish to remove the block variation, which reduces the residual variation and enables us to estimate more precisely the differences between treatments.

Example: Effect of drugs on lymphocyte levels in mice.

In an experiment to compare the effects of four drugs (A, B, C and a placebo D) on lymphocyte counts in mice, a randomised block design with four mice from each of five litters was used, the litters being regarded as blocks. The lymphocyte counts (in thousands per mm^3 of blood) are shown in the following table (the data are stored in `lymphocyte.csv`).

	Litter					
Drug	M	N	O	P	Q	$\bar{x}$
A	7.1	6.1	6.9	5.6	6.4	6.42
B	6.7	5.1	5.9	5.1	5.8	5.72
C	7.1	5.8	6.2	5.0	6.2	6.06
D	6.7	5.4	5.7	5.2	5.3	5.66
$\bar{x}$	6.90	5.60	6.18	5.23	5.93	5.97

An appropriate statistical model for these data is

$$\begin{aligned} \text{count} &= \text{overall mean} + \text{litter effect} + \text{drug effect} + \text{error} \\ y_{ij} &= \mu + l_j + \alpha_i + e_{ij} \end{aligned}$$

$i = 1, \ldots, 4, \;\; j = 1, \ldots, 5, \;\; e_{ij} \stackrel{d}{=} \mathrm{N}(0, \sigma)$.

The null hypothesis H_0 is tested against the alternative hypothesis H_1, where

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 \;\; (= 0)$, i.e. there are no differences between drugs.

$H_1 : H_0$ is not true.

There appear to be clear differences between the drug means, and also between the litter means (though that is not of much interest). This is confirmed in the ANOVA:

```
> lymphocyte <- read.csv("../data/lymphocyte.csv")
> lymphocyte.lm <- lm(count ~ litter + drug, data = lymphocyte)
> anova(lymphocyte.lm)

Analysis of Variance Table

Response: count
          Df Sum Sq Mean Sq F value    Pr(>F)
litter     4 6.4030 1.60075  30.155 3.544e-06 ***
drug       3 1.8455 0.61517  11.589 0.0007392 ***
Residuals 12 0.6370 0.05308
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Comparisons between pairs of means are performed in the same way as for completely randomised designs.

The general form of the ANOVA table for a randomised block design with t treatments and b blocks is:

Source	df	SS	MS	F	P
blocks	$b-1$	SS_{blk}	$SS_{blk}/(b-1)$		
treatments	$t-1$	SS_{trt}	$SS_{trt}/(t-1)$	MS_{trt}/MS_{res}	P
residual	$(b-1)(t-1)$	SS_{res}	$SS_{res}/[(b-1)(t-1)]$		
total	$bt-1$	SS_{tot}			

Everything in the ANOVA table works the same way as it did when the model included only treatments; there is simply an additional factor (blocks) to be accounted for in the partitioning of the total variation. Note that we haven't included an F-test for the effect of blocks in the general form; though some statistical packages (including R) perform the test, it is not especially useful.

What if blocks were (wrongly) ignored?

Suppose that the data were presented for analysis without reference to the different litters, i.e. all the mice were assumed to be unrelated to each other. The ANOVA table would then be as follows:

```
> lymphocyte.lm.1 <- lm(count ~ drug, data = lymphocyte)
> anova(lymphocyte.lm.1)

