Designs with One Source of Variation

3

3.1 Introduction

In working through the checklist in Chap. 2, the experimenter must choose an experimental design at step (c). A design is the rule that determines the assignment of the experimental units to treatments. The simplest possible design is the *completely randomized design*, where the experimental units are assigned to the treatments completely at random, subject to the number of observations to be taken on each treatment. Completely randomized designs involve no blocking factors.

Two ways of calculating the required number of observations (sample sizes) on each treatment are presented in Sects. 3.6 and 4.5. The first method chooses sample sizes to obtain desired powers of hypothesis tests, and the second chooses sample sizes to achieve desired lengths of confidence intervals. We sometimes refer to the list of treatments and the corresponding sample sizes as the design, with the understanding that the assignment of experimental units to treatments is to be done completely at random.

In this chapter, we discuss the random assignment procedure for the completely randomized design, we introduce the method of least squares for estimating model parameters, and we develop a procedure for testing equality of the treatment parameters. Analyses by the SAS and R software are described at the end of the chapter.

3.2 Randomization

In this section we provide a procedure for randomization that is very easily applied using a computer, but can equally well be done by hand. On a computer, the procedure requires the availability of software that stores data in rows and columns (like spreadsheet software, a SAS data set, or a data.frame or matrix in R), that includes a function that randomly generates real numbers between zero and one, and that includes the capacity to sort rows by the values in one column.

We use r_i to denote the number of observations to be taken on the ith treatment, and $n = \Sigma r_i$ to denote the total number of observations (and hence the required number of experimental units). We code the treatments from 1 to v and label the experimental units 1 to n.

Step 1: Enter into one column r_1 1's, then r_2 2's, …, and finally r_v v's, giving a total of $n = \Sigma r_i$ entries. These represent the treatment labels.

Step 2: Enter into another column $n = \Sigma r_i$ random numbers, including enough digits to avoid ties. (The random numbers can be generated by a computer program or read from Table A.1).

A. Dean et al., *Design and Analysis of Experiments*,
Springer Texts in Statistics, DOI 10.1007/978-3-319-52250-0_3

Table 3.1 Randomization

Unsorted treatments	Unsorted random numbers	Sorted treatments	Sorted random numbers	Experimental unit
1	0.533	3	0.139	1
1	0.683	2	0.379	2
2	0.702	3	0.411	3
2	0.379	1	0.533	4
3	0.411	1	0.683	5
3	0.962	2	0.702	6
3	0.139	3	0.962	7

Step 3: Reorder both columns so that the random numbers are put in ascending order. This arranges the treatment labels into a random order.

Step 4: Assign experimental unit t to the treatment whose label is in row t.

If the number n of experimental units is a k-digit integer, then the list in step 2 should be a list of k-digit random numbers. To obtain k-digit random numbers from Table A.1, a random starting place is found as described in Sect. 1.1.4, p. 3. The digits are then read across the rows in groups of k (ignoring spaces).

We illustrate the randomization procedure using the SAS software in Sect. 3.8.1, p. 52, and using the R software in Sect. 3.9.1, p. 59. The procedure can equally well be done using the random digits in Table A.1 and sorting by hand.

Example 3.2.1 `Randomization`
Consider a completely randomized design for three treatments and sample sizes $r_1 = r_2 = 2, r_3 = 3$. The unrandomized design (step 1 of the randomization procedure) is 1 1 2 2 3 3 3, and is listed in column 1 of Table 3.1. Suppose step 2 generates the random numbers in column 2 of Table 3.1. In step 3, columns 1 and 2 are sorted so that the entries in column 2 are in ascending order. This gives columns 3 and 4. In step 4, the entries in column 3 are matched with experimental units 1–7 in order, so that column 3 contains the design after randomization. Treatment 1 is in rows 4 and 5, so experimental units 4 and 5 are assigned to treatment 1. Likewise, units 2 and 6 are assigned to treatment 2, and units 1, 3 and 7 are assigned to treatment 3. The randomly ordered treatments are then 3 2 3 1 1 2 3, and the experimental units 1–7 are assigned to the treatments in this order. ☐

3.3 Model for a Completely Randomized Design

A model is an equation that shows the dependence of the response variable upon the levels of the treatment factors. (Models involving block effects or covariates are considered in later chapters.)

Let Y_{it} be a random variable that represents the response obtained on the tth observation of the ith treatment. Let the parameter μ_i denote the "true response" of the ith treatment, that is, the response that would always be obtained from the ith treatment if it could be observed under *identical* experimental conditions and measured without error. Of course, this ideal situation can never happen—there is always some variability in the experimental procedure even if only caused by inaccuracies in reading measuring instruments. Sources of variation that are deemed to be minor and ignored during the planning of the experiment also contribute to variation in the response variable. These sources of nuisance variation are usually represented by a single variable ϵ_{it}, called an *error variable*, which is a random variable with zero mean. The model is then

$$Y_{it} = \mu_i + \epsilon_{it}, \quad t = 1, \ldots, r_i, \quad i = 1, \ldots, v,$$

where v is the number of treatments and r_i is the number of observations to be taken on the ith treatment. An alternative way of writing this model is to replace the parameter μ_i by $\mu + \tau_i$, so that the model becomes

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, \quad t = 1, \ldots, r_i, \quad i = 1, \ldots, v.$$

In this model, $\mu + \tau_i$ denotes the true mean response for the ith treatment, and examination of differences between the parameters μ_i in the first model is equivalent to examination of differences between the parameters τ_i in the second model.

It will be seen in Sect. 3.4 that unique estimates of the parameters in the second formulation of the model cannot be obtained. Nevertheless, many experimenters prefer this model. The parameter μ is a constant, and the parameter τ_i represents the positive or negative deviation of the response from this constant when the ith treatment is observed. This deviation is called the “effect” on the response of the ith treatment.

The above models are *linear models*, that is, the response variable is written as a linear function of the parameters. Any model that is not, or cannot, be transformed into a linear model cannot be treated by the methods in this book. Linear models often provide reasonably good approximations to more complicated models, and they are used extensively in practice.

The specific forms of the distributions of the random variables in a model need to be identified before any statistical analyses can be done. The error variables represent all the minor sources of variation taken together, including all the measurement errors. In many experiments, it is reasonable to assume that the error variables are independent and that they have a normal distribution with zero mean and unknown variance σ^2, which must be estimated. We call these assumptions the *error assumptions.* It will be shown in Chap. 5 that plots of the experimental data give good indications of whether or not the error assumptions are likely to be true. Proceeding with the analysis when the constant variance, normality, or independence assumptions are violated can result in a totally incorrect analysis.

A complete statement of the model for any experiment should include the list of error assumptions. Thus, for a completely randomized design with v specifically selected treatments (fixed effects), the model is

$$\begin{gathered} Y_{it} = \mu + \tau_i + \epsilon_{it}\,, \\ \epsilon_{it} \sim N(0, \sigma^2)\,, \\ \epsilon_{it}\text{'s are mutually independent,} \\ t = 1, \ldots, r_i, \quad i = 1, \ldots, v, \end{gathered} \tag{3.3.1}$$

where “$\sim N(0, \sigma^2)$” denotes “has a normal distribution with mean 0 and variance σ^2.” This is sometimes called a *one-way analysis of variance model*, since the model includes only one major source of variation, namely the treatment effect, and because the standard analysis of data using this model involves a comparison of measures of variation.

Notice that it is unnecessary to specify the distribution of Y_{it} in the model, as it is possible to deduce this from the stated information. Since Y_{it} is modeled as the sum of a treatment mean $\mu + \tau_i$ and a normally distributed random variable ϵ_{it}, it follows that

$$Y_{it} \sim N(\mu + \tau_i, \sigma^2).$$

Also, since the ϵ_{it}’s are mutually independent, the Y_{it}’s must also be mutually independent. Therefore, if the model is a true representation of the behavior of the response variable, then the data values y_{it} for the ith treatment form a random sample from a $N(\mu + \tau_i, \sigma^2)$ distribution.

3.4 Estimation of Parameters

3.4.1 Estimable Functions of Parameters

A function of the parameters of any linear model is said to be *estimable* if and only if it can be written as the expected value of a linear combination of the response variables. Only estimable functions of the parameters have unique linear unbiased estimates. Since it makes no sense to work with functions that have an infinite possible number of values, it is important that the analysis of the experiment involve only the estimable functions. For the one-way analysis of variance model (3.3.1), every estimable function is of the form

$$\begin{aligned} E\left[\sum_i \sum_t a_{it} Y_{it}\right] &= \sum_i \sum_t a_{it} E[Y_{it}] \\ &= \sum_i \sum_t a_{it}(\mu + \tau_i) \;=\; \sum_i b_i(\mu + \tau_i)\,, \end{aligned}$$

where $b_i = \Sigma_t a_{it}$ and the a_{it}'s are real numbers. Any function not of this form is *nonestimable*.

Clearly, $\mu + \tau_1$ is estimable, since it can be obtained by setting $b_1 = 1$ and $b_2 = b_3 = \cdots = b_v = 0$. Similarly, each $\mu + \tau_i$ is estimable. If we choose $b_i = c_i$ where $\sum c_i = 0$, we see that $\sum c_i \tau_i$ is estimable. Any such function $\sum c_i \tau_i$ for which $\sum_i c_i = 0$ is called a *contrast*, so all contrasts are estimable in the one-way analysis of variance model. For example, setting $b_1 = 1$, $b_2 = -1$, $b_3 = \cdots = b_v = 0$ shows that $\tau_1 - \tau_2$ is estimable. Similarly, each $\tau_i - \tau_s$, $i \neq s$, is estimable. Notice that there are no values of b_i that give μ, $\tau_1, \tau_2, \ldots$, or τ_v separately as the expected value. Therefore, these parameters are not individually estimable.

3.4.2 Notation

We write the ith treatment sample mean as

$$\overline{Y}_{i.} = \frac{1}{r_i}\left(\sum_{t=1}^{r_i} Y_{it}\right)$$

and the corresponding observed sample mean as $\overline{y}_{i.}$. The "dot" notation means "add over all values of the subscript replaced with a dot," and the "bar" means "divide by the number of terms that have been added up." This notation will be extremely useful throughout this book. For example, in the next subsection we write

$$\frac{1}{n}\sum_{i=1}^{v}\sum_{t=1}^{r_i} y_{it} = \frac{1}{n}\sum_{i=1}^{v} y_{i.} = \frac{1}{n} y_{..} = \overline{y}_{..}\,, \text{ where } n = \sum_{i=1}^{v} r_i = r_.\,,$$

so that $\overline{y}_{..}$ is the average of all of the observations. Note that if the summation applies to a subscript on two variables, the dot notation cannot be used. For example, $\sum r_i \hat{\tau}_i$ cannot be written as $r_. \hat{\tau}_.$, since $r_. \hat{\tau}_.$ denotes $(\sum r_i)(\sum \hat{\tau}_i)$. Also note that when notation involves both a sum and a square, such as $y_{..}^2$ or $\bar{y}_{i.}^2$, the sum is taken first and then the sum is squared.

3.4.3 Obtaining Least Squares Estimates

The *method of least squares* is used to obtain estimates and estimators for estimable functions of parameters in linear models. We shall show that the ith treatment sample mean $\overline{Y}_{i.}$ and its observed value $\overline{y}_{i.}$ are the "least squares estimator" and "least squares estimate," respectively, of $\mu + \tau_i$. Least squares solutions for the parameters $\mu, \tau_1, \ldots, \tau_v$ are any set of corresponding values $\hat{\mu}, \hat{\tau}_1, \ldots, \hat{\tau}_v$ that minimize the sum of squared errors

$$\sum_{i=1}^{v}\sum_{t=1}^{r_i} e_{it}^2 = \sum_{i=1}^{v}\sum_{t=1}^{r_i}(y_{it} - \mu - \tau_i)^2. \tag{3.4.2}$$

The estimated model $\hat{y}_{it} = \hat{\mu} + \hat{\tau}_i$ is the model that best fits the data in the sense of minimizing (3.4.2).

Finding least squares solutions is a standard problem in calculus.[1] The sum of squared errors (3.4.2) is differentiated with respect to each of the parameters $\mu, \tau_1, \ldots, \tau_v$ in turn. Then each of the $v + 1$ resulting derivatives is set equal to zero, yielding a set of $v + 1$ equations. These $v + 1$ equations are called the *normal equations*. Any solution to the normal equations gives a minimum value of the sum of squared errors (3.4.2) and provides a set of least squares solutions for the parameters.

The reader is asked to verify in Exercise 6 that the normal equations for the one-way analysis of variance model (3.3.1) are those shown in (3.4.3). The first equation in (3.4.3) is obtained by setting the derivative of the sum of squared errors of (3.4.2) with respect to μ equal to zero, and the other v equations are obtained by setting the derivatives with respect to each τ_i in turn equal to zero. We put "hats" on the parameters at this stage to denote solutions. The $v + 1$ normal equations are

$$\begin{aligned} y_{..} - n\hat{\mu} - \sum_i r_i\hat{\tau}_i &= 0, \\ y_{i.} - r_i\hat{\mu} - r_i\hat{\tau}_i &= 0, \quad i = 1, \ldots, v, \end{aligned} \tag{3.4.3}$$

and include $v + 1$ unknown parameters. From the last v equations, we obtain

$$\hat{\mu} + \hat{\tau}_i = \overline{y}_{i.}, \quad i = 1, \ldots, v,$$

so the least squares solution for the ith treatment mean $\mu + \tau_i$ is the corresponding sample mean $\overline{y}_{i.}$.

