# Statistical Research Planning

## From basic concepts to complex experimental designs

Luc Duchateau

Academic year 2013-2014
Dept. Comparative Physiology and Biometrics
Ghent University
Belgium

# Part I

# Basic statistical concepts required for experimental design

# Chapter 1

# Basic concepts in statistical research planning

## 1.1 Introduction

In this first chapter, we will make a walk through the different steps typically taken when setting up an experiment. In that process, we will introduce important concepts in experimental design.

An important first step, discussed in Section 1.2, is to clearly define a research hypothesis, and translate it into a testable statistical hypothesis.
Next, the investigator has to think about the possible resources to run the experiment and test the relevant hypothesis. The experimental material is split up in different parts, and we discuss in Section 1.3 the concepts of experimental and observational units. We assess the response variable at the level of an observational unit, whereas treatments are randomly assigned to experimental units.
Another important step to consider is power analysis, i.e., assessing how many replications are required in the experiment in order to have a reasonable chance to reject the null hypothesis. The concept of replication is explained in Section 1.4.
The next question in setting up an experiment is whether the units can be grouped in a meaningful way, i.e., in blocks, so that part of the random variation between the experimental units can be explained. The reasoning behind blocking is explained in Section 1.5.
Once experimental units have been determined and possibly grouped in blocks, we need to assign treatments to experimental units in a random way. Randomisation is the cornerstone of experimental design and will be thoroughly discussed in Section 1.6. Randomisation enables the scientist to claim causal relationships.
Randomisation takes away part of the subjectivity in an experiment. Nevertheless, once the study is running, we must ensure that evaluation of the response variable is also done in an objective way. Blinding (of the investigator) and double blinding (including the subject) is discussed in Section 1.7. It is an excellent tool to ensure that no bias towards a particular treatment occurs.

## 1.2 Specifying statistical hypotheses

### 1.2.1 Scientific and statistical hypotheses

It is essential in experimental research to clearly define the hypothesis that needs to be tested, so that an adequate experiment can be set up that leads to a firm conclusion with respect to that hypothesis.
Once such an hypothesis is defined in scientific terms, we need to translate this in a statistical hypothesis (in terms of population parameters) that can be tested with an appropriate statistical test. This step is often not straightforward but needs to be taken before the start of the experiment to avoid useless experiments from which no conclusions can be drawn. We study in the next section different types of hypothesis, and also demonstrate that one and the same study can lead to different hypotheses.

**Example 1.1 Mastitis trial and relevant hypothesis specifications**

We want to investigate the effect of inoculation dose of *Escherichia coli* in the udder quarter on the somatic cell count (SCC) in the milk of dairy cows. We randomly assign 5 heifers to the low dose ($10^4$ colony forming units) and 5 heifers to the high dose ($10^6$ colony forming units). We assess the SCC in intervals of 3 hours in the next 24 hours. We wish to evaluate the effect of the inoculation dose on SCC. This is, however, not a workable hypothesis, because it is much too general and it does not enable us to write down a statistical hypothesis that can be tested. It is necessary to specify more precisely what is expected in terms of SCC if inoculation dose matters. Here is a list of a few possible hypotheses.

- We expect that the SCC over the 24 hours will be higher in the high dose group relative to the low dose group; we could therefore compare the averages (equals the area under the surface with equal time intervals) over the different measurements in time.
- We expect that the maximum SCC will be higher in the high dose group relative to the low dose group; we could therefore compare the maximum observed SCC over the 24 hours.
- We expect that the SCC will increase faster in the high dose group relative to the low dose group; we could therefore compare the (linear) increase in SCC in the two groups.
- We expect that the SCC will increase faster in the high dose group relative to the low dose group; as alternative for the previous hypothesis we could compare the time required before the SCC passes a certain (absolute or relative) threshold value.
- Sometimes it is not known upfront when to expect differences; an alternative is then to compare the two groups at each of the different time points. This is not efficient, as we have to correct for the fact that different hypotheses are tested simultaneously, and each individual hypothesis therefore needs to be tested at a more stringent significance level.

- In general, we can test in a global model whether there are any differences at all, comparing averages, and next whether the inoculation dose and time are interacting, i.e., whether SCC evolves differently in time according to the dose group.

It is obvious from this list of hypotheses that it is essential to define a clear hypothesis. The more specific the hypothesis, the more likely it will result in a significant result, if the alternative hypothesis is correct.

■

### 1.2.2 Types of hypotheses

#### Hypotheses concerning one parameter

Sometimes one wishes to test a hypothesis concerning one parameter. This parameter can be any characteristic of the population, but often it is a population mean or a population proportion. Other possible parameters are the slope of a linear regression line or the variance.

#### Example 1.2

Without vaccination, 30% of the animals survives an infection with a protozoan parasite, *Theileria parva*. In an experiment, we wish to prove that a particular vaccine protects more than 30% of the vaccinated animals. With $\pi$ the population proportion of survivors in the population after vaccination, the hypothesis becomes

$$\mathrm{H}_0 : \pi = 0.3 \text{ versus } \mathrm{H}_a : \pi > 0.3$$

■

#### Example 1.3

The white blood cells (WBC) are essential in the immune system defense of a cow against mastitis. The WBC should not only be present, but also active. Activity of WBC can be assessed by chemiluminesence. We wish to assess whether the activity of the WBC before infection has a positive influence on the evolution of mastitis. We investigate this hypothesis by correlating the WBC activity before infection with the proportional milk reduction measured 48 hours after infection. We wish to prove that milk production is less reduced with increasing activity of the WBC. Define $\beta$ as the proportional milk reduction per unit increase activity of the WBC (in units chemiluminesence); the hypothesis then corresponds to

$$\mathrm{H}_0 : \beta = 0 \text{ versus } \mathrm{H}_a : \beta < 0$$

■

**Example 1.4**

We need to assess the reproducibility of a new technique to quantify SCC. We therefore take a number of milk samples of the same cow at the same time and measure the SCC for each individual sample. We further assume that the measurements are well described by the normal distribution, and define the hypothesis in terms of the variance of the normal distribution. We wish to prove that the variance is smaller than a specific value, for instance $\sigma^2 = 100$, from which follows the hypothesis

$$\mathrm{H}_0 : \sigma^2 = 100 \text{ versus } \mathrm{H}_a : \sigma^2 < 100$$

■

**Hypotheses concerning two parameters**

In most experiments at least two treatments or populations are compared with one another. As in the previous section, most hypotheses are concerned with population means or proportions, but also the variances or the slopes of two populations can be compared with one another.

**Example 1.5**

Most vaccination experiments consist of two groups, a control group and a vaccinated group, and animals are randomly assigned to one of the two groups. One wishes to prove that more animals are surviving in the vaccinated group, compared to the control group. With $\pi_v$ the proportion survivors in the population of vaccinated animals and $\pi_c$ the proportion survivors in the population of control animals, the hypothesis is stated as

$$\mathrm{H}_0 : \pi_v - \pi_c = 0 \text{ versus } \mathrm{H}_a : \pi_v - \pi_c > 0$$

If one wants to prove that proportion survivors in the population of vaccinated animals is at least a certain percentage, e.g. 20%, above that of the population of control animals, the hypothesis becomes

$$\mathrm{H}_0 : \pi_v - \pi_c = 0.2 \text{ versus } \mathrm{H}_a : \pi_v - \pi_c > 0.2$$

■

**Example 1.6**

In the previous example concerning the relationship between the activity of the WBC before the infection and the reduction in milk production 48 hours after the infection with *E. coli*, measurements were taken both on heifers and multiparous cows. We wish to compare the relationship between heifers and multiparous cows. With $\beta_h$ ($\beta_m$) the proportional milk reduction per unit increase activity of the WBC (in units chemiluminesence) in heifers (multiparous cows), the hypothesis is stated as

$$\mathrm{H}_0 : \beta_h - \beta_m = 0 \text{ versus } \mathrm{H}_a : \beta_h - \beta_m \neq 0$$

■

**Example 1.7**

We have to choose between two techniques to quantify SCC, the microscopic technique or the automated Coulter Counter. For a number of milk samples of the same cow at the same time, SCC is assessed using the two techniques. We assume that the measurements are well described by a normal distribution, and we wish to test an hypothesis related to the variability of the two technqiues. With $\sigma_m^2$ ($\sigma_c^2$) the variance of the microscopic technique (Coulter Counter), the hypothesis becomes

$$\mathrm{H}_0 : \sigma_m^2 = \sigma_c^2 \text{ versus } \mathrm{H}_a : \sigma_m^2 \neq \sigma_c^2$$

■

**Hypotheses concerning more than two parameters**

There are often more than two treatments or populations included in an experiment. We can discern between experiments with only one factor, for instance dose of a drug, appearing at more than two levels and experiments with more than two factors, each at two or more levels. We first discuss the situation of one factor appearing at more than two levels.

*One factor with more than two levels*

In the case of one factor appearing at more than two levels, different relationships can exist -and can be tested for- between the different levels. The most general alternative hypothesis states that there is at least one pairwise difference amongst the different population means, without specifying which specific pair that is. If the levels are doses of a drug, we can test a slightly more specific hypothesis: we could test whether there exist differences between the control and any of the applied doses; we then need to adapt the significance level for each comparison as multiple comparisons are made. Different hypotheses will be put forward in the example below.

**Example 1.8**

The effect of progesterone on the diapedesis of white blood cells from the blood to the milk in the udder is investigated in an experimental setup, which consists of two compartments separated by an artificial membrane. A fixed number of white blood cells is put in the first compartment, whereas different concentrations of progesterone are established in the other compartment. We measure the number of migrated white blood cells after a defined amount of time. Apart from a control without progesterone, three different progesteron concentrations are used. Define the population mean of the number of migrated cells as $\mu_0, \mu_1, \mu_2$ and $\mu_3$ for control, first, second and third highest dose respectively.

The global null hypothesis is then

$$\mathrm{H}_0 : \mu_0 = \mu_1 = \mu_2 = \mu_3$$

versus the alternative hypothesis

$$\mathrm{H}_a : \mu_i \neq \mu_j \text{ for at least one pair } (i, j) \text{ with } i, j = 0, \ldots, 3$$

Any difference (at the level of the population) between concentrations should therefore lead to rejection of the null hypothesis. As this is a very general hypothesis, it is very well possible that a real and important difference, for instance between the highest dose and the control, is missed.

A more specific hypothesis consists of the comparison of each concentration with the control, specified as

$$\mathrm{H}_0 : \mu_0 = \mu_1; \mu_0 = \mu_2; \mu_0 = \mu_3$$

versus the alternative hypothesis

$$\mathrm{H}_a : \mu_i \neq \mu_0 \text{ for at least one concentration } i = 1, \ldots, 3$$

but we could also test each of the three hypotheses individually. In that case we have to correct for multiple comparisons as discussed before.
Finally, we could also decide to test whether there exists a linear relationship between the concentration and the number of migrated cells.
■

*More than one factor*

In the case of two or more factors, we distinguish between the main effects of the factors and the interaction effects between the factors. We should first test the interaction effect; in the case of no interaction, the main effects can be tested (see Chapter 3).

**Example 1.9**

For the induction of mastitis, two different doses, low and high, are used as before. The dosis, at a low and high level, is the first factor. Furthermore, we include the same number of heifers and multiparous cows in the experiment, and the parity, heifer versus multiparous cow, is the second factor. Define $\mu_{LH}$ en $\mu_{HH}$ as the mean SCC of the population of heifers with low (high) induction dose, and similarly $\mu_{LM}$ en $\mu_{HM}$ as the mean SCC of the population of multiparous cows with low (high) induction dose.

We first define the hypotheses related to the main effects. For the induction dose the hypothesis is

$$\mathrm{H}_0 : \frac{\mu_{LH} + \mu_{LM}}{2} = \frac{\mu_{HH} + \mu_{HM}}{2} \text{ versus } \mathrm{H}_a : \frac{\mu_{LH} + \mu_{LM}}{2} \neq \frac{\mu_{HH} + \mu_{HM}}{2}$$

We therefore test whether the induction dose has an effect, regardless the parity, or averaged over the parity.

For parity the hypothesis is

$$\mathrm{H}_0 : \frac{\mu_{LH} + \mu_{HH}}{2} = \frac{\mu_{LM} + \mu_{HM}}{2} \text{ versus } \mathrm{H}_a : \frac{\mu_{LH} + \mu_{HH}}{2} \neq \frac{\mu_{LM} + \mu_{HM}}{2}$$

We therefore test whether parity has an effect, regardless the induction dose, or averaged over the induction doses.

The hypothesis concerning the interaction between the two factors is

$$\mathrm{H}_0 : \mu_{LH} - \mu_{LM} = \mu_{HH} - \mu_{HM} \text{ versus } \mathrm{H}_a : \mu_{LH} - \mu_{LM} \neq \mu_{HH} - \mu_{HM}$$

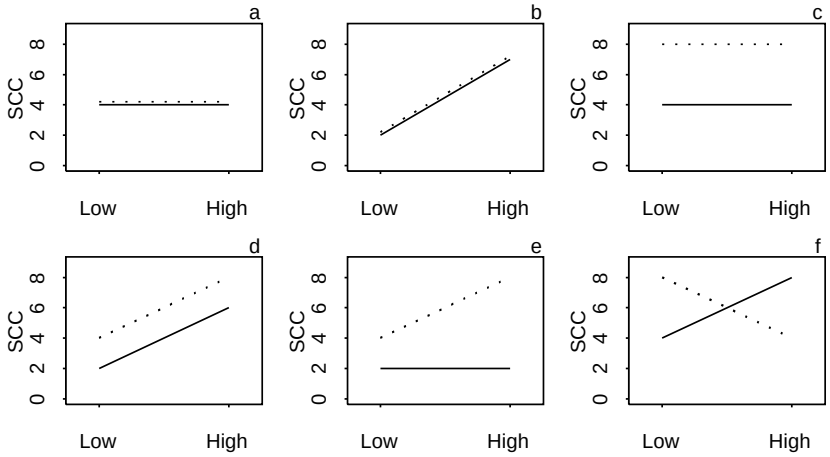
We therefore test whether the effect of parity is the same for the two induction doses. This interaction hypothesis can be reformulated as

$$\mathrm{H}_0 : \mu_{LH} - \mu_{HH} = \mu_{LM} - \mu_{HM} \text{ versus } \mathrm{H}_a : \mu_{LH} - \mu_{HH} \neq \mu_{LM} - \mu_{HM}$$

and now we test whether the effect of the induction doses is the same for heifers and multiparous cows. Both hypothesis tests give exactly the same result due to the symmetry of the two hypotheses.

We demonstrate cases of presence and absence of interaction in Figure 1.1.

Figure 1.1: Main effects and interactions in the mastitis experiment with parity and induction dose as factors. The solid line is related to the multiparous cows, the dotted line to heifers.




In Figure 1.1.a there is no effect of dose and no effect of parity, in Figures 1.1.b and 1.1.c there is only an effect of dose, respectively parity, in Figure 1.1.d there is an effect of dose and of parity, but no interaction between the two factors, in Figure 1.1.e both factors

have an effect and there is also interaction: there is a dose effect for heifers but not for multiparous cows. Finally a case is presented in Figure 1.1.f with interaction between the two factors in the absence of main effects.
■

## 1.3 Experimental and observational units

The experimental unit is an important concept in experimental design. It is the entity or part of the experimental material to which a treatment is randomly assigned. In many cases, the patient or the animal is the experimental unit, but it is not necessarily so. On the other hand, the observational unit is the entity on which the response variable is measured. Possibly, the experimental unit and the observational unit coincide. We demonstrate these wo concepts in Example 1.10.

**Example 1.10 Experimental and observational units**

We want to compare the effect of two drugs against trypanosomosis, Berenil and Samorin. The response variable corresponds to the increase in packed cell volume (PCV) observed in an animal. In a first setup, one herd is selected, and within the herd cows are randomly assigned to Berenil and Samorin. The cow is the experimental unit, and at the same time the observational unit. If it is logistically too complex to assign the two drugs within a farm to different cows, we can opt to select different herds, and randomly assign the herd as a whole to Berenil and Samorin, i.e., each cow within a herd receives the same treatment that is randomly assigned to the herd. In this setup, the herd is the experimental unit, whereas the cow remains the observational unit. ■

In many experiments, there is only one type of experimental unit, as in the two setups in Example 1.10. In more complex studies, however, random assignment can occur at two different levels, leading to two different types of experimental units. Experiment units of the first type are often grouped into an experimental unit of the second type, i.e., experimental units of the first type are nested in experimental units of the second type. We demonstrate this concept of nesting in Example 1.11.

**Example 1.11 Two types of experimental units in the same experiment**

We extend the setup of Example 1.10 with an additional factor, the dose of the drug: we wish to test the two drugs at a low and a high dose. The drugs Berenil and Samorin are assigned to the herd as a whole, but within a herd cows are randomly assigned to the high or the low dose. The cow remains the observational unit, but there are now two types of experimental units. The experimental unit of the first type is the cow, to which the dose is randomly assigned. The experimental unit of the first type, the cows, are nested within the experimental units of the second type, i.e., the herd. The nesting concept is used in experimental design in the following way. One type of experimental unit, say the first type, is nested within another type of experimental unit, say the second type, if an experimental unit of the second type consists of a number of experimental units of the first type, which are different from the experimental units of the first type included in another experimental unit of the second type. Applying this to our example, each herd consists

of a number of cows, and these cows are necessarily different from the cows in another herd.

The concept of nesting will also be used later when considering a treatment factor that is nested within another treatment factor, where it has exactly the same meaning as above. ■

## 1.4 Replication

### 1.4.1 Introduction

The objective of an experiment is to assess whether an observed treatment difference is caused by a real difference between the treatments. In order to do so, we need to have an idea of the variation that can be expected between such observed treatment differences if there is no difference at all between the treatments. Therefore, it is essential that at least two experimental units are randomly assigned to each of the treatments; this will enable us to estimate the inherent variability. Assume, for instance, that one wants to assess whether there is a difference between two diets, A and B, with respect to weight gain. One randomly selected chicken receives diet A, another one diet B, resulting in observations 1.8 kg and 2.0 kg respectively. The observed treatment difference thus is 0.2 kg. Having only these two observations, it is impossible to conclude that diets differ from each other as the difference could be merely due to the difference between the two chickens.

Assume that 4 chickens are used instead, with two randomly assigned to diet A and two to diet B, with the results given by 1.75 kg and 1.85 kg for diet A and 1.95 kg and 2.05 kg for diet B. The observed treatment difference equals 0.2 kg as before. We can now, however, evaluate this difference against the variability of chickens that have received the same treatment.

### 1.4.2 Replication versus repeated measurements

It is essential to distinguish between replications and repeated measures. When two treatments are compared in a simple experiment, then a random assignment of the treatment to an experimental unit results in a replication. If the experimental unit and the observational unit coincide, we have only replications and no repeated measurements. If, however, different observations are scored on an experimental unit, separated either in time or space, then the experimental unit consists of different observational units. These observations on one and the same experimental unit are repeated measurements and not replications. Such repeated measurements allow us to have a better assessment of the response variable for the experimental unit (a mean is better than a single observation), but do not provide us with information about the variability between experimental units against which we have to test the treatment effects.

As these concepts of replication and repeated measurement are crucial in setting up efficient experiments, two different examples of repeated measures are provided here, with repeated measurements either spread in time, as in Example 1.12, or spread in space, as in Example 1.13.

**Example 1.12 Repeated measures in time**

We want to investigate the effect of a vaccine against trypanosomosis. One randomly selected cow receives the vaccine, another cow a control injection. Both cows are now followed up daily for their PCV values. The observational unit is now a PCV measurement for a cow at a particular moment in time. The experimental unit, however, is the cow. Thus the PCV measurements within a cow are repeated measurements and not replications. No matter how large the difference between the PCV values of the two cows, or no matter how many repeated measurements are taken over time, we should never conclude from such an experiment without replications that the vaccine has an effect.
■

**Example 1.13 Repeated measures in space**

A vaccination trial for mastitis is set up in two farms. In a first farm, 15 heifers are chosen and vaccinated, whereas in a second farm, we randomly chose 15 multiparious cows and also vaccinate these cows. Assume that we want to compare the effect of the vaccine between heifers and multiparious cows. It is clear that no replications are available for this factor. Parity has been randomly assigned to a farm; thus the farm is the experimental unit. The cow is the observational unit, but the cows within the farm are repeated measurements, not replications. For a particular farm, with 15 cows or observational units, we can get a precise estimate of the vaccine effect, based on these repeated measurements. The repeated measurements, however, can not be used to assess the variability between the farms, i.e., the experimental units. If big differences are seen between the farms, this can either be due to the difference between heifers and multiparous cows, but also be due to the inherent differences between the two farms. In order to draw conclusions with respect to the difference between heifers and multiparous cows, we need to have at least two farms at each level of the parity factor. Alternatively, we could choose to have heifers and multiparous cows within the same farm, which makes the farm a block factor, and the cow at the same time the experimental and observational unit. ■

### 1.4.3 Sample size determination

**Objectives of sample size determination**

The objective of most experiments is to draw a clear conclusion based on the results generated by the experiment. A clear conclusion follows when the null hypothesis can be rejected. If the null hypothesis can not be rejected it does not mean that the null hypothesis is true. Assuming that we compare two treatments, non-rejection of the null hypothesis can be due to different reasons

1. the null hypothesis is correct.
2. bad luck, a result in accordance with the null hypothesis is observed, both sample means are the same, whereas there is a difference between the population means.
3. we observe a large difference between the sample means of the treatments but the sample size is too small (or the variance too big) to conclude that this difference is

caused by more than chance alone.

For reason [1], the right conclusion is drawn. We can never exclude reason [2], but its probability decreases with larger sample size or a larger difference between the population means. We have control over reason [3], as we can chose our sample size, or number of replications, in such a way that particular differences of interest between sample means will lead to rejection of the null hypothesis. We need to chose a sample size in such a way that we will reject the null hypothesis with high probability (typically 80 %) if the difference between the population means, $\Delta$, is equal to or more than a relevant difference. For instance, we are not interested in showing that diet A leads to 10 grams more weight gain than diet B, as this has no pratical relevance; it would be difficult anyway, requiring lots of animals. But differences starting from, say, 200 grams could be practically relevant.

**Parameters in sample size determination**

The probability that the null hypothesis is rejected depends on a number of parameters

1. The type I error $\alpha$.
2. The true difference under the alternative hypothesis, for the case of comparing two population means $\Delta = \mu_1 - \mu_2$.
3. The variance between the experimental units $\sigma^2$.
4. The sample size n.

We want this probability, presented by $1 - \beta$ and called power, to be sufficiently large. We wish to determine the optimal sample size for the experiment, not too small nor too large. The type I error is usually fixed at 5 %. The objective of the experiment is usally to accept a particular alternative hypothesis, from which follows the true difference between population means under the alternative hypothesis $\Delta$. The remaining parameter is $\sigma^2$, for which we have to propose a particular (set of) values based on previous experiments or experience.

**Sample size determination for the comparison of two population means**

The hypothesis for the comparison of two population means, $\mu_1$ en $\mu_2$, assuming that the underlying variable is normally distributed, is given by

$$\mathrm{H}_0 : \mu_1 - \mu_2 = \mu_0$$

We typically assume that the variances of both populations are known. We discern between one-sided and two-sided alternative hypotheses.

*One-sided alternative hypothesis*

The one-sided alternative hypothesis (we only consider one of the two possibilities here) is written as

$$\mathrm{H}_a : \mu_1 - \mu_2 > \mu_0$$

The relevant test statistic is the difference between the sample means $\bar{X}_1 - \bar{X}_2$, with distribution under the null hypothesis

$$\bar{X}_1 - \bar{X}_2 \sim \mathrm{N}\left(\mu_0; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

It follows that the decision rule for rejection of the null hypothesis at significance level $\alpha$ is

$$\text{Reject } \mathrm{H}_0 \text{ if } \frac{\bar{X}_1 - \bar{X}_2 - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq z_{1-\alpha}$$

where $z_{1-\alpha}$ corresponds to that value of the standard normal distribution for which there is a probability of $1-\alpha$ to find a smaller value, i.e., $\mathrm{P}(Z < z_{1-\alpha}) = 1-\alpha$.

This rejection rule can be reworked in terms of the difference between the two sample means

$$\text{Reject } \mathrm{H}_0 \text{ if } \bar{X}_1 - \bar{X}_2 \geq \mu_0 + z_{1-\alpha} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = C_u$$

The power can now be derived for a specific difference between the two population means under the alternative hypothesis, $\mu_1 - \mu_2 = \Delta$. It corresponds to the probablity that $\bar{X}_1 - \bar{X}_2$ is located in the critical region described in (1.4.3) given that $\mu_1 - \mu_2 = \Delta$.

The power for $\mu_1 - \mu_2 = \Delta$ is thus given by

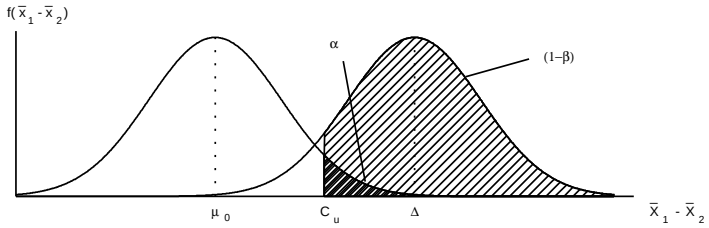
$$\mathrm{P}\left(\bar{X}_1 - \bar{X}_2 \geq \mu_0 + z_{1-\alpha} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \mid \mu_1 - \mu_2 = \Delta\right) \tag{1.1}$$

which is demonstrated in Figure 1.2.
The curve on the left hand side in Figure 1.2 is the distribution of $\bar{X}_1 - \bar{X}_2$ under the null hypothesis; the critical region corresponds to all values larger than $C_u$. The area under this distribution to the right of $C_u$, denoted by $\alpha$, is the probability of falsely rejecting the null hypothesis. The curve on the right hand side in Figure 1.2 is the distribution of $\bar{X}_1 - \bar{X}_2$ under the alternative hypothesis $\mu_1 - \mu_2 = \Delta$. The area under this curve to the right of $C_u$ corresponds to the probability that $\bar{X}_1 - \bar{X}_2$ takes a value in the critical region, and is thus the power.

