Table 8.4.2.1. *Values of the F ratio for which the c.d.f. $\Psi(F,\nu_1,\nu_2)$ has the value 0.95, for various choices of ν_1 and ν_2*

ν_2 \ ν_1	1	2	4	8	15
10	4.9646	4.1028	3.4781	3.0717	2.8450
20	4.3512	3.4928	2.8661	2.4471	2.2033
30	4.1709	3.3158	2.6896	2.2662	2.0148
40	4.0847	3.2317	2.6060	2.1802	1.9245
50	4.0343	3.1826	2.5572	2.1299	1.8714
60	4.0012	3.1504	2.5252	2.0970	1.8364
80	3.9604	3.1108	2.4859	2.0564	1.7932
100	3.9361	3.0873	2.4626	2.0323	1.7675
120	3.9201	3.0718	2.4472	2.0164	1.7505
150	3.9042	3.0564	2.4320	2.0006	1.7335
200	3.8884	3.0411	2.4168	1.9849	1.7167
300	3.8726	3.0259	2.4017	1.9693	1.6998
400	3.8648	3.0183	2.3943	1.9616	1.6914
600	3.8570	3.0107	2.3868	1.9538	1.6831
1000	3.8508	3.0047	2.3808	1.9477	1.6764

The marginal p.d.f. for F is obtained by integration of the joint p.d.f.,

$$\Phi(F) = \int_0^\infty \Phi_C\left(F|\chi_2^2\right)\Phi_M\left(\chi_2^2\right)\,\mathrm{d}\chi_2^2, \tag{8.4.2.3}$$

yielding the result

$$\Phi(F,\nu_1,\nu_2) = \frac{(\nu_1/\nu_2)F^{\nu_1/2-1}}{B(\nu_1/2,\nu_2/2)\left[1+(\nu_1/\nu_2)F\right]^{(\nu_1+\nu_2)/2}}. \tag{8.4.2.4}$$

This p.d.f. is known as the *F distribution with ν_1 and ν_2 degrees of freedom*. Table 8.4.2.1 gives the values of F for which the c.d.f. $\Psi(F,\nu_1,\nu_2)$ is equal to 0.95 for various choices of ν_1 and ν_2. Fortran code for the program from which the table was generated appears in Prince (1994).

The cumulative distribution function $\Psi(F,\nu_1,\nu_2)$ gives the probability that the F ratio will be less than some value by chance if the models are equally consistent with the data. It is therefore a necessary, but not sufficient, condition for concluding that the unconstrained model gives a significantly better fit to the data that $\Psi(F,\nu_1,\nu_2)$ be greater than $1-\alpha$, where α is the desired level of significance. For example, if $\Psi(F,\nu_1,\nu_2)=0.95$, the probability is only 0.05 that a value of F this large or greater would have been observed if the two models were equally good representations of the data.

Hamilton (1964) observed that the F ratio could be expressed in terms of the crystallographic weighted R index, which is defined, for refinement on |F| (and similarly for refinement on $|\mathrm{F}|^2$), by

$$R_w = \left[\sum w_i(|\mathrm{F}_o|_i - |\mathrm{F}_c|_i)^2 \Big/ \sum w_i|\mathrm{F}_o|_i^2\right]^{1/2}. \tag{8.4.2.5}$$

Denoting by R_c and R_u the weighted R indices for the constrained and unconstrained models, respectively,

$$F = (\nu_2/\nu_1)[(R_c/R_u)^2 - 1], \tag{8.4.2.6}$$

and a c.d.f. for R_c/R_u can be readily derived from this relation. A significance test based on R_c/R_u is known as *Hamilton's R-ratio test*; it is entirely equivalent to a test on the F ratio.

8.4.3. Comparison of different models

Tests based on F or the R ratio have several limitations. One important one is that they are applicable only when the parameters of one model form a subset of the parameters of the other. Also, the F test makes no distinction between improvement in fit as a result of small improvements throughout the entire data set and a large improvement in a small number of critically sensitive data points. A test that can be used for comparing arbitrary pairs of models, and that focuses attention on those data points that are most sensitive to differences in the models, was introduced by Williams & Kloot (1953; also Himmelblau, 1970; Prince, 1982).

Table 8.4.3.1. *Values of t for which the c.d.f. $\Psi(t,\nu)$ has the values given in the column headings, for various values of ν*

ν	0.75	0.90	0.95	0.99	0.995
1	1.0000	3.0777	6.3138	31.8206	63.6570
2	0.8165	1.8856	2.9200	6.9646	9.9249
3	0.7649	1.6377	2.3534	4.5407	5.8409
4	0.7407	1.5332	2.1319	3.7469	4.6041
6	0.7176	1.4398	1.9432	3.1427	3.7074
8	0.7064	1.3968	1.8596	2.8965	3.3554
10	0.6998	1.3722	1.8125	2.7638	3.1693
12	0.6955	1.3562	1.7823	2.6810	3.0546
14	0.6924	1.3450	1.7613	2.6245	2.9769
16	0.6901	1.3368	1.7459	2.5835	2.9208
20	0.6870	1.3253	1.7247	2.5280	2.8453
25	0.6844	1.3164	1.7081	2.4851	2.7874
30	0.6828	1.3104	1.6973	2.4573	2.7500
35	0.6816	1.3062	1.6896	2.4377	2.7238
40	0.6807	1.3031	1.6839	2.4233	2.7045
50	0.6794	1.2987	1.6759	2.4033	2.6778
60	0.6786	1.2958	1.6707	2.3901	2.6603
80	0.6776	1.2922	1.6641	2.3739	2.6387
100	0.6770	1.2901	1.6602	2.3642	2.6259
120	0.6765	1.2886	1.6577	2.3578	2.6174

Consider a set of observations, y_{0i}, and two models that predict values for these observations, y_{1i} and y_{2i}, respectively. We determine the slope of the regression line $z=\lambda x$, where $z_i = \left[y_{0i} - (1/2)(y_{1i}+y_{2i})\right]/\sigma_i$, and $x_i = (y_{1i}-y_{2i})/\sigma_i$. Suppose model 1 is a perfect fit to the data, which have been measured with great precision, so that $y_{0i}=y_{1i}$ for all i. Under these conditions, $\lambda=+1/2$. Similarly, if model 2 is a perfect fit, $\lambda=-1/2$. Real experimental data, of course, are subject to random error, and $|\lambda|$ in general would be expected to be less than 1/2. A least-squares estimate of λ is

$$\widehat{\lambda} = \frac{\sum_{i=1}^n z_i x_i}{\sum_{i=1}^n x_i^2}, \tag{8.4.3.1}$$

and it has an estimated variance

$$\widehat{\sigma}_{\lambda}^2 = \frac{\sum_{i=1}^n z_i^2 - \widehat{\lambda}^2\sum_{i=1}^n x_i^2}{(n-1)\sum_{i=1}^n x_i^2}. \tag{8.4.3.2}$$

The hypothesis that the two models give equally good fits to the data can be tested by considering $\widehat{\lambda}$ to be an unconstrained, one-parameter fit that is to be compared with a constrained, zero-parameter fit for which $\lambda=0$. A p.d.f. for making this comparison can be derived from an F distribution with $\nu_1=1$ and $\nu_2=\nu=(n-1)$.

$$\Phi(F,1,\nu) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\nu F}\,\Gamma(\nu/2)(1+F/\nu)^{(\nu+1)/2}}. \tag{8.4.3.3}$$

references