Regression theory may be approached in a number of ways. One of them, to derive it as the statistical behavour of one variable, conditional upon all the others, in a multivariate normal distribution, seems unnecessarily restricted. Another, as a direct application of the theory of least squares, is the least restricted and often the simplest to carry out, but one naturally prefers an approach that depends on more primitive principles of estimation.

Two approaches will be used here, first, the least squares derivation, second, one requiring that the estimators of the regression coefficients be best, linear, unbiased. The equivalence of the two derivations establishes the Gauss-Markoff theorem.

Notation

$x_0, x_1, \ldots, x_p$ will denote the independent variables of the system, called independent variables, fixed variables, predictors, selectors.

y will denote the dependent or statistical variable. Observations on these variables will be called

$$y_\alpha, x_{0\alpha}, x_{1\alpha}, \ldots, x_{p\alpha}, \quad \alpha = 1, 2, \ldots, N$$

The regression question

Any question of the sort : given values of the x's, what can be said about the statistical behavior of y? might be said to be a regression question. The essential feature, of course, being that it is a conditional question. The x's, here, are numerical, but have no other restrictions. They may be continuous or discontinuous, and need not come from, or be associated with, frequency distributions. They serve simply as selectors which, once assigned, select a y - population,

which will then be sampled one or more times.

The dependent variable, y, will be treated here as continuous, but the usual devices used to bring discontinuous variables within the scope of these procedures are available here also.

The usual restricted regression question

The question usually posed concerns the conditional mean of $(y|x_0, x_1, \ldots, x_p)$ in the form

$$E(y|x_0, x_1, \ldots, x_p) = f(x_0, x_1, \ldots, x_p),$$

where the function f is specified up to an set of parameters, whose values are to be estimated from the observations. The only form of f to be discussed here is one linear in the parameters.

$$E(y|x_0, x_1, \ldots, x_p) = \beta_0 x_0 + \beta_1 x_1 + \ldots + \beta_p x_p = \sum_{i=0}^{p} \beta_i x_i$$

This assumption about the form of f imposes severe restrictions, to be sure, but they are not as severe as may appear at first glance. Nothing stands in the way of the x's being functions of one another (apart from linear functions). We could, for example, choose $x_0 = 1$, $x_3 = x_1^2$, $x_4 = x_2^2$, $x_5 = x_1 x_2$, etc., in which case we are fitting a curved regression surface.

In most regression questions, it is appropriate to put $x_0 = 1$, i.e. to provide a constant term in the regression model. Most of the results that we will be reaching do not require this supposition. Those that do will be pointed out.

Another assumption will be made about the nature of the y - distributions.

$$\mathrm{Var}(y|x_0, x_1, \ldots, x_p) = \sigma^2,$$

i.e. the variance of y does not vary from one population to another. Practically speaking, the y - distributions are identical, apart from

location. This is a strong assumption, never to be taken lightly.

Since all our distributions are conditional, we may, to save writing, use Ey instead of $E(y|x_0, x_1, \ldots, x_p)$ and Var y instead of $\mathrm{Var}(y|x_0, x_1, \ldots, x_p)$ and so on.

Assumptions about the sample.

We envisage a sample obtained in the following way: a set of x-values is chosen, in any way we deem suitable, $x_{0\alpha}, x_{1\alpha}, \ldots, x_{p\alpha}$. These x-values select a y-population, which we proceed to sample randomly, obtaining y_α.

Sometimes, this is the most natural and convenient way of carrying out the sampling. On occasion, though, the x's and y all pertain to the same individual, and it is easier to select a sample of individuals and measure the x's and y on each of them. In this case, the x's as well as y may be considered to come randomly from frequency distributions, but this is irrelevant to the regression question we have proposed.

The assumptions made this far permit the following statements

$$E\,y = \sum \beta_i x_i\,, \qquad E\,y_\alpha = \sum \beta_i\, x_{i\alpha}\,,$$

$y_\alpha = \sum \beta_i\, x_{i\alpha} + \varepsilon_\alpha$, where ε_α represents an "error", coming randomly from some distribution of errors with

$$E\,\varepsilon_\alpha = 0$$

$$E\,\varepsilon_\alpha \varepsilon_\beta = \sigma^2\,\delta_{\alpha\beta}, \qquad \delta_{\alpha\beta} = 1, \quad \alpha = \beta$$

$$= 0, \quad \alpha \neq \beta.$$

Estimation of the regression coefficient and $E(y)$.

We shall write b_i as an estimator of β_i and Y as an estimator of $E(y)$. Then,

$$Y = \sum b_i x_i \text{ and } Y_\alpha = \sum b_i x_{i\alpha}.$$

The principle of least squares.

This principle asserts that the estimation be carried out by choosing the b_i so as to minimize the sum of squares of residuals, i.e.

$$E = S(y_\alpha - Y_\alpha)^2 \text{ is to be minimized.}$$

Thus, for each b_j, we write

$$\frac{\partial E}{\partial b_j} = -2\, S(y_\alpha - Y_\alpha)\, \frac{\partial Y_\alpha}{\partial b_j} = -2\, S(y_\alpha - Y_\alpha)\, x_{j\alpha} = 0.$$

We get, then, a set of $p + 1$ equations,

$$S(y_\alpha - Y_\alpha)\, x_{j\alpha} = 0, \quad j = 0, 1, \ldots, p.$$

These are the so-called normal equations. Writing them explicitly in terms of the b's,

$$S\, Y_\alpha\, x_{j\alpha} = S\, y_\alpha\, x_{j\alpha} \qquad \text{or}$$

$$S \sum_i b_i\, x_{i\alpha}\, x_{j\alpha} = S\, y_\alpha\, x_{j\alpha} \quad \text{or}$$

$$\sum_i b_i\, S\, x_{i\alpha}\, x_{j\alpha} = S\, y_\alpha\, x_{j\alpha}, \; j = 0, 1, \ldots, p.$$

The numbers calculated from the sample, $S\, x_{i\alpha}\, x_{j\alpha}$ and $S\, y_\alpha\, x_{j\alpha}$, will be given the symbols $a_{ji} \equiv a_{ij}$ and g_j.

Hence, we have the set of normal equations

$$\sum_{j=0}^{p} a_{ij}\, b_j = g_i, \quad i = 0, 1, \ldots, p.$$

It will be convenient to exhibit these equations in solved forms by introducing the elements of the matrix (c_{ij}), inverse to (a_{ij}), i.e. the solutions of the equations

$$\sum_{j=0}^{p} a_{ij}\, c_{jk} = \delta_{ik} \quad i,k = 0, 1, \ldots, p.$$

Then the solutions of the normal equations are

$$b_i = \sum c_{ij}\, g_j, \quad i = 0, 1, \ldots, p.$$

Best Linear Unbiased Estimates.

Let us seek values of the b's which have the following properties.

1. b_i is to be a linear function of the y's,

$$b_i = \sum_{\alpha=1}^{N} w_{i\alpha}\, y_\alpha = S\, w_{i\alpha}\, y_\alpha.$$

where the w's will, presumably, be functions of the x's

2. $E\, b_i = \beta_i$

3. Var b_i to be as small as possible.

To set up this question as a purely mathematical exercise, we note:

$$\beta_i = E\, b_i = S\, w_{i\alpha}\, E\, y_\alpha = S\, w_{i\alpha} \sum_j \beta_j\, x_{j\alpha}$$

$$= \sum_j \beta_j\, S\, w_{i\alpha}\, x_{j\alpha}, \text{ identically in the } \beta\text{'s.}$$

Hence equating coefficients of the β's,

$$S\, w_{i\alpha}\, x_{j\alpha} = \delta_{ij}, \quad i, j = 0, 1, \ldots, p,$$

also, $\text{Var}\, b_i = S\, w_{i\alpha}^2\, \text{Var}\, y_\alpha = \sigma^2\, S\, w_{i\alpha}^2$.

Hence, we wish, for each i, to minimize $S\, w_{i\alpha}^2$, subject to the restraints $S\, w_{i\alpha}\, x_{j\alpha} = \delta_{ij}$. We therefore introduce, for each i, $p + 1$ Lagrange multipliers λ_{ij} and seek the unrestricted minimum of

$$\psi_i = \frac{1}{2}\, S\, w_{i\alpha}^2 - \sum_j \lambda_{ij}\, (S\, w_{i\alpha}\, x_{j\alpha} - \delta_{ij}).$$

The rest is simply formal mathematics.

$$(1) \quad \frac{\partial \psi_i}{\partial w_{i\alpha}} = w_{i\alpha} - \sum_j \lambda_{ij}\, x_{j\alpha} = 0, \quad \alpha = 1, 2 \ldots, N.$$

$$(2) \quad \frac{\partial \psi_i}{\partial \lambda_{ij}} = S\, w_{i\alpha} - \delta_{ij} = 0, \quad j = 0, 1, \ldots, p.$$

These $N + p + 1$ equations are to be solved for the N $w_{i\alpha}$'s and the $p + 1$ λ_{ij}'s. We may proceed as follows. Multiply equations (1) by $x_{k\alpha}$ and sum.

$$S\, w_{i\alpha}\, x_{k\alpha} - \sum_j \lambda_{ij}\, S\, x_{j\alpha}\, x_{k\alpha} = 0,$$

whence, in virtue of (2) and noting that $S\, x_{j\alpha}\, x_{k\alpha} = a_{jk}$,

$$\sum \lambda_{ij}\, a_{jk} = \delta_{ik} \qquad i, k = 0, 1, \ldots, \ .$$

Thus the λ's are, in fact, the elements of the inverse matrix to (a_{ij}), i.e. $\lambda_{ij} = c_{ij}$.

Then

$$b_i = S\, w_{i\alpha}\, y_\alpha = S \sum_j c_{ij}\, x_{j\alpha}\, y_\alpha = \sum_j c_{ij}\, S\, x_{j\alpha}\, y_\alpha$$

$$= \sum_j c_{ij}\, g_j \ ,$$

the solutions obtained before by minimizing $S(y_\alpha - Y_\alpha)^2$. The least squares solution thus provides BLUE estimates of the β's. This is the Gauss-Markoff theorem.

Variances and covariances of the b's.

$\mathrm{Cov}(b_i, b_j) = E(b_i - \beta_i)(b_j - \beta_j)$. If, for the moment, we write $\delta(\ )$ for $(\ ) - E(\ )$,

$$\mathrm{Cov}\ b_i\, b_j = E\ \delta b_i\ \delta b_j \ .$$

Starting from $b_i = \sum_k c_{ik}\, g_k$,

$$\delta\ b_i = \sum c_{ik}\ \delta\ g_k \quad \text{and}$$

$$\delta\ b_i\ \delta b_j = \sum_k \sum_l c_{ik}\, c_{jl}\ \delta g_k\ \delta g_l$$

Now, $g_k = S\, x_{k\alpha}\, y_\alpha$, whence

$$\delta g_k = S\, x_{k\alpha}\, S\, y_\alpha = S\, x_{k\alpha}\, \varepsilon_\alpha \quad \text{and}$$

$$\delta g_k\, \delta g_l = \underset{\alpha}{S}\,\underset{\beta}{S}\, x_{k\alpha}\, x_{l\beta}\, \varepsilon_\alpha\, \varepsilon_\beta \,. \quad \text{Thus}$$

$$E\, \delta g_k\, \delta g_l = \underset{\alpha}{S}\,\underset{\beta}{S}\, x_{k\alpha}\, x_{l\beta}\, \sigma^2\, \delta_{\alpha\beta}$$

$$= \sigma^2\, \underset{\alpha}{S}\, x_{k\alpha}\, x_{l\alpha}$$

$$= \sigma^2\, a_{kl} \,. \quad \text{Hence}$$

$$E\, \delta b_i\, \delta b_j = \sigma^2\, \underset{k}{\Sigma}\,\underset{l}{\Sigma}\, c_{ik}\, c_{jl}\, a_{kl} = \sigma^2\, \underset{l}{\Sigma}\, c_{jl}\, \delta_{il} = \sigma^2\, c_{ji} \,.$$

Thus, it emerges that the values of the elements of the inverse matrix will be needed to specify the variances and covariances of the b's, even though we might not want them for calculating the b's. (I think there are better ways to get the b's.)

Estimation of the error variance.

It seems obvious that the residual sum of squares, $S(y_\alpha - Y_\alpha)^2$, will reflect error only if the assumed functional form of the regression is correct. This will be checked later. In any event, it seems likely that we will require its value. Computation from the definition is tedious, unless one is using a high-speed calculator. An alternative is provided by the following identity.

$$S(y_\alpha - Y_\alpha)^2 = S(y_\alpha - Y_\alpha)\, y_\alpha - S(y_\alpha - Y_\alpha)\, Y_\alpha$$

$$= S\, y_\alpha^2 - S\, y_\alpha\, Y_\alpha - S(y_\alpha - Y_\alpha)\, \Sigma b_i\, x_{i\alpha}$$

$$= S\, y_\alpha^2 - S\, y_\alpha\, \Sigma\, b_i\, x_{i\alpha} - 0 \quad \text{(normal equations)}$$

$$= S\, y_\alpha^2 - \sum_{i=0}^{p} b_i\, g_i \,.$$

Another identity, for the cases where $X_0 = 1$, may be checked in the same way,

$$S(y_\alpha - \bar{y})^2 = S(y_\alpha - Y_\alpha)^2 + S(Y_\alpha - \bar{y})^2.$$

Tests of significance.

If we are to develop exact tests of significance, we need the assumption, not hitherto required, that the errors are $N(0, \sigma^2)$. It follows, from this assumption, that b_i is $N(\beta_i, \sigma^2 c_{ii})$ and, indeed, that all the b's are distributed in a multivariate normal distribution, with covariance matrix $\sigma^2(c_{ij})$. Hence the quadratic form $\sum_{i,j=0}^{p} a_{ij} \delta b_i \delta b_j$ is a $\sigma^2 X^2_{(p+1)}$, distributed independently of $S(y_\alpha - Y_\alpha)^2$, which is $\sigma^2 X^2_{(N-p-1)}$. Furthermore, any subset of the b's, say, $b_{q+1}, \ldots, b_p$, will also be multivariate normal with the same covariance $\sigma^2 c_{ij}, i, j = q+1, \ldots, p$ and the quadratic form $\sum_{i,j=q+1}^{p} a^*_{ij} \delta b_i \delta b_j$ is $\sigma^2 X^2_{(p-q)}$, where $(a^*_{ij}) = (c_{ij})^{-1}$ $i, j = q+1, \ldots, p$

Any tests of significance we may require follow from these statements. The truth of such of these results as we need will be checked out as we go along.

The "straight" regression line.

Putting $x_0 = 1$, $x_1 = x$, $p = 1$, we discuss the fitting of a regression $Y = b_0 + b_1 x$ to the sample (x_α, y_α), $\alpha = 1, \ldots, N$. This yields some useful formulae and some hints for the development of distribution theory and other things.

The normal equations are

$$N b_0 + b_1 Sx = Sy$$
$$b_0 Sx + b_1 Sx^2 = Sxy.$$

These equations may be solved algebraically to provide formulae, but instead, observe that the first of the equations asserts that $\bar{y} = b_0 + b_1 \bar{x}$, i.e., the average point lies on the regression line. Hence we might, with profit, think of fitting the regression in the form $Y = b_0' + b_1(x-\bar{x})$, where $b_0 = b_0' - b_1\bar{x}$. The normal equations for this fitting are

$$N b_0' + b_1 S(x-\bar{x}) = S\, y,$$
$$b_0' S(x-\bar{x}) + b_1 S(x-\bar{x})^2 = S\, y(x-\bar{x}).$$

Since $S(x-\bar{x}) = 0$, the equations reduce to

$$N b_0' = S\, y \qquad\qquad b_0' = \bar{y}$$

$$b_1 S(x-\bar{x})^2 = S\, y(x-\bar{x}) \qquad\qquad b_1 = \frac{S\, y(x-\bar{x})}{S(x-\bar{x})^2}$$

We can read off, too: $c_{00} = \frac{1}{N}$, $c_{11} = \frac{1}{S(x-\bar{x})^2}$, $c_{01} = 0$.

