The reason here is similar to that advanced above. An estimate of variance s_2^2 on $\nu_2 = \infty$ DF must be exactly equal to the population variance σ^2. Thus, F may be written as $s_1^2/\sigma^2 = \chi^2_{(\nu_1)}/\nu_1$ (see p. 149).

The F test and the associated confidence limits provide an exact treatment of the comparison of two variance estimates from two independent *normal* samples. Unfortunately, the methods are rather sensitive to the assumption of normality—much more so than in the corresponding uses of the t distribution to compare two means. This defect is called a lack of *robustness*.

The methods described in this section are appropriate only for the comparison of two *independent* estimates of variances. Sometimes this condition fails because the observations in the two samples are paired, as in the first situation considered in §4.2. The appropriate method for this case makes use of a technique described in Chapter 7, and is therefore postponed until p. 203.

A different use of the F distribution has already been noted on p. 117.

5.2 Inferences from counts

Suppose that x is a count, say, of the number of events occurring during a certain period or a number of small objects observed in a biological specimen, which can be assumed to follow the Poisson distribution with mean μ (§3.7). What can be said about μ?

Suppose first that we wish to test a null hypothesis specifying that μ is equal to some value μ_0. On this hypothesis, x would follow a Poisson distribution with expectation μ_0. The departure of x from its expected value, μ_0, is measured by the extent to which x falls into either of the tails of the hypothesized distribution. The situation is similar to that of the binomial (§3.6). Thus if $x > \mu_0$ and the probabilities in the Poisson distribution are $P_0, P_1, \ldots$, the P value for a one-sided test will be

$$\begin{aligned} P_+ &= P_x + P_{x+1} + P_{x+2} + \ldots \\ &= 1 - P_0 - P_1 - \ldots - P_{x-1}. \end{aligned}$$

The possible methods of constructing a two-sided test follow the same principles as for the binomial in §3.6.

Again considerable simplification is achieved by approximating the Poisson distribution by the normal (§3.8). On the null hypothesis, and including a continuity correction,

$$z = \frac{|x - \mu_0| - \frac{1}{2}}{\sqrt{\mu_0}} \tag{5.5}$$

is approximately a standardized normal deviate. Excluding the continuity correction corresponds to the mid-P value obtained by including only $\frac{1}{2}P_x$ in the summation of Poisson probabilities.

The reason here is similar to that advanced above. An estimate of variance s_2^2 on $\nu_2 = \infty$ DF must be exactly equal to the population variance σ^2. Thus, F may be written as $s_1^2/\sigma^2 = \chi^2_{(\nu_1)}/\nu_1$ (see p. 149).

The F test and the associated confidence limits provide an exact treatment of the comparison of two variance estimates from two independent *normal* samples. Unfortunately, the methods are rather sensitive to the assumption of normality—much more so than in the corresponding uses of the t distribution to compare two means. This defect is called a lack of *robustness*.

The methods described in this section are appropriate only for the comparison of two *independent* estimates of variances. Sometimes this condition fails because the observations in the two samples are paired, as in the first situation considered in §4.2. The appropriate method for this case makes use of a technique described in Chapter 7, and is therefore postponed until p. 203.

A different use of the F distribution has already been noted on p. 117.

5.2 **Inferences from counts**

Suppose that x is a count, say, of the number of events occurring during a certain period or a number of small objects observed in a biological specimen, which can be assumed to follow the Poisson distribution with mean μ (§3.7). What can be said about μ?

Suppose first that we wish to test a null hypothesis specifying that μ is equal to some value μ_0. On this hypothesis, x would follow a Poisson distribution with expectation μ_0. The departure of x from its expected value, μ_0, is measured by the extent to which x falls into either of the tails of the hypothesized distribution. The situation is similar to that of the binomial (§3.6). Thus if $x > \mu_0$ and the probabilities in the Poisson distribution are $P_0, P_1, \ldots$, the P value for a one-sided test will be

$$\begin{aligned} P_+ &= P_x + P_{x+1} + P_{x+2} + \ldots \\ &= 1 - P_0 - P_1 - \ldots - P_{x-1}. \end{aligned}$$

The possible methods of constructing a two-sided test follow the same principles as for the binomial in §3.6.

