17 Random Effects and Variance Components

17.1 Introduction

Until now, we have looked only at treatment factors whose levels have been specifically chosen. We have tested hypotheses about, and calculated confidence intervals for, comparisons in the effects of these particular treatment factor levels. These treatment effects are known as *fixed effects*, since we represent them in the model as unknown constants (parameters). Models that contain only fixed effects are called *fixed-effects models*.

As mentioned in step (f) of the checklist in Sect. 2.2, p. 7, there are occasions when we are interested in a large population of possible levels of a treatment factor, and the levels that are actually used in the experiment are a random sample from this population. The effects of the levels used in the experiment are then represented as random variables whose distributions are the distributions of values in the population. Such treatment-factor effects are called *random effects*, and the corresponding models are called *random-effects models*. We are not interested in just the levels that happen to be in the experiment. Rather, we are concerned with the variability of the effects of all the levels in the population. Consequently, random effects are handled somewhat differently from fixed effects. Some examples of experiments involving random effects are given in Sect. 17.2.

In Sect. 17.3 we look at experiments with a single random effect. The selection of sample sizes and model assumption checking are discussed in Sects. 17.4 and 17.5. These ideas are extended to experiments with two or more random effects in Sect. 17.6.

An experiment may involve both random and fixed effects, and the corresponding model is then known as a *mixed model*. Such experiments are discussed in Sect. 17.7. Block effects may also be random effects, and these are discussed in Sect. 17.9. Rules for obtaining confidence intervals and testing hypotheses for random effects are given in Sect. 17.8. The use of SAS and R software is considered in Sects. 17.10 and 17.11, respectively.

17.2 Some Examples

Before running an experiment, the checklist (Sect. 2.2, p. 7) should be completed as usual. In the case of a random effect, the treatment factor will have an extremely large number of levels, only a very small proportion of which can be observed in the experiment. Throughout this chapter, we will assume that the total possible number of levels of each treatment factor will be at least 100 times larger than the numbers of levels that can be observed. Typically, the population of possible levels *will* meet this

A. Dean et al., *Design and Analysis of Experiments*,
Springer Texts in Statistics, DOI 10.1007/978-3-319-52250-0_17

requirement, and for the purposes of writing down a model, we may regard the population as infinite. Otherwise, one needs to make a correction for a "finite population" in all of the formulae, and this is beyond the scope of this book. Some examples of "infinite" populations are given in Example 17.2.1.

Example 17.2.1 `Infinite populations`

Suppose that a manufacturer of canned tomato soup wishes to reduce the variability in the thickness of the soup. Suppose that the most likely causes of the variability are the quality of the cornflour (cornstarch) received from the supplier and the actions of the machine operators. Let us consider two different scenarios:

Scenario 1: The machine operators are highly skilled and have been with the company for a long time. Thus, the most likely cause of variability is the quality of the cornflour delivered to the company. The treatment factor is "cornflour," and its possible levels are all the possible batches of cornflour that the supplier could deliver. Theoretically, this is an infinite population of batches. We are interested not only in the batches of cornflour that have currently been delivered, but also in all those that might be delivered in the future. If we assume that the batches delivered to the company are a random sample from all batches that could be delivered, and if we take a random sample of delivered batches to be observed in the experiment, then the effect of the cornflour on the thickness is a random effect and can be modeled by a random variable.

Scenario 2: It is known that the quality of the cornflour is extremely consistent, so the most likely cause of variability is due to the different actions of the machine operators. The company is large and machine operators change quite frequently. Consequently, those that are available to take part in the experiment are only a small sample of all operators employed by the company at present or that might be employed in the future. If we can assume that the operators available for the experiment are representative of the population, then we can assume that they are similar to a random sample from a very large population of possible operators, present and future. Since we would like to know about the variability of the entire population, we model the effect of the operators as random variables, and call them random effects. □

In the absence of any blocking factors, a completely randomized design would be used. The levels of the random-effects treatment factor are first selected at random from the population of all possible levels, and then the experimental units are randomly assigned to these selected levels as usual. At step (h) of the checklist, we need to calculate the number of levels v of the treatment factor to be observed in the experiment in addition to r, the number of observations on each level. Since this calculation uses the formulas for confidence intervals and hypothesis tests, we will postpone the discussion to Sect. 17.4. As a general rule, if the variability of the treatment effects is much greater than the error (measurement) variability, then v should be large and r small; and vice versa.

Example 17.2.2 `Clean wool experiment`

The clean wool experiment was reported by J.M. Cameron, of the National Bureau of Standards, in the 1951 volume of *Biometrics*. The following checklist has been compiled from the information given in the article.

(a) **Define the objectives of the experiment**.

Raw wool contains varying amounts of grease, dirt, and foreign material which must be removed before manufacturing begins. The purchase price and customs levy of a shipment are based on the actual amount of wool present, i.e., on the amount of wool present after thorough cleaning—the "clean content." The clean content is expressed as the percentage the weight of the clean wool is of the original weight of the raw wool.

The experiment was run in order to estimate the variability in "clean content" of bales of wool in a shipment.

(b) **Identify all sources of variation**.

(i) Treatment factors and their levels.
The treatment factor was "wool bale" and its levels were the entire population of bales in a particular shipment. Seven bales were observed in the experiment, and these were selected at random from the shipment. The shipment was large enough to allow the bales used in the experiment to be regarded as a random sample from an infinite population of bales. The treatment factor "wool bale" was therefore regarded as a random effect.

(ii) Experimental units.
The experimental units were time slots, so that allocation of these to the levels of the treatment factor determined the order in which the wool bales were observed.

(iii) Blocking factors, noise factors, and covariates.
No nuisance factors were identified as major sources of variation.

(c) **Choose a rule by which to assign the experimental units to the levels of the treatment factors**.

A completely randomized design was selected.

(d) **Specify the measurements to be made, the experimental procedure, and the anticipated difficulties**.

A machine was used to bore through a bale of wool and extract a core of wool. Several cores were taken from each of the seven selected bales so that several observations on clean content could be made on each bale. Each core of wool was weighed and then cleaned by scouring, removing burrs, etc. After cleaning, the wool was reweighed and the clean content calculated as the ratio of the clean wool to the initial weight, times 100%.
An anticipated difficulty was that the scouring process, which works well with large amounts of wool, proves difficult with a small core of wool, so that the experimental error observed in the experiment may be larger than would normally be observed in routine production.

The observations on the clean content of the seven bales are shown in Table 17.1. Model selection for this experiment and its analysis via SAS and R software are discussed in Sects. 17.10 and 17.11, respectively. □

Table 17.1 Data for the clean wool experiment

	Bale						
	1	2	3	4	5	6	7
Clean content	52.33	56.99	54.64	54.90	59.89	57.76	60.27
	56.26	58.69	57.48	60.08	57.76	59.68	60.30
	62.86	58.20	59.29	58.72	60.26	59.58	61.09
	50.46	57.35	57.51	55.61	57.53	58.08	61.45

Source Cameron (1951).

17.3 One Random Effect

17.3.1 The Random-Effects One-Way Model

For a completely randomized design, with v randomly selected levels of a treatment factor T, the random-effects one-way model is

$$\begin{gathered} Y_{it} = \mu + T_i + \epsilon_{it} \,, \\ \epsilon_{it} \sim N(0, \sigma^2)\,, \quad T_i \sim N(0, \sigma_T^2)\,, \\ \epsilon_{it}\text{'s and } T_i\text{'s are all mutually independent}\,, \\ t = 1, \ldots, r_i, \quad i = 1, \ldots, v\,. \end{gathered} \tag{17.3.1}$$

Compare this with the fixed-effects one-way analysis of variance model (3.3.1), p. 33. The form of the model and the error assumptions are exactly the same. The only difference is in the modeling of the treatment effect. Since the ith level of the treatment factor T observed in the experiment has been randomly selected from the "infinite" population, its observed effect is an observation of a random variable T_i. The distribution of T_i is the distribution of treatment effects in the whole population. We have assumed in (17.3.1) that the population of effects follows a normal distribution with variance σ_T^2, and this assumption will need to be checked along with the error assumptions. The mean of the treatment-effect population has been absorbed into the constant μ, so the distribution of T_i is listed as $N(0, \sigma_T^2)$. The variance σ_T^2 is the parameter of interest, since if the effects of all of the treatment-factor levels are the same, then σ_T^2 is zero. If the effects of the levels are quite different, then σ_T^2 is quite large.

Our final assumption is one of independence. If the treatment-factor levels are selected at random, then the assumption of independence of $T_1, T_2, \ldots, T_v$ is reasonable. However, if, as in Example 17.2.1 Scenario 2, the levels are a "convenient sample," then this assumption should be investigated carefully. Independence of the T_i and ϵ_{it} requires that the treatment factor not affect any source of variation that has been absorbed into the error variable.

In a random-effects model, the expected value of the response is μ, since

$$E[Y_{it}] = E[\mu] + E[T_i] + E[\epsilon_{it}] = \mu\,.$$

The variance of Y_{it} is

$$\text{Var}(Y_{it}) = \text{Var}(\mu + T_i + \epsilon_{it}) = \text{Var}(T_i) + \text{Var}(\epsilon_{it}) + 2\text{Cov}(T_i, \epsilon_{it}) = \sigma_T^2 + \sigma^2\,,$$

since T_i and ϵ_{it} are mutually independent and so have zero covariance. Therefore, the distribution of Y_{it} is

$$Y_{it} \sim N(\mu,\ \sigma_T^2 + \sigma^2). \tag{17.3.2}$$

The two components σ_T^2 and σ^2 of the variance of Y_{it} are known as *variance components*. Observations on the same treatment are correlated, with

$$\text{Cov}(Y_{it}, Y_{is}) = \text{Cov}(\mu + T_i + \epsilon_{it}, \mu + T_i + \epsilon_{is}) = \text{Var}(T_i) = \sigma_T^2\,.$$

17.3.2 Estimation of σ^2

In order to be able to test hypotheses about σ_T^2 or to calculate confidence intervals, we need an unbiased estimate of σ^2. The random-effects one-way model (17.3.1) is very similar to the fixed effects one-way analysis of variance model (3.3.1), p. 33, so a natural question is whether the fixed-effects mean square for error MSE provides an unbiased estimator for σ^2 in the random-effects model also. The answer, happily, is "yes," and we can check it by calculating $E[MSE]$ for the random-effects model, as shown below.

From (3.4.6), p. 39, the fixed-effects sum of squares for error is

$$SSE = \sum_{i=1}^{v}\sum_{t=1}^{r_i} Y_{it}^2 - \sum_{i=1}^{v} r_i \overline{Y}_{i.}^2 \,.$$

Remember that the variance of a random variable X is calculated as $\text{Var}(X) = E[X^2] - (E[X])^2$. So, we have

$$E[Y_{it}^2] = \text{Var}(Y_{it}) + (E[Y_{it}])^2 = (\sigma_T^2 + \sigma^2) + \mu^2 \,.$$

Now,

$$\overline{Y}_{i.} = \mu + T_i + \frac{1}{r_i}\sum_{t=1}^{r_i} \epsilon_{it} \,,$$

so

$$\text{Var}(\overline{Y}_{i.}) = \sigma_T^2 + \frac{\sigma^2}{r_i} \quad \text{and} \quad E[\overline{Y}_{i.}] = \mu \,. \tag{17.3.3}$$

Consequently,

$$E[\overline{Y}_{i.}^2] = \left(\sigma_T^2 + \frac{\sigma^2}{r_i}\right) + \mu^2.$$

Thus,

$$\begin{aligned} E[SSE] &= \sum_{i=1}^{v}\sum_{t=1}^{r_i} (\sigma_T^2 + \sigma^2 + \mu^2) - \sum_{i=1}^{v} r_i \left(\sigma_T^2 + \frac{\sigma^2}{r_i} + \mu^2\right) \\ &= n\sigma^2 - v\sigma^2 \quad \left(\text{where } n = \sum_{i=1}^{v} r_i\right) \\ &= (n - v)\sigma^2 \,, \end{aligned}$$

giving

$$E[MSE] = E[SSE/(n - v)] = \sigma^2.$$

So MSE is an unbiased estimator for σ^2, and the observed value of the mean square for error, msE, is an unbiased estimate for σ^2 in the random-effects one-way model, as well as in the fixed-effects one-way model.

Confidence bounds for σ^2 can be computed as under fixed-effects models (Sect. 3.4.6), that is,

$$\sigma^2 \le \frac{ssE}{\chi^2_{n-v,1-\alpha}}\,, \tag{17.3.4}$$

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.

17.3.3 Estimation of σ_T^2

Since the fixed-effects mean square for error msE provides an unbiased estimate of σ^2, the next question that is natural to ask is whether the fixed-effects mean square for treatments msT provides an unbiased estimate for σ_T^2. The answer is "not quite," but we can certainly use it to find an estimate. Now $msT = ssT/(v-1)$, and ssT is given in (3.5.11), p. 43, as

$$ssT = \sum_{i=1}^{v} r_i \overline{y}_{i.}^2 - n\overline{y}_{..}^2$$

with corresponding random variable

$$SST = \sum_{i=1}^{v} r_i \overline{Y}_{i.}^2 - n\overline{Y}_{..}^2\,.$$

Using the same type of calculation as in Sect. 17.3.2 above, we have

$$\overline{Y}_{..} = \mu \;+\; \frac{1}{n}\sum_i r_i T_i \;+\; \frac{1}{n}\sum_{i=1}^{v}\sum_{t=1}^{r_i}\epsilon_{it}\,.$$

So

$$E[\overline{Y}_{..}] = \mu \quad\text{and}\quad \text{Var}(\overline{Y}_{..}) = \frac{\sum r_i^2}{n^2}\sigma_T^2 + \frac{n}{n^2}\sigma^2\,.$$

Also, from (17.3.3),

$$E[\overline{Y}_{i.}] = \mu \quad\text{and}\quad \text{Var}(\overline{Y}_{i.}) = \sigma_T^2 + \frac{\sigma^2}{r_i}.$$

Therefore,

$$\begin{aligned} E[SST] &= \sum_{i=1}^{v} r_i\left(\sigma_T^2 + \frac{\sigma^2}{r_i} + \mu^2\right) - n\left(\frac{\sum r_i^2}{n^2}\sigma_T^2 + \frac{\sigma^2}{n} + \mu^2\right) \\ &= \left(n - \frac{\sum r_i^2}{n}\right)\sigma_T^2 + (v-1)\sigma^2\,. \end{aligned}$$

Since $MST = SST/(v-1)$, we have

$$E[MST] = c\sigma_T^2 + \sigma^2, \text{ where } c = \frac{n^2 - \sum r_i^2}{n(v-1)}.$$

Notice that if all r_i are equal to r, then $n = vr$ and $c = r$.

We see that MST is an unbiased estimator of $c\sigma_T^2 + \sigma^2$, not σ_T^2. Nevertheless, we can easily find an unbiased estimator of σ_T^2, since

$$E\left[\frac{MST - MSE}{c}\right] = \sigma_T^2 . \tag{17.3.5}$$

It is, unfortunately, possible for the observed value of this estimator to be negative even though σ_T^2 cannot be negative. This will occur when msE happens to be greater than msT, and this is most likely when σ_T^2 is close to zero. If msE is considerably greater than msT, then the model should be questioned, as it is unlikely to be a good description of the data.

Example 17.3.1 `Ice cream experiment`

The following experiment was run by Sue Hubbard in 1986 to determine whether or not different flavors of ice cream melt at different speeds. A random sample of three flavors was selected from a large population of flavors offered to the customer by a single manufacturer in May 1986. It is not obvious that the selected flavors are representative of all possible ice cream flavors, since some may include an ingredient that inhibits melting. The theoretical population is therefore the population of all flavors that could be made with ingredients similar to those flavors available.

The three flavors of ice cream were stored in the same freezer in similar-sized containers. For each observation, one teaspoonful of ice cream was taken from the freezer, transferred to a plate, and the melting time at room temperature was observed to the nearest second. Eleven observations were taken on each flavor. These are shown, together with their order of observation, in Table 17.2 and plotted in Fig. 17.1.

Now,

$$\begin{aligned} ssE &= \sum\sum y_{it}^2 - 11\sum \overline{y}_{i.}^2 \\ &= 30{,}206{,}485 - 30{,}003{,}028.8181 \\ &= 203{,}456.1819 . \end{aligned}$$

So an unbiased estimate of σ^2 is

$$msE = ssE/(33-3) = 6781.8727 \text{ seconds}^2 .$$

Table 17.2 Melting times for three randomly selected flavors of ice cream. Order of observation in parentheses

Flavor	Time in seconds (order of observation)					
1	924 (1)	876 (2)	1150 (5)	1053 (7)	1041 (10)	1037 (12)
	1125 (15)	1075 (16)	1066 (20)	977 (22)	886 (25)	
2	891 (3)	982 (4)	1041 (8)	1135 (13)	1019 (14)	1093 (18)
	994 (27)	960 (30)	889 (31)	967 (32)	838 (33)	
3	817 (6)	1032 (9)	844 (11)	841 (17)	785 (19)	823 (21)
	846 (23)	840 (24)	848 (26)	848 (28)	832 (29)	

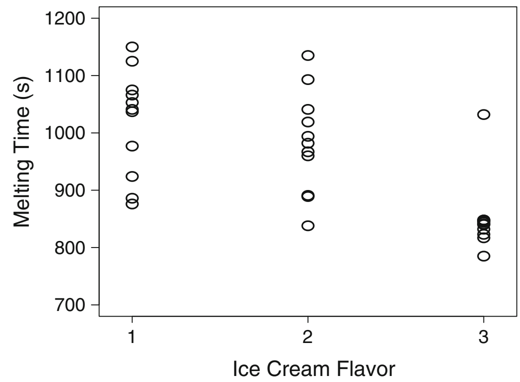

Fig. 17.1 Plot of data for the ice cream experiment

Similarly,

$$\begin{aligned} ssT &= 11 \sum \overline{y}_{i.}^2 - 33\overline{y}_{..}^2 \\ &= 30{,}003{,}028.8181 - 29{,}830{,}018.9393 \\ &= 173{,}009.8787\,. \end{aligned}$$

So $msT = ssT/(3-1) = 86{,}504.9394$ seconds^2, and an unbiased estimate of σ_T^2 is given by

$$\begin{aligned} \frac{msT - msE}{c} &= \frac{86{,}504.9394 - 6781.8727}{11} \\ &= 7247.5515 \text{ seconds}^2. \end{aligned}$$

□

17.3.4 Testing Equality of Treatment Effects

When the treatment factor has random effects, we are interested in the variability of the treatment effects in the entire population of levels, not just those in the experiment. Since the variance of the effects in the population is σ_T^2, the null hypothesis of interest is of the form

$$H_0^T : \sigma_T^2 = 0\,,$$

and the alternative hypothesis is

$$H_A^T : \sigma_T^2 > 0.$$

It would be very convenient if we could use the same hypothesis-testing rule as we used for testing equality of treatment effects in the fixed-effects model. The fixed-effects decision rule was to reject the hypothesis of no difference in the treatments if $msT/msE > F_{v-1,n-v,\alpha}$, (see (3.5.15), p. 43). Let us examine the ratio msT/msE in the random-effects one-way model (17.3.1). In Sect. 17.3.2 we showed that

$$E[MSE] = \sigma^2\,,$$

and in Sect. 17.3.3 we showed that

$$E[MST] = c\sigma_T^2 + \sigma^2 ,$$

where $c = (n^2 - \sum r_i^2)/n(v-1)$, and if all r_i are equal to r, then $c = r$.