The s.s. residuals $= Sy^2 - \bar{y}Sy - b_1 Sy(x-\bar{x})$

$$= S(y-\bar{y})^2 - \frac{[Sy(x-\bar{x})]^2}{S(x-\bar{x})^2}$$

Furthermore, since $Y_\alpha - \bar{y} = b_1(x_\alpha - \bar{x})$, we see that

$$S(Y_\alpha - \bar{y})^2 = b_1^2 \, S(x_\alpha - \bar{x})^2 = \frac{[S\, y_\alpha (x_\alpha - \bar{x})]^2}{S(x_\alpha - \bar{x})^2} .$$

It is convenient and customary to assemble these computations in an analysis of variance table.

	d.f.	s.s.
attributable to regression	1	$S(Y_\alpha - \bar{y})^2$
deviations from regression	$N-2$	$S(y_\alpha - Y_\alpha)^2$
total	$N-1$	$S(y_\alpha - \bar{y})^2$

If the d.f. are not obvious, they may be made so by embedding the whole computation in an orthogonal transformation.

	y_1	y_α	y_N		
z_1	1	1	1	$\sqrt{N}$	$z_1 = \sqrt{N}\, b_0'$
z_2	$x_1 - \bar{x}$	$x_\alpha - \bar{x}$	$x_N - \bar{x}$	$\sqrt{S(x_\alpha - \bar{x})^2}$	$z_2 = \sqrt{S(x_\alpha - \bar{x})^2}\, b_1$
z_3					
.					
.	orthogonal				
.					
z_N	a_1	a_α	a_N		

If any z_i, $i > 2$ is orthogonal with z_1 and z_2, say $z_i = \Sigma\, a_\alpha\, y_\alpha$, we must have $\Sigma\, a_\alpha = 0$, $\Sigma\, a_\alpha\, (x_\alpha - \bar{x}) = 0$, whence $\Sigma\, a_\alpha\, x_\alpha = 0$. Then $z_i = \Sigma\, a_\alpha\, y_\alpha = \Sigma\, a_\alpha\, (\beta_0 + \beta_1\, x_\alpha + \varepsilon_\alpha) = \Sigma\, a_\alpha\, \varepsilon_\alpha$, i.e., depends only on the errors therefore $E\, z_i = 0$.

Therefore, $z_3, z_4, \ldots, z_N$ are independent, normal variables, each $N(0, \sigma^2)$, hence $z_3^2 + z_4^2 + \ldots + z_N^2 = \sigma^2 X^2_{(N-2)}$ and is independent of z_1 and z_2. Evidently, $z_2^2 = S(Y_\alpha - \bar{y})^2$, $z_3^2 + z_4^2 + \ldots + z_N^2 = S(y_\alpha - Y_\alpha)^2$. It should be clear, too, that $\frac{S(y_\alpha - Y_\alpha)^2}{N-2} = s^2$ (say) estimates σ^2.

Thus $\frac{b_1 - \beta_1}{s\sqrt{c_{11}}}$ is $t_{(N-2)}$ and can be used to test any hypothesized β_1. $\beta_1 = 0$ may equivalently be tested by $\frac{S(Y_\alpha - \bar{y})^2/1}{S(y_\alpha - Y_\alpha)^2/N-2}$, which is $F_{(1,N-2)}$ when $\beta_1 = 0$.

Testing the assumption $E\, y = \beta_0 + \beta_1 x$.

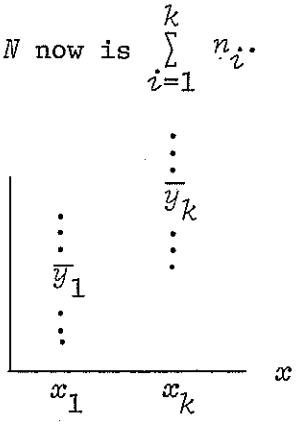
Clearly, everything we have done depends entirely on the correctness of the assumed functional form of the regression function. In particular, the sum of squares of residuals reflects, not only error, but also any systematic departures of the assumed function from the correct one. Here we see a way of checking on the question of correctness of the assumed function, provided we make provision in our sampling for an estimate of error that does not depend on the correctness of the fitted function. If, in our sampling, having chosen an x-value, we sample the selected y-population not once, but several times, differences among these y - observations will reflect error without any assumption about the regression to be fitted.

The formulae for calculating the regression require no changes. No assumption was made in developing them that the x's need be all different. However, it may be useful to change the notation to recognize that each x population is sampled several times. Let the different x's be labelled $x_1, \ldots, x_i, \ldots, x_k$ and the y - observations

corresponding to x_i be

$y_{i1}, y_{i2}, \cdots y_{in_i}$. N now is $\sum_{i=1}^{k} n_i$.

x	y	T	$\bar{y}$
x_1	$y_{11}, \cdots, y_{1n_1}$	T_1	$\bar{y}_1$
x_i	$y_{i1}, \cdots, y_{in_i}$	T_i	$\bar{y}_i$
x_k	$y_{k1}, \cdots, y_{kn_k}$	T_k	$\bar{y}_k$



The normal equations, written in this notation, are

$$N\, b_0 + b_1 \Sigma\, n_i x_i = Sy = \Sigma T_i = \Sigma n_i \bar{y}_i$$

$$b_0\, \Sigma n_i x_i + b_1 \Sigma n_i x_i^2 = Sxy = \Sigma x_i T_i = \Sigma n_i x_i \bar{y}_i$$

From those equations, we see that a regression fitted to the points $(x_i, \bar{y}_i)$, taken n_i times is the same as the regression fitted to the original observations. The s.s. residuals is, of course, different, being $\Sigma\, n_i\, \bar{y}_i^2 - b_0\, g_0 - b_1\, g_1$ instead of $Sy^2 - b_0\, g_0 - b_1\, g_1$. The difference between the two s.s. is $Sy^2 - \Sigma n_i \bar{y}_i^2$, that is, the within samples s.s.

We are, in fact, in the pattern called a completely randomized experiment, with a preliminary analysis of variance that reads:

	d.f.	s.s.		
among samples (x's)	k-1	$\frac{\Sigma\, T_i^2}{n_i} - \frac{G^2}{N}$	regression on x	1
			residuals	k-2
within samples (error)	N-k	by subtraction		
Total	N-1	$Sy^2 - \frac{G^2}{N}$		

The (linear) regression on x has s.s. $\frac{[\Sigma\ T_i(x_i-\bar{x})]^2}{\Sigma\ n_i(x_i-\bar{x})^2}$ and the residuals may be obtained by subtraction. The residuals s.s. may be tested against error, because if it contains nothing but error, the ratio $\frac{(\text{s.s. residuals})\ (k-2)}{(\text{s.s. error})/(N-k)}$ is $F(k-2,\ N-k)$. Of course, if N is at all large, this could be a very weak test and we might want to enquire more closely into the residuals.

An orthogonal transformation, corresponding to this partitioning of the total s.s., follows

	y_{11}	y_{12}	y_{1n_1}	$\cdots$	y_{i1}	y_{i2}	y_{in_i}	$\cdots$	y_{k1}	y_{k2}	y_{kn_k}	divisor
z_1	1	1	1		1	1	1		1	1	1	$\sqrt{N}$
z_2	$x_1-\bar{x}$	$x_1-\bar{x}$	$x_1-\bar{x}$		$x_i-\bar{x}$	$x_i-\bar{x}$	$x_i-\bar{x}$		$x_k-\bar{x}$	$x_k-\bar{x}$	$x_k-\bar{x}$	$\sqrt{\Sigma\ n_i(x_i-\bar{x})^2}$
z_3												
.												
.					orthogonal							
.												
z_k	a_1	a_1	a_1		a_i	a_i	a_i		a_k	a_k	a_k	
z_{k+1}	α_1	α_2	α_{n_1}		o	o	o		o	o	o	
.												
.			orthogonal									
.												
z_{k+n_1-1}	β_1	β_2	β_{n_1}		o	o	o		o	o	o	
z_N	o	o	o		o	o	o		γ_1	γ_2	γ_{n_k}	

$$\Sigma\ a_i = \Sigma\ \alpha_\alpha = \Sigma\ \beta_\alpha \cdots = \Sigma\ \gamma_\alpha = o\ .$$

$$z_1 = \sqrt{N}\ \bar{y}$$

$$z_2 = \frac{\Sigma\ T_i(x_i-\bar{x})}{\sqrt{\Sigma\ n_i(x_i-\bar{x})^2}} = \sqrt{\Sigma\ n_i(x_i-\bar{x})^2}\ \ b_1$$

$z_3 \ldots z_k$ are deviations of the sample averages from the fitted regression. $z_{k+1} \ldots z_n$ are within-sample deviations. The analysis of variance table is

	d.f.		s.s.
among samples	$k-1$	1 attr. to x	z_2^2
		$k-2$ residuals	$z_3^2 + \ldots + z_k^2$
within samples	$N-k$		$z_{k+1}^2 + \ldots + z_N^2$.

Distribution of the sum of squares of residuals Transformation of the independent variables Orthogonal Functions.

The device used to derive the distribution of the s.s. residuals about the fitting $Y = b_0 + b_1 x$ will now be extended to the general case.

Suppose we envisage a regression $Y = \sum_{i=0}^{p} b_i x_i$ and define a set of linear functions of x,

$$P_i = P_i (x_0, x_1, \ldots, x_p), \qquad i = 0, 1, \ldots, p.$$

Then, the regression equation may be re-written in the form $Y = \Sigma B_i P_i$, where the B's are linear functions of the b's.

If we think of fitting this regression directly to the data, we would have a set of normal equations $\Sigma A_{ij} B_j = G_i$, where

$$A_{ij} = \sum_{\alpha=1}^{N} P_{i\alpha} P_{j\alpha} = S P_{i\alpha} P_{j\alpha}, \quad G_i = S P_{i\alpha} y_\alpha,$$

where $P_{i\alpha} = P_i (x_{0\alpha}, x_{1\alpha}, \ldots, x_{p\alpha})$.

A rather special set of linear functions will be used, not because it is essential to this argument (they are sufficient for our purposes, though), but because they will be useful later.

Define $P_0 = P_0\ (x_0) = \lambda_{00}\ x_0$

$P_1 = P_1\ (x_0, x_1) = \lambda_{10}\ x_0 + \lambda_{11}\ x_1$

$\vdots$

$P_i = P_i\ (x_0, x_1, \ldots, x_i) = \lambda_{i0}\ x_0 + \lambda_{i1}\ x_1 + \ldots + \lambda_{ii}\ x_i$.

The λ's are constants, to be chosen according to our needs.

Then, $Y = b_0\ x_0 + b_1\ x_1 + \ldots + b_{p-1}\ x_{p-1} + b_p\ x_p$

$= B_0\ P_0 + B_1\ P_1 + \ldots + B_{p-1}\ P_{p-1} + B_p\ P_p$

$= B_0\ (\lambda_{00}\ x_0) + B_1\ (\lambda_{10}\ x_0 + \lambda_{11}\ x_1) +$

$B_{p-1}(\lambda_{p-1,0}\ x_0 + \ldots + \lambda_{p-1,p-1}\ x_{p-1}) + B_p(\lambda_{p0}\ x_0 + \ldots + \lambda_{pp}\ x_p)$.

Equating coefficients of the x's,

$b_p = \lambda_{pp}\ B_p$

$b_{p-1} = \lambda_{p-1,p-1}\ B_{p-1} + \lambda_{p,p-1}\ B_p$

$b_i = \lambda_{ii}\ B_i + \lambda_{i+1,i}\ B_{i+1} + \ldots + \lambda_{pi}\ B_p$

$b_0 = \lambda_{00}\ B_0 + \lambda_{10}\ B_1 + \ldots + \lambda_{p0}\ B_p$.

Now, let us think of choosing the λ's so that

$$A_{ij} = SP_{i\alpha}\ P_{j\alpha} = 0,\ i \neq j\ .$$

With such choice of λ's, the functions P will be said to be orthogonal over the set of observations on the x's.

It is easy to check that such a choice of λ's can be made and to see how to determine them. For example,

$S\ P_0\ P_1 = S\ \lambda_{00}\ x_{0\alpha}\ (\lambda_{10}\ x_{0\alpha} + \lambda_{11}\ x_{1\alpha})$

$= \lambda_{00}\ [\lambda_{10}\ S\ x_{0\alpha}^2 + \lambda_{11}\ S\ x_{0\alpha}\ x_{1\alpha}]$

$= \lambda_{00}\ [a_{00}\ \lambda_{10} + a_{01}\ \lambda_{11}] = 0$.

Indeed, we have a free choice of one λ - value in each function and the rest can be obtained by solving a set of equations much like the normal equations. Our concern here, though, is simply with the existence of these orthogonal linear functions.

The normal equations for fitting the regression to these functions are

$$A_{ii}B_i = G_i,\ A_{ii} = S\ P_{i\alpha}^2,\ G_i = S\ P_{i\alpha}\ y_\alpha\ .$$

The fitting can be embedded in an orthogonal transformation as follows.

	y_1	y_α	y_N	divisor	
z_1	P_{01}	$P_{0\alpha}$	P_{0N}	$\sqrt{S\ P_{0\alpha}^2}$	$z_1 = \sqrt{S\ P_{0\alpha}^2}B_0$
z_2	P_{11}	$P_{1\alpha}$	P_{1N}	$\sqrt{S\ P_{1\alpha}^2}$	$z_2 = \sqrt{S\ P_{1\alpha}^2}B_1$
z_{p+1}	P_{p1}	$P_{p\alpha}$	P_{pN}	$\sqrt{S\ P_{p\alpha}^2}$	$z_{p+2} = \sqrt{S\ P_{p\alpha}^2}B_p$
z_{p+2}	a_1	a_α	a_N		
.					
.		orthogonal.			
.					
z_N					

Now, it is easy to check that z_{p+2}, ..., z_N reflect error only, hence each has expectation zero.

Any component, $\Sigma\ a_\alpha\ y_\alpha$, orthogonal to z_1, z_2, ..., z_p, must satisfy the conditions $\Sigma\ a_\alpha\ P_{i\alpha} = 0$, $i = 0, 1, \ldots, p$. Hence

$\Sigma\ a_\alpha\ x_{i\alpha} = 0$, $i = 0, 1, \ldots, p$.

Then, $\sum\limits_\alpha a_\alpha\ y_\alpha = \sum\limits_\alpha a_\alpha \left(\sum\limits_i \beta_i\ x_{i\alpha} + \epsilon_\alpha\right) = \sum a_\alpha\ \epsilon_\alpha$.

Now, from the properties of the transformation, we read the following facts, granted the normality of the error system.

1. Each of the B's is normally distributed independently of one another.
2. The sum of squares $z_{p+2}^2 + \ldots + z_N^2$ is $\sigma^2 \chi^2_{(N-p-1)}$ and is independent of the B's.

The sum of squares $z_{p+2}^2 + \ldots + z_N^2 = S\, y^2 - z_1^2 - z_2^2 - \cdots - z_{p+1}^2$

$$= S\, y^2 - \sum_{i=0}^{p} B_i^2\, S\, P_{i\alpha}^2 = S\, y^2 - \sum_{i=0}^{p} B_i\, G_i$$

$= S\,(y_\alpha - Y_\alpha)^2$ by our general rule for s.s. residuals.

It follows that $s^2 = \dfrac{S(y_\alpha - Y_\alpha)^2}{N-p-1}$ estimates σ^2, that $\dfrac{b_i - \beta_i}{s\sqrt{c_{ii}}} = t_{(N-p-1)}$ and so on.

The determination of orthogonal functions to facilitate a fitting is not usually employed in practice, because more arithmetic is involved than in solving the original normal equations. However, some of the virtues of such an approach should be recognized. Of course, the normal equations solve themselves, but, more important, the decision to exclude some sets of the x's, such as x_p, x_p and x_{p-1}, x_p, x_{p-1}, ..., x_{p-q} amounts to excluding P_p, P_p and P_{p-1}, P_p, P_{p-1}, ..., P_{p-q} (in virtue of the triangular form to this transformation), and the remaining coefficients B_0, B_1, B_2 are unchanged.

We may, in passing, observe two things, the effect of including an x variable that is not needed (its $\beta = 0$) and the effect of leaving out an x - variable that <u>is</u> needed (its $\beta \neq 0$).

It will be sufficient to ask these questions about x_p. We have

$E\, b_i = \lambda_{ii}\, E\, B_i + \lambda_{i+1,i}\, E\, B_{i+1} + \ldots + \lambda_{pi}\, E\, B_p$ and

$E\, b_p = \lambda_{pp}\, E\, B_p$.

Now, if $E\, b_p = 0$, then $E\, B_p = 0$ and if x_p is, in fact, included in the regression, it simply adds a zero term to $E\, b_i$ and b_i is still on unbiased estimator of β_i.

On the other hand, if $\beta_p \neq 0$, then $E\, B_p \neq 0$, and if x_p is omitted from the regression, a non-zero term is omitted from $E\, b_i$, (the other B's are the same, whether x_p is included or not) and b_i is now seen to be biased.