Again considerable simplification is achieved by approximating the Poisson distribution by the normal (§3.8). On the null hypothesis, and including a continuity correction,

$$z = \frac{|x - \mu_0| - \frac{1}{2}}{\sqrt{\mu_0}} \tag{5.5}$$

is approximately a standardized normal deviate. Excluding the continuity correction corresponds to the mid-P value obtained by including only $\frac{1}{2}P_x$ in the summation of Poisson probabilities.

Example 5.2

In a study of asbestos workers a large group was followed over several years and 33 died of lung cancer. Making allowance for age, using national death rates, the expected number of deaths due to lung cancer was 20·0. How strong is this evidence that there is an excess risk of death due to lung cancer?

On the null hypothesis that the national death rates applied, the standard error of x is $\sqrt{20{\cdot}0} = 4{\cdot}47$. The observed deviation is $33 - 20{\cdot}0 = 13{\cdot}0$. With continuity correction, the standardized normal deviate is $(13{\cdot}0 - 0{\cdot}5)/4{\cdot}47 = 2{\cdot}80$, giving a one-sided normal tail area of 0·0026. The exact one-sided value of P, from the Poisson distribution, is 0·0047, so the normal test exaggerated the significance. Two-sided values may be obtained by doubling these values, and both methods show that the evidence of excess mortality due to lung cancer is strong.

The exact one-sided mid-P value is 0·0037 and the corresponding standardized normal deviate is $13{\cdot}0/4{\cdot}47 = 2{\cdot}91$, giving a one-sided level of 0·0018.

The 95% confidence limits for μ are the two values, μ_L and μ_U, for which x is just significant by a one-sided test at the $2\frac{1}{2}\%$ level. These values may be obtained from tables of the Poisson distribution (e.g. Pearson & Hartley, 1966, Table 7) and Bailar and Ederer (1964) give a table of confidence factors. Table VIII1 of Fisher and Yates (1963) may also be used.

The normal approximation may be used in similar ways to the binomial case.

1 The tail areas could be estimated from (5.5). Thus approximations to the 95% limits are given by

$$\frac{x - \mu_L - \frac{1}{2}}{\sqrt{\mu_L}} = 1{\cdot}96$$

and

$$\frac{x - \mu_U + \frac{1}{2}}{\sqrt{\mu_U}} = -1{\cdot}96.$$

2 If x is large the continuity correction in method **1** may be omitted.

3 Replace $\sqrt{\mu_L}$ and $\sqrt{\mu_U}$ by $\sqrt{x}$. This is only satisfactory for large values (greater than 100).

The exact limits may be obtained by using tables of the χ^2 distribution. This follows from the mathematical link between the Poisson and the χ^2 distributions (see Liddell, 1984). The limits are

$$\left.\begin{aligned} \mu_L &= \tfrac{1}{2}\chi^2_{2x,\,0{\cdot}975} \\ \text{and} \\ \mu_U &= \tfrac{1}{2}\chi^2_{2x+2,\,0{\cdot}025} \end{aligned}\right\}. \tag{5.6}$$

Example 5.2, continued

With $x = 33$, the exact 95% confidence limits are found to be 22·7 and 46·3. Method **1** gives 23·1 and 46·9, method **2** gives 23·5 and 46·3, and method **3** gives 21·7 and 44·3. In this example methods **1** and **2** are adequate. The 95% confidence limits for the relative death rate due to lung cancer, expressed as the ratio of observed to expected, are 22·7/20·0 and 46·3/20·0 = 1·14 and 2·32. The mid-P limits are obtained from method **2** as 23·5/20·0 and 46·3/20·0 = 132.

Example 5.3

As an example where exact limits should be calculated, suppose that, in a similar situation to Example 5.2, there were two deaths compared with an expectation of 0·5. Then

$$\mu_L = \tfrac{1}{2}\chi^2_{4,0.975} = 0.24$$

and

$$\mu_U = \tfrac{1}{2}\chi^2_{6,0.025} = 7.22.$$

The limits for the ratio of observed to expected deaths are 0·24/0·5 and 7·22/0·5 = 0·5 and 14·4. The mid-P limits of μ may be obtained by trial and error on a programmable calculator or personal computer as those values for which $P(x = 0) + P(x = 1) + \frac{1}{2}P(x = 2) = 0.975$ or 0·025. This gives $\mu_L = 0.335$ and $\mu_U = 6.61$ so that the mid-P limits of the ratio of observed to expected deaths are 0·7 and 13·2. The evidence of excess mortality is weak but the data do not exclude the possibility of a large excess.

