11. Analysis of Variance

The analysis of variance (or ANOVA), originally developed by R. A. FISHER, concerns testing the hypothesis of equal means of a number of samples. Such problems occur, for example, in the comparison of a series of measurements carried out under different conditions, or in quality control of samples produced by different machines. One tries to discover what influence the changing of *external variables* (e.g., experimental conditions, the number of a machine) has on a sample. For the simple case of only two samples, this problem can also be solved with Student's difference test (Sect. 8.3).

We speak of *one-way analysis of variance*, or also one-way classification, when only one external variable is changed. The evaluation of a series of measurements of an object micrometer performed with different microscopes can serve as an example. One has a *two- (or more) way analysis of variance* (two-way classification) when several variables are changed simultaneously. In the example above, if different observers carry out the series of measurements with each microscope, then a two-way analysis of variance can investigate influences of both the observer and the instrument on the result.

11.1 One-Way Analysis of Variance

Let us consider a sample of size n, which can be divided into t groups according to a certain criterion A. Clearly the criterion must be related to the sampling or measuring process. We say that the groups are constructed according to the *classification* A. We assume that the populations from which the t subsamples are taken are normally distributed with the same variance σ^2. We now want to test the hypothesis that the mean values of these populations are also equal. If this hypothesis is true, then all of the samples come from the same population. We can then apply the results of Sect. 6.4 (samples from subpopulations). Using the same notation as there, we have t groups of size n_i with

S. Brandt, *Data Analysis: Statistical and Computational Methods for Scientists and Engineers*,
DOI 10.1007/978-3-319-03762-2_11,

$$n = \sum_{i=1}^{t} n_i$$

and we write the jth element of the ith group as x_{ij}. The sample mean of the ith group is

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \tag{11.1.1}$$

and the mean of the entire sample is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{t} \sum_{j=1}^{n_i} x_{ij} = \frac{1}{n} \sum_{i=1}^{t} n_i \bar{x}_i \quad . \tag{11.1.2}$$

We now construct the *sum of squares*

$$\begin{aligned} Q &= \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i + \bar{x}_i - \bar{x})^2 \\ &= \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^{t} \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2 + 2 \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{x}) \quad . \end{aligned}$$

The last term vanishes because of (11.1.1) and (11.1.2). One therefore has

$$\begin{aligned} Q &= \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = \sum_{i=1}^{t} n_i (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \quad , \\ Q &= Q_A + Q_W \quad . \end{aligned} \tag{11.1.3}$$

The first term is the *sum of squares between the groups* obtained with the classification A. The second term is a sum over the *sums of squares within a group*. The sum of squares Q is decomposed into a sum of two sums of squares corresponding to different "sources" – the variation of means within the classification A and the variation of measurements within the groups. If our hypothesis is correct, then Q is a sum of squares from a normal distribution, i.e., Q/σ^2 follows a χ^2-distribution with $n-1$ degrees of freedom. Correspondingly, for each group the quantity

$$\frac{Q_i}{\sigma^2} = \frac{1}{\sigma^2} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

follows a χ^2-distribution with $n_i - 1$ degrees of freedom. The sum

$$\frac{Q_W}{\sigma^2} = \sum_{i=1}^{t} \frac{Q_i}{\sigma^2}$$

is then described by a χ^2-distribution with $\sum_i (n_i - 1) = n - t$ degrees of freedom (see Sect. 6.6). Finally, Q_A/σ^2 follows a χ^2-distribution with $t - 1$ degrees of freedom.

The expressions

$$\begin{aligned} s^2 &= \frac{Q}{n-1} = \frac{1}{n-1} \sum_i \sum_j (x_{ij} - \bar{x})^2 \quad , \\ s_A^2 &= \frac{Q_A}{t-1} = \frac{1}{t-1} \sum_i n_i (\bar{x}_i - \bar{x})^2 \quad , \\ s_W^2 &= \frac{Q_W}{n-t} = \frac{1}{n-t} \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 \end{aligned} \tag{11.1.4}$$

are unbiased estimators of the population variances. (In Sect. 6.5 we called such expressions mean squares.) The ratio

$$F = s_A^2 / s_W^2 \tag{11.1.5}$$

can thus be used to carry out an F-test.