So, if $H_0^T : \sigma_T^2 = 0$ is true, then the expected value of the numerator of the ratio *MST/MSE* is equal to σ^2, the same as the expected value of the denominator. Then, if H_0^T is true, the ratio should be in the region of 1.0. But if σ_T^2 is large, the expected value of the numerator is larger than the denominator, and the ratio should be large and positive. This situation is similar to that for the fixed-effects case. The only remaining question is whether *MST/MSE* has an F distribution with $v-1$ and $n-v$ degrees of freedom when H_0^T is true.

It can be shown that

$$SST/(c\sigma_T^2 + \sigma^2) \sim \chi^2_{v-1} \tag{17.3.6}$$

and

$$SSE/\sigma^2 \sim \chi^2_{n-v}$$

and that *SST* and *SSE* are independent. Consequently, we have

$$\frac{SST/((c\sigma_T^2+\sigma^2)(v-1))}{SSE/(\sigma^2(n-v))} = \frac{MST/(c\sigma_T^2+\sigma^2)}{MSE/\sigma^2} \sim \frac{\chi^2_{v-1}/(v-1)}{\chi^2_{n-v}/(n-v)} \sim F_{v-1,n-v} , \tag{17.3.7}$$

and when $\sigma_T^2 = 0$, then

$$\frac{MST}{MSE} \sim F_{v-1,n-v} .$$

Thus, to test $H_0 : \sigma_T^2 = 0$ against $H_A : \sigma_T^2 > 0$, our decision rule is to

$$\text{reject } H_0^T \text{ if } \frac{msT}{msE} > F_{v-1,n-v,\alpha} \tag{17.3.8}$$

for some chosen value of the significance level α. The test can be set out in an analysis of variance table in the usual way; see Table 17.3. We have included the expected mean squares in the table for easy reference.

Rather than testing whether or not the variance of the population of treatment effects is zero, it may be of more interest to test whether the variance is less than or equal to some proportion of the error

Table 17.3 Analysis of variance table for the random-effects one-way model

Source of variation	Degrees of freedom	Sum of squares	Mean squares	Ratio	Expected mean square
Treatments	$v-1$	ssT	$\frac{ssT}{v-1}$	$\frac{msT}{msE}$	$c\sigma_T^2+\sigma^2$
Error	$n-v$	ssE	$\frac{ssE}{n-v}$		σ^2
Total	$n-1$	$sstot$			

Computational formulae

$ssT = \sum_i r_i\overline{y}_{i.}^2 - n\overline{y}_{..}^2$	$ssE = \sum_i \sum_t y_{it}^2 - \sum_i r_i\overline{y}_{i.}^2$
$sstot = \sum_i \sum_t y_{it}^2 - n\overline{y}_{..}^2$	$c = \frac{n^2 - \sum r_i^2}{n(v-1)}$

variance, that is,

$$H_0^{\gamma T} : \sigma_T^2 \leq \gamma\sigma^2 \text{ and } H_A^{\gamma T} : \sigma_T^2 > \gamma\sigma^2,$$

for some constant γ. From (17.3.7), we see that if $H_0^{\gamma T}$ is true with $\sigma_T^2 = \gamma\sigma^2$, then

$$\frac{MST/(\sigma^2(c\sigma_T^2/\sigma^2 + 1))}{MSE/\sigma^2} = \frac{MST}{MSE(c\gamma + 1)} \sim F_{v-1,n-v}.$$

So, our decision rule (17.3.8) needs only the minor modification of including the constant $(c\gamma + 1)$, that is,

$$\text{reject } H_0^{\gamma T} \text{ if } \frac{msT}{msE} > (c\gamma + 1)F_{v-1,n-v,\alpha}\,. \tag{17.3.9}$$

If we choose $\gamma = 0$, then the decision rule (17.3.9) reduces to rule (17.3.8) for testing the null hypothesis $H_0^T : \sigma_T^2 = 0$ against its alternative hypothesis $H_A^T : \sigma_T^2 > 0$.

Example 17.3.2 `Ice cream experiment, continued`

The analysis of variance table for the ice cream experiment of Example 17.3.1 is shown in Table 17.4. If we test the null hypothesis that the variance of melting times in the population of ice creams is negligible against the alternative hypothesis that it is not (that is, $H_0^T : \sigma_T^2 = 0$ versus $H_A^T : \sigma_T^2 > 0$) with a Type I error probability of $\alpha = 0.05$, we would reject H_0^T, since

$$msT/msE = 12.76 \; > \; F_{2,30,0.05} = 3.32\,,$$

or equivalently, the p-value is less than 0.05.

In such an experiment there will clearly be considerable error variability in the data due to fluctuations of room temperature and the difficulty of determining the exact time at which the ice cream has melted completely. Variability in the melting time of different flavors is unlikely to be of interest to the experimenter unless it is larger than the error variability. Suppose, therefore, instead of testing the hypothesis H_0^T against H_A^T, we test the null hypothesis $H_0^{\gamma T} : \{\sigma_T^2 \leq \sigma^2\}$ against $H_A^{\gamma T} : \{\sigma_T^2 > \sigma^2\}$. Since there are $r = 11$ observations on each ice cream, the constant is $c = 11$, and the hypothesis-testing rule (17.3.9) with $\gamma = 1.0$ becomes

$$\text{reject } H_0^{\gamma T} \text{ if } \frac{msT}{msE} > (11 + 1)F_{2,30,\alpha}\,,$$

that is,

$$\text{reject } H_0^{\gamma T} \text{ if } 12.76 > 12\; F_{2,30,\alpha}.$$

Table 17.4 Analysis of variance table for the ice cream experiment

Source of variation	Degrees of freedom	Sum of squares	Mean squares	Ratio	p-value
Flavor	2	173009.8788	86504.9394	12.76	0.0001
Error	30	203456.1818	6781.8727		
Total	32	376466.0606			

It can be seen from the table in Appendix A.6 that for any practical choice of α, there is not sufficient evidence to reject the null hypothesis. Thus, although the variation in the melting times of the different flavors is significant, so apparently $\sigma_T^2 > 0$, sufficient evidence has not been gathered to be able to claim that the variation is significantly larger than the error variation in the data. □

17.3.5 Confidence Intervals for Variance Components

We showed in Sect. 17.3.1, p. 618, that the response variable Y_{it} in a random-effects one-way model (17.3.1) has a normal distribution with variance $\sigma^2 + \sigma_T^2$, where σ^2 is the variance of the error variables and σ_T^2 is the variance of the treatment effects in the population. In order to assess the variability of the treatment-effect population, we may wish to calculate a confidence interval for σ_T^2 or, alternatively, for σ_T^2/σ^2 if we want to assess the treatment variability relative to the error variability. Since the latter is the easier calculation, we investigate this first.

Confidence Intervals for σ_T^2/σ^2

From (17.3.7), p. 623, we know that

$$\frac{MST}{MSE(c\sigma_T^2/\sigma^2 + 1)} \sim F_{v-1,n-v}\,, \tag{17.3.10}$$

where $c = (n^2 - \Sigma r_i^2)/(n(v-1))$, and if the r_i are all equal to r, then $c = r$. From this, we can write down an interval in which *MST*/*MSE* lies with probability $1-\alpha$; that is,

$$P\left(F_{v-1,n-v,1-\alpha/2} \le \frac{MST}{MSE(c\sigma_T^2/\sigma^2 + 1)} \le F_{v-1,n-v,\alpha/2}\right) = 1-\alpha.$$

If we rearrange the left-hand inequality, we find that

$$c\sigma_T^2/\sigma^2 \le \frac{MST}{MSE\,F_{v-1,n-v,1-\alpha/2}} - 1\,,$$

and similarly for the right-hand inequality,

$$c\sigma_T^2/\sigma^2 \ge \frac{MST}{MSE\,F_{v-1,n-v,\alpha/2}} - 1\,.$$

So, replacing the random variables by their observed values, we obtain a $100(1-\alpha)\%$ confidence interval for σ_T^2/σ^2 as

$$\frac{1}{c}\left[\frac{msT}{msE\,F_{v-1,n-v,\alpha/2}} - 1\right] \le \frac{\sigma_T^2}{\sigma^2} \le \frac{1}{c}\left[\frac{msT}{msE\,F_{v-1,n-v,1-\alpha/2}} - 1\right]. \tag{17.3.11}$$

A drawback of this interval is that if *msT* is not much larger than *msE* (or perhaps smaller), then it is possible for the left-hand end of the interval to be negative even though σ_T^2/σ^2 can never be negative. Although we could replace a negative lower bound by zero, we will not do so, since it can result in a short interval, giving the misleading impression that the experiment was more accurate than it actually was.

For calculation of the interval, remember that $F_{v-1,n-v,\alpha/2}$ denotes the percentile of the $F_{v-1,n-v}$ distribution corresponding to a probability of $\alpha/2$ in the right-hand tail. Also, $F_{v-1,n-v,1-\alpha/2}$ denotes the percentile corresponding to a probability of $\alpha/2$ in the left-hand tail, that is, $1-\alpha/2$ in the right-hand tail. Since $F_{v-1,n-v,1-\alpha/2}$ is not tabulated in Appendix A.6, it is important to note that

$$F_{v-1,n-v,1-\alpha/2} = (F_{n-v,v-1,\alpha/2})^{-1}. \tag{17.3.12}$$

Example 17.3.3 `Ice cream experiment, continued`

In the ice cream experiment of Examples 17.3.1 and 17.3.2, pp. 621 and 624, the variance σ_T^2 in the melting times (in seconds) of the population of different flavors of ice cream is substantially greater than zero but not substantially greater than the error variance σ^2. A confidence interval for σ_T^2/σ^2 can be obtained using (17.3.11). The values $msT = 86504.9394$, $msE = 6781.8727$, $v = 3$, $c = r = 11$, and $n = 33$ are obtained from Example 17.3.2. From the table in Appendix A.6, we have

$$F_{2,30,.05} = 3.32 \quad \text{and} \quad F_{2,30,.95} = (F_{30,2,.05})^{-1} = (19.5)^{-1} = 0.0513\,.$$

Therefore, the confidence interval (17.3.11) becomes

$$\frac{1}{11}\left(\frac{86504.9394}{(6781.8727)(3.32)} - 1\right) \le \frac{\sigma_T^2}{\sigma^2} \le \frac{1}{11}\left(\frac{86504.9394}{(6781.8727)(0.0513)} - 1\right),$$

that is,

$$\sigma_T^2/\sigma^2 \in (0.258, 22.513)\,.$$

This interval is too wide to be of much practical use, since it says that with 95% confidence, σ_T^2 could be 4 times smaller or as much as 22 times bigger than σ^2. However, the result does agree with the test of the null hypothesis $H_0^{\gamma T}$ in Example 17.3.2, since the interval includes the value $\sigma_T^2/\sigma^2 = 1.0$. □

As can be seen from Example 17.3.3, a confidence interval for σ_T^2/σ^2 can be very wide. Not only do we need sufficient numbers of observations on each treatment in the experiment in order to keep a confidence interval narrow, but we also need a sufficiently large selection of treatments to represent the population. In Example 17.3.3, there were only $v = 3$ treatments to represent an entire population of ice cream flavors, and this has contributed to the lack of precision in the experiment. Calculation of sample sizes will be discussed in Sect. 17.4.

Confidence Intervals for σ_T^2

There are various methods of obtaining approximate $100(1-\alpha)\%$ confidence intervals for σ_T^2. The only method that we shall give here is one that is useful when σ_T^2 is not close to zero and that can be easily adapted when we have more complicated models.

First, remember that an unbiased estimator for σ_T^2 was obtained in Eq. (17.3.5), p. 621, as

$$U = c^{-1}(MST - MSE)\,, \tag{17.3.13}$$

where $c = (n^2 - \Sigma r_i^2)/(n(v-1))$, and $c = r$ when the sample sizes are equal. If we can determine the distribution of U, then we can easily find a confidence interval for σ_T^2. We know that for the random-effects one-way model, $SST/(c\sigma_T^2 + \sigma^2) \sim \chi^2_{v-1}$ and $SSE/\sigma^2 \sim \chi^2_{n-v}$ and that SST and SSE are independent. The exact distribution of U is therefore based on the difference of two chi-squared distributions each multiplied by a constant of unknown value, and this is not a standard tabulated distribution. However, it can be shown that a reasonable approximation to the true distribution of U/σ_T^2 is a chi-squared distribution divided by its degrees of freedom x, where x is estimated by

$$x = \frac{(msT - msE)^2}{msT^2/(v-1) + msE^2/(n-v)}. \tag{17.3.14}$$

In other words, the distribution of $xU/E[U]$ is approximately χ^2_x. This result is related to the Satterthwaite approximation that we used in Sect. 5.6.3, p. 115 (Scheffé 1959, Sect. 7.5, gives the general result). Using this approximation, we can write down the approximate probability statement

$$P\left(\chi^2_{x,1-\alpha/2} \le \frac{xU}{\sigma_T^2} \le \chi^2_{x,\alpha/2}\right) \approx 1 - \alpha\,.$$

If we rearrange the left-hand inequality, we obtain

$$\sigma_T^2 \le \frac{xU}{\chi^2_{x,1-\alpha/2}}\,,$$

and if we rearrange the right-hand inequality, we obtain

$$\frac{xU}{\chi^2_{x,\alpha/2}} \le \sigma_T^2\,.$$

Consequently, we obtain an approximate $100(1-\alpha)\%$ confidence interval for σ_T^2 as

$$\frac{xu}{\chi^2_{x,\alpha/2}} \le \sigma_T^2 \le \frac{xu}{\chi^2_{x,1-\alpha/2}}\,, \tag{17.3.15}$$

where u is the observed value of U; that is,

$$u = c^{-1}(msT - msE)\,. \tag{17.3.16}$$

Example 17.3.4 `Ice cream experiment, continued`

Suppose we require a 90% confidence interval for the variance of the melting times of the population of ice creams in the ice cream experiment of Examples 17.3.1 and 17.3.2, pp. 621 and 624. Using the information in those examples, we obtain the unbiased estimate (17.3.16) of σ_T^2 as $u = 7247.5526$ seconds^2. The degrees of freedom x of the approximate distribution of U are calculated using (17.3.14), that is,

$$x = \frac{(86504.9394 - 6781.8727)^2}{(86504.9394)^2/2 - (6781.8727)^2/30} \approx 1.7.$$

From Table A.5 we can guess at the approximate values of $\chi^2_{x,.05}$ and $\chi^2_{x,.95}$ as

$$\chi^2_{1.7,.05} \approx 5.3 \quad \text{and} \quad \chi^2_{1.7,.95} \approx 0.07\,.$$

So a 90% confidence interval for σ_T^2 is roughly

$$\sigma_T^2 \in \left(\frac{(1.7)(7247.5515)}{5.3}, \quad \frac{(1.7)(7247.55)}{0.07}\right)$$
$$= (2{,}324.69, \quad 176{,}011.97)$$

Taking square roots and converting to minutes, we obtain the approximate 90% confidence interval for the standard deviation of melting times as

$$\sigma_T \in (0.8,\ 7.0) \text{ minutes.}$$

Again, this interval is too wide for practical use, due to the small number of flavors examined from the population. □

17.4 Sample Sizes for an Experiment with One Random Effect

For the fixed-effects one-way analysis of variance model, we looked at two different ways of determining sample sizes. The first method (Sect. 3.6) was based on the required power of the hypothesis test for detecting whether two treatment effects differ by more than a chosen quantity Δ. The second method (Sect. 4.5) was based on the required length of confidence intervals for one or more treatment contrasts.

For the random-effects one-way model, we need to determine both the number v of levels of the treatment factor to be observed in the experiment and the number r of observations to be taken on each of these levels. A glance at the formulae (17.3.11) and (17.3.15) shows that a calculation of v and r based on the lengths of confidence intervals will not be straightforward. Both formulae depend not only on the value of *msE*, but also on that of *msT*, both of which are unknown prior to the experiment. However, consideration of the variances of the estimators used to develop the confidence intervals helps us determine an appropriate balance between "more treatments" and "more replication."

Consider first the confidence interval for σ_T^2 given in (17.3.15). The confidence interval should be tight if the variance of the unbiased estimator U is small. Assuming equal sample sizes, $U = r^{-1}(MST - MSE)$ has variance

$$\text{Var}(U) = \frac{2n^2(n\sigma_T^2/v + \sigma^2)^2}{v^2(v-1)} + \frac{2n^2\sigma^4}{v^2(n-v)}$$

for $n > v$. (This follows because *MST* and *MSE* are independent, $SST/(r\sigma_T^2 + \sigma^2) \sim \chi^2(v-1)$, $SSE/\sigma^2 \sim \chi^2(n-v)$, and the variance of a chi-squared random variable is twice its degrees of freedom.) We want this variance to be as small as possible. Suppose that the total number of observations $n = rv$ is fixed by budget considerations and we require $r \geq 2$ since replication is needed to estimate σ^2, so $v \leq n/2$. To minimize the first term of this variance, we require v as large as possible, corresponding to $v = n/2$. It can be shown that the second term is minimized by taking $v = 2n/3$, or by taking $v = n/2$ if we require equal sample sizes and $r \geq 2$.

In summary, assuming equal replication with $r \geq 2$, the variance of U is minimized by taking $v = n/2$ and $r = 2$. In this case, our estimator would be $U = (1/2)(MST - MSE)$, for which $\text{Var}(U) = 8[(2\sigma_T^4 + \sigma^2)^2/(v-1) + \sigma^4/v]$ is as small as possible given equal replication with $r \geq 2$.

We find a similar requirement resulting from a confidence interval for σ_T^2/σ^2. The mean of an F-distribution with $v-1$ and $n-v$ degrees of freedom is $(n-v)/(n-v-2)$. So, from (17.3.10),

p. 625, with $c = r$, an unbiased estimator of σ_T^2/σ^2 is given by

$$U = \frac{1}{r}\left[\frac{(n-v-2)}{(n-v)}\frac{MST}{MSE} - 1\right],$$

and a narrow confidence interval should be obtained if we choose v and r to make Var(U) small. The variance of an F-distribution with m and p degrees of freedom is

$$\frac{2p^2(m+p-2)}{m(p-2)^2(p-4)}.$$

It follows from (17.3.10), the definition of U, and $m = v - 1$ and $p = n - v$ that when the sample sizes are all equal to r,

$$\begin{aligned}\text{Var}(U) &= \left(r\frac{\sigma_T^2}{\sigma^2} + 1\right)^2 \frac{1}{r^2}\left(\frac{2(n-v)^2(n-3)(n-v-2)^2}{(v-1)(n-v-2)^2(n-v-4)(n-v)^2}\right) \\ &= \left(\frac{\sigma_T^2}{\sigma^2} + \frac{1}{r}\right)^2\left(\frac{2(n-3)}{(v-1)(n-v-4)}\right).\end{aligned}$$

So, if the number of observations n is fixed with $n = rv$ and $r \geq 2$, and if we expect that $\sigma_T^2/\sigma^2 > 1/2$, say, then the squared term $(\sigma_T^2/\sigma^2)^2$ from the first set of parentheses will be most important for determining the size of the variance—more important that the term $(1/r)^2$ or the cross-product term—and we need to minimize its coefficient, which is $2(n-3)/((v-1)(n-v-4))$. This requires that $v = (n-3)/2$, or $v = n/2$ and $r = 2$ in the equireplicate case. On the other hand, in the more unusual case when σ_T^2 is expected to be much smaller than σ^2, then the squared term $(1/r)^2$ from the first set of parentheses will be most important for determining the size of the variance, and we need the minimum value of

$$\frac{1}{r^2}\left(\frac{2(n-3)}{(v-1)(n-v-4)}\right) = \frac{2v^2(n-3)}{n^2(v-1)(n-v-4)},$$

and this occurs when v is as small as possible. However, it is unusual to be interested in σ_T^2/σ^2 if this ratio is expected to be very small, so we discount this case.