There is a problem in solving the normal equations to obtain least squares solutions for each parameter $\mu, \tau_1, \ldots, \tau_v$ individually. If the last v normal equations (3.4.3) are added together, the first equation results. This means that the $v+1$ equations are not distinct (not linearly independent). The last v normal equations *are* distinct, since they each contain a different τ_i. Thus, there are exactly v distinct normal equations in $v + 1$ unknown parameters, and there is no unique solution for the parameters. This is not surprising, in view of the fact that we have already seen in Sect. 3.4.1 that these parameters are not individually estimable. For practical purposes, any one of the infinite number of solutions will be satisfactory, since they lead to identical solutions for the estimable parameters. To obtain any one of these solutions, it is necessary to add a further equation to the set of normal equations. *Any* extra equation can be added, provided that it is not a linear combination of the equations already present. The trick is to add whichever equation will aid most in solving the entire set of equations.

---

[1] Readers without a background in calculus may note that the least squares solutions for the parameters, individually, are not unique and then may skip forward to Sect. 3.4.4.

One obvious possibility is to add the equation $\hat{\mu} = 0$, in which case the normal equations become

$$\begin{aligned} \hat{\mu} &= 0\,, \\ y_{..} - \sum_i r_i \hat{\tau}_i &= 0\,, \\ y_{i.} - r_i \hat{\tau}_i &= 0, \quad i = 1, \ldots, v. \end{aligned}$$

It is then a simple matter to solve the last v equations for the $\hat{\tau}_i$'s, yielding $\hat{\tau}_i = y_{i.}/r_i = \overline{y}_{i.}$. Thus, one solution to the normal equations is

$$\begin{aligned} \hat{\mu} &= 0\,, \\ \hat{\tau}_i &= \overline{y}_{i.}, \quad i = 1, \ldots, v. \end{aligned}$$

A more common solution is obtained by adding the extra equation $\sum_i r_i \hat{\tau}_i = 0$ to (3.4.3). In this case, the normal equations become

$$\begin{aligned} \sum_i r_i \hat{\tau}_i &= 0\,, \\ y_{..} - n\hat{\mu} &= 0\,, \\ y_{i.} - r_i \hat{\mu} - r_i \hat{\tau}_i &= 0, \quad i = 1, \ldots, v, \end{aligned}$$

from which we obtain the least squares solutions

$$\begin{aligned} \hat{\mu} &= \overline{y}_{..}\,, \\ \hat{\tau}_i &= \overline{y}_{i.} - \overline{y}_{..}, \quad i = 1, \ldots, v. \end{aligned}$$

Still another solution, used, for example, by the SAS software, is obtained by adding the equation $\hat{\tau}_v = 0$. Then the solutions to the normal equations are

$$\begin{aligned} \hat{\mu} &= \overline{y}_{v.}\,, \\ \hat{\tau}_i &= \overline{y}_{i.} - \overline{y}_{v.}, \quad i = 1, \ldots, v. \end{aligned}$$

The default solution for the R software is similar and obtained by adding the equation $\hat{\tau}_1 = 0$. In each of the sets of solutions just obtained, it is always true that

$$\hat{\mu} + \hat{\tau}_i = \overline{y}_{i.}.$$

No matter which extra equation is added to the normal equations, $\overline{y}_{i.}$ will *always* be the least squares solution for $\mu + \tau_i$. Thus, although it is not possible to obtain unique least squares solutions for μ and τ_i separately, the least squares solution for the estimable true treatment mean $\mu + \tau_i$ *is* unique. We call $\overline{y}_{i.}$ the *least squares estimate* and $\overline{Y}_{i.}$ the *least squares estimator* of $\mu + \tau_i$. The notation $\hat{\mu} + \hat{\tau}_i$ is used somewhat ambiguously to mean both the least squares estimator and estimate. It should be clear from the context which of these is meant.

3.4.4 Properties of Least Squares Estimators

An important property of a least squares estimator is that

> the least squares estimator of any estimable function of the parameters is the unique best linear unbiased estimator.

This statement, called the *Gauss–Markov Theorem*, is true for all linear models whose error variables are independent and have common variance σ^2. The theorem tells us that for the one-way analysis of variance model (3.3.1), the least squares estimator $\sum b_i \overline{Y}_{i.}$ of the estimable function $\sum b_i(\mu + \tau_i)$ is unique, is unbiased and has smallest variance. The theorem also tells us that τ_i cannot be estimable, since we have three different solutions for τ_i and none of the corresponding estimators has expected value equal to τ_i.

For the one-way analysis of variance model, Y_{it} has a normal distribution with mean $\mu + \tau_i$ and variance σ^2 (see Sect. 3.3), so $E[\overline{Y}_{i.}] = \mu + \tau_i$ and $\text{Var}(\overline{Y}_{i.}) = \sigma^2/r_i$. Therefore, the distribution of the least squares estimator $\overline{Y}_{i.}$ of $\mu + \tau_i$ is

$$\overline{Y}_{i.} \sim N(\mu + \tau_i\ ,\ \sigma^2/r_i)\,.$$

The $\overline{Y}_{i.}$'s are independent, since they are based on different Y_{it}'s. Consequently, the distribution of the least squares estimator $\sum c_i \overline{Y}_{i.}$ of the contrast $\sum c_i \tau_i$, with $\sum c_i = 0$, is

$$\sum c_i \overline{Y}_{i.} \sim N(\Sigma c_i \tau_i,\ \Sigma \frac{c_i^2}{r_i}\sigma^2).$$

Example 3.4.1 `Heart-lung pump experiment`

The following experiment was run by Richard Davis at The Ohio State University in 1987 to determine the effect of the number of revolutions per minute (rpm) of the rotary pump head of an Olson heart–lung pump on the fluid flow rate. The rpm was set directly on the tachometer of the pump console and PVC tubing of size 3/8" by 3/32" was used. The flow rate was measured in liters per minute. Five equally spaced levels of the treatment factor "rpm" were selected, namely, 50, 75, 100, 125, and 150 rpm, and these were coded as 1, 2, 3, 4, 5, respectively. The experimental design was a completely randomized design with $r_1 = r_3 = r_5 = 5$, $r_2 = 3$, and $r_4 = 2$. The data, in the order collected, are given in Table 3.2, and the summary information is

$$\begin{aligned}
y_{1.} &= 5.676,\ r_1 = 5,\ \overline{y}_{1.} = 1.1352,\\
y_{2.} &= 5.166,\ r_2 = 3,\ \overline{y}_{2.} = 1.7220,\\
y_{3.} &= 11.634,\ r_3 = 5,\ \overline{y}_{3.} = 2.3268,\\
y_{4.} &= 5.850,\ r_4 = 2,\ \overline{y}_{4.} = 2.9250,\\
y_{5.} &= 17.646,\ r_5 = 5,\ \overline{y}_{5.} = 3.5292.
\end{aligned}$$

The least squares estimate of the mean fluid flow rate when the pump is operating at 150 rpm is

$$(\hat{\mu} + \hat{\tau}_5) = \overline{y}_{5.} = 3.5292$$

liters per minute. The other mean fluid flow rates are estimated in a similar way. The experimenter expected the flow rate to increase as the rpm of the pump head was increased. Figure 3.1 supports this expectation.

Table 3.2 Fluid flow obtained from the rotary pump head of an Olson heart–lung pump

Observation	rpm	Level	Liters/minute
1	150	5	3.540
2	50	1	1.158
3	50	1	1.128
4	75	2	1.686
5	150	5	3.480
6	150	5	3.510
7	100	3	2.328
8	100	3	2.340
9	100	3	2.298
10	125	4	2.982
11	100	3	2.328
12	50	1	1.140
13	125	4	2.868
14	150	5	3.504
15	100	3	2.340
16	75	2	1.740
17	50	1	1.122
18	50	1	1.128
19	150	5	3.612
20	75	2	1.740

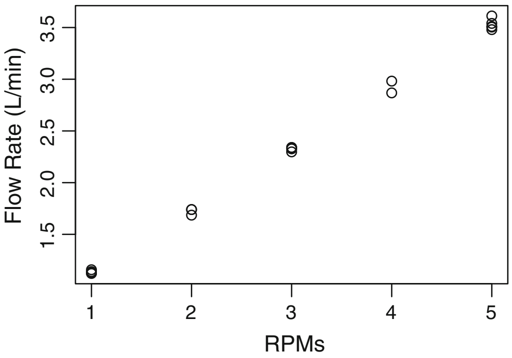


Fig. 3.1 Plot of data for the heart–lung pump experiment

Since the variance of the least squares estimator $\overline{Y}_{i.}$ of $\mu + \tau_i$ is σ^2/r_i, the first, third, and fifth treatment means are more precisely measured than the second and fourth.

The least squares estimate of the difference in fluid flow rate between 50 and 150 rpm is

$$(\hat{\tau}_5 - \hat{\tau}_1) = (\hat{\mu} + \hat{\tau}_5) - (\hat{\mu} + \hat{\tau}_1) = \overline{y}_{5.} - \overline{y}_{1.} = 2.394$$

liters per minute. The associated variance is

$$\sum_{i=1}^{5} \frac{c_i^2}{r_i}\sigma^2 = \left(\frac{1}{5} + \frac{1}{5}\right)\sigma^2 = 0.4\,\sigma^2.$$

3.4.5 Estimation of σ^2

The least squares estimates $\hat{\mu} + \hat{\tau}_i = \overline{y}_{i.}$ of $\mu + \tau_i$ $(i = 1, \ldots, v)$ minimize the sum of squared errors. Therefore, for the one-way analysis of variance model (3.3.1), the minimum possible value of the sum of squared errors (3.4.2), which we write as *ssE*, is equal to

$$ssE = \sum_i \sum_t \hat{e}_{it}^2 = \sum_i \sum_t (y_{it} - \hat{\mu} - \hat{\tau}_i)^2 .$$

Here, $\hat{e}_{it} = (y_{it} - \hat{\mu} - \hat{\tau}_i)$ is the deviation of the tth observation on the ith treatment from the estimated ith treatment mean. This is called the (it)th *residual*. Substituting the least squares estimates $\hat{\mu} + \hat{\tau}_i = \overline{y}_{i.}$ into the formula for *ssE*, we have

$$ssE = \sum_i \sum_t (y_{it} - \overline{y}_{i.})^2 . \tag{3.4.4}$$

The minimum sum of squared errors, *ssE*, is called the *sum of squares for error* or the *error sum of squares*, and is used below to find an unbiased estimate of the error variance σ^2. A useful computational formula for *ssE* is obtained by multiplying out the quantity in parentheses in (3.4.4); that is,

$$ssE = \sum_i \sum_t y_{it}^2 - \sum_i r_i \overline{y}_{i.}^2 . \tag{3.4.5}$$

Now, the random variable *SSE* corresponding to the minimum sum of squared errors *ssE* in (3.4.4) is

$$SSE = \sum_i \sum_t (Y_{it} - \overline{Y}_{i.})^2 = \sum_i (r_i - 1) S_i^2 , \tag{3.4.6}$$

where $S_i^2 = \sum_{t=1}^{r_i} (Y_{it} - \overline{Y}_{i.})^2 / (r_i - 1))$ is the sample variance for the ith treatment. In Exercise 3.11, the reader is asked to verify that S_i^2 is an unbiased estimator of the error variance σ^2. Then, the expected value of *SSE* is

$$E(SSE) = \sum_i (r_i - 1) E(S_i^2) = (n - v)\sigma^2 ,$$

giving an unbiased estimator of σ^2 as

$$\hat{\sigma}^2 = SSE/(n - v) = MSE. \tag{3.4.7}$$

The corresponding unbiased estimate of σ^2 is the observed value of *MSE*, namely $msE = ssE/(n - v)$. Both *MSE* and *msE* are called the *mean square for error* or *error mean square*. The estimate *msE* is sometimes called the "within groups (or within treatments) variation."

3.4.6 Confidence Bound for σ^2

If an experiment were to be repeated in the future, the estimated value of σ^2 obtained from the current experiment could be used at step (h) of the checklist to help calculate the number of observations that should be taken in the new experiment (see Sects. 3.6.2 and 4.5). However, the error variance in the new experiment is unlikely to be exactly the same as that in the current experiment, and in order not to underestimate the number of observations needed, it is advisable to use a larger value of σ^2 in the sample size calculation. One possibility is to use the upper limit of a one-sided confidence interval for σ^2.

It can be shown that the distribution of SSE/σ^2 is chi-squared with $n - v$ degrees of freedom, denoted by χ^2_{n-v}. Consequently,

$$P\left(\frac{SSE}{\sigma^2} \geq \chi^2_{n-v,1-\alpha}\right) = 1 - \alpha\,, \tag{3.4.8}$$

where $\chi^2_{n-v,1-\alpha}$ is the percentile of the chi-squared distribution with $n - v$ degrees of freedom and with probability of $1 - \alpha$ in the right-hand tail.