If the hypothesis of equal means is false, then the values $\bar{x}_i$ of the individual groups will be quite different. Thus s_A^2 will be relatively large, while s_W^2, which is the mean of the variances of the individual groups, will not change much. This means that the ratio (11.1.5) will be large. Therefore one uses a one-sided F-test. The hypothesis of equal means is rejected at the significance level α if

$$F = s_A^2 / s_W^2 > F_{1-\alpha}(t-1, n-t) \quad . \tag{11.1.6}$$

The sums of squares can be computed according to two equivalent formulas,

$$\begin{aligned} Q &= \sum_i \sum_j (x_{ij} - \bar{x})^2 = \sum_i \sum_j x_{ij}^2 - n\bar{x}^2 \quad , \\ Q_A &= \sum_i n_i (\bar{x}_i - \bar{x})^2 = \sum_i n_i \bar{x}_i^2 - n\bar{x}^2 \quad , \\ Q_W &= \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 = \sum_i \sum_j x_{ij}^2 - \sum_i n_i \bar{x}_i^2 \quad . \end{aligned} \tag{11.1.7}$$

The expression on the right of each line is usually easier to compute. Since each sum of squares is obtained by computing the difference of two relatively large numbers, one must pay attention to possible problems with errors in rounding. Although one only needs the sums Q_A and Q_W in order to compute the ratio F, it is recommended to compute Q as well, since one can then perform a check using (11.1.3), i.e., $Q = Q_A + Q_W$. The check is only meaningful when Q is computed with the left-hand form of (11.1.3). Usually the results of an analysis of variance are summarized in a so-called *analysis of variance table*, (or ANOVA table) as shown in Table 11.1.

Before carrying out an analysis of variance one must consider whether the requirements are met under which the procedure has been derived. In particular, one must check the assumption of a normal distribution for the measurements within each group. This is by no means certain in every case. If, for example, the measured valuesare always positive (e.g., the length or weight

Table 11.1: ANOVA table for one-way classification.

Source	SS (sum of squares)	DF (degrees of freedom)	MS (mean square)	F
Between the groups	Q_A	$t-1$	$s_A^2 = \frac{Q_A}{t-1}$	
Within the groups	Q_W	$n-t$	$s_W^2 = \frac{Q_W}{n-t}$	$F = \frac{s_A^2}{s_W^2}$
Sum	Q	$n-1$	$s^2 = \frac{Q}{n-1}$	

of an object) and if the standard deviation is of a magnitude comparable to the measured values, then the probability density can be asymmetric and thus not Gaussian. If, however, the original measurements (let us denote them for the moment by x') are transformed using a monotonic transformation such as

$$x = a \log(x' + b) \quad , \tag{11.1.8}$$

where a and b are appropriately chosen constants, then a normal distribution can often be sufficiently well approximated. Other transformations sometimes used are $x = \sqrt{x'}$ or $x = 1/x'$.

Example 11.1: One-way analysis of variance of the influence of various drugs

The spleens of mice with cancer are often attacked particularly strongly. The weight of the spleen can thus serve as a measure of the reaction to various drugs. The drugs (I–III) were used to treat ten mice each. Table 11.2 contains the measured spleen weights, which have already been transformed according to $x = \log x'$, where x' is the weight in grams. Most of the calculation is presented in Table 11.2. Table 11.3 contains the resulting ANOVA table. Since even at a significance level of 50 % the F-test gives $F_{0.5}(2,24) = 3.4$, one cannot reject the hypothesis of equal mean values. The experiment thus showed no significant difference in the effectiveness of the three drugs. ∎

Table 11.2: Data for Example 11.1.

Experiment number	Group I	II	III	
1	19	40	32	
2	45	28	26	
3	26	26	30	
4	23	15	17	
5	36	24	23	
6	23	26	24	
7	26	36	29	
8	33	27	20	
9	22	28	—	
10	—	19	—	$\sum_i \sum_j x_{ij}^2 = 20\,607$
$\sum_j x_j$	253	269	201	$\sum_i \sum_j x_{ij} = 723$
n_i	9	10	8	$n = 27$
				$n\bar{x}^2 = 19\,360$
$\bar{x}_i$	28.11	26.90	25.13	$\bar{x} = 26.78$
$\bar{x}_i^2$	790.23	723.61	631.52	$\sum_i n_i \bar{x}_i^2 = 19\,398$

Table 11.3: ANOVA table for Example 11.1.

Source	SS	DF	MS	F
Between the groups	38	2	19.0	0.377
Within the groups	1209	24	50.4	
Sum	1247	26	47.8	

11.2 Two-Way Analysis of Variance

Before we turn to analysis of variance with two external variables, we would like to examine more carefully the results obtained for one-way classification. We denoted the jth measurement of the quantity x in group i by x_{ij}. We now assume for simplicity that each group contains the same number of measurements, i.e., $n_i = J$. In addition we denote the total number of groups by I. The classification into individual groups was done according to the criterion A, e.g., the production number of a microscope, by which the groups can be distinguished. The labeling according to measurement and group is illustrated in Table 11.4.