In most cases, we have to determine the sample size to attain the proposed power. We then have to rework (1.1). We denote the required power by $1-\beta$. We then have

$$\begin{aligned} 1-\beta &= \mathrm{P}\left(\bar{X}_1 - \bar{X}_2 \geq \mu_0 + z_{1-\alpha} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \mid \mu_1 - \mu_2 = \Delta\right) \\ &= \mathrm{P}\left(\frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq z_{1-\alpha} + \frac{\mu_0 - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) \end{aligned}$$

Figure 1.2: Power calculation for a one-sided hypothesis




from which follows that

$$z_\beta = z_{1-\alpha} + \frac{(\mu_0 - \Delta)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

If we take equal sample size for the two populations, i.e., $n_1 = n_2 = n$, and equal variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$ we have

$$n = \frac{2\left(z_\beta + z_\alpha\right)^2 \sigma^2}{\left(\mu_0 - \Delta\right)^2}$$

*Two-sided alternative hypothesis*

The two-sided alternative hypothesis is written as

$$\mathrm{H}_a : \mu_1 - \mu_2 \neq \mu_0$$

The rejection region can be derived in a similar fashion as before and is

$$\text{Reject } \mathrm{H}_0 \text{ if } \mid \bar{X}_1 - \bar{X}_2 \mid \geq \mu_0 + z_{1-\alpha/2} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{1.2}$$

The power can be derived, as before, for a specific difference between the two population means under the alternative hypothesis, $\mu_1 - \mu_2 = \Delta$. It corresponds to the probablity that $\bar{X}_1 - \bar{X}_2$ is located in the critical region described in (1.2) given that $\mu_1 - \mu_2 = \Delta$.

The power for $\mu_1 - \mu_2 = \Delta$ is thus given by

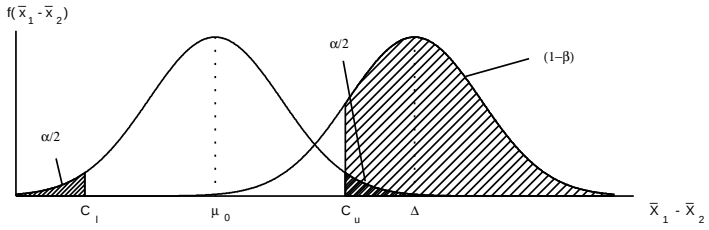
$$1 - \mathrm{P}\left(\mu_0 + z_{\alpha/2} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \bar{X}_1 - \bar{X}_2 \leq \mu_0 + z_{1-\alpha/2} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \mid \mu_1 - \mu_2 = \Delta\right) \tag{1.3}$$

which correponds to

$$\mathrm{P}\left(\bar{X}_1 - \bar{X}_2 \leq \mu_0 + z_{\alpha/2} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \mid \mu_1 - \mu_2 = \Delta\right) + \mathrm{P}\left(\bar{X}_1 - \bar{X}_2 \geq \mu_0 + z_{1-\alpha/2} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \mid \mu_1 - \mu_2 = \Delta\right)$$

as demonstrated in Figure 1.3.

Figure 1.3: Power calculation for a two-sided hypothesis




The biggest difference is that the critical regions splits up in two different regions: on both sides of the density function under the null hypothesis, an area under the curve equal to $\alpha/2$ is taken. This leads to lower power for the two-sided test for the same data set when the alternative hypothesis is true. The area denoted by $(1-\beta)$ in Figure 1.3 corresponds to the probability that the observed test statistic $\bar{X}_1 - \bar{X}_2$ is located in the critical region, and is thus the power. There is a very small probability that the observed test statistic is located in the critical region on the other side, left from the acceptance region, but that probability is negligible and is not added in the power calculations.
As before, the power expression (1.3) can be rewitten in terms of the required sample size

$$n = \frac{2\left(z_\beta + z_{\alpha/2}\right)^2 \sigma^2}{\left(\mu_0 - \Delta\right)^2}$$

**Example 1.14 Power calculation for the comparison of two means**

We want to set up an experiment to compare the effect of two trypanocidal drugs on the packed cell volume (PCV) of cows having trypanosomosis. We want to determine the required number of animals so that there is a probability (power) equal to 90 % to detect a significant difference at the 5% significance level if we assume that the real (population) difference in PCV equals 2% and that the population variance for both groups is equal to 2.
As $z_{0.1} = -1.28$ and $z_{0.025} = -1.96$, we have for a two-sided hypothesis

$$n = \frac{2\left(-1.28 - 1.96\right)^2 2}{\left(2\right)^2} = 10.5$$

We will therefore use 11 animals per group.

For a one-sided hypothesis test we find

$$n = \frac{2\left(-1.28 - 1.645\right)^2 2}{\left(2\right)^2} = 8.56$$

requiring a sample size of 9 animals per group.
■

## 1.5 Blocking

Blocking is an important tool in experimental design. It allows the investigator to reduce the variability against which the treatment effects need to be compared. We will explain the principle of blocking in this short section.

We typically have inherent variability amongst the experimental units. Even if all the experimental units would receive the same treatment, they would still differ with respect to the observed value for the response variable (assuming for the time being that the experimental unit coincides with the observational unit). This variability between the experimental units with the same treatment is the inherent variability in the experiment.

Experimental units, however, can sometimes be grouped in a meaningfull way into blocks, so that experimental units within a block share some characteristic(s) and are therefore more alike compared to experimental units from another block. If random assignment of the treatment factor is done in a proper way, it is possible to explain part of the variability between the experimental units, thereby reducing the variability against which treatment effects need to be compared. A proper random assignment of the treatment factor ensures that different treatments appear in the same block; this means that treatment differences can be assessed within a block of experimental units that resemble each other more. In the simplest setting, the randomised complete block design (see Chapter 4), each treatment occurs exactly once in each block. In Example 1.15 a somewhat more complex application of blocks is discussed.

**Example 1.15 Blocking**

We wish to study the effect of two doses of Berenil with respect to the recovery of cows after a trypanosoma infection. Different herds are included in the study as we wish to draw general conclusions. Cows within the same herd resemble each other more and are also kept under the same management, so that herd can be considered to be a block factor. Therefore, animals within a herd (block) are randomly assigned to low and high dose. If the number of animals assigned to each of the treatments is the same and higher than 1, as in this example, we have a generalised complete block design (see Chapter 4). The use of blocks will enable us to filter out the variation between herds. We thus investigate the effect of a low (L) and a high (H) dose and cows are evaluated for PCV one month after the treatment which results in the data in Table 1.1.

Table 1.1: Difference in PCV (%) just before and one month after treatment with a high or low dose of Berenil of cows with a trypanosoma infection

| Herd 1 | | Herd 2 | | Herd 3 | |
|---|---|---|---|---|---|
| L | H | L | H | L | H |
| 0.9 | 7.4 | 2.3 | 6.6 | 3.4 | 8.8 |
| 2.0 | 6.8 | 2.7 | 7.1 | 2.7 | 8.3 |
| 2.0 | 7.1 | 1.3 | 6.2 | 2.9 | 7.9 |
| 2.2 | 6.7 | 1.6 | 7.8 | 3.0 | 8.2 |
| 2.0 | 7.9 | 2.1 | 7.2 | 3.4 | 8.1 |

■

## 1.6 Randomisation

### 1.6.1 The concept of simple randomisation

The concept of randomisation has been introduced by Fisher (1935). Simple randomisation goes as follows. Assume we want to compare two treatments A and B in an experiment, without taking into account any other factors. We will then randomly assign a treatment to an experimental unit with a particular probability; often this probability of assignment is the same for both treatments, and thus equal to 50 %, but we might diverge from that if we would like to have relatively more observations on one treatment. In simple randomisation, the previous random assignments should not have an effect on the current random assignment. Therefore, each random assignment has exactly the same probability if the two treatments have the same assignment probability. For instance, the following 3 treatment assignments to 5 experimental units

A A A A A
B B B B B
A B A B A

have exactly the same probability of occurrence.

It is obvious that the two first sequences are useless if we want to compare the treatments A and B; that is why we use restricted randomisation schemes explained in the next section.

The use of random assignment ensures that other variables that could have an impact on the response variable are distributed randomly over the two treatment groups. Randomisation is necessary if an experimenter wants to demonstrate a causal relationship between a treatment factor and a response variable. If the treatment assignment would be chosen by the investigator, there is a serious risk that the investigator, consciously or not, assigns patients or subjects with particular characteristics to particular treatments. If a significant difference is found, it is unclear whether this is due to the treatment or to the subjective assignment of the treatment. This last phenomenon is called selection bias.

Alternatively, it can be said that the treatment factor is confounded with certain patient characteristics. In extreme cases, it could be that an investigator assigns the patients with a better prognosis to the new treatment to try to show that the new treatment is better than the standard.

It is a well known fact that non-randomised trials lead to more significant results than randomised trials in the same setting.

The same reasoning as above can be applied to observational studies, where risk factors are not randomly assigned, but merely observed. At best, we can conclude that there exists a significant association between the risk factor and the outcome variable, but not that there is a causal relationship: the risk factor might be correlated with the real causal factor.

### 1.6.2 Restricted randomisation

As stated in the previous section, we normally don't use the simple randomisation scheme, because it could lead to serious imbalances in sample size of the two treatment groups. For instance, in a group of 4 patients, the probability of having 2 in each treatment group can be derived from the binomial distribution. With $n_A$ the number of patients in treatment group A, and assuming equal assignment probability to groups A and B, we have

$$\mathrm{P}(n_A = 2) = \binom{4}{2}(0.5)^4 = 0.375 \tag{1.4}$$

So on average five in the eight trials will have an imbalance in sample size.

In a restricted randomisation scheme, we force the numbers of patients to be equal in the two groups. We can therefore no longer assign a treatment to each experimental unit independently. We rather look at all possible sequences with equal sample size in the two groups and chose one of the sequences randomly. For instance, for 4 experimental units with 2 treatments, we have the following six possible treatment assignments

A A B B
A B A B
A B B A
B B A A
B A B A
B A A B

and we chose one of these 6 sequences for the experiment. We have therefore reduced the number of possible sequences substantially, compared to the simple randomisation scheme, which has 16 possible sequences.

Restricted randomisation is also typically used in block designs. For instance, in a randomised complete block design (see Chapter 4), we want each treatment to occur exactly once in each block. For 4 treatment in blocks of size 4, for instance, there are in total 24 different possible sequences from which we have to chose.

### 1.6.3 Randomisation tests

The concept of randomisation also led to the first and most basic statistical tests, the Fisher exact test just being one of the famous examples.

Whenever the randomisation scheme is known, a randomisation test can be derived in a straightforward manner. The only requirement is to find an appropriate test statistic for the problem at hand. We demonstrate the principles of the randomisation test in Example 1.16.

**Example 1.16 Randomisation tests for insecticide resistance**

Assume that we want to compare the resistance of Anopheles mosquitos that transmits malaria against two insecticides A and B. We put 20 mosquitos in a recipient, to which we add one of the two insecticides in a random manner. Next, we count the number of survivors in each recipient. The more survivors, the more resistant the Anopheles moquito is. For 6 recipients, 3 with insecticide A and 3 with insecticide B, the data are shown in Table 1.2.

Table 1.2: Number of surviving mosquitos (out of 20) according to insecticide

| Insecticide | Number of surviving mosquitos | | |
|---|---|---|---|
| A | 8 | 13 | 18 |
| B | 3 | 7 | 5 |

Assume we want to test the hypothesis

$$\mathrm{H}_0 : \mu_A = \mu_B \text{ versus } \mathrm{H}_a : \mu_A > \mu_B$$

with $\mu_A$ ($\mu_B$) the average number of mosquitos surviving out of 20 when using insecticide A (B).

We will use the simplest summary statistic possible, i.e., the difference between the mean number of surviving mosquitos in the two treatment groups, denoted by $D$. In the experiment we observe $d = 13 - 5 = 8$.

Whenever $H_0$ is correct, i.e., there are no differences between the two insecticides, then each permutation of the treatment assignment is equally likely. The idea behind it is that the sequence of observations would have remained exactly the same, even if another treatment assignment would have been used. In total, there are $\binom{6}{3} = 20$ different permutation assignments, each with the same probability of occurrence according to the applied randomisation scheme, i.e., we need 3 experimental units assigned to one treatment and 3 to the other. Therefore, also each difference $d$, linked to a particular permutation has the same probability of occurrence under $H_0$. All 20 permutations are shown in Table 1.3, with corresponding difference $d$. Based on Table 1.3, we can construct the exact distribution of $D$ (Table 1.4).

Table 1.3: All 20 possible permutations of treatment assignment in the mosquito experiment

| | Insecticide | | | | | | |
|---|---|---|---|---|---|---|---|
| Permutation | A | | | B | | | d |
| 1 | 3 | 5 | 7 | 8 | 13 | 18 | -8.00 |
| 2 | 3 | 5 | 8 | 7 | 13 | 18 | -7.33 |
| 3 | 3 | 5 | 13 | 7 | 8 | 18 | -4.00 |
| 4 | 3 | 5 | 18 | 7 | 8 | 13 | -0.67 |
| 5 | 3 | 7 | 8 | 5 | 13 | 18 | -6.00 |
| 6 | 3 | 7 | 13 | 5 | 8 | 18 | -2.67 |
| 7 | 3 | 7 | 18 | 5 | 8 | 13 | 0.67 |
| 8 | 3 | 8 | 13 | 5 | 7 | 18 | -2.00 |
| 9 | 3 | 8 | 18 | 5 | 7 | 13 | 1.33 |
| 10 | 3 | 13 | 18 | 5 | 7 | 8 | 4.67 |
| 11 | 5 | 7 | 8 | 3 | 13 | 18 | -4.67 |
| 12 | 5 | 7 | 13 | 3 | 8 | 18 | -1.33 |
| 13 | 5 | 7 | 18 | 3 | 8 | 13 | 2.00 |
| 14 | 5 | 8 | 13 | 3 | 7 | 18 | -0.67 |
| 15 | 5 | 8 | 18 | 3 | 7 | 13 | 2.67 |
| 16 | 5 | 13 | 18 | 3 | 8 | 7 | 6.00 |
| 17 | 7 | 8 | 13 | 3 | 5 | 18 | 0.67 |
| 18 | 7 | 8 | 18 | 3 | 5 | 13 | 4.00 |
| 19 | 7 | 13 | 18 | 3 | 5 | 8 | 7.33 |
| 20 | 8 | 13 | 18 | 3 | 5 | 7 | 8.00 |

The P-value is given by the probability that we observe a similar or more extreme result

Table 1.4: Exact distribution for differences in mean number of mosquitos surviving when treated with the two insecticides

| d | P(D=d) |
|---|---|
| -8 | 0.05 |
| -7.33 | 0.05 |
| -6 | 0.05 |
| -4.67 | 0.05 |
| -4 | 0.05 |
| -2.67 | 0.05 |
| -2 | 0.05 |
| -1.33 | 0.05 |
| -0.67 | 0.1 |
| 0.67 | 0.1 |
| 1.33 | 0.05 |
| 2 | 0.05 |
| 2.67 | 0.05 |
| 4 | 0.05 |
| 4.67 | 0.05 |
| 6 | 0.05 |
| 7.33 | 0.05 |
| 8 | 0.05 |

than in the experiment, given that the null hypothesis is true, i.e., $D$ is distributed as in Table 1.4. The P-value is thus given by $\mathrm{P}(D \geq 8) = 0.05$. When testing at the 5 % significance level, we can just reject the null hypothesis.
Consider now the same experiment, but assume that it runs over 3 days. At each day, two recipients are taken and one is randomly assigned to insecticide A, the other to B leading to the data in Table 1.5.

Table 1.5: Number of surviving mosquitos (out of 20) according to insecticide and day

| | Number of surviving mosquitos | | |
|---|---|---|---|
| Insecticide | Day 1 | Day 2 | Day 3 |
| A | 8 | 13 | 18 |
| B | 3 | 7 | 5 |

The randomisation scheme has now changed, and fewer randomisation schemes are possible now, namely 8. We will use the same statistic, but the possible permutations with corresponding difference now leads to Table 1.6
It is clear that each value $d$ for $D$ appears only for one permutation, so that each unique value of $D$ has the same probability equal to 0.125. Therefore, the P-value is given by $\mathrm{P}(D \geq 8) = 0.125$ and we can no longer reject the null hypothesis.

Table 1.6: All 8 possible permutations of treatment assignment in mosquito experiment

| | Insecticide | | | | | | |
|---|---|---|---|---|---|---|---|
| Permutation | A | | | B | | | d |
| 1 | 8 | 13 | 18 | 3 | 5 | 7 | 8.00 |
| 2 | 3 | 13 | 18 | 5 | 7 | 8 | 4.67 |
| 3 | 7 | 8 | 18 | 3 | 5 | 13 | 4.00 |
| 4 | 5 | 8 | 13 | 3 | 7 | 18 | -0.67 |
| 5 | 3 | 7 | 18 | 5 | 8 | 13 | 0.67 |
| 6 | 3 | 5 | 13 | 7 | 8 | 18 | -4.00 |
| 7 | 5 | 7 | 8 | 3 | 13 | 18 | -4.67 |
| 8 | 3 | 5 | 7 | 8 | 13 | 18 | -8.00 |

■

### 1.6.4 And after randomisation?

After having randomised subjects to a particular treatment, it is important to ensure that no other factors in the course of the experiment are confounded with the treatment factor. This is probably one of the most frequently made errors. From randomisation up to the observation of the outcome, we have to avoid that biases are appearing in the treatment assessment. We consider two examples below where the assessment of the treatment was biased due to decisions taken after randomisation.

**Example 1.17 Grouping experimental units with the same treatment together**

Assume that we want to study the water uptake of chickens after vaccination. Ten chickens are randomly assigned to the vaccine, whereas 10 other chickens are injected with distilled water (sham control). As the vaccine is experimental and is based on an attenuated virus, the investigator decided to keep all the vaccinated chickens in one stable, and the control chickens in another stable. Although each chicken is housed individually to record the water uptake of the chicken, the vaccinated chickens all share the same environment, which is different from the control chickens. Therefore, the stable effect is fully confounded with the treatment effect. Observed differences between the two groups of chickens could be due to the treatment but also due to the stable.

■

**Example 1.18 Assessing yield according to treatment**

Three different wheat varieties are investigated for their yield. A randomised complete block designs is set up in the field. In each of the three blocks, there are three plots; one variety is randomly assigned to one of the plots within the block. At the time of harvest, it is impossible to harvest all plots at the same day. To avoid confusion, the investigator decides to first harvest variety A, on the next day variety B and finally on the third day

variety C. It is obvious that variety C had one or two extra growth days, so observed differences could be due to the longer growth period if it is found that variety C is best.

A correct approach is to use the block also as harvesting day. That way, the block factor contains extra variation, i.e., the variation due to harvesting at different days.
■

## 1.7 Blinding

Some outcomes, such as death, can be assessed in an unambiguous way. Other outcomes, however, have a more subjective nature. Consider, for instance, the wellfare of farm animals, where the investigator has to give a score according to the general appearance of the animal. Such a score is based on the subjective assessment of the observer. For such outcomes, it is important to blind the investigator to the treatment of the animal or the intervention that took place. Optimally, the person giving the treatment differs from the person assessing the outcome, and the last one is not aware of the treatment.
Obviously, such blinding is only possible if the applied treatment cannot be observed by the assessor. In human medicine, the golden standard is the double blind study, in which both the medical doctor and the patient are blinded for the assigned treatment. Placebo effects can be substantial in clinical trials, and the only correct way around this placebo effect is blinding. In cancer clinical trials, blinding is very well possible when comparing two different chemotherapies, but cannot be done when one wants to compare, for instance, surgery with chemotherapy.

# Chapter 2

# Fixed effects model with one factor - one way ANOVA

## 2.1 Introduction

In this chapter we discuss models which include only 1 factor, appearing at different levels. We investigate factors with at least 3 different levels; t-tests can be applied when only two levels are present.
When only fixed effect factors are considered, it is assumed that the effects of the particular levels are constant and of interest by themselves. These models are therefore named fixed effects models.

The fixed effects model is constructed in Section 2.2. In Section 2.3 estimators for the model parameters are proposed. Section 2.4 deals with testing of general hypotheses, i.e., are all factor effects equal or not? Section 2.5 deals with more specific hypotheses, using pairwise comparisons, contrasts and lineair combinations. Finally, diagnostic tools for investigating particular ANOVA model assumptions are discussed in Section 2.6.

## 2.2 Model specification for the fixed effects model with one factor

We introduce the model specification of the fixed effects model with 1 factor using the following example.

**Example 2.1 Weight gain in 4 breeds of chicken**

In an experiment, 4 breeds of chicken are compared with respect to their weight gain over 8 weeks. The experiment is based on 5 chickens per breed. The weights (in kg) are shown in Table 2.1.
■

Table 2.1: Weight gain of 5 chickens of 4 different breeds over 8 weeks.

| Breed 1 | Breed 2 | Breed 3 | Breed 4 |
|---|---|---|---|
| 1.56 | 1.38 | 1.49 | 1.46 |
| 1.54 | 1.41 | 1.54 | 1.49 |
| 1.50 | 1.44 | 1.48 | 1.44 |
| 1.49 | 1.37 | 1.51 | 1.52 |
| 1.51 | 1.40 | 1.48 | 1.49 |

The respons variable, weight gain, for the $j^{\text{th}}$ ($j = 1, \ldots, 5$) chicken of the $i^{\text{th}}$ ($i = 1, \ldots, 4$) breed is modeled as

$$Y_{ij} = \mu_i + e_{ij} \tag{2.1}$$

where $\mu_i$ is the population mean of the $i^{\text{th}}$ breed and $e_{ij}$ is the deviation of the observation from its population mean. A further assumption for the $e_{ij}$'s is that they are mutually independent and normally distributed N(0,$\sigma^2$).
This model is named the cell means model because the model parameters correspond to the population means.
The number of levels of the factor is denoted by $a$, and the number of observations of the $i^{\text{th}}$ level by $n_i$. The total number of observations is

$$n_{\cdot} = \sum_{i=1}^{a} n_i$$

Given the definition of the cell means model above, a number of model characteristics follow:

- The observed value $Y_{ij}$ is the sum of 2 components: a constant term $\mu_i$ and a random error term $e_{ij}$.
- As E($e_{ij}$)=0, it follows that E($Y_{ij}$) = $\mu_i$ (for the meaning of the operator E(.), see note 2.1).
- As $\mu_i$ is a constant, it follows that Var($Y_{ij}$) = Var($e_{ij}$) = $\sigma^2$. All observations have the same variance regardless the factor level.
- As $e_{ij}$ is normally distributed, it follows that $Y_{ij}$ is normally distributed, as $Y_{ij}$ is a lineair combination of $e_{ij}$ (see note 2.2).
- The random error terms $e_{ij}$ are assumed to be mutually independent. Therefore, the value of one random error term does not influence the value of another random error term. As the $e_{ij}$'s are mutually independent, it follows that the $Y_{ij}$'s are mutually independent.
- From the above model characteristics, it follows that the $Y_{ij}$'s are mutually independent with distribution N($\mu_i, \sigma^2$).

The model for Example 2.1 is represented graphically in Figure 2.1. We depict the population means (generally not known) corresponding to $\mu_1 = 1.53$, $\mu_2 = 1.39$, $\mu_3 = 1.50$ and $\mu_4 = 1.47$ in Figure 2.1 and use the standard deviation of the population $\sigma$ equal to 0.028.

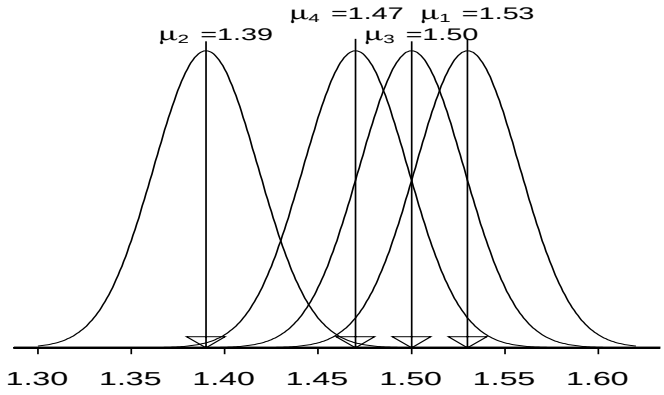



Figure 2.1: Representation of the cell means model for the fixed effects model with 1 factor. The 4 population means equal $\mu_1 = 1.53$, $\mu_2 = 1.39$, $\mu_3 = 1.50$ and $\mu_4 = 1.47$. The populations are characterized by a normal distribution with different population means but the same standard deviation $\sigma = 0.028$.

An alternative for the cell means model is the factor effects model. This alternative model representation is a somewhat more complex way to describe the same data. In the remainder, however, the factor effects model is the preferred model. In more complex data structures, it is far easier to define the hypotheses of interest in the factor effects model compared to the cell means model. The factor effects model is therefore introduced in the simple context of the fixed effects model with 1 factor.

The factor effects model is given by

$$Y_{ij} = \mu + \alpha_i + e_{ij} \tag{2.2}$$

where

| | |
|---|---|
| $\mu$ | a constant, common for all observations |
| $\alpha_i$ | a constant, the effect of the $i^{\text{th}}$ factor level |
| $e_{ij}$ | the random error term, independent and N(0,$\sigma^2$) |

We name this model the factor effects model because it is expressed in terms of the effects of the factor levels. These models are more complex because they are overparameterised;

it means that the model contains too many parameters compared to the information that needs to be described. This can be easily seen by comparing the factor effects model with the cell means model. In the latter model, each factor level leads to one parameter, the population mean. Therefore, the cell means model contains $a$ parameters. In the factor effects model, we have the general population mean $\mu$, and on top a parameter for each factor level leading to $a + 1$ parameters. The factor effects model thus has 1 parameter too much. Due to this reason, parameter restrictions need to be added to the model specification; otherwise the meaning of the parameters is unclear and not unique. Although many different types of restrictions can be used and are actually used in practice, we will consider only one restriction type, as it leads to the most straightforward parameter interpretation.

