Table 8.4.1.1. *Values of χ^2/ν for which the c.d.f. $\Psi(\chi^2,\nu)$ has the values given in the column headings, for various values of ν*

ν	0.5	0.9	0.95	0.99	0.995
1	0.4549	2.7055	3.8415	6.6349	7.8795
2	0.6931	2.3026	2.9957	4.6052	5.2983
3	0.7887	2.0838	2.6049	3.7816	4.2794
4	0.8392	1.9449	2.3719	3.3192	3.7151
6	0.8914	1.7741	2.0986	2.8020	3.0913
8	0.9180	1.6702	1.9384	2.5113	2.7444
10	0.9342	1.5987	1.8307	2.3209	2.5188
15	0.9559	1.4871	1.6664	2.0385	2.1868
20	0.9669	1.4206	1.5705	1.8783	1.9999
25	0.9735	1.3753	1.5061	1.7726	1.8771
30	0.9779	1.3419	1.4591	1.6964	1.7891
40	0.9834	1.2951	1.3940	1.5923	1.6692
50	0.9867	1.2633	1.3501	1.5231	1.5898
60	0.9889	1.2400	1.3180	1.4730	1.5325
80	0.9917	1.2072	1.2735	1.4041	1.4540
100	0.9933	1.1850	1.2434	1.3581	1.4017
120	0.9945	1.1686	1.2214	1.3246	1.3638
140	0.9952	1.1559	1.2044	1.2989	1.3346
160	0.9958	1.1457	1.1907	1.2783	1.3114
200	0.9967	1.1301	1.1700	1.2472	1.2763

$$\Gamma(x) = \int\limits_0^\infty t^{x-1} \exp(-t)\, {\rm d}t. \qquad (8.4.1.13)$$

Although this function is continuous for all $x > 0$, its value is of interest in the context of this analysis only for x equal to positive, integral multiples of 1/2. It can be shown that $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$, and $\Gamma(x+1) = x\Gamma(x)$. It follows that, for a positive integer, n, $\Gamma(n) = (n-1)!$, and that $\Gamma(3/2) = \sqrt{\pi}/2$, $\Gamma(5/2) = 3\sqrt{\pi}/4$, *etc*. The *beta function* is defined by

$$B(x, y) = \int\limits_0^1 t^{x-1}(1-t)^{y-1}\, {\rm d}t. \qquad (8.4.1.14)$$

It can be shown (Prince, 1994) that $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$. Making the substitution $t = z_1^2/s^2$, (8.4.1.12) becomes

$$\begin{aligned}\Phi(s^2) &= \frac{1}{2\pi}\exp\left(-\frac{s^2}{2}\right)\int\limits_0^1 [t(1-t)]^{-1/2}\,{\rm d}t\\ &= \frac{1}{2\pi}\exp\left(-\frac{s^2}{2}\right)B(1/2, 1/2)\\ &= \tfrac{1}{2}\exp\left(-\frac{s^2}{2}\right), \qquad s^2 \geq 0. \qquad (8.4.1.15)\end{aligned}$$

By a similar procedure, it can be shown that, if χ^2 is the sum of ν terms, $z_1^2, z_2^2, \ldots, z_\nu^2$, where all are drawn independently from a population with the p.d.f. given in (8.4.1.10), χ^2 has the p.d.f.

$$\begin{aligned}\Phi(\chi^2, \nu) &= \frac{(\chi^2)^{\nu/2-1}}{2^{\nu/2}\Gamma(\nu/2)}\exp\left(-\frac{\chi^2}{2}\right), \qquad \chi^2 > 0,\\ \Phi(\chi^2, \nu) &= 0, \qquad \chi^2 \leq 0.\end{aligned} \qquad (8.4.1.16)$$

The parameter ν is known as the number of *degrees of freedom*, but this use of that term must not be confused with the conventional use in physics and chemistry. The p.d.f. in (8.4.1.16) is the *chi-squared distribution with ν degrees of freedom*. Table 8.4.1.1 gives the values of χ^2/ν for which the *cumulative distribution function* (*c.d.f.*) $\Psi(\chi^2, \nu)$ has various values for various choices of ν. This table is provided to enable verification of computer codes that may be used to generate more extensive tables. It was generated using a program included in the statistical library DATAPAC (Filliben, unpublished). Fortran code for this program appears in Prince (1994).