We can write the individual means of the groups in the form

$$\begin{aligned} \bar{x}_{..} &= \bar{x} = \frac{1}{IJ}\sum_i\sum_j x_{ij} \quad , \\ \bar{x}_{i.} &= \frac{1}{J}\sum_j x_{ij} \quad , \\ \bar{x}_{.j} &= \frac{1}{I}\sum_i x_{ij} \quad . \end{aligned} \tag{11.2.1}$$

Table 11.4: One-way classification.

Measurement number	Classification A					
	A_1	A_2	$\ldots$	A_i	$\ldots$	A_I
1	x_{11}	x_{22}		x_{i1}		x_{I1}
2	x_{12}	x_{22}		x_{i2}		x_{I2}
$\vdots$						
j	x_{1j}	x_{2j}		x_{ij}		x_{Ij}
$\vdots$						
J	x_{1J}	x_{2J}		x_{iJ}		x_{IJ}

Here a point denotes summation over the index that it replaces. This notation allows a simple generalization to a larger number of indices. The analysis of variance with one-way classification is based on the assumption that the measurements within a group only differ by the measurement errors, which follow a normal distribution with a mean of zero and variance σ^2. That is, we consider the *model*

$$x_{ij} = \mu_i + \varepsilon_{ij} \quad . \tag{11.2.2}$$

The goal of an analysis of variance was to test the hypothesis

$$H_0(\mu_1 = \mu_2 = \ldots = \mu_I = \mu). \tag{11.2.3}$$

By choosing measurements out of a certain group i and by applying the maximum-likelihood method to (11.2.2) one obtains the estimator

$$\widetilde{\mu}_i = \bar{x}_{i.} = \frac{1}{J}\sum_j x_{ij} \quad . \tag{11.2.4}$$

If H_0 is true, then one has

$$\widetilde{\mu} = \bar{x} = \frac{1}{IJ}\sum_i\sum_j x_{ij} = \frac{1}{I}\sum_i \widetilde{\mu}_i \quad . \tag{11.2.5}$$

The (composite) alternative hypothesis is that not all of the μ_i are equal. We want, however, to retain the concept of the overall mean and we write

$$\mu_i = \mu + a_i \quad .$$

The model (11.2.2) then has the form

$$x_{ij} = \mu + a_i + \varepsilon_{ij} . \tag{11.2.6}$$

Between the quantities a_i, which represent a measure of the deviation of the mean for the ith group from the overall mean, one has the relation

$$\sum_i a_i = 0 \quad . \tag{11.2.7}$$

The maximum-likelihood estimators for the a_i are

$$\widetilde{a}_i = \bar{x}_{i.} - \bar{x} \quad . \tag{11.2.8}$$

The one-way analysis of variance of Sect. 11.1 was derived from the identity

$$x_{ij} - \bar{x} = (\bar{x}_{i.} - \bar{x}) + (x_{ij} - \bar{x}_{i.}) \quad , \tag{11.2.9}$$

which describes the deviation of the individual measurements from the overall mean. The sum of squares Q of these deviations could then be decomposed into the terms Q_A and Q_W; cf. (11.1.3).

After this preparation we now consider a two-way classification, where the measurements are divided into groups according to two criteria, A and B. The measurements x_{ijk} belong to class A_i, which is given by the classification according to A, and also to class B_j. The index k denotes the measurement number within the group that belongs to both class A_i and class B_j.

A two-way classification is said to be *crossed*, when a certain classification B_j has the same meaning for all classes A. If, for example, microscopes are classified by A and observers by B, and if each observer carries out a measurement with each microscope, then the classifications are crossed. If, however, one compares the microscopes in different laboratories, and if therefore in each laboratory a different group of J observers makes measurements with a certain microscope i, then the classification B is said to be *nested* in A. The index j then merely counts the classes B within a certain class A.