While these facts seem to be obvious, they are sometimes presented as a theorem and indeed without them the utility of regression theory would be seriously diminished.

The fitting of polynomial regression. Orthogonal polynomials.

Regression theory includes, as a particular case, the fitting of polynomial functions in one or more independent variables. The normal equations have coefficients which can be very large indeed, inasmuch as they are sums of powers of the recorded x's. In the case of one independent variable, we may put $x_0 = 1$, $x_1 = x$, $x_2 = x^2$, $x_3 = x^3$, and so on. Then $a_{ij} = S\, x^{i+j}$, $g_i = Sx^i\, y$.

If the degree of the polynomial is at all large, the normal equations can be very nasty indeed. Furthermore, more often than not we do not know beforehand the degree of the polynomial we wish to fit and we must proceed by adding powers until an adequate fit is attained. Each time as a higher power is added, a whole new set of normal equations must be solved. In this sort of situation, the

use of orthogonal functions takes on an added importance.

When the x - observations are equally spaced which often happens or can be arranged, they may, perhaps by a change of origin and such be made to read 0, 1, 2, ..., N-1. The coefficients of the orthogonal polynomials and the values taken by the polynomials then are functions of N only. It then is feasible to determine these polynomials, one set for each sample size, and to list the values taken by them.

If we write the polynomials in the form

$$P_0 = \lambda_{00}$$

$$P_1 = \lambda_{10} + \lambda_{11}\, x$$

$$P_2 = \lambda_{20} + \lambda_{21}\, x \; + \lambda_{22}\, x^2$$

$$\vdots$$

and choose the λ's so that $S\, P_{i\alpha}\, P_{j\alpha} = 0$, $i \neq j$, the normal equations are

$$S\, P_{i\alpha}^2\, B_i = G_i = SP_{i\alpha}\, y_\alpha \,. \qquad i = 0, 1, 2, \ldots$$

and if the values of the $P_{i\alpha}$ are tabulated, we need only compute $S\, P_{i\alpha}\, y_\alpha$.

We note that in each P, one coefficient is available for arbitrary disposal. R.A. Fisher, who was an early advocate of the use of these polynomials, included in his Statistical Methods for Research Workers an elegant finite difference procedure for carrying out the fitting without having the $P_{i\alpha}$ values. For his purposes, it was convenient to assign each λ_{ii} the value unity. The resulting polynomials he called ξ_0, ξ_1, ξ_2 and so on.

Later on, when the Fisher and Yates tables were assembled, values

of the $P_{i\alpha}$ were included, on Yates's insistence and against Fisher's wishes. For this purpose, the ξ - polynomals are not suitable, since their values are mostly fractions. Another disposal of the arbitrary constants is required, one that will yield integers values with no common factor. Hence we get $\xi_i' = \lambda(i,N)\ \xi_i$, with $\lambda(i, N)$ chosen to produce integer values in their lowest terms. Fisher and Yates, and all other tabulations, list values of the ξ', $S\ \xi'^2$, and the $\lambda(i, N)$.

Extension to two (or more) dimensions.

If we have two independent variables, u and v we may wish to fit a polynomial

$$Y = b_{00} + b_{10}\ u + b_{01}\ v + b_{11}\ u\ v + b_{20}\ u^2 + b_{02}\ v^2 + \text{etc.}$$

This may be re-written in the form

$$Y = B_{00}\ \xi_0'(u)\ \xi_0(v) + B_{10}\ \xi_1'(u)\ \xi_0'(v) + B_{11}\ \xi_1'(u)\ \xi_1'(v) + \text{etc.}$$

If the observations on u are equally spaced and also those on v, i.e. the observations are on a rectangular grid, the normal equations take the form

$$S\ \xi_i'^2(u)\ \xi_j'^2(v)\ B_{ij} = S\ y\ \xi_i'(u)\ \xi_j'(v)\ .$$

Details may be found in the tables Values and Integrals of the Orthogonal Polynomials up to $N = 26$ (D.B. DeLury (1950)).

The fundamental distribution Theorem.

The theorem is concerned with testing a specified subset of regression coefficients to see whether their inclusion results in a significantly better fit, i.e. whether their corresponding β's are not all zero.

Without loss of generality, this may be set up as follows. Suppose we have fitted a regression

$$Y = b_0 x_0 + b_1 x_1 + \ldots + b_q x_q + b_{q+1} x_{q+1} + \ldots + b_p x_p .$$

We ask whether this regression fits appreciably better than

$$Y' = b'_0 x_0 + b'_1 x_1 + \ldots + b'_q x_q .$$

The question will be explored by examining the differences between the two s.s. residuals, $S(y-Y)^2$ and $S(y-Y')^2$.

The following statements will be proved.

1. $S(y-Y')^2 - S(y-Y)^2$

2. $= S(Y-\bar{y})^2 - S(Y'-\bar{y})^2$ when $x_0 = 1$

3. $= S(Y-Y')^2$

4. $= \sum_{i,j=q+1}^{p} a^*_{ij} b_i b_j$ where $\sum_{j=q+1}^{p} a^*_{ij} c_{jk} = \delta_{ik}$ $i, k = q+1, \ldots, p$

5. $= \sum_{i=q+1}^{p} b_i g^*_i$ where $\sum_{j=q+1}^{p} a^*_{ij} b_j = g^*_i$ $i = q+1, \ldots, p$

6. $= \sum_{i,j=q+1}^{p} c_{ij} g^*_i g^*_j$

Each of these s.s. is distributed independently of $S(y-Y)^2$ and is $\sigma^2 \chi^2_{(p-q)}$ if $\beta_{q+1} = \ldots = \beta_p = 0$. This sum of squares may be called "the s.s. associated with b_{q+1}, b_{q+2} ... b_p" or "the s.s. attributable to x_{q+1}, x_{q+2}, ..., x_p".

The passage from (1) to (2) is immediate, in view of the identities $S(y-\bar{y})^2 = S(y-Y)^2 + S(Y-\bar{y})^2$ and

$$S(y-\bar{y})^2 = S(y-Y')^2 + S(Y'-\bar{y})^2 \ .$$

The passage from (1) to (3) could no doubt be checked by an appeal to the normal equations. It will be deduced in another way shortly. The passage from (1) to (4) is the crucial element in this list. The equivalence of (4), (5) and (6) is immediate.

Obviously, these statements should become rather obvious if we go over to the orthogonal functions we introduced earlier. The two equations become $Y = B_0 P_0 + \ldots + B_q P_q + B_{q+1} P_{q+1} + \ldots + B_p P_p$

$$Y' = B_0 P_0 + \ldots + B_q P_q \ .$$

Then, $S(y-Y)^2 = S\ y^2 - \sum_0^p B_i^2\ S\ P_{i\alpha}^2$

$$S(y-Y')^2 = Sy^2 - \sum_0^q B_i^2\ S\ P_{i\alpha}^2$$

$$S(y-Y')^2 - S(y-Y)^2 = \sum_{q+1}^p B_i^2\ S\ P_{i\alpha}^2 \ .$$

also, $Y-Y' = \sum_{q+1}^p B_i\ P_i$, $(Y-Y')^2 = \sum_{i,j=q+1}^p B_i\ B_j\ P_i\ P_j$.

$$S(Y-Y')^2 = \sum B_i\ B_j\ S\ P_i\ P_j = \sum B_i^2\ S\ P_i^2 \ ,$$

which establishes the equality of (1) and (3).

Now, B_i is normal, variance $\frac{\sigma^2}{S\ P_i^2}$, distributed independently of all other B's and of the s.s. residuals. If $E\ b_i = 0$, $i = q+1, \ldots, p$, then, from the relations connecting the b's and the B's, so also $E\ B_i = 0$, $i = q+1, \ldots p$ then, $B_i \sqrt{S\ P_i^2}$ is $N(0,\sigma^2)$ and $\sum_{q+1}^p B_i^2\ S\ P_i^2$

is $\sigma^2\ \chi^2_{(p-q)}$, independent of $S(y-Y)^2$, which is $\sigma^2\ \chi^2_{(N-p-1)}$. It remains only to establish the equivalence of (1) and (4). In the fitting based on the orthogonal functions,

$$\sum_{i,j=q+1}^{p} a^*_{ij}\ b_i\ b_j \quad \text{becomes} \quad \sum_{i=q+1}^{p} A_{ii}\ B_i^2\ , \quad \text{where } A_{ii} = S\ P_i^2\ .$$

In this system, then, (1) and (4) are identical. The proof is complete, then, if we establish that the quadratic form $\sum_{i,j=q+1}^{p} a^*_{ij}\ b_i\ b_j$ is invariant under the transformation used to pass from the original fitting to that based on the orthogonal functions.

This is a purely mathematical exercise and could well be omitted, but is included here for completness. It would be tedious without the devices of linear algebra, hence the following summary in matrix notation. Let x_p stand for the matrix

$$x_p = \begin{pmatrix} x_{01} & x_{11} & x_{q1} & x_{q+1\ 1} & x_{p1} \\ x_{02} & x_{12} & x_{q2} & x_{q+1\ 2} & x_{p2} \\ x_{0\alpha} & x_{1\alpha} & x_{q\alpha} & x_{q+1\ \alpha} & x_{p\alpha} \\ x_{0N} & x_{1N} & x_{qN} & x_{q+1\ N} & x_{pN} \end{pmatrix}$$

Let $\underset{\sim}{y}$ be the column vector

$$\underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

Then, the matrix of the coefficients of the normal equation is $a_p = x'_p\ x_p$ and that of the right sides is $\underset{\sim}{g}_p = x'\underset{\sim}{y}$. If $\underset{\sim}{b}_p$ is the column vector of the b's, the normal equations are

$$a_p \; \underset{\sim}{b}_p = \underset{\sim}{g}_p \; .$$

The inverse matrix is $c_p = a_p^{-1}$.

The matrix a^* is defined by partitioning the inverse matrix as follows.

$$\left(\begin{array}{ccc|ccc} c_{11} & \cdots\cdots & c_{1q} & c_{1(q+1)} & \cdots\cdots & c_{1p} \\ \vdots & & \vdots & \vdots & & \vdots \\ c_{q1} & \cdots\cdots & c_{qq} & c_{q(q+1)} & \cdots\cdots & c_{qp} \\ \hline c_{(q+1)1} & \cdots\cdots & c_{(q+1)q} & c_{(q+1)(q+1)} & \cdots & c_{(q+1)p} \\ \vdots & & \vdots & \vdots & & \vdots \\ c_{p1} & \cdots\cdots & c_{pq} & c_{p(q+1)} & \cdots\cdots & c_{pp} \end{array}\right) = \begin{pmatrix} c_q & r \\ r' & c_{p-q} \end{pmatrix} \quad \text{(say)}$$

where r is a matrix which need not be specified for this discussion. Then, $a^* = c_{p-q}^{-1}$. Similarly, partitioning $\underset{\sim}{b}_p$ into $\begin{pmatrix} \underset{\sim}{b}_q \\ \underset{\sim}{b}_{p-q}' \end{pmatrix}$, the quadratic form $\sum_{i,j=q+1}^{p} a^*_{ij} \; b_i \; b_j$ may be expressed as $\underset{\sim}{b}'_{p-q} \; c_{p-q}^{-1} \; \underset{\sim}{b}_{p-q}$.

Any linear transformation of the independent variables specified by a matrix λ_p, $|\lambda_p| \neq 0$, transforms the matrix of observations into $\xi'_p = \lambda_p \; X'_p$. The transformed normal equations will be written

$$A_p \; \underset{\sim}{B}_p = \underset{\sim}{G}_p \text{ where } A_p = \lambda_p \; a_p \; \lambda'_p \text{ and } \underset{\sim}{G}_p = \lambda_p \; \underset{\sim}{g}_p \; .$$

Then $\underset{\sim}{b}_p = \lambda'_p \; \underset{\sim}{B}_p$. The inverse matrix $C_p = (\lambda'_p)^{-1} \; c_p \; \lambda_p^{-1}$.

For the transformation we are using, $\lambda_p = \begin{pmatrix} \lambda_q & 0 \\ S & \lambda_{p-q} \end{pmatrix}$, although this is not essential to the proof. Then,

$$c_p = \begin{bmatrix} c_q & r \\ r' & c_{p-q} \end{bmatrix} = \lambda_p' \, C_p \, \lambda_p = \begin{bmatrix} \lambda_q' & S' \\ 0 & \lambda_{p-q}' \end{bmatrix} \begin{bmatrix} C_q & R \\ R' & C_{p-q} \end{bmatrix} \begin{bmatrix} \lambda_q & 0 \\ S & \lambda_{p-q} \end{bmatrix}$$

$$= \begin{bmatrix} x & x \\ x & \lambda_{p-q}' \, C_{p-q} \, \lambda_{p-q} \end{bmatrix} . \quad \text{Thus we have}$$

$$c_{p-q} = \lambda_{p-q}' \, C_{p-q} \, \lambda_{p-q}$$

In the same way, we get

$$\underset{\sim}{b}_{p-q} = \lambda_{p-q}' \, \underset{\sim}{B}_{p-q} . \quad \text{Hence,}$$

$$\underset{\sim}{b}_{p-q}' \, c_{p-q}^{-1} \, \underset{\sim}{b}_{p-q} = (\lambda_{p-q}' \, \underset{\sim}{B}_{p-q})' (\lambda_{p-q}' \, C_{p-q} \, \lambda_{p-q})^{-1} \, (\lambda'_{p-q} \, \underset{\sim}{B}_{p-q})$$

$$= \underset{\sim}{B}_{p-q}' \, C_{p-q}^{-1} \, \underset{\sim}{B}_{p-q} .$$

This demonstrates the invariance of the quadratic form $\sum a_{ij}^{*} \, b_i \, b_j$ and completes the proof.

The solution of normal equations.

The pattern of solution recommended here is not the most efficient, in a numerical sense but it has considerable merit, in some circumstances. It is called Chio's rule, the method of pivotal condensation and the method of sweep-out, with the additional feature that each pivot is used to reduce to zero the entries in the rows above the pivot as well as those in the rows below it. Also, augmenting the matrix with the additional last row produces sums of squares of residuals.

The example shows the fitting $Y = b_0 \, x_0 + b_1 \, x_1 + b_2 \, x_2 + b_3 \, x_3$. On the way to the solution, other fittings are provided.

1. $Y^{(0)} = b_0^{(0)} x_0 , c_{00}^{(0)}, S(y-Y^{(0)})^2$

2. $Y^{(1)} = b_0^{(1)} x_0 + b_1^{(1)} x_1, c_{00}^{(1)}, c_{11}^{(1)}, S(y-Y^{(1)})^2.$

3. $Y^{(2)} = b_0^{(2)} x_0 + b_1^{(2)} x_1 + b_2^{(2)} x_2, c_{00}^{(2)}$ etc., $S(y-Y^{(2)})^2$.

