PRACTICAL STATISTICS FOR MEDICAL RESEARCH

Douglas G. Altman

Chapman and Hall

$$t = \frac{316.6 - 256.4}{45.72 \times \sqrt{\frac{1}{8} + \frac{1}{9}}}$$

$$= 2.71 \text{ on 15 degrees of freedom.}$$

If we are comparing each pair of groups we will make three comparisons. The above t value of 2.71 corresponds to $P < 0.02$ (Table B4), with an exact value of $P = 0.016$. The corrected P value is $P' = 0.016 \times 3 = 0.048$ so it is just significant at the 5% level after adjustment. Neither of the other comparisons is significant. The main explanation for the difference between the groups that was identified in the analysis of variance (Table 9.11) is thus the difference between groups I and II.

9.8.5 Ordered groups

When the groups are ordered it is not reasonable to compare each pair of groups, but rather we should study the possibility that there is a trend across groups. For many purposes it will suffice to consider whether there is a *linear* trend.

Table 9.12 shows the mean and standard deviation of serum trypsin levels in healthy volunteers divided into six age groups. We can carry out one way analysis of variance from these summary statistics without having the raw observations, using the formulae given in section 9.9, to get the results shown in Table 9.13. (Unfortunately, very few statistical packages

Table 9.12 Serum levels of immunoreactive trypsin in healthy volunteers divided into six age groups (based on data given by Koehn and Mostbeck, 1981)

	Age					
	10–19	20–29	30–39	40–49	50–59	60–69
Number of subjects	32	137	38	44	16	4
Mean (ng/ml)	128	152	194	207	215	218
Standard deviation (ng/ml)	50.9	58.5	49.3	66.3	60.0	14.0

Table 9.13 One way analysis of variance of data in Table 9.12

Source of variation	df	Sums of squares	Mean squares	*F*	P
Between groups	5	224 103	44 820.6	13.5	< 0.0001
Within groups	265	879 272	3318.0		
Total	270	1 103 375			

will perform analysis of variance using means and standard deviations that are already calculated.) Clearly there is highly significant variation among the six age groups. However, we can go further by 'partitioning' the variability between groups into components. Here we would be more interested in whether there was a linear trend, that is whether serum trypsin values tend to rise at a constant rate with increasing age.

Using the formula given in section 9.9 we find that the sum of squares associated with a linear trend is 55 147 on one degree of freedom, so the analysis of variance table can be rewritten as shown in Table 9.14. There is a highly significant linear trend, showing that mean serum trypsin level rises with age. However, the non-linear variation between the age groups is also highly significant, indicating that the linear trend only explains some of the age effect. Fitting a linear trend in one way analysis of variance is equivalent to linear regression analysis, which is described in Chapter 11.

Table 9.14 Analysis of variance table showing test for linear trend

Source of variation	df	Sums of squares	Mean squares	F	P
Between groups:	5	224 103	44 820.6		
(a) linear	1	55 147	55 147.0	16.6	< 0.0001
(b) non-linear	4	168 956	42 239.0	12.7	< 0.0001
Within groups:	265	879 272	3 318.0		
Total	270	1 103 375			

9.8.6 Non-parametric one way analysis of variance – the Kruskal-Wallis test

Just as analysis of variance is a more general form of t test, so there is a more general form of the non-parametric Mann-Whitney test. The **Kruskal-Wallis test** is an obvious mathematical extension of the Mann-Whitney test, with the same problems of interpretation as were just discussed for one way analysis of variance.