The simplest case is a crossed classification with only *one observation*. Since then $k = 1$ for all observations x_{ijk}, we can drop the index k. One uses the model

$$x_{ij} = \mu + a_i + b_j + \varepsilon_{ij} \quad , \qquad \sum_i a_i = 0 \quad , \qquad \sum_j b_j = 0 \quad , \tag{11.2.10}$$

where ε is normally distributed with mean zero and variance σ^2. The null hypothesis says that by classification according to A or B, no deviation from the overall mean occurs. We write this in the form of two individual hypotheses,

$$H_0^{(A)}(a_1 = a_2 = \cdots = a_I = 0) \ , \quad H_0^{(B)}(b_1 = b_2 = \cdots = b_J = 0) \ . \tag{11.2.11}$$

The least-squares estimators for a_i and b_j are

$$\widetilde{a}_i = \bar{x}_{i.} - \bar{x} \quad , \qquad \widetilde{b}_j = \bar{x}_{.j} = \bar{x} \quad .$$

In analogy to Eq. (11.2.9) we can write

$$x_{ij} - \bar{x} = (\bar{x}_{i.} - \bar{x}) + (\bar{x}_{.j} - \bar{x}) + (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}) \quad . \tag{11.2.12}$$

In a similar way the sum of squares can be written

$$\sum_i \sum_j (x_{ij} - \bar{x})^2 = Q = Q_A + Q_B + Q_W \quad , \tag{11.2.13}$$

where

$$\begin{aligned} Q_A &= J \sum_i (\bar{x}_{i.} - \bar{x})^2 = J \sum_i \bar{x}_{i.}^2 - IJ\bar{x}^2 \quad , \\ Q_B &= I \sum_j (\bar{x}_{.j} - \bar{x})^2 = I \sum_j \bar{x}_{.j}^2 - IJ\bar{x}^2 \quad , \\ Q_W &= \sum_i \sum_j (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2 \\ &= \sum_i \sum_j x_{ij}^2 - J \sum_i \bar{x}_{i.}^2 - I \sum_j \bar{x}_{.j}^2 + IJ\bar{x}^2 \quad . \end{aligned} \tag{11.2.14}$$

When divided by the corresponding number of degrees of freedom, these sums are estimators of σ^2, providing that the hypotheses (11.2.11) are correct. The hypotheses $H_0^{(A)}$ and $H_0^{(B)}$ can be tested individually by using the ratio

$$F^{(A)} = s_A^2 / s_W^2 \quad , \qquad F^{(B)} = s_B^2 / s_W^2 \quad . \tag{11.2.15}$$

Here one uses one-sided F-tests as in Sect. 11.1. The overall situation can be summarized in an ANOVA table (Table 11.5).

If more than one observation is made in each group, then the crossed classification can be generalized in various ways. The most important generalization involves *interaction* between the classes. One then has the model

$$x_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \varepsilon_{ijk} \quad . \tag{11.2.16}$$

The quantity $(ab)_{ij}$ is called the *interaction* between the classes A_i and B_j. It describes the deviation from the group mean that occurs because of the specific interaction of A_i and B_j. The parameters a_i, b_j, $(ab)_{ij}$ are related by

$$\sum_i a_i = \sum_j b_j = \sum_i \sum_j (ab)_{ij} = 0 \quad . \tag{11.2.17}$$

Their maximum-likelihood estimators are

$$\begin{aligned} \widetilde{a}_i = \bar{x}_{i..} - \bar{x} \quad , \qquad \widetilde{b}_j = \bar{x}_{.j.} - \bar{x} \quad , \\ (\widetilde{ab})_{ij} = \bar{x}_{ij.} + \bar{x} - \bar{x}_{i..} - \bar{x}_{.j.} \quad . \end{aligned} \tag{11.2.18}$$

Table 11.5: Analysis of variance table for crossed two-way classification with only one observation.

Source	SS	DF	MS	F
Class. A	Q_A	$I-1$	$s_A^2 = \dfrac{Q_A}{I-1}$	$F^{(A)} = \dfrac{s_A^2}{s_W^2}$
Class. B	Q_B	$J-1$	$s_B^2 = \dfrac{Q_B}{J-1}$	$F^{(B)} = \dfrac{s_B^2}{s_W^2}$
Within groups	Q_W	$(I-1)(J-1)$	$s_W^2 = \dfrac{Q_W}{(I-1)(J-1)}$	
Sum	Q	$IJ-1$	$s^2 = \dfrac{Q}{IJ-1}$	