The following parameter restriction is used: put the sum of all factor effects equal to zero, or

$$\sum_{i=1}^{a} \alpha_i = 0 \tag{2.3}$$

from which it follows that the general population mean $\mu$ corresponds to the mean of the different population means

$$\mu = \sum_{i=1}^{a} \frac{\mu_i}{a}$$

as

$$\sum_{i=1}^{a} \mu_i = a\mu + \sum_{i=1}^{a} \alpha_i = a\mu$$

## 2.3 Estimating the model parameters of the fixed effects model with 1 factor

The traditional way to estimate the population means in the cell means model is based on the least squares (LS) criterion.

We first introduce some new notation.

The sum of the observations of the $i^{\text{th}}$ level of the factor is denoted as $Y_{i.}$, i.e.,

$$Y_{i.} = \sum_{j=1}^{n_i} Y_{ij}$$

The sample mean for the $i^{\text{th}}$ level of the factor is denoted as $\bar{Y}_{i.}$, i.e.,

$$\bar{Y}_{i.} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} = \frac{Y_{i.}}{n_i}$$

The sum of all the observations is denoted as $Y_{..}$, i.e.,

$$Y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{n_i} Y_{ij}$$

which leads to the overall sample mean $\bar{Y}_{..}$, i.e.,

$$\bar{Y}_{..} = \frac{\sum_{i=1}^{a} \sum_{j=1}^{n_i} Y_{ij}}{n_.} = \frac{Y_{..}}{n_.}$$

The LS estimator of $\mu_i$ can then be obtained by minimising the LS criterion

$$Q = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2$$

This criterion can be minimised by taking the first partial derivative

$$\frac{dQ}{d\mu_i} = \sum_{j=1}^{n_i} (-2) (Y_{ij} - \mu_i)$$

and equating to zero (where now $\mu_i$ is replaced by its estimator $\hat{\mu}_i$)

$$2 \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i) = 0$$

leading to

$$\hat{\mu}_i = \bar{Y}_{i.}$$

The LS estimator of the population mean thus equals the sample mean from that population.

In a similar way, and applying the restriction (2.3), we find for the factor effects model

$$\begin{aligned} \hat{\mu} &= \bar{Y}_{..} \\ \hat{\alpha}_i &= \bar{Y}_{i.} - \bar{Y}_{..} \text{ for } i = 1, \ldots, a \end{aligned}$$

## 2.4 The general hypothesis test

To test the general hypothesis of no differences between the different factor levels, we make use of sums of squares, more specifically of a ratio of sums of squares. We first introduce these sums of squares.

The starting point is the deviation of the observation from the overall mean, $Y_{ij} - \bar{Y}_{..}$. This deviation can be written as the sum of two terms

$$Y_{ij} - \bar{Y}_{..} = \left(\bar{Y}_{i.} - \bar{Y}_{..}\right) + \left(Y_{ij} - \bar{Y}_{i.}\right) \tag{2.4}$$

where the first term in the rhs of (2.4) corresponds to the deviation of the estimated $i^{\text{th}}$ sample mean from the overall mean, and the second term is to the deviation of the observation from its estimated sample mean, which corresponds to the residual term $\hat{e}_{ij}$.

If both sides of (2.4) are squared and summed over all observations, we obtain

$$\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_{..}\right)^2=\sum_{i=1}^{a}n_i\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_{i.}\right)^2 \tag{2.5}$$

Remark that the crossproduct in (2.5) equals zero. Indeed

$$\begin{aligned}\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)\left(Y_{ij}-\bar{Y}_{i.}\right) &= \sum_{i=1}^{a}\sum_{j=1}^{n_i}\bar{Y}_{i.}Y_{ij}-\sum_{i=1}^{a}\sum_{j=1}^{n_i}\bar{Y}_{i.}\bar{Y}_{i.}-\sum_{i=1}^{a}\sum_{j=1}^{n_i}Y_{ij}\bar{Y}_{..}+\sum_{i=1}^{a}\sum_{j=1}^{n_i}\bar{Y}_{..}\bar{Y}_{i.}\\ &= \sum_{i=1}^{a}\bar{Y}_{i.}Y_{i.}-\sum_{i=1}^{a}n_i\bar{Y}_{i.}\bar{Y}_{i.}-\sum_{i=1}^{a}Y_{i.}\bar{Y}_{..}+\sum_{i=1}^{a}n_i\bar{Y}_{..}\bar{Y}_{i.}\\ &= \sum_{i=1}^{a}n_i\bar{Y}_{i.}\bar{Y}_{i.}-\sum_{i=1}^{a}n_i\bar{Y}_{i.}\bar{Y}_{i.}-\sum_{i=1}^{a}n_i\bar{Y}_{i.}\bar{Y}_{..}+\sum_{i=1}^{a}n_i\bar{Y}_{..}\bar{Y}_{i.}=0\end{aligned}$$

The lhs in (2.5) measures the total variability of the observations, and is therefore called the total sums of squares, $\text{SS}_{\text{tot}}$,

$$\text{SS}_{\text{tot}}=\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_{..}\right)^2 \tag{2.6}$$

The first term of the rhs in (2.5) measures the separation between the different sample means. If all the sample means are exactly the same, this term equals zero. This term therefore contains information about the observed differences between the different levels of the factor. As the levels of the factor often correspond to different treatments, this sum of squares is called the treatment sum of squares, $\text{SS}_{\text{trt}}$,

$$\text{SS}_{\text{trt}}=\sum_{i=1}^{a}n_i\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)^2 \tag{2.7}$$

The second term of the rhs in (2.5) measures the spread of the observations around their estimated sample mean. If all observations within a factor level are the same, this term is zero. This term contains information about the random variation of the observations around their sample mean. This sum of squares is called the sum of squares of the error, $\text{SS}_{\text{err}}$,

$$\text{SS}_{\text{err}}=\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_{i.}\right)^2 \tag{2.8}$$

In hypothesis testing, we are not using these sums of squares, but rather the mean sums of squares. The mean sum of squares is obtained by dividing the sum of squares by the number of independent terms in the sum of squares, also called the degrees of freedom of

the sum of squares.

$\text{SS}_{\text{tot}}$ consists of $n_.$ terms, but only $n_.-1$ are independent. Indeed, we have that $\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_{..}\right)=0$ and one term can thus be written as a function of the other terms.

The mean total sum of squares is then

$$\text{MS}_{\text{tot}}=\frac{\text{SS}_{\text{tot}}}{n_.-1}$$

$\text{SS}_{\text{trt}}$ consists of $a$ terms, with only $a-1$ independent terms as $\sum_{i=1}^{a} n_i\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)=0$. The mean treatment sum of squares is thus

$$\text{MS}_{\text{trt}}=\frac{\text{SS}_{\text{trt}}}{a-1}$$

$\text{SS}_{\text{err}}$ consist of $n_.$ terms, but within each level $i$ of the factor we have that $\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_{i.}\right)=0$. Therefore, there are $a$ terms that can be written as a function of the other terms, leading to $n_.-a$ degrees of freedom for this sum of squares and the mean error sum of squares is therefore given by

$$\text{MS}_{\text{err}}=\frac{\text{SS}_{\text{err}}}{n_.-a}$$

The reason why mean sums of squares rather than sums of squares are used, can be understood by studying the expected value of the mean sums of squares.

The expected values of $\text{MS}_{\text{err}}$ and $\text{MS}_{\text{trt}}$ are given by (see Note 2.3):

$$\begin{aligned}\text{E}(\text{MS}_{\text{err}}) &= \sigma^2\\ \text{E}(\text{MS}_{\text{trt}}) &= \sigma^2+\frac{\sum_{i=1}^{a} n_i(\mu_i-\mu_.)^2}{a-1}\end{aligned}\tag{2.9}$$

where $\mu_.=\frac{\sum_{i=1}^{a} n_i\mu_i}{n_.}$.

Therefore, $\text{MS}_{\text{err}}$ is an unbiased estimator for $\sigma^2$, regardless whether treatment differences exist.

On the other hand, when all population means are equal, and thus also equal to $\mu_.$, it follows that $\text{E}(\text{MS}_{\text{trt}})=\sigma^2$ as the second term of the rhs of (2.9) equals zero. Whenever the population means differ from each other, the second term of the rhs of (2.9) will make a positive contribution to $\text{MS}_{\text{trt}}$, and thus $\text{E}(\text{MS}_{\text{trt}})>\sigma^2$.

Under the null hypothesis that all population means are equal, i.e.,

$$\mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_a$$

the expectations of $\mathrm{MS}_{\mathrm{trt}}$ and $\mathrm{MS}_{\mathrm{err}}$ are equal to each other and we expect for the ratio

$$F^* = \frac{\mathrm{MS}_{\mathrm{trt}}}{\mathrm{MS}_{\mathrm{err}}}$$

a value equal to 1.

Under the alternative hypothesis

$$\mathrm{H}_\mathrm{a} : \text{ Not all } \mu_i \text{ equal}$$

the expected value of $\mathrm{MS}_{\mathrm{trt}}$ will be larger than $\mathrm{MS}_{\mathrm{err}}$ and we expect a value larger than 1 for the ratio $F^*$.

Under the null hypothesis, the ratio $F^*$ has a F-distribution with $(a-1)$ and $(n_\cdot - a)$ degrees of freedom

$$F^* \sim \mathrm{F}[a-1, n_\cdot - a]$$

We will mainly use the P-value to take the decision to reject the null hypothesis. The P-value corresponds to the probability of observing the same or a more extreme result as in the experiment, given that the null hypothesis is true. Under the null hypothesis, the ratio $F^*$ is F-distributed, and more extreme results contradicting the null hypothesis correspond to higher values for the $F^*$ statistic than the one observed, $f^*$. The P-value is therefore given by

$$\mathrm{P}(\mathrm{F}[a-1, n_\cdot - a] \geq f^*)$$

with $f^*$ the actual value for the $F^*$ statistic. The null hypothesis is rejected when the P-value is smaller than the significance level $\alpha$.

The cumulatieve F-distribution, from which the P-value can be read, is given in Table 9.3.

We often present the SS, degrees of freedom, MS, F-statistics and P-values in an analysis of variance or ANOVA ('ANalysis Of VAriance') table, as demonstrated in Example 2.2

**Example 2.2 Analysis of variance for weight gain in chickens of 4 different breeds**

The sample means of the 4 different breeds equal $\bar{y}_{1\cdot} = 1.52$, $\bar{y}_{2\cdot} = 1.40$, $\bar{y}_{3\cdot} = 1.50$ and $\bar{y}_{4\cdot} = 1.48$. The ANOVA table is presented in Table 2.2.
The P-value equals P(F[3,16]$\geq$ 17.29)= 0.000028, and is much smaller than the default significance level of 5%. We can reject the null hypothesis that the 4 breeds are equal. A graphical representation of the P-value is given in Figure 2.2 as the area under the density function F[3,16] to the right of the value 17.29.
■

Table 2.2: ANOVA table for weight gain of 5 chickens of 4 different breeds after 8 weeks.

| Term | SS | df | MS | $f^*$ | P(F$\geq f^*$) |
|---|---|---|---|---|---|
| Breeds | 0.0415 | 3 | 0.0138 | 17.29 | 0.000028 |
| Error | 0.0128 | 16 | 0.0008 | | |
| Total | 0.0543 | 19 | 0.0029 | | |

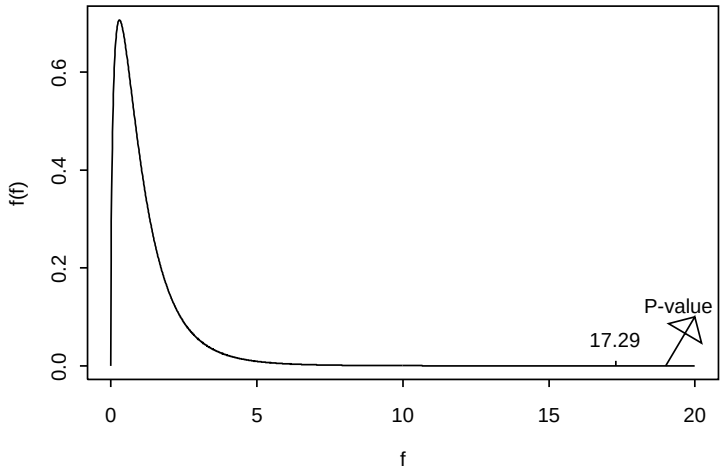



Figure 2.2: Representation of the P-value as the area under the density function F[3,16] to the right of the value 17.29.

## 2.5 Specific comparisons

If the general null hypothesis described in the previous section is rejected, we would like to know which treatments differ from one another. Three different types of comparisons are investigated in this section, the pairwise comparison, the contrast and the linear combination. The lineair combination is the most general comparison; the two other comparisons can be described in terms of a linear combination.

### 2.5.1 Pairwise comparison

Two levels of a factor are often compared by defining hypotheses based on the difference between the population means of the two levels

$$\Delta = \mu_i - \mu_j$$

Such a comparison is called a pairwise comparison. An unbiased estimator is given by

$$\hat{\Delta} = \bar{Y}_{i.} - \bar{Y}_{j.}$$

This estimator is unbiased as $\mathrm{E}(\hat{\Delta}) = \mu_i - \mu_j$.

Due to the independence between $\bar{Y}_{i.}$ and $\bar{Y}_{j.}$ we have

$$\begin{aligned} \mathrm{Var}\left(\hat{\Delta}\right) &= \mathrm{Var}\left(\bar{Y}_{i.}\right) + \mathrm{Var}\left(\bar{Y}_{j.}\right) \\ &= \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right) \end{aligned}$$

The estimated variance of $\hat{\Delta}$ is obtained by replacing $\sigma^2$ with its unbiased esimator

$$S^2\left(\hat{\Delta}\right) = \mathrm{MS}_{\mathrm{err}}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)$$

Because $\hat{\Delta}$ is a linear combination of independent normally distributed random variables, it follows that

$$\frac{\hat{\Delta} - \Delta}{\sqrt{\mathrm{Var}\left(\hat{\Delta}\right)}} \sim \mathrm{N}(0,1)$$

If $\mathrm{Var}\left(\hat{\Delta}\right)$ in this expression is replaced by its estimator $S^2\left(\hat{\Delta}\right)$, we also have to replace the standard normal distribution by the $T$-distribution with $n_{\cdot} - a$ degrees of freedom

$$\frac{\hat{\Delta} - \Delta}{S\left(\hat{\Delta}\right)} \sim T[n_{\cdot} - a]$$

It follows that the (1-$\alpha$)100% confidence interval is given by

$$\hat{\Delta} \pm T[1 - \alpha/2, n_{\cdot} - a]S\left\{\hat{\Delta}\right\}$$

Testing whether the population means of the levels differ is based on the following set of hypotheses

$$\mathrm{H}_0 : \mu_i - \mu_j = 0$$

and

$$\mathrm{H}_{\mathrm{a}} : \mu_i - \mu_j \neq 0$$

To test this hypothesis we use the following test statistic

$$T^* = \frac{\hat{\Delta} - 0}{S\left(\hat{\Delta}\right)}$$

which is distributed as $T[n_\cdot - a]$ under the null hypothesis. The P-value for this two-sided hypothesis is thus given by

$$2 \times \mathrm{P}(T[n_\cdot - a] \geq \mid t^* \mid)$$

If the P-value is smaller than the significance level $\alpha$ the null hypothesis is rejected.

**Example 2.3 Pairwise comparison between two chicken breeds**

We compare the first with the second breed. The first breed has a sample mean equal to 1.52, the second sample mean equals 1.40. From Example 2.2 we find that $\mathrm{MS}_{\mathrm{err}}$=0.0008. It follows that $s^2\left(\hat{\Delta}\right) = 0.00032$. The 95% confidence interval for the difference between breeds 1 and 2 is therefore given by

$$(1.52 - 1.40) \pm 2.120\sqrt{0.00032} \text{ and thus } [0.082; 0.158].$$

The P-value equals

$$2 \times \mathrm{P}\left(T[16] \geq \frac{1.52 - 1.40}{\sqrt{0.00032}}\right) = 2 \times \mathrm{P}\left(T[16] \geq 6.708\right)$$

and this P-value is smaller than 0.002 (see Table 9.2). The null hypothesis is rejected.
■

### 2.5.2 Contrasts

A contrast is a comparison related to two or more levels of the factor. We write a contrast as $L$; a contrast is defined as a linear combination of population means of the different levels of the factor, $\mu_i$, with the coefficients $c_i$ summing to zero:

$$L = \sum_{i=1}^{a} c_i \mu_i \text{ with } \sum_{i=1}^{a} c_i = 0 \tag{2.10}$$

The pairwise comparison is thus a special case of a contrast with one population mean having the coefficient 1 and the other the coefficient -1, all other coefficients being equal to 0.

An unbiased estimator of $L$ is given by

$$\hat{L} = \sum_{i=1}^{a} c_i \bar{Y}_{i\cdot}$$

Due to the independence of the $\bar{Y}_{i\cdot}$'s it follows that

$$\begin{aligned}\mathrm{Var}\left(\hat{L}\right) &= \sum_{i=1}^{a} c_i^2 \mathrm{Var}\left(\bar{Y}_{i.}\right) \\ &= \sum_{i=1}^{a} c_i^2 \left(\frac{\sigma^2}{n_i}\right) \\ &= \sigma^2 \sum_{i=1}^{a} \frac{c_i^2}{n_i}\end{aligned}$$

The estimated variance of $\hat{L}$ is obtained by replacing $\sigma^2$ with its unbiased estimator

$$S^2\left(\hat{L}\right) = \mathrm{MS}_{\mathrm{err}} \sum_{i=1}^{a} \frac{c_i^2}{n_i}$$

As $\hat{L}$ is a linear combination of independent normally distributed variables it follows that

$$\frac{\hat{L} - L}{\mathrm{Var}\left(\hat{L}\right)} \sim \mathrm{N}(0,1)$$

If $\mathrm{Var}\left(\hat{L}\right)$ in this expression is replaced by its estimator $S^2\left(\hat{L}\right)$, we also have to replace the standard normal distribution by the $T$-distribution with $n_{\cdot} - a$ degrees of freedom

$$\frac{\hat{L} - L}{S\left(\hat{L}\right)} \sim T[n_{\cdot} - a]$$

It follows that the (1-$\alpha$)100% confidence interval is given by

$$\hat{L} \pm T[1 - \alpha/2, n_{\cdot} - a] S\left(\hat{L}\right)$$

Testing whether the population means of the levels differ is based on the following set of hypotheses

$$\mathrm{H}_0 : L = 0$$

and

$$\mathrm{H}_{\mathrm{a}} : L \neq 0$$

To test this hypothesis we use the following test statistic

$$T^* = \frac{\hat{L}}{S\left(\hat{L}\right)}$$

which is distributed as $T[n_{\cdot} - a]$ under the null hypothesis. The P-value for this two-sided hypothesis is thus given by

$$2 \times \mathrm{P}(T[n_\cdot - a] \geq \mid t^* \mid)$$

If the P-value is smaller than the significance level $\alpha$ the null hypothesis is rejected.

**Example 2.4 Contrast between four chicken breeds**

We compare the mean of the first and the second breed with the mean of the third and the fourth breed. We choose as coefficients $c_1 = 0.5$, $c_2 = 0.5$, $c_3 = -0.5$ and $c_4 = -0.5$. This is indeed a contrast as the sum of the coefficients $c_i$ equals zero.
The hypotheses for this contrast are written as

$$\mathrm{H}_0 : \frac{\mu_1 + \mu_2}{2} = \frac{\mu_3 + \mu_4}{2}$$

and

$$\mathrm{H}_\mathrm{a} : \frac{\mu_1 + \mu_2}{2} \neq \frac{\mu_3 + \mu_4}{2}$$

The estimate of this contrast equals -0.03; the estimated standard deviation of the contrast equals $s = 0.01265$.

The P-value is thus

$$2 \times \mathrm{P}\left(T[16] > \mid \frac{-0.03}{0.01265} \mid\right) = 2 \times \mathrm{P}\left(T[16] > 2.37\right)$$

and this P-value is between 0.02 and 0.05 (see Table 9.2). The null hypothesis is rejected at a significance level of 5%.
■

### 2.5.3 Linear combinations

In some cases, interest is in a linear combination of two or more levels of a factor that is not a contrast. We write a linear combination also as $L$, but without the restriction $\sum_{i=1}^{a} c_i = 0$. Exactly the same techniques as in the previous section on contrasts can be used to find confidence intervals or P-values.

### 2.5.4 Multiple comparisons

In the previous sections, we described how different hypotheses could be tested using linear combinations, contrasts and pairwise differences. If only 1 hypothesis is tested, the significance level $\alpha$ is respected, i.e., on average only 100 $\alpha$% of the experiments will lead to a significant result, although the null hypothesis is true (type I error). However, if we test different hypotheses, all at a significance level $\alpha$, the probability of at least one significant result will increase with increasing number of hypotheses tested, and the type I error will be (much) larger than $\alpha$. This problem occurs when testing multiple comparisons. It is therefore required to test each hypothesis at a significance level $\alpha_g$, in order to ensure that

the type I error for the joint set of hypotheses does not exceed the global significance level $\alpha$: the comparisonwise significance level $\alpha_g$ needs to be chosen so that the probability that one or more comparisons are significant at the $\alpha_g$ significance level does not exceed $\alpha$ under the null hypothesis.

Different techniques have been developed depending on the comparisons of interest. The Tukey method is typically used when all pairwise comparisons are of interest. The Dunnett method is most efficient if one wants to compare all treatments with a control group. If Scheffé's method is used, one is allowed to construct any set of comparisons.

We only describe one technique in this section, namely the Bonferroni technique, because it is the simplest technique and most generally applicable. If the global significance level should not exceed $\alpha$, and if one is interested in $g$ hypotheses, then each single hypothesis must be tested at a significance level equal to $\alpha_g = \alpha/g$. We can calculate the P-value of a specific comparison as before, but the P-value needs to be smaller than $\alpha_g$ before a significant difference can be claimed.

**Example 2.5 Three pairwise comparisons between vier chicken breeds**

We compare the first breed with the three other breeds; we therefore have three comparisons, that need to be tested at a significance level of 0.05/3=0.0167 according to the Bonferroni technique. As an example we demonstrate how we can compare the first and second breed. We can still use the P-value from Example 2.3; the P-value was smaller than 0.002. The P-value is therefore also smaller than 0.0167 and we can claim a significant difference between breed 1 and breed 2.

■

## 2.6 Diagnostic tests for the fixed effects model with 1 factor

Analysis of variance is based on a number of assumptions regarding the data. It is good statistical practice to evaluate to what degree the model assumptions are supported by the data. We discern graphical techniques, based to a large extent on residuals, on the one hand and more formal tests on the other hand. We will first define below different types of residual values, that will be useful to evaluate different model assumptions. In the next sections, we investigate how the homogeneity of the variance, the normality and the independence of the observations can be evaluated. Finally, we describe how outliers can be found in a dataset.

### 2.6.1 Residual values

De residual values for the ANOVA model

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

are obtained by replacing $\mu_i$ with its estimator $\bar{Y}_{i.}$ from which follows

$$e_{ij} = Y_{ij} - \bar{Y}_{i.}$$

Such residual values are often standardised. The semistudentized residual $e_{ij}^*$ is obtained as

$$e_{ij}^* = \frac{e_{ij}}{\sqrt{\text{MS}_{\text{err}}}}$$

We call $e_{ij}^*$ the semistudentized residual because it is not based on the correct estimation of the variance of $e_{ij}$.

The studentized residual is given by

$$r_{ij} = \frac{e_{ij}}{s\left(e_{ij}\right)}$$

where

$$s\left(e_{ij}\right) = \sqrt{\text{MS}_{\text{err}} \frac{n_i - 1}{n_i}}$$

In calculating the previous types of residual values, the observation itself is also used to estimate the parameters. If the effect of a specific observation needs to be known, it is more obvious to remove the observation when estimating the required parameters. The deleted residual for the $j^{\text{th}}$ observation is obtained as

$$d_{ij} = Y_{ij} - \bar{Y}_{i(-j)} \tag{2.11}$$

where $\bar{Y}_{i(-j)}$ represents the $i^{\text{th}}$ sample mean without making use of the $j^{\text{th}}$ observation.

The studentized deleted residual is then given by

$$t_{ij} = e_{ij} \sqrt{\frac{n_{\cdot} - a - 1}{\text{SS}_{\text{err}}\left(1 - \frac{1}{n_i}\right) - e_{ij}^2}}$$

**Example 2.6 Different types of residual values for the weight gain of 4 chicken breeds**

Figure 2.3 shows 4 different types of residual values as a functon of breed. We will use this figure in the remainder to test the different model assumptions.

■

### 2.6.2 Homogeneity of the variance

In the ANOVA model we assume that the observations for each level of the factor are characterized by the same variance $\sigma^2$. A visual control of this assumption is based on a plot of the residual values at each level of the factor. This way, it can be assessed whether the spread of the observations for each level of the factor is similar (see Figuur 2.3).

There exist also formal tests, such as the Hartley test, which can be applied as follows.