Suppose that in Example 5.3 there had been no deaths, then there is some ambiguity on the calculation of a 95% confidence interval. The point estimate of μ is zero and, since the lower limit cannot exceed the point estimate and also cannot be negative, its only possible value is zero. There is a probability of zero that the lower limit exceeds the true value of μ instead of the nominal value of $2\frac{1}{2}$%, and a possibility is to calculate the upper limit as $\mu_U = \frac{1}{2}\chi^2_{2,0.05} = 3.00$, rather than as $\frac{1}{2}\chi^2_{2,0.025} = 3.69$, so that the probability that the upper limit is less than the true value is approximately 5%, and the interval has approximately 95% coverage. Whilst this is logical and provides the narrowest 95% confidence interval it seems preferable that the upper limit corresponds to $2\frac{1}{2}$% in the upper tail to give a uniform interpretation. It turns out that the former value, $\mu = 3.00$, is the upper mid-P limit. Whilst it is impossible to find a lower limit with this interpretation, this is clear from the fact that the limit equals the point estimate and that both are at the extreme of possible values. This rationale is similar to that in our recommendation that a two-sided significance level should be double the one-sided level.

Comparison of two counts

Suppose that x_1 is a count which can be assumed to follow a Poisson distribution with mean μ_1. Similarly let x_2 be a count independently following a Poisson distribution with mean μ_2. How might we test the null hypothesis that $\mu_1 = \mu_2$?

One approach would be to use the fact that the variance of $x_1 - x_2$ is $\mu_1 + \mu_2$ (by virtue of (3.19) and (4.9)). The best estimate of $\mu_1 + \mu_2$ on the basis of the available information is $x_1 + x_2$. On the null hypothesis $\mathrm{E}(x_1 - x_2) = \mu_1 - \mu_2 = 0$, and $x_1 - x_2$ can be taken to be approximately normally distributed unless μ_1 and μ_2 are very small. Hence,

$$z = \frac{x_1 - x_2}{\sqrt{(x_1 + x_2)}} \tag{5.7}$$

can be taken as approximately a standardized normal deviate.

A second approach has already been indicated in the test for the comparison of proportions in paired samples (§4.5). Of the total frequency $x_1 + x_2$, a portion x_1 is observed in the first sample. Writing $r = x_1$ and $n = x_1 + x_2$ in (4.17) we have

$$z = \frac{x_1 - \frac{1}{2}(x_1 + x_2)}{\frac{1}{2}\sqrt{(x_1 + x_2)}} = \frac{x_1 - x_2}{\sqrt{(x_1 + x_2)}}$$

as in (5.7). The two approaches thus lead to exactly the same test procedure.

A third approach uses a rather different application of the χ^2 test from that described for the 2×2 table in §4.5, the total frequency of $x_1 + x_2$ now being divided into two components rather than four. Corresponding to each observed frequency we can consider the expected frequency, on the null hypothesis, to be $\frac{1}{2}(x_1 + x_2)$:

Observed	x_1	x_2
Expected	$\frac{1}{2}(x_1 + x_2)$	$\frac{1}{2}(x_1 + x_2)$

Applying the usual formula (4.30) for a χ^2 statistic, we have

$$\begin{aligned} X^2 &= \frac{[x_1 - \frac{1}{2}(x_1 + x_2)]^2}{\frac{1}{2}(x_1 + x_2)} + \frac{[x_2 - \frac{1}{2}(x_1 + x_2)]^2}{\frac{1}{2}(x_1 + x_2)} \\ &= \frac{(x_1 - x_2)^2}{x_1 + x_2}. \end{aligned} \tag{5.8}$$

As for (4.30) X^2 follows the $\chi^2_{(1)}$ distribution, which we already know to be the distribution of the square of a standardized normal deviate. It is therefore not surprising that X^2 given by (5.8) is precisely the square of z given by (5.7). The third approach is thus equivalent to the other two, and forms a particularly useful method of computation since no square root is involved

Consider now an estimation problem. What can be said about the ratio μ_1/μ_2? The second approach described above can be generalized, when the null hypothesis is not necessarily true, by saying that x_1 follows a binomial distribution with parameters $x_1 + x_2$ (the n of §3.7) and $\mu_1/(\mu_1 + \mu_2)$ (the π of §3.6). The methods of §4.4 thus provide confidence limits for $\pi = \mu_1/(\mu_1 + \mu_2)$, and hence for μ_1/μ_2 which is merely $\pi/(1 - \pi)$. The method is illustrated in Example 5.4.