The procedure starts with the element in the upper left corner as pivot, reduces it to 1 by a division, then, with the requisite multiplications and subtractions, reduces all the elements in its column to zero. We get, then, $a_{ij.0} = a_{ij} - \frac{a_{i0}\, a_{0j}}{a_{00}}$, $g_{i.0} = g_i - \frac{a_{i0}}{a_{00}} g_0$

More generally, $a_{ij.k} = a_{ij.(k-1)} - \frac{a_{ik.(k-1)}\, a_{kj.(k-1)}}{a_{kk.(k-1)}}$

$$g_{i.k} = g_{i.(k-1)} - \frac{a_{ik.(k-1)}}{a_{kk.(k-1)}} g_{k.(k-1)}$$

a_{00}	a_{01}	a_{02}	a_{03}	g_0	1	0	0	0	$\Sigma a_{ij} + g_0 + 1$
a_{10}	a_{11}	a_{12}	a_{13}	g_1	0	1	0	0	$\Sigma a_{1j} + g_1 + 1$
a_{20}	a_{21}	a_{22}	a_{23}	g_2	0	0	1	0	$\Sigma a_{2j} + g_2 + 1$
a_{30}	a_{31}	a_{32}	a_{33}	g_3	0	0	0	1	$\Sigma a_{3j} + g_3 + 1$
g_0	g_1	g_2	g_3	Sy^2					$\Sigma g_i + Sy^2$
1	*	*	*	$b_0^{(0)}$	$c_{00}^{(0)}$	0	0	0	✓
0	$a_{11.0}$	$a_{12.0}$	$a_{13.0}$	$g_{1.0}$	*	1	0	0	✓
0	$a_{21.0}$	$a_{22.0}$	$a_{23.0}$	$g_{2.0}$	*	0	1	0	✓
0	$a_{31.0}$	$a_{32.0}$	$a_{33.0}$	$g_{3.0}$	*	0	0	1	✓
0	$g_{1.0}$	$g_{2.0}$	$g_{3.0}$	$S(y-Y^{(0)})^2$					✓
1	0	*	*	$b_0^{(1)}$	$c_{00}^{(1)}$	$c_{01}^{(1)}$	0	0	✓
0	1	*	*	$b_1^{(1)}$	$c_{10}^{(1)}$	$c_{11}^{(1)}$	0	0	✓
0	0	$a_{22.1}$	$a_{23.1}$	$g_{2.1}$	*	*	1	0	✓
0	0	$a_{32.1}$	$a_{33.1}$	$g_{3.1}$	*	*	0	1	✓
0	0	$g_{2.1}$	$g_{3.1}$	$S(y-Y^{(1)})^2$					✓
1	0	0	*	$b_0^{(2)}$	$c_{00}^{(2)}$	$c_{01}^{(2)}$	$c_{02}^{(2)}$	0	✓
0	1	0	*	$b_1^{(2)}$	$c_{10}^{(2)}$	$c_{11}^{(2)}$	$c_{13}^{(2)}$	0	✓
0	0	1	*	$b_2^{(2)}$	$c_{20}^{(2)}$	$c_{21}^{(2)}$	$c_{22}^{(2)}$	0	✓
0	0	0	$a_{33.2}$	$g_{3.2}$	*	*	*	1	✓
0	0	0	$g_{3.2}$	$S(y-Y^{(2)})^2$					✓
1	0	0	0	b_0	c_{00}	c_{01}	c_{02}	c_{03}	✓
0	1	0	0	b_1	c_{10}	c_{11}	c_{12}	c_{13}	✓
0	0	1	0	b_2	c_{20}	c_{21}	c_{22}	c_{23}	✓
0	0	0	1	b_3	c_{30}	c_{31}	c_{32}	c_{33}	✓
0	0	0	0	$S(y-Y)^2$					✓

Some comments on this form of solution.

1. The matrix a_{ij} is symmetric and this symmetry is maintained throughout the solution. Likewise, the c matrices are symmetric. Therefore, several numbers have been calculated twice. One could, of course, avoid this duplication, but the internal checks are worth having too, because one can see the effects of round-off mistakes by comparing duplicates.

2. The Chio rule has a property described as "preserving the elements of the inverse matrix". By this is meant the following.

$$\begin{pmatrix} a_{11.0} & a_{12.0} & a_{13.0} \\ a_{21.0} & a_{22.0} & a_{23.0} \\ a_{31.0} & a_{32.0} & a_{33.0} \end{pmatrix}^{-1} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}$$

$$\begin{pmatrix} a_{22.1} & a_{23.1} \\ a_{32.1} & a_{33.1} \end{pmatrix}^{-1} = \begin{pmatrix} c_{22} & c_{23} \\ c_{32} & c_{33} \end{pmatrix}$$

$$\begin{pmatrix} a_{33.2} \end{pmatrix}^{-1} = \begin{pmatrix} c_{33} \end{pmatrix}$$

It is sufficient to check this for the first cycle

$$\sum_{j=1}^{p} a_{ij.0}\, c_{jk} = \sum_{j=0}^{p} \left(a_{ij} - \frac{a_{i0}\, a_{0j}}{a_{00}} \right) c_{jk}$$

$$= \sum_{j=0}^{p} a_{ij}\, c_{jk} - \frac{a_{i0}}{a_{00}} \sum_{j=0}^{p} a_{0j}\, c_{jk}$$

$$= \delta_{ik} - 0 \text{ for } i, k = 1, 2, \ldots, p.$$

This feature has useful consequences. The matrices $(a_{ij.k})$ and $(g_{i.k})$ satisfy the conditions of the (a^*_{ij}) and (g^*) of the fundamental

theorem. If, for example, we ask for the s.s. attributable to x_2 and x_3, we could express it in several convenient ways.

(a) $S(y-Y^{(1)})^2 - S(y-Y)^2$. This is the most convenient way.

(b) $b_2\, g_{2.1} + b_3\, g_{3.1}$

(c) $a_{22.1}\, b_2^2 + 2\, a_{23.1}\, b_2\, b_3 + a_{33.1}\, b_3^2$

(d) $c_{22}\, g_{2.1}^2 + 2\, c_{23}\, g_{2.1}\, g_{3.1} + c_{33}\, g_{3.1}^2$

If one should ask for the s.s. attributable to x_1 and x_3, it would probably be simplest to compute $\begin{pmatrix} c_{11} & c_{13} \\ c_{31} & c_{33} \end{pmatrix}^{-1}$ to get the a^*'s and proceed as in (c).

3. Occasionally we may want to use calculations of this kind when the solution has been reached algebraically, perhaps without following formally the Chio rule. We can check the results to see if they are those yielded by the Chio rule, because under this rule, the coefficients of the g's with the largest subscripts are equal to unity.

4. In the usual case, when $x_0 = 1$, the coefficients after the first cycle, $a_{ij.0}$, are sums of squares and products of deviations from the averages.

$$a_{ij.0} = S\,(x_{i\alpha} - x_i)\;(x_{j\alpha} - x_j)\ .$$

This fact will be used in the analysis of covariance.

The fitting of a pair of parallel straight lines.

A number of situations call for the fitting of two or more lines, constrained to be parallel.

(1) An experiment with animals, the response being final weight, with initial weight recorded. In such instances, the object is chiefly to remove from error that part attributable to variation in initial weight.

(2) One may wish to study the regression of one variable on another, in circumstances in which one has to use several samples. For example a study of the dependence of reading ability on IQ may use several classes and even several schools.

The numerial procedures developed for such use are called the analysis of covariance. As we shall see, the analysis of covariance is simply regression theory, adapted to take advantage of whatever orthogonality is present by making part of the computations with the simple procedures of the analysis of variance.

(3) The fitting of parallel lines occurs also in bioassay.

Let $(y_{i\alpha}, x_{i\alpha})$, $i = 1, 2$, $\alpha = 1, 2, \ldots, n_i$, be two samples, to each of which we wish to fit a linear regression, with the constraint that the lines must be parallel. We could, of course, write the equations

$$Y^{(1)} = b_0^{(1)} + b_1 x$$

$$Y^{(2)} = b_0^{(2)} + b_1 x$$

and minimize the s.s. $\Sigma (y_{1\alpha} - Y^{(1)})^2 + \Sigma (y_{2\alpha} - Y^{(2)})^2$. Preferably, we can reduce the fitting to the fitting of a single regression by the

elegant device (first used, as far as I know, by Wilks, Metron, 1938) of introducing an indicator variable $\delta_{i\alpha} = 0 \quad i = 1$

$= 1 \quad i = 2$

and fitting $Y = b_0 + b_1\delta + b_2 x$ to both samples, regarded as a single sample $(y_{i\alpha}, x_{i\alpha}, \delta_{i\alpha})$.

The normal equations may be written down at once

$$
\begin{array}{lllll}
n_1 + n_2 & S\delta & Sx & Sy & \\
S\delta & S\delta 2 & S\delta x & S\delta y & \\
Sx & S\delta x & Sx^2 & Sxy & \\
n_1 + n_2 & n_2 & \Sigma x_1 + \Sigma x_2 & \Sigma y_1 + \Sigma y_2 & (g_0) \\
n_2 & n_2 & \Sigma x_2 & \Sigma y_2 & (g_1) \\
\Sigma x_1 + \Sigma x_2 & \Sigma x_2 & \Sigma x_1^2 + \Sigma x_2^2 & \Sigma x_1 y_1 + \Sigma x_2 y_2 & (g_2)
\end{array}
$$

We may note, in passing, that the second of these equations may be written $b_0 + b_1 + b_2 \bar{x}_2 = \bar{y}_2$ and, subtracting the second from the first, $b_0 + b_2 \bar{x}_1 = \bar{y}_1$. Thus, the two lines pass through $(\bar{x}_1, \bar{y}_1)$ and $(\bar{x}_2, \bar{y}_2)$.

The object, here, is to solve these equations algebraically to obtain formulae for b_0, b_1, b_2 and also for the elements of the inverse matrix. This will be carried out by replacing the right sides by symbols, g_0, g_1, g_2, solving, then replacing the g's by the proper values, obtain b_0, b_1, b_2 and by replacing the g's by proper selections of 1's and 0's obtain the inverse matrix.

The solution turns out to be

$$W_{xx}\ "b_2" = g_2 - \frac{g_1 - g_2}{n_1}\ \Sigma x_1 - \frac{g_1}{n_2}\ \Sigma x_2$$

$$"b_1" = \frac{g_1}{n_2} - \frac{g_0 - g_1}{1} - "b_2"\ (\bar{x}_2 - \bar{x}_1)$$

$$"b_0" = \frac{g_0 - g_1}{n_1} - "b_2"\ \bar{x}_1$$

where $W_{xx} = \Sigma(x_{1\alpha} - \bar{x}_1)^2 + \Sigma(x_{2\alpha} - \bar{x}_2)^2$, the within-samples sum of squares. Substituting for the g's

$$W_{xx}\ b_2 = \Sigma x_1 y_1 + \Sigma x_2 y_2 - \frac{\Sigma x_1\ \Sigma y_1}{n_1} - \frac{\Sigma x_2 \Sigma y_2}{n_2}$$

$$= \Sigma(x_1 - \bar{x}_1)(y_1 - \bar{y}_1) + \Sigma(x_2 - \bar{x}_2)(y_2 - \bar{y}_2)$$

$$= W_{xy}\ ,$$

the within-samples sum of products of deviations

$$b_1 = \bar{y}_2 - \bar{y}_1 - b_2\ (\bar{x}_2 - \bar{x}_1)$$

$$b_0 = \bar{y}_1 - b_2\ \bar{x}_1$$

Putting $g_2 = 1$, $g_1 = g_0 = 0$ in "b_2" yields

$$c_{22} = \frac{1}{W_{xx}}$$

Putting $g_1 = 1$, $g_2 = g_0 = 0$ in "b_1" yields

$$c_{11} = \frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_2 - \bar{x}_1)^2}{W_{xx}}$$

The other elements of the inverse matrix may be found in the same way, if they should be needed.

The residual s.s. is

$$Sy^2 - (y_1 - b_1 x_1)(\Sigma y_1 + \Sigma y_2) - [\bar{y}_2 - \bar{y}_1 - b_2(x_2 - x_1)] - b_2[\Sigma x_1 y_1 + \Sigma x_2 y_2]$$

$$= \Sigma y_1^2 + \Sigma y_2^2 - \frac{(\Sigma y_1)^2}{n_1} - \frac{(\Sigma y_2)^2}{n_2} - b_2 \left[\Sigma x_1 y_1 + \Sigma x_2 y_2 - \frac{\Sigma x_1 \, \Sigma y_1}{n_1} - \frac{\Sigma x_2 \, \Sigma y_2}{n_2}\right]$$

$$= \Sigma (y_1 - \bar{y}_1)^2 + \Sigma (y_2 - \bar{y}_2)^2 - b_2 \left[\Sigma (x_1 - \bar{x}_1)(y_1 - \bar{y}_1) + \Sigma (x_2 - \bar{x}_2)(y_2 - \bar{y}_2)\right]$$

$$= W_{yy} - b_2 W_{xy}$$

$$= W_{yy} - \frac{(W_{xy})^2}{W_{xx}} \quad .$$

This is the error s.s., with $n_1 + n_2 - 3$ d.f.

It is called the reduced error, inasmuch as the variation attributable to variation in x has been removed from it.

$b_1 = \bar{y}_2 - \bar{y}_1 - b_2 (\bar{x}_2 - \bar{x}_1)$ is called the adjusted difference of sample means, that is, any difference between $\bar{y}_1$ and $\bar{y}_2$ resulting from a difference in the x averages has been adjusted.

If there were no x's to be concerned about the analysis of variance table, corresponding to the arrangement, would be

	d.f.	s.s.
between samples	1	$B_{yy} = \Sigma n_i (\bar{y}_i - \bar{y})^2 = \sum \frac{T_i^2(y)}{n_i} - \frac{T^2(y)}{n_1 + n_2}$
within samples	$n_1 + n_2 - 2$	$W_{yy} = \Sigma\Sigma (y_{i\alpha} - \bar{y}_i)^2$ by subtraction
Total	$n_1 + n_2 - 1$	$T_{yy} = Sy_{i\alpha}^2 - \frac{T^2(y)}{n_1 + n_2}$.

The formulae reached above dictate that we make the same computation for x and, no doubt, for the xy products as well. It is easy to check that the same rules apply. The "total" sum of products is

$$T_{xy} = \sum_{i=1}^{2} \sum_{\alpha=1}^{n_i} (x_{i\alpha} - \bar{x})\ (y_{i\alpha} - \bar{y})$$

$$= \sum \sum (x_{i\alpha} - \bar{x}_i + \bar{x}_i - \bar{x})\ (y_{i\alpha} - \bar{y}_i + \bar{y}_i - \bar{y})$$

$$= \sum \sum (x_{i\alpha} - \bar{x}_i)\ (y_{i\alpha} - \bar{y}_i) + \sum_i n_i\ (\bar{x}_i - \bar{x})\ (\bar{y}_i - \bar{y})\ ,$$

$$= W_{xy} + B_{xy}\ .$$

B_{xy} is easily seen to be equal to

$$\sum_i \frac{T_i(x)\ T_i(y)}{n_i} - \frac{T(x)\ T(y)}{n_1+n_2}\ .$$

We therefore calculate the table

	d.f.	(xx)	(xy)	(yy)
between samples	1	B_{xx}	B_{xy}	B_{yy}
within samples	$n_1 + n_2 - 2$	W_{xx}	W_{xy}	W_{yy}
Total	$n_1 + n_2 - 1$	T_{xx}	T_{xy}	T_{yy}

The numbers in the within-samples row would be obtained by subtraction.

The numbers we require from this are

slope $b_2 = \frac{W_{xy}}{W_{xx}}$.

s.s. attributable to regression on x, $\frac{b_2^2}{c_{22}} = \frac{W_{xy}^2}{W_{xx}}$, 1 d.f.

s.s. deviations (error) $W_{yy} - \frac{W_{xy}^2}{W_{xx}}$, $n_1 + n_2 - 3$ d.f.

Tests of significance

Sometimes there is a need to test $E\ b_2 = 0$. If it is,

$$\frac{W_{xy}^2/W_{xx}}{\left(W_{yy} - W_{xy}^2/W_{xx}\right)/n_1+n_2-3} = F(1, n_2 + n_2 - 3)\ .$$

Sometimes we may want to test $E\, b_1 = 0$. If it is,

$$\frac{b_1}{s\sqrt{c_{11}}} = t_{(n_1+n_2-3)}, \text{ where } s^2 = \frac{1}{n_1+n_2-3}\left(W_{yy} - \frac{W_{xy}^2}{W_{xx}}\right).$$

Another, less obvious, approach is now convenient. It stems from another of the rules for finding the s.s. attributable to set of independent variables - fit with them included, fit with them excluded and calculate the difference between the residuals. Here, we are asking for the s.s. attributable to δ. Fitting without δ amounts to fitting both samples with a single line, the results being, obviously,

attributable to x	T_{xy}^2/T_{xx}	1 d.f.
deviations	$T_{yy} - \frac{T_{xy}^2}{T_{xx}}$	$n_1 + n_2 - 2$ d.f.

The required s.s., attributable to δ, is then

$$\left(T_{yy} - \frac{T_{xy}^2}{T_{xx}}\right) - \left(W_{xy} - \frac{W_{xy}^2}{W_{xx}}\right) \text{ with 1 d.f.}$$

$$= B_{yy} + \frac{W_{xy}^2}{W_{xx}} - \frac{T_{xy}^2}{T_{xx}}.$$

This is called the adjusted between-samples s.s..

These computations are usually carried out in the following pattern.

	attributable to x		deviations from regression	
	d.f.	s.s.	d.f.	s.s.
error	1	W_{xy}^2/W_{xx}	n_1+n_2-3	$W_{yy} - \frac{W_{xy}^2}{W_{xx}} = E$
error + between samples	1	T_{xy}^2/T_{xx}	n_1+n_2-2	$T_{yy} - \frac{T_{xy}^2}{T_{xx}} = G$
adjusted between samples			1	$G-E$.

The adjusted s.s. may be compared with s^2 in an F-test.

In (1), we would want to test b_1, not b_2.

In (2), we would want to test b_2, not b_1.

Occasionally, we may wish to test both b_1 and b_2.