Manipulating the inequalities in (3.4.8), and replacing SSE by its observed value ssE, gives a one-sided $100(1 - \alpha)\%$ confidence bound for σ^2 as

$$\sigma^2 \leq \frac{ssE}{\chi^2_{n-v,1-\alpha}}\,. \tag{3.4.9}$$

This upper bound is called a $100(1 - \alpha)\%$ *upper confidence limit* for σ^2.

Example 3.4.2 `Battery experiment, continued`
The data of the battery experiment (Sect. 2.5.2, p. 24) are summarized in Table 3.3. The sum of squares for error is obtained from (3.4.5); that is,

$$\begin{aligned} ssE &= \sum_i \sum_t y_{it}^2 - \sum_i r_i \overline{y}_{i.}^2 \\ &= 6{,}028{,}288 - 4(570.75^2 + 860.50^2 + 433.00^2 + 496.25^2) \\ &= 28{,}412.5. \end{aligned}$$

An unbiased estimate of the error variance is then obtained as

$$msE = ssE/(n - v) = 28{,}412.5/(16 - 4) = 2367.71\,.$$

A 95% upper confidence limit for σ^2 is given by

$$\sigma^2 \leq \frac{ssE}{\chi^2_{12,0.95}} = \frac{28{,}412.5}{5.23} = 5432.60\,,$$

and taking the square root of the confidence limit, a 95% upper confidence limit for σ is 73.71 minutes per dollar. If the experiment were to be repeated in the future, the calculation for the number of observations at step (h) of the checklist might take the largest likely value for σ to be around 70–75 minutes per dollar. ☐

Table 3.3 Data for the battery experiment

Battery type	Life per unit cost (minutes per dollar)				$\overline{y}_{i.}$
1	611	537	542	593	570.75
2	923	794	827	898	860.50
3	445	490	384	413	433.00
4	476	569	480	460	496.25

3.5 One-Way Analysis of Variance

3.5.1 Testing Equality of Treatment Effects

In an experiment involving v treatments, an obvious question is whether or not the treatments differ at all in terms of their effects on the response variable. Thus one may wish to test the null hypothesis

$$H_0 : \{\tau_1 = \tau_2 = \cdots = \tau_v\}$$

that the treatment effects are all equal against the alternative hypothesis

$$H_A : \{\text{at least two of the } \tau_i\text{'s differ}\}.$$

At first glance, the null hypothesis appears to involve nonestimable parameters. However, we can easily rewrite it in terms of $v - 1$ estimable contrasts, as follows:

$$H_0 : \{\tau_1 - \tau_2 = 0 \;\text{ and }\; \tau_1 - \tau_3 = 0 \;\text{ and }\; \cdots \;\text{ and }\; \tau_1 - \tau_v = 0\}.$$

This is not the only way to rewrite H_0 in terms of estimable contrasts. For example, we could use the contrasts $\tau_i - \overline{\tau}_.$ (where $\overline{\tau}_. = \sum \tau_i / v$) and write the null hypothesis as follows:

$$H_0 : \{\tau_1 - \overline{\tau}_. = 0 \text{ and } \tau_2 - \overline{\tau}_. = 0 \text{ and } \;\cdots\; \text{ and } \;\tau_v - \overline{\tau}_. = 0\}\,.$$

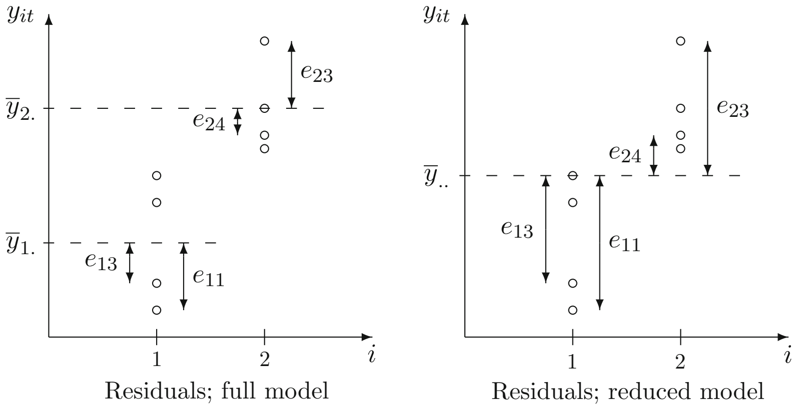
Now $\overline{\tau}_.$ is the average of the τ_i's, so the $\tau_i - \overline{\tau}_.$'s add to zero. Consequently, if $\tau_i - \overline{\tau}_. = 0$ for $i = 1, \ldots, v - 1$, then $\tau_v - \overline{\tau}_.$ must also be zero. Thus, this form of the null hypothesis could be written in terms of just the first $v - 1$ estimable functions $\tau_1 - \overline{\tau}_., \ldots, \tau_{v-1} - \overline{\tau}_.$.

Any way that we rewrite H_0 in terms of estimable functions of the parameters, it will always depend on $v - 1$ distinct contrasts. The number $v - 1$ is called the *treatment degrees of freedom*.

The basic idea behind an analysis of variance test is that the sum of squares for error measures how well the model fits the data. Consequently, a way of testing H_0 is to compare the sum of squares for error under the original one-way analysis of variance model (3.3.1), known as the *full model*, with that obtained from the modified model, which assumes that the null hypothesis is true. This modified model is called the *reduced model*.

Under H_0, the τ_i's are equal, and we can write the common value of $\tau_1, \ldots, \tau_v$ as τ. If we incorporate this into the one-way analysis of variance model, we obtain the reduced model

Fig. 3.2 Residuals under the full and reduced models when H_0 is false



$$\begin{aligned}Y_{it} &= \mu + \tau + \epsilon^0_{it}\,,\\ \epsilon^0_{it} &\sim N(0,\sigma^2)\,,\\ \epsilon^0_{it}{}'&\text{s are mutually independent}\,,\\ t &= 1,\ldots,r_i,\quad i = 1,\ldots,v,\end{aligned}$$

where we write ϵ^0_{it} for the (it)th error variable in the reduced model. To calculate the sum of squares for error, ssE_0, we need to determine the value of $\mu + \tau$ that minimizes the sum of squared errors

$$\sum_i \sum_t (y_{it} - \mu - \tau)^2\,.$$

Using calculus, the reader is asked to show in Exercise 7 that the unique least squares estimate of $\mu+\tau$ is the sample mean of all the observations; that is, $\hat{\mu} + \hat{\tau} = \overline{y}_{..}$. Therefore, the error sum of squares for the reduced model is

$$\begin{aligned}ssE_0 &= \sum_i \sum_t (y_{it} - \overline{y}_{..})^2\\ &= \sum_i \sum_t y_{it}^2 - n\overline{y}_{..}^2\,.\end{aligned} \tag{3.5.10}$$

If the null hypothesis $H_0 : \{\tau_1 = \tau_i = \cdots = \tau_v\}$ is false, and the treatment effects differ, the sum of squares for error ssE under the full model (3.3.1) is considerably smaller than the sum of squares for error ssE_0 for the reduced model. This is depicted in Fig. 3.2. On the other hand, if the null hypothesis is true, then ssE_0 and ssE will be very similar. The analysis of variance test is based on the difference $ssE_0 - ssE$, relative to the size of ssE; that is, the test is based on $(ssE_0 - ssE)/ssE$. We would want to reject H_0 if this quantity is large.

We call $ssT = ssE_0 - ssE$ the *sum of squares for treatments* or the *treatment sum of squares*, since its value depends on the differences between the treatment effects. Using formulas (3.5.10) and (3.4.5) for ssE_0 and ssE, the treatment sum of squares is

$$\begin{aligned} ssT &= ssE_0 - ssE \qquad (3.5.11)\\ &= \left(\sum_i \sum_t y_{it}^2 - n\overline{y}_{..}^2\right) - \left(\sum_i \sum_t y_{it}^2 - \sum_i r_i \overline{y}_{i.}^2\right)\\ &= \sum_i r_i \overline{y}_{i.}^2 - n\overline{y}_{..}^2 . \qquad (3.5.12) \end{aligned}$$

An equivalent formulation is

$$ssT = \sum_i r_i (\overline{y}_{i.} - \overline{y}_{..})^2 . \qquad (3.5.13)$$

The reader is invited to multiply out the parentheses in (3.5.13) and verify that (3.5.12) is obtained. There is a shortcut method of expanding (3.5.13) to obtain (3.5.12). First write down each term in y and square it. Then associate with each squared term the signs in (3.5.13). Finally, precede each term with the summations and constant outside the parentheses in (3.5.13). This quick expansion will work for all terms like (3.5.13) in this book. Formula (3.5.13) is probably the easier form of ssT to remember, while (3.5.12) is easier to manipulate for theoretical work and use for computations.

Since we will reject H_0 if ssT/ssE is large, we need to know what “large” means. This in turn means that we need to know the distribution of the corresponding random variable SST/SSE when H_0 is true, where

$$SST = \sum_i r_i (\overline{Y}_{i.} - \overline{Y}_{..})^2 \quad \text{and} \quad SSE = \sum_i \sum_t (Y_{it} - \overline{Y}_{i.})^2 . \qquad (3.5.14)$$

Now, as mentioned in Sect. 3.4.6, it can be shown that SSE/σ^2 has a chi-squared distribution with $n-v$ degrees of freedom, denoted by χ^2_{n-v}. Similarly, it can be shown that when H_0 is true, SST/σ^2 has a χ^2_{v-1} distribution, and that SST and SSE are independent. The ratio of two independent chi-squared random variables, each divided by their degrees of freedom, has an F distribution. Therefore, if H_0 is true, we have

$$\frac{SST/\sigma^2(v-1)}{SSE/\sigma^2(n-v)} \sim F_{v-1,n-v} .$$

We now know the distribution of SST/SSE multiplied by the constant $(n-v)/(v-1)$, and we want to reject the null hypothesis $H_0 : \{\tau_1 = \cdots = \tau_v\}$ in favor of the alternative hypothesis H_A : {at least two of the treatment effects differ} if this ratio is large. Thus, if we write $msT = ssT/(v-1)$, $msE = ssE/(n-v)$, where ssT and ssE are the observed values of the treatment sum of squares and error sum of squares, respectively, our decision rule is to

$$\text{reject } H_0 \quad \text{if} \quad \frac{msT}{msE} > F_{v-1,n-v,\alpha} , \qquad (3.5.15)$$

where $F_{v-1,n-v,\alpha}$ is the critical value from the F distribution with $v-1$ and $n-v$ degrees of freedom with α in the right-hand tail. The probability α is often called the *significance level* of the test and is the probability of rejecting H_0 when in fact it is true (a Type I error). Thus, α should be selected to be small if it is important not to make a Type I error ($\alpha = 0.01$ and 0.001 are typical choices); otherwise, α can be chosen to be a little larger ($\alpha = 0.10$ and 0.05 are typical choices). Critical values $F_{v-1,n-v,\alpha}$ for the F distribution are given in Table A.6. Due to lack of space, only a few typical values of α have been tabulated.

Table 3.4 One-way analysis of variance table

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio	Expected mean square
Treatments	v-1	ssT	$\frac{ssT}{v-1}$	$\frac{msT}{msE}$	$\sigma^2 + Q(\tau_i)$
Error	n-v	ssE	$\frac{ssE}{n-v}$		σ^2
Total	n-1	$sstot$			
Computational formulae					
$ssT = \sum_i r_i \overline{y}_{i.}^2 - n\overline{y}_{..}^2$			$ssE = \sum_i \sum_t y_{it}^2 - \sum_i r_i \overline{y}_{i.}^2$		
$sstot = \sum_i \sum_t y_{it}^2 - n\overline{y}_{..}^2$					
$Q(\tau_i) = \sum_i r_i(\tau_i - \sum_h r_h \tau_h / n)^2 / (v-1)$					

The calculations involved in the test of the hypothesis H_0 against H_A are usually written as an *analysis of variance table* as shown in Table 3.4. The last line shows the *total sum of squares* and *total degrees of freedom*. The total sum of squares, *sstot*, is $(n-1)$ times the sample variance of all of the data values. Thus,

$$sstot = \sum_i \sum_t (y_{it} - \overline{y}_{..})^2 = \sum_i \sum_t y_{it}^2 - n\overline{y}_{..}^2. \tag{3.5.16}$$

From (3.5.10), we see that *sstot* happens to be equal to ssE_0 for the one-way analysis of variance model, and from (3.5.11) we see that

$$sstot = ssT + ssE.$$

Thus, the total sum of squares consists of a part *ssT* that is explained by differences between the treatment effects and a part *ssE* that is not explained by any of the parameters in the model.

Example 3.5.1 `Battery experiment, continued`
Consider the battery experiment introduced in Sect. 2.5.2, p. 24. The sum of squares for error was calculated in Example 3.4.2, p. 40, to be $ssE = 28{,}412.5$. The life per unit cost responses and treatment averages are given in Table 3.3, p. 41. From these, we have $\Sigma\Sigma y_{it}^2 = 6{,}028{,}288$, $\overline{y}_{..} = 590.125$, and $r_i = 4$. Hence, the sums of squares *ssT* (3.5.12) and *sstot* (3.5.16) are

$$\begin{aligned} ssT &= \sum r_i \overline{y}_{i.}^2 - n\overline{y}_{..}^2 \\ &= 4(570.75^2 + 860.50^2 + 433.00^2 + 496.25^2) - 16(590.125)^2 \\ &= 427{,}915.25, \\ sstot = ssE_0 &= \sum\sum y_{it}^2 - n\overline{y}_{..}^2 \\ &= 6{,}028{,}288 - 16(590.125)^2 = 456{,}327.75, \end{aligned}$$

and we can verify that $sstot = ssT + ssE$.