The null hypothesis can be divided into three individual hypotheses,

$$\begin{aligned} H_0^{(A)}(a_i = 0;\ i = 1,2,\ldots,I) \quad , \qquad H_0^{(B)}(b_j = 0;\ j = 1,2,\ldots,J) \quad , \\ H_0^{(AB)}((ab)_{ij} = 0;\ i = 1,2,\ldots,I;\ j = 1,2,\ldots,J) \quad , \end{aligned} \tag{11.2.19}$$

which can then be tested individually. The analysis of variance is based on the identity

$$\begin{aligned} x_{ijk} - \bar{x} &= (\bar{x}_{i..} - \bar{x}) + (\bar{x}_{.j.} - \bar{x}) \\ &\quad + (\bar{x}_{ij.} + \bar{x} - \bar{x}_{i..} - \bar{x}_{.j.}) + (x_{ijk} - \bar{x}_{ij.}) \quad , \end{aligned} \tag{11.2.20}$$

which allows the decomposition of the sum of squares of deviations into four terms,

$$\begin{aligned} Q &= \sum_i \sum_j \sum_k (x_{ijk} - \bar{x})^2 \\ &= Q_A + Q_B + Q_{AB} + Q_W \quad , \qquad (11.2.21) \\ Q_A &= JK \sum_i (\bar{x}_{i..} - \bar{x})^2 \quad , \\ Q_B &= IK \sum_j (\bar{x}_{.j.} - \bar{x})^2 \quad , \\ Q_{AB} &= K \sum_i \sum_j (\bar{x}_{ij.} + \bar{x} - \bar{x}_{i..} - \bar{x}_{.j.})^2 \quad , \\ Q_W &= \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2 \quad . \end{aligned}$$

The degrees of freedom and mean squares as well as the F-ratio, which can be used for testing the hypotheses, are given in Table 11.6.

Table 11.6: Analysis of variance table for crossed two-way classification.

Source	SS	DF	MS	F
Class. A	Q_A	$I-1$	$s_A^2 = \frac{Q_A}{I-1}$	$F^{(A)} = \frac{s_A^2}{s_W^2}$
Class. B	Q_B	$J-1$	$s_B^2 = \frac{Q_B}{J-1}$	$F^{(B)} = \frac{s_B^2}{s_W^2}$
Interaction	Q_{AB}	$(I-1)(J-1)$	$s_{AB}^2 = \frac{Q_{AB}}{(I-1)(J-1)}$	$F^{(AB)} = \frac{s_{AB}^2}{s_W^2}$
Within groups	Q_W	$IJ(K-1)$	$s_W^2 = \frac{Q_W}{IJ(K-1)}$	
Sum	Q	$IJK-1$	$s^2 = \frac{Q}{IJK-1}$	

Finally, we will give the simplest case of a *nested two-way classification*. Because the classification B is only defined within the individual classes of A, the terms b_j and $(ab)_{ij}$ from Eq. (11.2.10) are not defined, since they imply a sum over i for fixed j. Therefore one uses the model

$$x_{ijk} = \mu + a_i + b_{ij} + \varepsilon_{ijk} \qquad (11.2.22)$$

with

$$\begin{aligned} \textstyle\sum_i a_i &= 0 \quad , & \textstyle\sum_i \sum_j b_{ij} &= 0 \quad , \\ \widetilde{a}_i &= \bar{x}_{i..} - \bar{x} \quad , & \widetilde{b}_{ij} &= \bar{x}_{ij.} - \bar{x}_{i..} \quad . \end{aligned}$$

The term b_{ij} is a measure of the deviation of the measurements of class B_j within class A_i from the overall mean of class A_i. The null hypothesis consists of

$$\begin{aligned} &H_0^{(A)}(a_i = 0;\ i = 1,2,\ldots,I) \quad , \\ &H_0^{(B(A))}(b_{ij} = 0;\ i = 1,2,\ldots,I;\ j = 1,2,\ldots,J) \quad . \end{aligned} \qquad (11.2.23)$$

An analysis of variance for testing these hypotheses can be carried out with the help of Table 11.7. Here one has

$$\begin{aligned} Q_A &= JK\sum_i (\bar{x}_{i..} - \bar{x})^2 \quad , \\ Q_{B(A)} &= K\sum_i\sum_j (\bar{x}_{ij.} - \bar{x}_{i..})^2 \quad , \\ Q_W &= \sum_i\sum_j\sum_k (x_{ijk} - \bar{x}_{ij.})^2 \quad , \\ Q &= Q_A + Q_{B(A)} + Q_W = \sum_i\sum_j\sum_k (x_{ijk} - \bar{x})^2 \quad . \end{aligned}$$

In a similar way one can construct various models for two-way or multiple classification. For each model the total sum of squares is decomposed into a certain sum of individual sums of squares, which, when divided by the corresponding number of degrees of freedom, can be used to carry out an F-test. With this, the hypotheses implied by the model can be tested.