If we wish to present adjusted means,

$$\overline{y}'_i = \overline{y}_i - b_2 (\overline{x}_i - \overline{x}).$$

There are many occasions in which we should check whether parallel lines are warranted, before embarking on analysis of the sort just discussed. The approach to this check should be obvious: fit lines to the individual samples, allowing each its own slope, then fit with the lines constrained to the parallel, then compare the s.s. residuals.

	d.f.	(xx)	(xy)	(yy)
sample A	$n_1 - 1$	A_{xx}	A_{xy}	A_{yy}
sample C	$n_2 - 1$	C_{xx}	C_{xy}	C_{yy}
within samples	$n_1 + n_2 - 2$	W_{xx}	W_{xy}	W_{yy}

	attributable to x		deviations	
	d.f.	s.s.	d.f.	s.s.
sample A	1	A_{xy}^2/A_{xx}	$n_1 - 2$	$A_{yy} - A_{xy}^2/A_{xx} = E_1$
sample C	1	C_{xy}^2/C_{xx}	$n_2 - 2$	$C_{yy} - C_{xy}^2/C_{xx} = E_2$
within samples	1	W_{xy}^2/W_{xy}	$n_1 + n_2 - 3$	$W_{yy} - W_{xy}^2/W_{xx} = E$

The $E - (E_1 + E_2)$, with 1 d.f., reflects the improvement in fit when the lines are permitted their individual slopes. One would expect $E - (E_1 + E_2)$ to be comparable with $(E_1 + E_2)/(n_1 + n_2 - 4)$ in an F-test.

To check on this, we may represent the fitting of separate lines by a single regression by writing

$$Y = b_0 + b_1 x + \delta(a_0 + a_1 x)$$
$$= B_0 + B_1 \delta + B_2 x + B_3 \delta x,$$

which differs from that used to fit parallel lines in the term in δx. To test whether this gives a better fit than that using parallel lines is to test $E B_3 = 0$.

The normal equations for fitting this regression are

$$\begin{aligned}
(n_1 + n_2) B_0 + B_1 S\delta &+ B_2 Sx &+ B_3 S\delta x &= Sy \\
B_0 S\delta &+ B_1 S\delta^2 &+ B_2 S\delta x &+ B_3 S\delta^2 x &= S\delta y \\
B_0 Sx &+ B_1 S\delta x &+ B_2 Sx^2 &+ B_3 S\delta x^2 &= Sxy \\
B_0 S\delta x &+ B_1 S\delta^2 x &+ B_2 S\delta x^2 &+ B_3 S\delta^2 x^2 &= S\delta xy
\end{aligned}$$

$n_1 + n_2$	n_2	Sx	Σx_2	Sy
n_2	n_2	Σx_2	Σx_2	Σy_2
Sx	Σx_2	Sx^2	Σx_2^2	Sxy
Σx_2	Σx_2	Σx_2^2	Σx_2^2	$\Sigma x_2 y_2$

From these, we see that the normal equations separate into two sets:

$$n_1 B_0 + B_2 \Sigma x_1 = \Sigma y_1 \qquad n_2 (B_0 + B_1) + (B_2 + B_3) \Sigma x_2 = \Sigma y_2$$
$$B_0 \Sigma x_1 + B_2 \Sigma x_1^2 = \Sigma x_1 y_1 \qquad (B_0 + B_1)\Sigma x_2 + (B_2 + B_3) \Sigma x_2^2 = \Sigma x_2 y_2$$

One set represents the fitting of a line to sample 1, with residual s.s. $\Sigma y_1^2 - B_0 \Sigma y_1 - B_2 \Sigma x_1 y_1 = E_1$; the other is the fitting of a line to sample 2, with residual s.s. $\Sigma y_2^2 - (B_0 + B_1) \Sigma y_2 -$

$(B_2 + B_3)\ \Sigma x_2 y_2 = E_2$. The residual about the entire regression is

$$Sy^2 - B_0 Sy - B_1 \Sigma y_2 - B_2 Sxy - B_3\ \Sigma x_2 y_2$$

$$= \Sigma y_1^2 + \Sigma y_2^2 - B_0 \Sigma y_1 - B_2\ \Sigma x_1 y_1 - (B_0 + B_1)\ \Sigma y_2 - (B_2 + B_3)\ \Sigma x_2 y_2$$

$$= E_1 + E_2, \text{ with } n_1 + n_2 - 4 \text{ d.f.}$$

The s.s. attributable to δx is, by one of our rules, $E - (E_1 + E_2)$, with 1 d.f. distributed independently of $E_1 + E_2$. This confirms the test for parallelism.

Extension of the use of indicator variables

The indicator variable used in the discussion shows how it could be used to bring a qualitative difference (between samples) within the scope of regression theory. If we suppress the x-variable in the discussion, i.e. fit $Y = b_0 + b_1\ \delta$, we see that we obtain the standard analysis for this pattern, the completely randomized experiment. We might use indicator variables in other ways, too. For example, we might have used two, δ_1 and δ_2, with $\delta_{1i\alpha} = 0$ in sample 1 and 1 in sample 2, $\delta_{2i\alpha} = 1$ in sample 1, 0 in sample 2. Then, fit the regression $Y = c_1\ \delta_1 + c_2\ \delta_2$. Note that we cannot include a constant in this regression, because it would imply a linear relation among the independent variables, $\delta_1 + \delta_2 = 1 = x_0$.

This way of using indicator variables extends at once to any number of samples. If the observations are $y_{i\alpha}$, $i = 1,2,\ldots,k$, $\alpha = 1,2,\ldots,n_i$, we may postulate $y_{i\alpha} = \gamma_i + \varepsilon_{i\alpha}$, $\varepsilon : N(0,\sigma^2)$.

To get formal identifications with regression theory, we introduce indicator functions $\delta_1, \delta_2, \ldots, \delta_k$ with

$$\delta_i = 1 \quad \text{in sample } i$$
$$= 0 \quad \text{in all others.}$$

The regression equation is then

$$Y = c_1 \delta_1 + \ldots. c_i \delta_i + \ldots. c_k \delta_k.$$

The normal equations are

$$c_1 S \delta_1^2 + c_2 S \delta_1 \delta_2 + \ldots. c_k S \delta_1 \delta_k = S \delta_1 y,$$

and so on.

These reduce to

$$n_i c_i = \sum_\alpha y_{i\alpha}, \qquad i = 1,2,\ldots, k.$$

The residual s.s. is

$$Sy^2 = \sum_i c_i \sum_\alpha y_{i\alpha} = Sy^2 - \sum_i \frac{\left(\sum_\alpha y_{i\alpha}\right)^2}{n_i}$$

which is the within sample s.s.

This is perhaps the simplest way of bringing regression theory into the picture, but it suffers from the fact that none of the contrasts we may wish to study appears directly and the s.s. attributable to regression on the δ's has no usefulness.

Occasionally, it may be worth while to transform the δ's in order that the regression coefficients will exhibit contrasts we want to see. As an example, only, we might transform $\delta_1, \ldots, \delta_k$ into $u_0, \ldots, u_{k-1}$ by the linear transformation

$$u_1 = \delta_1 - \delta_2$$

$$u_2 = \delta_1 + \delta_2 - 2\delta_3$$

$$\vdots$$

$$u_{k-1} = \delta_1 + \delta_2 + \delta_3 + \ldots + \delta_{k-1} - (k-1)\,\delta_k$$

$$u_0 = \delta_1 + \delta_2 + \ldots + \delta_{k-1} + \delta_k = 1$$

The regression $Y = b_0 + b_1 u_1 + \ldots + b_{k-1}\, u_{k-1}$ has normal equations

$$Nb_0 + (n_1 - n_2)b_1 + (n_1 + n_2 - 2n_3)b_2 + \ldots + (n_1 + n_2 + \ldots - (k-1)n_k)b_{k-1} = Sy$$

and so on.

If most of the n's are equal, most of the coefficients will be zero and the normal equations will be easy to solve. Thus this device may be useful in situations in which the pattern is only slightly non-orthogonal. It is seldom used, however. See DeLury, the analysis of latin squares when some observations are missing, J.A.S.A. 1946.

On the whole, it is simpler not to press the regression pattern too hard, but rather to invoke the principle of least squares directly, in this case, by minimizing the s.s. $S(y_{i\alpha} - \gamma_i)^2$. This leads at once to the same normal equations.

Another formulation of the same question, which corresponds more closely with our objectives in the analysis, is

$$y_{i\alpha} = \mu + \gamma_i + \varepsilon_{i\alpha} \qquad \Sigma\gamma_i = 0 \ .$$

Now, the s.s. attributable to the γ's is the among-samples s.s. and the s.s. residuals in the within-samples s.s. We introduce a Lagrange multiplier λ and write

$$\psi = \frac{1}{2} S(y_{i\alpha} - \mu - \gamma_i)^2 + \lambda \Sigma \gamma_i .$$

The unrestricted minimum of ψ, treating λ as a variable yields the minimum of $S(y_{i\alpha}-\mu-\gamma_i)^2$ with $\Sigma\gamma_i = 0$.

$$\frac{\partial\psi}{\partial\mu} = -S(y_{i\alpha}-\mu-\gamma_i) = 0$$

$$\frac{\partial\psi}{\partial\gamma} = -S(y_{i\alpha}-\mu-\gamma_i)\frac{\partial\gamma_i}{\partial\gamma_j} + \lambda = 0$$

$$\frac{\partial\psi}{\partial\lambda} = \Sigma\gamma_i = 0$$

Introducing latin letters to denote estimators

1 $$Nm + \Sigma n_i c_i = Sy = g_0$$

2 $$n_j m + n_j c_j + \lambda = \sum_{\alpha=1}^{n_j} y_{j\alpha} = g_j \qquad j = 1,2,\ldots,k$$

$$\Sigma c_i = 0$$

Write $g_0, g_1, \ldots, g_k$ as the right side of these equations and solve.

$$m + c_j + \lambda/n_j = g_j/n_j \quad \text{adding,} \quad j = 1, \ldots, k$$

$$km + \lambda\Sigma_j\, 1/n_j = \Sigma_j\, g_j/n_j$$

also, adding equations 2 and subtracting from 1

$$k\lambda = \Sigma g_i - g_0$$

We have, then,

$$\lambda = \frac{1}{k}(\Sigma g_i - g_0)$$

$$\text{"}m\text{"} = \frac{1}{k}\Sigma\frac{g_j}{n_j} - \frac{\lambda}{k}\Sigma\frac{1}{n_j}$$

$$\text{"}c_j\text{"} = \frac{g_j}{n_j} - \frac{1}{k}\Sigma\frac{g_i}{n_i} - \lambda\left(\frac{1}{n_j} - \frac{1}{k}\Sigma\frac{1}{n_i}\right)$$

Substituting for the g's,

$$\lambda = 0$$

$$c_j = \bar{y}_j - \frac{1}{k} \sum_{i=1}^{k} \bar{y}_i$$

Now, putting $g_j = 1$, all other g's = 0, we get

$$c_{jj} = \frac{1}{n_j}\left(1 - \frac{2}{k}\right) + \frac{1}{k^2} \Sigma \frac{1}{n_i}$$

and putting $g_l = 1$, all others zero, we get

$$c_{jl} = -\frac{1}{k}\left(\frac{1}{n_j} + \frac{1}{n_l}\right) + \frac{1}{k^2} \Sigma \frac{1}{n_i}$$

We get, then,

$$Var(c_j - c_l) = \sigma^2(c_{jj} + c_{ll} - 2c_{jl}) = \sigma^2\left(\frac{1}{n_j} + \frac{1}{n_l}\right)$$

which we know to be correct because $c_j - c_l = \bar{y}_j - \bar{y}_l$.

The residual s.s. is

$$Sy^2 - mSy - \sum_j c_j \sum_\alpha y_{i\alpha} = Sy^2 - mSy - \sum_j \sum_\alpha y_{j\alpha}\,(\bar{y}_j - m)$$

$$= Sy^2 - \sum_j n_j \bar{y}_j^2 \text{ , the within-samples s.s.}$$

None of this is needed, of course, since we can manage the completely randomized pattern quite nicely without regression.

The two-way classification

Suppose we have a set of observations that may be classified, through the manner in which they were obtained, in two ways, rows and columns, with varying numbers of observations in each cell. They may be symbolized

$$y_{ij\alpha}, \quad i = 1, 2, \ldots r$$
$$j = 1, 2, \ldots c$$
$$\alpha = 1, 2, \ldots n_{ij}$$

This could come about, for example, as a $r \times c$ factorial arrangement, in a completely randomized experiment, or as a randomized block experiment, rows (or columns) representing replications.

We have, to start with, variation among cells and within cells, the latter reflecting error (perhaps sampling error) only. Only analysis of variance calculations are needed to get the corresponding s.s.

among cells	$rc - 1$	$\sum_i \sum_j n_{ij}(\overline{y}_{ij}-\overline{y})^2$
within cells	$N - rc$	$\sum_i \sum_j \sum_\alpha (y_{ij\alpha}-\overline{y}_{ij})^2$

If the n_{ij} are all equal, the experiment is wholly orthogonal and simple calculations yield independent sums of squares corresponding to

rows	$r - 1$
columns	$c - 1$
$r \times c$	$(r-1)(c-1)$
among cells	$rc - 1$

with further separation into particular contrasts. Presumably, when the n_{ij} are not all equal, we have the same objectives, but in general, regression methods are needed to attain them.

The first of these computations corresponds to the fitting of a regression

$$E\, y_{ij\alpha} = \mu + \varphi_{ij}, \quad \Sigma\Sigma\varphi_{ij} = 0.$$

There are rc disposable constants and the fit to the cell averages is perfect.

The second breakdown corresponds to the assumption $\varphi_{ij} = \rho_i + \gamma_j + \pi_{ij}$, where $\Sigma\rho_i = 0$, $\Sigma\gamma_j = 0$, $\sum_i \pi_{ij} = 0$, all j, $\sum_j \pi_{ij} = 0$, all i. Again there are rc constants and the fit to the cell averages is perfect. The s.s. attributable to regression is precisely the s.s. among cells.

We can set up the function to be minimized in the form

$$\psi = \frac{1}{2} S\left(y_{ij\alpha}-\mu-\rho_i-\gamma_j-\pi_{ij}\right)^2+\xi\Sigma\rho_i+\eta\Sigma\gamma_j+\sum_{j=1}^{c} \zeta_j \sum_{i=1}^{r} \pi_{ij}+\sum_{i=1}^{r-1} \tau_i \sum_{j=1}^{c} \pi_{ij}$$

$$\frac{\partial\psi}{\partial\mu} = -S\left(y_{ij\alpha}-\mu-\rho_i-\gamma_j-\pi_{ij}\right) = 0$$

$$\frac{\partial\psi}{\partial\rho_k} = -S\left(y_{ij\alpha}-\mu-\rho_i-\gamma_j-\pi_{ij}\right)\delta_{ik} + \xi = 0, \quad k = 1, 2, \ldots, r$$

$$\frac{\partial\psi}{\partial\gamma_l} = -S\left(y_{ij\alpha}-\mu-\rho_i-\gamma_j-\pi_{ij}\right)\delta_{jl} + \eta = 0, \quad l = 1, 2, \ldots, c$$

$$\frac{\partial\psi}{\partial\pi_{uv}} = -S\left(y_{ij\alpha}-\mu-\rho_i-\gamma_j-\pi_{ij}\right)\delta_{iu}\,\delta_{jv} + \zeta_v + \tau_u(1-\delta_{ur}) = 0$$

$$u = 1, 2, \ldots r, \; v = 1, 2, \ldots c$$

$$\frac{\partial\psi}{\partial\xi} = \Sigma\rho_i = 0, \text{ etc}$$

Put $N = \Sigma\Sigma n_{ij}$, $T_{ij} = \sum_{\alpha=1}^{n_{ij}} y_{ij\alpha}$, $T_{i\cdot} = \sum_{j=1}^{c} T_{ij}$,

$T_{\cdot j} = \sum_{i=1}^{r} T_{ij}$. Using m, r_i etc to denote estimates, the normal equations are

(1) $$Nm + \sum_i n_{i\cdot} r_i + \sum_j n_{\cdot j} c_j + \Sigma\Sigma n_{ij} p_{ij} = T$$

(2) $$n_{k\cdot} m + n_{k\cdot} r_k + \sum_j n_{kj} c_j + \sum_j n_{kj} p_{kj} + \xi = T_{k\cdot}$$

(3) $$n_{\cdot l} m + \sum_i n_{il} r_i + n_{\cdot l} c_l + \sum_i n_{il} p_{il} + \eta = T_{\cdot l}$$

(4) $$n_{uv} m + n_{uv} r_u + n_{uv} c_v + n_{uv} p_{uv} + \zeta_v + \tau_u(1-\delta_{ur}) = T_{uv}$$

Adding together the equations in (2) and subtracting from (1), we see $\xi = 0$. Similiarly all other Lagrange multipliers are zero. Then, set (4) contains all the other equations. Writing them as

$$m + r_u + c_v + p_{uv} = \frac{T_{uv}}{n_{uv}} = \bar{y}_{uv},$$

and summing over u and v, using $\Sigma r_u = 0$ etc,

$$rcm = \sum_u \sum_v \bar{y}_{uv}$$

Summing over v,

$$cm + cr_u = \sum_v \bar{y}_{uv}$$

$$r_u = \frac{1}{c} \sum_v y_{uv} - \frac{1}{rc} \Sigma\Sigma \bar{y}_{uv} .$$

Similiarly,

$$c_v = \frac{1}{r} \sum_u \bar{y}_{uv} - \frac{1}{rc} \Sigma\Sigma \bar{y}_{uv},$$

$$p_{uv} = \bar{y}_{uv} - \frac{1}{r} \sum_u \bar{y}_{uv} - \frac{1}{c} \sum_v \bar{y}_{uv} + \frac{1}{rc} \Sigma\Sigma \bar{y}_{uv} .$$

Observe that these estimates are all formed from unweighted averages of cell averages.