The decision rule for testing the null hypothesis $H_0 : \{\tau_1 = \tau_2 = \tau_3 = \tau_4\}$ that the four battery types have the same average life per unit cost against the alternative hypothesis that at least two of the battery types differ, at significance level α, is

$$\text{reject } H_0 \text{ if } msT/msE = 60.24 > F_{3,12,\alpha}.$$

Table 3.5 One-way analysis of variance table for the battery experiment

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio	p-value
Type	3	427,915.25	142,638.42	60.24	0.0001
Error	12	28,412.50	2,367.71		
Total	15	456,327.75			

From Table A.6, it can be seen that $60.24 > F_{3,12,\alpha}$ for any of the tabulated values of α. For example, if α is chosen to be 0.01, then $F_{3,12,0.01} = 5.95$. Thus, for any tabulated choice of α, the null hypothesis is rejected, and it is concluded that at least two of the battery types differ in mean life per unit cost. In order to investigate which particular pairs of battery types differ, we would need to calculate confidence intervals. This will be done in Chap. 4. ☐

3.5.2 Use of p-Values

The *p-value* of a test is the smallest choice of α that would allow the null hypothesis to be rejected. For convenience, computer packages usually print the p-value as well as the ratio *msT/msE*. Having information about the p-value saves looking up $F_{v-1,n-v,\alpha}$ in Table A.6. All we need to do is to compare the p-value with our selected value of α. Therefore, the decision rule for testing $H_0 : \{\tau_1 = \cdots \tau_v\}$ against $H_A : \{\text{not all of } \tau_i\text{'s are equal}\}$ can be written as

$$\text{reject } H_0 \text{ if } p < \alpha.$$

Example 3.5.2 `Battery experiment, continued`
In the battery experiment of Example 3.5.1, the null hypothesis $H_0 : \{\tau_1 = \tau_2 = \tau_3 = \tau_4\}$ that the four battery types have the same average life per unit cost was tested against the alternative hypothesis that they do not. The p-value generated by SAS software for the test is shown in Table 3.5 as $p = 0.0001$. A value of 0.0001 in the SAS computer output indicates that the p-value is less than or equal to 0.0001. Smaller values are not printed explicitly. If α were chosen to be 0.01, then the null hypothesis would be rejected, since $p < \alpha$. ☐

3.6 Sample Sizes

Before an experiment can be run, it is necessary to determine the number of observations that should be taken on each treatment. This forms step (h) of the checklist in Sect. 2.2. In order to make this determination, the experimenter must first ascertain the approximate cost, in both time and money, of taking each observation and whether the cost differs for different levels of the treatment factor(s). There will probably be a fixed budget for the entire experiment. Therefore, remembering to set aside sufficient resources for the analysis of the experimental data, a rough calculation can be made of the maximum number, N, of observations that can be afforded. After having worked through steps (a)–(g) of the checklist, the experimenter will have identified the objectives of the experiment and the type of analysis required. It must now be ascertained whether or not the objectives of the experiment can be achieved within the budget. The calculations at step (h) may show that it is unnecessary to take as many as N observations, in which case valuable resources can be saved. Alternatively, and unfortunately the more likely, it may be found that more than N observations are needed in order to fulfill all the experimenter's requirements of the experiment. In this case, the experimenter needs to go back and review the decisions made so far in order to try to relax some of the requirements. Otherwise, an increase in budget needs to be obtained. There is little point in running the experiment with smaller sample sizes than those required without finding out what effect this will have on the analysis. The following quotation from J. N. R. Jeffers in his article "Acid rain and tree roots: an analysis" in *The Statistical Consultant in Action* (1987) is worth careful consideration:

> There is a quite strongly held view among experimenters that statisticians always ask for more replication than can be provided, and hence jeopardize the research by suggesting that it is not worth doing unless sufficient replication can be provided. There is, of course, some truth in this allegation, and equally, some truth in the view that, unless an experiment can be done with adequate replication, and with due regard to the size of the difference which it is important to be able to detect, the research may indeed not be worth doing.

We will consider two methods of determining the number of observations on each treatment (the sample sizes). One method, which involves specifying the desired length of confidence intervals, will be presented in Sect. 4.5. The other method, which involves specifying the power required of the analysis of variance, is the topic of this section. Since the method uses the expected value of the mean square for treatments, we calculate this first.

3.6.1 Expected Mean Squares for Treatments

The formula for *SST*, the treatment sum of squares, was given in (3.5.14) on p. 43. Its expected value is

$$\begin{aligned} E[SST] &= E[\sum r_i(\overline{Y}_{i.} - \overline{Y}_{..})^2] \\ &= E[\Sigma r_i \overline{Y}_{i.}^2 - n\overline{Y}_{..}^2] \\ &= \sum r_i E[\overline{Y}_{i.}^2] - nE[\overline{Y}_{..}^2]\,. \end{aligned}$$

From the definition of the variance of a random variable, we know that $\text{Var}(X) = E[X^2] - (E[X])^2$, so we can write $E[SST]$ as

$$E[SST] = \Sigma r_i[\text{Var}(\overline{Y}_{i.}) + (E[\overline{Y}_{i.}])^2] - n[\text{Var}(\overline{Y}_{..}) + (E[\overline{Y}_{..}])^2]\,.$$

For the one-way analysis of variance model (3.3.1), the response variables Y_{it} are independent, and each has a normal distribution with mean $\mu + \tau_i$ and variance σ^2. So,

$$\begin{aligned} E[SST] &= \sum r_i\left(\sigma^2/r_i + (\mu+\tau_i)^2\right) \\ &\quad -n\left(\sigma^2/n + \left(\mu + \sum r_i\tau_i/n\right)^2\right) \\ &= v\sigma^2 + n\mu^2 + 2\mu\sum r_i\tau_i + \sum r_i\tau_i^2 \\ &\quad -\sigma^2 - n\mu^2 - 2\mu\sum r_i\tau_i - (\sum r_i\tau_i)^2/n \\ &= (v-1)[\sigma^2 + Q(\tau_i)]\,, \end{aligned}$$

where

$$Q(\tau_i) = \Sigma_i r_i \left(\tau_i - \Sigma_h r_h \tau_h / n\right)^2 / (v-1)\,, \qquad (3.6.17)$$

which reduces to $Q(\tau_i) = r\sum_i(\tau_i - \overline{\tau}_.)^2/(v-1)$ when $r_1 = r_2 = \ldots r_v = r$. The expected value of the mean square for treatments $MST = SST/(v-1)$ is

$$E[MST] = \sigma^2 + Q(\tau_i)\,,$$

which is the quantity we listed in the analysis of variance table, Table 3.4. We note that when the treatment effects are all equal, $Q(\tau_i) = 0$, and $E[MST] = \sigma^2$.

3.6.2 Sample Sizes Using Power of a Test

Suppose that one of the major objectives of an experiment is to examine whether or not the treatments all have a similar effect on the response. The null hypothesis is actually somewhat unrealistic. The effects of the treatments are almost certainly not *exactly* equal, and even if they were, the nuisance variability in the experimental data would mask this fact. In any case, if the different levels produce only a very small difference in the response variable, the experimenter may not be interested in discovering this fact. For example, a difference of 5 minutes in life per dollar in two different batteries would probably not be noticed by most users. However, a larger difference such as 60 minutes may well be noticed. Thus the experimenter might require H_0 to be rejected with high probability if $\tau_i - \tau_s > 60$ minutes per dollar for some $i \neq s$ but may not be concerned about rejecting the null hypothesis if $\tau_i - \tau_s \leq 5$ minutes per dollar for all $i \neq s$. In most experiments, there is some value Δ such that if the difference in the effects of any two of the treatments exceeds Δ, the experimenter would like to reject the null hypothesis in favor of the alternative hypothesis with high probability.

The *power* of the test at Δ, denoted by $\pi(\Delta)$, is the probability of rejecting H_0 when the effects of at least two of the treatments differ by Δ. The power of the test $\pi(\Delta)$ is a function of Δ and also of the sample sizes, the number of treatments, the significance level α, and the error variance σ^2. Consequently, the sample sizes can be determined if $\pi(\Delta)$, v, α, and σ^2 are known. The values of Δ, $\pi(\Delta)$, v, and α are chosen by the experimenter, but the error variance has to be guessed using data from a pilot study or another similar experiment. In general, the largest likely value of σ^2 should be used. If the guess for σ^2 is too small, then the power of the test will be lower than the specified $\pi(\Delta)$. If the guess for σ^2 is too high, then the power will be higher than needed, and differences in the τ_i's smaller than Δ will cause H_0 to be rejected with high probability.

The rule for testing the null hypothesis $H_0 : \{\tau_1 = \cdots = \tau_v\}$ against H_A: {at least two of the τ_i's differ}, given in (3.5.15), on p. 43, is

$$\text{reject } H_0 \text{ if } \frac{msT}{msE} > F_{v-1,n-v,\alpha}\,.$$

As stated in Sect. 3.5.1, the test statistic *MST/MSE* has an F distribution if the null hypothesis is correct. But if the null hypothesis is false, then *MST/MSE* has a related distribution called a noncentral F distribution. The noncentral F distribution is denoted by $F_{v-1,n-v,\delta^2}$, where δ^2 is called the *noncentrality parameter* and is defined to be

$$\delta^2 = (v-1)Q(\tau_i)/\sigma^2\,, \qquad (3.6.18)$$

where $Q(\tau_i)$ was calculated in (3.6.17). When $Q(\tau_i) = 0$, then $\delta^2 = 0$, and the distribution of *MST/MSE* becomes the usual F-distribution. Otherwise, δ^2 is greater than zero, and the mean and spread of the distribution of *MST/MSE* are larger than those of the usual F-distribution. For equal sample sizes $r_1 = r_2 = \cdots = r_v = r$, we see that δ^2 is

$$\delta^2 = r \sum_i (\tau_i - \overline{\tau}_.)^2/\sigma^2 .$$

The calculation of the sample size r required to achieve a power $\pi(\Delta)$ at Δ for given v, α, and σ^2 rests on the fact that the hardest situation to detect is that in which the effects of two of the factor levels (say, the first and last) differ by Δ, and the others are all equal and midway between; that is,

$$\mu + \tau_2 = \mu + \tau_3 = \cdots = \mu + \tau_{v-1} = c ,$$

$$\mu + \tau_1 = c + \Delta/2 , \quad \text{and} \quad \mu + \tau_v = c - \Delta/2 ,$$

for some constant c. In this case,

$$\delta^2 = r \sum_i \frac{(\tau_i - \overline{\tau}_.)^2}{\sigma^2} = \frac{r\Delta^2}{2\sigma^2} . \tag{3.6.19}$$

The power of the test depends on the sample size r through the distribution of *MST/MSE*, which depends on δ^2. Since the power of the test is the probability of rejecting H_0, we have

$$\pi(\Delta) = P\left(\frac{MST}{MSE} > F_{v-1,n-v,\alpha}\right) .$$

The noncentral F distribution is tabulated in Table A.7, with power π given as a function of $\phi = \delta/\sqrt{v}$ for various values of $\nu_1 = v - 1$, $\nu_2 = n - v$, and α. Using (3.6.19),

$$\phi^2 = \frac{\delta^2}{v} = \frac{r\Delta^2}{2v\sigma^2} ,$$

so

$$r = \frac{2v\sigma^2\phi^2}{\Delta^2} . \tag{3.6.20}$$

Hence, given α, Δ, v, and σ^2, the value of r can be determined from Table A.7 to achieve a specified power $\pi(\Delta)$. The determination has to be done iteratively, since the denominator degrees of freedom, $\nu_2 = n - v = v(r - 1)$, depend on the unknown r. The procedure is as follows:

(a) Find the section of Table A.7 for the numerator degrees of freedom $\nu_1 = v - 1$ and the specified α (only $\alpha = 0.05$ is shown).
(b) Calculate the denominator degrees of freedom using $\nu_2 = 1000$ in the first iteration and $\nu_2 = n - v = v(r - 1)$ in the following iterations, and locate the appropriate row of the table, taking the smaller listed value of ν_2 if necessary.
(c) For the required power $\pi(\Delta)$, use interpolation to determine the corresponding value of ϕ, or take the larger listed value if necessary.

(d) Calculate $r = 2v\sigma^2\phi^2/\Delta^2$, rounding up to the nearest integer. (The first iteration gives a lower bound for r.)
(e) Repeat steps (b)–(d) until the value of r is unchanged or alternates between two values. Select the larger of alternating values.

Example 3.6.1 `Soap experiment, continued`
The first part of the checklist for the soap experiment is given in Sect. 2.5.1, p. 20, and is continued in Sect. 3.7, below. At step (h), the experimenter calculated the number of observations needed on each type of soap as follows.