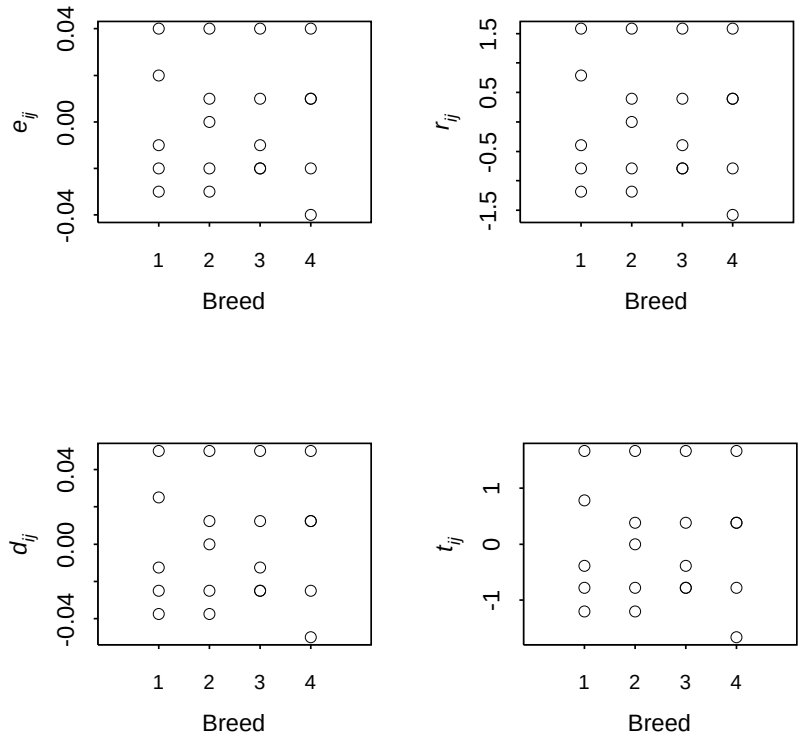



Figure 2.3: Four types of residual values as a function of the level of the factor breed, where $e_{ij}$ is the residual value, $r_{ij}$ the studentized residual, $d_{ij}$ the deleted residual and $t_{ij}$ the deleted studentized residual.

The null hypothesis

$$\mathrm{H}_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_a^2$$

is tested relative to the alternative hypothesis

$$\mathrm{H}_\mathrm{a} : \text{ not all } \sigma_i^2 \text{ equal}$$

The test statistic is based on the largest and smallest sample variance

$$H^* = \frac{\max(s_i^2)}{\min(s_i^2)}$$

with values close to 1 supporting the null hypothesis. The cumulatieve distribution of $H^*$ under de null hypothesis is presented in Table 9.4. The distribution of $H^*$ depends on

the number of levels of the factor $a$ and the degrees of freedom available to estimate the sample variance of one population, $df = n_i - 1$, where we assume that alle samples have the same size.
The decision rule for a significance level $\alpha$ is given by

$$\begin{aligned} &\text{If } H^* \leq H[1-\alpha; a; df], \quad \text{decide } \mathrm{H_0} \\ &\text{If } H^* > H[1-\alpha; a; df], \quad \text{decide } \mathrm{H_a} \end{aligned}$$

where $H[1-\alpha; a; df]$ corresponds to the $(1-\alpha)100$ percentile of the distribution of $H^*$ under the null hypothesis.

**Example 2.7 Homogeneity of the variance for the weight gain of 4 chicken breeds**

We first use the first plot of residual values in Figure 2.3. It seems that the spread of the residual values is similar for each breed. To execute the Hartley test, we first need to derive the sample variances of each of the breeds, with results $s_1^2 = 0.00085$, $s_2^2 = 0.00075$, $s_3^2 = 0.00065$ and $s_4^2 = 0.00095$. The test statistic is thus

$$H^* = \frac{\max(s_i^2)}{\min(s_i^2)} = \frac{0.00095}{0.00065} = 1.461$$

This test statistic needs to be evaluated in Table 9.4 in column $a = 4$ and row $df = 4$. The critical value corresponds to 20.6. The test statistic is much smaller than the critical value; we can not reject the null hypothesis of homogeneity of variances.
■

### 2.6.3 Normality of the observations

Another model assumption is that the residual terms $e_{ij}$ are distributed normally, and therefore also the observations. The normal probability plot is often used as a diagnostic plot for this model assumption. If data are normally distributed and the model is correct, then the residual values constitute a sample from a normal distribution. The residual values are then ordered, and for each residual value, the expected value is then derived according to its order and the normal distribution assumption. As the mean of the residual values is zero, and the variance can be estimated by $\mathrm{MS_{err}}$, a good approximation of the expected value of the $i^{\text{th}}$ smallest observation is

$$\sqrt{\mathrm{MS_{err}}}\left[z\left(\frac{i-0.375}{n+0.25}\right)\right] \tag{2.12}$$

where $z(A)$ is the $(A)100$ percentile of the standard normal distribution.

The residual values are then depicted as a function of their expected values, what should lead to a set of points close to a straight line if the normality assumption holds.

### 2.6.4 Independence of the observations

We further assume in an ANOVA model that the observations, and therefore also the residual values, are independent from each other. Using a graphical presentation of the residual values, we can evaluate whether this assumption is reasonable if the residual values can be plotted sequentially in a meaningful way. For instance, in the case of a diagnostic test, we can plot the residual values as a function of the time when the diagnostic test was executed.

**Example 2.8 Independence of observations of weight gain of chickens**

To evaluate whether the observed weights change as a functon of the time of measurement, we plot the residual values in Figure 2.4 according to the time of measurement. There does not seem to be a trend in the residual values, which shows that the measurement sequence has little to no effect on the weight.

■

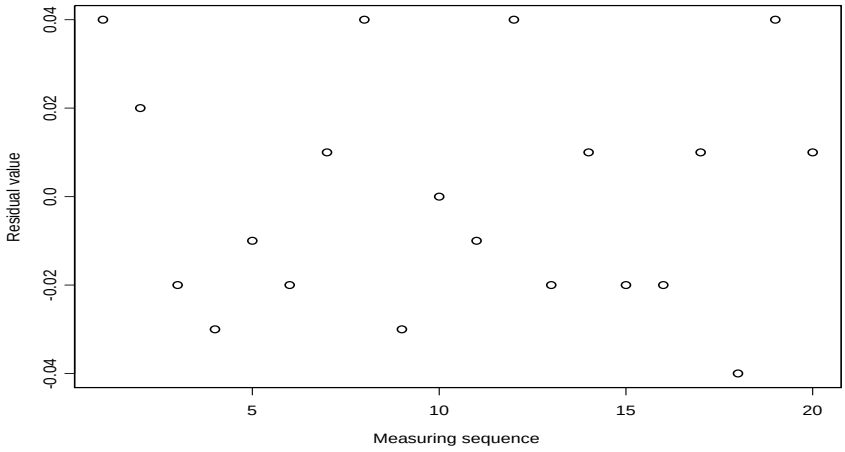



Figure 2.4: The residual values of the weights of four breeds of chicken as a functon of the measurement sequence.

### 2.6.5 Outliers

An outlier is an observation with a large absolute value for the studentized residual; this means that the model does not predict this observation well. We use the deleted studentized residuals to detect outliers. To propose limits of acceptable values, the Bonferroni principle can be used. The deleted studentized residuals are $T$ distributed with $n_{\cdot} - a - 1$ degrees of freedom if the model is correct. As the total number of observations correspond to $n_{\cdot}$ and we want to use a two-sided interval (both very small negative and very large positive deviations can be outliers), the limit values are $T[1 - \alpha/2n_{\cdot}; n_{\cdot} - a - 1]$.

**Example 2.9 Outliers in observations of weight gain of chickens**

To evaluate whether there are outliers in the observed weights of the 4 breeds, we plot the deleted studentized residuals according to breed in Figure 2.5, together with the Bonferroni corrected limit values. The limit values are given by $T(0.99875; 15)$=3.62. All observations are located within the limits; no outliers are present in the data.

■

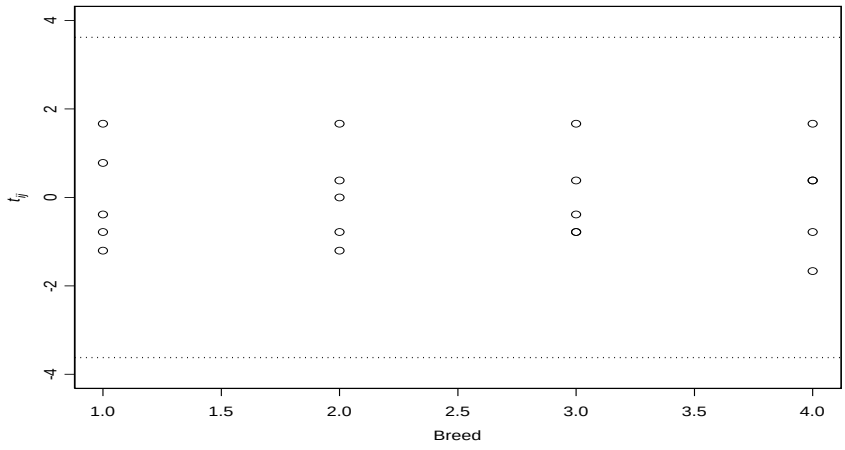



Figure 2.5: The deleted studentized residuals of the weight gains of four chicken breeds according to breed.

## 2.7 Notes

**Note 2.1 Expected value operator E(.)**

The operator E(.) is an important tool in statistical reasoning. We first give a formal definition, thereafter a more intuitive explanation follows.

We can define the expected value for any random variable or function of random variables. We first consider the random variable itself.

For a discrete random variable $Y$, taking values $y_1, \ldots, y_a$ with probabilities $p_1, \ldots, p_a$, the expected value is given by

$$\mathrm{E}(Y) = \sum_{i=1}^{a} y_i p_i \tag{2.13}$$

which corresponds to a weighted mean of all possible values, with weights given to values according to their probability of occurence.

For instance, for $Y \sim \mathrm{B}(n;\pi)$, we have

$$\mathrm{E}(Y) = \sum_{i=1}^{n} y_i \mathrm{P}(Y = y_i) = \sum_{i=1}^{n} y_i \binom{n}{y_i} \pi^{y_i}(1-\pi)^{n-y_i} = n\pi$$

For a continuous random variable $Y$ with density function $f(y)$, we have that

$$\mathrm{E}(Y) = \int_{-\infty}^{+\infty} y f(y) dy \tag{2.14}$$

For instance, with $Y \sim \mathrm{N}(\mu, \sigma^2)$, we have (without proof)

$$\begin{aligned} \mathrm{E}(Y) &= \int_{-\infty}^{+\infty} y f_Y(y) dy \\ &= \int_{-\infty}^{+\infty} y \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right) dy = \mu \end{aligned}$$

An alternative and more intuitive interpretation is based on sampling. Assume that a sample of size $n$ is taken from a normally distributed random variable $Y \sim \mathrm{N}(\mu, \sigma^2)$. Then the sample mean will tend to the expected value if the sample size $n$ tends to infinity. For small sample sizes, the sample means will be scattered around the expected value. The expected value therefore corresponds to the sample mean: if the sample size goes to $N$, the population size for a finite popultion, or to $\infty$ for an infinite population, then the sample mean goes to the population mean.
The population variance is also defined in terms of the expected value operator, i.e.,

$$\mathrm{Var}(Y) = \mathrm{E}(Y - E(Y))^2$$

For instance, for $Y \sim \mathrm{N}(\mu, \sigma^2)$, we have (without proof)

$$\mathrm{Var}(Y) = \int_{-\infty}^{+\infty} (Y - E(Y))^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right) dy = \sigma^2$$

Before continuing with the expected value of a function of random variables, we consider the following important properties of the E(.) operator. With $X$ and $Y$ two random variables and $a$ and $b$ two constants, we have

$$\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$$

This follows immediately from definitions 2.13 and 2.14.
Furthermore, as for $Y \sim \mathrm{N}(\mu, \sigma^2)$, $\mathrm{E}(Y) = \mu$, it follows that for $e_{ij} \sim \mathrm{N}(0, \sigma^2)$, $\mathrm{E}(e_{ij}) = 0$.
Finally, for a constant, such as the global mean $\mu$, we have that $\mathrm{E}(\mu) = \mu$.

We now apply the E operator to the sample variance, $S^2 = \frac{\sum(Y_i-\bar{Y}_\cdot)^2}{n-1}$. We first rewrite the numerator.
As $\sum_{i=1}^{n}(Y_i - \bar{Y}_\cdot) = \sum_{i=1}^{n}(Y_i - \mu) - (\bar{Y} - \mu)$ we have

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \sum_{i=1}^{n}(Y_i - \mu)^2 + \sum_{i=1}^{n}(\bar{Y} - \mu)^2 - 2\sum_{i=1}^{n}(Y_i - \mu)(\bar{Y} - \mu) \\ &= \sum_{i=1}^{n}(Y_i - \mu)^2 + n(\bar{Y} - \mu)^2 - 2n(\bar{Y} - \mu)^2 \\ &= \sum_{i=1}^{n}(Y_i - \mu)^2 - n(\bar{Y} - \mu)^2\end{aligned}$$

Taking the expected value, and using the fact that the variance of $\bar{Y}_\cdot$ equals $\sigma^2/n$ (see Note 2.2), we have

$$\begin{aligned}\mathrm{E}\left(\sum_{i=1}^{n}(Y_i - \bar{Y}_\cdot)^2\right) &= \mathrm{E}\left(\sum_{i=1}^{n}(Y_i - \mu)^2\right) - \mathrm{E}\left(n(\bar{Y}_\cdot - \mu)^2\right) \\ &= n\sigma^2 - \sigma^2 = (n-1)\sigma^2\end{aligned}$$

Therefore, $S^2$ is an unbiased estimator of $\sigma^2$ as

$$\mathrm{E}\left(S^2\right) = \frac{\mathrm{E}\left(\sum_{i=1}^{n}(Y_i - \bar{Y}_\cdot)^2\right)}{n-1} = \frac{(n-1)\sigma^2}{n-1} = \sigma^2$$

■

**Note 2.2 The variance of a linear combination of independently distributed random variables**

Assume that we have a series of mutually independent random variables $X_1, \ldots, X_a$. The variance of a linear combination of these random variables, $\sum_{i=1}^{a} a_i X_i$, is given by $\sum_{i=1}^{a} a_i^2 \mathrm{Var}(X_i)$.
A useful example is the sample mean. Assume that the sample mean is based on $n$ independent observations from the same probability or density function with $\mathrm{Var}(X)$, then the variance of the sample mean is given by

$$\mathrm{Var}(\bar{X}_\cdot) = \mathrm{Var}(\sum_{i=1}^{n}\frac{X_i}{n}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X) = \frac{\mathrm{Var}(X)}{n}$$

In the case of mutually independent normally distributed random variables $X_1, \ldots, X_a$, with $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $i = 1, \ldots, a$, we have that a linear combination of these random variables, $\sum_{i=1}^{a} a_i X_i$, is also normally distributed, $\sum_{i=1}^{a} a_i X_i \sim \mathrm{N}(a_i\mu_i, a_i^2\sigma_i^2)$.

In the case of a sample mean of $n$ independent observations from the normal distribution $\mathrm{N}(\mu, \sigma^2)$, it follows that

$$\bar{X}_{\cdot} \sim \mathrm{N}(\mu, \sigma^2/n) \tag{2.15}$$

■

**Note 2.3 Expected values of mean sum of squares**

We derive the expected values of $\mathrm{MS}_{\mathrm{err}}$ and $\mathrm{MS}_{\mathrm{trt}}$.

We first rewrite the $\mathrm{MS}_{\mathrm{err}}$

$$\begin{aligned}
\mathrm{MS}_{\mathrm{err}} &= \frac{1}{n_{\cdot} - a} \sum_{i=1}^{a} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2 \\
&= \frac{1}{n_{\cdot} - a} \sum_{i=1}^{a} (n_i - 1) \sum_{j=1}^{n_i} \frac{\left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2}{n_i - 1} \\
&= \frac{1}{n_{\cdot} - a} \sum_{i=1}^{a} (n_i - 1) s_i^2
\end{aligned}$$

Taking the expected value, we find

$$\mathrm{E}\left(\mathrm{MS}_{\mathrm{err}}\right) = \frac{1}{n_{\cdot} - a} \sum_{i=1}^{a} (n_i - 1) \mathrm{E}\left(s_i^2\right) = \sigma^2 \tag{2.16}$$

The derivation of the expected value of $\mathrm{MS}_{\mathrm{trt}}$ is somewhat more involved. We consider the more simple case of balanced data, i.e., $n_i \equiv n$, the same number of observations for each treatment

$$\mathrm{MS}_{\mathrm{trt}} = \frac{n \sum_{i=1}^{a} \left(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}\right)^2}{a - 1}$$

Since $Y_{ij} = \mu_i + e_{ij}$ it follows that

$$\begin{aligned}
\bar{Y}_{i\cdot} &= \mu_i + \bar{e}_{i\cdot} \text{ with } \bar{e}_{i\cdot} = (\sum_{j=1}^{n_i} e_{ij})/n \\
\bar{Y}_{\cdot\cdot} &= \mu + \bar{e}_{\cdot\cdot} \text{ where } \bar{e}_{\cdot\cdot} = (\sum_{i=1}^{a} \sum_{j=1}^{n_i} e_{ij})/an, \mu = (\sum_{i=1}^{a} \mu_i)/a
\end{aligned}$$

We next rewrite part of the numerator

$$\begin{aligned}\sum_{i=1}^{a}\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)^2 &= \sum_{i=1}^{a}\left((\mu_i+\bar{e}_{i.})-(\mu_.+\bar{e}_{..})\right)^2 \\ &= \sum_{i=1}^{a}\left((\mu_i-\mu)+(\bar{e}_{i.}-\bar{e}_{..})\right)^2 \\ &= \sum_{i=1}^{a}(\mu_i-\mu)^2+\sum_{i=1}^{a}(\bar{e}_{i.}-\bar{e}_{..})^2+2\sum_{i=1}^{a}(\mu_i-\mu)(\bar{e}_{i.}-\bar{e}_{..}) \quad (2.17)\end{aligned}$$

We now find the expected values of each term in the rhs of (2.17).
As the $\mu_i$'s and $\mu$ are constants we have

$$\mathrm{E}(\sum_{i=1}^{a}(\mu_i-\mu)^2)=\sum_{i=1}^{a}(\mu_i-\mu)^2 \quad (2.18)$$

Consider the random variables $\bar{e}_{i.}$. As $e_{ij}\sim \mathrm{N}(0,\sigma^2)$, it follows that $\bar{e}_{i.}\sim \mathrm{N}(0,\sigma^2/n)$, and its sample variance is given by

$$\frac{\sum_{i=1}^{a}(\bar{e}_{i.}-\bar{e}_{..})^2}{a-1}$$

It follows that

$$\mathrm{E}\left(\frac{\sum_{i=1}^{a}(\bar{e}_{i.}-\bar{e}_{..})^2}{a-1}\right)=\frac{\sigma^2}{n}$$

and thus

$$\mathrm{E}\left(\sum_{i=1}^{a}(\bar{e}_{i.}-\bar{e}_{..})^2\right)=\frac{(a-1)\sigma^2}{n} \quad (2.19)$$

Finally, for the last term in the rhs of (2.17) we have

$$\mathrm{E}\left(\sum_{i=1}^{a}(\mu_i-\mu)(\bar{e}_{i.}-\bar{e}_{..})\right)=\sum_{i=1}^{a}(\mu_i-\mu)\left(\mathrm{E}(\bar{e}_{i.})-\mathrm{E}(\bar{e}_{..})\right)=0 \quad (2.20)$$

since both $\mathrm{E}(\bar{e}_{i.})$ and $\mathrm{E}(\bar{e}_{i.})$ are equal to zero.

Now we have

$$
\begin{aligned}
\mathrm{E}\left(\mathrm{MS}_{\mathrm{trt}}\right) &= \frac{n}{a-1}\mathrm{E}\left(\sum_{i=1}^{a}\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)^2\right) \\
&= \frac{n}{a-1}\mathrm{E}\left(\sum_{i=1}^{a}(\mu_i-\mu)^2\right)+\mathrm{E}\left(\sum_{i=1}^{a}(\bar{e}_{i.}-\bar{e}_{..})^2\right)+\mathrm{E}\left(2\sum_{i=1}^{a}(\mu_i-\mu)(\bar{e}_{i.}-\bar{e}_{..})\right) \\
&= \frac{n}{a-1}\left(\sum_{i=1}^{a}(\mu_i-\mu)^2+\frac{(a-1)\sigma^2}{n}\right) \\
&= \frac{n}{a-1}\sum_{i=1}^{a}(\mu_i-\mu)^2+\sigma^2
\end{aligned}
$$

The second step is based on (2.17), the third step on plugging in (2.18-2.20)
■

# Chapter 3

# Fixed effects models with two factors

## 3.1 Introduction

The one way ANOVA model is extended to the ANOVA model with two 2 factors, both of interest to the investigator. In Section 3.2 we discuss the interpretation of the different terms required in the model to describe the observations and we also explain why it is more efficient to investigate the effect of different factors simultaneously. In Section 3.3 the model specification is given. In this chapter, we only discuss the analysis of variance for balanced data, i.e., the same number of observations for each treatment combination. In Section 3.4 we investigate the situation where more than 1 observation per treatment combination is available, followed by Section 3.5 with only 1 and exactly 1 observation per treatment combination.

## 3.2 Interpreting the terms in the fixed effects model with 2 factors

### 3.2.1 Introduction

We first investigate in this section which terms are required in the model to describe all the observations when two factors are included simultaneously in a study. We often study more than one factor in an experiment; we will demonstrate why it is a better strategy to investigate factors simultaneously in an experiment. We first give two data examples that will be used in the remainder of this section.

**Example 3.1 Treatment effect on PCV for Boran and Holstein cows with trypanosomosis**

Trypanosomosis or sleeping sickness is still a common disease in certain regions in Africa. Cows having trypanosomosis are often anemic, which is translated into a low packed cell volume (PCV). We wish to investigate the effect of two different drugs, Berenil and Samorin, on the evolution of the disease in two different cow breeds, Boran and Holstein. We choose 6 Boran and 6 Holstein cows, and randomly allocate them to the two drugs

in such a way that 3 Boran (Holstein) cows receive Berenil and 3 Samorin. The data are presented in Table 3.1. We use as response variabele the difference in PCV before and after the treatment. ■

Table 3.1: Packed cell volume (PCV) before and after treatment with Berenil or Samorin in Holstein and Boran cows

| Cowid | Breed | Drug | PCV-before | PCV-after | PCV-difference |
|---|---|---|---|---|---|
| 1 | Boran | Berenil | 18.4 | 26.3 | 7.9 |
| 2 | Boran | Berenil | 20.3 | 28.1 | 7.8 |
| 3 | Boran | Berenil | 22.2 | 27.8 | 5.6 |
| 4 | Boran | Samorin | 16.3 | 30.1 | 13.8 |
| 5 | Boran | Samorin | 15.4 | 27.3 | 11.9 |
| 6 | Boran | Samorin | 19.2 | 32.7 | 13.5 |
| 7 | Holstein | Berenil | 21.3 | 28.3 | 7.0 |
| 8 | Holstein | Berenil | 17.4 | 26.8 | 9.4 |
| 9 | Holstein | Berenil | 18.2 | 25.8 | 7.6 |
| 10 | Holstein | Samorin | 22.2 | 38.1 | 15.9 |
| 11 | Holstein | Samorin | 19.8 | 32.3 | 12.5 |
| 12 | Holstein | Samorin | 20.4 | 30.8 | 10.4 |

**Example 3.2 Mastitis in cows as a function of parity and inoculation dose**

Udder infection or mastitis is globally a common disease in cows. An efficient experimental model is required to evaluate the effect of a vaccine in controled conditions. The animals are experimentally infected with *Escherichia Coli* in such an experimental model by infusing the bacteria in an udder quarter.

To develop such an experimental model, we need to assess the effect of two different factors: the inoculation dose and the parity. We make use of three different inoculation doses in the experiment: $10^2$ (low), $10^4$ (medium) and $10^6$ (high) colony forming units (CFU). We also want to assess whether differences exist between heifers (cows which experienced only 1 calving) and multiparous cows (cows which experienced more than 1 calving). We wish to determine the most efficient setup with regard to inoculation dose and parity for future vaccination experiments.

Six treatment combinations occur. For each treatment combination, we have two cows. Thus, 6 heifers are randomly allocated to the 3 inoculation doses in such a way that each dose appears twice ('restricted randomisation'). The 6 multiparous cows are assigned to the 3 inoculation doses in a similar way.

The proportional reduction in milk production in the non-inoculated udder quarters 48 hours after infection is the response variable. The milk reductions for the different cows

are given in Table 3.2. ■

Table 3.2: Milk production in non-infected udder quarters just before (Milk0) and 48 hours after (Milk48) infection as a function of the two factors parity and inoculation dose.

| Cowid | Parity | Inoculation dose | Milk0 | Milk48 | Reduction |
|---|---|---|---|---|---|
| 1 | heifer | high | 32.4 | 30.2 | 6.79 |
| 2 | heifer | high | 33.6 | 32.3 | 3.87 |
| 3 | heifer | medium | 29.3 | 20.5 | 30.03 |
| 4 | heifer | medium | 34.4 | 21.3 | 38.08 |
| 5 | heifer | low | 31.3 | 14.5 | 53.67 |
| 6 | heifer | low | 35.3 | 13.4 | 62.04 |
| 7 | multiparous | high | 42.4 | 39.5 | 6.84 |
| 8 | multiparous | high | 43.3 | 39.7 | 8.31 |
| 9 | multiparous | medium | 45.2 | 23.9 | 47.12 |
| 10 | multiparous | medium | 44.4 | 24.8 | 44.14 |
| 11 | multiparous | low | 41.5 | 6.7 | 83.86 |
| 12 | multiparous | low | 45.2 | 4.1 | 90.93 |

An investigator sometimes decides to study two factors separately, i.e., vary one factor at a time, keep the others constant. In the previous example, one could compare heifers and multiparous cows first at the highest inoculation dose. If multiparous cows experience a higher milk reduction (which makes them more appropriate for the experimental infection model), then the investigator could compare the 3 inoculation doses only for multiparous cows.

This 'one factor at a time' approach is, however, inefficient, due to different reasons.

- In this approach, not all treatment combinations are evaluated. We might therefore not choose the optimal treatment combination.
- We can not assess whether the two factors interact, i.e., whether the differences between levels in 1 factor change according to the level of another factor. For instance, large differences between heifers and multiparous cows could exist at the low inoculation dose, but not so at the high inoculation dose.
- The treatment combinations are not completely randomly assigned because they occur in two different experiments. For instance, in the experimental situation described above, the heifers with the high inoculation dose can not be compared with the multiparous cows with the low inoculation dose because the data come from two different experiments.
- The 'one factor at a time' approach is also logistically more demanding, because the study consists of two parts, and the results from the first part must be available

to proceed to the second part, as the choice of the treatment combinations in the second part depend on the results observed in the first part.