Analysis of Variance Table

Response: count
          Df Sum Sq Mean Sq F value Pr(>F)
drug       3 1.8455 0.61517  1.3981 0.2797
Residuals 16 7.0400 0.44000
```

The P-value is 0.28, from which we would conclude that the drugs are not significantly different in their effect. Clearly, not accounting for the litters has led to less precise inference, and in this case, incorrect conclusions.

The residual mean square, which estimates σ^2, has increased greatly, from 0.053 to 0.440. In the previous analysis (assuming randomised blocks), σ^2 is related to differences between two mice *within a litter*. In this analysis (assuming no blocking), σ^2 is related to differences between *any* two mice, and so is much larger.

Question: Can you produce this ANOVA table from the ANOVA which included litters?

A special case: 2 experimental units per block

A randomised block design sometimes has just 2 treatments and 2 experimental units per block—for example, a medical study involving identical twins. This can be analysed as in the example above, using a two-way ANOVA, or as a **paired t-test**; the outputs look a little different, but the analyses are equivalent.

10.4 Latin square designs

The analysis of a Latin square design is a straightforward extension of the analysis of a randomised block design; it just has an additional blocking factor. The analysis therefore has three factors, usually denoted rows, columns and treatments.

An appropriate model for the ith row, the jth column and the kth treatment is

$$\begin{aligned} \text{response} &= \text{overall mean} + \text{row effect} + \text{column effect} + \text{treatment effect} + \text{error} \\ y_{ijk} &= \mu + r_i + c_j + \alpha_k + e_{ijk} \end{aligned}$$

where as usual the errors are assumed to be normally distributed with the same standard deviation σ, and independent.

The general form of the ANOVA table for a Latin square design with t treatments is:

Source	df	SS	MS	F	P
rows	$t-1$	SS_{row}	$SS_{row}/(t-1)$		
columns	$t-1$	SS_{col}	$SS_{col}/(t-1)$		
treatments	$t-1$	SS_{trt}	$SS_{trt}/(t-1)$	MS_{trt}/MS_{res}	P
residual	$(t-1)(t-2)$	SS_{res}	$SS_{res}/[(t-1)(t-2)]$		
total	t^2-1	SS_{tot}			

Example: Food supplements and milk yield of cows

Cow	Period					
	I		II		III	
1	A	608	B	885	C	940
2	B	715	C	1087	A	766
3	C	844	A	711	B	832

```
> milk.yield <- data.frame(cow = factor(rep(1:3, each = 3)), period = rep(c("I",
+     "II", "III"), 3), supp = c("A", "B", "C", "B", "C", "A",
+     "C", "A", "B"), yield = c(608, 885, 940, 715, 1087, 766,
+     844, 711, 832))
> milk.yield

  cow period supp yield
1   1      I    A   608
2   1     II    B   885
3   1    III    C   940
4   2      I    B   715
5   2     II    C  1087
6   2    III    A   766
7   3      I    C   844
8   3     II    A   711
9   3    III    B   832

> milk.yield.lm <- lm(yield ~ cow + period + supp, data = milk.yield)
> anova(milk.yield.lm)

Analysis of Variance Table
```

```
Response: yield
          Df Sum Sq Mean Sq F value  Pr(>F)
cow        2   5900    2950  1.2183 0.45079
period     2  47214   23607  9.7490 0.09303 .
supp       2 103436   51718 21.3584 0.04473 *
Residuals  2   4843    2421
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> summary(milk.yield.lm)$coef
```

```
             Estimate Std. Error    t value    Pr(>|t|)
(Intercept) 586.55556   43.39753 13.5158739 0.005429545
cow2         45.00000   40.17831  1.1200074 0.379153384
cow3        -15.33333   40.17831 -0.3816321 0.739464919
periodII    172.00000   40.17831  4.2809171 0.050471130
periodIII   123.66667   40.17831  3.0779462 0.091325625
suppB       115.66667   40.17831  2.8788338 0.102451791
suppC       262.00000   40.17831  6.5209319 0.022718594
```