The calculation of the test statistic is simple. The complete set of N observations are ranked from 1 to N regardless of which group they are in, and for each group the sum of the ranks is calculated. If the sum of the ranks of n_i observations in the ith group is R_i, then the average rank in each group is given by $\bar{R}_i$. We calculate the statistic H defined by

$$H = \frac{12\Sigma n_i(\bar{R}_i - \bar{R})^2}{N(N+1)}$$

where $\bar{R}$ is the average of all the ranks, and is always equal to $(N+1)/2$. The summation term in this formula is very similar to the between group

(b) Means and standard deviations available
If we already have the mean (M_i) and standard deviation (s_i) for each group of size n_i we can use the above formulae for T and B together with a simpler method of calculating the within groups sum of squares, W, as

$$W = \sum_{i=1}^{k} (n_i - 1)s_i^2.$$

9.9.2 Linear trend

If there is a natural ordering of the groups, the between groups sum of squares can be partitioned into a component due to a linear trend, and the remaining (non-linear) component. We give scores l_i to the groups, where the values of the l_i are equally spaced and chosen so that their sum is zero. We then calculate

$$L = \sum l_i \bar{y}_i$$

and its standard error

$$se(L) = s_{res}\sqrt{\Sigma(l_i^2/n_i)}.$$

A one sample t test can be performed by comparing $L/se(L)$ to the t distribution with the number of degrees of freedom within groups.

Alternatively, the sum of squares due to L can be calculated as

$$SS(L) = L^2/\sum(l_i^2/n_i)$$

and the analysis of variance table recalculated by partitioning the between group sum of squares into linear and non-linear components. The F test for the linear contrast is exactly equivalent to the above t test.

(This method is equivalent to performing a regression analysis with the l_i as explanatory variable – see section 11.10.)

9.9.3 Worked example

For the serum trypsin data in Table 9.12 the sum of the 271 observations is given by

$$T = 32 \times 128 + 137 \times 152 + 38 \times 194 + 44 \times 207 + 16 \times 215 + 4 \times 218$$
$$= 45712.$$

The within groups sum of squares is obtained from the formula based on standard deviations as

$$W = 31 \times 50.9^2 + 136 \times 58.5^2 + \ldots + 3 \times 14.0^2 = 879271.9$$

and the quantity $\Sigma n_i M_i^2$ is

$$32 \times 128^2 + 137 \times 152^2 + \ldots + 4 \times 218^2 = 7934756.$$

The between groups sum of squares is thus

$$B = 7934756 - 45712^2/271$$
$$= 224103.1.$$

The complete analysis of variance table is shown in Table 9.13. The residual standard deviation is $\sqrt{3318} = 57.602$. To evaluate a possible linear trend we give the groups scores l_i which are equally spaced and add to zero, such as −5, −3, −1, 1, 3, and 5. The value of the linear contrast is then

$$L = -5 \times 128 - 3 \times 152 - 1 \times 194 + 1 \times 207 + 3 \times 215 + 5 \times 218$$
$$= 652$$

and its standard error is

$$se(L) = 57.602 \times \sqrt{\frac{(-5)^2}{32} + \frac{(-3)^2}{137} + \frac{(-1)^2}{38} + \frac{1^2}{44} + \frac{3^2}{16} + \frac{5^2}{4}}$$
$$= 159.93.$$

The calculations for fitting a linear trend across age groups are shown in Table 9.17. The t test for the linear contrast gives $t = 652/159.93 = 4.08$ (P = 0.00006).

Alternatively, the sum of squares for L is $652.0^2/7.7085 = 55147$, as shown in Table 9.14. The F test is exactly equivalent to the t test above, as is shown by the value of F (16.6) being equal to the square of the value of t (4.08).

Table 9.17 Calculating the sum of squares for linear trend in serum trypsin data from Table 9.12

Group	n	$\bar{y}_i$	l_i	$l_i\bar{y}_i$	l_i^2/n_i
1	32	128	−5	−640.0	0.78125
2	137	152	−3	−456.0	0.06569
3	38	194	−1	−194.0	0.02632
4	44	207	1	207.0	0.02273
5	16	215	3	645.0	0.56250
6	4	218	5	1090.0	6.25000
Total				652.0	7.70849

9.10 PRESENTATION OF RESULTS

It is never sufficient to present the results of a statistical analysis solely as a P value, or even as a test statistic and P value. Some actual results should be quoted. This chapter has been concerned with continuous data, for