Studying two factors simultaneously in an experiment has several advantages

- Despite the fact that all efforts are concentrated on just one factor in the 'one factor at a time' approach, it does not mean that more precise information is available about that factor compared to the factorial experiment. This is linked to the phenomenon of hidden replication. If the experiment is set up as in Example 3.2, we can still compare 6 multiparous cows with 6 heifers, although we have also varied the dose compared to the 'one factor at a time' approach. We can use all that information, if we can assume that there is no interaction between the two factors. This can not be done if important interactions exist, because under those circumstances, the difference between heifers and multiparous cows depends on the dose and it does not make sense to consider a general effect of parity. The 'one factor at a time' approach is not capable of picking up such an interaction, which is even more problematic.
- The factorial experiment enables the investigator to test for interaction and quantify the differences.
- The conclusions based on a factorial experiment are more general. In Example 3.2, we can estimate a general difference between heifers and multiparous cows in the absence of interaction, which is valid for the different inoculation dose included in the trial. This is a more general conclusion than the conclusion based on the application of one single inoculation dose.

### 3.2.2 Terms in the fixed effects model with 2 factors

We first discuss the different terms that appear in the factor effects model. Next, we will make use of this information to construct the model in a formal way. *We assume in this section that all population means are known.*

#### Population means of treatment combinations

The population mean of a treatment combination in a factorial study is denoted as $\mu_{ij}$, where $i$ refers to the level of factor $A$ ($i = 1, \ldots, a$) and $j$ to the level of factor $B$ ($j = 1, \ldots, b$). The population mean $\mu_{11}$ corresponds to the mean of all possible heifers infused with a low dose. A set of population means $\mu_{ij}$ for Example 3.2 is given in Table 3.3. The population mean $\mu_{11} = 75$ means for instance that the mean milk reduction for heifers with a low inoculation dose equals 75%.
We can deduce from Table 3.3 that the milk reduction is equal for heifers and multiparous cows. On the other hand, the milk reduction reduces with increasing inoculation dose. Parity has therefore no effect in this example, whereas inoculation dose has. We come to the same conclusion if we compare the column means and the row means.

#### Population means of factor levels

Based on the population means of the treatment combinations, we can define population means for the different levels of a factor. This type of population means is important for

Table 3.3: Population means for the reduction of milk production in non-infected udder quarters as a function of the two factors parity and inoculation dose. There is no difference between heifers and multiparous cows, but there is between inoculation doses.

| | Factor B - inoculation dose | | | |
|---|---|---|---|---|
| Factor A - pariteit | $j = 1$, low | $j = 2$, medium | $j = 3$, high | Row mean |
| $i = 1$, heifer | 75 ($\mu_{11}$) | 42 ($\mu_{12}$) | 3 ($\mu_{13}$) | 40 ($\mu_{1.}$) |
| $i = 2$, multiparous | 75 ($\mu_{21}$) | 42 ($\mu_{22}$) | 3 ($\mu_{22}$) | 40 ($\mu_{2.}$) |
| Column mean | 75 ($\mu_{.1}$) | 42 ($\mu_{.2}$) | 3 ($\mu_{.3}$) | 40 ($\mu_{..}$) |

the further development of the model.

The column mean of the $j^{\text{th}}$ column is given by

$$\mu_{.j} = \frac{\sum_{i=1}^{a} \mu_{ij}}{a}$$

which corresponds to the overal mean of the $j^{\text{th}}$ level of factor $B$. The population mean $\mu_{.1}$, for instance, corresponds to the overal mean of cows with a low inoculation dose, averaged over heifers and multiparous cows.

Similarly, we have the row mean of the $i^{\text{th}}$ row

$$\mu_{i.} = \frac{\sum_{j=1}^{b} \mu_{ij}}{b}$$

The overal population mean can be expressed in different ways

$$\mu_{..} = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} \mu_{ij}}{ab} = \frac{\sum_{i=1}^{a} \mu_{i.}}{a} = \frac{\sum_{j=1}^{b} \mu_{ij}}{b} \tag{3.1}$$

**Main effects**

In the factor effects model, we do not use the population means of the treatment combinations as parameters, but rather parameters that code for the direct effects of the factor levels. We can define such factors as follows.

We start with the effect of parity

$$\alpha_i = \mu_{i.} - \mu_{..}$$

with $\alpha_i$ the main effect of the $i^{\text{th}}$ level of factor $A$. For instance, $\alpha_1$ is the effect of heifer; it corresponds to the difference between the overal population mean for heifers and the overal population mean. For the mastitis example presented in Table 3.3, it follows that $\alpha_1 = \alpha_2 = 0$.

The main effects for the inoculation doses are defined in a similar fashion as

$$\beta_j = \mu_{.j} - \mu_{..}$$

From (3.1) it follows that

$$\sum_{i=1}^{a} \alpha_i = 0 \qquad \sum_{j=1}^{b} \beta_j = 0$$

**Additive factor effects**

When we consider the set of population means $\mu_{ij}$ in Table 3.3, it seems to have an interesting property. All the population means can be obtained by adding the respective factor effects (one for parity and one for dose) to the overal population mean $\mu_{..}$, i.e.,

$$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j$$

If all population means can be written in such a way, we can conclude that there is no interaction between the two factors. It means that the effect of one factor does not depend on the level of the other factor. This is indeed true for Table 3.3, as the difference between heifers and multiparous cows is always zero, regardless the inoculation dose. Or otherwise stated, the difference between two inoculation doses is the same, regardless the parity. This can be also demonstrated graphically as in Figure 3.1. The fact that the three lines corresponding to the three inoculation doses are parallel, signifies that there is no interaction between parity and inoculation doses.

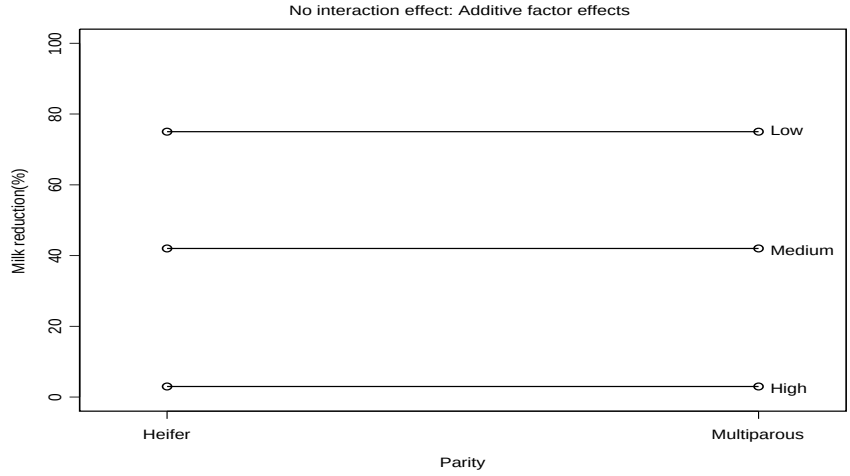



Figure 3.1: The representation of the population means of the mastitis experiment, without parity effect but with strong inoculation dose effect. The parallelism of the lines means that there is no interaction between the two factors.

The lines are not necessarily having slope zero in the absence of interaction. An alternative set of population means is given in Table 3.4 and depicted in Figure 3.2. The milk reduction in this example reduces with increasing inoculation dose, and the line for the multiparous cows lies higher than the one for heifers. Also in this example, there is no interaction between the two factors because the difference between multiparous cows and heifers at each each inoculation dose is exactly the same. This corresponds to the lines in Figure 3.2 being parallel.

**Interacting factors**

Two factors often interact. It means that the difference between two levels of one factor depends on the level of the other factor. Such an example is given in Table 3.5.
It is clear from the table that the difference between heifers and multiparous cows decreases with higher inoculation dose. To describe the population means in Table 3.5 we will need extra parameters.

There is no interaction between two factors when all the population means can be written as

$$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j$$

The factor effects are additive if that is the case, and there is no interaction between the two factors. The difference between the population mean $\mu_{ij}$ and the value $\mu_{..} + \alpha_i + \beta_j$, which is expected if the two factors are additive, is called the interaction effect, or the interaction between the $i^{\text{th}}$ level of factor $A$ and the $j^{\text{th}}$ level of factor $B$. It is denoted as $(\alpha\beta)_{ij}$. The interaction is formally defined as

$$(\alpha\beta)_{ij} = \mu_{ij} - (\mu_{..} + \alpha_i + \beta_j) \tag{3.2}$$

As the model is overparameterised, we also need to add restrictions for the interaction terms.

The restrictions are given by (see Note **??**)

$$\sum_{i=1}^{a} (\alpha\beta)_{ij} = 0 \quad j = 1, \ldots, b$$
$$\sum_{j=1}^{b} (\alpha\beta)_{ij} = 0 \quad i = 1, \ldots, a$$

from which follows that

$$\sum_{i=1}^{a} \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0$$

It is often easier to determine in a figure whether interactions occur. The population means of Table 3.5 are depicted in Figure 3.3. It is clear that the two lines are not

Table 3.4: Population means for reduction of the milk production in non-infected udder quarters as a function of the two factors parity and inoculation dose. There is a difference between heifers and multiparous cows, and between the three inoculation doses, but there is no interaction between the two factors.

| | Factor B - inoculation dose | | | |
|---|---|---|---|---|
| Factor A - parity | $j = 1$, low | $j = 2$, medium | $j = 3$, high | Row mean |
| $i = 1$ heifer | 64 ($\mu_{21}$) | 36 ($\mu_{22}$) | 5 ($\mu_{22}$) | 35 ($\mu_{2.}$) |
| $i = 2$ multiparous | 74 ($\mu_{11}$) | 46 ($\mu_{12}$) | 15 ($\mu_{13}$) | 45 ($\mu_{1.}$) |
| Column mean | 69 ($\mu_{.1}$) | 41 ($\mu_{.2}$) | 10 ($\mu_{.3}$) | 40 ($\mu_{..}$) |

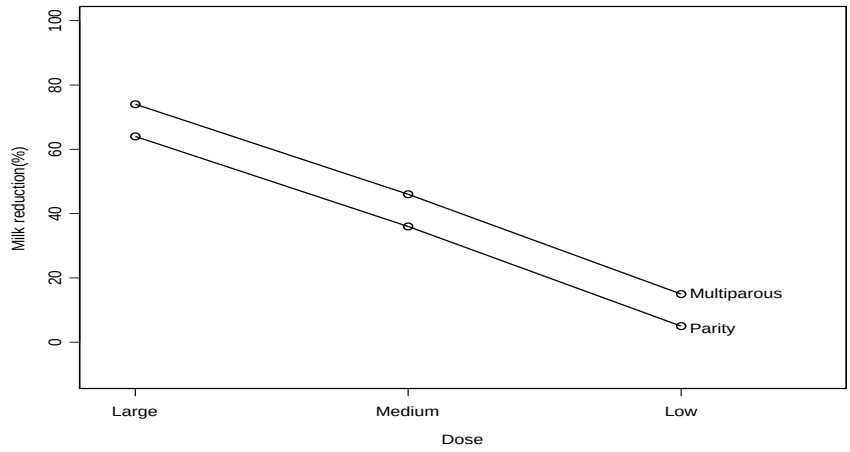



Figure 3.2: The representation of the population means of the mastitis experiment, with effect of parity and inoculation dose. The parallelism of the lines means that there is no interaction between the two factors.

parallel; it follows that the factors parity and inoculation dose are interacting for this set of population means.
Interactions between two factors can take different forms; interaction occurs whenever the difference between two levels of a factor is not the same at each level of the other factor. Consider for instance the interaction presented in Figure 3.4. It can be deduced from this figure that there is almost no difference between heifers and multiparous cows at the medium inoculation dose, that the multiparous cows have a higher milk reduction at the low dose, whereas this effect reverses at the high dose. It is extremely important in such a scenario to test for interactions; if only the main effects of parity and inoculation dose

Table 3.5: Population means for reduction of the milk production in non-infected udder quarters as a function of the two factors parity and inoculation dose. There is a difference between heifers and multiparous cows, and between the three inoculation doses, and also an interaction between the two factors.

| | Factor B - inoculation dose | | | |
|---|---|---|---|---|
| Factor A - parity | $j = 1$, low | $j = 2$, medium | $j = 3$, high | Row means |
| $i = 1$ heifer | 60 ($\mu_{11}$) | 31 ($\mu_{12}$) | 5 ($\mu_{13}$) | 32 ($\mu_{1.}$) |
| $i = 2$ multiparous | 86 ($\mu_{21}$) | 45 ($\mu_{22}$) | 7 ($\mu_{22}$) | 46 ($\mu_{2.}$) |
| Column mean | 73 ($\mu_{.1}$) | 38 ($\mu_{.2}$) | 6 ($\mu_{.3}$) | 39 ($\mu_{..}$) |

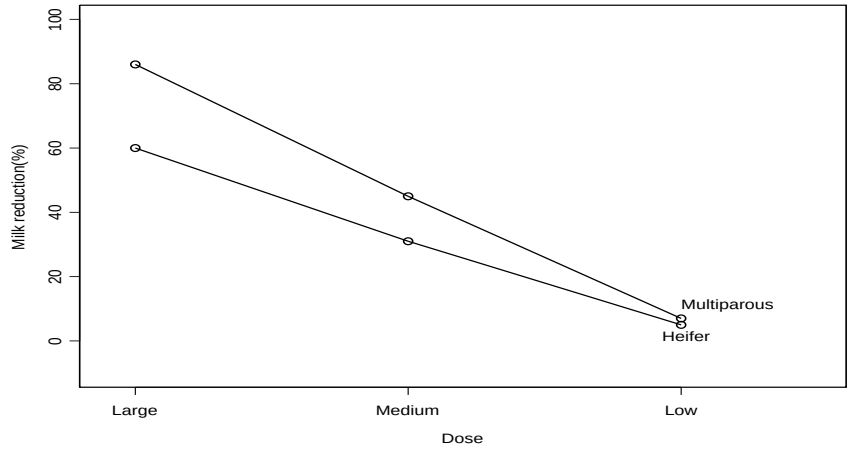



Figure 3.3: Population means for reduction of the milk production in non-infected udder quarters as a function of the two factors parity and inoculation dose. There is a difference between heifers and multiparous cows, and between the three inoculation doses, and also an interaction between the two factors. This corresponds to the fact that the lines are not parallel.

would be evaluated, we would conclude incorrectly that there are no effetcs of these two factors.

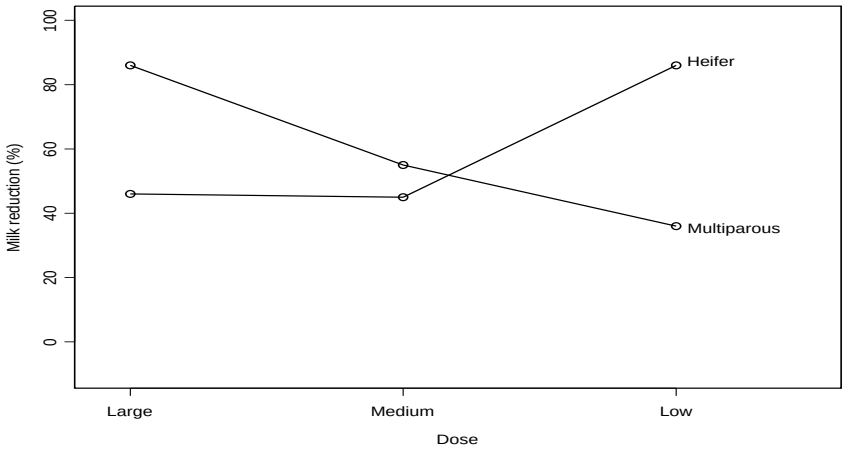



Figure 3.4: Population means for reduction of the milk production in non-infected udder quarters as a function of the two factors parity and inoculation dose. The presence of interaction makes that the milk reduction is highest for multiparous cows if a low inoculation dose is used, and is highest for heifers if a high inoculation dose is used.

The interaction has to be taken into account when reporting the results when the interaction presents itself as in Figure 3.4. We then have to give a description of the differences between heifers and multiparous cows at each of the three inoculation doses separately, or alternatively, the three different inoculation doses can be compared separately for heifers and multiparous cows. This is obviously much more involved than the simple reporting of the main effects of dose and parity. The interaction is sometimes so small to an extent that it is negligible, and that the reporting of main effects remains relevant. We give as an example the set of population means in Table 3.6. In this table, the difference between multiparous cows and heifers equals 11% for the high and low dose and 10% for the medium dose. It therefore makes sense to report that the overal difference between multiparous cows and heifers equals 10.66%; this estimate is a good description for each of the doses.

Table 3.6: Population means for reduction of the milk production in non-infected udder quarters as a function of the two factors parity and inoculation dose. There is an effect of dose and parity, and a small but negligible interaction between the two factors.

| | Factor B - inoculation dose | | | |
|---|---|---|---|---|
| Factor A - parity | $j = 1$, low | $j = 2$, medium | $j = 3$, high | Row mean |
| $i = 1$ heifer | 64 ($\mu_{21}$) | 36 ($\mu_{22}$) | 5 ($\mu_{22}$) | 35 ($\mu_{2.}$) |
| $i = 2$ multiparous | 75 ($\mu_{11}$) | 46 ($\mu_{12}$) | 16 ($\mu_{13}$) | 45.7 ($\mu_{1.}$) |
| Column mean | 69.5 ($\mu_{.1}$) | 41 ($\mu_{.2}$) | 10.5 ($\mu_{.3}$) | 40.3 ($\mu_{..}$) |

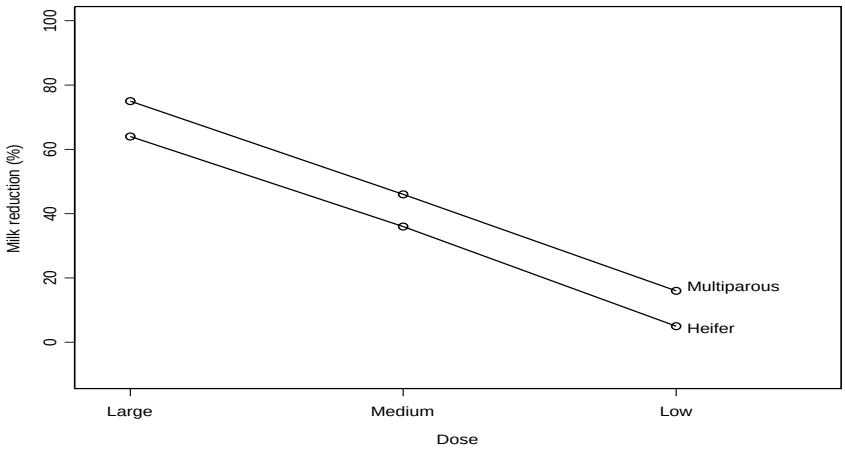



Figure 3.5: Population means for reduction of the milk production in non-infected udder quarters as a function of the two factors parity and inoculation dose. There is an effect of dose and parity, and a small but negligible interaction between the two factors.

## 3.3 Specification of the fixed effects model with 2 factors

The fixed effects model with 2 factors contains a factor $A$ with $a$ levels and a factor $B$ with $b$ levels. All $ab$ treatment combinations are included in the study, and the treatment combination with level $i$ of factor $A$ and level $j$ of factor $B$ appears $n_{ij}$ times. The total number of observations equals $n_{..} = \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}$. The investigator is typically interested in each specific level of the factors $A$ and $B$. The cell means model can also be used here in a similar way as in the previous chapter where it was defined for the fixed effects model with 1 factor. This leads essentially to exactly the same model

$$Y_{ijk} = \mu_{ij} + e_{ijk} \tag{3.3}$$

with $Y_{ijk}$ the $k^{\text{th}}$ observation of the treatment combination with factor $A$ at level $i$ and factor $B$ at level $j$, $\mu_{ij}$ the corresponding population mean and $e_{ijk}$ the deviation of the observation from its population mean. Furthermore, it is assumed that the $e_{ijk}$'s are mutually independent and normally distributed N(0,$\sigma^2$).
We rewrite this model as a factor effects model since we would like to evaluate the main effects of the two factors, and also the interaction between the two factors

$$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} \tag{3.4}$$

where

$$\begin{aligned} \mu_{..} &= \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\mu_{ij}}{ab} \\ \alpha_i &= \mu_{i.} - \mu_{..} \\ \beta_j &= \mu_{.j} - \mu_{..} \\ (\alpha\beta)_{ij} &= \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..} \end{aligned}$$

If we replace the population means in the cell means model (3.3) with (3.4), we obtain the factor effects model

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} \tag{3.5}$$

where

| | |
|---|---|
| $\mu_{..}$ | the overall population mean (constant) |
| $\alpha_i$ | main effect of level $i$ of factor $A$, $\quad i = 1, \ldots, a$ |
| | constants with restriction $\sum_{i=1}^{a}\alpha_i = 0$ |
| $\beta_j$ | main effect of level $j$ of factor $B$, $\quad j = 1, \ldots, b$ |
| | constants with restriction $\sum_{j=1}^{b}\beta_j = 0$ |
| $(\alpha\beta)_{ij}$ | interaction between level $i$ of factor $A$ and level $j$ of factor $B$ |
| | constants with restrictions $\sum_{i=1}^{a}(\alpha\beta)_{ij} = 0 \quad \sum_{j=1}^{b}(\alpha\beta)_{ij} = 0$ |
| $e_{ijk}$ | independent random error term $\sim \mathrm{N}(0, \sigma^2)$, $\quad k = 1, \ldots, n_{ij}$ |

## 3.4 ANOVA for the fixed effects model with 2 factors balanced data with replication

We investigate in this section the simple case of balanced data with replication, i.e., $n_{ij} \equiv n$, with $n > 1$.

The factor effects model was constructed in the previous section assuming that the population means were known. In practice, however, this is never the case and we use the observations to estimate the parameters of the proposed model and to test particular hypotheses stated in terms of the population parameters. We first introduce new notation for the fixed effects model with 2 factors.

The sum of the observations of the $i^{\text{th}}$ level of factor $A$ and the $j^{\text{th}}$ level of factor $B$ is denoted by $Y_{ij.}$, i.e.,

$$Y_{ij.} = \sum_{k=1}^{n} Y_{ijk}$$

The corresponding sample mean is denoted by $\bar{Y}_{ij.}$, i.e.,

$$\bar{Y}_{ij.} = \frac{Y_{ij.}}{n}$$

The sum of the observations of the $i^{\text{th}}$ level of factor $A$ is denoted by $Y_{i..}$, i.e.,

$$Y_{i..} = \sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk}$$

with corresponding sample mean

$$\bar{Y}_{i..} = \frac{Y_{i..}}{bn}$$

In a similar way, the sum of the observations of the $j^{\text{th}}$ level of factor $B$ is denoted by $Y_{.j.}$, i.e.,

$$Y_{.j.} = \sum_{i=1}^{a} \sum_{k=1}^{n} Y_{ijk}$$

with corresponding sample mean

$$\bar{Y}_{.j.} = \frac{Y_{.j.}}{an}$$

The sum of all observations is denoted by $Y_{...}$, i.e.,

$$Y_{...} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} Y_{ijk}$$

which leads to the overall sample mean $\bar{Y}_{...}$, i.e.,

$$\bar{Y}_{...} = \frac{Y_{...}}{nab}$$

We use the least squares (LS) criterion to obtain estimators of the parameters of the factor effects model. We minimize the LS criterion

$$Q = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \mu_{..} - \alpha_i - \beta_j - (\alpha\beta)_{ij})^2$$

subject to the restrictions

$$\sum_{i=1}^{a}\alpha_i = 0 \quad \sum_{j=1}^{b}\beta_j = 0 \quad \sum_{i=1}^{a}(\alpha\beta)_{ij} = 0 \quad \sum_{j=1}^{b}(\alpha\beta)_{ij} = 0$$

This leads to the following estimators (see Note **??**)

$$\begin{aligned}
\hat{\mu}_{..} &= \bar{Y}_{...} \\
\hat{\alpha}_i &= \bar{Y}_{i..} - \bar{Y}_{...} \\
\hat{\beta}_j &= \bar{Y}_{.j.} - \bar{Y}_{...} \\
\hat{(\alpha\beta)}_{ij} &= \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}
\end{aligned}$$

The predicted value of an observation is thus given by

$$\begin{aligned}
\hat{Y}_{ijk} &= \hat{\mu}_{..} + \hat{\alpha}_i + \hat{\beta}_j + \hat{(\alpha\beta)}_{ij} \\
&= \bar{Y}_{...} + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) \\
&= \bar{Y}_{ij.}
\end{aligned}$$

from which follows the estimated residual value

$$\hat{e}_{ijk} = Y_{ijk} - \bar{Y}_{ij.}$$

As before, we use sums of squares, more specifically ratios of mean sums of squares, to test general hypotheses. We now derive the relevant sums of squares for the fixed effects model with 2 factors.

The starting point is the deviation of an observation from the overall sample mean, $(Y_{ijk} - \bar{Y}_{...})$. This deviation can be rewritten as the sum of two terms

$$Y_{ijk} - \bar{Y}_{...} = \left(\bar{Y}_{ij.} - \bar{Y}_{...}\right) + \left(Y_{ijk} - \bar{Y}_{ij.}\right) \tag{3.6}$$

where the first term of the rhs of (3.6) corresponds to the deviation of the estimated sample mean of a particular treatment combination from the overall sample mean, and the seconde term to the deviation of the observation from its estimated sample mean, which corresponds to the residual term $\hat{e}_{ijk}$.