The error variance was estimated to be about 0.007 grams squared from the pilot experiment. In testing the hypothesis $H_0 : \{\tau_1 = \tau_2 = \tau_3\}$, the experimenter deemed it important to be able to detect a difference in weight loss of at least $\Delta = 0.25$ g between any two soap types, with a probability 0.90 of correctly doing so, and a probability 0.05 of a Type I error. This difference was considered to be the smallest discrepancy in the weight loss of soaps that would be noticeable.

Using a one-way analysis of variance model, for $v = 3$ treatments, with $\Delta = 0.25$, $r = 2v\sigma^2\phi^2/\Delta^2 = 0.672\phi^2$, and $\nu_2 = v(r-1) = 3(r-1)$, r was calculated as follows. Using Table A.7 for $\nu_1 = v - 1 = 2$, $\alpha = 0.05$, and $\pi(\Delta) = 0.90$:

r	$\nu_2 = 3(r-1)$	ϕ	$r = 0.672\phi^2$	Action
	1000	2.25	3.40	Round up to $r = 4$
4	9	2.50	4.20	Round up to $r = 5$
5	12	2.50	4.20	Stop, and use $r = 4$ or 5.

The experimenter decided to take $r = 4$ observations on each soap type. Sections 3.8.3 and 3.9.4 show how to make these calculations using the SAS and R software, respectively. ☐

3.7 A Real Experiment—Soap Experiment, Continued

The objective of the soap experiment described in Sect. 2.5.1, p. xx, was to compare the extent to which three different types of soap dissolve in water. The three soaps selected for the experiment were a regular soap, a deodorant soap, and a moisturizing soap from a single manufacturer, and the weight-loss after 24 h of soaking and 4 days drying is reproduced in Table 3.6. Steps (a)–(d) of the checklist were given in Sect. 2.5.1. The remaining steps and part of the analysis of the experimental data are described below. The first part of the description is based on the written report of the experimenter, Suyapa Silvia.

Table 3.6 Data for the soap experiment

Soap	Weight-loss (grams)				$\overline{y}_{i.}$
1	−0.30	−0.10	−0.14	0.40	−0.0350
2	2.63	2.61	2.41	3.15	2.7000
3	1.86	2.03	2.26	1.82	1.9925

3.7.1 Checklist, Continued

(e) **Run a pilot experiment**.

A pilot experiment was run and used for two purposes. First, it helped to identify the difficulties listed at step (d) of the checklist. Secondly, it provided an estimate of σ^2 for step (h). The error variance was estimated to be about 0.007 g^2. The value 0.007 gm^2 was the value of msE in the pilot experiment. In fact, this is an underestimate, and it would have been better to have used the one-sided confidence bound (3.4.9) for σ^2.

(f) **Specify the model**.

Since care will be taken to control all extraneous sources of variation, it is assumed that the following model will be a reasonable approximation to the true model.

$$\begin{gathered} Y_{it} = \mu + \tau_i + \epsilon_{it} \,, \\ \epsilon_{it} \sim N(0, \sigma^2) \,, \\ \epsilon_{it}\text{'s are mutually independent} \\ i = 1, 2, 3; \quad t = 1, \ldots r_i ; \end{gathered}$$

where τ_i is the (fixed) effect on the response of the ith soap, μ is a constant, Y_{it} is the weight loss of the tth cube of the ith soap, and ϵ_{it} is a random error.
Before analyzing the experimental data, the assumptions concerning the distribution of the error variables will be checked using graphical methods. (Assumption checking will be discussed in Chap. 5).

(g) **Outline the analysis**.

In order to address the question of differences in weights, a one-way analysis of variance will be computed at $\alpha = 0.05$ to test

$$\begin{array}{ll} & H_0 : \{\tau_1 = \tau_2 = \tau_3\} \\ \text{versus} & H_A : \{\text{ the effects of at least two pairs of soap types differ}\}. \end{array}$$

To find out more about the differences among pairs of treatments, 95% confidence intervals for the pairwise differences of the τ_i will be calculated using Tukey's method (Tukey's method will be discussed in Sect. 4.4.4).

(h) **Calculate the number of observations that need to be taken**.

Four observations will be taken on each soap type. (See Example 3.6.1, p. 49, for the calculation.)

(i) **Review the above decisions. Revise if necessary**.

It is not difficult to obtain 4 observations on each of 3 soaps, and therefore the checklist does not need revising. Small adjustments to the experimental procedure that were found necessary during the pilot experiment have already been incorporated into the checklist.

3.7.2 Data Collection and Analysis

The data collected by the experimenter are plotted in Fig. 2.2, p. 24, and reproduced in Table 3.6. The assumptions that the error variables are independent and have a normal distribution with constant

Table 3.7 One-way analysis of variance table for the soap experiment

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio	p-value
Soap	2	16.1220	8.0610	104.45	0.0001
Error	9	0.6946	0.0772		
Total	11	16.8166			

variance were checked (using methods to be described in Chap. 5) and appear to be satisfied. The least squares estimates, $\hat{\mu} + \hat{\tau}_i = \overline{y}_{i.}$, of the average weight loss values (in grams) are

$$\overline{y}_{1.} = -0.0350\,, \qquad \overline{y}_{2.} = 2.7000\,, \qquad \overline{y}_{3.} = 1.9925.$$

The hypothesis of no differences in weight loss due to the different soap types is tested below using an analysis of variance test.

Using the values $\overline{y}_{i.}$ given above, together with $\sum\sum y_{it}^2 = 45.7397$ and $r_1 = r_2 = r_3 = 4$, the sums of squares for Soap and Total are calculated using (3.5.12) and (3.5.16), pp. 43 and 44, as

$$\begin{aligned} ssT &= \sum r_i \overline{y}_{i.}^2 - n\overline{y}_{..}^2 \\ &= \left[4(-0.0350)^2 + 4(2.7000)^2 + 4(1.9925)^2\right] - \left[12(1.5525)^2\right] = 16.1220\,, \\ sstot = ssE_0 &= \sum\sum y_{it}^2 - n\overline{y}_{..}^2 \\ &= 45.7397 - 12(1.5525)^2 = 16.8166. \end{aligned}$$

The sum of squares for error can be calculated by subtraction, giving $ssE = sstot - ssT = 0.6946$, or directly from (3.4.5), p. 39, as

$$\begin{aligned} ssE &= \sum\sum y_{it}^2 - \sum r_i \overline{y}_{i.}^2 \\ &= 45.7397 - \left[4(-0.0350)^2 + 4(2.7000)^2 + 4(1.9925)^2\right] = 0.6946\,. \end{aligned}$$

The estimate of error variability is then

$$\hat{\sigma}^2 = msE = ssE/(n - v) = 0.6945/(12 - 3) = 0.0772\,.$$

The sums of squares and mean squares are shown in the analysis of variance table, Table 3.7. Notice that the estimate of σ^2 is ten times larger than the estimate of 0.007 g^2 provided by the pilot experiment. This suggests that the pilot experiment was not sufficiently representative of the main experiment. As a consequence, the actual power of detecting a difference of $\Delta = 0.25$ g between the weight losses of the soaps is, in fact, somewhat below the desired probability of 0.90.

The decision rule for testing $H_0 : \{\tau_1 = \tau_2 = \tau_3\}$ against the alternative hypothesis, that at least two of the soap types differ in weight loss, using a significance level of $\alpha = 0.05$, is to reject H_0 if $msT/msE = 104.45 > F_{2,9,0.05}$. From Table A.6, $F_{2,9,0.05} = 4.26$. Consequently, the null hypothesis is rejected, and it is concluded that at least two of the soap types do differ in their weight loss after 24 h in water (and 4 days drying time). This null hypothesis would have been rejected for most practical choices of α. If α had been chosen to be as small as 0.005, $F_{2,9,\alpha}$ is still only 10.1. Alternatively, if the analysis is done by computer, the p-value would be printed in the computer output. Here the p-value is less than 0.0001, and H_0 would be rejected for any choice of α above this value.

The experimenter was interested in estimating the contrasts $\tau_i - \tau_u$ for all $i \neq u$, that is, she was interested in comparing the effects on weight loss of the different types of soaps. For the one-way analysis of variance model (3.3.1) and a completely randomized design, all contrasts are estimable, and the least squares estimate of $\tau_i - \tau_u$ is

$$\hat{\tau}_i - \hat{\tau}_u = (\hat{\mu} + \hat{\tau}_i) - (\hat{\mu} + \hat{\tau}_u) = \overline{y}_{i.} - \overline{y}_{u.} \,.$$

Hence, the least square estimates of the differences in the treatment effects are

$$\hat{\tau}_2 - \hat{\tau}_3 = 0.7075\,, \qquad \hat{\tau}_2 - \hat{\tau}_1 = 2.7350\,, \qquad \hat{\tau}_3 - \hat{\tau}_1 = 2.0275\,.$$

Confidence intervals for the differences will be evaluated in Example 4.4.5.

3.7.3 Discussion by the Experimenter

> The results of this experiment were unexpected in that the soaps reacted with the water in very different ways, each according to its ingredients. An examination of the soap packages showed that for the deodorant soap and the moisturizing soap, water is listed as the third ingredient, whereas the regular soap claims to be 99.44% pure soap. Information on the chemical composition of soaps revealed that soaps are sodium and/or potassium salts of oleic, palmitic, and coconut oils and therefore in their pure form (without water) should float as the regular soap bars do. The other two soaps under discussion contain water and therefore are more dense and do not float.
>
> One possible reason for the regular soap's actual increase in weight is that this "dry" soap absorbed and retained the water and dissolved to a lesser extent during the soaking period. The deodorant soap and the moisturizing soap, on the other hand, already contained water and did not absorb as much as the regular soap. They dissolved more easily during the soaking phase as a consequence. This is somewhat supported by the observation that the dissolved soap gel that formed extensively around the deodorant soap and the moisturizing soap did not form as much around the regular soap. Furthermore, the regular soap appeared to increase in size and remain larger, even at the end of the drying period.

3.7.4 Further Observations by the Experimenter

> The soaps were weighed every day for one week after the experimental data had been collected in order to see what changes continued to occur. The regular soap eventually lost most of the water it retained, and the average loss of weight (due to dissolution) was less than that for the other two soaps.
>
> If this study were repeated, with a drying period of at least one week, I believe that the results would indicate that regular soap loses less weight due to dissolution than either of the deodorant soap or the moisturizing soap

3.8 Using SAS Software

3.8.1 Randomization

A simple procedure for randomizing a completely randomized design was given in Sect. 3.2, p. 31. This procedure is easily implemented using the SAS software, as we now illustrate. Consider a completely randomized design for two treatments and $r = 3$ observations on each, giving a total of $n = 6$ observations. The following SAS statements create and print a data set named `DESIGN`, which includes the lists of values of the two variables `TRTMT` and `RANNO` as required by steps 1 and 2 of the randomization procedure in Sect. 3.2. The statements `INPUT` and `LINES` are instructions to SAS that the values of `TRTMT` are being input on the lines that follow rather than from an external data file. Inclusion of "`@@`" in the `INPUT` statement allows the levels of `TRTMT` to be entered on one line as opposed to one per line. For each treatment label entered for the variable `TRTMT`, a corresponding value of `RANNO` is generated using the SAS random number generating function `RANUNI` which generates uniform random numbers between 0 and 1.

```
DATA DESIGN;
  INPUT TRTMT @@;
  RANNO=RANUNI(0);
  LINES;
  1 1 1 2 2 2
;
PROC PRINT; RUN;
```

The statement `PROC PRINT` then prints the following output. The column labeled `OBS` (observation number) is generated by the SAS software for reference.

```
The SAS System
Obs    TRTMT     RANNO
 1       1      0.74865
 2       1      0.62288
 3       1      0.87913
 4       2      0.32869
 5       2      0.47360
 6       2      0.72967
```

The following additional statements sort the data set by the values of `RANNO`, as required by step 3 of the randomization procedure, and print the randomized design along with the ordered experimental unit labels 1–6 under the heading `OBS`.

```
PROC SORT; BY RANNO;
PROC PRINT; RUN;
```

The resulting output is as follows.

```
The SAS System
Obs    TRTMT     RANNO
 1       2      0.32869
 2       2      0.47360
 3       1      0.62288
 4       2      0.72967
 5       1      0.74865
 6       1      0.87913
```

Experimental units 3, 5, and 6 are assigned to treatment 1, and experimental units 1, 2, and 4 are assigned to treatment 2.

Table 3.8 SAS program for the soap experiment

```
Line   SAS Program
  1    OPTIONS LINESIZE = 72;

  2    DATA SOAP;
  3      INPUT WTLOSS SOAP;
  4      LINES;
  5      -0.30 1
  6      -0.10 1
  7      -0.14 1
  8        :   :
  9       1.82 3
 10    ;

 11    PROC PRINT;

 12    PROC SGPLOT;
 13      SCATTER X = SOAP Y = WTLOSS;
 14      XAXIS TYPE = DISCRETE LABEL = 'Soap';
 15      YAXIS LABEL = 'Weight Loss (grams)';

 16    PROC GLM;
 17      CLASS SOAP;
 18      MODEL WTLOSS = SOAP;
 19      LSMEANS SOAP;

 20    RUN; QUIT;
```

3.8.2 Analysis of Variance

In this section we illustrate how SAS software can be used to conduct a one-way analysis of variance test for equality of the treatment effects, assuming that model (3.3.1) is appropriate. We use the data in Table 2.7, p. 23, from the soap experiment.