In summary, the general recommendation again is to set $v = n/2$ and $r = 2$. The exception is the extraordinary case where one plans such an experiment to study σ_T^2/σ^2 but expects σ_T^2 to be much smaller than σ^2, in which case one should choose v to be small.

We can get a feel for how many observations $n = rv$ are needed in total if we examine the power of the hypothesis test for testing $H_0^{\gamma T} : \sigma_T^2 \leq \gamma\sigma^2$ against the alternative hypothesis $H_A^{\gamma T} : \sigma_T^2 > \gamma\sigma^2$ (for a chosen $\gamma \geq 0$). The decision rule was given in (17.3.9), p. 624, as

$$\text{reject } H_0^{\gamma T} \text{ if } \frac{msT}{msE} > (c\gamma + 1)F_{v-1,n-v,\alpha} = k, \text{ say.} \tag{17.4.17}$$

What is the probability of rejecting $H_0^{\gamma T}$ if the true value of σ_T^2/σ^2 is Δ? In other words, what is the probability that $MST/MSE > k$, when σ_T^2/σ^2 is equal to Δ? This is the power of the test at the value Δ. We can calculate the power from the knowledge that

$$\frac{MST}{MSE(c\sigma_T^2/\sigma^2+1)} \sim F_{v-1,n-v}\,,$$

see (17.3.12), p. 626. If σ_T^2/σ^2 is equal to Δ, then

$$P\left(\frac{MST}{MSE} > k\right) = P\left(\frac{MST}{MSE(c\Delta+1)} > \frac{k}{c\Delta+1}\right).$$

Suppose we stipulate that the power must be π when $\sigma_T^2/\sigma^2 = \Delta$. Then, we must have that

$$\frac{k}{c\Delta+1} = F_{v-1,n-v,\pi}.$$

Remembering from (17.4.17) that $k = (c\gamma+1)F_{v-1,n-v,\alpha}$, and that $(n-v) = v(r-1)$ and $c = r$ for equal sample sizes, we obtain the equality

$$\frac{F_{v-1,v(r-1),\alpha}}{F_{v-1,v(r-1),\pi}} = \frac{r\Delta+1}{r\gamma+1}\,.$$

So, we need to select γ and α for testing $H_0^{\gamma T}$ together with Δ and π. Then we can try to determine v and r by trial and error as illustrated in Example 17.4.1. Since $F_{v-1,v(r-1),\pi} = (F_{v(r-1),v-1,1-\pi})^{-1}$, we try to find values of v and r such that

$$(F_{v-1,v(r-1),\alpha})(F_{v(r-1),v-1,1-\pi}) \le \frac{r\Delta+1}{r\gamma+1}\,. \tag{17.4.18}$$

Example 17.4.1 `Ice cream experiment, continued`

In Example 17.3.2, p. 624, we were unable to reject the hypothesis $H_0^{\gamma T}: \sigma_T^2 \le \sigma^2$ in favor of the hypothesis $H_A^{\gamma T}: \sigma_T^2 > \sigma^2$ at a significance level of $\alpha = 0.05$. Suppose we wish to repeat this experiment, still with $\gamma = 1.0$ and a Type I error probability of $\alpha = 0.05$. Suppose further that we would like to reject the hypothesis with high probability (say $\pi = 0.95$) if the true value of σ_T^2/σ^2 is at least $\Delta = 2.0$. How many ice cream flavors should we look at and how many observations should we take on each?

From (17.4.18), we need to find v and r such that

$$(F_{v-1,v(r-1),.05})(F_{v(r-1),v-1,.05}) \le (2r+1)/(r+1)\,.$$

For the moment set $r = 11$, which is the value used by the experimenter in the ice cream experiment. Then $(2r+1)/(r+1) = 23/12 \approx 1.92$, and we have

v	$F_{v-1,10v,.05}$	$F_{10v,v-1,.05}$	Product		$\frac{2r+1}{r+1}$	Action
4	2.84	8.59	24.40	$>$	1.92	Increase v
100	1.26	1.30	1.64	$<$	1.92	Decrease v
80	1.30	1.34	1.74	$<$	1.92	Decrease v
60	1.34	1.41	1.89	$\approx$	1.92	Stop

So, v around 60 should be reasonable, requiring 660 observations in total, including 60 ice cream varieties.

Let us now reduce r to 3 and compute the required value of v. Then $(2r+1)/(r+1) = 1.75$, so

v	$F_{v-1,2v,.05}$	$F_{2v,v-1,.05}$	Product		$\frac{2r+1}{r+1}$	Action
80	1.37	1.39	1.90	$>$	1.75	Increase v
100	1.32	1.34	1.78	$>$	1.75	Increase v
105	1.31	1.33	1.75	$=$	1.75	Stop

So v in the region of 105 would be fine, requiring only $n = 315$ observations. In Exercise 2, the reader is asked to determine whether the use of $r = 2$ would require a smaller total number of observations. To greatly reduce the required number of observations, we would need to relax our requirement of such a high power to reject $H_0^{\gamma T} : \sigma_T^2 \leq \sigma^2$ when $\sigma_T^2 = 2\sigma^2$. □

17.5 Checking Assumptions on the Model

The simplest way to check the assumptions on the one-way random-effects model is to use residual plots in much the same way as for a fixed-effects one-way model. We need to check that the error assumptions are valid, that is,

$$\epsilon_{it} \sim N(0, \sigma^2), \quad t = 1, \ldots, r_i \,,$$

for each treatment factor level i ($i = 1, \ldots, v$), and also that the assumptions on the random effect T_i are valid, that is,

$$T_i \sim N(0, \sigma_T^2), \quad i = 1, \ldots, v \,,$$

and that all random variables are mutually independent.

Checking the error assumptions is straightforward, since we proceed in exactly the same way as for the fixed-effects one-way model. We replace T_i in the model, temporarily, by the fixed effect τ_i. Then the residuals are defined as usual as

$$\widehat{e}_{it} = y_{it} - \widehat{y}_{it} = y_{it} - \overline{y}_{i.} \,.$$

These are then standardized to obtain the standardized residuals z_{it} with standard deviation 1.0. We plot the standardized residuals versus treatment-factor levels, versus $\widehat{y}_{it}$, versus order, and versus normal scores, as in Chap. 5, to check for outliers, independence, constant variance, and normality. Non-independence between the ϵ_{it}'s and the T_i's is not easy to detect, but unequal variances of the ϵ_{it}'s indicates one form of the problem.

The normality assumption on the random effect T_i can be checked when the sample sizes are equal, unless v is too small. The treatment averages $\overline{Y}_{i.}$ should have a $N(\mu, \sigma_T^2 + \sigma^2/r)$ distribution. So, if we standardize the observed averages $\overline{y}_{i.}$ to have average value zero and sample standard deviation one, and we plot these standardized averages against their corresponding normal scores, we should roughly obtain a straight line—one that cuts the vertical axis at about zero and that has slope about one. It is important to check the normality assumption, since the analysis for random-effects models is not robust to nonnormality of the random effects. We can also use this normal probability plot to check for treatment effect outliers among the observed treatments. In an experiment such as the ice cream experiment, where only $v = 3$ levels of the treatment factor were observed, there is not enough data to be able to examine the distribution of the T_i's in any detail. In Sects. 17.10 and 17.11 we will illustrate the assumption-checking procedures using the SAS and R software packages, respectively, and the data from the clean wool experiment that was described in Sect. 17.2.2, p. 615.

17.6 Two or More Random Effects

17.6.1 Models and Examples

In the ice cream experiment of Example 17.3.1, p. 621, we modeled the ice cream effect as a random effect, since we were interested in the variability of the melting rates of varieties of a large population of all possible ice creams with similar ingredients. If the experimenter had also been interested in whether or not the container affects the melting time, then she might have randomly selected a number b of containers from the population of all possible containers. If one ice cream melts faster than another ice cream in one container, then it might be safe to assume that it melts faster, and by the same amount, in another container. In other words, the assumption of no ice cream$\times$container interaction might be reasonable. In this case a random two-way main-effects model (with no interaction) would be a possible model; that is,

$$\begin{gathered}Y_{ijt} = \mu + A_i + B_j + \epsilon_{ijt}\,, \qquad (17.6.19)\\ A_i \sim N(0, \sigma_A^2),\;\; B_j \sim N(0, \sigma_B^2),\;\; \epsilon_{ijt} \sim N(0, \sigma^2)\,,\\ A_i\text{'s},\; B_j\text{'s and } \epsilon_{ijt}\text{'s are all mutually independent}\\ t = 1, \ldots, r_{ij},\;\; i = 1, \ldots, a,\;\; j = 1, \ldots, b.\end{gathered}$$

where A_i is the effect of the ith ice cream randomly selected from the population of ice creams whose effects on melting times follow a normal distribution with variance σ_A^2 for each container, and where B_j is the effect of the jth container randomly selected from the population of containers whose effects on the melting times follow a normal distribution with variance σ_B^2 for each ice cream. The number of observations r_{ij} to be taken on the (ij)th ice cream–container combination needs to be determined. Normally, we would select the r_{ij}'s to be equal if possible.

Alternatively, it may be expected that a slightly thicker container would show a greater difference in melting times of ice creams than would a thinner container. In other words, an interaction may be expected. In this case, we would add to model (17.6.19) a random effect representing the interaction, as shown in the random-effects two-way complete model (17.6.20):

$$\begin{gathered}Y_{ijt} = \mu + A_i + B_j + (AB)_{ij} + \epsilon_{ijt} \qquad (17.6.20)\\ A_i \sim N(0, \sigma_A^2),\;\; B_j \sim N(0, \sigma_B^2)\\ (AB)_{ij} \sim N(0, \sigma_{AB}^2),\;\; \epsilon_{ijt} \sim N(0, \sigma^2)\\ A_i\text{'s},\; B_j\text{'s},\; (AB)_{ij}\text{'s and } \epsilon_{ijt}\text{'s are mutually independent}\\ t = 1, \ldots, r_{ij},\;\; i = 1, \ldots, a,\;\; j = 1, \ldots, b.\end{gathered}$$

If σ_{AB}^2 is positive, then there are AB *effects* present—namely, main effects and interactions for the factors A and B. If σ_A^2 or σ_B^2 is positive, then the corresponding main effects are present.

Example 17.6.1 `Ammunition experiment`

W.A. Thompson, Jr. and J.R. Moore in the 1963 volume of *Technometrics* describe an experiment concerning the muzzle velocity characteristics of ammunition for a field artillery weapon. They describe the ammunition as follows:

Table 17.5 Data for the ammunition experiment

		Charge lot			
		1	2	3	4
Projectile lot	1	63	56	69	78
		78	58	63	79
	2	71	60	64	65
		70	65	68	77
	3	72	58	69	63
		55	55	71	72
	4	70	60	66	73
		64	71	68	79

Source Thompson and Moore (1963).

> Propelling charges and projectiles for this type of weapon are manufactured and stored separately in a such a way that any charge might be employed by the user to propel any projectile…. Both projectiles and charges are grouped into lots at the time of manufacture, each lot consisting of a large number of individual units assembled during a short period of time using essentially uniform components. Thus, it is hoped that the round to round dispersion [variability] in velocity will be reduced by using charges and projectiles from within lots.

The experiment involved the examination of a random sample of four charge lots (factor A with levels 1, 2, 3, 4) selected at random from a large population of charge lots, and four projectile lots (factor B with levels 1, 2, 3, 4) selected at random from a large population of projectile lots. A weapon surveillance test was conducted using one weapon under uniform ballistic conditions. The muzzle velocities were measured to the nearest foot per second. These are shown in Table 17.5, except that a constant has been added to each recorded velocity.

Since the lots involved in the experiment were randomly selected from large populations, a random-effects two-way complete model (17.6.20) was used in the analysis. □

In an experiment with more than two random-effects treatment factors, variables representing all of the main effects of the factors and some or all of their interactions would be included in the model in the obvious way. For example, an experiment with five random-effects treatment factors A, B, C, D, G, in which interactions AB, AC, BC, CD, and ABC were thought to be nonnegligible, would be modeled as follows:

$$\begin{aligned} Y_{ijklmt} &= \mu + A_i + B_j + C_k + D_l + G_m \\ &\quad + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + (CD)_{kl} + (ABC)_{ijk} + \epsilon_{ijklmt}\,, \end{aligned}$$

$$A_i \sim N(0, \sigma_A^2),\ \ B_j \sim N(0, \sigma_B^2),\ \ C_k \sim N(0, \sigma_C^2),\ \ D_l \sim N(0, \sigma_D^2),\ \ G_m \sim N(0, \sigma_G^2),$$

$$(AB)_{ij} \sim N(0, \sigma_{AB}^2),\ (AC)_{ik} \sim N(0, \sigma_{AC}^2),\ (BC)_{jk} \sim N(0, \sigma_{BC}^2),\ (CD)_{kl} \sim N(0, \sigma_{CD}^2),$$

$$(ABC)_{ijk} \sim N(0, \sigma_{ABC}^2),\ \ \epsilon_{ijklmt} \sim N(0, \sigma^2),$$

all random variables on the right-hand side
of the model are mutually independent,

$$t = 1, \ldots, r_{ijklm},\ \ i = 1, \ldots, a,\ \ j = 1, \ldots, b,$$

$$k = 1, \ldots, c,\ \ l = 1, \ldots, d,\ \ m = 1, \ldots, g.$$

As for the fixed-effects models, if a high-order interaction is included in the model, then so are all of its "subinteractions" and constituent main effects; that is, if $(ABC)_{ijk}$ is in the model, so are $(AB)_{ij}$, $(AC)_{ik}$, $(BC)_{jk}$, A_i, B_j, and C_k.

17.6.2 Checking Model Assumptions

We may check the error assumptions by replacing temporarily all of the random effects by fixed effects, calculating the standardized residuals, and examining the residual plots in the usual way. Checking the assumptions of each random effect is not easy, since in a two-way or higher-way model there are generally few levels of each treatment factor observed, and the cell averages are not independent. Consequently, we will omit the model checks for the random-effect assumptions, and hope that any severe problems will show up through the analysis of the residuals.

17.6.3 Estimation of σ^2

In Sect. 17.3.2 we found that for the one-way random-effects model, an unbiased estimate of σ^2 was given by *msE*, where *msE* was calculated exactly as for the fixed-effects one-way model. Perhaps this should not be surprising, since *msE* measures the variability in the data that is not accounted for by those sources of variation that were ignored in the experiment. An unbiased estimate for σ^2 in *any* random-effects model can be obtained from its fixed-effects model counterpart.

Example 17.6.2 `Unbiased estimate of` σ^2

We will show that an unbiased estimate of σ^2 in the random-effects two-way complete model is $msE = ssE/(n - v)$, where

$$ssE = \left[\sum_i \sum_j \sum_t y_{ijt}^2 - \sum_i \sum_j r_{ij}\overline{y}_{ij.}^2\right],$$

as in (6.4.16) for the fixed-effects two-way complete model. First, note that

$$E[Y_{ijt}] = \mu \text{ and } \mathrm{Var}(Y_{ijt}) = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma^2$$

for the random-effects two-way complete model. Also,

$$\overline{Y}_{ij.} = \mu + A_i + B_j + (AB)_{ij} + \sum_t \epsilon_{ijt}/r_{ij}\,.$$

So,

$$E[\overline{Y}_{ij.}] = \mu \quad \text{and} \quad \mathrm{Var}(\overline{Y}_{ij.}) = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma^2/r_{ij}.$$

Thus, the expected value of the random variable *SSE* is

$$\begin{aligned} E\,[SSE] &= E\left[\sum_i \sum_j \sum_t Y_{ijt}^2 - \sum_i \sum_j r_{ij}\overline{Y}_{ij.}^2\right] \\ &= \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{t=1}^{r_{ij}} (\mathrm{Var}(Y_{ijt}) + E[Y_{ijt}]^2) \end{aligned}$$

$$- \sum_{i=1}^{a} \sum_{j=1}^{b} r_{ij} (\text{Var}(\overline{Y}_{ij.}) + E[\overline{Y}_{ij.}]^2)$$

$$= \left(\sum_{i=1}^{a} \sum_{j=1}^{b} r_{ij} \right) \sigma^2 - \sum_{i=1}^{a} \sum_{j=1}^{b} \sigma^2 = (n - v)\sigma^2 ,$$

where $n = \Sigma\Sigma r_{ij}$ and $v = ab$. Consequently,

$$E[MSE] = E[SSE/(n - v)] = \sigma^2.$$

□

17.6.4 Estimation of Variance Components

In Sect. 17.3.3, p. 620, we found that for the random-effects one-way model,

$$E[MST] = c\sigma_T^2 + \sigma^2 \quad , \quad \text{where } c = \frac{n^2 - \sum r_i^2}{n(v - 1)} ,$$

where MST is the mean square for treatments from the fixed-effects one-way model, and where $c = r$ if all the sample sizes are equal. From this, we were able to find an unbiased estimator for σ_T^2, namely $(MST - MSE)/c$.

For more complicated models, we will also be able to find unbiased estimators for the variance components using the fixed-effects mean squares, but each estimator must be calculated individually.

Example 17.6.3 `Unbiased estimate of` σ_B^2

Suppose an experiment involves three random-effects treatment factors A, B, and D having a, b, and d levels, respectively, and suppose r observations are taken on each of the $v = abd$ combinations. If the only interactions that are expected to be nonnegligible are AB and BD, then, the model is

$$Y_{ijk} = \mu + A_i + B_j + D_k + (AB)_{ij} + (BD)_{jk} + \epsilon_{ijkt} ,$$

$$t = 1, \ldots, r, \quad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, d$$

with the usual assumptions about the distributions of the random treatment effects and error variables.

Suppose we want an unbiased estimator for σ_B^2. We start by investigating $E[MSB]$, where $MSB = SSB/(b - 1)$. Using rule 4, p. 209, the sum of squares for B is

$$ssB = adr \sum_{j=1}^{b} \overline{y}_{.j..}^2 - abdr\overline{y}_{....}^2 .$$

Now,

$$E[\overline{Y}_{.j..}] = E[\overline{Y}_{....}] = \mu$$

and

$$\text{Var}[\overline{Y}_{.j..}] = \frac{\sigma_A^2}{a} + \sigma_B^2 + \frac{\sigma_D^2}{d} + \frac{\sigma_{AB}^2}{a} + \frac{\sigma_{BD}^2}{d} + \frac{\sigma^2}{adr} ,$$
$$\text{Var}[\overline{Y}_{....}] = \frac{\sigma_A^2}{a} + \frac{\sigma_B^2}{b} + \frac{\sigma_D^2}{d} + \frac{\sigma_{AB}^2}{ab} + \frac{\sigma_{BD}^2}{bd} + \frac{\sigma^2}{abdr} .$$

Consequently,

$$\begin{aligned} E[SSB] &= adr \sum_j \left[Var(\overline{Y}_{.j..}) + E[\overline{Y}_{.j..}]^2 \right] - abdr \left[Var(\overline{Y}_{....}) + E[\overline{Y}_{....}]^2 \right] \\ &= adr(b-1)\sigma_B^2 + dr(b-1)\sigma_{AB}^2 + ar(b-1)\sigma_{BD}^2 + (b-1)\sigma^2 . \end{aligned}$$

So,

$$E[MSB] = adr\sigma_B^2 + dr\sigma_{AB}^2 + ar\sigma_{BD}^2 + \sigma^2 . \tag{17.6.21}$$

Thus, if we wish to find an unbiased estimator for σ_B^2, we must find unbiased estimators also for σ_{AB}^2 and σ_{BD}^2. The logical place to look for these is at $E[MS(AB)]$ and $E[MS(BD)]$. We have

$$ss(AB) = dr \sum_{i=1}^{a} \sum_{j=1}^{b} \overline{y}_{ij..}^2 - bdr \sum_{i=1}^{a} \overline{y}_{i...}^2 - adr \sum_{j=1}^{b} \overline{y}_{.j..}^2 + abdr\overline{y}_{....}^2 ,$$
$$ss(BD) = ar \sum_{j=1}^{b} \sum_{k=1}^{d} \overline{y}_{.jk.}^2 - adr \sum_{j=1}^{b} \overline{y}_{.j..}^2 - abr \sum_{k=1}^{d} \overline{y}_{..k.}^2 + abdr\overline{y}_{....}^2 ,$$

and

$$\begin{aligned} E[\overline{Y}_{ij..}] &= E[\overline{Y}_{i...}] = E[\overline{Y}_{.j..}] = E[\overline{Y}_{.jk.}] = E[\overline{Y}_{..k.}] = [\overline{Y}_{....}] = \mu , \\ \text{Var}[\overline{Y}_{ij..}] &= \sigma_A^2 + \sigma_B^2 + \frac{\sigma_D^2}{d} + \sigma_{AB}^2 + \frac{\sigma_{BD}^2}{d} + \frac{\sigma^2}{dr} , \\ \text{Var}[\overline{Y}_{i...}] &= \sigma_A^2 + \frac{\sigma_B^2}{b} + \frac{\sigma_D^2}{d} + \frac{\sigma_{AB}^2}{b} + \frac{\sigma_{BD}^2}{bd} + \frac{\sigma^2}{bdr} , \\ \text{Var}[\overline{Y}_{.jk.}] &= \frac{\sigma_A^2}{a} + \sigma_B^2 + \sigma_D^2 + \frac{\sigma_{AB}^2}{a} + \sigma_{BD}^2 + \frac{\sigma^2}{ar} , \\ \text{Var}[\overline{Y}_{..k.}] &= \frac{\sigma_A^2}{a} + \frac{\sigma_B^2}{b} + \sigma_D^2 + \frac{\sigma_{AB}^2}{ab} + \frac{\sigma_{BD}^2}{b} + \frac{\sigma^2}{abr} , \end{aligned}$$

as well as

$$\begin{aligned} E[SS(AB)] &= \left(dr \sum_i \sum_j \overline{Y}_{ij..}^2 - bdr \sum_i \overline{Y}_{i...}^2 \right) - (E[SSB]) \\ &= adr(b-1)\sigma_B^2 + adr(b-1)\sigma_{AB}^2 + ar(b-1)\sigma_{BD}^2 + a(b-1)\sigma^2 \\ &\quad - adr(b-1)\sigma_B^2 - dr(b-1)\sigma_{AB}^2 - ar(b-1)\sigma_{BD}^2 - (b-1)\sigma^2 \\ &= dr(a-1)(b-1)\sigma_{AB}^2 + (a-1)(b-1)\sigma^2 . \end{aligned}$$

So,

$$E[MS(AB)] = dr\sigma_{AB}^2 + \sigma^2.$$

Similarly,

$$E[MS(BD)] = ar\sigma_{BD}^2 + \sigma^2 .$$

Thus, an unbiased estimator for σ_B^2 is

$$U = (MSB - MS(AB) - MS(BD) + MSE)/(adr) ,$$

and an unbiased estimate for σ_B^2 is therefore

$$u = (msB - ms(AB) - ms(BD) + msE)/(adr) .$$

□

Calculation of expected mean squares is quite time-consuming, as was seen in Example 17.6.3. However, when sample sizes are all equal, we can exploit the pattern that emerges in studying such examples. All of the variance components that are involved in $E[MSB]$ in (17.6.21) are those whose random effects include the same subscript as for B in the model. Specifically, B has subscript j in the model, and a j also occurs as subscript in $(AB)_{ij}$, $(BD)_{jk}$, and ϵ_{ijkt}. The constant in front of each variance component is the number of observations taken on each combination of subscripts; that is, there are adr observations on each of the b levels of B, there are dr observations on each of the ab levels of AB, and so on.

A similar pattern can be seen for $E[MS(AB)]$ and $E[MS(BD)]$. This gives us a general rule when sample sizes are equal (which we add to the 16 rules in Chap. 7):

17. To obtain the expected mean square for a main effect or interaction in a random-effects model, first note the subscripts on the term representing that effect in the model. Write down a variance component σ^2 for the effect of interest, for the error, and for every interaction whose term in the model includes the noted set of subscripts. Multiply each variance component except σ^2 by the number of observations taken on each level or combination of levels of the corresponding main effect or interaction. Add up the terms.

17.6.5 Confidence Intervals for Variance Components

In the previous subsection a rule was given for calculating the expected mean square corresponding to each term in the model, when the sample sizes are equal. For unequal sample sizes, the mean squares and expected mean squares are best calculated by a computer program.

From the list of expected mean squares, we can find an unbiased estimator for any given variance component, say σ_*^2. Again, this was illustrated in Example 17.6.3. The estimator can always be a linear combination of mean squares, which, in general, we can write as $U = \Sigma k_i (MS)_i$, where k_i is the constant in front of the ith mean square in the linear combination. Then, an approximation to the distribution of xU/σ_*^2 is a chi-squared distribution with x degrees of freedom, where

$$x = \frac{(\Sigma k_i (ms)_i)^2}{\Sigma k_i^2 (ms)_i^2 / x_i} \qquad (17.6.22)$$

and where x_i is the number of degrees of freedom corresponding to the ith mean square and $(ms)_i$ is the observed value of the ith mean square in the linear combination. An example of this formula was given in Sect. 17.3.5, p. 625, for the one-way model. A more complicated example is given below.

Example 17.6.4 `Calculation of degrees of freedom`

We continue Example 17.6.3, which involved a random-effects model with five random effects A_i, B_j, and D_k (corresponding to main effects of factors A, B, and D), and $(AB)_{ij}$ and $(BD)_{jk}$ (corresponding to interactions AB and BD). An unbiased estimator for σ_B^2 was shown to be

$$U = \Sigma k_i (MS)_i = MSB/(adr) - MS(AB)/(adr) - MS(BD)/(adr) + MSE/(adr).$$

An approximation to the distribution of xU/σ_B^2 is a χ_x^2 distribution, where x is given by (17.6.22), that is,

$$\begin{aligned} x &= \frac{[(msB - ms(AB) - ms(BD) + msE)/(adr)]^2}{\frac{msB^2}{(adr)^2(b-1)} + \frac{ms(AB)^2}{(adr)^2(a-1)(b-1)} + \frac{ms(BD)^2}{(adr)^2(b-1)(d-1)} + \frac{msE^2}{(adr)^2 df}} \\ &= \frac{[msB - ms(AB) - ms(BD) + msE]^2}{\frac{msB^2}{(b-1)} + \frac{ms(AB)^2}{(a-1)(b-1)} + \frac{ms(BD)^2}{(b-1)(d-1)} + \frac{msE^2}{df}}, \end{aligned}$$

where *df* is the number of degrees of freedom for error, which can be obtained, as usual, by subtraction. In this example, *df* is equal to

$$\begin{aligned} df &= (abdr - 1) - (a-1) - (b-1) - (d-1) \\ &\quad - (a-1)(b-1) - (b-1)(d-1) \\ &= ab(dr-1) - b(d-1) + 1\,. \end{aligned}$$

□

Once we know an approximate distribution for a variance component estimator, we can easily write down a probability statement and convert it to a confidence interval. Suppose that $U = \Sigma k_i (MS)_i$ is an unbiased estimator for σ_*^2 and that xU/σ_*^2 has approximately a χ_x^2 distribution; then

$$P\left(\chi^2_{x,1-\alpha/2} \le xU/\sigma_*^2 \le \chi^2_{x,\alpha/2}\right) \approx 1 - \alpha.$$

Then, if we observe the value of U to be $u = \Sigma k_i (ms)_i$, by manipulating the two inequalities in the probability statement we can obtain the following approximate $100(1-\alpha)\%$ confidence interval:

$$\frac{xu}{\chi^2_{x,\alpha/2}} \le \sigma_*^2 \le \frac{xu}{\chi^2_{x,1-\alpha/2}}, \tag{17.6.23}$$

where x is calculated as in (17.6.22). If the estimate u is negative or the calculated degrees of freedom x is extremely small, then this approximate confidence interval procedure should not be used.

Example 17.6.5 `Ammunition experiment, continued`

The ammunition experiment was described in Example 17.6.1, p. 632, and the data were given in Table 17.5. A random-effects two-way complete model (17.6.20) was used. The mean squares for this

Table 17.6 Two-way analysis of variance table for the ammunition experiment

Source of variation	Degrees of freedom	Sum of squares	Mean squares	Expected mean square
Charge (A)	3	669.12	223.04	$8\sigma_A^2 + 2\sigma_{AB}^2 + \sigma^2$
Projectile (B)	3	92.12	30.71	$8\sigma_B^2 + 2\sigma_{AB}^2 + \sigma^2$
Interaction (AB)	9	257.63	28.63	$2\sigma_{AB}^2 + \sigma^2$
Error	16	516.00	32.25	σ^2
Total	31	1534.87		

model are calculated in exactly the same way as for the fixed-effects two-way complete model, and these are shown in the analysis of variance table, Table 17.6. Also listed in the table are the expected mean squares calculated as in rule 17, p. 637.

For example, to calculate the expected mean square for A, we note that the subscript for the term A_i in the model is i, and also that i is included among the subscripts of the terms $(AB)_{ij}$ and ϵ_{ijt}. This means that the expected mean square must include the three variance components

$$\sigma_A^2, \quad \sigma_{AB}^2, \quad \text{and} \quad \sigma^2 .$$

The constant in front of σ_A^2 is 8, the number of observations on each charge lot, whereas the constant in front of σ_{AB}^2 is 2, the number of observations on each combination of charge lot and projectile lot.

The expected mean square for AB, $E[MS(AB)] = 2\sigma_{AB}^2 + \sigma^2$, contains only two terms, since only the two terms $(AB)_{ij}$ and ϵ_{ijt} in the model contain both i and j as subscripts. An unbiased estimator for σ_A^2 is given by $U = (MSA - MS(AB))/8$. Also, xU/σ_A^2 has approximately a χ_x^2 distribution, where x is calculated as in (17.6.22). Thus, an unbiased estimate of σ_A^2 from this experiment is

$$u = (msA - ms(AB))/8 = (223.04 - 28.63)/8 = 24.30 ,$$

and the number of degrees of freedom of the associated chi-squared distribution is

$$x = \frac{24.30^2}{\frac{223.04^2}{(8^2)(3)} + \frac{28.63^2}{(8^2)(9)}} = 2.27 .$$

Therefore, an approximate 90% confidence interval (17.6.23) for σ_A^2, the variance of velocities arising from the population of charge lots, is

$$\frac{(2.27)(24.3)}{\chi^2_{2.27,0.05}} \le \sigma_A^2 \le \frac{(2.27)(24.3)}{\chi^2_{2.27,0.95}} ,$$

and since $\chi^2_{2.27,0.05} \approx 6.5$ and $\chi^2_{2.27,0.05} \approx 0.17$, the approximate 90% confidence interval, in units of (feet per second)^2, is

$$8.49 \le \sigma_A^2 \le 324.48 .$$

An approximate 90% confidence interval for the standard deviation of the velocity (in feet per second), obtained by taking square roots, is

$$2.91 \le \sigma_A \le 18.01 .$$

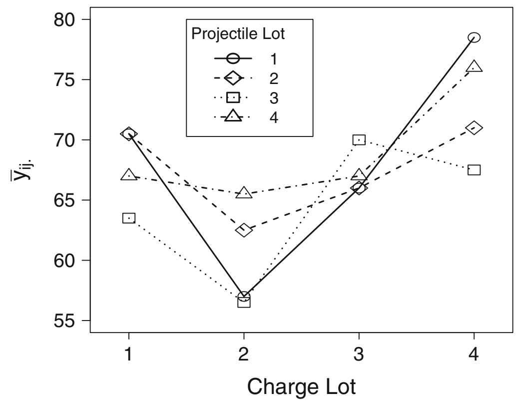

Fig. 17.2 Plot of average velocity against charge lot by projectile lot for the ammunition experiment

Before leaving this example, we note that an unbiased estimate for σ^2_{AB} calculated in this way is actually negative, since

$$u = (ms(AB) - msE)/2 = -1.81 \,,$$

and the calculation for x, the number of degrees of freedom of the associated χ^2 distribution, is

$$x = \frac{(-1.81)^2}{\frac{28.63^2}{(2^2)(9)} + \frac{32.25^2}{(2^2)(16)}} = 0.084 \,.$$

Thus, we are not able to say anything sensible about the variance of the interaction, other than that it appears to be very small. The interaction plot in Fig. 17.2 for the lots included in the experiment supports this conclusion. □

17.6.6 Hypothesis Tests for Variance Components

In order to focus the discussion, we will use the random-effects model of Example 17.6.3, p. 635; that is,

$$Y_{ijkt} = \mu + A_i + B_j + D_k + (AB)_{ij} + (BD)_{jk} + \epsilon_{ijkt} \,,$$

$$t = 1, \ldots, r \,; \quad i = 1, \ldots, a \,; \quad j = 1, \ldots, b \,; \quad k = 1, \ldots, d \,,$$

together with the usual assumptions about the distributions of the random variables. Some of the expected mean squares for this model were calculated in Example 17.6.3 and are listed, together with the remaining mean squares, in Table 17.7.

Testing the hypothesis $H_0^{AB} : \{\sigma^2_{AB} = 0\}$ against its alternative hypothesis $H_A^{AB} : \{\sigma^2_{AB} > 0\}$ is straightforward, since the corresponding expected mean square looks very similar to the situation that we had in the one-way model. If H_0^{AB} is true, then the numerator of the ratio $ms(AB)/msE$ is expected to be σ^2, the same as the denominator. Otherwise, the numerator is expected to be larger. Consequently, the decision rule is

$$\text{reject } H_0^{AB} \text{ if } \frac{ms(AB)}{msE} > F_{(a-1)(b-1), df, \alpha}$$

Table 17.7 Expected mean squares and degrees of freedom for a random-effects three-way model with two interactions

Effect	Degrees of freedom	Expected mean square
A	$a-1$	$bdr\sigma_A^2 + dr\sigma_{AB}^2 + \sigma^2$
B	$b-1$	$adr\sigma_B^2 + dr\sigma_{AB}^2 + ar\sigma_{BD}^2 + \sigma^2$
D	$d-1$	$abr\sigma_D^2 + ar\sigma_{BD}^2 + \sigma^2$
AB	$(a-1)(b-1)$	$dr\sigma_{AB}^2 + \sigma^2$
BD	$(b-1)(d-1)$	$ar\sigma_{BD}^2 + \sigma^2$
Error	df	σ^2

as usual, where the number of error degrees of freedom is

$$df = ab(dr-1) - b(d-1) + 1 .$$

We could modify this test as in (17.3.9), p. 624, so that the decision rule for testing $H_0^{\gamma AB} : \{\sigma_{AB}^2 \le \gamma\sigma^2\}$ against $H_A^{\gamma AB} : \{\sigma_{AB}^2 > \gamma\sigma^2\}$ is

$$\text{reject } H_0^{\gamma AB} \text{ if } \frac{ms(AB)}{msE} > (1 + dr\gamma)F_{(a-1)(b-1),df,\alpha} .$$

We have similar tests for H_0^{BD} against H_A^{BD}, and $H_0^{\gamma BD}$ against $H_A^{\gamma BD}$.

Testing $H_0^A : \sigma_A^2 = 0$ against $H_A^A : \sigma_A^2 > 0$ is more complicated. Until now, we have used the same test statistics as we used in the fixed-effects case. But if we try to use msA/msE to test H_0^A, we have a problem. If H_0^A is true, so that $\sigma_A^2 = 0$, the expected value of the numerator is $E[MSA] = dr\sigma_{AB}^2 + \sigma^2$, while that of the denominator is $E[msE] = \sigma^2$. This suggests two things:

(i) we should use $ms(AB)$ as the denominator, not msE, and
(ii) we should question whether it makes sense to test H_0^A if the interaction AB is significant.

The second point is, of course, exactly the same point that arose in the fixed-effects model, and the answer is usually "no, it makes no sense." Consequently, we generally test a main effect only when that factor is not involved in any significant interactions. Nevertheless, we shall still use the interaction mean square as the denominator in case an incorrect decision was made regarding the interaction. Consequently, the decision rule for testing H_0^A against H_A^A is

$$\text{reject } H_0^A \text{ if } MSA/MS(AB) > F_{a-1,(a-1)(b-1),\alpha} .$$

Notice that the second set of degrees of freedom for the F-distribution is the degrees of freedom corresponding to the denominator of the ratio. The test for $H_0^D : \{\sigma_D^2 = 0\}$ is similar.

Obtaining a suitable denominator for testing the null hypothesis $H_0^B : \{\sigma_B^2 = 0\}$ versus the alternative hypothesis $H_0^B : \{\sigma_B^2 > 0\}$ is harder again. If H_0^B is true, so that $\sigma_B^2 = 0$, then the expected value of MSB is

$$E[MSB] = dr\sigma_{AB}^2 + ar\sigma_{BD}^2 + \sigma^2 .$$

We would generally want to test this hypothesis only if we believed that the interactions AB and BD were both negligible. Yet to be on the safe side, we would like a denominator with the same expected value. It can be verified that

$$E[U] = E[MS(AB) + MS(BD) - MSE] = dr\sigma^2_{AB} + ar\sigma^2_{BD} + \sigma^2.$$

As in Sect. 17.3.5, $xU/(dr\sigma^2_{AB} + ar\sigma^2_{BD} + \sigma^2)$ has approximately a chi-squared distribution with degrees of freedom x calculated as in (17.6.22), p. 637; that is,

$$x = \frac{[ms(AB) + ms(BD) - msE]^2}{\frac{[ms(AB)]^2}{(a-1)(b-1)} + \frac{[ms(BD)]^2}{(b-1)(d-1)} + \frac{[msE]^2}{df}}\,.$$

Therefore, if H_0 is true, msB/U has approximately an $F_{b-1,x}$ distribution. So to test H_0^B against H_A^B, the decision rule is

$$\text{reject } H_0^B \text{ if } \frac{msB}{ms(AB) + ms(BD) - msE} > F_{(b-1),x,\alpha}.$$

Example 17.6.6 `Ammunition experiment, continued`

An unbiased estimate for the variance of the muzzle velocities due to the population of charge lots (factor A) was calculated to be $u = 24.3$ (feet per second)^2 in Example 17.6.5 for the ammunition experiment. A question is whether this value could be due to random error or whether the variance is really sizable; that is, we wish to test the hypothesis $H_0^A : \{\sigma^2_A = 0\}$ against the alternative hypothesis $H_A^A : \{\sigma^2_A > 0\}$. The interaction variability was found to be very small in Example 17.6.5, so the main-effect hypothesis makes sense. The expected mean squares for A and AB are listed in Table 17.6 as

$$\text{E}[MSA] \;=\; 8\sigma^2_A + 2\sigma^2_{AB} + \sigma^2$$

and

$$\text{E}[MS(AB)] \;=\; 2\sigma^2_{AB} + \sigma^2\,,$$

with 3 and 9 corresponding degrees of freedom, respectively. The decision rule, therefore, is

$$\text{reject } H_0^A \text{ if } \frac{msA}{ms(AB)} \;>\; F_{3,9,\alpha}\,.$$

If we select a Type I error probability of $\alpha = 0.05$, then $F_{3,9,0.05} = 3.86$. Since $msA/ms(AB) = 223.04/28.63 = 7.79$, we can conclude that $\sigma^2_A > 0$. □

17.6.7 Sample Sizes

If we test main effects and interactions only when the higher-order interactions involving those factors are negligible, then we can adapt (17.4.18) by changing the degrees of freedom to match those in the decision rule being used.

17.7 Mixed Models

Models that contain both random and fixed treatment effects are called *mixed models*. The analysis of random effects proceeds in exactly the same way as described in the previous sections. All that is needed is a way to write down the expected mean squares. The fixed effects can be analyzed as in Chaps. 3–7, except that, here, too, we may need to replace the mean square for error by a different appropriate mean square. We show how to calculate the expected mean squares for a mixed model in Sect. 17.7.1.

An interaction between two or more factors any of which has random effects will be regarded as a random effect, since the combination of levels observed in the experiment depends upon the random selection of levels of those factors that have random effects.

17.7.1 Expected Mean Squares and Hypothesis Tests

Expected mean squares can be obtained for a mixed model when the sample sizes are equal by modifying rule 17 on p. 637. We start by writing out the expected mean squares as though all the factors were random. We then collect all of the fixed effects and list them together as one "quadratic form." The quadratic form is a function of fixed-effect parameters such as $\alpha_i^* = \alpha_i + \overline{(\alpha\beta)}_{i.}$ (see Example 17.7.1) that typically feature in fixed-effects models.

As an example, consider a model containing the main effects of factors A, B, and D and the interactions AB and BD. Suppose that factors A and B have fixed effects, so that all of their levels of interest are observed in the experiment, and factor D has random effects, so that its levels form a large population of which only a random selection are observed in the experiment. Then interaction AB is a fixed effect, but interaction BD is a random effect.

We use α_i to represent the effect of the ith level of A, β_j to represent the effect of the jth level of B, and $(\alpha\beta)_{ij}$ to represent their interaction. The effect of the kth randomly selected level of D is represented by the random variable D_k, and the effect of the interaction between the jth specifically selected level of B and the kth randomly selected level of D is denoted by the random variable $(\beta D)_{jk}$. The model is then as follows:

$$\begin{gathered}
Y_{ijkt} = \mu + \alpha_i + \beta_j + D_k + (\alpha\beta)_{ij} + (\beta D)_{jk} + \epsilon_{ijkt}\,, \\
D_k \sim N(0, \sigma_D^2), \quad (\beta D)_{jk} \sim N(0, \sigma_{BD}^2), \quad \epsilon_{ijkt} \sim N(0, \sigma^2)\,, \\
D_k\text{'s}, \ (\beta D)_{jk}\text{'s and, } \epsilon_{ijkt}\text{'s are all mutually independent}, \\
t = 1, \ldots, r, \quad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, d.
\end{gathered} \tag{17.7.24}$$

The expected mean squares for the corresponding random-effects model were calculated in Example 17.6.3 and are reproduced in the second column of Table 17.8. The expected mean squares for the above mixed model are given in the third column of Table 17.8 and are obtained by collecting the terms in the expected mean squares corresponding to the fixed effects into one quadratic form.

Table 17.8 Expected mean squares for a three-way mixed model

Effect	For random-effects model	For mixed model
A	$bdr\sigma_A^2 + dr\sigma_{AB}^2 + \sigma^2$	$Q(A, AB) + \sigma^2$
B	$adr\sigma_B^2 + dr\sigma_{AB}^2 + ar\sigma_{BD}^2 + \sigma^2$	$Q(B, AB) + ar\sigma_{BD}^2 + \sigma^2$
D	$abr\sigma_D^2 + ar\sigma_{BD}^2 + \sigma^2$	$abr\sigma_D^2 + ar\sigma_{BD}^2 + \sigma^2$
AB	$dr\sigma_{AB}^2 + \sigma^2$	$Q(AB) + \sigma^2$
BD	$ar\sigma_{BD}^2 + \sigma^2$	$ar\sigma_{BD}^2 + \sigma^2$
Error	σ^2	σ^2

The expected mean squares can all be verified by direct calculation. We illustrate the calculation for B in the following example. The term $Q(B, AB)$ in *E[MSB]* corresponds to a quadratic (i.e., squared) function of $\beta_j^* = \beta_j + \overline{(\alpha\beta)}_{.j}$, a quantity that we are used to dealing with in fixed-effects models.

Example 17.7.1 `Calculation of expected mean squares`

Consider an experiment with two fixed-effects treatment factors A and B, and one random-effects treatment factor D, and suppose that (17.7.24) is thought to be a reasonable model. Using rule 4 of Chap. 7, the fixed-effect sum of squares for B is

$$SSB = adr\sum_{j=1}^{b}\overline{Y}_{.j..}^2 - abdr\overline{Y}_{....}^2 .$$

Now,

$$\begin{aligned}\overline{Y}_{.j..} &= \sum_{i=1}^{a}\sum_{k=1}^{d}\sum_{t=1}^{r} Y_{ijkt}/adr \\ &= \mu + \overline{\alpha}_. + \beta_j + \overline{D}_. + \overline{(\alpha\beta)}_{.j} + \overline{(\beta D)}_{j.} + \overline{\epsilon}_{.j..} .\end{aligned}$$

So,

$$E[\overline{Y}_{.j..}] = \mu + \overline{\alpha}_. + \beta_j + \overline{(\alpha\beta)}_{.j} \;\text{ and }\; \text{Var}(\overline{Y}_{.j..}) = \frac{\sigma_D^2}{d} + \frac{\sigma_{BD}^2}{d} + \frac{\sigma^2}{adr} .$$

Similarly,

$$E[\overline{Y}_{....}] = \mu + \overline{\alpha}_. + \overline{\beta}_. + \overline{(\alpha\beta)}_{..} \;\text{ and }\; \text{Var}(\overline{Y}_{....}) = \frac{\sigma_D^2}{d} + \frac{\sigma_{BD}^2}{bd} + \frac{\sigma^2}{abdr} .$$

Using the facts that $MSB = SSB/(b-1)$ and $E[X^2] = \text{Var}(X) + (E[X])^2$, we obtain

$$\begin{aligned}E[MSB] &= \frac{adr}{(b-1)}\sum_{j=1}^{b}\left(\frac{\sigma_D^2}{d} + \frac{\sigma_{BD}^2}{d} + \frac{\sigma^2}{adr} + [\mu + \overline{\alpha}_. + \beta_j + \overline{(\alpha\beta)}_{.j}]^2\right) \\ &\quad - \frac{abdr}{(b-1)}\left(\frac{\sigma_D^2}{d} + \frac{\sigma_{BD}^2}{bd} + \frac{\sigma^2}{abdr} + [\mu + \overline{\alpha}_. + \overline{\beta}_. + \overline{(\alpha\beta)}_{..}]^2\right) \\ &= ar\sigma_{BD}^2 + \sigma^2 + Q(B, AB) ,\end{aligned}$$

where

$$Q(B, AB) = \frac{adr}{(b-1)}\sum_j \left[(\beta_j + \overline{(\alpha\beta)}_{.j}) - (\overline{\beta}_. + \overline{(\alpha\beta)}_{..})\right]^2 .$$

□

Notice that in Example 17.7.1, the quadratic form $Q(B, AB)$ is equal to zero when all the $\beta_j^* = \beta_j + \overline{(\alpha\beta)}_{.j}$ are equal. We can make use of this fact when looking for an appropriate denominator for the test ratio for testing $H_0^B : \{\beta_j + \overline{(\alpha\beta)}_{.j}$ are all equal$\}$. If this hypothesis is true, then

Table 17.9 Test ratios for a three-way mixed model

Effect	$E[MS]$	Ratio
A	$Q(A, AB) + \sigma^2$	msA/msE
B	$Q(B, AB) + ar\sigma^2_{BD} + \sigma^2$	$msB/ms(BD)$
D	$abr\sigma^2_D + ar\sigma^2_{BD} + \sigma^2$	$msD/ms(BD)$
AB	$Q(AB) + \sigma^2$	$ms(AB)/msE$
BD	$ar\sigma^2_{BD} + \sigma^2$	$ms(BD)/msE$

$$E[MSB] = ar\sigma^2_{BD} + \sigma^2 \,.$$

Consequently, a sensible denominator would be *MS(BD)*, which has the same expected value (see Table 17.8). Thus, the decision rule for testing H_0^B against the alternative hypothesis that the β_j^* are not all equal is

$$\text{reject } H_0^B \text{ if } \frac{msB}{ms(BD)} > F_{(b-1),(b-1)(d-1),\alpha}.$$

From Table 17.8 we can construct tests for the other relevant hypotheses in a similar manner. For example, to test the hypothesis

$$H_0^{AB} : \{(\alpha\beta)_{ij} - \overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{.j} + \overline{(\alpha\beta)}_{..} = 0\,,\ \text{for all } i, j\}$$

against the alternative hypothesis that the interaction contrasts are not all zero, the decision rule is

$$\text{reject } H_0^{AB} \text{ if } \frac{ms(AB)}{msE} > F_{(a-1)(b-1),df,\alpha}\,,$$

where *df* is the number of error degrees of freedom.

To test the hypothesis $H_0^D : \{\sigma^2_D = 0\}$ against the alternative hypothesis $H_A^D : \{\sigma^2_D > 0\}$, the decision rule is

$$\text{reject } H_0^D \text{ if } \frac{msD}{ms(BD)} > F_{d-1,(b-1)(d-1),\alpha}.$$

The test ratios are summarized in Table 17.9. Generally, we would not test a main-effect or interaction hypothesis unless all higher-order interactions involving these same factors were believed to be negligible. For some mixed models, as for random-effects models, the appropriate denominator for the test statistic may not be listed among the expected mean squares for the factors in the model. In this case, it would be necessary to calculate it, and the corresponding degrees of freedom, using (17.6.22), p. 637.

17.7.2 Confidence Intervals in Mixed Models

Confidence Intervals for Fixed Effects

For fixed effects in a mixed model with equal sample sizes, we may use all of the rules of Sect. 7.3, p. 209, exactly as if there were no random effects in the model, *except that we replace msE by the same mean square that was identified for hypothesis testing*—namely, used in the denominator of the test ratio—and the error degrees of freedom are also replaced. The necessity of doing this replacement is highlighted in Example 17.7.2. Apart from this, we may use the Bonferroni, Scheffé, Tukey, and Dunnett methods of multiple comparisons in the usual way. When the sample sizes are unequal, computing least squares estimates and appropriate standard errors is more complicated. Appropriate methods will be illustrated in Chaps. 18 and 19 using `PROC MIXED` in SAS software and using the `lmer` and `lsmeans` functions in R.

Example 17.7.2 `Calculation of confidence intervals`

Consider an experiment with two fixed-effects treatment factors and a third treatment factor with random effects, for which the following model is thought to be reasonable (this is the same model that has been discussed throughout this subsection):

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + D_k + (\alpha\beta)_{ij} + (\beta D)_{jk} + \epsilon_{ijkt}.$$

The fixed part of the model is

$$\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}\,,$$

which looks exactly like one of the two-way analysis of variance models that was studied in Chap. 6.

Suppose we need confidence intervals for pairwise comparisons in the levels of A and of B. Then, as usual, the least squares estimates for pairwise differences are

$$\widehat{\alpha}_i^* - \widehat{\alpha}_p^* = \left(\widehat{\alpha}_i + \widehat{\overline{(\alpha\beta)}}_{i.}\right) - \left(\widehat{\alpha}_p - \widehat{\overline{(\alpha\beta)}}_{p.}\right) = \overline{y}_{i...} - \overline{y}_{p...}$$

and

$$\widehat{\beta}_j^* - \widehat{\beta}_u^* = \left(\widehat{\beta}_j + \widehat{\overline{(\alpha\beta)}}_{.j}\right) - \left(\widehat{\beta}_u + \widehat{\overline{(\alpha\beta)}}_{.u}\right) = \overline{y}_{.j..} - \overline{y}_{.u..}\,.$$

Tables 17.8 (p. 643) and 17.9 (p. 645) suggest that *msE* should be used in the formulae for confidence intervals for $\alpha_i^* - \alpha_p^*$, as usual, but that *ms(BD)* should be used in place of *msE* in the formulae for confidence intervals for $\beta_j^* - \beta_u^*$. All confidence intervals are of the form

$$\text{(least squares estimate)} \pm (w) \times \text{(standard error)}\,.$$

The standard error is the square root of the estimated variance of the least squares estimator. Now,

$$\text{Var}(Y_{ijkt}) = \sigma_D^2 + \sigma_{BD}^2 + \sigma^2 \quad \text{and} \quad \text{Var}(\overline{Y}_{.j..}) = \frac{\sigma_D^2}{d} + \frac{\sigma_{BD}^2}{d} + \frac{\sigma^2}{adr}\,.$$

The Y_{ijkt}'s are not independent. Observations on the same level of D are correlated. If two observations are taken on the same levels of B and D, we have

$$\text{Cov}(Y_{ijkt}\,,\; Y_{pjks}) = \sigma_D^2 + \sigma_{BD}^2\,.$$

If two observations are taken on the same level of D, but different levels of B, then

$$\text{Cov}(Y_{ijkt}\,,\; Y_{puks}) = \sigma_D^2\,.$$

All other pairs of response variables are independent. Consequently,

$$\begin{aligned}\text{Cov}(\overline{Y}_{.j..}\,,\ \overline{Y}_{.u..}) &= \frac{1}{a^2d^2r^2}\left[\sum_{i=1}^{a}\sum_{p=1}^{a}\sum_{k=1}^{d}\sum_{t=1}^{r}\sum_{s=1}^{r}\text{Cov}(Y_{ijkt}, Y_{puks})\right]\\ &= \frac{1}{a^2d^2r^2}\left[a^2dr^2\sigma_D^2\right]\\ &= \frac{\sigma_D^2}{d}\,,\end{aligned}$$

and

$$\begin{aligned}\text{Var}(\overline{Y}_{.j..} - \overline{Y}_{.u..}) &= \text{Var}(\overline{Y}_{.j..}) + \text{Var}(\overline{Y}_{.u..}) - 2\text{Cov}(\overline{Y}_{.j..}\,,\ \overline{Y}_{.u..})\\ &= 2\left(\frac{\sigma_D^2}{d} + \frac{\sigma_{BD}^2}{d} + \frac{\sigma^2}{adr}\right) - 2\left(\frac{\sigma_D^2}{d}\right)\\ &= \frac{2}{adr}\left(ar\sigma_{BD}^2 + \sigma^2\right)\,,\end{aligned}$$

which is of the form $(\Sigma c_i^2/(adr))(ar\sigma_{BD}^2 + \sigma^2)$. Thus, we need to estimate $(ar\sigma_{BD}^2 + \sigma^2)$ rather than σ^2, and an unbiased estimate is given by *ms(BD)*. So, the standard error for $\hat{\beta}_j^* - \hat{\beta}_u^* = \overline{Y}_{.j..} - \overline{Y}_{.u..}$ is $((2/(adr))\ ms(BD))^{1/2}$ with corresponding degrees of freedom $(b-1)(d-1)$. □

In some models, the necessary mean square will not be listed in the expected mean squares table, and (17.6.22), p. 637, and methods discussed there will need to be used to find an approximate mean square and degrees of freedom.

Confidence Intervals for Variance Components

In obtaining confidence intervals for variance components, only the random part of the model is used, or, equivalently, only the mean squares corresponding to random effects. Consequently, the formulae of Sect. 17.6.5 are used exactly as described for random-effects models.

17.8 Rules for Analysis of Random-Effects and Mixed Models

Rules 1–7 of Sect. 7.3, p. 209, are valid for calculating degrees of freedom, sums of squares, and mean squares in random-effects and mixed models as well as in fixed-effects models. In addition, rules 8–16 are valid for analyzing fixed effects, except that σ^2 and *msE* may need to be replaced. Rules 17–22 below summarize the results of this chapter. Rule 17 is an expanded version of rule 17 on p. 637.

17.8.1 Rules—Equal Sample Sizes

17. To obtain the expected mean square for a particular main effect or interaction, first make a note of the subscripts on the term representing that particular effect in the model. Write down variance components for the effect of interest, for the error, and for every interaction whose term in the

model includes the noted set of subscripts. Gather up all variance components corresponding to fixed effects into one quadratic form Q. Multiply any remaining variance component except σ^2 by the number of observations taken on each level or combination of levels of the corresponding effect (main effect or interaction). Add up the terms.

18. To obtain the denominator of the test statistic for testing the null hypothesis that a main effect or interaction effect is zero, write down the expected mean square for the effect of interest (see rule 17). Cross out the term that would be zero if the null hypothesis were true. The denominator of the test statistic is the mean square, or linear combination of mean squares, u, whose expected value is equal to the remaining expression.
19. For a random effect, let $U = \Sigma k_i MS_i$ be the mean square or linear combination of mean squares whose expected value is equal to the variance component corresponding to the random effect. An exact or approximate $100(1-\alpha)\%$ confidence interval for this variance component is

$$\left(\frac{xu}{\chi^2_{x,\alpha/2}} \; , \; \frac{xu}{\chi^2_{x,1-\alpha/2}} \right),$$

where

$$x = \frac{[\Sigma k_i (ms_i)]^2}{\Sigma [k_i (ms_i)]^2 / x_i}$$

and where u is the observed value of U, ms_i is the observed value of MS_i, and x_i is the number of degrees of freedom corresponding to ms_i.

20. For a fixed effect, confidence intervals are obtained as in rule 14, p. 211, except that msE is replaced by the denominator u from rule 18, and the number of error degrees of freedom is replaced by x in rule 19.
21. For a fixed effect, the decision rule for testing the hypothesis that the effect is zero is the same as that in rule 8, p. 210, for fixed-effects models, except that msE is replaced by the denominator u from rule 18, and the number of error degrees of freedom is replaced by x in rule 19.
22. For a random effect, the decision rule for testing the hypothesis H_0 that the corresponding variance component is zero against the alternative hypothesis that it is not zero is

$$\text{reject } H_0 \text{ if } \frac{ms}{u} > F_{\nu,x,\alpha} \; ,$$

where ms is the mean square for the effect of interest and ν the corresponding degrees of freedom, u is the observed value of the denominator as in rule 18, and x is the corresponding degrees of freedom calculated as in rule 19.

17.8.2 Controversy (Optional)

Before proceeding, we should mention that some other textbooks may present slightly different tables of expected mean squares. For example, the expected mean square for D in Table 17.8, which we have calculated as

$$E[MSD] = abr\sigma^2_D + ar\sigma^2_{BD} + \sigma^2,$$

may in other texts be listed as

$$E[MSD] = abr\sigma^2_D + \sigma^2 \; .$$

This alternative listing occurs when constraints are placed on the model parameters involving fixed-effects factors, and it suggests use of the denominator *msE* rather than *ms(BD)* in testing $H_0^D : \{\sigma_D^2 = 0\}$ against $H_A^D : \{\sigma_D^2 > 0\}$. A number of articles in the statistical literature have been written advocating one denominator rather than the other, and there still appears to be no consensus.

If we follow the line of reasoning that we have followed to this point, that normally we will examine main effects only when there is no interaction, then some of the controversy disappears. If σ_{BD}^2 is really zero, then $E[MSD] = abr\sigma_D^2 + \sigma^2$ in both cases. Of course, due to variability of the data and uncertainty about whether or not σ_{BD}^2 is really zero (or close to it), we still have to make the choice in practice. We have recommended using *ms(BD)* as the denominator if the objective is to test $H_0^D : \sigma_D^2 = 0$. However, if interest is really in testing

$$H_0^{D+BD} : \{\sigma_D^2 + b^{-1}\sigma_{BD}^2 = 0\}\,,$$

or equivalently

$$H_0^{D+BD} : \{\sigma_D^2 = \sigma_{BD}^2 = 0\}\,,$$

then we would use *msE* as the denominator.

The controversy originally arose from the formulation of the model. In our example, the model was given in (17.7.24), p. 643, and the controversy surrounds the random effect $(\beta D)_{jk}$. We have modeled this as a normally distributed random variable. Some authors add to the model the restriction $\Sigma_j(\beta D)_{jk} = 0$, and this leads to the canceling of the term in σ_{BD}^2 when the expected mean square of D is calculated.

Hocking (1996, p. 569) shows that under this restriction, the hypothesis H_0^D is actually our hypothesis H_0^{D+BD}. An explanation for this is as follows. If constraints are placed on the parameters, then the $(\beta D)_{jk}$ effects truly represent interaction effects, and σ_{BD}^2 measures precisely variability in *BD*-interaction effects. However, if no constraints are placed on the parameters, then σ_{BD}^2 being positive implies the presence of main effects of B and D as well as the presence of *BD*-interaction effects. In other words, the parameters $(\beta D)_{jk}$ represent "*BD* effects" in model (17.7.24), though we have referred to them as *BD*-interaction effects. Thus, under our model (17.7.24), the hypothesis $H_0^{D+BD} : \{\sigma_D^2 = \sigma_{BD}^2 = 0\}$ is that there are no main effects of D (or *BD* interactions). Also, there are no *BD* interactions if $\sigma_{BD}^2 = 0$, and there are no main effects of B (or *BD* interactions) if $\beta_1 = \beta_2 = \cdots = \beta_b$ and $\sigma_{BD}^2 = 0$. From this viewpoint, the hypothesis $H_0^D : \sigma_D^2 = 0$ is that there are no main effects of D if σ_{BD}^2 is believed to be zero; otherwise, it is the hypothesis that main effects of D are no less negligible than *BD* interactions.

Since there are problems inherent in placing restrictions on the model parameters, we prefer not to do so, and we prefer to use the set of expected mean squares in Table 17.7. If the parameters in the model are properly interpreted, then there is no controversy, and the appropriate test is determined by what is most sensible for the experiment at hand.

17.9 Block Designs and Random Block Effects

In certain types of experiments, it is extremely common for the levels of a blocking factor to be randomly selected. For example, in medical, psychological, educational, or pharmaceutical experiments, blocks frequently represent subjects that have been selected at random from a large population of similar subjects. In agricultural experiments, blocks may represent different fields selected from a large variable population of fields. In industrial experiments, different machine operators may represent different levels of the blocking factor and may be similar to a random sample from a large population of possible operators. Raw material may be delivered to the factory in batches, a random selection of which are used as blocks in the experiment.

Since we are not interested in the blocking factor itself, its designation as random rather than fixed will affect the analysis only if the model includes a block×treatment interaction. For example, suppose that factor D in Table 17.8 represents a random-effects blocking factor, and that A and B are two fixed-effects treatment factors. The analysis of factor A, which has no interaction with D, is unaffected by the designation of D as a random effect. However, the analysis of factor B, which interacts with blocks, *is* affected, since msE in hypothesis tests and confidence intervals for contrasts in the levels of B will be replaced by $ms(BD)$.

Example 17.9.1 `Temperature experiment`

The temperature experiment was run by M. Bowe, J. Cooper, J. Donato, S. Giust, and H. Schieman in 1994 to compare the times required for three different digital thermometers (factor A at $a = 3$ levels) to register body temperature at two different sites—in the mouth and under the arm—(factor B at $b = 2$ levels). Thus, there were six treatment combinations. Four subjects were selected at random from the American statistics graduate students at The Ohio State University, and each treatment combination was measured once for each subject. The experiment was designed as a randomized complete block design, with subjects representing blocks. The recorded times are shown in Table 17.10.

The four subjects used in the experiment are not themselves of interest. Of more interest is how the thermometers react on average over a large population of subjects. The population of American statistics graduate students at the university is large, but not infinite. However, the four subjects used in the experiment are, hopefully, representative of all possible American graduate students, and it is reasonable to model the subject (block) effect as a random effect.

Since subjects vary in body heat, it is possible that factor B (site) might interact with subject. It is also possible that different thermometers might act differently at the two different sites. Consequently the following model might be reasonable for this experiment.

$$Y_{hij} = \mu + S_h + \alpha_i + \beta_j + (\alpha\beta)_{ij} + (S\beta)_{hj} + \epsilon_{hij}\,, \tag{17.9.25}$$
$$h = 1, 2, 3, 4, \quad i = 1, 2, 3, \quad j = 1, 2,$$
$$S_h \sim N(0, \sigma_S^2), \quad (S\beta)_{hj} \sim N(0, \sigma_{SB}^2), \quad \epsilon_{hij} \sim N(0, \sigma^2)\,,$$
$$S_h\text{'s}, \ (S\beta)_{hj}\text{'s and } \epsilon_{hij}\text{'s are all mutually independent}\,, \tag{17.9.26}$$

where all random variables on the right-hand side of the model are mutually independent, and where S_h represents the effect of the hth randomly selected subject (block), α_i represents the effect of the ith specifically selected thermometer, and β_j represents the effect of the ith specifically selected site. This model is similar to mixed model (17.7.24) with S_h replacing D_k. Consequently, the expected mean squares will be similar to those in Table 17.8, p. 643. The analysis of variance table is shown in Table 17.11.

Table 17.10 Data (in seconds) for the temperature experiment

Subject	Treatment combination					
	11	12	21	22	31	32
1	62.16	61.53	154.42	310.46	95.98	225.65
2	65.63	63.70	132.30	284.64	98.50	241.63
3	63.12	61.34	105.52	315.61	110.05	364.07
4	61.51	61.54	94.88	294.16	107.93	304.58

Table 17.11 Analysis of variance table for the mixed model temperature experiment

Source of variation	Degrees of freedom	Mean square	p-value	Expected mean square
Subject (block)	3	570.04	–	–
Thermometer (A)	2	52879.34	0.0001	$Q(A, AB) + \sigma^2$
Site (B)	1	86029.60	0.0035	$Q(B, AB) + 3\sigma^2_{SB} + \sigma^2$
Therm*Site (AB)	2	21897.23	0.0001	$Q(AB) + \sigma^2$
Subject*Site (SB)	3	1210.67	0.2625	$3\sigma^2_{SB} + \sigma^2$
Error	12	802.57		σ^2
Total	23			

We start by testing the two interaction hypotheses. To test the hypothesis $H_0^{SB} : \{\sigma^2_{SB} = 0\}$, that the subject by site interaction variance is negligible, against the alternative hypothesis that it is not negligible, using a significance level of 0.01 (so that the overall significance level will be at most 0.05), we

$$\text{reject } H_0^{SB} \text{ if } \frac{ms(SB)}{msE} > F_{3,12,0.01} = 5.95\,.$$

Since $ms(SB)/msE = 1.51$, there is not sufficient evidence to conclude that the interaction variance is greater than zero (equivalently, the p-value is greater than 0.01). Before we can examine the site main effect, however, we also need to look at the thermometer by site interaction.

To test the hypothesis

$$H_0^{AB} : \{(\alpha\beta)_{ij} - (\alpha\beta)_{ip} - (\alpha\beta)_{uj} + (\alpha\beta)_{up},\ \text{for all } i, j, u, p\}$$

against the alternative hypothesis that the interaction is not negligible, we

$$\text{reject } H_0^{AB} \text{ if } \frac{ms(AB)}{msE} > F_{2,12,0.01} = 6.93\,.$$

Since $ms(AB)/msE = 27.28$, we reject H_0^{AB} and conclude that there is a thermometer×site interaction. Thus, it is unlikely that the thermometer and site main effects are of interest. However, for illustration purposes, we ask whether the *average* time taken for these three digital thermometers to register is the same whether used in the mouth or under the arm. Thus, we will test the hypothesis

$$H_0^B : \{\beta_1 + \overline{(\alpha\beta)}_{.1} = \beta_2 + \overline{(\alpha\beta)}_{.2}\}\,.$$

To test this hypothesis at significance level 0.01, we

$$\text{reject } H_0^B \text{ if } \frac{msB}{ms(SB)} > F_{1,3,0.01} = 29.5\,.$$

Since $msB/ms(SB) = 71.06$, we reject H_0^B and conclude that it does make a difference in registering temperature (on average for these three thermometers) as to whether the thermometer is used in the mouth or under the arm. This conclusion is made on average over the three thermometers and over the whole population of similar graduate students. □

17.10 Using SAS Software

Section 17.10.1 illustrates the use of SAS software to check model assumptions on random effects. Then in Sect. 17.10.2, the analysis of mixed models using `PROC GLM` and `PROC MIXED` is illustrated, followed by an example of analysis of covariance to deal with a quadratic time trend. Finally in Sect. 17.10.3, the SAS `DATA` step and functions are used to do the sample size calculations of Example 17.4.1.

17.10.1 Checking Assumptions on the Model

Using the data of Table 17.1, p. 617, for the clean wool experiment, we illustrate some methods of checking model assumptions for a random-effects one-way model. The experimenters took observations on $r = 4$ cores of wool from each of $v = 7$ randomly selected wool bales.

We let the random variable T_i represent the true clean content of the ith randomly selected bale of wool from the shipment, and let $Y_{it} = T_i + \epsilon_{it}$ represent the observed clean content of the tth core (observation) from the ith bale, where the error variable ϵ_{it} includes the deviation from the true average clean content of the tth core from the ith bale, the measurement error, environmental conditions, etc.

First, we check the error assumptions by calculating and plotting the standardized residuals obtained as though the bale effects were fixed. The standardized residuals are calculated in the usual way and plotted against the levels of the treatment factor and the predicted values (see Sect. 5.8, p. 119). The latter plot, obtained by `PROC SGPLOT`, is shown in Fig. 17.3.

The most noticeable feature is that bale 1 gives rise to one very large standardized residual (an outlier). This means one of several things: Perhaps the data value is in error, so that this value is an outlier, or perhaps bale 1 is extremely more variable than the other bales in the population, or perhaps the error variables are not normally distributed. Let us suppose that we could go back to the original experimenters and that indeed, something unusual happened at this point during the time at which the observations were taken. If so, we could exclude this value. The new residual plot is shown in Fig. 17.4.

All standardized residuals now lie within the expected range for normally distributed errors. The plot gives us quite a lot of information about our sample of bales and possibly about the shipment of bales from which they were drawn. First, the average clean content of bale 1 is around 53, considerably

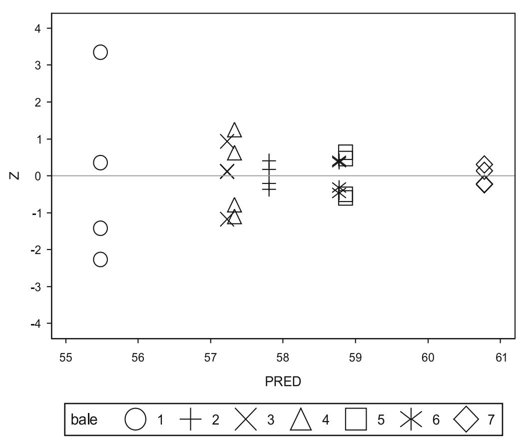

Fig. 17.3 Residuals versus predicted values by bale type for the clean wool experiment

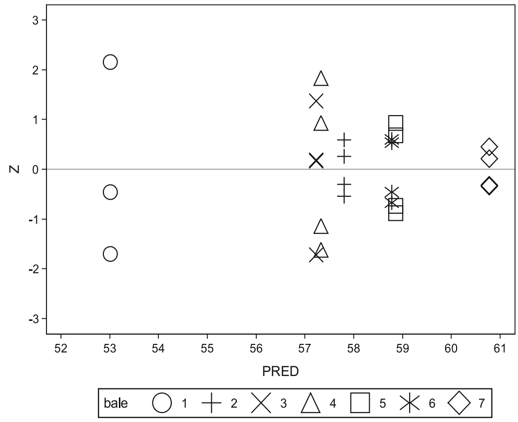

Fig. 17.4 Residuals versus predicted values for the clean wool experiment, excluding the outlier

Table 17.12 SAS program to plot standardized treatment averages against their normal scores

```
PROC SORT; BY BALE;
PROC MEANS NOPRINT; BY BALE;
  VAR CONTENT;
  OUTPUT OUT = WOOL2 MEAN = AVCONT;
PROC STANDARD STD = 1.0 MEAN = 0.0;
  VAR AVCONT;
PROC RANK NORMAL = BLOM;
  VAR AVCONT;
  RANKS NSCORE;
PROC SGPLOT;
  SCATTER X = NSCORE Y = AVCONT / GROUP = BALE;
```

below the others. This was the bale that had the supposed outlier. One might suspect that this bale either did not come from the same shipment or was contaminated at some point before being measured. On the other hand, the shipment may contain a number of "rogue bales," and this ought to be investigated. At the other end of the range, we see that bale 7 had the highest clean content and was least variable. Perhaps this is not too surprising, since a bale with 100% clean content would probably show no variability in the measurements taken on it. Thus, one might suspect that our model that includes normally distributed errors is not ideal for this situation. However, the plot of standardized residuals against normal scores does not show any anomalies (figure not shown).

In a one-way random-effects model, we can check the assumption that the treatment effects have a normal distribution by making a normal probability plot of the standardized treatment averages $\overline{Y}_{i.}$ against their normal scores. (This cannot be done for models with more than one random effect, since the treatment averages are not independent.) For the clean wool experiment, the normal probability plot is obtained by means of the statements in Table 17.12, and the resulting plot is shown in Fig. 17.5. If the normality assumption for the population of bales is satisfied, the standardized bale averages should roughly lie along a line (with slope 1.0) through (0, 0). In Fig. 17.5, we see that this is roughly the case.

In summary, the random-effects one-way model with the standard distribution assumptions does not fit these data too well, since variances apparently are not constant or there is an outlier. Nevertheless, we have established that the population of bales in this shipment is extremely variable. Selected bales 1 and 7 appear to be somewhat different from the other five selected bales. Perhaps the shipment is

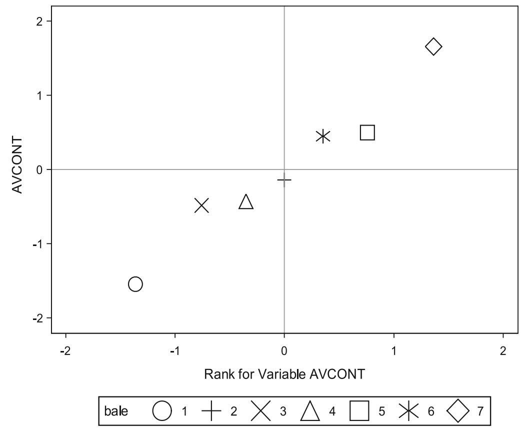

Fig. 17.5 Normal probability plot of the standardized treatment averages for the clean wool experiment

Table 17.13 SAS program for the temperature experiment

```
DATA TEMPR;
  INPUT THERM SITE SUBJ TIME;
  LINES;
  1 1 1  62.16
  1 2 1  61.53
  : : :     :
  3 2 4 304.58
;
PROC GLM;
  ODS EXCLUDE LSMeanCL;
  CLASS THERM SITE SUBJ;
  MODEL TIME = SUBJ THERM SITE THERM*SITE SUBJ*SITE;
  RANDOM SUBJ SUBJ*SITE / TEST;
  CONTRAST 'SITE1-SITE2' SITE 1 -1 / E = SUBJ*SITE;
  LSMEANS SITE / CL PDIFF E = SUBJ*SITE;
```

made up of dissimilar subpopulations (perhaps from different sources). This should be checked, since it may give a clue as to how to improve the wool clean content in the future.

17.10.2 Estimation and Hypothesis Testing

PROC GLM

Analysis of variance tables for random-effects and mixed models are obtained using `PROC GLM` in exactly the same way as for fixed-effects models. The additional expected mean squares column can be obtained very easily by inserting a `RANDOM` statement immediately after the model statement. All random effects should be listed in the `RANDOM` statement, as shown, for example, in Table 17.13 for the temperature experiment of Example 17.9.1, p. 650. The denominators, calculated as explained throughout this chapter, can be obtained by adding the option `TEST` to the `RANDOM` statement, as shown in the following example. The actual denominators are printed out as well as the p-values.

The output is shown in Fig. 17.6. The first few lines reproduce the expected mean squares that were calculated by hand in Table 17.7, p. 641. The remainder of the output gives the `TYPE III` sums of squares, but instead of calculating the usual test ratios with *msE* as the denominator, the `TEST` option on the `RANDOM` statement has caused the denominator *ms*(subj×site) to be used where appropriate.

All pairwise comparisons of fixed effects can be estimated using the `LSMEANS` statement, which allows the correct variance estimator to be specified. For example, the statement

```
LSMEANS SITE / CL PDIFF E = SUBJ*SITE;
```

will use the subj×site interaction mean square as the variance estimate for comparing sites, rather than the error mean square. The `LSMEANS` statement provides a confidence interval for the pairwise comparison of sites, and also a *p*-value for testing equality of the two site effects. The `ODS` statement is used to exclude printing of the confidence intervals (limits) for site level means, since the standard errors would be incorrectly estimated using *ms*(subj×site); the reader is asked in Exercise 9 to verify that $\text{Var}(\overline{Y}_{.\,.j}) = (3\sigma_S^2 + 3\sigma_{SB}^2 + \sigma^2)/12$ for the *j*th site under model (17.9.25), p. 650.

Any contrast for the fixed effects can be estimated as usual using the `ESTIMATE` statement, though the standard error is computed using *MSE* whether or not this is appropriate. Confidence intervals can be calculated by hand and the mean squared error replaced by the denominator used in the test procedures if necessary. For testing individual contrasts, the `CONTRAST` statement can be used and the required denominator can be specified. For example, the statement

```
CONTRAST 'SITE1-SITE2' SITE 1 -1 / E = SUBJ*SITE;
```

will use the subj×site interaction mean square as the variance estimate for comparing sites as appropriate.

The deficiencies of `PROC GLM` in computing standard errors are overcome by `PROC MIXED`, introduced next.

PROC MIXED

The SAS package includes an alternative procedure `PROC MIXED`, explicitly designed to cope with random-effects and mixed models. The statements that generate the same set of information as in Fig. 17.6 are

```
PROC MIXED METHOD = TYPE3;
  CLASS THERM SITE SUBJ;
  MODEL TIME = THERM SITE THERM*SITE / DDFM = SAT;
  RANDOM SUBJ SUBJ*SITE;
  LSMEANS SITE / CL DIFF;
```

For `PROC MIXED`, the `MODEL` statement only contains fixed effects, with random effects specified in the `RANDOM` statement. The option `METHOD=TYPE3` causes the model to be fit by the method of least squares, as used by `PROC GLM`. Estimates of the variance components are also calculated by this procedure and, under the option `METHOD=TYPE3`, they match those one could compute by hand from `GLM` output. An important advantage of `PROC MIXED` is that standard errors of means and contrasts are automatically correctly estimated under mixed models, even if composite estimates are needed, as is the case here for site treatment means, for example. The `DDFM=SAT` option in the `MODEL` statement causes Satterthwaite's approximation to be used to compute degrees of freedom associated with any composite variance estimates, as may be needed for either *F*-statistic denominators or standard error estimates.

Throughout this chapter we have discussed the analysis of mixed and random effects models using the *analysis of variance approach*—namely, fitting the model by least squares as if all effects were fixed, obtaining a corresponding analysis of variance table, then using the expected mean squares to obtain unbiased estimates of the variance components and to determine appropriate *F*-statistic denominators and standard error estimates. The variance component estimates so obtained are called *analysis of variance estimates*. For balanced data under normality, the least squares estimates of any

Fig. 17.6 SAS software analysis of variance for the temperature experiment

Results Viewer - sashtml.htm

The GLM Procedure

Source	Type III Expected Mean Square
SUBJ	Var(Error) + 3 Var(SITE*SUBJ) + 6 Var(SUBJ)
THERM	Var(Error) + Q(THERM,THERM*SITE)
SITE	Var(Error) + 3 Var(SITE*SUBJ) + Q(SITE,THERM*SITE)
THERM*SITE	Var(Error) + Q(THERM*SITE)
SITE*SUBJ	Var(Error) + 3 Var(SITE*SUBJ)

The GLM Procedure
Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: TIME

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
	SUBJ	3	1710.122846	570.040949	0.47	0.7240
*	SITE	1	86030	86030	71.06	0.0035
	Error: MS(SITE*SUBJ)	3	3632.016912	1210.672304		
* This test assumes one or more other fixed effects are zero.						

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
*	THERM	2	105759	52879	65.89	<.0001
	THERM*SITE	2	43794	21897	27.28	<.0001
	SITE*SUBJ	3	3632.016912	1210.672304	1.51	0.2625
	Error: MS(Error)	12	9630.819867	802.568322		
* This test assumes one or more other fixed effects are zero.						

estimable fixed effects are best linear unbiased estimates, and the analysis of variance estimates of variance components are minimum variance unbiased estimates, but the variance component estimates can be negative, as happens in the above example where the subjects variance component estimate is $\hat{\sigma}_S^2 = [ms(\text{SUBJ}) - ms(\text{SITE*SUBJ})]/6 \approx -106.77$.

There are other more sophisticated statistical methods for estimating variance components that prevent the estimates from ever being negative, and that are generally preferable for unbalanced data. One such approach involves estimation of the variance components by *restricted maximum likelihood* (ReML). This approach, implemented in `PROC MIXED`, will be discussed in Chap. 19.

Covariates

Before leaving this section, we will examine a more complicated model. The plot of standardized residuals against order of observation by flavor for the ice cream experiment (Example 17.3.1, p. 621) is shown in Fig. 17.7. This plot suggests that there may be a quadratic time trend in the data.

We define two extra variables `X` and `X2` in the `DATA` statement as follows,

```
DATA ICE;
  INPUT FLAVOR MELTTIME ORDER;
  X=ORDER-16.5;
  X2=X*X;
  LINES;
  1 924 1
  :  :  :
```

and we add these variables to the model statement, so the code for PROC GLM becomes

```
PROC GLM;
  CLASS FLAVOR;
  MODEL MELTTIME = X X2 FLAVOR;
  RANDOM FLAVOR / TEST;
```

The variable X is just the same as ORDER, except that we have subtracted the average order 16.5. This helps to reduce computational problems in the model fitting. The Type III sums of squares and the expected mean squares are shown in Fig. 17.8. We see that the quadratic effect of time order is quite substantial and that from the list of expected mean squares, our estimate of the variance of melting times due to flavor (var(FLAVOR)) must be calculated as

$$\hat{\sigma}_T^2 = \frac{94179.139 - 4497.426}{9.6478} = 9359.62 \text{ seconds}^2$$

or $\hat{\sigma}_T = 96.75$ seconds, which is a little larger than the estimate of $\hat{\sigma}_T = 85.13$ seconds that we obtained in Example 17.3.4, p. 627. Examination of the residuals in the new model shows that the error assumptions are fairly well satisfied. In Exercise 8, the reader is asked to recalculate the confidence intervals for σ_T^2 and σ_T^2/σ^2 using the new model.

17.10.3 Sample Size Calculations

In this section, a program in the SAS DATA step is used to do the sample size calculations of Example 17.4.1, p. 630, for the ice cream experiment. Recall, in testing the hypothesis $H_0^{\gamma T} : \sigma_T^2 \leq \sigma^2$ against the hypothesis $H_A^{\gamma T} : \sigma_T^2 > \sigma^2$ at a significance level of $\alpha = 0.05$, the goal is to be able to reject the null hypothesis with probability $\pi = 0.95$ if the true value of σ_T^2/σ^2 is at least $\Delta = 2.0$. In Example 17.4.1, trial and error was used to determine the number v of ice cream flavors we should look at if we took $r = 11$ or $r = 3$ observations on each. Recall, to achieve the desired power for given r, we need to find v such that

$$(F_{v-1,v(r-1),.05})(F_{v(r-1),v-1,.05}) \leq (2r + 1)/(r + 1)\,.$$

The calculations are illustrated in the SAS program in Table 17.14, and the corresponding output is displayed in Fig. 17.9. In the SAS program, for each value $r = 11$ and $r = 3$, the SAS function FINV

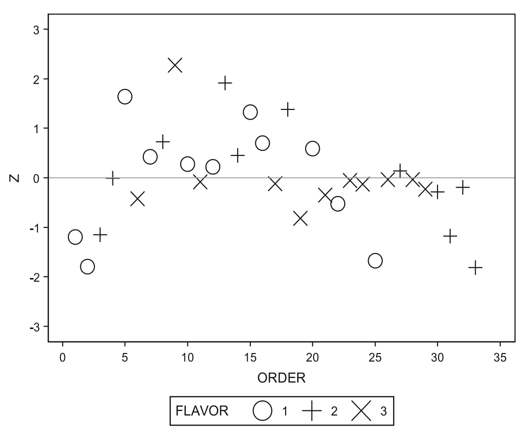

Fig. 17.7 Plot of the standardized residuals against order of observation by flavor for the ice cream experiment

Fig. 17.8 SAS software analysis of variance for the ice cream experiment

Results Viewer - sashtml.htm

The GLM Procedure
Dependent Variable: MELTTIME

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	250538.1210	62634.5302	13.93	<.0001
Error	28	125927.9396	4497.4264		
Corrected Total	32	376466.0606			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X	1	6923.4352	6923.4352	1.54	0.2250
X2	1	66292.6430	66292.6430	14.74	0.0006
FLAVOR	2	188394.2787	94197.1394	20.94	<.0001

Source	Type III Expected Mean Square
X	Var(Error) + Q(X)
X2	Var(Error) + Q(X2)
FLAVOR	Var(Error) + 9.6478 Var(FLAVOR)

Table 17.14 SAS program doing sample size calculations for the ice cream experiment

```
DATA POWER;
  ALPHA=0.05; POWER=0.95;
  DO R=11,3;
    RATIO = (2*R+1)/(R+1);
    RESULT="Need more data";
    V=1;
    DO WHILE (RESULT="Need more data" and V < 201);
      V=V+1;
      DF1=V-1; DF2=V*(R-1);
      F1 = FINV(1-ALPHA,DF1,DF2); * Compute F(DF1,DF2,ALPHA);
      F2 = FINV(POWER,DF2,DF1);   * Compute F(DF2,DF1,1-POWER);
      PRODUCT=F1*F2;
      IF PRODUCT < RATIO THEN RESULT="Enough data";
    END;
    OUTPUT; * Output results to data set;
  END; * End R loop;
;
PROC PRINT;
  VAR R V RESULT POWER ALPHA F1 F2 PRODUCT RATIO DF1 DF2;
```

is used to compute the quantile values $F_{v-1,v(r-1),.05}$ and $F_{v(r-1),v-1,.05}$ for each value $v = 2, 3, \ldots,$ either stopping when the above inequality is satisfied, giving the required value of v and the result "Enough data", or reaching $v = 200$ and giving the result "Need more data". The results displayed in Fig. 17.9 are a bit more precise than those given in Example 17.4.1.

Results Viewer - SAS Output

The SAS System

Obs	R	V	RESULT	POWER	ALPHA	F1	F2	PRODUCT	RATIO	DF1	DF2
1	11	58	Enough data	0.95	0.05	1.35009	1.41935	1.91625	1.91667	57	580
2	3	106	Enough data	0.95	0.05	1.31132	1.33137	1.74585	1.75000	105	212

Fig. 17.9 SAS software sample size calculations for the ice cream experiment

17.11 Using R Software

Section 17.11.1 illustrates the use of R to check model assumptions on random effects. Then in Sect. 17.11.2, the analysis of fixed effects in mixed models using `aov` is illustrated. Finally in Sect. 17.11.3, an R function is defined to do the sample size calculations of Example 17.4.1.

17.11.1 Checking Assumptions on the Model

Using the data of Table 17.1, p. 617, for the clean wool experiment, we illustrate some methods of checking model assumptions for a random-effects one-way model. The experimenters took observations on $r = 4$ cores of wool from each of $v = 7$ randomly selected wool bales.

We let the random variable T_i represent the true clean content of the ith randomly selected bale of wool from the shipment, and let $Y_{it} = T_i + \epsilon_{it}$ represent the observed clean content of the tth core (observation) from the ith bale, where the error variable ϵ_{it} includes the deviation from the true average clean content of the tth core from the ith bale, the measurement error, environmental conditions, etc.

First, we check the error assumptions by calculating and plotting the standardized residuals obtained as though the bale effects were fixed. The standardized residuals are calculated in the usual way and plotted against the levels of the treatment factor and the predicted values (see Sect. 5.9, p. 126). The latter plot, obtained from the statements

```
plot(z ~ ypred, data=wool.data, ylab="Standardized Residuals", las=1,
     type="n") # Suppress plotting of circles
  text(z ~ ypred, bale, cex=0.75, data=wool.data) # Plot bale number
  mtext("Plotting symbol is bale", side=3, adj=1, line=1) # Margin text
  abline(h=0)
```

is shown in Fig. 17.10. The plotted symbols indicate the bales from which the residuals arose.

The most noticeable feature is that bale 1 gives rise to one very large standardized residual (an outlier). This means one of several things: Perhaps the data value is in error, so that this value is an outlier, or perhaps bale 1 is extremely more variable than the other bales in the population, or perhaps the error variables are not normally distributed. Let us suppose that we could go back to the original experimenters and that indeed, something unusual happened at this point during the time at which the observations were taken. If so, we could exclude this value. The new residual plot is shown in Fig. 17.11.

All standardized residuals now lie within the expected range for normally distributed errors. The plot gives us quite a lot of information about our sample of bales and possibly about the shipment of bales from which they were drawn. First, the average clean content of bale 1 is around 53, considerably below the others. This was the bale that had the supposed outlier. One might suspect that this bale either did not come from the same shipment or was contaminated at some point before being measured. On the other hand, the shipment may contain a number of "rogue bales," and this ought to be investigated. At the other end of the range, we see that bale 7 had the highest clean content and was least variable. Perhaps this is not too surprising, since a bale with 100% clean content would probably show no variability in the measurements taken on it. Thus, one might suspect that our model that includes normally distributed errors is not ideal for this situation. However, the plot of standardized residuals against normal scores does not show any anomalies (figure not shown).

In a one-way random-effects model, we can check the assumption that the treatment effects have a normal distribution by making a normal probability plot of the standardized treatment averages $\overline{Y}_{i.}$ against their normal scores. (This cannot be done for models with more than one random effect, since the treatment averages are not independent.) For the clean wool experiment, the normal probability plot is obtained by means of the statements in Table 17.15, and the resulting plot is shown in Fig. 17.12. If the normality assumption for the population of bales is satisfied, the standardized bale averages should roughly lie along a line (with slope 1.0) through (0, 0). In Fig. 17.12, we see that this is roughly the case.

In summary, the random-effects one-way model with the standard distribution assumptions does not fit these data too well, since variances apparently are not constant or there is an outlier. Nevertheless, we have established that the population of bales in this shipment is extremely variable. Selected bales 1 and 7 appear to be somewhat different from the other five selected bales. Perhaps the shipment is

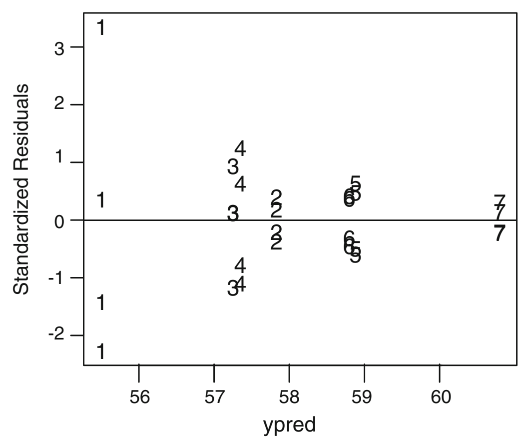

Fig. 17.10 Residuals versus predicted values for the clean wool experiment

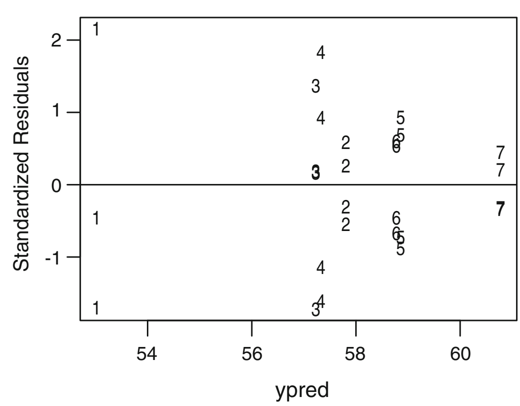

Fig. 17.11 Residuals versus predicted values for the clean wool experiment, excluding the outlier

Table 17.15 R program to plot standardized treatment averages against their normal scores

```
AvgContent = by(wool.data$y, wool.data$bale, mean)  # Col of means (by bale)
# Standardized treatment means
StdzdAvg = (AvgContent - mean(AvgContent))/sd(AvgContent - mean(AvgContent))
# Normal scores
nscore = qqnorm(StdzdAvg)$x
plot(StdzdAvg ~ nscore, ylab="Standardized Trtmt Means", las=1)
  qqline(StdzdAvg)  # Line through 1st and 3rd quantile points
```

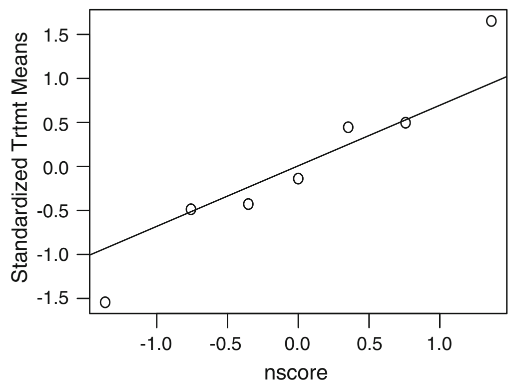

Fig. 17.12 Normal probability plot of the standardized treatment averages for the clean wool experiment

made up of dissimilar subpopulations (perhaps from different sources). This should be checked, since it may give a clue as to how to improve the wool clean content in the future.

17.11.2 Estimation and Hypothesis Testing

Analysis of Fixed Effects in Mixed Models

Analysis of variance F-tests for fixed effects in mixed models are obtained in essentially the same way as for fixed-effects models. The `aov` function is used to fit the linear model by least squares, but with the following changes: random effects are entered into the model as error terms, and the `summary` command is used to generate the analysis of variance F tests for fixed effects.

Sample R code is shown in Table 17.16 for the temperature experiment of Example 17.9.1, p. 650. F-tests for fixed effects are generated by the second block of code. In the model specified for the `aov` function, the term

```
Error(fSubj + fSubj:fSite)
```

calls the `Error` function in R, modeling subject main effects and subject–site interactions as random effects, and allowing their use as error terms for tests and for estimating standard errors for confidence intervals for means and fixed effect contrasts as appropriate. Each model can include only one `Error` function call, but the call may include multiple random effects, with two in this example. The expected mean squares are not displayed, but correct F-tests are generated for fixed effects. In particular, the denominators, calculated as explained throughout this chapter, are automatically obtained by modeling the random effects as error terms as in this sample code. Tests for fixed effects are provided, including the appropriate test statistic denominator used.

Table 17.16 R program for the temperature experiment

```
tempr.data = read.table("data/temperature.txt", header=T)
tempr.data = within(tempr.data, {
   fTherm = factor(Therm); fSite = factor(Site); fSubj = factor(Subj) })
head(tempr.data, 3)

# Least squares anova, specifying random effects as error terms
options(contrasts = c("contr.sum", "contr.poly"))
model1 = aov(Time ~ fTherm + fSite + fTherm:fSite
                   + Error(fSubj + fSubj:fSite), data=tempr.data)
summary(model1)

# Means and contrasts: estimates, CIs, tests
library(lsmeans)
lsmSite = lsmeans(model1, ~ fSite) # Compute and save lsmeans
confint(lsmSite, level=0.95) # Display lsmeans and 95% CIs
# Pairwise comparison
summary(contrast(lsmSite, method="pairwise"),
        infer=c(T,T), level=0.95, side="two-sided")
```

Corresponding output is shown in Table 17.17. The first block of output code shows information for subjects, though it is not used for any tests. The second block of output displays the F test for the main effect of site, using the mean square for subject–site interaction as the denominator of the F statistic. In the third block of output, thermometer main effects and thermometer–site interactions are each tested using the usual mean squared error as the denominator of the corresponding F statistics.