If we square both sides of (3.6) sum these squared terms over all observations, we obtain

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\bar{Y}_{...}\right)^2 = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{Y}_{ij.}-\bar{Y}_{...}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\bar{Y}_{ij.}\right)^2 \tag{3.7}$$

The term on the lhs of (3.7) measures the total variation of the observations, and is therefore called the total sum of squares, $\mathrm{SS_{TOT}}$,

$$\mathrm{SS_{TOT}} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\bar{Y}_{...}\right)^2 \tag{3.8}$$

The first term on the rhs of (3.7) measures in how far the different sample means differ from one another and is therefore called the sum of squares of the treatment, $\mathrm{SS_{TRT}}$:

$$\mathrm{SS_{TRT}} = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{Y}_{ij.}-\bar{Y}_{...}\right)^2 \tag{3.9}$$

The second term on the rhs of (3.7) measures in how far the different observations are spread around their sample mean and is therefore called the sum of squares of the error, $\mathrm{SS_{ERR}}$,

$$\mathrm{SS_{ERR}} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\bar{Y}_{ij.}\right)^2 \tag{3.10}$$

The sum of squares of the treatment describes the variability of the sample means of the treatment combinations; in a factor effects model, however, we wish to split up this variability further into the main factor effects and possibly their interaction. Therefore, the deviation of the sample mean from the overall mean is split up further as follows

$$\bar{Y}_{ij.}-\bar{Y}_{...} = \left(\bar{Y}_{i..}-\bar{Y}_{...}\right) + \left(\bar{Y}_{.j.}-\bar{Y}_{...}\right) + \left(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...}\right) \tag{3.11}$$

We also take the squares on both sides and sum over all the observations. The cross products disappear in the case of balanced data, i.e., the same number of replications for each treatment combination. This is an important and useful property as the original sum of squares of the treatment, $\mathrm{SS_{TRT}}$, can now be written as

$$\mathrm{SS_{BEH}} = \mathrm{SS_A} + \mathrm{SS_B} + \mathrm{SS_{AB}} \tag{3.12}$$

where

$$\mathrm{SS_A} = nb\sum_{i=1}^{a}\left(\bar{Y}_{i..}-\bar{Y}_{...}\right)^2 \tag{3.13}$$

$$\mathrm{SS_B} = na\sum_{j=1}^{b}\left(\bar{Y}_{.j.}-\bar{Y}_{...}\right)^2 \tag{3.14}$$

$$\mathrm{SS_{AB}} = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...}\right)^2 \tag{3.15}$$

It follows from (3.7) and (3.12) that the sum of squares of the interaction can also be obtained as

$$\mathrm{SS_{AB}} = \mathrm{SS_{TOT}} - \mathrm{SS_{ERR}} - \mathrm{SS_A} - \mathrm{SS_B}$$

The sums of squares have the following meaning. $\mathrm{SS_A}$, the factor $A$ sum of squares, measures de variability of the sample means of the different levels of factor $A$, $\bar{Y}_{i..}$. The larger the differences among them, the larger $\mathrm{SS_A}$ will be. The same is true for $\mathrm{SS_B}$ but then for factor $B$. Finally $\mathrm{SS_{AB}}$ measures the variability of the estimated interaction terms $(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})$. A large sum of squares $\mathrm{SS_{AB}}$ reflects the presence of interaction between the two factors.

The split up of $\mathrm{SS_{TRT}}$ in the three sums of squaresmen $\mathrm{SS_A}$, $\mathrm{SS_B}$ and $\mathrm{SS_{AB}}$ is called an orthogonal decomposition because the sum of the three sums of squares equals $\mathrm{SS_{TRT}}$. We do not make use of these sums of squares themselves, but rather of the mean sums of squares. The mean sums of squares are obtained by dividing the sums of squares by their corresponding independent terms in the sum of squares. The number of independent terms is called the degrees of freedom of the SS.

$\mathrm{SS_{TRT}}$ consists of $ab$ terms, of which $ab-1$ are independent from each other as $\sum_{i=1}^{a}\sum_{j=1}^{b} n\left(\bar{Y}_{ij.} - \bar{Y}_{...}\right) = 0$. The mean sum of squares of the treatment is therefore given by

$$\mathrm{MS_{TRT}} = \frac{\mathrm{SS_{TRT}}}{ab-1}$$

$\mathrm{SS_{ERR}}$ consists of $nab$ terms, but within each treatment combination we have that $\sum_{k=1}^{n}\left(Y_{ijk} - \bar{Y}_{ij.}\right) = 0$. Therefore, $ab$ terms can be written as a function of the other terms; this sum of squares has $(n-1)ab$ degrees of freedom, from which follows that

$$\mathrm{MS_{ERR}} = \frac{\mathrm{MS_{ERR}}}{(n-1)ab}$$

$\mathrm{SS_A}$ and $\mathrm{SS_B}$ consist of $a$, resp. $b$ terms, but in each case one of the terms can be written as a function of the other terms using the restriction $\sum_{i=1}^{a}\left(\bar{Y}_{i..} - \bar{Y}_{...}\right) = 0$, resp. $\sum_{j=1}^{b}\left(\bar{Y}_{.j.} - \bar{Y}_{...}\right) = 0$ . Therefore $\mathrm{SS_A}$ and $\mathrm{SS_B}$ have $a-1$, resp. $b-1$ degrees of freedom, from which follows that

$$\mathrm{MS_A} = \frac{\mathrm{SS_A}}{a-1} \tag{3.16}$$

and

$$\mathrm{MS_B} = \frac{\mathrm{SS_B}}{b-1} \tag{3.17}$$

Finally there is $\mathrm{SS_{AB}}$, a sum of squares consisting of $ab$ termes. These terms are subject to $b$ restrictions as

$$\sum_{i=1}^{a}\left(Y_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}\right) = 0 \quad j = 1, \ldots, b$$

and another $a$ restrictions

$$\sum_{j=1}^{b} \left(Y_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}\right) = 0 \quad i = 1, \ldots, a$$

but from the last set of restriction, only $a-1$ are independent as the last restriction follows from the first $b$ restrictions (see Note **??**). Therefore, there are $ab-(b+a-1)=(a-1)(b-1)$ independent terms and we have for the mean sum of squares of the interaction

$$\mathrm{SS_{AB}} = \frac{\mathrm{SS_{AB}}}{(a-1)(b-1)} \tag{3.18}$$

Why we rather use mean sums of squares becomes obvious when we look at the expected values of these mean sums of squares

$$\begin{aligned}
\mathrm{E(MS_{ERR})} &= \sigma^2 \\
\mathrm{E(MS_A)} &= \sigma^2 + nb\frac{\sum_{i=1}^{a} \alpha_i^2}{a-1} \\
\mathrm{E(MS_B)} &= \sigma^2 + na\frac{\sum_{j=1}^{b} \beta_j^2}{b-1} \\
\mathrm{E(MS_{AB})} &= \sigma^2 + n\frac{\sum_{i=1}^{a}\sum_{j=1}^{b} (\alpha\beta)_{ij}^2}{(a-1)(b-1)}
\end{aligned}$$

The mean sum of squares for the factor $A$, resp. $B$, has the same expected value as $\mathrm{MS_{ERR}}$ in the absence of an effect of factor $A$, resp. $B$. Similarly, the sum of squares for the interaction $\mathrm{MS_{AB}}$ has the same expected value as $\mathrm{MS_{ERR}}$ in the absence of an interaction effect, i.e., all $(\alpha\beta)_{ij}$'s equal to zero.

In the statistical analysis of a two way ANOVA, we usually start with testing for interaction.

Under the null hypothesis that the two factors do not interact, written formally as

$$\mathrm{H_0} : \text{all } (\alpha\beta)_{ij} = 0$$

the expected values of $\mathrm{MS_{AB}}$ and $\mathrm{MS_{ERR}}$ are equal, and we thus expect for the ratio

$$F^* = \frac{\mathrm{MS_{AB}}}{\mathrm{MS_{ERR}}}$$

a value close to 1.

Under the alternative hypothesis

$$\mathrm{H_a} : \text{ Not all } (\alpha\beta)_{ij} \text{ equal to zero}$$

the expected value of $\mathrm{MS_{AB}}$ will be larger than the expected value of $\mathrm{MS_{ERR}}$ and we thus expect for the ratio $F^*$ a value larger than 1.

Under the null hypothesis, the ratio $F^*$ has a F-distribution with $(a-1)(b-1)$ and $(n-1)ab$ degrees of freedom

$$F^* \sim \mathrm{F}\left[(a-1)(b-1), (n-1)ab\right]$$

We will mainly use the P-value to decide whether the null hypothesis can be rejected. The P-value is given by

$$\mathrm{P}(\mathrm{F}[(a-1)(b-1), (n-1)ab] \geq f*)$$

with $f*$ the actual value of the test statistic. The null hypothesis is rejected when the P-value is smaller than the significance level $\alpha$.

If the null hypothesis of no interaction can not be rejected, we further test the factors $A$ and $B$.

Under the null hypothesis for the factor $A$, written as

$$\mathrm{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0$$

the expected values of $\mathrm{MS_A}$ and $\mathrm{MS_{ERR}}$ are equal, and we expect for the ratio

$$F^* = \frac{\mathrm{MS_A}}{\mathrm{MS_{ERR}}}$$

a value close to 1.

The ratio $F^*$has an F-distribution with $a-1$ and $(n-1)ab$ degrees of freedom under the null hypothesis

$$F^* \sim \mathrm{F}\left[a-1, (n-1)ab\right]$$

The P-value is given by

$$\mathrm{P}(\mathrm{F}[a-1, (n-1)ab] \geq f*)$$

with $f*$ the actual value of the test statistic. The null hypothesis is rejected when the P-value is smaller than the significance level $\alpha$.

A similar development follows for factor $B$.
The SS, degrees of freedom, MS, F-statistics and P-values are often presented in an analysis of variance or ANOVA ('ANalysis Of VAriance') table, as demonstrated in Example 3.3.

**Example 3.3 Analysis of variance for trypanosomis data set**

We analyse the data of Example 3.1. We fit the fixed effects model with the factors breed and drug and their interaction as terms and the different in PCV before and after the treatment as response variabele. The ANOVA table is given in Table 3.7.

Table 3.7: ANOVA table for the effect of drug and breed on PCV in cows having trypanosomosis.

| Term | SS | df | MS | $f^*$ | P(F$\geq f^*$) |
|---|---|---|---|---|---|
| Breed | 0.441 | 1 | 0.441 | 0.150 | 0.711 |
| Drug | 89.107 | 1 | 89.107 | 29.710 | 0.0006 |
| Breed*Drug | 0.801 | 1 | 0.801 | 0.270 | 0.619 |
| Error | 23.993 | 8 | 2.999 | | |
| Total | 114.342 | 11 | | | |

According to this ANOVA table, the P-value for drug, P($F[1,8] \geq 29.71$)= 0.0006, is much smaller than the default significance level of 5%. We can thus reject the null hypothesis that Berenil and Samorin do not differ. There does not seem to be a difference between the two breeds (P=0.711) and also the interaction between breed and drug is not significant (P=0.619).

■

**Specific comparisons in the fixed effects model with 2 factors without interaction**

The interpretation of the data is quite straightforward if the null hypothesis of no interaction can not be rejected. In such a situation, we can assess the general effect of a particular level of the factor $A$ regardless the level of factor $B$. We therefore merely compare the different levels of factor $A$ with one another, and similarly for factor $B$. We then have much fewer relevant comparisons compared to the situation where interaction is present. The hypotheses can be defined in terms of the $\alpha_i$'s and $\beta_j$'s, or alternatively in terms of the $\mu_{i.}$'s and $\mu_{.j}$'s because for instance $\mu_{i.} = \mu_{..} + \alpha_{i.}$.

As seen before, the estimators of the $\mu_{i.}$'s and $\mu_{.j}$'s are based on the sample means $\bar{Y}_{i..}$ and $\bar{Y}_{.j.}$. We therefore first derive the distributional properties of these estimators as we will need them to define test statistics for hypotheses related to the $\mu_{i.}$'s and $\mu_{.j}$'s.

For the variance of sample mean we have

$$\text{Var}\left(\bar{Y}_{i..}\right) = \frac{\sigma^2}{bn} \qquad \text{Var}\left(\bar{Y}_{.j.}\right) = \frac{\sigma^2}{an}$$

as $\bar{Y}_{i..}$ and $\bar{Y}_{.j.}$ are based on $an$, resp. $bn$ independent observations each with variance $\sigma^2$.

The unknown $\sigma^2$ is replaced by the unbiased estimator $\text{MS}_{\text{ERR}}$ which leads to the sample variances

$$S^2\left(\bar{Y}_{i..}\right) = \frac{\text{MS}_{\text{ERR}}}{bn} \qquad S^2\left(\bar{Y}_{.j.}\right) = \frac{\text{MS}_{\text{ERR}}}{an}$$

Using the fact that

$$\frac{\bar{Y}_{i..} - \mu_{i.}}{S\left(\bar{Y}_{i..}\right)} \sim T[(n-1)ab] \qquad \frac{\bar{Y}_{.j.} - \mu_{.j}}{S\left(\bar{Y}_{.j.}\right)} \sim T[(n-1)ab]$$

we can derive the confidence intervals for the parameters $\mu_{i.}$ and $\mu_{.j}$ as

$$\bar{Y}_{i..} \pm t\left[1-\alpha/2;(n-1)ab\right] s\left(\bar{Y}_{i..}\right)$$
$$\bar{Y}_{.j.} \pm t\left[1-\alpha/2;(n-1)ab\right] s\left(\bar{Y}_{.j.}\right)$$

In the remainder of this section, we will only discuss how hypotheses concerning linear combinations of the set of parameters can be tested, as both contrasts and pairwise comparisons are special cases of linear combinations.

$L$ refers to a linear combination of the population means of the different levels of the factor $A$, $\mu_{i.}$, or the factor $B$, $\mu_{.j}$. We develop the hypothesis test here for factor $A$. The process is identical for factor $B$.

A linear combination of the population means $\mu_{i.}$ is given by

$$L = \sum_{i=1}^{a} c_i \mu_{i.}$$

An unbiased estimator of $L$ is

$$\hat{L} = \sum_{i=1}^{a} c_i \bar{Y}_{i..}$$

Due to the independence of the $\bar{Y}_{i..}$'s it follows that

$$\begin{aligned} \mathrm{Var}\left(\hat{L}\right) &= \sum_{i=1}^{a} c_i^2 \mathrm{Var}\left(\bar{Y}_{i..}\right) \\ &= \frac{\sigma^2}{bn} \sum_{i=1}^{a} c_i^2 \end{aligned}$$

The estimated variance of $\hat{L}$ is then obtained by replacing $\sigma^2$ with its unbiased estimator

$$S^2\left(\hat{L}\right) = \frac{\mathrm{MS_{ERR}}}{bn} \sum_{i=1}^{a} c_i^2$$

As $\hat{L}$ is also a linear combination of independent normally distributed variables it follows that

$$\frac{\hat{L} - L}{s\left(\hat{L}\right)} \sim T[(n-1)ab]$$

The (1-$\alpha$)100 % confidence interval is therefore given by

$$\hat{L} \pm t[1 - \alpha/2; (n-1)ab] s\left(\hat{L}\right)$$

Testing the hypothesis whether the linear combination equals a specific value $c$ is written as the set of hypotheses

$$\mathrm{H}_0 : L = c$$

and

$$\mathrm{H}_\mathrm{a} : L \neq c$$

We use the following test statistic to test this hypothesis

$$T^* = \frac{L - c}{s\left(\hat{L}\right)}$$

which is distributed as $T[(n-1)ab]$ under the null hypothesis. The P-value is given by

$$2 \times \mathrm{P}(T[(n-1)ab] >\mid t^* \mid)$$

We reject the bull hypothesis if the P-value is smaller than the significance level $\alpha$.

**Example 3.4 Comparing the levels of factors for the trypanosomosis data set**

The number of meaningfull contrasts or linear combinations is quite limited as each factor is only occuring at two levels. As an example, we will first test whether the effect of Samorin is larger than that of Berenil. With $\mu_{1.}$ ($\mu_{2.}$) the population mean of cows treated with Samorin (Berenil), we test the following hypotheses

$$\mathrm{H}_0 : \mu_{1.} - \mu_{2.} = 0 \text{ versus } \mathrm{H}_\mathrm{a} : \mu_{1.} - \mu_{2.} > 0$$

The estimated means are given by $\hat{\mu}_{1.} = \bar{Y}_{1..} = 13.00$ and $\hat{\mu}_{2.} = \bar{Y}_{2..} = 7.55$. It follows that $\hat{L} = 5.45$ with $s\left(\hat{L}\right) = 0.999$ and the test statistic thus equals 5.45. The P-value is $\mathrm{P}(T[8] > 5.45)$. We can deduce from Table 9.2 that this P-value is smaller than 0.001, which leads to rejection of the null hypothesis at a significance level of 5%. Remark that testing a two-sided hypothesis with this technique leads to exactly the same P-value as testing the main effect of drug using the F-test; there are only two levels and therefore we test the same hypothesis in both cases.
We could alternatively test whether Samorin increases PCV at least 2% more compared to Berenil; this could be a relevant hypothesis when Samorin would be a much more expensive drug. We test

$$\mathrm{H}_0 : \mu_{1.} - \mu_{2.} = 2 \text{ versus } \mathrm{H}_\mathrm{a} : \mu_{1.} - \mu_{2.} > 2$$

The test statistic now equals 3.45, with corresponding P-value $\mathrm{P}(T[8] > 3.45)$. This P-value is located between 0.001 and 0.005 and thus smaller than 0.05. We reject the null hypothesis and conclude with a confidence of 95% that Samorin increases PCV at least 2% more than Berenil.

■

**Specific comparisons in the fixed effects model with 2 factors in the presence of interaction**

The interpretation of the results is more complex if significant interaction is present. It is no longer meaningfull to summarize the results in terms of general effects of a factor level because the effect depends on the particular level of another factor. We therefore base our comparisons on the $\mu_{ij}$'s, the population means of the treatment combinations. A typical method of analysis evaluates the effects of factor $A$ separately at each level of factor $B$.

We have already shown before that the estimators of the $\mu_{ij}$'s are given by the sample means $\bar{Y}_{ij.}$. We now determine the distributional properties of these estimators as we will need them to test hypotheses regarding the $\mu_{ij}$'s.

The variance of these sample means is

$$\mathrm{Var}\left(\bar{Y}_{ij.}\right) = \frac{\sigma^2}{n}$$

as $\bar{Y}_{ij.}$ is based on $n$ independent observations, each with variance $\sigma^2$.

The unknown parameter $\sigma^2$ is replaced by the unbiased estimator $\mathrm{MS}_{\mathrm{ERR}}$ which leads to the estimator of the variance

$$s^2\left(\bar{Y}_{ij.}\right) = \frac{\mathrm{MS}_{\mathrm{ERR}}}{n}$$

Using the fact that

$$\frac{\bar{Y}_{ij.} - \mu_{ij}}{S\left(\bar{Y}_{ij.}\right)} \sim T[(n-1)ab]$$

we can derive the confidence interval for $\mu_{ij}$ as

$$\bar{Y}_{ij.} \pm t\left[1 - \alpha/2; (n-1)ab\right] s\left(\bar{Y}_{ij.}\right)$$

Also here, we will only discuss how hypotheses concerning linear combinations of the set of parameters can be tested. $L$ refers to a linear combination of the population means of the treatment combinations, $\mu_{ij}$,

$$L = \sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}\mu_{ij.}$$

An unbiased estimator of $L$ is

$$\hat{L} = \sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}\bar{Y}_{ij.}$$

Due to the independence of the $\bar{Y}_{ij.}$'s it follows that

$$\begin{aligned}\mathrm{Var}\left(\hat{L}\right) &= \sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}^2 \mathrm{Var}\left(\bar{Y}_{ij.}\right) \\ &= \frac{\sigma^2}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}^2\end{aligned}$$

The estimated variance of $\hat{L}$ is then obtained by replacing $\sigma^2$ with its unbiased estimator

$$S^2\left(\hat{L}\right) = \frac{\mathrm{MS_{ERR}}}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}^2$$

As $\hat{L}$ is also a linear combination of independent normally distributed variables it follows that

$$\frac{\hat{L} - L}{s\left(\hat{L}\right)} \sim T[(n-1)ab]$$

The 1-$\alpha$ 100 % confidence interval is therefore given by

$$\hat{L} \pm t[1-\alpha/2;(n-1)ab]s\left(\hat{L}\right)$$

Testing that the linear combination equals a specific value $c$ is based on the set of hypotheses

$$\mathrm{H_0} : L = c$$

and

$$\mathrm{H_a} : L \neq c$$

We use the following test statistic to test this hypothesis

$$T^* = \frac{L - c}{s\left(\hat{L}\right)}$$

which is distributed as $T[(n-1)ab]$ under the null hypothesis. The P-value is given by

$$2 \times \mathrm{P}(T[(n-1)ab] >\mid t^* \mid)$$

**Example 3.5 Comparing levels of one factor at a particular level of another factor for the mastitis data set**

We typically start the analysis of variance with testing the main effects and the interaction. We fit the model with parity, dose and their interaction. The ANOVA table is presented in Table 3.8.

Table 3.8: ANOVA tabel for the effect of parity and dose on the reduction of milk production for cows having mastitis.

| Term | SS | df | MS | $f^*$ | P(F $\geq f^*$) |
|---|---|---|---|---|---|
| Dose | 8757.846 | 2 | 4378.923 | 257.06 | <0.0001 |
| Parity | 626.670 | 1 | 626.670 | 36.79 | 0.0009 |
| Dose*Parity | 384.936 | 2 | 192.468 | 11.30 | 0.0092 |
| Error | 102.206 | 6 | 17.034 | | |
| Total | 9871.684 | 11 | | | |

The P-values for dose (P<0.0001), parity (P=0.0009) and the interaction (P=0.0092) are all smaller than 5%, and all have thus a significant effect. We will no longer consider the main effects of dose and parity as the interaction is significant. Similarly, it does not make sense anymore to test hypotheses regarding these main effects. We will now compare the levels of one factor at a particular level of the other factor. We can, for instance, compare the effect of doses separately in heifers and multiparous cows. We will consider here as an example the differences between the three doses for the multiparous cows.

With $\mu_{11}$, $\mu_{12}$ and $\mu_{13}$ the population means of the multiparous cows treated with the low, medium and high dose, resp., we can first test the three pairwise comparisons.
The estimated means are $\hat{\mu}_{11} = \bar{Y}_{11.} = 7.575$, $\hat{\mu}_{12} = \bar{Y}_{12.} = 45.630$ and $\hat{\mu}_{13} = \bar{Y}_{13.} = 87.395$.
The differences, test statistics and P-values are given in Table 3.9.

Table 3.9: Pairwise comparisons between the three doses for multiparous cows.

| Comparison | Difference | $t^*$ | P($T[6] \geq t^*$) |
|---|---|---|---|
| High-Medium | 41.765 | 10.12 | <0.0001 |
| High-Low | 79.820 | 19.34 | <0.0001 |
| Medium-Low | 38.055 | 9.22 | <0.0001 |

The P-values for the three pairwise comparisons are all smaller than the Bonferroni adjusted comparisonwise significance level of 0.05/3=0.0167. All doses differ from each other. ■

## 3.5 ANOVA for the fixed effects model with 2 factors balanced data without replication

We investigate in this section a special case of balanced data with two factors; there is only one observation per treatment combination, i.e., $n_{ij} \equiv n$, with $n = 1$.

Assume that we would fit the same model (3.5) as in the previous section, including the main effects and the interaction

$$\begin{aligned} Y_{ij} &= \mu_{ij} + e_{ij} \\ &= \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ij} \end{aligned}$$

Remark that we can drop the third index because we have only one observation for each treatment combination.

According to the definition of $\mathrm{SS}_{\mathrm{ERR}}$ in (3.10), this leads in this situation to

$$\mathrm{SS}_{\mathrm{ERR}} = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij} - \bar{Y}_{ij}\right)^2$$

but this is equal to zero because the average of the observations of the $ij^{\mathrm{th}}$ treatment combination is the observation $Y_{ij}$ itself as there is only one such observation.

Similarly,

$$df_{\mathrm{ERR}} = ab(n-1) = 0$$

from which follows that

$$\mathrm{MS}_{\mathrm{ERR}} = \frac{\mathrm{MS}_{\mathrm{ERR}}}{df_{\mathrm{ERR}}} = \frac{0}{0}$$

which is undefined.

An alternative way to look at it is that each observation in model (3.5) is correctly estimated by the model without any need for the random error term $e_{ij}$. Therefore, we do not have any observations left to estimate the underlying variance of the observations, $\sigma^2$, which is required to define F-tests.

We can, however, fit a simplified model to such data. Assume that there is no interaction between the two factors, i.e., $(\alpha\beta)_{ij} = 0, i = 1\ldots, a; j = 1, \ldots, b$, then the model simplifies to

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + e_{ij}$$

Furthermore, the expected value of different sums of squares under the assumption of no interaction are given by

$$\begin{aligned} \mathrm{E}(\mathrm{MS_A}) &= \sigma^2 + nb\frac{\sum_{i=1}^{a}\alpha_i^2}{a-1} \\ \mathrm{E}(\mathrm{MS_B}) &= \sigma^2 + na\frac{\sum_{j=1}^{b}\beta_j^2}{b-1} \\ \mathrm{E}(\mathrm{MS_{AB}}) &= \sigma^2 \end{aligned}$$

Therefore, $\mathrm{E}(\mathrm{MS_{AB}})$ can now be used as an estimator of $\sigma^2$ in the denominator of the F-test.

Under the null hypothesis for the factor $A$, written as

$$\mathrm{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0$$

the expected values of $\mathrm{MS_A}$ and $\mathrm{MS_{AB}}$ are equal and we have for the ratio

$$F^* = \frac{\mathrm{MS_A}}{\mathrm{MS_{AB}}} \sim \mathrm{F}\left[a-1, (n-1)ab\right]$$

The P-value is given by $\mathrm{P}(\mathrm{F}[a-1,(n-1)ab] \geq f*)$ with $f*$ the actual value of the test statistic. The null hypothesis is rejected when the P-value is smaller than the significance level $\alpha$.

A similar development follows for factor $B$.
This analysis is thus based on a model without interaction, and the interaction terms thus become the random error terms. It is therefore to asess what happens when the assumption of no interaction is violated. In such a case we have

$$\mathrm{E}(\mathrm{MS_{AB}}) = \sigma^2 + n\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$$

We thus expect -and on average have- a larger value for $\mathrm{E}(\mathrm{MS_{AB}})$ than the underlying variance $\sigma^2$. This results in a larger value of the denominator of the F-statistic, and a smaller value for the F-statistic. We therefore will reject the null hypothesis less frequently in the presence of interaction. This type of test is called a conservative test. In the presence of interaction we pay the price of less power. Such a conservative test, however, is acceptable as we can only draw a conclusion if the null hypothesis is rejected.