A sample SAS program to analyze the data is given in Table 3.8. Line numbers have been included for reference, but the line numbers are not part of the SAS program and if included would cause SAS software to generate error messages. SAS programs and data files for this edition are available at the following website.

http://www.wright.edu/~dan.voss/DeanVossDraguljic.html

The option `LINESIZE = 72` in the `OPTIONS` statement in line 1 of the program causes all list output generated by the program to be restricted to 72 characters per line, which is convenient for printing list output on 8.5 by 11 inch paper in the portrait orientation. This option has no effect on the standard html output, however, so can be ignored by readers running the SAS software in a windows environment, which is most likely to be the norm. Some of our SAS programs, including those using `PROC SGPLOT` for example, assume the user is running SAS in a windows environment, whereas `PROC PLOT` might be used instead if running SAS in a command line mode. All SAS statements are ended by a semicolon.

Lines 2–10 of the program create a SAS data set named `SOAP` that includes as variables the response variable `WTLOSS` and the corresponding level of the treatment factor `SOAP`. The `LINES` statement indicates that subsequent lines contain data to be read directly from the program, until data entry is

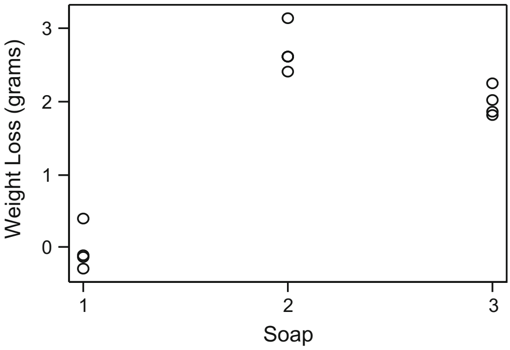


Fig. 3.3 Data plot from the SAS software for the soap experiment

stopped by the next semicolon (line 10). Line 8 must be replaced by the additional data not shown here.

Alternatively, the same data could be read from a file, `soap.txt` say, by replacing lines 2–10 with the following code, including a correct file path.

```
INFILE 'c:\path\soap.txt' FIRSTOBS = 2; INPUT WTLOSS SOAP;
```

Include the option `FIRSTOBS = 2` if the data file contains headers on line one then data starting on line two, but delete it if the data starts on line one with no headers.

The `PRINT` procedure (line 11) prints the data. While this is good practice to verify that the data were read correctly, the `PRINT` procedure will not routinely be shown in subsequent programs.

The `SGPLOT` procedure (lines 12–15) generates a scatterplot of `WTLOSS` versus `SOAP` like that shown in Fig. 3.3. The x-axis option `TYPE = DISCRETE` instructs the SAS software to use integer values for x-axis tick marks. The `LABEL` option sets the desired label for each axis.

The resulting scatterplot is displayed in a SAS output window, by default (on a PC). Alternatively, one could redirect the scatterplot to be saved in the file `ch3soap.pdf` in pdf format, for example, as illustrated by the following code.

```
ODS GRAPHICS / RESET IMAGENAME ='ch3soap' IMAGEFMT = PDF
               HEIGHT = 1.5in WIDTH = 2in;
ODS LISTING GPATH ='c:\path\figs';
* insert PROC SGPLOT and its subcommands here;
RUN; * Run PROC SGPLOT before closing output to pdf file;
ODS GRAPHICS / RESET;
```

The statements beginning "*" and ending ";" are comments which are not executed by the SAS software. The first `ODS GRAPHICS` command redirects graphics output to the file `ch3soap.pdf` using pdf as the image format, and specifies the dimension of the graphic image to be saved. The `ODS LISTING` command then specifies the directory where the SAS software is to store the file, so the user must replace "`c:\path\figs`" with an existing path and directory on the user's computer. Following `PROC SGPLOT` and its statements, the second `ODS GRAPHICS` command resets the graphics defaults, so graphics output reverts again to the SAS output window. However, before doing so, the `RUN` command causes `SGPLOT` to execute while output is still directed to the file `ch3soap.pdf`.

The General Linear Models procedure `PROC GLM` (lines 16–19) generates an analysis of variance table. The `CLASS` statement identifies `SOAP` as a major source of variation to be modeled as a *classification variable*, so a parameter is associated with each of its levels. The `MODEL` statement defines the response variable as `WTLOSS`, and the only source of variation included in the model is `SOAP`. The

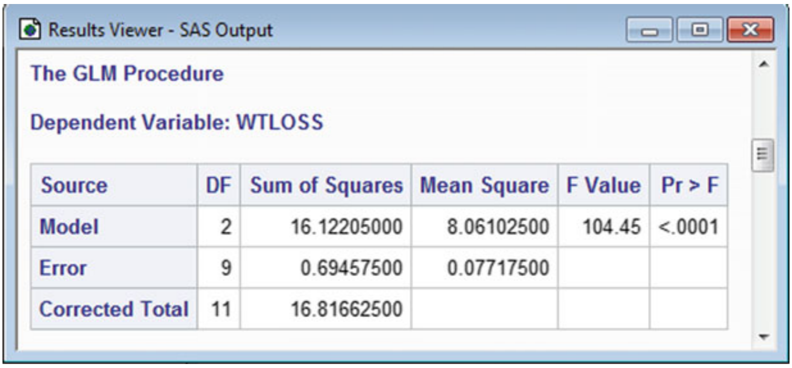


Fig. 3.4 Sample SAS output from `PROC GLM` for the soap experiment

The GLM Procedure

Dependent Variable: WTLOSS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	16.12205000	8.06102500	104.45	<.0001
Error	9	0.69457500	0.07717500		
Corrected Total	11	16.81662500			

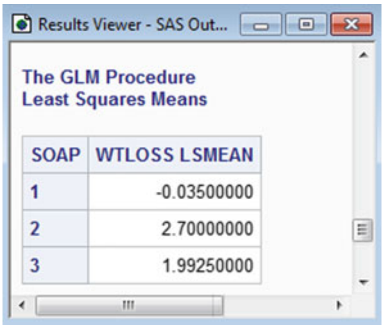


Fig. 3.5 Output from the `LSMEANS` statement for the soap experiment

The GLM Procedure
Least Squares Means

SOAP	WTLOSS LSMEAN
1	-0.03500000
2	2.70000000
3	1.99250000

parameter μ and the error variables are automatically included in the model. The `MODEL` statement causes the analysis of variance table shown in Fig. 3.4 to be calculated. The `F Value` is the value of the ratio *msT/msE* for testing the null hypothesis that the three treatment effects are all equal. The value `Pr > F` is the p-value of the test to be compared with the chosen significance level. When the p-value is listed as $<.0001$, the null hypothesis would be rejected for any chosen significance level larger than 0.0001.

The `LSMEANS` statement (line 19 of Table 3.8), which is part of the `GLM` procedure, causes the least squares means, $\hat{\mu} + \hat{\tau}_i = \overline{y}_{i.}$, to be printed. The output from this statement is shown in Fig. 3.5.

The `RUN` statement in line 20 is needed to cause the last procedure to be executed when the program is run in an interactive line mode, typical of a PC Windows environment for example, and the `QUIT` statement ends the procedure. Though necessary for interactive program processing, the `RUN` and `QUIT` statements will not be shown from now on in any programs.

3.8.3 Calculating Sample Size Using Power of a Test

In Table 3.9, we show a sample SAS program which calculates the power of the test of the null hypothesis $H_0 : \{\tau_1 = \cdots = \tau_v\}$ against H_A: {at least two of the τ_i's differ}. The program uses a `DO` statement, which allows the calculation to be done for a selected range of sample sizes r, using the formulae in Sect. 3.6.2. The line `DO R = 3 TO 6 BY 1;` asks the SAS software to do the calculations for each value of r between 3 and 6, increasing r by 1 each time.

The code shown is for the soap experiment in Example 3.6.1, but is easily modified for other experiments by changing the values of the number of levels of the treatment factor (`V`), the difference

Table 3.9 Calculating sample sizes using power of the test

```
DATA POWER;
  V = 3;
  DEL = 0.25;
  SIGMA2 = 0.007;
  ALPHA = 0.05;
  NU1 = V - 1;
  LHTPB = 1 - ALPHA;

DO R = 3 TO 6 BY 1;
  NU2 = V*(R - 1);
  PHI = (SQRT(R / (2*V*SIGMA2)))*DEL;
  FVALUE = FINV(LHTPB, NU1, NU2);
  NONCN = V*PHI**2;
  POWER = 1 - PROBF(FVALUE, NU1, NU2, NONCN);
  OUTPUT;
END;

PROC PRINT;
  VAR R POWER;
```

(`DEL`) to be detected (i.e. Δ), the assumed largest value of the error variance (`SIGMA2`), the significance level of the test (`ALPHA`), and the range of values of r to be investigated.

In Table 3.9, the degrees of freedom $\nu_1 = v - 1$ and $\nu_2 = n - v = v(r - 1)$ for the F-distribution are donated by `NU1` and `NU2`. The "left-hand tail probability" $1 - \alpha$ is called `LHTPB`, and is used in calculating the critical value $F_{v-1,n-v,\alpha}$, called `FVALUE`. From (3.6.20), the value of ϕ, labelled `PHI` is calculated as $\sqrt{r\Delta^2/(2v\sigma^2)}$. The "non-centrality parameter" `NONCN` is $\delta^2 = v\phi^2$ and this is needed by the non-central F distribution in the calculation of the power for the range of values of r specified. The output, generated by the `PROC PRINT` statement is

```
Obs    R     POWER
 2     3    0.70934
 3     4    0.89565
 4     5    0.96715
 5     6    0.99058
```

and we can see that, just as in Example 3.6.1, to achieve a power of approximately 0.9, the number of observations needed is $r = 4$ per level of the treatment factor. To see the values of all the variables calculated at each step, remove the line `VAR R POWER;` which restricts which variables are printed.

3.9 Using R Software

Preliminaries

R is a free software environment for statistical computing and graphics, used extensively in this book. Readers can install the R software after downloading it from http://cran.us.r-project.org, for example, choosing the appropriate version for the computer and operating system. RStudio is free software providing an enhanced environment for running R. After installing R, readers are recommended to also download and install RStudio; it can be downloaded from http://www.rstudio.com, choosing again the appropriate version for the computer and operating system. Run either R or RStudio, as RStudio invokes R.

Throughout the book, we shall assume that the reader has set up a working directory for R called `RCode` and that R program files are either in the working directory or in specified subdirectories of the working directory. For example, we assume that RCode includes a subdirectory called `data` containing any data set that is to be read by R and a subdirectory called `figs` in which R will store any plots generated. R programs and data files for this edition are available at the following website.

http://www.wright.edu/~dan.voss/DeanVossDraguljic.html

We shall assume that the user will execute the following commands or similar each time that R (or RStudio) is started.

```
rm(list = ls())   # Remove all objects (start with clean slate)
opar = par()      # Save default graphics parameters as opar
setwd("/RCode")   # Set the working directory
getwd()           # Confirm working directory
options(show.signif.stars = FALSE) # Show no stars for significance tests
options(width = 72, digits = 5, scipen = 2) # Control printed output
ooptions = options(width = 72, digits = 5, scipen = 2) # Save print options
```

The first command removes all existing R objects created previously by the user, clearing the slate for a new session. The symbol "#" starts a line comment, used for program documentation. The second command assigns R's current graphics parameters `par()`—initially the default graphics parameter values—to the object `opar`, saving them so they can be restored later via the command `par(opar)` if desired. We will routinely use "=" for assignment, though use of "<–" is more traditional in R. The `setwd` command in the third line sets `/RCode` as the *working directory*, where the software reads and writes files by default. If RCode is not in the root directory, then use `setwd("/path/RCode")` but specify the correct directory path to RCode. The `getwd()` command in the fourth line displays the working directory, to confirm it is now `/RCode`. For functions that conduct hypothesis tests, the `options` command in the fifth line suppresses printing of stars for various levels of significance. The `options` command in line six controls printed output, restricting it to be at most 72 columns wide with five significant digits, and penalizing use of scientific notation. We have initialized R in this way or similarly when running our programs, though these commands will generally not be shown in our subsequent program code. The last line saves these print options as `ooptions`, so if changed they can be restored by the command `options(ooptions)`.

While the above commands can be typed into the R Console and executed one by one, it is more convenient to save them in a file, `startup.r` say, in the working directory. Then the single command

```
source("/RCode/startup.r")
```

will execute the commands in the `startup.r` file. We routinely executed this code line each time we started R (or RStudio) to produce R program output in this book, though we do not show this code line in our programs.

As will be seen in the following sections, when R is waiting for the next command, a prompt > is displayed, and if the user command is not complete when a line is entered (for example if the final parenthesis is missing), the prompt will change to + on the next line, prodding the user to enter the rest of the command. To end an R session, type `q()`.

It is prudent to use the latest production version of R. On a Windows operating system, the `updateR` command of the `installr` package will detect if there is a new R version available, and if so it will download and install it and update previously installed add-on packages. The following commands, when executed from within R, install and load the `installr` package and execute the `updateR` command.

```
install.packages("installr"); library(installr); updateR()
```

Assuming the user has installed and set up R as noted above, we are ready to use the software.