The difference $\mu_1 - \mu_2$ is estimated by $x_1 - x_2$, and the usual normal theory can be applied as an approximation, with the standard error of $x_1 - x_2$ estimated as in (5.7) by $\sqrt{(x_1 + x_2)}$.

Example 5.4

Equal volumes of two bacterial cultures are spread on nutrient media and after incubation the numbers of colonies growing on the two plates are 13 and 31. We require confidence limits for the ratio of concentrations of the two cultures.

The estimated ratio is $13/31 = 0{\cdot}4194$. From the Geigy tables a binomial sample with 13 successes out of 44 provides the following 95% confidence limits for π: 0·1676 and 0·4520. Calculating $\pi/(1 - \pi)$ for each of these limits gives the following 95% confidence limits for μ_1/μ_2:

$$0{\cdot}1676/0{\cdot}8324 = 0{\cdot}2013$$

and

$$0{\cdot}4520/0{\cdot}5480 = 0{\cdot}8248.$$

The mid-P limits for π, calculated exactly as described in §4.4, are 0·1752 and 0·4418, leading to mid- P limits for μ_1/μ_2 of 0·2124 and 0·7915.

The normal approximations described in §4.4 can, of course, be used when the frequencies are not too small.

Example 5.5

Just as the distribution of a proportion, when n is large and π is small, is well approximated by assuming that the number of successes, r, follows a Poisson distribution, so a comparison of two proportions under these conditions can be effected by the methods of this section. Suppose, for example, that, in a group of 1000 men observed during a particular year, 20 incurred a certain disease, whereas, in a second group of 500 men, four cases occurred. Is there a significant difference between these proportions? This question could be answered by the methods of §4.5. As an approximation we could compare the observed proportion of deaths falling into group 2, $p = 4/24$, with the theoretical proportion $\pi = 500/1500 = 0{\cdot}3333$. The equivalent χ^2 test would run as follows:

	Group 1	Group 2	Total
Observed cases	20	4	24
Expected cases	$\frac{1000 \times 24}{1500} = 16$	$\frac{500 \times 24}{1500} = 8$	24

With continuity correction

$$\begin{aligned} X_c^2 &= (3\tfrac{1}{2})^2/16 + (3\tfrac{1}{2})^2/8 \\ &= 0{\cdot}766 + 1{\cdot}531 \\ &= 2{\cdot}30 \ (P = 0{\cdot}13). \end{aligned}$$

The difference is not significant. Without the continuity correction, $X^2 = 3{\cdot}00$ $(P = 0{\cdot}083)$.

If the full analysis for the 2×2 table is written out it will become clear that this abbreviated analysis differs from the full version in omitting the contributions to X^2 from the non-affected individuals. Since these are much more numerous than the cases, their contributions to X^2 have large denominators and are therefore negligible in comparison with the terms used above. This makes it clear that the short method described here must be used only when the proportions concerned are very small.

Example 5.6

Consider a slightly different version of Example 5.5. Suppose that the first set of 20 cases occurred during the follow-up of a large group of men for a total of 1000 man-years, whilst the second set of four cases occurred amongst another large group followed for 500 man-years. Different men may have different risks of disease, but, under the assumptions that each man has a constant risk during his period of observation and that the lengths of follow-up are unrelated to the individual risks, the number of cases in each group will approximately follow a Poisson distribution. As a test of the null hypothesis that the mean risks per unit time in the two groups are equal, the χ^2 test shown in Example 5.5 may be applied.

Note, though, that a significant difference may be due to failure of the assumptions. One possibility is that the risk varies with time, and that the observations for one group are concentrated more heavily at the times of high risk than is the case for the other group; an example would be the comparison of infant deaths, where one group might be observed for a shorter period after birth, when the risk is high. Another possibility is that lengths of follow-up are related to individual risk. Suppose, for example, that individuals with high risk were observed for longer periods than those with low risk; the effect would be to increase the expected number of cases in that group.

Further methods for analysing follow-up data are described in Chapter 17.

5.3 Ratios and other functions

We saw, in §4.2, that inferences about the population mean are conveniently made by using the standard error of the sample mean. In §§4.4 and 5.2, approximate methods for proportions and counts made use of the appropriate standard errors, invoking the normal approximations to the sampling distributions. Similar normal approximations are widely used in other situations, and it is therefore useful to obtain formulae for standard errors (or, equivalently, their