**Example 3.6 ANOVA for the mastitis data set with one observation per treatment combination**

Assume that the mastitis experiment was set up with only one observation for each combination of parity and inoculation dose, resulting in the data set presented in Table 3.10.

Table 3.10: Milk production in non-infected udder quarters just before (Milk0) and 48 hours after (Milk48) infection as a function of the two factors parity and inoculation dose with only one observation per treatment combination.

| Cowid | Parity | Inoculation dose | Milk0 | Milk48 | Reduction |
|---|---|---|---|---|---|
| 1 | heifer | high | 32.4 | 30.2 | 6.79 |
| 2 | heifer | medium | 29.3 | 20.5 | 30.03 |
| 3 | heifer | low | 31.3 | 14.5 | 53.67 |
| 4 | multiparous | high | 42.4 | 39.5 | 6.84 |
| 5 | multiparous | medium | 45.2 | 23.9 | 47.12 |
| 6 | multiparous | low | 41.5 | 6.7 | 83.86 |

Table 3.11: ANOVA tabel for the effect of parity, dose and their interaction on the reduction of milk production for cows having mastitis.

| Term | SS | df | MS | $f^*$ | P(F $\geq f^*$) |
|---|---|---|---|---|---|
| Dose | 3838.62 | 2 | 1919.31 | . | . |
| Parity | 373.35 | 1 | 373.35 | . | . |
| Dose*Parity | 228.40 | 2 | 114.20 | . | . |
| Error | 0 | 0 | . | | |
| Total | 4440.38 | 5 | | | |

If we fit a model including the interaction terms, we obtain the ANOVA Table 3.11.
It is clear that no degrees of freedom are left over for the sum of squares of the error, and therefore no F-tests can be shown for the factors of interest. Excluding the interaction terms from the model leads to the ANOVA Table 3.12.

Table 3.12: ANOVA tabel for the effect of parity and dose on the reduction of milk production for cows having mastitis.

| Term | SS | df | MS | $f^*$ | P(F $\geq f^*$) |
|---|---|---|---|---|---|
| Dose | 3838.62 | 2 | 1919.31 | 16.81 | 0.0562 |
| Parity | 373.35 | 1 | 373.35 | 3.27 | 0.2123 |
| Error | 228.40 | 2 | 114.20 | | |
| Total | 4440.38 | 5 | | | |

The sum of squares of the error in Table 3.12 corresponds to the sum of squares of the interaction in Table 3.11. Thus, we now have an error sum of squares and F-statistics

can be obtained. It seems that neither parity (P=0.2123) nor dose (P=0.0562) have a significant effect on milk reduction. We have to take into account, however, that there is not much replication, which might be the main reason why we can not reject the null hypothesis.

■

Although we can not test whether there is interaction using the full model (3.5), we can fit a model which includes a specific type of interaction. Assume that the interaction is of form

$$(\alpha\beta)_{ij} = \delta\alpha_i\beta_j$$

with $D$ a constant, then we have the folllowing model

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + \delta\alpha_i\beta_j + e_{ij} \tag{3.19}$$

In this model, we only use 1 degree of freedom to estimate the parameter $\delta$ for the interaction instead of the usual $(a-1)(b-1)$ degrees of freedom for the full model. This leads to the name of the test, the "Tukey one degree of freedom" test.
This model can be fitted (given that $a > 2$ or $b > 2$).

We will test the hypothesis

$$\mathrm{H}_0 : \delta = 0 \text{ versus } \mathrm{H}_\mathrm{a} : \delta \neq 0$$

The development of the test is based on similar reasoning as in the regression model

$$Y_i = \mu + \gamma x_i + e_i$$

For such a regression model, the sum of squares related to the parameter $\gamma$ is given by

$$SS_\gamma = \hat{\gamma}^2 \sum (x_i - \bar{x}_.)^2 \text{ with } \hat{\gamma} = \frac{\sum (x_i - \bar{x}_.) y_i}{\sum (x_i - \bar{x}_.)^2}$$

Analogously, we find for slope $\delta$ in model (3.19) and assuming that the $\alpha_i$'s and $\beta_j$'s are known

$$\begin{aligned} SS_\delta &= \hat{\delta}^2 \sum_{i=1}^{a} \sum_{j=1}^{b} \left( \alpha_i\beta_j - \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} \alpha_i\beta_j}{ab} \right)^2 \\ &= \hat{\delta}^2 \sum_{i=1}^{a} \sum_{j=1}^{b} (\alpha_i\beta_j)^2 \\ &= \hat{\delta}^2 \sum_{i=1}^{a} \alpha_i^2 \sum_{j=1}^{b} \beta_j^2 \end{aligned} \tag{3.20}$$

with

$$\begin{aligned}\hat{\delta} &= \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\alpha_i\beta_j - \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\alpha_i\beta_j}{ab}\right)y_{ij}}{\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\alpha_i\beta_j - \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\alpha_i\beta_j}{ab}\right)^2} \\ &= \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\alpha_i\beta_j y_{ij}}{\sum_{i=1}^{a}\alpha_i^2\sum_{j=1}^{b}\beta_j^2} \end{aligned} \tag{3.21}$$

Plugging in (3.21)in (3.20) and replacing the $\alpha_i$'s and $\beta_j$'s with their estimates

$$\begin{aligned}\hat{\alpha}_i &= \bar{y}_{i.} - \bar{y}_{..} \\ \hat{\beta}_j &= \bar{y}_{.j} - \bar{y}_{..}\end{aligned}$$

we obtain

$$\begin{aligned}SS_\delta &= \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\alpha_i\beta_j y_{ij}\right)^2}{\left(\sum_{i=1}^{a}\alpha_i^2\sum_{j=1}^{b}\beta_j^2\right)^2}\sum_{i=1}^{a}\alpha_i^2\sum_{j=1}^{b}\beta_j^2 \\ &= \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}\alpha_i\beta_j y_{ij}\right)^2}{\sum_{i=1}^{a}\alpha_i^2\sum_{j=1}^{b}\beta_j^2} \\ &= \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{i.}-\bar{y}_{..})(\bar{y}_{.j}-\bar{y}_{..})y_{ij}\right)^2}{\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2\sum_{j=1}^{b}(\bar{y}_{.j}-\bar{y}_{..})^2}\end{aligned}$$

The total sum of squares is now split up as

$$\mathrm{SS_{TOT}} = \mathrm{SS_A} + \mathrm{SS_B} + \mathrm{SS}_\delta + \mathrm{SS_{ERR}} \tag{3.22}$$

with resulting relevannt mean sum of squares

$$\begin{aligned}\mathrm{MS}_\delta &= \mathrm{SS}_\delta \\ \mathrm{MS_{ERR}} &= \frac{\mathrm{SS_{ERR}}}{ab-a-b}\end{aligned}$$

Under the null hypothesis $H_0 : \delta = 0$, the expected value of both $MS_\delta$ and $MS_{ERR}$ equals $\sigma^2$ and for the ratio of the two mean sums of squares we have

$$F^* = \frac{MS_\delta}{MS_{ERR}} \sim F\,[1, ab - a - b]$$

with P-value given by $P(F[1, (ab - a - b)] \geq f^*)$.

**Example 3.7 Testing for interaction for the mastitis data set with one observation per treatment combination**

We fit the model () to the mastitis data with dose and parity as main factors resulting in the ANOVA Table 3.13.

Table 3.13: ANOVA tabel for the effect of parity, dose and simple interaction on the reduction of milk production for cows having mastitis.

| Term | SS | df | MS | $f^*$ | $P(F \geq f^*)$ |
|---|---|---|---|---|---|
| Dose | 3838.62 | 2 | | | |
| Parity | 373.35 | 1 | | | |
| $\delta$ | 227.56 | 1 | 227.56 | 270.75 | 0.038 |
| Error | 0.84 | 1 | 0.84 | | |
| Total | 4440.38 | 5 | | | |

■

**Note 3.4 Restrictions op de interactietermen** ■

**Note 3.5 Schatters van modelparameters** ■

**Note 3.6 Restrictions op de interactietermen** ■

**Note 3.7 sum of squaresmen gebaseerd op KKW gemiddelden** ■

# Chapter 4

# Balanced block designs

## 4.1 Introduction

The block concept is introduced in this chapter, and its usefulness in experimental design demonstrated for the case of balanced block designs. In Section 4.2 we further elaborate on the block concept. Next, we move to the different types of balanced block designs in Section 4.3, comprising the randomised and generalised complete block designs. The analysis of variance of such data is explained in Section 4.4. We will fall back on the balancedness of the data to use the orthogonal decomposition of the total sum of squares, which leads to a similar analysis as the one for a fixed effects model with two factors.

## 4.2 The blocking principle in experimental design

Experimental units are often heterogeneous. This heterogeneity is one of the main factors determining the power of the experiment, as it provides the background variability against which the factor that was randomly assigned to the experimental unit needs to be assessed. Sometimes, the heterogeneity can be decreased by consideing technical improvements. For instance, the measuring device can be made more precise so that the measurement error, and thus the variability between the experimental units, decreases.
Other factors, however, are closely linked to characteristics for the experimental unit and make that the experimental units differ from each other. For instance, in animal experiments, we might have animals coming from different litters, with litter having an important effect on the variable of interest. As we need a sufficiently large number of animals, we can not just do with one litter.
In cases where a factor that causes heterogeneity between experimental units is known, such as the litter in the example above, the use of blocks makes that the part of the heterogeneity between the experimental units related to that factor can be removed from the background variability by making proper use of the blocking principle.

Experimental units which are more alike because they share the level of such a factor, e.g., the same litter, are collected together in a block. Next, in the most optimal block setting, all treatments occur at least once in each block, so that the comparison between treatments can be based on experimental units residing in the same block, that is the

mechanics by which the variability between blocks is taken out.

Blocking can be done in many different ways, and we consider some in Example 4.1 below. Furthermore, we study in this chapter only designs in which all treatments occur at least once in each block. Such blocking structurs, however, are not always possible in practice. If less experimental units per block are available than the number of different treatments, it is important that the treatments are randomly assigned in a particular fashion. The balanced incomplete block (BIB) design is such a design with some optimal properties. The BIB design, however, is not a balanced block design; its discussion is postponed to Chapter **??**.

In some other cases, there are two blocking factors. Although all treatments might occur at least at each level of the two blocking factors separately, it is not necesarilly true for the combinations of the two blocking factors. To handle two such blocking factors, we need other designs with other analysis strategies. Latin square designs typically have two blocking factors, and will be discussed in Chapter **??**.

**Example 4.1 Different block designs for the study of genetically modified rice plants**

We want to assess the effect of the insertion of a gene into a rice plant. Due to Mendelian seggregation, rice plants come in three different types. The wild type (W), not containing the insertion at all; the haploid type (H), containing the insertion on one of the two chromosomes, and finally the diploid type (D), containing the insertion on both of its chromosomes. We want to compare these three plant types. We have a greenhouse that can be split up in different ways to obtain plots and blocks. We assume that there is a temperature gradient from the door of the greenhouse at the left towards the right side of the greenhouse.

We consider first a situation in which we make 9 different plots. In the most simple setting, we randomly assign each of the treatments to 3 plots, as shown in Figure 4.1.a. This design leads to a proper statistical analysis due to the randomisation, but does not take care of the temperature gradient in the greenhouse. In Figure 4.1.b, we present the situation where we first make blocks, orthogonal to the source of variation, and next we assign each of the three treatments to one of the plots within a block. This results in a design called the randomised complete block design: each block contains each treatment exactly once. Finally, assume there is also a light gradient going from front to rear, so that an additional blocking factor needs to be taken into consideration. A proper design for such situation is presented in Figure 4.1.c which will be discussed further in Chapter **??**. For the time being, remark that for the blocks defined by the temperature gradient solely, and also for the blocks defined by the light gradient solely, we have a randomised complete block design, but obviously not for the combination of the blocks defined by temperature gradient and light gradient, as there is then only one experimental unit for such temperature-light combination.

Next consider the situation where we double the number of plots to 18. There are now dif-

ferent options to assign treatments to plots. First consider Figure 4.2.a, where we actually assigned a particular treatment randomly either to the two higher, middle or lower plots. Therefore, the experimental unit consists here of a collection of two plots, and the two plots provide two repeated measurements within that experimental unit. A much better design is presented in Figure 4.2.b, where each treatment is assigned twice at random to two plots. In this setting the plot is the experimental unit, and this type of design is called a generalised complete block design because each treatment is assigned more than once (but each treatment the same number of times) to a plot in a block. Alternatively, we could have made 6 blocks, as in Figure 4.2.c, which makes the design again a randomised complete block design.

Finally we consider the situation where the number of plots in a block is not a multiple of the number of treatments. This can either mean that there are less or more treatments than experimental units in a block. First consider Figure 4.3.a, where the blocks contain only two experimental units. Remark that the random assignment has taken place in a way so that each pairwise comparison appears the same number of times together in a block, in this case twice. Such designs are called balanced incomplete block designs and will be studied in Chapter **??**. Alternatively, we might have more experimental units in a block than treatments, but less than a multiple of the treatments. Instead of not using these plots, it might be advantageous to repeat one of the treatments in each block. In Figure 4.3.b, the treatment W is appearing twice in each block; as we want to compare both H and D with W, it is best to have an extra replication for W. Compared to the BIB design, this design is still a balanced block design, as will be discussed in Section 4.3.3.
■

Figure 4.1: Three designs to compare treatments W, H and D with (a) completely randomised design, (b) the randomised complete block design and (c) the latin square.

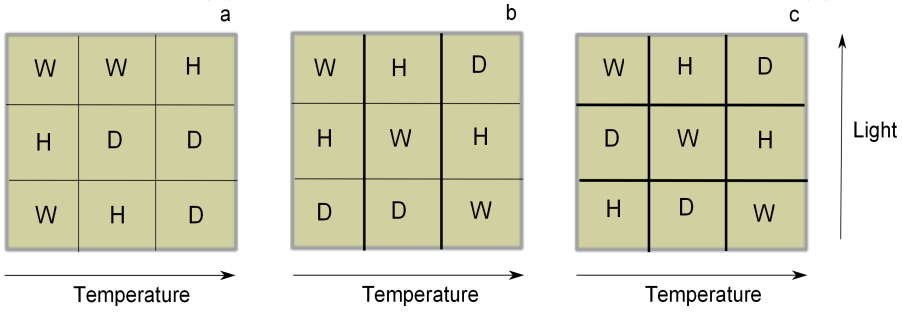



Figure 4.2: Three designs to compare treatments W, H and D with (a) the randomised complete block design with repeated measurements, (b) the generalised block design and (c) the randomised complete block design.

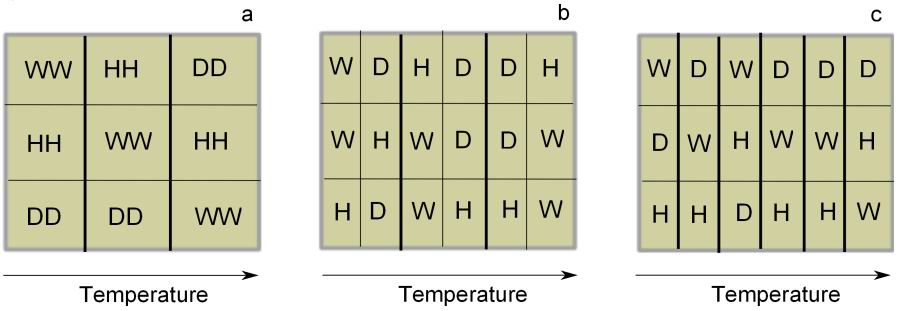
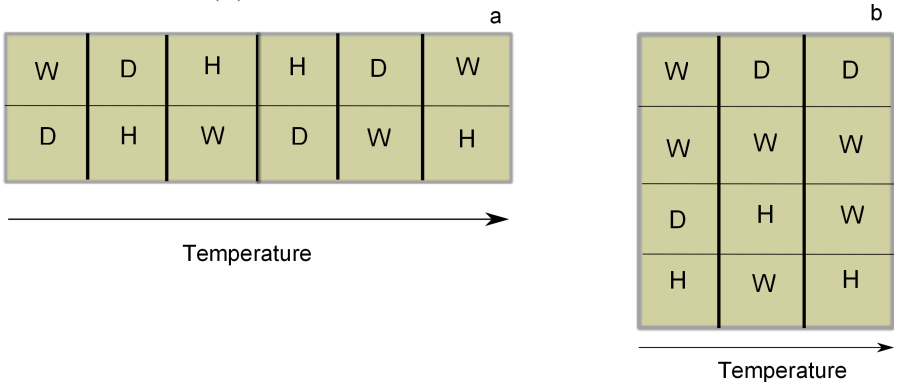



Figure 4.3: Two designs to compare treatments W, H and D with (a) balanced incomplete block design and (b) a balanced block design with W appearing twice in each block.



## 4.3 Balanced block design types

Block designs come in many different forms. Once experimental units are grouped together in blocks, we can talk about a block design. There are, however, some standard block designs that are often used because they have some optimal properties. The randomised complete block design is certainly the most popular block design. An extension of this design is the generalised complete block design. We will only discuss balanced block designs in this chapter and below we will study the requirements for a balanced block design.

### 4.3.1 Randomised complete block design

The randomised complete block design is the simplest block design. The block has the same number of experimental units as the number of treatments and each treatment is assigned once and only once to one of the experimental units in a random fashion. Examples of randomised complete block designs are given in Figures 4.1.b and 4.2.c. We present data for the setup presented in Figure 4.1.b in Table 4.1.

Table 4.1: Randomised complete block design to compare different rice plant types with 1000-kernel weight as response variable

| Block | Plant type | 1000-kernel weight (g) |
|---|---|---|
| 1 | W | 22.4 |
| 1 | H | 24.8 |
| 1 | D | 25.2 |
| 2 | W | 27.3 |
| 2 | H | 28.6 |
| 2 | D | 28.4 |
| 3 | W | 24.5 |
| 3 | H | 25.8 |
| 3 | D | 26.2 |

### 4.3.2 Generalised complete block design

The generalised complete block design is an extension of the randomised complete block design, the simplest block design. The number of experimental units in the block is a multiple $d$ of the number of treatments and each treatment is assigned $d$ times to an experimental unit in the block in a random fashion. So each treatment is appearing the same number of times in each block. An example of a generalised complete block design is given in Figure 4.2.b. We present data for the setup presented in Figure 4.2.b in Table 4.2.

Table 4.2: Generalised complete block design to compare different rice plant types with 1000-kernel weight as response variable

| | | 1000-kernell gewicht (g) | |
|---|---|---|---|
| Block | Plant type | Observation 1 | Observation 2 |
| 1 | W | 22.4 | 23.0 |
| 1 | H | 24.8 | 25.2 |
| 1 | D | 25.2 | 25.6 |
| 2 | W | 27.3 | 26.3 |
| 2 | H | 28.6 | 27.5 |
| 2 | D | 28.4 | 29.3 |
| 3 | W | 24.5 | 24.8 |
| 3 | H | 25.8 | 25.9 |
| 3 | D | 26.2 | 25.3 |

### 4.3.3 Other balanced block designs

The requirements on the block size in the two previous block designs is quite strict. Each block needs to contain exactly the same number of experimental units, and the numer of experiments needs to be a multiple of the number of treatments. Blocks do not always come in that format. In this section, we address the question whether it is possible to set up balanced block designs that go beyond the RCB and GCB. We first need to explain what we mean by balanced block designs. A block design is balanced whenever we cannot deduce information for the treatment effects by comparing block means. This can be achieved by ensuring that the proportion of occurence of each treatment is the same in each block. This principle allows us to have blocks of different sizes and also to have blocks in which one treatment occurs more frequently than another treatment. Consider, for instance, the design in Figure 4.3.b. The treatment W is appearing twice in each block, its proportion of occurence in each block is therefore 50 %. The block mean differences are not informative for the treatment effect. Compare this with the balanced incomplete block design presented in Figure 4.3.a. The difference between the first block mean and the second block mean contains information on the difference between W and H. Therefore, this is not a balanced block design according to the definition above.

Using the principle of balance, we can also have blocks of different sizes. Consider the design in Figure 4.4. We have three blocks, one of size 8 and two of size 4. This is a balanced block design as the proportion of occurence of each treatment is the same in each block.

We present data for the setup presented in Figure 4.3.c in Table 4.3.

Figure 4.4: A balanced block design to compare treatments W, H and D. The block sizes differ and W is occuring twice as much as the two other treatments. The proportion of occurence of each treatment, however, is the same in each block.

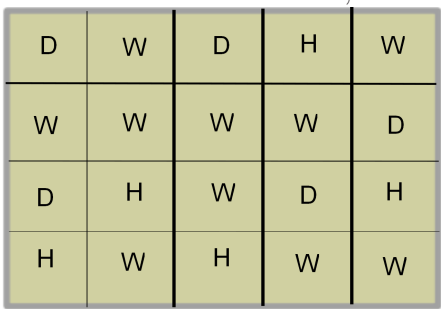



Table 4.3: Generalised complete block design to compare different rice plant types with 1000-kernel weight as response variable

| Block | Plant type | 1000-kernell gewicht (g) |
|---|---|---|
| 1 | W | 22.4 |
| 1 | W | 23.0 |
| 1 | H | 24.8 |
| 1 | D | 25.2 |
| 2 | W | 27.3 |
| 2 | W | 26.3 |
| 2 | H | 28.6 |
| 2 | D | 28.4 |
| 3 | W | 24.5 |
| 3 | W | 24.8 |
| 3 | H | 25.8 |
| 3 | D | 26.2 |

## 4.4 ANOVA for balanced block designs

The analysis of variance of balanced block designs is rather straightforward because it is similar to the analysis of variance of a fixed effects model with two factors. The orthogonal decomposition of the total sum of squares and the corresponding ANOVA table is exactly the same as in Section 3.5 for the randomised complete block design and as in Section 3.4 for the generalised complete block design. The main difference is the interpretation and use of the block effect: we are not really interested in the block effect; we include it in the model to get rid of part of the heterogeneity between the experimental units.
We start with the analysis of the randomised complete block design. Obviously, we can not include an interaction term because there is only 1 observation for each block-treatment combination. We therefore have to assume that blocks are not interacting with treatment and have the following model

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + e_{ij} \tag{4.1}$$

where

| | |
|---|---|
| $\mu_{..}$ | the overall population mean (constant) |
| $\alpha_i$ | main effect of level $i$ of factor $A$, $\quad i = 1, \ldots, a$ |
| | constants with restriction $\sum_{i=1}^{a} \alpha_i = 0$ |
| $\beta_j$ | effect of block $j$, $\quad j = 1, \ldots, b$ |
| | constants with restriction $\sum_{j=1}^{b} \beta_j = 0$ |
| $e_{ij}$ | independent random error term $\sim \mathrm{N}(0, \sigma^2)$ |

Under the null hypothesis for the factor $A$, written as

$$\mathrm{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0$$

the expected values of $\mathrm{MS_A}$ and $\mathrm{MS_{AB}}$ are equal and we have for the ratio

$$F^* = \frac{\mathrm{MS_A}}{\mathrm{MS_{AB}}} \sim \mathrm{F}\left[a-1, (n-1)ab\right]$$

The P-value is given by $\mathrm{P}(\mathrm{F}[a-1, (n-1)ab] \geq f*)$ with $f*$ the actual value of the test statistic. The null hypothesis is rejected when the P-value is smaller than the significance level $\alpha$.

The block factor is normally not tested, but a similar test can be defined as for factor A if required.

**Example 4.2 ANOVA for randomised complete block design for rice plants data set**

We analyse the data of Table 4.1. We fit the model with plant type and block as factor and 1000-kernel weight as response variable. The ANOVA table is given in Table 4.4 According to this ANOVA table, the P-value for plant type is 0.0118. We can thus reject

Table 4.4: ANOVA table for the effect of plant type on 1000-kernel weight taking into account a block effect

| Term | SS | df | MS | $f^*$ | $\mathrm{P}(\mathrm{F} \geq f^*)$ |
|---|---|---|---|---|---|
| Plant type | 5.787 | 2 | 2.893 | 10.17 | 0.0118 |
| Block | 36.036 | 3 | 12.012 | 42.23 | 0.0002 |
| Error | 1.707 | 6 | 0.284 | | |
| Total | 43.529 | 11 | 3.957 | | |

the null hypothesis that there is no difference between the plant types.

■

In the case of a generalised complete block design on the other hand, we have two or more observations for each block-treatment combination. Therefore, the block-treatment interaction can be added to the model resulting in the same model as in Section 3.3

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} \tag{4.2}$$

where

| | |
|---|---|
| $\mu_{..}$ | the overall population mean (constant) |
| $\alpha_i$ | main effect of level $i$ of factor $A$, $\quad i = 1, \ldots, a$ |
| | constants with restriction $\sum_{i=1}^{a} \alpha_i = 0$ |
| $\beta_j$ | effect of block $j$, $\quad j = 1, \ldots, b$ |
| | constants with restriction $\sum_{j=1}^{b} \beta_j = 0$ |
| $(\alpha\beta)_{ij}$ | interaction between level $i$ of factor $A$ and block $j$ |
| | constants with restrictions $\sum_{i=1}^{a}(\alpha\beta)_{ij} = 0 \quad \sum_{j=1}^{b}(\alpha\beta)_{ij} = 0$ |
| $e_{ijk}$ | independent random error term $\sim \mathrm{N}(0, \sigma^2)$, $\quad k = 1, \ldots, n_{ij}$ |

F-tests and corresponding P-values are the same as in Section 3.4; its use is demonstrated in Example 4.3

**Example 4.3 ANOVA for generalised complete block design for rice plants data set**

We analyse the data of Table 4.2. We fit the model with plant type and block as factor and 1000-kernel weight as response variable. The ANOVA table is given in Table 4.5. According to this ANOVA table, the P-value for plant type is 0.0008. We can thus reject

Table 4.5: ANOVA table for the effect of plant type on 1000-kernel weight taking into account a block effect

| Term | SS | df | MS | $f^*$ | P(F$\geq f^*$) |
|---|---|---|---|---|---|
| Plant type | 14.131 | 2 | 7.065 | 17.77 | 0.0003 |
| Block | 53.202 | 3 | 17.734 | 44.61 | <0.0001 |
| Plant type*Block | 3.416 | 6 | 0.569 | 1.43 | 0.280 |
| Error | 4.770 | 12 | 0.397 | | |
| Total | 75.518 | 23 | 3.283 | | |

the null hypothesis that there is no difference between the plant types.