3.9.1 Randomization

A simple procedure for randomizing a completely randomized design was given in Sect. 3.2, p. 31. This procedure is easily implemented using the R software, as we now illustrate. Consider a completely randomized design for two treatments and $r = 3$ observations on each, giving a total of $n = 6$ observations.

The following R statements create and display a data frame named `design`, a *data frame* being a data set consisting of equal-length columns of information. Here, the columns of `design` are the lists of values of the two variables `trtmt` and `ranno` as required by steps 1 and 2 of the randomization procedure in Sect. 3.2. In particular, the first statement creates the column `trtmt` of treatment labels. The second statement creates the column `ranno` consisting of six uniform random numbers between 0 and 1, six being the length of the column `trtmt`. The third statement puts the columns `trtmt` and `ranno` into a data frame, and assigns this object the name `design`. Then the fourth statement displays `design`. The R output is shown immediately following the R statements.

```
> trtmt = c(1, 1, 1, 2, 2, 2)  # Create column trtmt = (1, 1, 1, 2, 2, 2)
> ranno = runif(length(trtmt))  # Create column of 6 unif(0, 1) RVs
> design = data.frame(trtmt, ranno) # Create data.frame "design"
> design  # Display the data.frame design

  trtmt    ranno
1     1 0.447827
2     1 0.494462
3     1 0.174414
4     2 0.894132
5     2 0.473540
6     2 0.010771
```

We digress to provide additional information about R, before finishing the randomization process in the next paragraph. The command `trtmt = rep(c(1, 2), each = 3)` would yield the same column `trtmt`, but by replicating 1 and 2 three times each—a more convenient approach for larger designs. Each time the above R commands are run, a different set of random numbers will result. Typing the command `design` causes the entire data frame `design` to be displayed. One could display only what is in column 1 named `trtmt`, for example, by typing `design$trtmt`, `design[, 1]`, or `design[, "trtmt"]`. The column `trtmt` also still exists as a separate object, that can be displayed simply by typing `trtmt`. One could remove this redundant object by the command `rm(trtmt)`. Note that R is “case-sensitive”, so if the column name is `trtmt`, then R will not be able to locate a column called, say, `Trtmt` with a capital `T`. The details about any of the commands used can be found by typing `?commandName`, for example typing `?runif` will bring up the command help file containing many details about the use of `runif`.

The following additional statements sort the `trtmt` column of the data frame `design` by the values of `ranno`, as required by step 3 of the randomization procedure. Specifically, the statement `order(ranno)` yields the order 6 3 1 5 2 4, since the smallest random variate is in row 6, the second smallest is in row 3, etc. So, the first statement below redefines the data frame `design` to have its rows reordered accordingly, effectively sorting the rows based on the values of the random numbers (RNs). The second statement below defines a new column of `design` named `EU` containing the integers from

Table 3.10 R program for the soap experiment: reading and plotting data

```
Line  R Code and Output
  1   > # Read the data into the data.frame "soap.data"
  2   > soap.data = read.table("data/soap.txt", header = TRUE)
  3   > head(soap.data, 5)  # Display first 5 lines of soap.data

  4     Soap Cube PreWt PostWt WtLoss
  5   1    1    1 13.14  13.44  -0.30
  6   2    1    2 13.17  13.27  -0.10
  7   3    1    3 13.17  13.31  -0.14
  8   4    1    4 13.17  12.77   0.40
  9   5    2    5 13.03  10.40   2.63

 10   > # Add factor variable fSoap to soap.data for later ANOVA
 11   > soap.data$fSoap = factor(soap.data$Soap)

 12   > # Plot WtLoss vs Soap, specify axis labels, suppress x-axis.
 13   > plot(WtLoss ~ Soap, data = soap.data, xlab = "Soap",
 14   +      ylab = "Weight Loss (grams)", las = 1, xaxt = "n")
 15   > # Insert x-axis (axis 1) with tick marks from 1 to 3 by 1.
 16   > axis(1, at = seq(1,3,1))
```

1 to 6 as labels for the experimental units. Then the last statement asks for the sorted `design` to be displayed.

```
> design = design[order(ranno), ]  # Sort rows by RNs, save
> design$EU = c(1:6)  # Add col EU = (1,2,3,4,5,6) to design
> design  # Display the results of the randomization

  trtmt    ranno EU
6     2 0.010771  1
3     1 0.174414  2
1     1 0.447827  3
5     2 0.473540  4
2     1 0.494462  5
4     2 0.894132  6
```

Experimental units 2, 3, and 5 are to be assigned to treatment 1, and experimental units 1, 4, and 6 are to be assigned to treatment 2.

3.9.2 Reading and Plotting Data

A sample R program to input, display, and plot the data is given in Table 3.10. Line numbers have been included for reference, but they are not part of the R program and if included in the R code would yield error messages. The prompt ">" and the continuation prompt "+" are supplied by R and should not be typed by the user.

We use the data in Table 2.7, p. 23, from the soap experiment, and assume that the data are stored in the file `soap.txt` in the `data` subdirectory of the working directory; that is, in `data/soap.txt`.

Line 2 of Table 3.10 reads the data from the file `soap.txt` and puts it into an R data set (data frame) called `soap.data`. Alternatively, one could enter the data via the keyboard, which is the standard input device `stdin()`, by replacing line 2 with the following.

```
> soap.data = read.table(stdin(), header = TRUE)
0: Soap Cube PreWt PostWt WtLoss
1:    1    1 13.14  13.44  -0.30
2:    1    2 13.17  13.27  -0.10
 :    :    :   :      :      :
12:   3   12 13.00  11.18   1.82
13:
```

Keyboard data entry is ended by hitting the return key twice.

The `head(soap.data, 5)` command in line 3 displays the first five lines of the data set shown in lines 4–9. Alternatively, the command `head(soap.data)` would display the first six lines by default, and the command `soap.data` would display the full data set. The first column displayed indicates the data is from rows 1–5 of the data set, and the other five columns show the five variables in the data set, including the response variable `WtLoss` and the corresponding level of the treatment factor `Soap`. In line 2, the statement `header = TRUE` (i.e. `header = T`) tells R that the columns of data in the file `soap.txt` have headings, and these can be seen in line 4. If a data file has no headings, the header statement may be omitted, the default being `header = FALSE` (i.e. `header = F`).

In line 11, the data set `soap.data` is augmented with a new variable, `fSoap`, created by converting the numerical variable `Soap` to a factor variable, needed later for the analysis of variance.

The remaining code generates a plot of the data. The `plot` command in lines 13–14 generates a scatterplot of `WtLoss` versus `Soap` like that shown in Fig. 3.6, with labels specified for each axis by `xlab` and `ylab`. The option `data = soap.data` indicates that the variables being plotted are in the data set `soap.data`. Alternatively, one could use the syntax

```
plot(soap.data$WtLoss ~ soap.data$Soap, xlab ="Soap",
```

in line 13. The dollar sign identifies specific columns of the data set, so, for example, `soap.data$Soap` just means "read the column labeled `Soap` from the data set `soap.data`". The + prompt in line 14 indicates that the prior command is not yet complete. The option `las = 1` sets *labels style* 1, making tick mark labels horizontal, impacting y-axis labeling. The option `xaxt = "n"` in line 14 suppresses the automatically generated x-axis (which would have five tick marks), then the `axis` command in line 16 includes instead an x-axis with three tick marks specified to be at 1, 2 and 3—namely, a sequence starting at 1 and ending at 3 in steps of size 1. If the numerical variable `Soap` was replaced by the factor variable `fSoap` in line 13, then a box plot would be obtained instead of a scatterplot.

The resulting scatterplot is displayed in a graphics window, by default (on a PC). Alternatively, one can redirect the scatterplot to be saved in the file `ch3soap.pdf` in pdf format, for example, as illustrated by the following code.

```
pdf("figs/ch3soapplot.pdf", width = 5, height = 3)  # Open a pdf file
# Insert plot command and its subcommands here
dev.off()  # Close the pdf file
```

The `pdf` command opens the pdf file `ch3soapplot.pdf` in the `figs` subdirectory of the working directory (as specified by the user upon startup), and specifies the dimension of the graphic image to be saved. Once this has been done, `plot` function calls will send output to the pdf file.

Fig. 3.6 R data plot for the soap experiment

Then the `dev.off()` command closes the pdf file, so graphical output reverts to the default graphics window.

3.9.3 Analysis of Variance

In this section we illustrate how the R software can be used to conduct a one-way analysis of variance test for equality of the treatment effects, assuming model (3.3.1) is appropriate. Table 3.11 contains a sample program and output, with line numbers again included for reference only. We continue to use the soap experiment data, which in line 1 is read from the file `soap.txt` into the data set `soap.data`.

In line 3, we convert the numerical variable `Soap` to a factor variable, saving it as a new variable `fSoap` of the `soap.data` data set. The statement in line 4 generates summary statistics for the variables in columns 1, 5 and 6 of the data set, with the output shown in lines 5–11. If all six variables are to be summarized, the statement `summary(soap.data)` without column numbers is sufficient. One can see that the `summary` command treats the numeric variable `Soap` and the factor variable `fSoap` differently, providing summary statistics for the `Soap` values, but levels and frequencies for `fSoap`. A *factor variable* is treated as a qualitative variable by R, analogous to a `CLASS` variable in SAS software.

The `aov` function in line 12 fits a linear model to the soap data, specifying `WtLoss` as the response variable and `fSoap` as the primary source of variation, the symbol "$\sim$" separating and distinguishing these, saving the resulting information as the object `model1`. Because `fSoap` is a factor variable, it is modeled as a classification variable as desired. The parameter μ and the error variables are automatically included in the model. The `anova(model1)` command in line 13 displays the one-way analysis of variance information shown in lines 14–19. In lines 17–18, the `F value` is the value of the ratio *msT/msE* for testing the null hypothesis that the three treatment effects are all equal, and `Pr(>F)` is the p-value of the test to be compared with the chosen significance level. The listed p-value is 5.91×10^{-7}, so the null hypothesis is rejected for any chosen significance level larger than this small value.

The `library(lsmeans)` command in line 21 loads the package `lsmeans` from the user's library for subsequent use; (this assumes the reader has already installed the package as discussed in the next paragraph). The `lsmeans` command in line 22 generates the least squares means for the three levels of the factor `fSoap`, plus further statistical information which will be discussed in Chap. 4. The output is shown in lines 23–28.

Table 3.11 R program for the soap experiment: analysis of variance and least squares means

```
Line  R Code or Output
  1   > soap.data = read.table("data/soap.txt", header = TRUE)
  2   > # Add factor variable fSoap to soap.data for ANOVA
  3   > soap.data$fSoap = factor(soap.data$Soap)
  4   > summary(soap.data[,c(1,5:6)]) # Summarize data in cols 1, 5, 6

  5         Soap       WtLoss       fSoap
  6    Min.   :1   Min.   :-0.300   1:4
  7    1st Qu.:1   1st Qu.: 0.275   2:4
  8    Median :2   Median : 1.945   3:4
  9    Mean   :2   Mean   : 1.552
 10    3rd Qu.:3   3rd Qu.: 2.460
 11    Max.   :3   Max.   : 3.150

 12   > model1 = aov(WtLoss ~ fSoap, data = soap.data)
 13   > anova(model1)

 14   Analysis of Variance Table
 15
 16   Response: WtLoss
 17             Df Sum Sq Mean Sq F value  Pr(>F)
 18   fSoap      2  16.12    8.06     104 5.9e-07
 19   Residuals  9   0.69    0.08

 20   > # install.packages("lsmeans")
 21   > library(lsmeans)
 22   > lsmeans(model1, "fSoap")

 23    fSoap  lsmean     SE df lower.CL upper.CL
 24    1     -0.0350 0.1389  9 -0.34922  0.27922
 25    2      2.7000 0.1389  9  2.38578  3.01422
 26    3      1.9925 0.1389  9  1.67828  2.30672
 27
 28    Confidence level used: 0.95
```

Add-On Packages

Installation of the R software (see Sect. 1.2) installs the base software, including some base packages providing limited functionality. There are thousands of additional user-defined packages that the user may freely download, the `lsmeans` package introduced above being one example. To use any such function not included in the base software installation, the "add-on" package containing the function must first be installed and loaded. For example, the command `install.packages("lsmeans")` in line 20 installs the `lsmeans` package, permanently saving it in a library of packages on the user's computer, so the command `library(lsmeans)` can load the `lsmeans` package from the user's library. A package only needs to be installed once. However, any add-on package must be loaded by the user prior to its first use in any new R session. As such, when our programs require an add-on package, we will routinely include the necessary `library` command to load the package. Furthermore, when we use an add-on package for the first time, the corresponding program will include the corresponding `install.packages` command, but commented out. Before running such a program the first time,

Table 3.12 Calculating sample sizes using power of the test

```
> #install.packages(pwr)
> library(pwr)
> v = 3; del = 0.25; sig2 = 0.007; alpha = 0.05; pwr = 0.90
> pwr.anova.test(k = v, sig.level = alpha, power = pwr,
+                f = sqrt(del^2/(2*v*sig2)))

     Balanced one-way analysis of variance power calculation

              k = 3
              n = 4.038656
              f = 1.219875
      sig.level = 0.05
          power = 0.9

NOTE: n is number in each group
```

the reader can simply delete the comment character, "#", so the package gets installed. On most systems, the process is automatic.

