8.4. Statistical significance tests

By E. Prince and C. H. Spiegelman

In Chapter 8.1, we discussed the method of least squares and procedures for estimating the values of the adjustable parameters of a model that predicts the mean of a population from which experimental observations are drawn at random. Any model, however, will have some set of parameter values that gives the best least-squares fit. We must now address the question of whether that best fit is adequate, that is, whether it is plausible, given the precision of the data, to accept the hypothesis that the model really is a correct representation of the phenomena that have been measured in the collection of the data. In this chapter, we discuss the probability density function for the sum of squared residuals if the individual residuals are drawn from a normal distribution, the χ^2 distribution, and the conditions under which this p.d.f. may be assumed to approximate a practical case. Next, we discuss the F distribution, which is the distribution of the ratio of two independent, random variables, each of which has a χ^2 distribution, and its use in comparing the fits of constrained and unconstrained versions of a model. We also discuss a test that is useful for a more general comparison of models. Finally, we discuss the variation among data points of their effectiveness in improving the precision of parameter estimates and the application of this analysis to the optimum design of experiments.

8.4.1. The χ^2 distribution

We have seen [equation (8.1.2.1)] that the least-squares estimate is derived by finding the minimum value of a sum of terms of the form

$$R_i = w_i[y_i - M_i(\mathbf{x})]^2, \tag{8.4.1.1}$$

and, further, that the precision of the estimate is optimized if the weight, w_i, is the reciprocal of the variance of the population from which the observation is drawn, $w_i = 1/\sigma_i^2$. Using this relation, (8.4.1.1) can be written

$$R_i = \left\{\left[y_i - M_i(\mathbf{x})\right]/\sigma_i\right\}^2. \tag{8.4.1.2}$$

Each term is the square of a difference between observed and calculated values, expressed as a fraction of the standard uncertainty of the observed value. But, by definition,

$$\sigma_i^2 = \left\langle [y_i - M_i(\mathbf{x})]^2 \right\rangle, \tag{8.4.1.3}$$

where $\mathbf{x}$ has its unknown 'correct' value, so that $\langle R\rangle = 1$, and the expected value of the sum of n such terms is n. It can be shown (Draper & Smith, 1981) that each parameter estimated reduces this expected sum by one, so that, for p estimated parameters,

$$\langle S\rangle = \left\langle \sum_{i=1}^{n} \left\{\left[y_i - M_i(\widehat{\mathbf{x}})\right]/\sigma_i\right\}^2 \right\rangle = n - p, \tag{8.4.1.4}$$

where $\widehat{\mathbf{x}}$ is the least-squares estimate. The *standard uncertainty of an observation of unit weight*, also referred to as the *goodness-of-fit parameter*, is defined by

$$G = \left[\frac{S}{n-p}\right]^{1/2} = \left[\frac{\left\{\sum_{i=1}^{n} w_i\left[y_i - M_i(\widehat{\mathbf{x}})\right]^2\right\}}{n-p}\right]^{1/2}. \tag{8.4.1.5}$$

From (8.4.1.4), it follows that $\langle G\rangle = 1$ for a correct model with weights assigned in accordance with (8.4.1.2).

A value of G that is close to one, if the weights have been assigned by $w_i = 1/\sigma_i^2$, is an indicator that the model is consistent with the data. It should be noted that it is not necessarily an indicator that the model is 'correct', because it does not rule out the existence of an alternative model that fits the data as well or better. An assessment of the adequacy of the fit of a given model depends, however, on what is meant by 'close to one', which depends in turn on the spread of a probability density function for G. We saw in Chapter 8.1 that least squares with this weighting scheme would give the best, linear, unbiased estimate of the model parameters, with no restrictions on the p.d.f.s of the populations from which the observations are drawn except for the implicit assumption that the variances of these p.d.f.s are finite. To construct a p.d.f. for G, however, it is necessary to make an assumption about the shapes of the p.d.f.s for the observations. The usual assumption is that these p.d.f.s can be described by the *normal p.d.f.*,

$$\Phi_N(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{8.4.1.6}$$

The justification for this assumption comes from the *central-limit theorem*, which states that, under rather broad conditions, the p.d.f. of the arithmetic mean of n observations drawn from a population with mean μ and variance σ^2 tends, for large n, to a normal distribution with mean μ and variance σ^2/n. [For a discussion of the central limit theorem, see Cramér (1951).]

If we make the assumption of a normal distribution of errors and make the substitution $z = (x-\mu)/\sigma$, (8.4.1.6) becomes

$$\Phi_N(z, 0, 1) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right). \tag{8.4.1.7}$$

The probability that z^2 will be less than χ^2 is equal to the probability that z will lie in the interval $-\chi \le z \le \chi$, or

$$\Psi(\chi^2) = \int_0^{\chi^2} \Phi(z^2)\,\mathrm{d}z^2 = \int_{-\chi}^{+\chi} \Phi(z)\,\mathrm{d}z. \tag{8.4.1.8}$$

Letting $t = z^2/2$ and substituting in (8.4.1.7), this becomes

$$\Psi(\chi^2) = \frac{1}{\sqrt{\pi}}\int_0^{\chi^2/2} t^{-1/2}\exp(-t)\,\mathrm{d}t. \tag{8.4.1.9}$$

$\Phi(\chi^2) = \mathrm{d}\Psi(\chi^2)/\mathrm{d}\chi^2$, so that

$$\begin{aligned} \Phi\left(\chi^2\right) &= \left(2\pi\chi^2\right)^{-1/2}\exp\left(-\chi^2/2\right), && \chi^2 > 0, \\ \Phi\left(\chi^2\right) &= 0, && \chi^2 \le 0. \end{aligned} \tag{8.4.1.10}$$

The joint p.d.f. of the squares of two random variables, z_1 and z_2, drawn independently from the same population with a normal p.d.f. is

$$\Phi_J\left(z_1^2, z_2^2\right) = \frac{1}{2\pi z_1 z_2}\exp\left[-\frac{z_1^2 + z_2^2}{2}\right], \tag{8.4.1.11}$$

and the p.d.f. of the sum, s^2, of these two terms is the integral over the joint p.d.f. of all pairs of z_1^2 and z_2^2 that add up to s^2.

$$\Phi(s^2) = \frac{1}{2\pi}\exp\left(-\frac{s^2}{2}\right)\left[z_1^2(s^2 - z_1^2)\right]\mathrm{d}z_1^2. \tag{8.4.1.12}$$

This integral can be evaluated by use of the gamma and beta functions. The *gamma function* is defined for positive real x by

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

$$\Gamma(x) = \int_0^\infty t^{x-1}\exp(-t)\,\mathrm{d}t. \qquad (8.4.1.13)$$

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

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t. \qquad (8.4.1.14)$$

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

$$\begin{aligned}\Phi(s^2) &= \frac{1}{2\pi}\exp\left(-\frac{s^2}{2}\right)\int_0^1 [t(1-t)]^{-1/2}\,\mathrm{d}t \\ &= \frac{1}{2\pi}\exp\left(-\frac{s^2}{2}\right)B(1/2,1/2) \\ &= \tfrac{1}{2}\exp\left(-\frac{s^2}{2}\right), \qquad s^2 \geq 0. \end{aligned} \qquad (8.4.1.15)$$

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 $(\boldsymbol{A}^T\boldsymbol{W}\boldsymbol{A})^{-1}$ (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

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

and the marginal p.d.f. for χ_2^2 is

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

references