■

biomedische statistiek 2013-2014/figtabz.eps
biomedische statistiek 2013-2014/figtabz.eps

Table 9.1: Standaard normale verdeling

| | Tweede decimaal van $z$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $z$ | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| 0.9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 1.3 | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| 1.4 | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9279 | .9292 | .9306 | .9319 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| 1.7 | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| 1.8 | .9641 | .9649 | .9656 | .9664 | .9671 | .9678 | .9686 | .9693 | .9699 | .9706 |
| 1.9 | .9713 | .9719 | .9726 | .9732 | .9738 | .9744 | .9750 | .9756 | .9761 | .9767 |
| 2.0 | .9772 | .9778 | .9783 | .9788 | .9793 | .9798 | .9803 | .9808 | .9812 | .9817 |
| 2.1 | .9821 | .9826 | .9830 | .9834 | .9838 | .9842 | .9846 | .9850 | .9854 | .9857 |
| 2.2 | .9861 | .9864 | .9868 | .9871 | .9875 | .9878 | .9881 | .9884 | .9887 | .9890 |
| 2.3 | .9893 | .9896 | .9898 | .9901 | .9904 | .9906 | .9909 | .9911 | .9913 | .9916 |
| 2.4 | .9918 | .9920 | .9922 | .9925 | .9927 | .9929 | .9931 | .9932 | .9934 | .9936 |
| 2.5 | .9938 | .9940 | .9941 | .9943 | .9945 | .9946 | .9948 | .9949 | .9951 | .9952 |
| 2.6 | .9953 | .9955 | .9956 | .9957 | .9959 | .9960 | .9961 | .9962 | .9963 | .9964 |
| 2.7 | .9965 | .9966 | .9967 | .9968 | .9969 | .9970 | .9971 | .9972 | .9973 | .9974 |
| 2.8 | .9974 | .9975 | .9976 | .9977 | .9977 | .9978 | .9979 | .9979 | .9980 | .9981 |
| 2.9 | .9981 | .9982 | .9982 | .9983 | .9984 | .9984 | .9985 | .9985 | .9986 | .9986 |
| 3.0 | .9987 | .9987 | .9987 | .9988 | .9988 | .9989 | .9989 | .9989 | .9990 | .9990 |

Table 9.2: t-verdeling

| | P | | | | | | |
|---|---|---|---|---|---|---|---|
| $vg$ | .750 | .900 | .950 | .975 | .990 | .995 | .999 |
| 1 | 1.000 | 3.078 | 6.314 | 12.706 | 31.821 | 63.656 | 318.289 |
| 2 | 0.816 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.328 |
| 3 | 0.765 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.214 |
| 4 | 0.741 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 |
| 5 | 0.727 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.894 |
| 6 | 0.718 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 |
| 7 | 0.711 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 |
| 8 | 0.706 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 |
| 9 | 0.703 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 |
| 10 | 0.700 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 |
| 11 | 0.697 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 |
| 12 | 0.695 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 |
| 13 | 0.694 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 |
| 14 | 0.692 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 |
| 15 | 0.691 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 |
| 16 | 0.690 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 |
| 17 | 0.689 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 |
| 18 | 0.688 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 |
| 19 | 0.688 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 |
| 20 | 0.687 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 |
| 21 | 0.686 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 |
| 22 | 0.686 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 |
| 23 | 0.685 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 |
| 24 | 0.685 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 |
| 25 | 0.684 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 |
| 26 | 0.684 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 |
| 27 | 0.684 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 |
| 28 | 0.683 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 |
| 29 | 0.683 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 |
| 30 | 0.683 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 |
| 40 | 0.681 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 |
| 60 | 0.679 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 |
| 120 | 0.677 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.160 |
| $\infty$ | 0.674 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 |

Tabel 9.2 voortgezet

| | P | | | | | | |
|---|---|---|---|---|---|---|---|
| $vg$ | 0.983 | 0.988 | 0.992 | 0.993 | 0.994 | 0.995 | 0.996 |
| 1 | 19.081 | 25.452 | 38.189 | 44.557 | 50.922 | 63.656 | 76.392 |
| 2 | 5.339 | 6.205 | 7.649 | 8.277 | 8.860 | 9.925 | 10.886 |
| 3 | 3.740 | 4.177 | 4.857 | 5.138 | 5.392 | 5.841 | 6.232 |
| 4 | 3.186 | 3.495 | 3.961 | 4.148 | 4.315 | 4.604 | 4.851 |
| 5 | 2.912 | 3.163 | 3.534 | 3.681 | 3.810 | 4.032 | 4.219 |
| 6 | 2.749 | 2.969 | 3.287 | 3.412 | 3.521 | 3.707 | 3.863 |
| 7 | 2.642 | 2.841 | 3.128 | 3.238 | 3.335 | 3.499 | 3.636 |
| 8 | 2.566 | 2.752 | 3.016 | 3.117 | 3.206 | 3.355 | 3.479 |
| 9 | 2.510 | 2.685 | 2.933 | 3.028 | 3.111 | 3.250 | 3.364 |
| 10 | 2.466 | 2.634 | 2.870 | 2.960 | 3.038 | 3.169 | 3.277 |
| 11 | 2.431 | 2.593 | 2.820 | 2.906 | 2.981 | 3.106 | 3.208 |
| 12 | 2.403 | 2.560 | 2.779 | 2.863 | 2.934 | 3.055 | 3.153 |
| 13 | 2.380 | 2.533 | 2.746 | 2.827 | 2.896 | 3.012 | 3.107 |
| 14 | 2.360 | 2.510 | 2.718 | 2.796 | 2.864 | 2.977 | 3.069 |
| 15 | 2.343 | 2.490 | 2.694 | 2.770 | 2.837 | 2.947 | 3.036 |
| 16 | 2.328 | 2.473 | 2.673 | 2.748 | 2.813 | 2.921 | 3.008 |
| 17 | 2.316 | 2.458 | 2.655 | 2.729 | 2.793 | 2.898 | 2.984 |
| 18 | 2.304 | 2.445 | 2.639 | 2.712 | 2.775 | 2.878 | 2.963 |
| 19 | 2.294 | 2.433 | 2.625 | 2.697 | 2.759 | 2.861 | 2.944 |
| 20 | 2.285 | 2.423 | 2.613 | 2.683 | 2.744 | 2.845 | 2.927 |
| 21 | 2.278 | 2.414 | 2.601 | 2.671 | 2.732 | 2.831 | 2.912 |
| 22 | 2.270 | 2.405 | 2.591 | 2.661 | 2.720 | 2.819 | 2.899 |
| 23 | 2.264 | 2.398 | 2.582 | 2.651 | 2.710 | 2.807 | 2.886 |
| 24 | 2.258 | 2.391 | 2.574 | 2.642 | 2.700 | 2.797 | 2.875 |
| 25 | 2.252 | 2.385 | 2.566 | 2.634 | 2.692 | 2.787 | 2.865 |
| 26 | 2.247 | 2.379 | 2.559 | 2.626 | 2.684 | 2.779 | 2.856 |
| 27 | 2.243 | 2.373 | 2.552 | 2.619 | 2.676 | 2.771 | 2.847 |
| 28 | 2.238 | 2.368 | 2.546 | 2.613 | 2.669 | 2.763 | 2.839 |
| 29 | 2.234 | 2.364 | 2.541 | 2.607 | 2.663 | 2.756 | 2.832 |
| 30 | 2.231 | 2.360 | 2.536 | 2.601 | 2.657 | 2.750 | 2.825 |
| 40 | 2.204 | 2.329 | 2.499 | 2.562 | 2.616 | 2.704 | 2.776 |
| 60 | 2.178 | 2.299 | 2.463 | 2.524 | 2.575 | 2.660 | 2.729 |
| 120 | 2.153 | 2.270 | 2.428 | 2.486 | 2.536 | 2.617 | 2.683 |
| $\infty$ | 2.128 | 2.241 | 2.394 | 2.450 | 2.498 | 2.576 | 2.638 |

Table 9.3: F-verdeling

| | | Vrijheidsgraden teller | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 20 | 40 | $\infty$ |
| 1 | .750 | 5.83 | 7.50 | 8.20 | 8.58 | 8.82 | 8.98 | 9.19 | 9.32 | 9.58 | 9.71 | 9.85 |
| | .900 | 39.9 | 49.5 | 53.6 | 55.8 | 57.2 | 58.2 | 59.4 | 60.2 | 61.7 | 62.5 | 63.3 |
| | .950 | 161 | 199 | 216 | 225 | 230 | 234 | 239 | 242 | 248 | 251 | 254 |
| 2 | .750 | 2.57 | 3.00 | 3.15 | 3.23 | 3.28 | 3.31 | 3.35 | 3.38 | 3.43 | 3.45 | 3.48 |
| | .900 | 8.53 | 9.00 | 9.16 | 9.24 | 9.29 | 9.33 | 9.37 | 9.39 | 9.44 | 9.47 | 9.49 |
| | .950 | 18.5 | 19.0 | 19.2 | 19.2 | 19.3 | 19.3 | 19.4 | 19.4 | 19.4 | 19.5 | 19.5 |
| | .990 | 98.5 | 99.0 | 99.2 | 99.3 | 99.3 | 99.3 | 99.4 | 99.4 | 99.4 | 99.5 | 99.5 |
| 3 | .750 | 2.02 | 2.28 | 2.36 | 2.39 | 2.41 | 2.42 | 2.44 | 2.44 | 2.46 | 2.47 | 2.47 |
| | .900 | 5.54 | 5.46 | 5.39 | 5.34 | 5.31 | 5.28 | 5.25 | 5.23 | 5.18 | 5.16 | 5.13 |
| | .950 | 10.1 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.85 | 8.79 | 8.66 | 8.59 | 8.53 |
| | .990 | 34.1 | 30.8 | 29.5 | 28.7 | 28.2 | 27.9 | 27.5 | 27.2 | 26.7 | 26.4 | 26.1 |
| | .999 | 167 | 148 | 141 | 137 | 135 | 133 | 131 | 129 | 126 | 125.0 | 123.5 |
| 4 | .750 | 1.81 | 2.00 | 2.05 | 2.06 | 2.07 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 | 2.08 |
| | .900 | 4.54 | 4.32 | 4.19 | 4.11 | 4.05 | 4.01 | 3.95 | 3.92 | 3.84 | 3.80 | 3.76 |
| | .950 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.04 | 5.96 | 5.80 | 5.72 | 5.63 |
| | .990 | 21.2 | 18.0 | 16.7 | 16.0 | 15.5 | 15.2 | 14.8 | 14.5 | 14.0 | 13.7 | 13.5 |
| | .999 | 74.1 | 61.2 | 56.2 | 53.4 | 51.7 | 50.5 | 49.0 | 48.1 | 46.1 | 45.1 | 44.0 |
| 5 | .750 | 1.69 | 1.85 | 1.88 | 1.89 | 1.89 | 1.89 | 1.89 | 1.89 | 1.88 | 1.88 | 1.87 |
| | .900 | 4.06 | 3.78 | 3.62 | 3.52 | 3.45 | 3.40 | 3.34 | 3.30 | 3.21 | 3.16 | 3.11 |
| | .950 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.82 | 4.74 | 4.56 | 4.46 | 4.37 |
| | .990 | 16.3 | 13.3 | 12.1 | 11.4 | 11.0 | 10.7 | 10.3 | 10.1 | 9.55 | 9.29 | 9.02 |
| | .999 | 47.2 | 37.1 | 33.2 | 31.1 | 29.8 | 28.8 | 27.6 | 26.9 | 25.4 | 24.6 | 23.8 |
| 6 | .750 | 1.62 | 1.76 | 1.78 | 1.79 | 1.79 | 1.78 | 1.78 | 1.77 | 1.76 | 1.75 | 1.74 |
| | .900 | 3.78 | 3.46 | 3.29 | 3.18 | 3.11 | 3.05 | 2.98 | 2.94 | 2.84 | 2.78 | 2.72 |
| | .950 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.15 | 4.06 | 3.87 | 3.77 | 3.67 |
| | .990 | 13.7 | 10.9 | 9.78 | 9.15 | 8.75 | 8.47 | 8.10 | 7.87 | 7.40 | 7.14 | 6.88 |
| | .999 | 35.5 | 27.0 | 23.7 | 21.9 | 20.8 | 20.0 | 19.0 | 18.4 | 17.1 | 16.4 | 15.7 |
| 7 | .750 | 1.57 | 1.70 | 1.72 | 1.72 | 1.71 | 1.71 | 1.70 | 1.69 | 1.67 | 1.66 | 1.65 |
| | .900 | 3.59 | 3.26 | 3.07 | 2.96 | 2.88 | 2.83 | 2.75 | 2.70 | 2.59 | 2.54 | 2.47 |
| | .950 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.73 | 3.64 | 3.44 | 3.34 | 3.23 |
| | .990 | 12.2 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.84 | 6.62 | 6.16 | 5.91 | 5.65 |
| | .999 | 29.2 | 21.7 | 18.8 | 17.2 | 16.2 | 15.5 | 14.6 | 14.1 | 12.9 | 12.3 | 11.7 |
| 8 | .750 | 1.54 | 1.66 | 1.67 | 1.66 | 1.66 | 1.65 | 1.64 | 1.63 | 1.61 | 1.59 | 1.58 |
| | .900 | 3.46 | 3.11 | 2.92 | 2.81 | 2.73 | 2.67 | 2.59 | 2.54 | 2.42 | 2.36 | 2.29 |
| | .950 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.44 | 3.35 | 3.15 | 3.04 | 2.93 |
| | .990 | 11.3 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.03 | 5.81 | 5.36 | 5.12 | 4.86 |
| | .999 | 25.4 | 18.5 | 15.8 | 14.4 | 13.5 | 12.9 | 12.0 | 11.5 | 10.5 | 9.92 | 9.33 |
| 9 | .750 | 1.51 | 1.62 | 1.63 | 1.63 | 1.62 | 1.61 | 1.60 | 1.59 | 1.56 | 1.54 | 1.53 |
| | .900 | 3.36 | 3.01 | 2.81 | 2.69 | 2.61 | 2.55 | 2.47 | 2.42 | 2.30 | 2.23 | 2.16 |
| | .950 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.23 | 3.14 | 2.94 | 2.83 | 2.71 |
| | .990 | 10.6 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.47 | 5.26 | 4.81 | 4.57 | 4.31 |
| | .999 | 22.9 | 16.4 | 13.9 | 12.6 | 11.7 | 11.1 | 10.4 | 9.89 | 8.90 | 8.37 | 7.81 |
| 10 | .750 | 1.49 | 1.60 | 1.60 | 1.59 | 1.59 | 1.58 | 1.56 | 1.55 | 1.52 | 1.51 | 1.48 |
| | .900 | 3.29 | 2.92 | 2.73 | 2.61 | 2.52 | 2.46 | 2.38 | 2.32 | 2.20 | 2.13 | 2.06 |
| | .950 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.07 | 2.98 | 2.77 | 2.66 | 2.54 |
| | .990 | 10.0 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.06 | 4.85 | 4.41 | 4.17 | 3.91 |
| | .999 | 21.0 | 14.9 | 12.6 | 11.3 | 10.5 | 9.93 | 9.20 | 8.75 | 7.80 | 7.30 | 6.76 |

Tabel 9.3 voortgezet

| | | Vrijheidsgraden teller | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 10 | 20 | 40 | ∞ |
| 12 | .750 | 1.46 | 1.56 | 1.56 | 1.55 | 1.54 | 1.53 | 1.51 | 1.50 | 1.47 | 1.45 | 1.42 |
| | .900 | 3.18 | 2.81 | 2.61 | 2.48 | 2.39 | 2.33 | 2.24 | 2.19 | 2.06 | 1.99 | 1.90 |
| | .950 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.85 | 2.75 | 2.54 | 2.43 | 2.30 |
| | .990 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.50 | 4.30 | 3.86 | 3.62 | 3.36 |
| | .999 | 18.6 | 13.0 | 10.8 | 9.63 | 8.89 | 8.38 | 7.71 | 7.29 | 6.40 | 5.93 | 5.42 |
| 14 | .750 | 1.44 | 1.53 | 1.53 | 1.52 | 1.51 | 1.50 | 1.48 | 1.46 | 1.43 | 1.41 | 1.38 |
| | .900 | 3.10 | 2.73 | 2.52 | 2.39 | 2.31 | 2.24 | 2.15 | 2.10 | 1.96 | 1.89 | 1.80 |
| | .950 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.70 | 2.60 | 2.39 | 2.27 | 2.13 |
| | .990 | 8.86 | 6.51 | 5.56 | 5.04 | 4.69 | 4.46 | 4.14 | 3.94 | 3.51 | 3.27 | 3.00 |
| | .999 | 17.1 | 11.8 | 9.73 | 8.62 | 7.92 | 7.44 | 6.80 | 6.40 | 5.56 | 5.10 | 4.60 |
| 16 | .750 | 1.42 | 1.51 | 1.51 | 1.50 | 1.48 | 1.47 | 1.45 | 1.44 | 1.40 | 1.37 | 1.34 |
| | .900 | 3.05 | 2.67 | 2.46 | 2.33 | 2.24 | 2.18 | 2.09 | 2.03 | 1.89 | 1.81 | 1.72 |
| | .950 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.59 | 2.49 | 2.28 | 2.15 | 2.01 |
| | .990 | 8.53 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 3.89 | 3.69 | 3.26 | 3.02 | 2.75 |
| | .999 | 16.1 | 11.0 | 9.01 | 7.94 | 7.27 | 6.80 | 6.20 | 5.81 | 4.99 | 4.54 | 4.06 |
| 18 | .750 | 1.41 | 1.50 | 1.49 | 1.48 | 1.46 | 1.45 | 1.43 | 1.42 | 1.38 | 1.35 | 1.32 |
| | .900 | 3.01 | 2.62 | 2.42 | 2.29 | 2.20 | 2.13 | 2.04 | 1.98 | 1.84 | 1.75 | 1.66 |
| | .950 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.51 | 2.41 | 2.19 | 2.06 | 1.92 |
| | .990 | 8.29 | 6.01 | 5.09 | 4.58 | 4.25 | 4.01 | 3.71 | 3.51 | 3.08 | 2.84 | 2.57 |
| | .999 | 15.4 | 10.4 | 8.49 | 7.46 | 6.81 | 6.35 | 5.76 | 5.39 | 4.59 | 4.15 | 3.67 |
| 20 | .750 | 1.40 | 1.49 | 1.48 | 1.47 | 1.45 | 1.44 | 1.42 | 1.40 | 1.36 | 1.33 | 1.29 |
| | .900 | 2.97 | 2.59 | 2.38 | 2.25 | 2.16 | 2.09 | 2.00 | 1.94 | 1.79 | 1.71 | 1.61 |
| | .950 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.45 | 2.35 | 2.12 | 1.99 | 1.84 |
| | .990 | 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.56 | 3.37 | 2.94 | 2.69 | 2.42 |
| | .999 | 14.8 | 10.0 | 8.10 | 7.10 | 6.46 | 6.02 | 5.44 | 5.08 | 4.29 | 3.86 | 3.38 |
| 30 | .750 | 1.38 | 1.45 | 1.44 | 1.42 | 1.41 | 1.39 | 1.37 | 1.35 | 1.30 | 1.27 | 1.23 |
| | .900 | 2.88 | 2.49 | 2.28 | 2.14 | 2.05 | 1.98 | 1.88 | 1.82 | 1.67 | 1.57 | 1.46 |
| | .950 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.27 | 2.16 | 1.93 | 1.79 | 1.62 |
| | .990 | 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.17 | 2.98 | 2.55 | 2.30 | 2.01 |
| | .999 | 13.3 | 8.77 | 7.05 | 6.12 | 5.53 | 5.12 | 4.58 | 4.24 | 3.49 | 3.07 | 2.59 |
| 40 | .750 | 1.36 | 1.44 | 1.42 | 1.40 | 1.39 | 1.37 | 1.35 | 1.33 | 1.28 | 1.24 | 1.19 |
| | .900 | 2.84 | 2.44 | 2.23 | 2.09 | 2.00 | 1.93 | 1.83 | 1.76 | 1.61 | 1.51 | 1.38 |
| | .950 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.18 | 2.08 | 1.84 | 1.69 | 1.51 |
| | .990 | 7.31 | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 2.99 | 2.80 | 2.37 | 2.11 | 1.80 |
| | .999 | 12.6 | 8.25 | 6.59 | 5.70 | 5.13 | 4.73 | 4.21 | 3.87 | 3.15 | 2.73 | 2.23 |
| 60 | .750 | 1.35 | 1.42 | 1.41 | 1.38 | 1.37 | 1.35 | 1.32 | 1.30 | 1.25 | 1.21 | 1.15 |
| | .900 | 2.79 | 2.39 | 2.18 | 2.04 | 1.95 | 1.87 | 1.77 | 1.71 | 1.54 | 1.44 | 1.29 |
| | .950 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.10 | 1.99 | 1.75 | 1.59 | 1.39 |
| | .990 | 7.08 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.82 | 2.63 | 2.20 | 1.94 | 1.60 |
| | .999 | 12.0 | 7.77 | 6.17 | 5.31 | 4.76 | 4.37 | 3.86 | 3.54 | 2.83 | 2.41 | 1.89 |
| 100 | .750 | 1.34 | 1.41 | 1.39 | 1.37 | 1.35 | 1.33 | 1.30 | 1.28 | 1.23 | 1.18 | 1.11 |
| | .900 | 2.76 | 2.36 | 2.14 | 2.00 | 1.91 | 1.83 | 1.73 | 1.66 | 1.49 | 1.38 | 1.21 |
| | .950 | 3.94 | 3.09 | 2.70 | 2.46 | 2.31 | 2.19 | 2.03 | 1.93 | 1.68 | 1.52 | 1.28 |
| | .990 | 6.90 | 4.82 | 3.98 | 3.51 | 3.21 | 2.99 | 2.69 | 2.50 | 2.07 | 1.80 | 1.43 |
| | .999 | 11.5 | 7.41 | 5.86 | 5.02 | 4.48 | 4.11 | 3.61 | 3.30 | 2.59 | 2.17 | 1.62 |
| ∞ | .750 | 1.32 | 1.39 | 1.37 | 1.35 | 1.33 | 1.31 | 1.28 | 1.26 | 1.19 | 1.14 | 1.01 |
| | .900 | 2.71 | 2.30 | 2.08 | 1.95 | 1.85 | 1.77 | 1.67 | 1.60 | 1.42 | 1.30 | 1.01 |
| | .950 | 3.84 | 3.00 | 2.61 | 2.37 | 2.21 | 2.10 | 1.94 | 1.83 | 1.57 | 1.40 | 1.01 |
| | .990 | 6.64 | 4.61 | 3.78 | 3.32 | 3.02 | 2.80 | 2.51 | 2.32 | 1.88 | 1.59 | 1.03 |
| | .999 | 10.8 | 6.91 | 5.43 | 4.62 | 4.11 | 3.75 | 3.27 | 2.96 | 2.27 | 1.84 | 1.02 |

Table 9.4: $H^*$ statistiek met 95 percentielen onder de nulhypothese

| | | | | | | $r$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $vg$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 2 | 39.0 | 87.5 | 142 | 202 | 266 | 333 | 403 | 475 | 550 | 626 | 704 |
| 3 | 15.4 | 27.8 | 39.2 | 50.7 | 62.0 | 72.9 | 83.5 | 93.9 | 104 | 114 | 124 |
| 4 | 9.6 | 15.5 | 20.6 | 25.2 | 29.5 | 33.6 | 37.5 | 41.1 | 44.6 | 48.0 | 51.4 |
| 5 | 7.15 | 10.8 | 13.7 | 16.3 | 18.7 | 20.8 | 22.9 | 24.7 | 26.5 | 28.2 | 29.9 |
| 6 | 5.82 | 8.38 | 10.4 | 12.1 | 13.7 | 15.0 | 16.3 | 17.5 | 18.6 | 19.7 | 20.7 |
| 7 | 4.99 | 6.94 | 8.44 | 9.70 | 10.8 | 11.8 | 12.7 | 13.5 | 14.3 | 15.1 | 15.8 |
| 8 | 4.43 | 6.00 | 7.18 | 8.12 | 9.03 | 9.78 | 10.5 | 11.1 | 11.7 | 12.2 | 12.7 |
| 9 | 4.03 | 5.34 | 6.31 | 7.11 | 7.80 | 8.41 | 8.95 | 9.45 | 9.91 | 10.3 | 10.7 |
| 10 | 3.72 | 4.85 | 5.67 | 6.34 | 6.92 | 7.42 | 7.87 | 8.28 | 8.66 | 9.01 | 9.34 |
| 12 | 3.28 | 4.16 | 4.79 | 5.30 | 5.72 | 6.09 | 6.42 | 6.72 | 7.00 | 7.25 | 7.48 |
| 15 | 2.86 | 3.54 | 4.01 | 4.37 | 4.68 | 4.95 | 5.19 | 5.40 | 5.59 | 5.77 | 5.93 |
| 20 | 2.46 | 2.95 | 3.29 | 3.54 | 3.76 | 3.94 | 4.10 | 4.24 | 4.37 | 4.49 | 4.59 |
| 30 | 2.07 | 2.40 | 2.61 | 2.78 | 2.91 | 3.02 | 3.12 | 3.21 | 3.29 | 3.36 | 3.39 |
| 60 | 1.67 | 1.85 | 1.96 | 2.04 | 2.11 | 2.17 | 2.22 | 2.26 | 2.30 | 2.33 | 2.36 |
| $\infty$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |