

The Analysis of Variance (ANOVA)

While the linear regression of Chapter 1 goes back to the nineteenth century, the Analysis of Variance of this chapter dates from the twentieth century, in applied work by Fisher motivated by agricultural problems (see §2.6). We begin this chapter with some necessary preliminaries, on the special distributions of Statistics needed for small-sample theory: the chi-square distributions $\chi^2(n)$ (§2.1), the Fisher F -distributions $F(m, n)$ (§2.3), and the independence of normal sample means and sample variances (§2.5). We shall generalise linear regression to multiple regression in Chapters 3 and 4 – which use the Analysis of Variance of this chapter – and unify regression and Analysis of Variance in Chapter 5 on Analysis of Covariance.

2.1 The Chi-Square Distribution

We now define the *chi-square distribution* with n *degrees of freedom* (df), $\chi^2(n)$. This is the distribution of

$$X_1^2 + \dots + X_n^2,$$

with the X_i iid $N(0, 1)$.

Recall (§1.5, Fact 9) the definition of the MGF, and also the definition of the *Gamma function*,

$$\Gamma(t) := \int_0^\infty e^{-x} x^{t-1} dx \quad (t > 0)$$

(the integral converges for $t > 0$). One may check (by integration by parts) that

$$\Gamma(n+1) = n! \quad (n = 0, 1, 2, \dots),$$

so the Gamma function provides a continuous extension to the factorial. It is also needed in Statistics, as it comes into the normalisation constants of the standard distributions of small-sample theory, as we see below.

Theorem 2.1

The chi-square distribution $\chi^2(n)$ with n degrees of freedom has

- (i) mean n and variance $2n$,
- (ii) MGF $M(t) = 1/(1-2t)^{\frac{1}{2}n}$ for $t < \frac{1}{2}$,
- (iii) density

$$f(x) = \frac{1}{2^{\frac{1}{2}n} \Gamma(\frac{1}{2}n)} x^{\frac{1}{2}n-1} \exp\left(-\frac{1}{2}x\right) \quad (x > 0).$$

Proof

(i) For $n = 1$, the mean is 1, because a $\chi^2(1)$ is the square of a standard normal, and a standard normal has mean 0 and variance 1. The variance is 2, because the fourth moment of a standard normal X is 3, and

$$\text{var}(X^2) = E[(X^2)^2] - [E(X^2)]^2 = 3 - 1 = 2.$$

For general n , the mean is n because means add, and the variance is $2n$ because variances add over independent summands (Haigh (2002), Th 5.5, Cor 5.6).

(ii) For X standard normal, the MGF of its square X^2 is

$$M(t) := \int e^{tx^2} \phi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(1-2t)x^2} dx.$$

So the integral converges only for $t < \frac{1}{2}$; putting $y := \sqrt{1-2t}.x$ gives

$$M(t) = 1/\sqrt{1-2t} \quad \left(t < \frac{1}{2}\right) \quad \text{for } X \sim N(0, 1).$$

Now when X, Y are independent, the MGF of their sum is the product of their MGFs (see e.g. Haigh (2002), p.103). For e^{tX} , e^{tY} are independent, and the mean of an independent product is the product of the means. Combining these, the MGF of a $\chi^2(n)$ is given by

$$M(t) = 1/(1-2t)^{\frac{1}{2}n} \quad \left(t < \frac{1}{2}\right) \quad \text{for } X \sim \chi^2(n).$$

(iii) First, $f(\cdot)$ is a density, as it is non-negative, and integrates to 1:

$$\begin{aligned} \int f(x) dx &= \frac{1}{2^{\frac{1}{2}n} \Gamma(\frac{1}{2}n)} \int_0^\infty x^{\frac{1}{2}n-1} \exp\left(-\frac{1}{2}x\right) dx \\ &= \frac{1}{\Gamma(\frac{1}{2}n)} \int_0^\infty u^{\frac{1}{2}n-1} \exp(-u) du \quad (u := \frac{1}{2}x) \\ &= 1, \end{aligned}$$

by definition of the Gamma function. Its MGF is

$$\begin{aligned} M(t) &= \frac{1}{2^{\frac{1}{2}n} \Gamma(\frac{1}{2}n)} \int_0^\infty e^{tx} x^{\frac{1}{2}n-1} \exp\left(-\frac{1}{2}x\right) dx \\ &= \frac{1}{2^{\frac{1}{2}n} \Gamma(\frac{1}{2}n)} \int_0^\infty x^{\frac{1}{2}n-1} \exp\left(-\frac{1}{2}x(1-2t)\right) dx. \end{aligned}$$

Substitute $u := x(1-2t)$ in the integral. One obtains

$$M(t) = (1-2t)^{-\frac{1}{2}n} \frac{1}{2^{\frac{1}{2}n} \Gamma(\frac{1}{2}n)} \int_0^\infty u^{\frac{1}{2}n-1} e^{-u} du = (1-2t)^{-\frac{1}{2}n},$$

by definition of the Gamma function. □

Chi-square Addition Property. If X_1, X_2 are independent, $\chi^2(n_1)$ and $\chi^2(n_2)$, $X_1 + X_2$ is $\chi^2(n_1 + n_2)$.

Proof

$X_1 = U_1^2 + \dots + U_{n_1}^2$, $X_2 = U_{n_1+1}^2 + \dots + U_{n_1+n_2}^2$, with U_i iid $N(0, 1)$. So $X_1 + X_2 = U_1^2 + \dots + U_{n_1+n_2}^2$, so $X_1 + X_2$ is $\chi^2(n_1 + n_2)$. □

Chi-Square Subtraction Property. If $X = X_1 + X_2$, with X_1 and X_2 independent, and $X \sim \chi^2(n_1 + n_2)$, $X_1 \sim \chi^2(n_1)$, then $X_2 \sim \chi^2(n_2)$.

Proof

As X is the independent sum of X_1 and X_2 , its MGF is the product of their MGFs. But X, X_1 have MGFs $(1-2t)^{-\frac{1}{2}(n_1+n_2)}$, $(1-2t)^{-\frac{1}{2}n_1}$. Dividing, X_2 has MGF $(1-2t)^{-\frac{1}{2}n_2}$. So $X_2 \sim \chi^2(n_2)$. □

2.2 Change of variable formula and Jacobians

Recall from calculus of several variables the change of variable formula for multiple integrals. If in

$$I := \int \dots \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n = \int_A f(\mathbf{x}) d\mathbf{x}$$

we make a one-to-one change of variables from \mathbf{x} to \mathbf{y} — $\mathbf{x} = \mathbf{x}(\mathbf{y})$ or $x_i = x_i(y_1, \dots, y_n)$ ($i = 1, \dots, n$) — let B be the region in \mathbf{y} -space corresponding to the region A in \mathbf{x} -space. Then

$$I = \int_A f(\mathbf{x}) d\mathbf{x} = \int_B f(\mathbf{x}(\mathbf{y})) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| d\mathbf{y} = \int_B f(\mathbf{x}(\mathbf{y})) |J| d\mathbf{y},$$

where J , the determinant of partial derivatives

$$J := \frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} := \det \left(\frac{\partial x_i}{\partial y_j} \right)$$

is the *Jacobian* of the transformation (after the great German mathematician C. G. J. Jacobi (1804–1851) in 1841 – see e.g. Dineen (2001), Ch. 14). Note that in one dimension, this just reduces to the usual rule for change of variables: $dx = (dx/dy).dy$. Also, if J is the Jacobian of the change of variables $\mathbf{x} \rightarrow \mathbf{y}$ above, the Jacobian $\partial \mathbf{y} / \partial \mathbf{x}$ of the inverse transformation $\mathbf{y} \rightarrow \mathbf{x}$ is J^{-1} (from the product theorem for determinants: $\det(AB) = \det A \cdot \det B$ – see e.g. Blyth and Robertson (2002a), Th. 8.7).

Suppose now that \mathbf{X} is a random n -vector with density $f(\mathbf{x})$, and we wish to change from \mathbf{X} to \mathbf{Y} , where \mathbf{Y} corresponds to \mathbf{X} as \mathbf{y} above corresponds to \mathbf{x} : $\mathbf{y} = \mathbf{y}(\mathbf{x})$ iff $\mathbf{x} = \mathbf{x}(\mathbf{y})$. If \mathbf{Y} has density $g(\mathbf{y})$, then by above,

$$P(\mathbf{X} \in A) = \int_A f(\mathbf{x}) d\mathbf{x} = \int_B f(\mathbf{x}(\mathbf{y})) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| d\mathbf{y},$$

and also

$$P(\mathbf{X} \in A) = P(\mathbf{Y} \in B) = \int_B g(\mathbf{y}) d\mathbf{y}.$$

Since these hold for all B , the integrands must be equal, giving

$$g(\mathbf{y}) = f(\mathbf{x}(\mathbf{y})) |\partial \mathbf{x} / \partial \mathbf{y}|$$

as the density g of \mathbf{Y} .

In particular, if the change of variables is linear:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}, \quad \mathbf{x} = \mathbf{A}^{-1}\mathbf{y} - \mathbf{A}^{-1}\mathbf{b}, \quad \partial \mathbf{y} / \partial \mathbf{x} = |\mathbf{A}|, \quad \partial \mathbf{x} / \partial \mathbf{y} = |\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}.$$

2.3 The Fisher F-distribution

Suppose we have two independent random variables U and V , chi-square distributed with degrees of freedom (df) m and n respectively. We divide each by its df, obtaining U/m and V/n . The distribution of the *ratio*

$$F := \frac{U/m}{V/n}$$

will be important below. It is called the *F-distribution* with *degrees of freedom* (m, n) , $F(m, n)$. It is also known as the (Fisher) *variance-ratio distribution*.

Before introducing its density, we define the *Beta function*,

$$B(\alpha, \beta) := \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx,$$

wherever the integral converges ($\alpha > 0$ for convergence at 0, $\beta > 0$ for convergence at 1). By *Euler's integral for the Beta function*,

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

(see e.g. Copson (1935), §9.3). One may then show that the density of $F(m, n)$ is

$$f(x) = \frac{m^{\frac{1}{2}m} n^{\frac{1}{2}n}}{B(\frac{1}{2}m, \frac{1}{2}n)} \cdot \frac{x^{\frac{1}{2}(m-2)}}{(mx + n)^{\frac{1}{2}(m+n)}} \quad (m, n > 0, \quad x > 0)$$

(see e.g. Kendall and Stuart (1977), §16.15, §11.10; the original form given by Fisher is slightly different).

There are two important features of this density. The first is that (to within a normalisation constant, which, like many of those in Statistics, involves ratios of Gamma functions) it behaves near zero like the power $x^{\frac{1}{2}(m-2)}$ and near infinity like the power $x^{-\frac{1}{2}n}$, and is smooth and unimodal (has one peak). The second is that, like all the common and useful distributions in Statistics, its percentage points are *tabulated*. Of course, using tables of the *F-distribution* involves the complicating feature that one has *two* degrees of freedom (rather than one as with the chi-square or Student *t*-distributions), and that these must be taken in the correct *order*. It is sensible at this point for the reader to take some time to gain familiarity with use of tables of the *F-distribution*, using whichever standard set of statistical tables are to hand. Alternatively, all standard statistical packages will provide percentage points of F , t , χ^2 , etc. on demand. Again, it is sensible to take the time to gain familiarity with the statistical package of your choice, including use of the online Help facility.

One can derive the density of the *F* distribution from those of the χ^2 distributions above. One needs the formula for the density of a quotient of random variables. The derivation is left as an exercise; see Exercise 2.1. For an introduction to calculations involving the *F* distribution see Exercise 2.2.

2.4 Orthogonality

Recall that a square, non-singular ($n \times n$) matrix A is *orthogonal* if its inverse is its transpose:

$$A^{-1} = A^T.$$

We now show that the property of being independent $N(0, \sigma^2)$ is preserved under an orthogonal transformation.

Theorem 2.2 (Orthogonality Theorem)

If $X = (X_1, \dots, X_n)^T$ is an n -vector whose components are independent random variables, normally distributed with mean 0 and variance σ^2 , and we change variables from X to Y by

$$Y := AX$$

where the matrix A is orthogonal, then the components Y_i of Y are again independent, normally distributed with mean 0 and variance σ^2 .

Proof

We use the Jacobian formula. If $A = (a_{ij})$, since $\partial Y_i / \partial X_j = a_{ij}$, the Jacobian $\partial Y / \partial X = |A|$. Since A is orthogonal, $AA^T = AA^{-1} = I$. Taking determinants, $|A| \cdot |A^T| = |A| \cdot |A| = 1$: $|A| = 1$, and similarly $|A^T| = 1$. Since length is preserved under an orthogonal transformation,

$$\sum_1^n Y_i^2 = \sum_1^n X_i^2.$$

The joint density of (X_1, \dots, X_n) is, by independence, the product of the marginal densities, namely

$$f(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} x_i^2 \right\} = \frac{1}{(2\pi)^{\frac{1}{2}n}} \exp \left\{ -\frac{1}{2} \sum_1^n x_i^2 \right\}.$$

From this and the Jacobian formula, we obtain the joint density of (Y_1, \dots, Y_n) as

$$f(y_1, \dots, y_n) = \frac{1}{(2\pi)^{\frac{1}{2}n}} \exp \left\{ -\frac{1}{2} \sum_1^n y_i^2 \right\} = \prod_1^n \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} y_i^2 \right\}.$$

But this is the joint density of n independent standard normals – and so (Y_1, \dots, Y_n) are independent standard normal, as claimed. \square

Helmert's Transformation.

There exists an orthogonal $n \times n$ matrix P with first row

$$\frac{1}{\sqrt{n}}(1, \dots, 1)$$

(there are many such! Robert Helmert (1843–1917) made use of one when he introduced the χ^2 distribution in 1876 – see Kendall and Stuart (1977), Example 11.1 – and it is convenient to use his name here for any of them.) For, take this vector, which spans a one-dimensional subspace; take $n-1$ unit vectors not in this subspace and use the Gram–Schmidt orthogonalisation process (see e.g. Blyth and Robertson (2002b), Th. 1.4) to obtain a set of n orthonormal vectors.

2.5 Normal sample mean and sample variance

For X_1, \dots, X_n independent and identically distributed (iid) random variables, with mean μ and variance σ^2 , write

$$\bar{X} := \frac{1}{n} \sum_1^n X_i$$

for the *sample mean* and

$$S^2 := \frac{1}{n} \sum_1^n (X_i - \bar{X})^2$$

for the *sample variance*.

Note 2.3

Many authors use $1/(n-1)$ rather than $1/n$ in the definition of the sample variance. This gives S^2 as an *unbiased* estimator of the population variance σ^2 . But our definition emphasizes the parallel between the bar, or average, for sample quantities and the expectation for the corresponding population quantities:

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_1^n X_i \leftrightarrow EX, \\ S^2 &= \overline{(X - \bar{X})^2} \leftrightarrow \sigma^2 = E[(X - EX)^2], \end{aligned}$$

which is mathematically more convenient.

Theorem 2.4

If X_1, \dots, X_n are iid $N(\mu, \sigma^2)$,

- (i) the sample mean \bar{X} and the sample variance S^2 are independent,
- (ii) \bar{X} is $N(\mu, \sigma^2/n)$,
- (iii) nS^2/σ^2 is $\chi^2(n-1)$.

Proof

(i) Put $Z_i := (X_i - \mu)/\sigma$, $Z := (Z_1, \dots, Z_n)^T$; then the Z_i are iid $N(0, 1)$,

$$\bar{Z} = (\bar{X} - \mu)/\sigma, \quad nS^2/\sigma^2 = \sum_1^n (Z_i - \bar{Z})^2.$$

Also, since

$$\begin{aligned} \sum_1^n (Z_i - \bar{Z})^2 &= \sum_1^n Z_i^2 - 2\bar{Z} \sum_1^n Z_i + n\bar{Z}^2 \\ &= \sum_1^n Z_i^2 - 2\bar{Z} \cdot n\bar{Z} + n\bar{Z}^2 = \sum_1^n Z_i^2 - n\bar{Z}^2 : \\ \sum_1^n Z_i^2 &= \sum_1^n (Z_i - \bar{Z})^2 + n\bar{Z}^2. \end{aligned}$$

The terms on the right above are quadratic forms, with matrices A, B say, so we can write

$$\sum_1^n Z_i^2 = Z^T A Z + Z^T B Z. \quad (*)$$

Put $W := PZ$ with P a Helmert transformation – P orthogonal with first row $(1, \dots, 1)/\sqrt{n}$:

$$W_1 = \frac{1}{\sqrt{n}} \sum_1^n Z_i = \sqrt{n}\bar{Z}; \quad W_1^2 = n\bar{Z}^2 = Z^T B Z.$$

So

$$\sum_2^n W_i^2 = \sum_1^n W_i^2 - W_1^2 = \sum_1^n Z_i^2 - Z^T B Z = Z^T A Z = \sum_1^n (Z_i - \bar{Z})^2 = nS^2/\sigma^2.$$

But the W_i are independent (by the orthogonality of P), so W_1 is independent of W_2, \dots, W_n . So W_1^2 is independent of $\sum_2^n W_i^2$. So nS^2/σ^2 is independent of $n(\bar{X} - \mu)^2/\sigma^2$, so S^2 is independent of \bar{X} , as claimed.

(ii) We have $\bar{X} = (X_1 + \dots + X_n)/n$ with X_i independent, $N(\mu, \sigma^2)$, so with MGF $\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$. So X_i/n has MGF $\exp(\mu t/n + \frac{1}{2}\sigma^2 t^2/n^2)$, and \bar{X} has MGF

$$\prod_1^n \exp\left(\mu t/n + \frac{1}{2}\sigma^2 t^2/n^2\right) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2/n\right).$$

So \bar{X} is $N(\mu, \sigma^2/n)$.

(iii) In $(*)$, we have on the left $\sum_1^n Z_i^2$, which is the sum of the squares of n standard normals Z_i , so is $\chi^2(n)$ with MGF $(1 - 2t)^{-\frac{1}{2}n}$. On the right, we have

two independent terms. As \bar{Z} is $N(0, 1/n)$, $\sqrt{n}\bar{Z}$ is $N(0, 1)$, so $n\bar{Z}^2 = Z^T B Z$ is $\chi^2(1)$, with MGF $(1 - 2t)^{-\frac{1}{2}}$. Dividing (as in chi-square subtraction above), $Z^T A Z = \sum_1^n (Z_i - \bar{Z})^2$ has MGF $(1 - 2t)^{-\frac{1}{2}(n-1)}$. So $Z^T A Z = \sum_1^n (Z_i - \bar{Z})^2$ is $\chi^2(n-1)$. So nS^2/σ^2 is $\chi^2(n-1)$. \square

Note 2.5

1. This is a remarkable result. We quote (without proof) that this property actually *characterises* the normal distribution: if the sample mean and sample variance are independent, then the population distribution is normal (Geary's Theorem: R. C. Geary (1896–1983) in 1936; see e.g. Kendall and Stuart (1977), Examples 11.9 and 12.7).

2. The fact that when we form the sample mean, the mean is unchanged, while the variance decreases by a factor of the sample size n , is true generally. The point of (ii) above is that normality is preserved. This holds more generally: it will emerge in Chapter 4 that normality is preserved under any linear operation.

Theorem 2.6 (Fisher's Lemma)

Let X_1, \dots, X_n be iid $N(0, \sigma^2)$. Let

$$Y_i = \sum_{j=1}^n c_{ij} X_j \quad (i = 1, \dots, p, \quad p < n),$$

where the row-vectors (c_{i1}, \dots, c_{in}) are orthogonal for $i = 1, \dots, p$. If

$$S^2 = \sum_1^n X_i^2 - \sum_1^p Y_i^2,$$

then

- (i) S^2 is independent of Y_1, \dots, Y_p ,
- (ii) S^2 is $\chi^2(n-p)$.

Proof

Extend the $p \times n$ matrix (c_{ij}) to an $n \times n$ orthogonal matrix $C = (c_{ij})$ by Gram–Schmidt orthogonalisation. Then put

$$Y := CX,$$

so defining Y_1, \dots, Y_p (again) and Y_{p+1}, \dots, Y_n . As C is orthogonal, Y_1, \dots, Y_n are iid $N(0, \sigma^2)$, and $\sum_1^n Y_i^2 = \sum_1^n X_i^2$. So

$$S^2 = \left(\sum_1^n - \sum_1^p \right) Y_i^2 = \sum_{p+1}^n Y_i^2$$

is independent of Y_1, \dots, Y_p , and S^2/σ^2 is $\chi^2(n-p)$. \square

2.6 One-Way Analysis of Variance

To compare two normal means, we use the Student t -test, familiar from your first course in Statistics. What about comparing r means for $r > 2$?

Analysis of Variance goes back to early work by Fisher in 1918 on mathematical genetics and was further developed by him at Rothamsted Experimental Station in Harpenden, Hertfordshire in the 1920s. The convenient acronym ANOVA was coined much later, by the American statistician John W. Tukey (1915–2000), the pioneer of exploratory data analysis (EDA) in Statistics (Tukey (1977)), and coiner of the terms hardware, software and bit from computer science.

Fisher's motivation (which arose directly from the agricultural field trials carried out at Rothamsted) was to compare yields of several varieties of crop, say – or (the version we will follow below) of one crop under several fertiliser *treatments*. He realised that if there was more variability between groups (of yields with different treatments) than within groups (of yields with the same treatment) than one would expect if the treatments were the same, then this would be evidence against believing that they were the same. In other words, Fisher set out to *compare means by analysing variability* ('variance' – the term is due to Fisher – is simply a short form of 'variability').

We write μ_i for the mean yield of the i th variety, for $i = 1, \dots, r$. For each i , we draw n_i independent readings X_{ij} . The X_{ij} are independent, and we assume that they are normal, all with the same unknown variance σ^2 :

$$X_{ij} \sim N(\mu_i, \sigma^2) \quad (j = 1, \dots, n_i, \quad i = 1, \dots, r).$$

We write

$$n := \sum_1^r n_i$$

for the total sample size.

With two suffices i and j in play, we use a bullet to indicate that the suffix in that position has been averaged out. Thus we write

$$X_{i\bullet}, \quad \text{or} \quad \bar{X}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \quad (i = 1, \dots, r)$$

for the i th *group mean* (the sample mean of the i th sample)

$$X_{\bullet\bullet}, \quad \text{or} \quad \bar{X} := \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} X_{ij} = \frac{1}{n} \sum_{i=1}^r n_i \bar{X}_i$$

for the *grand mean* and,

$$S_i^2 := \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - X_{i\bullet})^2$$

for the i th sample variance.

Define the *total sum of squares*

$$SS := \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - X_{\bullet\bullet})^2 = \sum_i \sum_j [(X_{ij} - X_{i\bullet}) + (X_{i\bullet} - X_{\bullet\bullet})]^2.$$

As

$$\sum_j (X_{ij} - X_{i\bullet}) = 0$$

(from the definition of $X_{i\bullet}$ as the average of the X_{ij} over j), if we expand the square above, the cross terms vanish, giving

$$\begin{aligned} SS &= \sum_i \sum_j (X_{ij} - X_{i\bullet})^2 \\ &\quad + \sum_i \sum_j (X_{ij} - X_{i\bullet})(X_{i\bullet} - X_{\bullet\bullet}) \\ &\quad + \sum_i \sum_j (X_{i\bullet} - X_{\bullet\bullet})^2 \\ &= \sum_i \sum_j (X_{ij} - X_{i\bullet})^2 + \sum_i \sum_j (X_{i\bullet} - X_{\bullet\bullet})^2 \\ &= \sum_i n_i S_i^2 + \sum_i n_i (X_{i\bullet} - X_{\bullet\bullet})^2. \end{aligned}$$

The first term on the right measures the amount of variability *within* groups. The second measures the variability *between* groups. We call them the *sum of squares for error* (or *within groups*), SSE , also known as the *residual sum of squares*, and the *sum of squares for treatments* (or *between groups*), respectively:

$$SS = SSE + SST,$$

where

$$SSE := \sum_i n_i S_i^2, \quad SST := \sum_i n_i (X_{i\bullet} - X_{\bullet\bullet})^2.$$

Let H_0 be the null hypothesis of no treatment effect:

$$H_0 : \quad \mu_i = \mu \quad (i = 1, \dots, r).$$

If H_0 is true, we have merely one large sample of size n , drawn from the distribution $N(\mu, \sigma^2)$, and so

$$SS/\sigma^2 = \frac{1}{\sigma^2} \sum_i \sum_j (X_{ij} - X_{\bullet\bullet})^2 \sim \chi^2(n-1) \quad \text{under } H_0.$$

In particular,

$$E[SS/(n-1)] = \sigma^2 \quad \text{under } H_0.$$

Whether or not H_0 is true,

$$n_i S_i^2 / \sigma^2 = \frac{1}{\sigma^2} \sum_j (X_{ij} - X_{i\bullet})^2 \sim \chi^2(n_i - 1).$$

So by the Chi-Square Addition Property

$$SSE / \sigma^2 = \sum_i n_i S_i^2 / \sigma^2 = \frac{1}{\sigma^2} \sum_i \sum_j (X_{ij} - X_{i\bullet})^2 \sim \chi^2(n - r),$$

since as $n = \sum_i n_i$,

$$\sum_{i=1}^r (n_i - 1) = n - r.$$

In particular,

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

Next,

$$SST := \sum_i n_i (X_{i\bullet} - X_{\bullet\bullet})^2, \quad \text{where} \quad X_{\bullet\bullet} = \frac{1}{n} \sum_i n_i X_{i\bullet}, \quad SSE := \sum_i n_i S_i^2.$$

Now S_i^2 is independent of $X_{i\bullet}$, as these are the sample variance and sample mean from the i th sample, whose independence was proved in Theorem 2.4. Also S_i^2 is independent of $X_{j\bullet}$ for $j \neq i$, as they are formed from different independent samples. Combining, S_i^2 is independent of all the $X_{j\bullet}$, so of their (weighted) average $X_{\bullet\bullet}$, so of SST , a function of the $X_{j\bullet}$ and of $X_{\bullet\bullet}$. So $SSE = \sum_i n_i S_i^2$ is also independent of SST .

We can now use the Chi-Square Subtraction Property. We have, under H_0 , the independent sum

$$SS / \sigma^2 = SSE / \sigma^2 +_{ind} SST / \sigma^2.$$

By above, the left-hand side is $\chi^2(n - 1)$, while the first term on the right is $\chi^2(n - r)$. So the second term on the right must be $\chi^2(r - 1)$. This gives:

Theorem 2.7

Under the conditions above and the null hypothesis H_0 of no difference of treatment means, we have the sum-of-squares decomposition

$$SS = SSE +_{ind} SST,$$

where $SS / \sigma^2 \sim \chi^2(n - 1)$, $SSE / \sigma^2 \sim \chi^2(n - r)$ and $SST / \sigma^2 \sim \chi^2(r - 1)$.

When we have a sum of squares, chi-square distributed, and we divide by its degrees of freedom, we will call the resulting ratio a *mean sum of squares*, and denote it by changing the SS in the name of the sum of squares to MS. Thus the mean sum of squares is

$$MS := SS/\text{df}(SS) = SS/(n-1)$$

and the mean sums of squares for treatment and for error are

$$\begin{aligned} MST &:= SST/\text{df}(SST) = SST/(r-1), \\ MSE &:= SSE/\text{df}(SSE) = SSE/(n-r). \end{aligned}$$

By the above,

$$SS = SST + SSE;$$

whether or not H_0 is true,

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

under H_0 ,

$$E[MS] = E[SS]/(n-1) = \sigma^2, \quad \text{and so also} \quad E[MST]/(r-1) = \sigma^2.$$

Form the F -statistic

$$F := MST/MSE.$$

Under H_0 , this has distribution $F(r-1, n-r)$. Fisher realised that comparing the size of this F -statistic with percentage points of this F -distribution gives us a way of testing the truth or otherwise of H_0 . Intuitively, if the treatments do differ, this will tend to inflate SST , hence MST , hence $F = MST/MSE$. To justify this intuition, we proceed as follows. Whether or not H_0 is true,

$$\begin{aligned} SST &= \sum_i n_i (X_{i\bullet} - X_{\bullet\bullet})^2 = \sum_i n_i X_{i\bullet}^2 - 2X_{\bullet\bullet} \sum_i n_i X_{i\bullet} + X_{\bullet\bullet}^2 \sum_i n_i \\ &= \sum_i n_i X_{i\bullet}^2 - nX_{\bullet\bullet}^2, \end{aligned}$$

since $\sum_i n_i X_{i\bullet} = nX_{\bullet\bullet}$ and $\sum_i n_i = n$. So

$$\begin{aligned} E[SST] &= \sum_i n_i E[X_{i\bullet}^2] - nE[X_{\bullet\bullet}^2] \\ &= \sum_i n_i [\text{var}(X_{i\bullet}) + (EX_{i\bullet})^2] - n[\text{var}(X_{\bullet\bullet}) + (EX_{\bullet\bullet})^2]. \end{aligned}$$

But $\text{var}(X_{i\bullet}) = \sigma^2/n_i$,

$$\begin{aligned} \text{var}(X_{\bullet\bullet}) &= \text{var}\left(\frac{1}{n} \sum_{i=1}^r n_i X_{i\bullet}\right) = \frac{1}{n^2} \sum_{i=1}^r n_i^2 \text{var}(X_{i\bullet}), \\ &= \frac{1}{n^2} \sum_{i=1}^r n_i^2 \sigma^2 / n_i = \sigma^2 / n \end{aligned}$$

(as $\sum_i n_i = n$). So writing

$$\begin{aligned}\bar{\mu} &:= \frac{1}{n} \sum_i n_i \mu_i = EX_{\bullet\bullet} = E \frac{1}{n} \sum_i n_i X_{i\bullet}, \\ E(SST) &= \sum_1^r n_i \left[\frac{\sigma^2}{n_i} + \mu_i^2 \right] - n \left[\frac{\sigma^2}{n} + \bar{\mu}^2 \right] \\ &= (r-1)\sigma^2 + \sum_i n_i \mu_i^2 - n\bar{\mu}^2 \\ &= (r-1)\sigma^2 + \sum_i n_i (\mu_i - \bar{\mu})^2\end{aligned}$$

(as $\sum_i n_i = n$, $n\bar{\mu} = \sum_i n_i \mu_i$). This gives the inequality

$$E[SST] \geq (r-1)\sigma^2,$$

with equality iff

$$\mu_i = \bar{\mu} \quad (i = 1, \dots, r), \quad \text{i.e.} \quad H_0 \text{ is true.}$$

Thus when H_0 is *false*, the mean of SST *increases*, so *larger* values of SST , so of MST and of $F = MST/MSE$, are evidence *against* H_0 . It is thus appropriate to use a *one-tailed* F -test, rejecting H_0 if the value F of our F -statistic is *too big*. How big is too big depends, of course, on our chosen significance level α , and hence on the tabulated value $F_{tab} := F_\alpha(r-1, n-r)$, the upper α -point of the relevant F -distribution. We summarise:

Theorem 2.8

When the null hypothesis H_0 (that all the treatment means μ_1, \dots, μ_r are equal) is true, the F -statistic $F := MST/MSE = (SST/(r-1))/(SSE/(n-r))$ has the F -distribution $F(r-1, n-r)$. When the null hypothesis is false, F increases. So large values of F are evidence against H_0 , and we test H_0 using a one-tailed test, rejecting at significance level α if F is too big, that is, with critical region

$$F > F_{tab} = F_\alpha(r-1, n-r).$$

Model Equations for One-Way ANOVA.

$$X_{ij} = \mu_i + \epsilon_{ij} \quad (i = 1, \dots, r, \quad j = 1, \dots, r), \quad \epsilon_{ij} \text{ iid } N(0, \sigma^2).$$

Here μ_i is the *main effect* for the i th treatment, the null hypothesis is $H_0: \mu_1 = \dots = \mu_r = \mu$, and the unknown variance σ^2 is a nuisance parameter. The point of forming the ratio in the F -statistic is to cancel this nuisance parameter σ^2 , just as in forming the ratio in the Student t -statistic in one's first course in Statistics. We will return to nuisance parameters in §5.1.1 below.

Calculations.

In any calculation involving variances, there is cancellation to be made, which is worthwhile and important numerically. This stems from the definition and ‘computing formula’ for the variance,

$$\sigma^2 := E[(X - EX)^2] = E[X^2] - (EX)^2$$

and its sample counterpart

$$S^2 := \overline{(X - \bar{X})^2} = \overline{X^2} - \bar{X}^2.$$

Writing T , T_i for the grand total and group totals, defined by

$$T := \sum_i \sum_j X_{ij}, \quad T_i := \sum_j X_{ij},$$

so $X_{\bullet\bullet} = T/n$, $nX_{\bullet\bullet}^2 = T^2/n$:

$$SS = \sum_i \sum_j X_{ij}^2 - T^2/n,$$

$$SST = \sum_i T_i^2/n_i - T^2/n,$$

$$SSE = SS - SST = \sum_i \sum_j X_{ij}^2 - \sum_i T_i^2/n_i.$$

These formulae help to reduce rounding errors and are easiest to use if carrying out an Analysis of Variance by hand.

It is customary, and convenient, to display the output of an Analysis of Variance by an ANOVA table, as shown in Table 2.1. (The term ‘Error’ can be used in place of ‘Residual’ in the ‘Source’ column.)

Source	df	SS	Mean Square	F
Treatments	$r - 1$	SST	$MST = SST/(r - 1)$	MST/MSE
Residual	$n - r$	SSE	$MSE = SSE/(n - r)$	
Total	$n - 1$	SS		

Table 2.1 One-way ANOVA table.

Example 2.9

We give an example which shows how to calculate the Analysis of Variance tables by hand. The data in Table 2.2 come from an agricultural experiment. We wish to test for different mean yields for the different fertilisers. We note that

Fertiliser	Yield
A	14.5, 12.0, 9.0, 6.5
B	13.5, 10.0, 9.0, 8.5
C	11.5, 11.0, 14.0, 10.0
D	13.0, 13.0, 13.5, 7.5
E	15.0, 12.0, 8.0, 7.0
F	12.5, 13.5, 14.0, 8.0

Table 2.2 Data for Example 2.9

we have six treatments so $6 - 1 = 5$ degrees of freedom for treatments. The total number of degrees of freedom is the number of observations minus one, hence 23. This leaves 18 degrees of freedom for the within-treatments sum of squares. The total sum of squares can be calculated routinely as $\sum(y_{ij} - \bar{y})^2 = \sum y_{ij}^2 - n\bar{y}^2$, which is often most efficiently calculated as $\sum y_{ij}^2 - (1/n)(\sum y_{ij})^2$. This calculation gives $SS = 3119.25 - (1/24)(266.5)^2 = 159.990$. The easiest next step is to calculate SST , which means we can then obtain SSE by subtraction as above. The formula for SST is relatively simple and reads $\sum_i T_i^2/n_i - T^2/n$, where T_i denotes the sum of the observations corresponding to the i th treatment and $T = \sum_{ij} y_{ij}$. Here this gives $SST = (1/4)(42^2 + 41^2 + 46.5^2 + 47^2 + 42^2 + 48^2) - 1/24(266.5)^2 = 11.802$. Working through, the full ANOVA table is shown in Table 2.3.

Source	df	Sum of Squares	Mean Square	F
Between fertilisers	5	11.802	2.360	0.287
Residual	18	148.188	8.233	
Total	23	159.990		

Table 2.3 One-way ANOVA table for Example 2.9

This gives a non-significant p -value compared with $F_{3,16}(0.95) = 3.239$. R calculates the p -value to be 0.914. Alternatively, we may place bounds on the p -value by looking at statistical tables. In conclusion, we have no evidence for differences between the various types of fertiliser.

In the above example, the calculations were made more simple by having equal numbers of observations for each treatment. However, the same general procedure works when this no longer continues to be the case. For detailed worked examples with unequal sample sizes see Snedecor and Cochran (1989) §12.10.

S-Plus/R[®].

We briefly describe implementation of one-way ANOVA in *S-Plus/R*[®]. For background and details, see e.g. Crawley (2002), Ch. 15. Suppose we are studying the dependence of yield on treatment, as above. [Note that this requires that we set treatment to be a *factor* variable, taking discrete rather than continuous values, which can be achieved by setting `treatment <- factor(treatment)`.] Then, using `aov` as short for ‘Analysis of Variance’, `<-` for the assignment operator in *S-Plus* (read as ‘goes to’ or ‘becomes’) and `~` as short for ‘depends on’ or ‘is regressed on’, we use

```
model <- aov (yield ~ treatment)
```

to do the analysis, and ask for the summary table by

```
summary(model)
```

A complementary `anova` command is summarised briefly in Chapter 5.2.1.

2.7 Two-Way ANOVA; No Replications

In the agricultural experiment considered above, problems may arise if the growing area is not homogeneous. The plots on which the different treatments are applied may differ in fertility – for example, if a field slopes, nutrients tend to leach out of the soil and wash downhill, so lower-lying land may give higher yields than higher-lying land. Similarly, differences may arise from differences in drainage, soil conditions, exposure to sunlight or wind, crops grown in the past, etc. If such differences are not taken into account, we will be unable to distinguish between differences in yield resulting from differences in *treatment*, our object of study, and those resulting from differences in growing conditions – *plots*, for short – which are not our primary concern. In such a case, one says that treatments are *confounded* with plots – we would have no way of separating the effect of one from that of the other.

The only way out of such difficulties is to subdivide the growing area into plots, each of which can be treated as a homogeneous growing area, and then subdivide each plot and apply different treatments to the different sub-plots or blocks. In this way we will be ‘comparing like with like’, and avoid the pitfalls of confounding.

When allocating treatments to blocks, we may wish to *randomise*, to avoid the possibility of inadvertently introducing a treatment-block linkage. Relevant here is the subject of *design of experiments*; see §9.3.

In the sequel, we assume for simplicity that the block sizes are the same and the number of treatments is the same for each block. The model equations will now be of the form

$$X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad (i = 1, \dots, r, \quad j = 1, \dots, n).$$

Here μ is the *grand mean* (or *overall mean*); α_i is the *i*th *treatment effect* (we take $\sum_i \alpha_i = 0$, otherwise this sum can – and so should – be absorbed into μ ; β_j is the *j*th *block effect* (similarly, we take $\sum_j \beta_j = 0$); the errors ϵ_{ij} are iid $N(0, \sigma^2)$, as before.

Recall the terms $X_{i\bullet}$ from the one-way case; their counterparts here are similarly denoted $X_{\bullet j}$. Start with the algebraic identity

$$(X_{ij} - X_{\bullet\bullet}) = (X_{ij} - X_{i\bullet} - X_{\bullet j} + X_{\bullet\bullet}) + (X_{i\bullet} - X_{\bullet\bullet}) + (X_{\bullet j} - X_{\bullet\bullet}).$$

Square and add. One can check that the cross terms cancel, leaving only the squared terms. For example, $(X_{ij} - X_{i\bullet} - X_{\bullet j} + X_{\bullet\bullet})$ averages over i to $-(X_{\bullet j} - X_{\bullet\bullet})$, and over j to $-(X_{i\bullet} - X_{\bullet\bullet})$, while each of the other terms on the right involves only one of i and j , and so is unchanged when averaged over the other. One is left with

$$\begin{aligned} \sum_{i=1}^r \sum_{j=1}^n (X_{ij} - X_{\bullet\bullet})^2 &= \sum_{i=1}^r \sum_{j=1}^n (X_{ij} - X_{i\bullet} - X_{\bullet j} + X_{\bullet\bullet})^2 \\ &\quad + n \sum_{i=1}^r (X_{i\bullet} - X_{\bullet\bullet})^2 \\ &\quad + r \sum_{j=1}^n (X_{\bullet j} - X_{\bullet\bullet})^2. \end{aligned}$$

We write this as

$$SS = SSE + SST + SSB,$$

giving the total sum of squares SS as the sum of the sum of squares for error (SSE), the sum of squares for treatments (SST) (as before) and a new term, the sum of squares for *blocks*, (SSB). The degrees of freedom are, respectively, $nr - 1$ for SS (the total sample size is nr , and we lose one df in estimating σ), $r - 1$ for treatments (as before), $n - 1$ for blocks (by analogy with treatments – or equivalently, there are n block parameters β_j , but they are subject to one constraint, $\sum_j \beta_j = 0$), and $(n - 1)(r - 1)$ for error (to give the correct total in the df column in the table below). Independence of the three terms on the right follows by arguments similar to those in the one-way case. We can accordingly construct a two-way ANOVA table, as in Table 2.4.

Here we have *two* F -statistics, $FT := MST/MSE$ for treatment effects and $FB := MSB/MSE$ for block effects. Accordingly, we can test *two* null hypotheses, one, $H_0(T)$, for presence of a treatment effect and one, $H_0(B)$, for presence of a block effect.

Source	df	SS	Mean Square	F
Treatments	$r - 1$	SST	$MST = \frac{SST}{r-1}$	MST/MSE
Blocks	$n - 1$	SSB	$MSB = \frac{SSB}{n-1}$	MSB/MSE
Residual	$(r - 1)(n - 1)$	SSE	$MSE = \frac{SSE}{(r-1)(n-1)}$	
Total	$rn - 1$	SS		

Table 2.4 Two-way ANOVA table**Note 2.10**

In educational psychology (or other behavioural sciences), ‘treatments’ might be different questions on a test, ‘blocks’ might be *individuals*. We take it for granted that individuals differ. So we need not calculate MSB nor test $H_0(B)$ (though packages such as S-Plus will do so automatically). Then $H_0(T)$ as above tests for differences between mean scores on questions in a test. (Where the questions carry equal credit, such differences are undesirable – but may well be present in practice!)

Implementation. In S-Plus, the commands above extend to

```
model <- aov(yield ~ treatment + block)
summary(model)
```

Example 2.11

We illustrate the two-way Analysis of Variance with an example. We return to the agricultural example in Example 2.9, but suppose that the data can be linked to growing areas as shown in Table 2.5. We wish to test the hypothesis that there are no differences between the various types of fertiliser. The

Fertiliser	Area 1	Area 2	Area 3	Area 4
A	14.5	12.0	9.0	6.5
B	13.5	10.0	9.0	8.5
C	11.5	11.0	14.0	10.0
D	13.0	13.0	13.5	7.5
E	15.0	12.0	8.0	7.0
F	12.5	13.5	14.0	8.0

Table 2.5 Data for Example 2.11

sum-of-squares decomposition for two-way ANOVA follows in an analogous way to the one-way case. There are relatively simple formulae for SS , SST , and SSB , meaning that SSE can easily be calculated by subtraction. In detail, these formulae are

$$\begin{aligned} SS &= \sum_{ij} X_{ij}^2 - \frac{1}{nr} \left(\sum X_{ij} \right)^2, \\ SST &= (X_{1\bullet}^2 + \dots + X_{r\bullet}^2) / n - \frac{1}{nr} \left(\sum X_{ij} \right)^2, \\ SSB &= (X_{\bullet 1}^2 + \dots + X_{\bullet n}^2) / r - \frac{1}{nr} \left(\sum X_{ij} \right)^2, \end{aligned}$$

with $SSE = SS - SST - SSB$. Returning to our example, we see that

$$\begin{aligned} SS &= 3119.25 - (1/24)(266.5)^2 = 159.990, \\ SST &= (42^2 + 41^2 + 46.5^2 + 47^2 + 42^2 + 48^2)/4 - (1/24)(266.5)^2 = 11.802, \\ SSB &= (80^2 + 71.5^2 + 67.5^2 + 47.5^2)/6 - (1/24)(266.5)^2 = 94.865. \end{aligned}$$

By subtraction $SSE = 159.9896 - 11.80208 - 94.86458 = 53.323$. These calculations lead us to the ANOVA table in Table 2.6. Once again we have no evidence for differences amongst the 6 types of fertiliser. The variation that does occur is mostly due to the effects of different growing areas.

Source	df	S.S.	MS	F	p
Fertilisers	5	11.802	2.360	0.664	0.656
Area	3	94.865	31.622	8.895	0.001
Residual	15	53.323	3.555		
Total	23	159.990			

Table 2.6 Two-way ANOVA table for Example 2.11

2.8 Two-Way ANOVA: Replications and Interaction

In the above, we have one reading X_{ij} for each *cell*, or combination of the i th treatment and the j th block. But we may have more. Suppose we have m *replications* – independent readings – per cell. We now need three suffices rather than two. The model equations will now be of the form

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad (i = 1, \dots, r, \quad j = 1, \dots, n, \quad k = 1, \dots, m).$$

Here the new parameters γ_{ij} measure possible *interactions* between treatment and block effects. This allows one to study situations in which effects are *not additive*. Although we use the word interaction here as a technical term in Statistics, this is fully consistent with its use in ordinary English. We are all familiar with situations where, say, a medical treatment (e.g. a drug) may interact with some aspect of our diet (e.g. alcohol). Similarly, two drugs may interact (which is why doctors must be careful in checking what medication a patient is currently taking before issuing a new prescription). Again, different alcoholic drinks may interact (folklore wisely counsels against mixing one's drinks), etc.

Arguments similar to those above lead to the following sum-of-squares decomposition:

$$\begin{aligned} \sum_{i=1}^r \sum_{j=1}^n (X_{ijk} - X_{\bullet\bullet\bullet})^2 &= \sum_i \sum_j \sum_k (X_{ijk} - X_{ij\bullet})^2 \\ &\quad + nm \sum_i (X_{i\bullet\bullet} - X_{\bullet\bullet\bullet})^2 \\ &\quad + rm \sum_j (X_{\bullet j\bullet} - X_{\bullet\bullet\bullet})^2 \\ &\quad + m \sum_i \sum_j (X_{ij\bullet} - X_{i\bullet\bullet} - X_{\bullet j\bullet} + X_{\bullet\bullet\bullet})^2. \end{aligned}$$

We write this as

$$SS = SSE + SST + SSB + SSI,$$

where the new term is the sum of squares for *interactions*. The degrees of freedom are $r - 1$ for treatments as before, $n - 1$ for blocks as before, $(r - 1)(n - 1)$ for interactions (the product of the effective number of parameters for treatments and for blocks), $rn - 1$ in total (there are rn readings), and $rn(m - 1)$ for error (so that the df totals on the right and left above agree).

Implementation. The S-Plus/R[®] commands now become

```
model <- aov(yield ~ treatment * block)

summary(model)
```

This notation is algebraically motivated, and easy to remember. With *additive* effects, we used a $+$. We now use a $*$, suggestive of the possibility of ‘product’ terms representing the interactions. We will encounter many more such situations in the next chapter, when we deal with multiple regression.

The summary table now takes the form of Table 2.7. We now have *three* F -statistics, FT and FB as before, and now FI also, which we can use to test for the presence of interactions.

Source	df	SS	Mean Square	F
Treatments	$r - 1$	SST	$MST = \frac{SST}{r-1}$	MST/MSE
Blocks	$n - 1$	SSB	$MSB = \frac{SSB}{n-1}$	MSB/MSE
Interaction	$(r - 1)(n - 1)$	SSI	$MSI = \frac{SSI}{(r-1)(n-1)}$	MSI/MSE
Residual	$rn(m - 1)$	SSE	$MSE = \frac{SSE}{rn(m-1)}$	
Total	$rmn - 1$	SS		

Table 2.7 Two-way ANOVA table with interactions**Example 2.12**

The following example illustrates the procedure for two-way ANOVA with interactions. The data in Table 2.8 link the growth of hamsters of different coat colours when fed different diets.

	Light coat	Dark coat
Diet A	6.6, 7.2	8.3, 8.7
Diet B	6.9, 8.3	8.1, 8.5
Diet C	7.9, 9.2	9.1, 9.0

Table 2.8 Data for Example 2.12

The familiar formula for the total sum of squares gives $SS = 805.2 - (97.8^2/12) = 8.13$. In a similar manner to Example 2.11, the main effects sum-of-squares calculations give

$$SST = \sum_{nm} \frac{y_{i\bullet\bullet}^2}{nm} - \frac{\left(\sum_{ijk} y_{ijk}\right)^2}{rmn},$$

$$SSB = \frac{y_{\bullet\bullet\bullet}^2}{rm} - \frac{\left(\sum_{ijk} y_{ijk}\right)^2}{rmn},$$

and in this case give $SST = (1/4)(30.8^2 + 31.8^2 + 35.2^2) - (97.8^2/12) = 2.66$ and $SSB = (1/6)(46.1^2 + 51.7^2) - (97.8^2/12) = 2.613$. The interaction sum of squares can be calculated as a sum of squares corresponding to every cell in the table once the main effects of SST and SSB have been accounted for. The calculation is

$$SSI = \frac{1}{m} \sum y_{ij\bullet}^2 - SST - SSB - \frac{\left(\sum_{ijk} y_{ijk}\right)^2}{rmn},$$

which in this example gives $SSI = (1/2)(13.8^2 + 17^2 + 15.2^2 + 16.6^2 + 17.1^2 + 18.1^2) - 2.66 - 2.613 - (97.8^2/12) = 0.687$. As before, SSE can be calculated by subtraction, and the ANOVA table is summarised in Table 2.9. The results

Source	df	SS	MS	F	p
Diet	2	2.66	1.33	3.678	0.091
Coat	1	2.613	2.613	7.226	0.036
Diet:Coat	2	0.687	0.343	0.949	0.438
Residual	5	2.17	0.362		
Total	11	8.13			

Table 2.9 Two-way ANOVA with interactions for Example 2.12.

suggest that once we take into account the different types of coat, the effect of the different diets is seen to become only borderline significant. The diet:coat interaction term is seen to be non-significant and we might consider in a subsequent analysis the effects of deleting this term from the model.

Note 2.13 (Random effects)

The model equation for two-way ANOVA with interactions is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk},$$

with $\sum_i \alpha_i = \sum_j \beta_j = \sum_{ij} \gamma_{ij} = 0$. Here the α_i , β_j , γ_{ij} are constants, and the randomness is in the errors ϵ_{ijk} . Suppose, however, that the β_i were themselves random (in the examination set-up above, the suffix i might refer to the i th question, and suffix j to the j th candidate; the candidates might be chosen at random from a larger population). We would then use notation such as

$$y_{ijk} = \mu + \alpha_i + b_j + c_{ij} + \epsilon_{ijk}.$$

Here we have both a fixed effect (for questions, i) and a random effect (for candidates, j). With both fixed and random effects, we speak of a *mixed* model; see §9.1.

With only random effects, we have a *random effects model*, and use notation such as

$$y_{ijk} = \mu + a_i + b_j + c_{ij} + \epsilon_{ijk}.$$

We restrict for simplicity here to the model with no interaction terms:

$$y_{ijk} = \mu + a_i + b_j + \epsilon_{ijk}.$$

Assuming independence of the random variables on the right, the variances add (see e.g. Haigh (2002), Cor. 5.6):

$$\sigma_y^2 = \sigma_a^2 + \sigma_b^2 + \sigma_\epsilon^2,$$

in an obvious notation. The terms on the right are called *variance components*; see e.g. Searle, Casella and McCulloch (1992) for a detailed treatment.

Variance components can be traced back to work of Airy in 1861 on astronomical observations (recall that astronomy also led to the development of Least Squares by Legendre and Gauss).

EXERCISES

- 2.1. (i) Show that if X, Y are positive random variables with joint density $f(x, y)$ their quotient $Z := X/Y$ has density

$$h(z) = \int_0^\infty y f(yz, y) dy \quad (z > 0).$$

So if X, Y are independent with densities f, g ,

$$h(z) = \int_0^\infty y f(yz) g(y) dy \quad (z > 0).$$

- (ii) If X has density f and $c > 0$, show that X/c has density

$$f_{X/c}(x) = cf(cx).$$

- (iii) Deduce that the Fisher F-distribution $F(m, n)$ has density

$$h(z) = m^{\frac{1}{2}m} n^{\frac{1}{2}n} \frac{\Gamma(\frac{1}{2}m + \frac{1}{2}n)}{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}n)} \cdot \frac{z^{\frac{1}{2}m-1}}{(n + mz)^{\frac{1}{2}(m+n)}} \quad (z > 0).$$

- 2.2. Using tables or S-Plus/R[®] produce bounds or calculate the exact probabilities for the following statements. [Note. In S-Plus/R[®] the command `pf` may prove useful.]

- (i) $P(X < 1.4)$ where $X \sim F_{5,17}$,
- (ii) $P(X > 1)$ where $X \sim F_{1,16}$,
- (iii) $P(X < 4)$ where $X \sim F_{1,3}$,
- (iv) $P(X > 3.4)$ where $X \sim F_{19,4}$,
- (v) $P(\ln X > -1.4)$ where $X \sim F_{10,4}$.

Fat 1	Fat 2	Fat 3	Fat 4
164	178	175	155
172	191	193	166
168	197	178	149
177	182	171	164
156	185	163	170
195	177	176	168

Table 2.10 Data for Exercise 2.3.

- 2.3. *Doughnut data.* Doughnuts absorb fat during cooking. The following experiment was conceived to test whether the amount of fat absorbed depends on the type of fat used. Table 2.10 gives the amount of fat absorbed per batch of doughnuts. Produce the one-way Analysis of Variance table for these data. What is your conclusion?
- 2.4. The data in Table 2.11 come from an experiment where growth is measured and compared to the variable *photoperiod* which indicates the length of daily exposure to light. Produce the one-way ANOVA table for these data and determine whether or not growth is affected by the length of daily light exposure.

Very short	Short	Long	Very long
2	3	3	4
3	4	5	6
1	2	1	2
1	1	2	2
2	2	2	2
1	1	2	3

Table 2.11 Data for Exercise 2.4

- 2.5. *Unpaired t -test with equal variances.* Under the null hypothesis the statistic t defined as

$$t = \frac{\sqrt{\frac{n_1 n_2}{n_1 + n_2}} (\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2))}{s}$$

should follow a t distribution with $n_1 + n_2 - 2$ degrees of freedom, where n_1 and n_2 denote the number of observations from samples 1 and 2 and s is the pooled estimate given by

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$$

where

$$\begin{aligned}s_1^2 &= \frac{1}{n_1 - 1}(\sum x_1^2 - (n_1 - 1)\bar{x}_1^2), \\ s_2^2 &= \frac{1}{n_2 - 1}(\sum x_2^2 - (n_2 - 1)\bar{x}_2^2).\end{aligned}$$

(i) Give the relevant statistic for a test of the hypothesis $\mu_1 = \mu_2$ and $n_1 = n_2 = n$.

(ii) Show that if $n_1 = n_2 = n$ then one-way ANOVA recovers the same results as the unpaired t -test. [Hint. Show that the F -statistic satisfies $F_{1,2(n-1)} = t_{2(n-1)}^2$].

2.6. Let Y_1, Y_2 be iid $N(0, 1)$. Give values of a and b such that

$$a(Y_1 - Y_2)^2 + b(Y_1 + Y_2)^2 \sim \chi_2^2.$$

2.7. Let Y_1, Y_2, Y_3 be iid $N(0, 1)$. Show that

$$\frac{1}{3} \left[(Y_1 - Y_2)^2 + (Y_2 - Y_3)^2 + (Y_3 - Y_1)^2 \right] \sim \chi_2^2.$$

Generalise the above result for a sample Y_1, Y_2, \dots, Y_n of size n .

2.8. The data in Table 2.12 come from an experiment testing the number of failures out of 100 planted soyabean seeds, comparing four different seed treatments, with no treatment ('check'). Produce the two-way ANOVA table for this data and interpret the results. (We will return to this example in Chapter 8.)

Treatment	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5
Check	8	10	12	13	11
Arasan	2	6	7	11	5
Sperguson	4	10	9	8	10
Semesan, Jr	3	5	9	10	6
Fermate	9	7	5	5	3

Table 2.12 Data for Exercise 2.8

2.9. *Photoperiod example revisited.* When we add in knowledge of plant genotype the full data set is as shown in Table 2.13. Produce the two-way ANOVA table and revise any conclusions from Exercise 2.4 in the light of these new data as appropriate.

Genotype	Very short	Short	Long	Very Long
A	2	3	3	4
B	3	4	5	6
C	1	2	1	2
D	1	1	2	2
E	2	2	2	2
F	1	1	2	3

Table 2.13 Data for Exercise 2.9

2.10. *Two-way ANOVA with interactions.* Three varieties of potato are planted on three plots at each of four locations. The yields in bushels are given in Table 2.14. Produce the ANOVA table for these data. Does the interaction term appear necessary? Describe your conclusions.

Variety	Location 1	Location 2	Location 3	Location 4
A	15, 19, 22	17, 10, 13	9, 12, 6	14, 8, 11
B	20, 24, 18	24, 18, 22	12, 15, 10	21, 16, 14
C	22, 17, 14	26, 19, 21	10, 5, 8	19, 15, 12

Table 2.14 Data for Exercise 2.10

2.11. *Two-way ANOVA with interactions.* The data in Table 2.15 give the gains in weight of male rats from diets with different sources and different levels of protein. Produce the two-way ANOVA table with interactions for these data. Test for the presence of interactions between source and level of protein and state any conclusions that you reach.

Source	High Protein	Low Protein
Beef	73, 102, 118, 104, 81, 107, 100, 87, 117, 111	90, 76, 90, 64, 86, 51, 72, 90, 95, 78
Cereal	98, 74, 56, 111, 95, 88, 82, 77, 86, 92	107, 95, 97, 80, 98, 74, 74, 67, 89, 58
Pork	94, 79, 96, 98, 102, 102, 108, 91, 120, 105	49, 82, 73, 86, 81, 97, 106, 70, 61, 82

Table 2.15 Data for Exercise 2.11

Regression

Linear Models in Statistics

Bingham, N.H.; Fry, J.M.

2010, XIII, 284 p. 50 illus., Softcover

ISBN: 978-1-84882-968-8