```
> tapply(milk.yield$yield, milk.yield$supp, mean)
```

```
       A        B        C
695.0000 810.6667 957.0000
```

The ANOVA table shows that the effect of food supplement is significant at the 5% level. The effects of cow and period are not significant, but this is not of interest anyway, because they are blocking factors.

The treatments are quite different in their mean milk yields; B produced 17% more milk than A, and treatment C produced 38% more (surely economically important differences). Yet the treatment effect was only just significant at the 0.05 level, and the means for treatments A and B were not significantly different ($P = 0.10$ for the `suppB` coefficient). This lack of power to detect significant differences can easily occur in a 3×3 Latin square, for two reasons:

1. The design has only 3 replicates, and so the random variation needs to be small to obtain significant differences. Here, it is not small: $\hat{\sigma} = \sqrt{2421} = 49.2$. For example, we would expect 95% of the yields for a particular treatment combination (e.g. cow 2 in period III receiving treatment A) to be within about 100 grams/day of the mean—quite a wide interval.

2. There are only 2 error DF in the ANOVA, which results in a large t value (4.303) for constructing confidence intervals, etc. This reflects the high level of uncertainty in estimating σ with so few degrees of freedom. Experimental designs should, in general, aim to have at least 10 error DF.

Therefore, it is wise to obtain more replicates if possible. For example, if there were six cows available, a pair of Latin squares could be used, as noted in lab 9 (top of page 9).

10.5 Factorial experiments

A **factorial experiment** is one whose purpose is to investigate the effect of two or more *factors*, each of which has at least two levels. The total number of **treatment combinations** is equal to the product of the number of levels of each factor. For example, if factor A has 3 levels, factor B has 2 levels and factor C has 4 levels, then the number of treatment combinations is $3 \times 2 \times 4 = 24$, and the trial is often called a $3 \times 2 \times 4$ factorial experiment. When each of the factors has the same number of levels, it is usual to write this in the form a^b. For example, an experiment with 3 factors each at 2 levels is called a 2^3 factorial experiment. The term "factorial" is usually attributed to R. A. Fisher, who used it in his 1935 book *The Design of Experiments*.

Factorial experiments are a way of setting up experiments so that we make use of the *structure* in the treatments, which is called a **factorial treatment structure**. Factors arranged in a factorial structure are often said to be **crossed**. For analysis, it is always possible to ignore such structure and perform a one-way ANOVA, with the number of groups equal to the number of treatment combinations. However,

this is often not very useful, particularly if we find that the effects of the factors are *additive*, because we would like separate information on each factor.

For the experimental designs examined in the previous lab, any factorial treatment structure doesn't make any difference to the design. For example, a 2^3 factorial experiment has 8 treatment combinations, which can be randomised to the experimental units in the same way as 8 levels of a single factor. The designs we have considered so far can incorporate factorial treatments; the distinctive features of the different types of designs relate to their *blocking* structures. The distinctive features of a factorial experiment relates to its *treatment* structure.

10.5.1 Main effects and interaction

Example: Quality of pancakes.

The effect of two factors—a baking supplement and the amount of whey—on the quality of pancakes was studied in an experiment with a completely randomized design. Three batches of each supplement $\times$ whey combination were made, and the quality rated by a panel of experts. The quality ratings were as follows:

	Amount of whey				
	0%	10%	20%	30%	mean
no supplement	4.4	4.6	4.5	4.6	
	4.5	4.5	4.8	4.7	4.63
	4.3	4.8	4.8	5.1	
	3.3	3.8	5.0	5.4	
supplement	3.2	3.7	5.3	5.6	4.34
	3.1	3.6	4.8	5.3	
mean	3.80	4.17	4.87	5.12	4.49

Overall, there appears to be some detrimental effect of the supplement, and a substantial beneficial effect relating to the amount of whey. But it would also be helpful to look at the effect of either factor at each level of the other factor. One way is to construct a table of means for each treatment combination.

```
> pancakes <- data.frame(supp = rep(c("no supplement", "supplement"),
+     each = 12), whey = rep(rep(c("0%", "10%", "20%", "30%"),
+     each = 3), 2), quality = c(4.4, 4.5, 4.3, 4.6, 4.5, 4.8,
+     4.5, 4.8, 4.8, 4.6, 4.7, 5.1, 3.3, 3.2, 3.1, 3.8, 3.7, 3.6,
+     5, 5.3, 4.8, 5.4, 5.6, 5.3))
> round(tapply(pancakes$quality, pancakes[, 1:2], mean), 2)
```

```
               whey
supp             0%  10%  20%  30%
  no supplement 4.4 4.63 4.70 4.80
  supplement    3.2 3.70 5.03 5.43
```

(Note that `[, 1:2]` means select all rows and the first two columns of the data frame.)

A more informative picture emerges. While the average effect of the supplement is detrimental, the extent to which it reduces the quality of the pancakes diminishes as the percentage of whey increases, and for the maximum percentage (30%), the supplement appears quite beneficial. Or, looking at it the other way, the effect of increasing the percentage of whey is minimal in the absence of the supplement, but substantial in the presence of the supplement. In statistical terminology, the two factors *interact*.

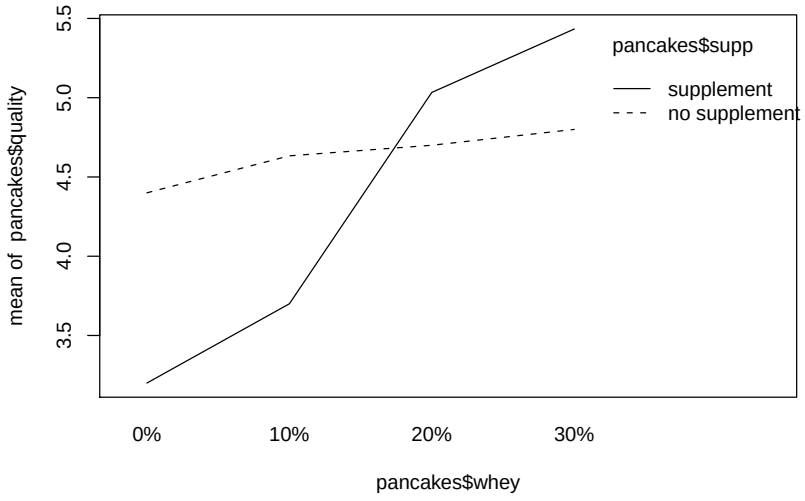
> **Interaction** between two factors occurs when the differences between levels of one factor depend on the level of the other factor.

The presence of interaction means that the effects are not *additive*, but there is a synergism to take into account as well.

An informative graph is what is sometimes called an *interaction plot*. The response variable is plotted on the vertical axis, and one of the factors (usually the one with the most levels) is put on the horizontal axis. For each level of the other factor, a line is drawn joining the means for each level of the first factor. In such a graph, no interaction appears as parallel lines. The greater the departure from parallel lines (or additivity), the greater the interaction.

The graph is produced in R by:

```
> library(stats)
> interaction.plot(pancakes$whey, pancakes$supp, pancakes$quality)
```



The plot shows visually and clearly the effects of the factors, and their substantial interaction. In particular, it reinforces the point that we can't consider one factor in isolation from the other.

The model we fit to the data is

$$\begin{aligned} \text{quality} &= \text{overall mean} + \text{supplement effect} + \text{whey effect} + \text{interaction effect} + \text{error} \\ y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \end{aligned}$$

The supplement effects ($\alpha_i,\ i = 1, 2$) are adjustments to the overall mean due to the absence or presence of the supplement. Similarly for the whey effects, $\beta_j,\ j = 1$ to 4. The third subscript k (ranging from 1 to 3) represents the different batches for each supplement $\times$ whey combination.

The supplement and whey effects, α_i and β_j, are called **main effects**. The **interaction effects**, γ_{ij}, are the additional effects arising from each particular combination of supplement and whey. For example, γ_{23} is the particular effect of baking with a supplement and 20% whey, which is added to the overall effect of baking with a supplement and the overall effect of baking with 20% whey. In terms of the interaction plot, the interaction effects are estimated by the amount of departure from parallel lines. In testing the significance of the interaction, the issue is whether the observed departure from parallelism (or additivity) is more than would be expected just by chance.

The model can be tested in an analysis of variance:

```
> pancakes.lm <- lm(quality ~ supp * whey, data = pancakes)
> anova(pancakes.lm)

Analysis of Variance Table

Response: quality
          Df Sum Sq Mean Sq F value    Pr(>F)
supp       1 0.5104 0.51042  17.014 0.0007942 ***
whey       3 6.6912 2.23042  74.347 1.304e-09 ***
supp:whey  3 3.7246 1.24153  41.384 9.130e-08 ***
Residuals 16 0.4800 0.03000
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

(Note that `supp*whey` fits both main effects and the interaction. It is equivalent to `supp + whey + supp:whey`, where `supp:whey` is the interaction.)