For linux users, when the `install.packages` command is invoked for the first time, you may get a warning that says in effect that a library is not writable and it will ask you whether you would like to use a personal library. If you answer `y`, it will ask you if you would like to create one. If you answer `y` again, it will ask you to select a CRAN "mirror" (i.e. a site from which to download the package for installation). Select (give the number of) any site near your location, and then R will create the personal library and download the package. After this, the `install.packages` command will proceed automatically.

3.9.4 Calculating Sample Size Using Power of a Test

In Table 3.12, we show a sample R program which calculates the power of the test of the null hypothesis $H_0 : \{\tau_1 = \cdots = \tau_v\}$ against H_A: {at least two of the τ_i's differ}. The code shown is for the soap experiment in Example 3.6.1, but is easily modified for other experiments by changing the values of the number of levels of the treatment factor (`v`), the difference (`del`) to be detected (i.e. Δ), the assumed largest value of the error variance (`sig2`), the significance level of the test (`alpha`), and the desired power of the test `pwr`.

The R program from Table 3.12 uses function `pwr.anova.test` which can be found in the package `pwr`. The function's inputs `k = v`, `sig.level = alpha`, and `power = pwr` are self-explanatory. The degrees of freedom $\nu_1 = v - 1$, $\nu_2 = n - v = v(r - 1)$, and the critical value $F_{v-1,n-v,\alpha}$ for the F-distribution corresponding to the required significance level are calculated internally by R. From (3.6.20), the value of $\phi/\sqrt{r}$, labeled as input `f`, is calculated as $\sqrt{\Delta^2/(2v\sigma^2)}$ and is needed by the non-central F distribution in the calculation of the power for different values of r. The output is also shown in Table 3.12. and we can see that, just as in Example 3.6.1, to achieve a power of approximately 0.9, the number of observations needed is $r = 4$ per level of the treatment factor. Notice that r is labelled `n` in the R output.

Exercises

1. Suppose that you are planning to run an experiment with one treatment factor having four levels and no blocking factors. Suppose that the calculation of the required number of observations has given $r_1 = r_2 = r_3 = r_4 = 5$. Assign at random 20 experimental units to the $v = 4$ levels of the treatments, so that each treatment is assigned 5 units.
2. Suppose that you are planning to run an experiment with one treatment factor having three levels and no blocking factors. It has been determined that $r_1 = 3$, $r_2 = r_3 = 5$. Assign at random 13 experimental units to the $v = 3$ treatments, so that the first treatment is assigned 3 units and the other two treatments are each assigned 5 units.
3. Suppose that you are planning to run an experiment with three treatment factors, where the first factor has two levels and the other two factors have three levels each. Write out the coded form of the 18 treatment combinations. Assign 36 experimental units at random to the treatment combinations so that each treatment combination is assigned two units.
4. For the one-way analysis of variance model (3.3.1), p. 33, the solution to the normal equations used by the SAS software is $\hat{\tau}_i = \overline{y}_{i.} - \overline{y}_{v.}$ $(i = 1, \ldots, v)$ and $\hat{\mu} = \overline{y}_{v.}$.

   (a) Is τ_i estimable? Explain.
   (b) Calculate the expected value of the least squares estimator for $\tau_1 - \tau_2$ corresponding to the above solution. Is $\tau_1 - \tau_2$ estimable? Explain.

5. Consider a completely randomized design with observations on three treatments (coded 1, 2, 3). For the one-way analysis of variance model (3.3.1), p. 33, determine which of the following are estimable. For those that are estimable, state the least squares estimator.

   (a) $\tau_1 + \tau_2 - 2\tau_3$.
   (b) $\mu + \tau_3$.
   (c) $\tau_1 - \tau_2 - \tau_3$.
   (d) $\mu + (\tau_1 + \tau_2 + \tau_3)/3$.

6. (requires calculus) Show that the normal equations for estimating μ, $\tau_1, \ldots, \tau_v$ are those given in Eq. (3.4.3) on p. 35.
7. (requires calculus) Show that the least squares estimator of $\mu + \tau$ is $\overline{Y}_{..}$ for the linear model $Y_{it} = \mu + \tau + \epsilon^0_{it}$ $(t = 1, \ldots, r_i; i = 1, 2, \ldots, v)$, where the ϵ^0_{it}'s are independent random variables with mean zero and variance σ^2. (This is the reduced model for the one-way analysis of variance test, Sect. 3.5.1, p. 41.)
8. For the model in the previous exercise, find an unbiased estimator for σ^2. (Hint: first calculate $E[ssE_0]$ in (3.5.10), p. 42.)
9. (requires calculus) Find the least squares estimates of $\mu_1, \mu_2, \ldots, \mu_v$ for the linear model $Y_{it} = \mu_i + \epsilon_{it}$ $(t = 1, \ldots, r_i; i = 1, 2, \ldots, v)$, where the ϵ_{it}'s are independent random variables with mean zero and variance σ^2. Compare these estimates with the least squares estimates of $\mu + \tau_i$ $(i = 1, 2, \ldots, v)$ in model (3.3.1), p. 33.
10. For the model in the previous exercise, find an unbiased estimator for σ^2. Compare the estimator with that in (3.4.7), p. 39.
11. Verify, for the one-way analysis of variance model (3.3.1), p. 33, that each treatment sample variance S_i^2 is an unbiased estimator of the error variance σ^2, so that

$$E(SSE) = \sum_i (r_i - 1)E(S_i^2) = (n - v)\sigma^2.$$

Table 3.13 Times (in seconds) for the balloon experiment

Time order	1	2	3	4	5	6	7	8
Coded color	1	3	1	4	3	2	2	2
Inflation time	22.0	24.6	20.3	19.8	24.3	22.2	28.5	25.7
Time order	9	10	11	12	13	14	15	16
Coded color	3	1	2	4	4	4	3	1
Inflation time	20.2	19.6	28.8	24.0	17.1	19.3	24.2	15.8
Time order	17	18	19	20	21	22	23	24
Coded color	2	1	4	3	1	4	4	2
Inflation time	18.3	17.5	18.7	22.9	16.3	14.0	16.6	18.1
Time order	25	26	27	28	29	30	31	32
Coded color	2	4	2	3	3	1	1	3
Inflation time	18.9	16.0	20.1	22.5	16.0	19.3	15.9	20.3

12. **Balloon experiment**

Prior to 1985, the experimenter (Meily Lin) had observed that some colors of birthday balloons seem to be harder to inflate than others. She ran this experiment to determine whether balloons of different colors are similar in terms of the time taken for inflation to a diameter of 7 inches. Four colors were selected from a single manufacturer. An assistant blew up the balloons and the experimenter recorded the times (to the nearest 1/10 second) with a stop watch. The data, in the order collected, are given in Table 3.13, where the codes 1, 2, 3, 4 denote the colors pink, yellow, orange, blue, respectively.

(a) Plot inflation time versus color and comment on the results.
(b) Estimate the mean inflation time for each balloon color, and add these estimates to the plot from part (a).
(c) Construct an analysis of variance table and test the hypothesis that color has no effect on inflation time.
(d) Plot the data for each color in the order that it was collected. Are you concerned that the assumptions on the model are not satisfied? If so, why? If not, why not?
(e) Is the analysis conducted in part (c) satisfactory?

13. **Heart–lung pump experiment, continued**

The heart–lung pump experiment was described in Example 3.4.1, p. 37, and the data were shown in Table 3.2, p. 38.

(a) Calculate an analysis of variance table and test the null hypothesis that the different number of revolutions per minute have the same effects on the fluid flow rate.
(b) Are you happy with your conclusion? Why or why not?
(c) Calculate a 90% upper confidence limit for the error variance σ^2.

14. **Meat cooking experiment**

(L. Alvarez, M. Burke, R. Chow, S. Lopez, and C. Shirk, 1998)

Table 3.14 Post-cooking weight data (in grams) for the meat cooking experiment

Frying			Grilling		
Fat content			Fat content		
10%	15%	20%	10%	15%	20%
81	85	71	84	83	78
88	80	77	84	88	75
85	82	72	82	85	78
84	80	80	81	86	79
84	82	80	86	88	82

Table 3.15 Data for the trout experiment

Code	Hemoglobin (grams per 100 ml)									
1	6.7	7.8	5.5	8.4	7.0	7.8	8.6	7.4	5.8	7.0
2	9.9	8.4	10.4	9.3	10.7	11.9	7.1	6.4	8.6	10.6
3	10.4	8.1	10.6	8.7	10.7	9.1	8.8	8.1	7.8	8.0
4	9.3	9.3	7.2	7.8	9.3	10.2	8.7	8.6	9.3	7.2

Source: Gutsell (1951). Copyright © 1951 International Biometric Society. Reprinted with permission

An experiment was run to investigate the amount of weight lost (in grams) by ground beef hamburgers after grilling or frying, and how much the weight loss is affected by the percentage fat in the beef before cooking. The experiment involved two factors: cooking method (factor A, with two levels frying and grilling, coded 1, 2), and fat content (factor B, with three levels 10, 15, and 20%, coded 1, 2, 3). Thus there were six treatment combinations 11, 12, 13, 21, 22, 23, relabeled as treatment levels 1, 2, …, 6, respectively. Hamburger patties weighing 110 g each were prepared from meat with the required fat content. There were 30 "cooking time slots" which were randomly assigned to the treatments in such a way that each treatment was observed five times ($r = 5$). The patty weights after cooking are shown in Table 3.14.

(a) Plot the data and comment on the results.
(b) Write down a suitable model for this experiment.
(c) Calculate the least squares estimate of the mean response for each treatment. Show these estimates on the plot obtained in part (a).
(d) Test the null hypothesis that the treatments have the same effect on patty post-cooking weight.
(e) Estimate the contrast $\tau_1 - (\tau_2 + \tau_3)/2$ which compares the effect on the post-cooked weight of the average of the two higher fat contents versus the leanest meat for the fried hamburger patties.
(f) Calculate the variance associated with the contrast in part (e). How does the value of the variance compare with the variance σ^2 of the random error variables?

15. **Trout experiment** (Gutsell 1951, Biometrics)

The data in Table 3.15 show the measurements of hemoglobin (grams per 100 ml) in the blood of brown trout. The trout were placed at random in four different troughs. The fish food added to the troughs contained, respectively, 0, 5, 10, and 15 g of sulfamerazine per 100 pounds of fish (coded 1, 2, 3, 4). The measurements were made on ten randomly selected fish from each trough after 35 days.

(a) Plot the data and comment on the results.
(b) Write down a suitable model for this experiment, assuming trough effects are negligible.
(c) Calculate the least squares estimate of the mean response for each treatment. Show these estimates on the plot obtained in (a). Can you draw any conclusions from these estimates?
(d) Test the hypothesis that sulfamerazine has no effect on the hemoglobin content of trout blood.
(e) Calculate a 95% upper confidence limit for σ^2.

16. **Trout experiment, continued**

Suppose the trout experiment of Exercise 3.15 is to be repeated with the same $v = 4$ treatments, and suppose that the same hypothesis, that the treatments have no effect on hemoglobin content, is to be tested.

(a) For calculating the number of observations needed on each treatment, what would you use as a guess for σ^2?
(b) Calculate the sample sizes needed for an analysis of variance test with $\alpha = 0.05$ to have power 0.95 if: (i) $\Delta = 1.5$; (ii) $\Delta = 1.0$; (iii) $\Delta = 2.0$.

17. **Meat cooking experiment, continued**

Suppose the meat cooking experiment of Exercise 3.14 is to be repeated with the same $v = 6$ treatments, and suppose the same hypothesis, that the treatments have the same effect on burger patty weight loss, is to be tested.

(a) Calculate an unbiased estimate of σ^2 and a 90% upper confidence limit for it.
(b) Calculate the sample sizes needed for an analysis of variance test with $\alpha = 0.05$ to have power 0.90 if:
(i) $\Delta = 5.0$; (ii) $\Delta = 10.0$.

18. The diameter of a ball bearing is to be measured using three different calipers. How many observations should be taken on each caliper type if the null hypothesis H_0:{effects of the calipers are the same} is to be tested against the alternative hypothesis that the three calipers give different average measurements. It is required to detect a difference of 0.01 mm in the effects of the caliper types with probability 0.98 and a Type I error probability of $\alpha = 0.05$. It is thought that σ is about 0.03 mm.
19. An experiment is to be run to determine whether or not time differences in performing a simple manual task are caused by different types of lighting. Five levels of lighting are selected ranging from dim colored light to bright white light. The one-way analysis of variance model (3.3.1), p. 33 is thought to be a suitable model, and $H_0 : \{\tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5\}$ is to be tested against the alternative hypothesis H_A:{the τ_i's are not all equal} at significance level 0.05. How many observations should be taken at each light level given that the experimenter wishes to reject H_0 with probability 0.90 if the difference in the effects of any two light levels produces a 4.5-second time difference in the task? It is thought that σ is at most 3.0 seconds.