The residual s.s. is, of course, the within - cells s.s. This is easily checked.

$$\text{s.s. residual} = Sy^2 - mT - \Sigma r_i T_{i\cdot} - \Sigma c_j T_{\cdot j} - \Sigma\Sigma p_{ij} T_{ij} .$$

If, in the last sum, we put

$$p_{ij} = y_{ij} - r_i - c_j - m,$$

the s.s. residuals reduces to $Sy^2 - \Sigma\Sigma n_{ij} \bar{y}_{ij}^2$

In situations where the within-cells s.s. provides an appropriate estimate of error, the p_{uv} components can be tested for significance. If some of them are significantly different from zero, there would be no point in testing the r's and c's. On the other hand, if it turns out that there are no row × column interactions, we would be concerned with the r_i and c_j. However, r_i and c_j have been estimated assuming the presence of interactions and if there are none, we should be forming our estimates under the assumption

$$\varphi_{ij} = \rho_i + \gamma_j$$

This leads to a different set of normal equations, which we may get simply by deleting the equations of set (4) and dropping the p_{ij} in the rest.

$$Nm + \Sigma n_{i\cdot} r_i + \Sigma n_{\cdot j} c_j = T$$

$$n_{k\cdot} m + n_{k\cdot} r_k + \sum_j n_{kj} c_j = T_{k\cdot}, \quad k = 1, 2, \ldots r$$

$$n_{\cdot l} m + \sum_i n_{il} r_i + n_{\cdot l} c_l = T_{\cdot l}, \quad l = 1, 2, \ldots c$$

These equations do not, in general, yield usable algebraic solutions, so we are forced into numerical solutions. Suppose this is done, so that numerical values of m, r_i, c_j are at hand. We can then evaluate the s.s. attributable to regression, i.e. the r's and c's. It will have $(r-1) + (c-1) = r+c-2$ d.f.. This s.s. may be called the "rows and columns" s.s..

Actually, we can derive a formula for this s.s. which will require, of course, the numerical values of the regression constants. Eliminating m from the normal equations by pivotal condensation, we get

$$\sum_i n_{k\cdot}\left(\delta_{ik} - \frac{n_{i\cdot}}{N}\right) r_i + \sum_j \left(n_{jk} - \frac{n_{k\cdot}\, n_{\cdot j}}{N}\right) c_j = T_{k\cdot} - \frac{n_{k\cdot}}{N} T,$$

$$\sum_i \left(n_{il} - \frac{n_{i\cdot} n_{\cdot l}}{N}\right) r_i + \sum_j n_{\cdot l}\left(\delta_{jl} - \frac{n_{\cdot j}}{N}\right) c_j = T_{\cdot l} - \frac{n_{\cdot l}}{N} T\ .$$

Actually, all we need here are the right sides of these equations, to put in the formula for the s.s. for rows and columns.

$$\text{(rows and columns) s.s.} = \sum_{k=1}^{r} r_k \left(T_{k\cdot} - \frac{n_{k\cdot}}{N} T\right) + \sum_{l=1}^{c} c_l \left(T_{\cdot l} - \frac{n_{\cdot l}}{N} T\right)$$

$$= mT - \frac{T^2}{N} + \Sigma r_k\, T_{k\cdot} + \Sigma c_l\, T_{\cdot l}\ .$$

The residual s.s. of this fitting, regarded as fitted to the cell averages, must be the rows × columns interaction s.s.. Its value is $\Sigma\Sigma n_{ij}\bar{y}_{ij}^2 - mT - \Sigma r_k\, T_{k\cdot} - \Sigma c_l\, T_{\cdot l}$. The sum of these two s.s. is $\Sigma\Sigma n_{ij}\bar{y}_{ij}^2 - \frac{T^2}{N}$, the among-samples s.s. The regression partitions the among-samples into two independent s.s..

	d.f.	s.s.
rows and columns	$r+c-2$	$mT-T^2/N+\Sigma r_k T_{k\cdot}+\Sigma c_l T_{\cdot l}$
rows × columns	$(r-1)(c-1)$	by subtraction
among samples	$rc-1$	$\Sigma\Sigma n_{ij} \bar{y}_{ij}^2 - T^2/N$

Here again we could test the $r \times c$ interaction if that is appropriate, or, if the $r \times c$ is, by definition, an error term (experimental error), we would want to look more closely into components of the rows and columns s.s.

If we should want s.s. for rows and for columns, they are not difficult to get. To get a rows s.s., we fit with the row constants included, then fit with them omitted and take the difference between the residuals or between the s.s. attributable regression.

Fitting with rows constants omitted leaves us with only one criterion of classification and it seems obvious that the corresponding s.s. must be simply the among-columns s.s. i.e.

$$\Sigma \frac{T_{\cdot l}^2}{n_{\cdot l}} - \frac{T^2}{N} .$$

This is easily checked. The normal equations are

$$Nm + \Sigma n_{\cdot j} c_j = T$$

$$n_{\cdot l} m + n_{\cdot l} c_l = T_{\cdot l} \qquad l = 1, 2, \ldots c$$

Eliminating m,

$$n_{\cdot l} \Sigma \left(\delta_{jl} - \frac{n_{\cdot l}}{N} \right) c_j = T_{\cdot l} - \frac{n_{\cdot l}}{N} T$$

The s.s. columns is therefore

$$\Sigma c_l \left(T_{\cdot l} - \frac{n_{\cdot l}}{N} T \right) = \Sigma c_l T_{\cdot l} - \frac{T}{N} \Sigma n_{\cdot l} c_l$$

$$= \Sigma \left(\frac{T_{\cdot l}}{n_l} - m \right) T_{\cdot l} - \frac{T}{N} (T - Nm)$$

$$= \Sigma \frac{T^2_{\cdot l}}{n_l} - mT - \frac{T^2}{N} + mT$$

$$= \Sigma \frac{T^2_{\cdot l}}{n_l} - \frac{T^2}{N} \quad .$$ This s.s. will be labeled columns, ignoring rows.

Thus we get

	d.f.	s.s.
columns (ignoring rows)	$c-1$	$\Sigma \frac{T^2_{\cdot l}}{n_{\cdot l}} - \frac{T^2}{N}$
rows (eliminating columns)	$r-1$	by subtraction
rows and columns	$r+c-2$	$mT - \frac{T^2}{N} + \Sigma r_k T_{k\cdot} + \Sigma c_l T_{\cdot l}$

In the same way, we can compute

rows (ignoring columns)	$r-1$	
columns (eliminating rows)	$c-1$	by subtraction
rows and columns	$r+c-2$	

Often, though, these s.s. for rows and for columns will serve no useful purpose. The whole point of the calculation is to get the error s.s. and the estimates of the row and column constants.

This discussion covers the analysis of a randomized block design. If the n_{ij} are all equal,

$$m = \bar{y}$$

$$r_i = \bar{y}_{i\cdot} - \bar{y}$$

$$c_j = \bar{y}_{\cdot j} - \bar{y}$$

$$p_{ij} = \bar{y}_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}$$

Proportional Frequencies

A special case, which is unlikely to occur in practice, but which is of some interest, occurs when the n_{ij} are all of the form

$$n_{ij} = \lambda_i \mu_j,$$

$$n_{i\cdot} = \lambda_i \Sigma\mu_j = \lambda_i M$$

$$n_{\cdot j} = \mu_j \Sigma\lambda_i = \mu_j \Lambda$$

$$N = \Lambda M.$$

The normal equations for rows and columns become

$$\sum_i \lambda_k M\left(\delta_{ik} - \lambda_i \frac{M}{N}\right) r_i + \sum_j \left(\lambda_k \mu_j - \frac{\lambda_k M \mu_j \Lambda}{N}\right) c_j$$

$$= T_{k\cdot} - \frac{\lambda_k M}{N} T, \quad k = 1, \ldots r.$$

$$\sum_i \left(\lambda_i \mu_l - \frac{\lambda_i M \mu_l \Lambda}{N}\right) r_i + \sum_j \mu_l \Lambda\left(\delta_{jl} - \frac{\mu_j \Lambda}{N}\right) c_j$$

$$= T_{\cdot l} - \frac{\mu_l \Lambda}{N} T, \quad l = 1, \ldots c.$$

We note first that the matrix of these equations is of the form $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$. It follows that the r's and c's are uncorrelated.

The formula for the rows and columns s.s. is

$$\Sigma r_k\left(T_{k\cdot} - \frac{\lambda_k}{\Lambda} T\right) + \Sigma c_l\left(T_{\cdot l} - \frac{\mu_l}{M} T\right)$$

and, in view of the independence of the r's and the c's, the first term yields the rows s.s. and the second the columns s.s.

This, in itself, offers only modest simplification, but we shall show that it is not necessary to solve for the r's and c's to use this formula and, indeed, we can get simple algebraic solutions for the r's and c's.

The normal equations, specialized to this case, are

$$\Lambda M m + M\Sigma\lambda_i r_i + \Lambda\Sigma\mu_j c_j = T$$

$$M\lambda_k m + M\lambda_k r_k + \lambda_k\Sigma\mu_j c_j = T_{k\cdot} \qquad k = 1,2....r.$$

$$\Lambda\mu_l m + \mu_l\Sigma\lambda_i r_i + \Lambda\mu_l c_l = T_{\cdot l} \qquad l = 1,2....c.$$

Write the equations of the second set in the form

$$(1) \qquad m + r_k + \frac{1}{M}\Sigma\mu_j c_j = \frac{T_{k\cdot}}{M\lambda_k}$$

adding these equations yields

$$(2) \qquad rm + \frac{r}{M}\Sigma\mu_j c_j = \frac{1}{M}\Sigma\frac{T_{k\cdot}}{\lambda_k} \quad \text{or}$$

$$(3) \qquad m + \frac{1}{M}\Sigma\mu_j c_j = \frac{1}{rM}\Sigma\frac{T_{k\cdot}}{\lambda_k}$$

Subtracting (3) from (1)

$$r_k = \frac{T_{k\cdot}}{M\lambda_k} - \frac{1}{rM}\Sigma\frac{T_{i\cdot}}{\lambda_i}$$

Similarly

$$c_l = \frac{T_{\cdot l}}{\Lambda\mu_l} - \frac{1}{c\Lambda}\Sigma\frac{T_{\cdot j}}{\mu_j}$$

The columns s.s. now becomes

$$\Sigma c_l \left(T_{\cdot l} - \frac{\mu_l}{M} T\right) = \Sigma c_l T_{\cdot l} - \frac{T}{M} \Sigma \mu_l c_l$$

$$= \sum_l \left(\frac{T_{\cdot l}}{\Lambda \mu_l} - \frac{1}{c\Lambda} \sum_j \frac{T_{\cdot j}}{\mu_j}\right) T_{\cdot l} - \frac{T}{M} \sum_l \left(\frac{T_{\cdot l}}{\Lambda \mu_l} - \frac{1}{c\Lambda} \sum_j \frac{T_{\cdot j}}{\mu_j}\right) \mu_l$$

$$= \Sigma \frac{T_{\cdot l}^2}{\Lambda \mu_l} - \frac{T^2}{\Lambda M}$$

Hence we have a formula for s.s. columns and, indeed, it is the same formula we have always used. A similar formula provides the rows s.s., the rows × columns s.s. comes by a subtraction. The rows, columns and $r \times c$ s.s. are independent of each other. The row constants are, in general, correlated as are the column constants.

The analysis of covariance

The one instance of covariance encountered this far is rather special in that it arose in a completely randomized pattern, that each sample provided a number of points, which could be studied for straightness and parallelism. The extension to three or more samples is immediate and obvious and need not concern us here. What must concern us is the fact that in many, perhaps most, experiments, there is only one (xy) pair in each sample, and the direct approach to checking straightness and parallelism is not available. More on this later.

It will be sufficient to discuss a particular example, say a randomized block pattern, yielding responses y_{ij}, which would be

embedded in regression theory with the model

$$y_{ij} = \mu + \rho_i + \gamma_j + \varepsilon_{ij} \ , \ \Sigma\rho_i = \Sigma\gamma_j = 0 \ ,$$

and yielding estimates

$$m = \bar{y} \ , \ r_i = \bar{y}_{i\cdot} - \bar{y} \ , \ c_j = \bar{y}_{\cdot j} - \bar{y} \ .$$

If, now, we have a concomitant variable x_{ij}, we can put it into the model, assuming a linear dependence of y on x (i.e. straightness and parallelism), by adding a term βx_{ij}. We have, then,

$$y_{ij} = \mu + \rho_i + \gamma_j + \beta x_{ij} + \varepsilon_{ij} \ , \ \Sigma\rho_i = \Sigma\gamma_j = 0.$$

We can now minimize the s.s. residuals. This is much simplified by introducing the following linear transformation of the independent variables and hence the regression coefficients. (Cochran, Biometrics, Vol. 12, no. 1, 1956)

$$x'_{ij} = x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}$$

$$\mu' = \mu + \beta\bar{x}$$

$$\rho'_i = \rho_i + \beta(\bar{x}_{i\cdot} - \bar{x})$$

$$\gamma'_j = \gamma_j + \beta(\bar{x}_{\cdot j} - \bar{x})$$

We observe that $\sum_i x'_{ij} = 0$, all j, $\sum_j x'_{ij} = 0$, all i, $\Sigma\rho'_i = 0$, $\Sigma\gamma'_j = 0$.

We therefore set up the function to be minimized,

$$\psi = \frac{1}{2} S(y_{ij} - \mu' - \rho'_i - \gamma'_j - \beta x'_{ij})^2 + \lambda\Sigma\rho'_i + \xi\Sigma\gamma'_j$$

$$= \frac{1}{2} S(y_{ij} - \mu' - \rho'_i - \gamma'_j)^2 - \beta S y_{ij} x'_{ij} + \frac{1}{2}\beta^2 S x'^2_{ij} + \lambda\Sigma\rho'_i + \xi\Sigma\gamma'_j$$

$\frac{\partial\psi}{\partial\beta} = -S y_{ij} x'_{ij} + \beta S x'^2_{ij} = 0$. Thus,

$$b = \frac{Sy_{ij}x'_{ij}}{Sx'^2_{ij}}$$

The estimate of b is recognized as the regression calculated from sums of squares and products taken from the error row of the analysis of variance and covariance table, E_{xy} and E_{xx}, say. Note that $Sy_{ij}x'_{ij} = S(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y})(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})$. The rest of the normal equations are the usual ones

$$rcm' = T$$

$$cm' + cr'_i = T_{i.}$$

$$rm' + rc'_j = T_{.j}$$

$$E_{xx}b = E_{xy}$$

$m' = \bar{y}$, $r'_i = \bar{y}_{i.} - \bar{y}$, $c'_j = \bar{y}_{.j} - \bar{y}$, and the residual s.s. is

$$Sy^2 - m'T - \Sigma r'_i T_{i.} - \Sigma c'_j T_{.j} - bE_{xy}$$

$$= S(y - \bar{y})^2 - \Sigma\left(\frac{T^2_{i.}}{c} - \frac{T^2}{rc}\right) - \Sigma\left(\frac{T^2_{.j}}{r} - \frac{T^2}{rc}\right) - \frac{E^2_{xy}}{E_{xx}}$$

$$= E_{yy} - E^2_{xy}/E_{xx}.$$

This is the reduced error s.s., with $(r-1)(c-1) - 1$ d.f.