Three null hypotheses can be tested using the ANOVA table:

$H_0 : \alpha_1 = \alpha_2 = 0$ (no main effect of supplement);
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (no main effect of whey);
$H_0 : \gamma_{ij} = 0$ for all i, j (no interaction).

The small P-values would lead us to reject all three null hypotheses. However, because of the significant interaction, it is not very meaningful to test the main effects, since any statement about one of the factors has to be qualified by stating which levels of the other factor are being considered. For example, the significant supplement main effect implies that the mean for pancakes baked without a supplement, 4.63, is significantly larger than the mean for pancakes baked with a supplement, 4.34. But these figures are averaged across all the levels of the whey factor, and so are not very useful. It is clear that the null hypothesis concerning the interaction should be tested first. If it is accepted, then the null hypotheses about the main effects can be tested.

Creating a single factor

Suppose we create a new factor, "cookmethod", with the levels of this factor being the eight combinations of supplement $\times$ whey. An ANOVA with cookmethod as the only factor in the model results in:

```
> pancakes$cookmethod <- factor(rep(1:8, each = 3))
> pancakes.lm.1 <- lm(quality ~ cookmethod, data = pancakes)
> anova(pancakes.lm.1)

Analysis of Variance Table

Response: quality
           Df Sum Sq Mean Sq F value   Pr(>F)
cookmethod  7 10.926  1.5609   52.03 7.94e-10 ***
Residuals  16  0.480  0.0300
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Compare this table with the ANOVA table having the two factors and their interaction. The SS for cookmethod is simply the sum of the SS for supp, whey, and the supp × whey interaction. In other words, the variation among the eight treatment combinations here *has been partitioned* into three components—the main effect of supplement, the main effect of whey, and the interaction.

Degrees of freedom for analysing a factorial experiment

The following rules apply to the degrees of freedom in an ANOVA, regardless of the experimental design:

- The total df is 1 fewer than the total number of observations.
- The df for a main effect is 1 fewer than the number of levels of the factor.
- The df for an interaction term is the product of the df of the factors making up the interaction (provided there are no missing treatment combinations).

10.5.2 Inference when the interaction is not significant

Differences in food consumption among rats were assessed when rancid lard was substituted for fresh lard in their diet. Food eaten (in grams) during 73 days by 12 rats was recorded. There were two factors: lard (fresh vs rancid) and gender (M, F). Three rats of each gender were randomly allocated to fresh lard, and three to rancid lard. The results were as follows:

	Lard		
Gender	Fresh	Rancid	row mean
M	709	592	
	679	538	
	699	476	
mean	695.7	535.3	615.5
F	657	508	
	594	505	
	677	539	
mean	642.7	517.3	580.0
column mean	669.2	526.3	

Looking at the means, there is a substantial difference between fresh and rancid lard, a much smaller difference between male and female rats, and there appears to be not much interaction. These observations are borne out in the interaction plot and the ANOVA table:

```
> ratfood <- data.frame(gender = rep(c("M", "F"), each = 6), lard = rep(rep(c("fresh",
+     "rancid"), each = 3), 2), food = c(709, 679, 699, 592, 538,
+     476, 657, 594, 677, 508, 505, 539))
> interaction.plot(ratfood$gender, ratfood$lard, ratfood$food)
```

```
> ratfood.lm <- lm(food ~ gender * lard, data = ratfood)
> anova(ratfood.lm)

Analysis of Variance Table

Response: food
            Df Sum Sq Mean Sq F value    Pr(>F)
gender       1   3781    3781  2.5925 0.1460358
lard         1  61204   61204 41.9685 0.0001925 ***
gender:lard  1    919     919  0.6300 0.4502546
Residuals    8  11667    1458
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Because the interaction is not significant, it is valid to compare means for the main effects.