Some models are, at least formally, contained within others. For example one finds by comparing Tables 11.6 and 11.7 the relation

$$Q_{B(A)} = Q_B + Q_{AB} \quad . \qquad (11.2.24)$$

A similar relation holds for the corresponding number of degrees of freedom,

$$f_{B(A)} = f_B + f_{AB} \quad . \qquad (11.2.25)$$

Table 11.7: Analysis of variance table for nested two-way classification.

Source	SS	DF	MS	F
Class. A	Q_A	$I-1$	$s_A^2 = \dfrac{Q_A}{I-1}$	$F^{(A)} = \dfrac{s_A^2}{s_W^2}$
Within A	$Q_{B(A)}$	$I(J-1)$	$s_{B(A)}^2 = \dfrac{A_{B(A)}}{I(J-1)}$	$F^{(B(A))} = \dfrac{s_{B(A)}^2}{s_W^2}$
Within groups	Q_W	$IJ(K-1)$	$s_W^2 = \dfrac{Q_W}{IJ(K-1)}$	
Sum	Q	$IJK-1$	$s^2 = \dfrac{Q}{IJK-1}$	

Example 11.2: Two-way analysis of variance in cancer research

Two groups of rats are injected with the amino acid thymidine containing traces of tritium, a radioactive isotope of hydrogen. In addition, one of the groups receives a certain carcinogen. The incorporation of thymidine into the skin of the rats is investigated as a function of time by measuring the number of tritium decays per unit area of skin. The classifications are crossed since the time dependence is controlled in the same way for both series of test animals. The measurements are compiled in Table 11.8. The numbers are already transformed from the original counting rates x' according to $x = 50 \log x' - 100$. The results, obtained with the class AnalysisOfVariance, are shown in Table 11.9. There is no doubt that the presence or absence of the carcinogen (classification A)has an influence on the result, since the ratio $F^{(A)}$ is very large. We now want to test the existence of a time dependence (classification B) and of an interaction between A and B at a significance level of $\alpha = 0.01$. From Table I.8 we find $F_{0.99} = 2.72$. The hypotheses of time independence and vanishing interaction must therefore be rejected. Table 11.9 also contains the values of α for which the hypothesis would not need to be rejected. They are very small. ■

Table 11.8: Data for Example 11.2.

Obs.	Injection	Time after injection (h)									
no.		4	8	12	16	20	24	28	32	36	48
1		34	54	44	51	62	61	59	66	52	52
2	Thymidine	40	57	52	46	61	70	67	59	63	50
3		38	40	53	51	54	64	58	67	60	44
4		36	43	51	49	60	68	66	58	59	52
1		28	23	42	43	31	32	25	24	26	26
2	Thymidine	32	23	41	48	45	38	27	26	31	27
3	and	34	29	34	36	41	32	27	32	25	27
4	Carcinogen	27	30	39	43	37	34	28	30	26	30

Table 11.9: Printout from Example 11.2.

Analysis of variance table

Source	Sum of squares	Degrees of freedom	Mean square	F Ratio	Alpha
A	9945.80	1	9945.80	590.547 25	0.00E−10
B	1917.50	9	213.06	12.650 50	0.54E−10
INT.	2234.95	9	248.33	14.744 85	0.03E−10
W	1010.50	60	16.84		
TTL.	15 108.75	79	191.25		

11.3 Java Class and Example Programs

Java Class

AnalysisOfVariance performs a crossed as well as a nested two-way analysis of variance.

Example Program 11.1: The class E1Anova demonstrates the use of AnalysisOfVariance

The short program analyses the data of Example 11.2. Data and output are presented as in Table 11.9.

Example Program 11.2: The class E2Anova simulates data and performs an analysis of variance on them

The program allows interactive input of numerical values for σ, the quantities I, J, K and three further parameters: $\Delta_i, \Delta_j, \Delta_k$. It generates data of the simple form

$$x_{ijk} = i\Delta_i + j\Delta_j + k\Delta_k + \varepsilon_{ijk} \quad .$$

Here the quantities ε_{ijk} are taken from a normal distribution with zero mean standard deviation σ. An analysis of variance is performed on the data and the results are presented for crossed and for nested two-way classification.

Suggestion: Take a value $\neq 0$ for only one of the parameters $\Delta_i, \Delta_j, \Delta_k$ and interpret the resulting analysis of variance table.