Treatment means and pairwise comparisons of fixed effects can be estimated using the `lsmeans` function of the `lsmeans` package, as illustrated by the last block of code in Table 17.16. The correct variance estimators are automatically chosen. For example, for confidence limits for site level means, the reader is asked in Exercise 9 to verify that $\text{Var}(\overline{Y}_{.j}) = (3\sigma_S^2 + 3\sigma_{SB}^2 + \sigma^2)/12$ for the jth site under model (17.9.25), p. 650. R computes the corresponding estimate, using Satterthwaite's method to compute the corresponding number of degrees of freedom, and provides corresponding lower and upper 95% confidence limits.

The last two lines of code in Table 17.16 concern the site main-effect contrast, providing information for estimation and testing this pairwise comparison. The corresponding output, displayed last in Table 17.17, correctly uses the subject–site interaction mean square to estimate the standard error for comparing sites. If site had more than two levels, one could apply Tukey's method, for example, by including the option `adjust="tukey"` in the `contrast` function as follows.

```
summary(contrast(lsmSite, method="pairwise", adjust="tukey"),
        infer=c(T,T), level=0.95, side="two-sided")
```

Also, any contrast for the fixed effects can be estimated using the following more generic syntax, and the list could be expanded to included multiple named contrasts separated by commas.

```
summary(contrast(lsmSite, list(Pairwise=c( 1, -1))),
        infer=c(T,T), level=0.95, side="two-sided
```

In each case, R will automatically compute an appropriate standard error estimate, which for the pairwise site contrast involves the subject–site interaction mean square.

Table 17.17 R analysis of fixed effects for the temperature experiment

```
> model1 = aov(Time ~ fTherm + fSite + fTherm:fSite
+              + Error(fSubj + fSubj:fSite), data=tempr.data)
> summary(model1)

Error: fSubj
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  3   1710     570

Error: fSubj:fSite
          Df Sum Sq Mean Sq F value Pr(>F)
fSite      1  86030   86030    71.1 0.0035
Residuals  3   3632    1211

Error: Within
             Df Sum Sq Mean Sq F value  Pr(>F)
fTherm        2 105759   52879    65.9 3.4e-07
fTherm:fSite  2  43794   21897    27.3 3.4e-05
Residuals    12   9631     803

> # Means and contrasts: estimates, CIs, tests
> library(lsmeans)
> lsmSite = lsmeans(model1, ~ fSite) # Compute and save lsmeans
NOTE: Results may be misleading due to involvement in interactions
> confint(lsmSite, level=0.95) # Display lsmeans and 95% CIs

 fSite lsmean     SE   df lower.CL upper.CL
 1      96.00 8.6137 5.31   74.244   117.76
 2     215.74 8.6137 5.31  193.986   237.50

Results are averaged over the levels of: fTherm
Confidence level used: 0.95

> # Pairwise comparison
> summary(contrast(lsmSite, method="pairwise"),
+          infer=c(T,T), level=0.95, side="two-sided")

 contrast estimate     SE df lower.CL upper.CL t.ratio p.value
 1 - 2     -119.74 14.205  3  -164.95  -74.536   -8.43  0.0035

Results are averaged over the levels of: fTherm
Confidence level used: 0.95
```

Analysis of Random Effects

Tests for random effects are not provided by the code discussed above for analysis fixed effects in mixed models. Still, the necessary means squares are provided, or they can be obtained by fitting a model treating all effects, including random effects, as fixed. So, based on expected mean squares, one could conduct the appropriate tests of random effects by hand for any balanced mixed- or random-effects model. The following section mentions another approach.

Other Approaches

Throughout this chapter we have discussed the analysis of mixed and random effects models using the *analysis of variance approach*—namely, fitting the model by least squares as if all effects were fixed, obtaining a corresponding analysis of variance table, then using the expected mean squares to obtain unbiased estimates of the variance components and to determine appropriate F-statistic denominators and standard error estimates. The variance component estimates so obtained are called *analysis of variance estimates*. For balanced data under normality, the least squares estimates of any estimable fixed effects are best linear unbiased estimates, and the analysis of variance estimates of variance components are minimum variance unbiased estimates, but the variance component estimates can be negative, as happens in the recently discussed temperature experiment, where the subjects variance component estimate is $\hat{\sigma}_S^2 = [ms(\text{fSubj}) - ms(\text{fSubj:fSite})]/6 \approx -106.83$.

There are other more sophisticated statistical methods for estimating variance components that prevent the estimates from ever being negative, and that are generally preferable for unbalanced data. One such approach involves estimation of the variance components by *restricted maximum likelihood* (ReML). This approach will be discussed in Chap. 18.

17.11.3 Sample Size Calculations

In this section, an R function is defined to do the sample size calculations of Example 17.4.1, p. 630, for the ice cream experiment. Recall, in testing the hypothesis $H_0^{\gamma T} : \sigma_T^2 \leq \sigma^2$ against the hypothesis $H_A^{\gamma T} : \sigma_T^2 > \sigma^2$ at a significance level of $\alpha = 0.05$, the goal is to be able to reject the null hypothesis with probability $\pi = 0.95$ if the true value of σ_T^2/σ^2 is at least $\Delta = 2.0$. In Example 17.4.1, trial and error was used to determine the number v of ice cream flavors we should look at if we took $r = 11$ or $r = 3$ observations on each. Recall, to achieve the desired power for given r, we need to find v such that

$$(F_{v-1,v(r-1),.05})(F_{v(r-1),v-1,.05}) \leq (2r+1)/(r+1)\,.$$

Table 17.18 contains the R program and corresponding output. A user-defined function, `compute.v.given.r`, is defined to do the computations. This function inputs an r value and the specified significance level and power. Given this information, the new function uses the standard R function `qf` to compute the quantile values $F_{v-1,v(r-1),.05}$ and $F_{v(r-1),v-1,.05}$ for each value $v = 2, 3, \ldots$, either stopping when the above inequality is satisfied, giving the required value of v and the result "Power = 0.95" (for power = 0.95, say), or reaching $v = 200$ and giving the result "Power < 0.95". Once done, the function returns all pertinent information. After the new function `compute.v.given.r` has been defined in Table 17.18, it is called first for $r = 11$ and then for $r = 3$, each time using significance level 0.05 and power 0.95. The results returned by the function calls, displayed in the bottom of Table 17.18, are a bit more precise than those given in Example 17.4.1.

Table 17.18 R function doing sample size calculations for the ice cream experiment

```
> # Create a user-defined function for sample size calculations
> compute.v.given.r = function(r=5,alpha=0.05,power=0.95){
+     # Initialize variables
+     ratio = (2*r+1)/(r+1);
+     result="Power <"
+     v=1;
+     while((result=="Power <")&(v<201)){
+         v=v+1;
+         df1=v-1; df2=v*(r-1);
+         F1 = qf(1-alpha,df1,df2) # Compute F(df1,df2,alpha)
+         F2 = qf(power,df2,df1)   # Compute F(df2,df1,1-power)
+         product=F1*F2
+         if (product<ratio) {result="Power ="}
+     } # End while loop, either finding v or power less than target
+     data.frame(r,v,result,power,alpha,F1,F2,product,ratio,df1,df2)
+ } # end function
>
> compute.v.given.r(r=11,alpha=0.05,power=0.95) # Call the function
   r  v  result power alpha     F1     F2 product  ratio df1 df2
1 11 58 Power =  0.95  0.05 1.3501 1.4193  1.9162 1.9167  57 580
>
> compute.v.given.r(r=3,alpha=0.05,power=0.95) # Call the function
  r   v  result power alpha     F1     F2 product ratio df1 df2
1 3 106 Power =  0.95  0.05 1.3113 1.3314  1.7459  1.75 105 212
```

Table 17.19 Data for the alcohol experiment

Bottle	Concentration (mg/ml)			
1	1.4357	1.4348	1.4336	1.4309
2	1.4244	1.4232	1.4213	1.4256
3	1.4153	1.4137	1.4176	1.4164
4	1.4331	1.4325	1.4312	1.4297
5	1.4252	1.4261	1.4293	1.4272
6	1.4179	1.4217	1.4191	1.4204

Exercises

1. **Alcohol experiment**

   Solutions of alcohol are used for calibrating Breathalyzers. The data in Table 17.19 show the alcohol concentrations (mg/ml) of samples of alcohol solutions taken from six bottles of alcohol solution randomly selected from a large batch. Concentrations are determined by gas chromatography.

   (a) Check the assumptions on the random-effects one-way model for these data.
   (b) Calculate a 95% upper confidence bound for the error variance.
   (c) Calculate a 95% confidence interval for the variance of the alcohol concentrations in the population of bottles in this large batch.

(d) Test the hypothesis that the variance of the alcohol concentrations is at most five times the error variance versus the alternative hypothesis that it is not. Use a significance level of $\alpha = 0.05$.

2. **Ice cream experiment, continued**

As in Example 17.4.1, p. 630, suppose the ice cream experiment is to be repeated, with $\gamma = 1.0$ and with a Type I error probability of $\alpha = .05$. Suppose that we would like to reject the null hypothesis $H_0^{\gamma T} : \{\sigma_T^2 \leq \sigma^2\}$ with probability $\pi = 0.95$ if the true value of σ_T^2/σ^2 is greater than $\Delta = 2.0$. How many ice cream flavors should be included in the experiment if $r = 2$ observations are to be taken on each? How many observations are needed? Is this an improvement over the result in the example for $r = 3$? (Note: $F_{150,150,.05} = 1.309$, $F_{160,160,.05} = 1.298$, $F_{170,170,.05} = 1.288$, $F_{180,180,.05} = 1.279$).

3. **Random effects model**

Consider the following random-effects model:

$$\begin{gathered}
Y_{ijkmt} = \mu + A_i + B_j + C_k + D_m \\
\quad + (AB)_{ij} + (BC)_{jk} + (BD)_{jm} + \epsilon_{ijkmt} \,, \\
i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, c, \\
m = 1, \ldots, d, \quad t = 1, \ldots, r, \\
A_i \sim N(0, \sigma_A^2), \quad B_j \sim N(0, \sigma_B^2), \quad C_k \sim N(0, \sigma_C^2)\,, \\
D_m \sim N(0, \sigma_D^2), \quad (AB)_{ij} \sim N(0, \sigma_{AB}^2), \quad (BC)_{jk} \sim N(0, \sigma_{BC}^2)\,, \\
(BD)_{jm} \sim N(0, \sigma_{BD}^2), \quad \epsilon_{ijkmt} \sim N(0, \sigma^2)\,,
\end{gathered}$$

where all random variables on the right hand side of the model are mutually independent.

(a) Write out the expected mean squares for all main effects and interactions in the model.
(b) How would you test the null hypothesis $H_0^A : \{\sigma_A^2 = 0\}$ against the alternative hypothesis $H_A^A : \{\sigma_A^2 > 0\}$?
(c) How would you test the null hypothesis $H_0^B : \{\sigma_B^2 = 0\}$ against the alternative hypothesis $H_A^B : \{\sigma_B^2 > 0\}$?
(d) Give formulae for unbiased estimates of σ_{BD}^2 and σ_B^2.
(e) Give formulae for individual 95% confidence intervals for σ_{BD}^2 and σ_B^2. What is the overall confidence level?

4. **Buttermilk biscuit experiment**

The buttermilk biscuit experiment was run by Stacie Taylor in 1995 to find out which brands of refrigerated buttermilk biscuit give rise to the fluffiest biscuits. Three brands were examined (factor A, 3 levels, fixed effect), all of which had claims to be light, fluffy, or flaky in their advertising campaigns. The biscuits were baked on a baking tray for 7 min in the center of an oven set to 425°F. Since only six biscuits could be baked at a time, the experiment was run as a general complete block design with blocks of size $k = 6$.

(a) Use a mixed model with interaction to represent the data, where the random effect represents the block (run of the oven) and the fixed effect represents the biscuit brand. Write out the model including all of the assumptions.

Table 17.20 Treatments and percentage change in height for the buttermilk biscuit experiment

Block	Position					
	1	2	3	4	5	6
1	2 (150.0)	1 (188.2)	2 (177.8)	3 (166.7)	3 (187.5)	1 (182.4)
2	1 (183.3)	2 (183.3)	2 (183.3)	3 (176.5)	1 (160.0)	3 (187.5)
3	1 (178.9)	3 (182.4)	2 (193.8)	3 (176.5)	2 (188.9)	1 (188.9)
4	2 (177.8)	1 (145.5)	3 (155.0)	1 (173.7)	3 (200.0)	2 (187.5)
5	1 (205.6)	3 (188.2)	3 (142.9)	2 (161.9)	2 (177.8)	1 (159.1)

Table 17.21 Data for the candle experiment (seconds)

Person	Color							
	Red		White		Blue		Yellow	
1	989	1032	1044	979	1011	951	974	998
	1077	1019	987	1031	928	1022	1033	1041
2	899	912	847	880	899	800	886	859
	911	943	879	830	820	812	901	907
3	898	840	840	952	909	790	950	992
	955	1005	961	915	871	905	920	890
4	993	957	987	960	864	925	949	973
	1005	982	920	1001	824	790	978	938

(b) The data collected by the experimenter are shown in Table 17.20. As far as possible, check the assumptions on the model for these data.
(c) Write out the expected mean squares for all terms in the model.
(d) Draw a block×brand interaction plot for those blocks observed in the experiment.
(e) Test the hypothesis that the variance in height of the biscuits due the population of block×brand interactions is negligible against the alternative hypothesis that it is not negligible. Interpret your conclusions in terms of the plot in part (d).
(f) Calculate a set of 95% simultaneous confidence intervals for the pairwise comparisons between the brands. State your conclusions.

5. **Candle experiment**

An experiment to determine whether different colored candles (red, white, blue, yellow) burn at different speeds was conducted by Hsing-Chuan Tsai, Mei-Chiao Yang, Derek Wheeler, and Tom Schultz in 1989. Each experimenter collected four observations on each color in a random order, and "experimenter" was used as a blocking factor. Thus, the design was a general complete block design with $v = 4$, $k = 16$, $b = 4$, and $s = 4$. The resulting burning times (in seconds) are shown in Table 17.21. A pilot experiment indicated that treatments and blocks do interact. The candles used in the experiment were cake candles made by a single manufacturer. Analyze the experiment as though the experimenters represent a random sample from a large population of people who might use these candles in practice. Use a two-way mixed model with interaction.

6. **Golf ball experiment**

An experiment was planned by Tim Kelaghan in 1995 to examine whether different brands of golf balls travel on average the same distances when hit by amateur golfers. The experiment was planned with a specific selection of $v = 3$ golf balls and some number b of golfers to be determined. The experiment was to be run as a general complete block design with fixed treatment effects and random golfer effects. Since the golfer is aware of which brand of ball he or she is hitting, there may well be a golfer×brand interaction. However, the differences between brands averaged over the interaction is important here.
A small pilot experiment was conducted. There were only two golfers, and each hit $s = 6$ balls of each brand in a random order. Mis-hits were ignored. The distances that the balls traveled were recorded in yards and are shown in Table 17.22.

(a) Use the pilot experiment data to calculate a 95% upper bound for the error variance σ^2.
(b) The experimenter wanted the main experiment to be able to calculate a set of simultaneous 95% confidence intervals for the pairwise differences in the brands, and he wanted the widths of these intervals to be at most 20 yards.
Assuming that each golfer would hit about 18 balls in total, as in the pilot experiment, how many randomly selected golfers would be needed?

7. **Mixed model**

Consider the following mixed model:

$$\begin{gathered} Y_{ijkmt} = \mu + \alpha_i + B_j + C_k + \delta_m + (\alpha B)_{ij} + (\alpha\delta)_{im} \\ + (B\delta)_{jm} + (C\delta)_{km} + (\alpha B\delta)_{ijm} + \epsilon_{ijkmt}\,, \\ i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, c, \\ m = 1, \ldots, d, \quad t = 1, \ldots, r, \\ B_j \sim N(0, \sigma_B^2), \quad C_k \sim N(0, \sigma_C^2), \quad (\alpha B)_{ij} \sim N(0, \sigma_{AB}^2), \\ (B\delta)_{jm} \sim N(0, \sigma_{BD}^2), \quad (C\delta)_{km} \sim N(0, \sigma_{CD}^2), \\ (\alpha B\delta)_{ijm} \sim N(0, \sigma_{ABD}^2), \quad \epsilon_{ijkmt} \sim N(0, \sigma^2)\,, \end{gathered}$$

where α_i and δ_m are fixed effects, all other effects are random effects, and all random variables on the right-hand side of the model are mutually independent.

Table 17.22 Distances (in yards) traveled by balls in the golf experiment

Golfer	Brand	Distance					
		1	2	3	4	5	6
1	1	209	204	179	230	233	245
	2	188	211	242	222	187	233
	3	219	204	247	215	197	161
2	1	240	207	192	190	226	188
	2	216	195	240	215	219	238
	3	195	221	205	192	183	230

(a) Write out the expected mean squares for all main effects and interactions in the model.
(b) How would you test the hypothesis $H_0 : \{\delta_m + \overline{(\alpha\delta)}_{.m}$ all equal$\}$ against the alternative hypothesis that these parameters are not all equal?
(c) Give a formula for an unbiased estimate of σ_B^2.
(d) Give a formula for a 95% confidence interval for σ_B^2.

8. **Ice cream experiment, continued**

The ice cream experiment was described in Example 17.3.1, p. 621, and was analyzed in Examples 17.3.2–17.3.4 and 17.4.1. In Sect. 17.10, p. 657, a new model was suggested that involved a quadratic time trend.

(a) What could account for a quadratic time trend?
(b) Investigate the assumptions on the models with and without the quadratic time trend.
(c) Redo the analyses of Examples 17.3.3 and 17.3.4 for the new model and compare your answers with the original model.
(d) Which model do you prefer and why?

9. **Temperature experiment, continued**

The temperature experiment was described and analyzed in Example 17.9.1, with corresponding SAS and R software analysis provided in Sects. 17.10.2 and 17.11.2, respectively. The experiment was to compare the times required for three different digital thermometers (factor A at three levels) to register body temperature at two different sites—in the mouth and under the arm—(factor B at two levels). A randomized complete block design was used, with each of the six treatment combinations observed once on each of four subjects (factor S at four levels). Using the mixed model (17.9.25), p. 650, consider estimation of the treatment mean $\bar{\mu}_{..j} = \mu + \bar{\alpha}_{.} + \beta_j + \overline{(\alpha\beta)}_{.j}$ for the jth site.

(a) Verify that $\text{Var}(\overline{Y}_{..j}) = (3\sigma_S^2 + 3\sigma_{SB}^2 + \sigma^2)/12$.
(b) Provide an unbiased estimate of $\text{Var}(\overline{Y}_{..j})$.
(c) Compute the number of degrees of freedom associated with the estimate in part (b).
(d) Given that $\overline{y}_{..1} = 96.00$, construct a 95% confidence interval for $\bar{\mu}_{..1}$.