The value of factorial experiments

This example illustrates one of the advantages of factorial experiments—although there are only 3 replicates of each treatment combination, 6 replicates are used to calculate the means for either treatment factor on its own because of the lack of significant interaction. This is sometimes called "hidden replication", because the additional replication for the main effects "appears" only after the lack of interaction is established. Of course, if the interaction is significant, the hidden replication does not apply, so the experimental design should not count on it being there.

Occasionally you will hear statements such as "in an experiment, only change one factor at a time". This *sounds* reasonable, but is not good advice—there is no need to follow such a restriction. R. A. Fisher argued against the prevailing view in 1926, when he wrote: "No aphorism is more frequently repeated in connection with field trials, than that we must ask Nature few questions, or, ideally, one question, at a time. The writer is convinced that this view is wholly mistaken. Nature will best respond to a logical and carefully thought out questionnaire."

In some situations, one or more of the treatment combinations may be impossible, nonsensical, or ethically dubious, and so the full factorial treatment structure is not appropriate. For example, consider a trial examining the effect of three factors on vitamin D levels in the blood: (1) fairness of skin; (2) exposure to the sun; and (3) vitamin D supplementation. It may be appropriate for those with the fairest skin not to be subject to the greatest sun exposure because of the increased risk of skin cancer. But in most experiments, the full factorial structure should be used.

10.6 Exercises

1. Here are the weight gains of chicks on four feeds. Using the ANOVA methods formulate, estimate, check and interpret an appropriate model.

Feed A	42	68	85			
Feed B	42	97	81	95	61	103
Feed C	61	112	30	89	63	
Feed D	169	137	169	111	154	

2. The following data were obtained from an experiment to determine the effect of storage conditions on the moisture content of white pine timber. A completely randomised design was used, but with unequal replication of the three conditions.

Conditions	Moisture Content (%)					Mean
1	7.3	8.3	7.6	8.6	8.3	8.02
2	5.4	7.4	7.1			6.63
3	8.5	9.5	10.0			9.33

   (a) Read the data into R, and test the null hypothesis of equal means using analysis of variance.

   (b) Manually calculate a 95% confidence interval for the mean difference between conditions 1 and 2. (A reminder that `qt(.975,df)` gives the relevant t distribution quantile in R.)

3. Rainfall and cloud seeding. The data are in `cloud seeding.csv`.

   (a) Calculate the mean, median and standard deviation for both the control and seeded groups. What do these summary statistics tell you about the distributions of the two groups?

   (b) Use the `anova()` function to compare the means of the control and seeded groups.

   (c) Perform the same analysis using a t-test, checking that the results are equivalent.

   (d) When larger standard deviations accompany larger means, a log transformation is often useful in addressing the issue of unequal variances. Apply a log transformation to the rainfall data, check the summary statistics, and repeat the analysis.

   (e) From your analysis of the data, would you recommend cloud seeding?

4. The following data were obtained from an experiment to compare four types of tip for measuring the hardness of metal. Five strips of metal were used and the depth of penetration measured with each tip used on each strip.

	Type of Tip				
Strip	1	2	3	4	Mean
1	9.3	9.4	9.2	9.7	9.40
2	9.4	9.3	9.4	9.6	9.43
3	9.6	9.8	9.5	10.0	9.73
4	10.0	9.9	9.7	10.2	9.95
5	9.7	9.7	9.5	9.9	9.70
Mean	9.60	9.62	9.46	9.88	9.64

```
> metal <- data.frame(strip = factor(rep(1:5, each = 4)), tip = factor(rep(1:4,
+     5)), hardness = c(9.3, 9.4, 9.2, 9.7, 9.4, 9.3, 9.4, 9.6,
+     9.6, 9.8, 9.5, 10, 10, 9.9, 9.7, 10.2, 9.7, 9.7, 9.5, 9.9))
```

   (a) Run an analysis of variance, performing any appropriate significance tests.

   (b) Calculate the 95% confidence interval for the mean difference in hardness between tips 3 and 4.

   (c) Construct diagnostic plots to examine the assumptions of the analysis.

(d) Using the ANOVA table from (a), and without further using R (except as a calculator), construct an ANOVA table for an analysis with strip omitted from the model. (Don't bother about the P-value.) Confirm your ANOVA table by running the analysis in R, and examine the P-value.