To get the s.s. attributable to rows (say), we fit omitting the row constants, get the s.s. residuals, from which we subtract the s.s. residuals when the row constants are included.

The analysis of variance table, with rows omitted, reads

	d.f.
columns	$c-1$
residuals	$c(r-1)$
total	$rc-1$

and the entries in the residual row must be rows + error, which we may call S_{xx}, S_{xy}, S_{yy}. The same computation as before then leads to a residual s.s.

$$S_{yy} - S_{xy}^2/S_{xx} \text{ with } c(r-1) - 1 \text{ d.f.}$$

The s.s. for rows (eliminating columns) is then

$$(S_{yy} - S_{xy}^2/S_{xx}) - (E_{yy} - E_{xy}^2/E_{xx})$$

$$= R_{yy} - \frac{E_{xy}^2}{E_{xx}} + \frac{S_{xy}^2}{S_{xx}} \text{ with } c(r-1) - 1 - [(r-1)(c-1) - 1] = r-1 \text{ d.f.}$$

This is the adjusted rows s.s., with $r-1$ d.f.

	d.f.	(xx)	(xy)	(yy)
rows	$r-1$	R_{xx}	R_{xy}	R_{yy}
columns	$c-1$	C_{xx}	C_{xy}	C_{yy}
error	$(r-1)(c-1)$	E_{xx}	E_{xy}	E_{yy}
rows + error	$c(r-1)$	S_{xx}	S_{xy}	S_{yy}

	attributable to regression on x		deviations from regression on x	
	d.f.	s.s.	d.f.	s.s.
error	1	E_{xy}^2/E_{xx}	$(r-1)(c-1) - 1$	$E_{yy} - E_{xy}^2/E_{xx} = E$
rows + error	1	S_{xy}^2/S_{xx}	$c(r-1) - 1$	$S_{yy} - S_{xy}^2/S_{xx} = G$
	adjusted row s.s.		$r-1$	$G - E$

Returning to the estimates of the original parameters

$$m = \bar{y} - b\bar{x},\ r_i = \bar{y}_{i\cdot} - \bar{y} - b(\bar{x}_{i\cdot} - \bar{x}),\ c_j = \bar{y}_{\cdot j} - \bar{y} - b(\bar{x}_{\cdot j} - \bar{x}).$$

If we want to exhibit adjusted row and column averages, we simply drop $\bar{y}$ from the r_i and c_j.

The procedure for obtaining the adjusted rows s.s. is obviously applicable to any component. If $u = \Sigma a_{ij}y_{ij}$, $v = \Sigma a_{ij}x_{ij}$, the entries in the analysis of variance and covariance table are v^2 uv u^2. The adjusted value of u^2 is obtained by calculating

$$S_{xx} = E_{xx} + v^2,\ S_{xy} = E_{xy} + uv,\ S_{yy} = E_{yy} + u^2,$$

and the adjusted value of the component is

$$(S_{yy} - S_{xy}^2/S_{xx}) - (E_{yy} - E_{xy}/E_{xx}^2) \text{ with 1 d.f.}$$

These procedures evidently apply whatever the design of the experiment, except that we have still to discuss the split-plot kind of experiment, where we have more than one error term.

The assumptions of straightness and parallelism, implicit in this development, may occasionally be called into question. When the concomitant variable is carried largely to control error, entering the experiment randomly, we usually have enough control of the experimental material to keep its range small so that any lack of straightness or of parallelism has little chance of affecting the results. On the other hand, when the treatments may affect the concomitant variable, its range may be large and may vary from treatment to treatment. Adjustment of the treatment averages may reflect an extrapolation into a range that may, in extreme cases, be non-existent, and in any event, the assumed straightness and parallelism plays a large role and could, if unwarranted, lead to gross mistakes. Caution is required here.

A check for constancy of the regression coefficient

If the regression coefficient is to vary over an experiment, it would presumably do so because it varies with the treatments. Our model would then take the form, for a randomized block experiment,

$$y_{ij} = \mu + \rho_i + \gamma_j + \beta_j x_{ij} + \varepsilon_{ij}.$$

The error term in the analysis of variance, where rows and columns have been removed, is $S(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y})^2$. Let us calculate this s.s.

$$y_{ij} = \mu + \rho_i + \gamma_j + \beta_j x_{ij} + \varepsilon$$

$$\bar{y}_{i.} = \mu + \rho_i + 0 + \tfrac{1}{c}\Sigma\beta_j x_{ij} + \varepsilon$$

$$\bar{y}_{.j} = \mu + 0 + \gamma_j + \beta_j \bar{x}_{.j} + \varepsilon$$

$$\bar{y} = \mu + 0 + 0 + \tfrac{1}{c}\Sigma\beta_j \bar{x}_{ij} + \varepsilon.$$

$$y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y} = \beta_j(x_{ij} - \bar{x}_{.j}) + \frac{1}{c}\sum_{k=1}^{c} \beta_k(x_{ik} - \bar{x}_{.k}) + \varepsilon.$$

Thus, we want to minimize, by appropriate choice of the β's, the s.s.

$$S\left\{y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y} - \sum_{k=1}^{c} (\delta_{kj} - \tfrac{1}{c})\beta_k(x_{ik} - \bar{x}_{.k})\right\}^2$$

Differentiating with respect to β_p:

$$S\left\{y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y} - \Sigma(\delta_{kj} - \tfrac{1}{c})\beta_k(x_{ik} - \bar{x}_{.k})\right\}(\delta_{kj} - \tfrac{1}{c})(x_{ik} - \bar{x}_{.k}) = 0$$

Replacing β's by b's and rearranging:

$$S \sum_k b_k(x_{ik} - \bar{x}_{.k})(x_{ip} - \bar{x}_{.p})(\delta_{kj} - \tfrac{1}{c})(\delta_{pj} - \tfrac{1}{c}) =$$

$$S(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y})(x_{ip} - \bar{x}_{.p})(\delta_{pj} - \tfrac{1}{c}) \quad , \; p = 1,2,\ldots,c \; .$$

These equations may be simplified somewhat. Put

$$a_{pk} = \sum_i (x_{ik} - \bar{x}_{\cdot k})(x_{ip} - \bar{x}_{\cdot p})$$

$$\lambda_{pk} = \sum_j (\delta_{kj} - \frac{1}{c})(\delta_{pj} - \frac{1}{c})$$

$$= -\frac{1}{c}\ , \quad k \neq p$$

$$= 1 - \frac{1}{c}\ , \quad k = p$$

$$= \delta_{pk} - \frac{1}{c}\ .$$

Also, write $g_{pj} = \sum_i (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y})(x_{ip} - \bar{x}_{\cdot p})$. Then, the normal equations may be written

$$\sum_{k=1}^{c} \lambda_{pk} a_{pk} b_k = \sum_{j=1}^{c} g_{pj}(\delta_{pj} - \frac{1}{c}) = G_p \text{ (say)} \quad p = 1,2,\ldots,c\ .$$

We may note that $\sum_{j=1}^{c} g_{pj} = 0$, all p. Hence

$$G_p = \sum_{j=1}^{c} g_{pj}\delta_{pj} - \frac{1}{c} \sum_j g_{pj}$$

$$= g_{pp}$$

$$= \sum_i (y_{ip} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y})(x_{ip} - \bar{x}_{\cdot p})$$

$$= \sum_i (y_{ip} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y})x_{ip}.$$

Rules for calculating the a_{pk} and G_p are easily developed.

Having solved for the b's, the s.s. residuals is

$E_{yy} - \sum_{p=1}^{c} b_p G_p$, with $(r-1)(c-1) - c$ d.f. Call this F. Recall that

$E = E_{yy} - E_{xy}^2/E_{xx}$ with $(r-1)(c-1) - 1$ d.f. Then $E - F$, with $c-1$ d.f.,

is the reduction in s.s. when we go from a single coefficient β to the coefficients $\beta_1, \beta_2, \ldots, \beta_c$. If $\beta_1 = \beta_2 = \ldots = \beta_c$, $E - F$ is $\sigma^2 X^2_{(c-1)}$, independent of F, and

$$\frac{(E - F)/(c-1)}{F/[(r-1)(c-1) - c]} \quad \text{is} \quad F_{(c-1,\ (r-1)(c-1) - c)}.$$

If we have assurance that treatments do not affect the regression, we could, no doubt, raise the question of linearity of the regression, by assuming a model

$$y_{ij} = \mu + \rho_i + \gamma_j + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \varepsilon_{ij}$$

and testing the reduction in residual s.s. resulting from the introduction of the quadratic term. This would require a procedure for an analysis of covariance with more than one concomitant variable.

Covariance with two or more concomitant variables

Surely no further theoretical development is needed here. The directions must be the same as in the case of a single covariate: Fit a regression using sums of squares and products from the error row. With a single covariate, this is accomplished by means of formulae derived by the algebraic solutions of the normal equations. With two or more covariates, these formulae are complicated and it is better to set up the normal equations, using sums of squares and products from the error row and solve them numerically. An example should suffice. Suppose we have a randomized block experiment, with a response y and two concomitant variables, u and v.

We first compute the analysis of variance and covariance table

	d.f.	(uu)	(uv)	(vv)	(uy)	(vy)	(yy)
replications	$r-1$	R_{uu}	R_{uv}	R_{vv}	R_{uy}	R_{vy}	R_{yy}
treatments	$c-1$	C_{uu}	C_{uv}	C_{vv}	C_{uy}	C_{vy}	C_{yy}
error	$(r-1)(c-1)$	E_{uu}	E_{uv}	E_{vv}	E_{uy}	E_{vy}	E_{yy}

We envisage a model

$$y_{ij} = \mu + \rho_i + \gamma_j + \beta_1 u_{ij} + \beta_2 v_{ij} + \varepsilon_{ij}.$$

The sums of squares and products in the error row involve only b_1 and b_2, estimates of β_1 and β_2. b_1 and b_2 satisfy the normal equations

$$E_{uu}b_1 + E_{uv}b_2 = E_{uy}$$

$$E_{uv}b_1 + E_{vv}b_2 = E_{vy}$$

We solve these equations and compute the s.s. residuals,

$E_{yy} - b_1E_{uy} - b_2E_{vy}$. This is the reduced error s.s. with $(r-1)(c-1) - 2$ d.f.

To get the adjusted s.s. for treatments, we compute $S_{uu} = C_{uu} + E_{uu}$ etc. and set up the equations

$$S_{uu}c_1 + S_{uv}c_2 = S_{uy}$$

$$S_{uv}c_1 + S_{vv}c_2 = S_{vy},$$

solve for c_1, c_2 and compute s.s. residuals from

$S_{yy} - c_1S_{uy} - c_2S_{vy}$, with $(r-1)(c-1) + (c-1) - 2$ d.f. The difference between these two residuals, with $c-1$ d.f., is the adjusted s.s. for treatments. The adjusted treatment means are obviously

$$\bar{y}_{i\cdot}' = \bar{y}_{i\cdot} - b_1(\bar{u}_{i\cdot} - \bar{u}) - b_2(\bar{v}_{i\cdot} - \bar{v}).$$

References for the analysis of covariance

Biometrics, Vol. 13, no. 3, Sept. 1957 is wholly devoted to covariance. Cochran's expository paper, in this issue, is the best account of covariance I have seen.

Truett and Smith, Adjustment by Covariance and consequent tests of significance in Split Plot Experiments, Biometrics, Vol. 12, no. 1, 1956, is worth reading.

Wilks, The analysis of variance and covariance in non-orthogonal data, Metron 13, No. 2, 1938.

Missing Values

When an orthogonal design is rendered non-orthogonal through the loss of one or more observations, we can, of course, conduct the analysis by means of regression theory, a vastly more elaborate and complicated job. We may ask whether regression theory may be adjusted (as it is in covariance) so that the added computation is not disproportionate to the amount of orthogonality lost, in particular, whether we can avoid the solving of a large set of normal equations. This is the "missing value" problem. It was first raised by Fisher, I think. The approach used here, formally embedded in regression theory, is closer to Yates than to Fisher.

This discussion will be directed to a randomized block experiment, but it applies equally to any design.

We postulate

$$y_{ij} = \mu + \rho_i + \gamma_j + \varepsilon_{ij}, \quad \Sigma\rho_i = \Sigma\gamma_j = 0.$$

For the intact design, a fitting of the regression $Y_{ij} = m + r_i + c_j$, $\Sigma r_i = \Sigma c_j = 0$, yields $m = \bar{y}$, $r_i = \bar{y}_{i\cdot} - \bar{y}$, $c_j = \bar{y}_{\cdot j} - \bar{y}$,

$$\text{s.s. residuals (errors)} = Sy^2 - mT - \Sigma r_i T_{i\cdot} - \Sigma c_j T_{\cdot j}$$
$$= S(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y})^2, \text{ with } (r-1)(c-1)\text{df}.$$

If some observations are missing, our objectives are the same, but the normal equations would not yield simple algebraic solutions and would require numerical solutions. Suppose this is done, so that values of m, r_i, c_j are at hand. We can then use the regression $Y_{ij} = m + r_i + c_j$ to calculate a Y-value for each of the missing observations. Suppose

this is done and let us think of fitting a new regression $Y'_{ij} = m' + r'_i + c'_j$ to the completed table of observations. This amounts to seeking Y' values such that

$$S(y_{ij} - Y'_{ij})^2 + \Sigma(Y_{ij} - Y'_{ij})^2 \quad \text{is minimized.}$$

S here stands for summation over the actual observations and Σ is over the filled in values.

Clearly, this s.s. is minimized by $Y' = Y$, so we get the same regression as before. Hence $m' = m$, $r'_i = r_i$, $c'_j = c_j$, and the sum of squares is left unchanged and therefore correct. The fitting to the completed table can be carried out according to the solution of the intact design, where the symbols m, $r_i = \bar{y}_{i\cdot} - \bar{y}$, $c_j = \bar{y}_{\cdot j} - \bar{y}$ may involve the substituted Y-values. Hence, if a way can be found to obtain these Y-values without actually fitting the first regression, methods appropriate to the intact design i.e. analysis of variance computations may be used.

Suppose that a single observation is missing in row l and column m. Write the symbol y_{lm} in this cell and proceed with the fitting of

$$\begin{aligned} Y_{ij} &= m + r_i + c_j \\ &= \frac{T}{rc} + \left(\frac{T_{i\cdot}}{c} - \frac{T}{rc}\right) + \left(\frac{T_{\cdot j}}{r} - \frac{T}{rc}\right). \end{aligned}$$

Now some of the symbols T, $T_{i\cdot}$, $T_{\cdot j}$ involve the symbol y_{lm}. The fitting is to be such that $Y_{lm} = y_{lm}$. We have, then,

$$\begin{aligned} y_{lm} &= \frac{T_{l\cdot}}{c} + \frac{T_{\cdot m}}{r} - \frac{T}{rc} \\ &= \frac{T'_{l\cdot} + y_{lm}}{c} + \frac{T'_{\cdot m} + y_{lm}}{r} - \frac{T' + y_{lm}}{rc} \end{aligned}$$

where T''s represent sums over the actual observations. Solving for

y_{lm} yields

$$y_{lm} = \frac{rT'_{l\cdot} + cT'_{\cdot m} - T'}{(r-1)(c-1)} ,$$

a missing-value formula for the randomized block design. Each design will have its own missing-value formula, derived as above or in a similar fashion. Always the number provided by the formula minimizes the error term, in fact, leaves it unchanged and entirely correct. (It is obvious that, in a completely randomized design, the numbers substituted to regain orthogonality, will be the sample averages.)

These facts can be used in several ways to cope with missing values. We could in any design, substitute symbols for the missing values, carry out the appropriate analysis of variance and choose the values of the symbols that minimize the error s.s., which of course is a quadratic form in the symbols.

If we have a convenient algebraic expression for the error s.s., in the randomized block design it is $S(y_{ij} - \overline{y}_{i\cdot} - \overline{y}_{\cdot j} + \overline{y})^2$, we can designate y_{lm} to represent the missing observation and choose its value to minimize the error s.s. This leads to the same missing-value formula.

If more than one missing value occur, we can derive a number of equations, equal in number to the number of missing values, to be solved for the numbers to be substituted for the missing values. Alternately, we may substitute for all but one of the missing values some number presumably $\overline{y}$ and use the missing value formula to fill in the cell left empty. Then, remove one of the $\overline{y}$'s and replace it with the value from the formula, and so on. This iterative procedure

usually converges rapidly.