The quantity $(n-p)G$ is the sum of n terms that have mean value $(n-p)/n$. Because the process of determining the least-squares fit establishes p relations among them, however, only $(n-p)$ of the terms are independent. The number of degrees of freedom is therefore $\nu = (n-p)$, and, if the model is correct, and the terms have been properly weighted, $\chi^2 = (n-p)G^2$ has the chi-squared distribution with $(n-p)$ degrees of freedom. In crystallography, the number of degrees of freedom tends to be large, and the p.d.f. for G correspondingly sharp, so that even rather small deviations from $G^2 = 1$ should cause one or both of the hypotheses of a correct model and appropriate weights to be rejected. It is common practice to assume that the model is correct, and that the weights have correct *relative* values, that is that they have been assigned by $w_i = k/\sigma_i^2$, where k is a number different from, usually greater than, one. G is then taken to be an estimate of k, and all elements of $(\boldsymbol{A}^T\boldsymbol{WA})^{-1}$ (Section 8.1.2) are multiplied by G^2 to get an estimated variance–covariance matrix. The range of validity of this procedure is limited at best. It is discussed further in Chapter 8.5.

8.4.2. The *F* distribution

Consider an unconstrained model with p parameters and a constrained one with q parameters, where $q < p$. We wish to decide whether the constrained model represents an adequate fit to the data, or if the additional parameters in the unconstrained model provide, in some important sense, a better fit to the data. Provided the $(p-q)$ additional columns of the design matrix, $\boldsymbol{A}$, are linearly independent of the previous q columns, the sum of squared residuals must be reduced by some finite amount by adjusting the additional parameters, but we must decide whether this improved fit would have occurred purely by chance, or whether it represents additional information.

Let s_c^2 and s_u^2 be the weighted sums of squared residuals for the constrained and unconstrained models, respectively. If the constrained and unconstrained models are equally good representations of the data, and the weights have been assigned by $w_i = 1/\sigma_i^2$, the expected values of the sums of squares are $\langle s_c^2\rangle = (n-q)$ and $\langle s_u^2\rangle = (n-p)$, and, further, they should be distributed as χ^2 with $(n-q)$ and $(n-p)$ degrees of freedom, respectively. Also, $\langle s_c^2 - s_u^2\rangle = (p-q)$, and $(s_c^2 - s_u^2)$ is distributed as χ^2 with $(p-q)$ degrees of freedom. s_c^2 and s_u^2 are not independent, but $(s_c^2 - s_u^2)$ is the squared magnitude of a vector in a $(p-q)$-dimensional subspace that is orthogonal to the $(n-p)$-dimensional space of s_u^2. Therefore, s_u^2 and $(s_c^2 - s_u^2)$ are independent, random variables, each with a χ^2 distribution. Let $\chi_1^2 = (s_c^2 - s_u^2)$, $\chi_2^2 = s_u^2$, $\nu_1 = p - q$, and $\nu_2 = n - p$. The ratio $F = (\chi_1^2/\nu_1)/(\chi_2^2/\nu_2)$ should have a value close to one, even if the weights have relative rather than absolute values, but we need a measure of how far away from one this ratio can be before we must reject the hypothesis that the two models are equally good representations of the data. The conditional p.d.f. for F, given a value of χ_2^2 , is

$$\Phi_C(F|\chi_2^2) = \frac{[(\nu_1/\nu_2)\chi_2^2]^{\nu_1/2}F^{\nu_1/2-1}}{2^{\nu_1/2}\Gamma(\nu_1/2)}\exp[-(\nu_1/\nu_2)\chi_2^2F/2], \qquad (8.4.2.1)$$

and the marginal p.d.f. for χ_2^2 is

$$\Phi_M(\chi_2^2) = \frac{(\chi_2^2)^{\nu_2/2-1}}{2^{\nu_2/2}\Gamma(\nu_2/2)}\exp(-\chi_2^2/2). \qquad (8.4.2.2)$$