(e) Comment on whether the blocking (by strip) in the experiment has been worthwhile.

5. Six varieties of turnip (A to F) were grown in 36 plots arranged in a Latin Square. The response variable was the fresh weight of turnips in kg per plot. The arrangement and fresh weights were as follows:

E	1.000	F	0.502	D	0.712	A	0.784	B	0.574	C	0.223
B	0.612	A	1.021	E	0.426	C	0.323	D	1.145	F	0.430
F	0.591	B	1.031	C	0.447	D	1.000	A	0.932	E	0.437
A	1.086	D	1.083	F	0.832	E	0.670	C	0.361	B	0.729
D	0.859	C	0.457	B	1.076	F	0.901	E	0.677	A	0.973
C	0.419	E	0.450	A	1.179	B	0.698	F	0.646	D	0.997

(a) The data are in `turnip.csv`. Read the data into R, and calculate means for the different varieties.

(b) Perform an analysis of variance incorporating the rows and columns of the experimental design, and test the hypothesis that the six means are equal. Note that rows and columns must be modelled as factors, not numerical variables.

(c) Suppose that you were told that the experiment had been arranged, not as a Latin square, but as a randomised block design, with the rows being blocks. Using the analysis from (b), calculate the residual MS that would result. Is it very different? Check your calculation by running the ANOVA as a randomised block design. Would the conclusions be very different compared to the Latin square analysis?

6. An experiment was conducted to investigate the effects of vitamin B_{12} and antibiotics on the weight gain of pigs. Three litters were used, each comprising 4 pigs, to which the 4 treatment combinations were randomised. The response measured was the weight gain (in kg) over a four week period. The results were as follows:

	Antibiotics		
B_{12}	No	Yes	Litter
No	1.30	1.05	1
	1.19	1.00	2
	1.08	1.05	3
Yes	1.26	1.52	1
	1.21	1.56	2
	1.19	1.55	3

(a) What type of experimental design was this?

(b) Enter the data into R, and construct a table of means for the treatment combinations, and an interaction plot. Use them to make some tentative conclusions about the interaction.

(c) Perform an analysis of variance on the weight gains. The model should include litter, and the main effects and interaction of both treatment factors. Use the ANOVA table to confirm your impressions about the interaction.

(d) Suppose that the treatments had been described as a single factor at 4 levels, corresponding to the four treatment combinations. Use the ANOVA table to determine whether this factor had a significant effect on weight gain. Create a single factor in R and run an ANOVA to confirm your calculation.

7. An experiment was conducted to examine the effect of three thinning regimes on the height of mountain ash trees. Five provenances were used, thus creating 15 treatment combinations. There were only enough resources to have two replicates, arranged in a completely randomised design. Complete the ANOVA table below, and calculate the P-values using the `pf()` function in R: $P = 1-$ `pf(F,df1,df2)`.

SOURCE	DF	SS	MS	F	P
Thinning		140			
Provenance			300		
Interaction					
Error		120			
Total		1890			

8. To ascertain the stability of vitamin C in reconstituted frozen orange juice concentrate stored in a refrigerator, an experiment was conducted at the Virginia Polytechnic Institute. Two varieties of orange (navel and valencia) and two types of storage (1 month and 3 months) were compared, with four samples of each treatment combination being tested in a completely randomised design. The amount of vitamin C, in milligrams of ascorbic acid per litre, was recorded as follows:

navel 1 month	valencia 1 month	navel 3 months	valencia 3 months
52.6 54.2 49.8 51.5	51.0 48.0 49.6 48.4	52.5 52.0 51.8 53.6	50.8 48.8 52.1 49.9

The data are stored in `orange juice.csv`.

(a) Assess the interaction using a table of means and an interaction plot.

(b) Run an ANOVA to test the main effects and the interaction.

(c) Estimate $\hat{\sigma}$, the standard deviation of the error distribution.

(d) Assuming that higher levels of vitamin C in orange juice are desirable, make a recommendation based on the results of this experiment.