After supplying the missing values, we make the usual computations

	d.f.	s.s.
replications	$r-1$	$\Sigma \frac{T_{i\cdot}^2}{c} - \frac{T^2}{rc}$
treatments	$t-1$	$\Sigma \frac{T_{\cdot j}^2}{r} - \frac{T^2}{rc}$
$r \times t$ (error)	$(r-1)(t-1)-n$	by subtraction
total	$rt-n-1$	$\Sigma y_{ij}^2 - \frac{T^2}{rc}$

n stands for the number of missing values. The totals used in the computations include the numbers supplied for the missing values.

As we have seen, the error s.s. so calculated is correct. All other s.s. are increased above their correct values, because adding numbers to the original observations can only increase the total s.s. but none of the increase appears in the error s.s. One can study ways of correcting these s.s., but usually we have no direct interest in them anyway. The whole missing-value precedure is really just a simple way of getting the correct error s.s.

The standard error of any contrasts we may wish to study, if they involve the missing values, are easily adjusted for this fact.

Missing values by covariance

Again, as an example, let us discuss the randomized block design, with one observation missing in cell (l,m). Put $y_{lm} = 0$, and introduce a concomitant variable x_{ij}, defined to be zero everywhere except for $x_{lm} = -1$. Then, fit the regression

$$Y_{ij} = m + r_i + c_j + bx_{ij},\ \Sigma r_i = \ c_j = 0.$$

Thus we minimize, subject to the restraints,

$S(y_{ij} - m - r_i - c_j)^2 + (0 - m - r_l - c_m + b)^2$; where S represents summation over the actual observations and is the s.s. minimized by fitting to the actual observations. The whole s.s. is minimized, and indeed attains the same minimum value, by choosing $b = m + r_l + c_m$. This is the same argument used before. b is seen to give the same missing value as was reached earlier. One can say that this device transfers non-orthogonality in y to non-orthogonality in x and there are no missing values.

What we do see here that is new is that the whole exercise can be put into the form of a covariance calculation and the correct s.s. for the reduced and adjusted s.s. may be found in this way.

The covariance table takes the form

	d.f.	(xx)	(xy)	(yy)
rows	$r-1$	$\frac{1}{c}(1 - \frac{1}{r}) = R_{xx}$	$-\frac{1}{c}(T_{l\cdot} - \frac{T}{r}) = R_{xy}$	R_{yy}
columns	$c-1$	$\frac{1}{r}(1 - \frac{1}{c}) = C_{xx}$	$-\frac{1}{r}(T_{\cdot m} - \frac{T}{c}) = C_{xy}$	C_{yy}
$r \times c$	$(r-1)(c-1)$	$(1 - \frac{1}{r} - \frac{1}{c} + \frac{1}{rc}) = E_{xx}$	$\frac{T_{l\cdot}}{c} + \frac{T_{\cdot m}}{r} - \frac{T}{rc} = E_{xy}$	E_{yy}
total	$rc-1$	$1 - \frac{1}{rc} = T_{xy}$	$\frac{T}{rc}$	Y_{yy}

We get $b = E_{xy}/E_{xx} = \dfrac{\frac{T_{l\cdot}}{c} + \frac{T_{\cdot m}}{r} - \frac{T}{rc}}{1 - \frac{1}{r} - \frac{1}{c} + \frac{1}{rc}}$, which is the missing-value formula reached earlier, because the T's, here, including as they do the "observation" $y_{lm} = 0$, are the T's calculated before. The residual s.s., i.e. the correct error s.s., is

$$E_{yy} - bE_{xy} = E_{yy} - E_{xy}^2/E_{xx} \text{ with } (r-1)(c-1) - 1 \text{ d.f.}$$

If we want the correct rows s.s., we get it by the standard covariance procedure, setting up (rows + error) and so on.

This approach is obviously available whatever the design of the experiment.

Split Plots

The discussion up to this point, according to which the analysis of experiments is embedded in regression theory, assumes the presence of only one error term, obtained by minimizing the residual s.s. When we have more than one error term, reflecting comparisons of different precisions, some extension is needed. To have something definite to discuss, think of a factorial experiment in which each level of the first factor requires a different subject, while all levels of the second factor may be applied to the same subject. Suppose the experiment is carried out in a number of replications. Let y_{ijk} be the observation corresponding to level i of the first factor, level j of the second factor, in replication k. Then, i and k run over subjects and j operates within subjects. We may now set up a model

$$y_{ijk} = \mu + \delta_i + \gamma_j + \pi_{ij} + \rho_k + \eta_{ik} + \varepsilon_{ijk},$$

with $\Sigma\delta = \Sigma\gamma = \Sigma\pi_{ij} = \Sigma\rho = 0$, η_{ik} represents a subject to subject component of error and ε_{ijk} a within-subject error. We assume ε: $N(0,\sigma^2)$, η : $N(0,\sigma_\eta^2)$, ε, η independent.

Recalling an earlier result, within-subject comparisons will have error variance σ^2 and between-subjects comparisons $\sigma^2 + q\sigma_\eta^2$, $(j = 1,\ldots,q)$.

Separating out the total variation into the within and between subjects components,

$$S(y_{ijk} - \bar{\bar{y}})^2 = S(y_{ijk} - \bar{y}_{i\cdot k})^2 + q \sum_i \sum_k (\bar{y}_{i\cdot k} - \bar{\bar{y}})^2$$

Now, $\bar{y}_{i\cdot k} = \mu + \delta_i + \rho_k + \eta_{ik} + \bar{\varepsilon}_{i\cdot k}$ and

$y_{ijk} - \bar{y}_{i\cdot k} = \gamma_j + \pi_{ij} + \varepsilon_{ijk} - \bar{\varepsilon}_{i\cdot k}$.

The weighted s.s. of residuals, with weights inversely proportional to the variances, is

$$W\, S(y_{ijk} - \bar{y}_{i\cdot k} - \gamma_i - \pi_{ij})^2 + W' \sum_i \sum_k \{\sqrt{q}(\bar{y}_{i\cdot k} - \mu - \delta_i - \rho_k)\}^2,$$

with $W = 1/\sigma^2$, $W' = 1/(\sigma^2 + q\sigma_\eta^2)$.

The normal equations obtained by minimizing this s.s., for the intact design, separate into two independent sets and yield the standard split-plot analysis, without requiring the values of W and W'. When some non-orthogonality is present, e.g. when a covariate is present, the normal equations do not separate and require the values of W and W', which must be estimated. We are thus led to complexity in the normal equations and to difficult distribution problems due to the estimation of W and W'.

The s.s. to be minimized, arising from

$$y_{ijk} = \mu + \delta_i + \gamma_j + \pi_{ij} + \rho_k + \beta x_{ijk} + \varepsilon_{ijk} \text{ is}$$

$$W\, S(y_{ijk} - \bar{y}_{i\cdot k} - \gamma_j - \pi_{ij} - \beta(x_{ijk} - \bar{x}_{i\cdot k}))^2$$

$$+ W' \sum_i \sum_k \{\sqrt{q}(\bar{y}_{i\cdot k} - \mu - \delta_i - \rho_k - \beta\bar{x}_{i\cdot k})\}^2.$$

Probably the simplest practical advice here is to carry out the covariance analysis separately in the main plot and in the sub-plot analysis. This amounts to minimizing individually the two s.s., leading to different estimates of β in the two parts, presumably each with lower efficiency than could be reached if they could somehow be combined. This is the only penalty paid for separate analyses. All the distribution theory remains valid.

The missing value problem may be regarded as a special instance of covariance and the same considerations apply.

The Rejection of Observations

The decision to reject observations should never be reached lightly. The decision to reject is a decision that the error system is out of control and we lose the essential basis for reaching assured conclusions. In a way, the concern is less about the observations we remove than with the ones we retain. How trustworthy are they if the error system is not to be trusted? No matter what the grounds for rejection (except when reasons for rejection external to the sample are at hand) or the procedure we may use to justify the rejection, it remains true that we reject them because we do not like them and retain them because we do like them; a fragile base from which to claim to prove something. The occurrence of observations we do not like is the commonest feature of all experimental and other statistical inquiries. There has been an inordinate amount of writing on the subject, going back at least 200 years. Many rules and procedures have been put forward for rejecting observations and for protecting ourselves, wholly or in part, from their contributions. Probably it is best not to use any rejection rule and such devices as "robust" methods and non-parametric methods should be used with judgement and caution. Not simply because the data are ragged and untrustworthy.

Incomplete Blocks

The notion of an incomplete block, i.e. the block is too small to hold a full replication, has already been encountered in factorial

experiments. The devices used there depend on the fact that, in a factorial experiment, especially a large one, some of the contrasts can be judged to have no real importance and can be sacrificed through being confounded with blocks. This device is no longer useful when all contrasts have the same importance, unless one is willing to employ balanced confounding. This discussion, then, has to do with balanced incomplete blocks. The condition of balance may be eased somewhat, leading to partially-balanced incomplete blocks, but this question will not be pursued here.

The objective in blocking is, of course, to ensure that every contrast we study is made within as uniform a set of conditions as possible. When the blocks are complete, every contrast is made within each block. This, evidently, is the condition that must be relaxed when the blocks are too small to contain all the contrasts. We must settle for arrangements in which every contrast is made within some blocks. If the arrangement is to be balanced, all contrasts must be perceptible within the same number of blocks.

A Historical Example

Seven strains of tobacco mosaic virus were to be compared by applying them to the leaves of young tobacco plants. It was known that individual plants differed substantially from one another in sensitivity to the virus and further that there were important differences among the leaves of each plant, in as much as each leaf was younger than the one immediately below it.

In these circumstances, we would naturally think of a latin square arrangement, with plants corresponding to columns and leaf

position to rows. For such an arrangement, we would require plants with 7 leaves, but it was judged that the largest number feasible was 3. Nevertheless, we still need what the latin square would provide, namely, contrasts made within plants and within leaf positions.

We might think of taking 3 rows out of a 7×7 latin square, but this does not necessarily yield a usable arrangement. If, for example, we take the first 3 rows of the latin square below,

$$
\begin{array}{ccccccc}
A & B & C & D & E & F & G \\
B & C & D & E & F & G & A \\
C & D & E & F & G & A & B \\
D & E & F & G & A & B & C \\
E & F & G & A & B & C & D \\
F & G & A & B & C & D & E \\
G & A & B & C & D & E & F
\end{array}
$$

thus making up seven groups of three, each to be applied to a single plant, we find that treatment A, for example, appears on the same plant once with C, F, twice with B, G and not at all with D, E. This arrangement is entirely unsuitable, On the other hand, if we select rows 1,2,4, we find an arrangement in which each treatment appears on the same plant once and only once with every other treatment. The condition of balance is met. The groups of three constitute incomplete blocks. This arrangement has another feature, usually not sought, that the observations are grouped into replications (top leaves, second leaves, third leaves). Such special incomplete block arrangements are called incomplete latin squares or Youden squares, after W.J. Youden who first published an account of such arrangements.

We may remark that the complementary set of rows, 3,5,6,7 furnish an arrangement in which each treatment appears in the same group with every other treatment in two groups and would provide a suitable allocation if plants with four leaves were to be used. Again, if we were using plants with six leaves, deleting any row from the 7×7 square would provide a grouping in which each treatment appears in a group 6 times with every other treatment.

If our plants had only two leaves, we would require 21 plants to attain balance and if we wanted the grouping into complete replications we would need 42 plants.

Balanced Incomplete Blocks

We shall speak of an experiment in which v varieties are tested, in b blocks, each of which contains k different varieties, each variety tested r times and each variety appearing in a block with every other variety λ times ($k < v$, λ: the number of times each pair of varieties occurs together in a block, i.e. the number of blocks containing each pair of varieties).

Clearly, these parameters cannot be assigned numerical values arbitrarily.

1. The total number of observations is

$$rv = bk = N \text{ (say)}$$

2. The total number of pairs (within blocks) is

$$\lambda \frac{v(v-1)}{2} = b \frac{k(k-1)}{2} . \quad \text{Hence,}$$

$$\lambda = \frac{b\ k(k-1)}{v(v-1)} = r\frac{(k-1)}{(v-1)} \quad \text{by} \quad (1).$$

These conditions are necessary, but are by no means sufficient, to ensure a suitable design. That some design exists is clear. We need only make up all k-subsets of v, $\binom{v}{k}$ in number, and assign each to a block. This would yield a design with $r = \binom{v-1}{k-1}$, $b = \binom{v}{k}$. Usually these numbers would be too large to be suitable. The enumerative problem, then, is to find arrangements that meet the conditions and that are not too large.

The assumptions made earlier for the randomized complete block are suitable here also. We envisage differences among blocks, among varieties and error.

$$y_{ij} = \mu + \beta_i + \varphi_j + \varepsilon_{ij}, \quad \sum_{i=1}^{b} \beta_i = \sum_{j=1}^{v} \varphi_j = 0.$$

The only way in which the treatment of this assumed structure differs from that in the complete block is the set of values taken by $j = j(i)$ varies with i, i.e. from block to block, according to the groups in the incomplete blocks.

$$E = \frac{1}{2} S(y_{ij} - \mu - \beta_i - \varphi_j)^2 + \xi \Sigma \beta_i + \eta \Sigma \varphi_j$$

$$\frac{\partial E}{\partial \mu} = -S(y_{ij} - \mu - \beta_i - \varphi_j) = 0$$

$$\frac{\partial E}{\partial \beta_p} = -S(y_{ij} - \mu - \beta_i - \varphi_j)\delta_{ip} + \xi = 0 \quad p = 1,2,\ldots,b$$

$$\frac{\partial E}{\partial \varphi_q} = -S(y_{ij} - \mu - \beta_i - \varphi_j)\delta_{jq} + \eta = 0 \quad q = 1,2,\ldots,v$$

$$\frac{\partial E}{\partial \xi} = \Sigma \beta_i = 0$$

$$\frac{\partial E}{\partial \eta} = \Sigma \varphi_j = 0$$

Using latin letters to denote estimates, these equations can be rearranged to read:

$$N_m + k \Sigma b_i + r \Sigma v_j = Sy$$

$$k_m + k\, b_p + (\text{sum of } v\text{'s in block } p) + \xi = \sum_j y_{pj}$$

$$r_m + (\text{sum of } b\text{'s for blocks containing } v_q) + rv_q + \eta = \sum_i y_{iq}$$

$$\Sigma b_i = 0$$

$$\Sigma v_i = 0$$

That is:

(1) $N_m = G$

(2) $k_m + kb_p + (\text{sum of } v\text{'s in block } p) + \xi = B_p,\ p = 1,2,\ldots,b$

(3) $r_m + (\text{sum of } b\text{'s containing } v_q) + rv_q = V_q,\ q = 1,2,\ldots,v$

(4) $\Sigma b_i = 0$

(5) $\Sigma v_j = 0$

Adding (2)

$$bk_m + k \Sigma b_p + r \Sigma v_j + b\, \xi = \Sigma B_p = G$$

hence $\xi = 0$ in virtue of (1), (4), (5); similarly, $\eta = 0$.

Now, add those equations in (2) over $p = p(q)$, i.e. over those blocks containing variety q. There are r such blocks. We get:

$$rk_m + k(\text{sum of } b\text{'s containing } v_q)$$

$$+ \sum_{p=p(q)} (\text{sum of } v\text{'s in block } p) = \sum_{p=p(q)} B_p = T_q$$

Now, $\sum_{p=p(q)}$ (sum of v's in block p), i.e. the sum of all the v's in blocks containing v_q, is seen to be

$$rv_q + \lambda(\text{sum of all other } v\text{'s})$$

$$= (r-\lambda)v_q + \lambda(\text{sum of all } v\text{'s}) = (r-\lambda)v_q.$$

Hence, $rm + \sum_{p=p(q)} b_p + \frac{r-\lambda}{k} v_q = \frac{T_q}{k}$.

Subtracting from (3),

$$v_q\left(r - \frac{r-\lambda}{k}\right) = V_q - \frac{T_q}{k} = Q_q,\ q = 1,2,\ldots,v$$

Now, $r - \frac{r-\lambda}{k} = r\,\frac{1 - \frac{1}{k}}{1 - \frac{1}{v}} = \frac{v\lambda}{k}$.

If we define $E = \frac{1 - \frac{1}{k}}{1 - \frac{1}{v}} = \frac{v\lambda}{rk} < 1$, since $k < v$, $v_q = \frac{Q_q}{rE}$.

Notation

$G = S_y$

V_s is the sum of the observations on variety s

B_p is the total of block p

$T_s = \sum_{p=p(s)} B_p$, the total of blocks containing variety s

$Q_s = V_s - T_s/k$

If we are to use this solution to deduce standard errors, we must solve the equations without making use of the relations among the right sides. We get:

$$\xi = \frac{1}{b}(\Sigma\, B_p - G),\quad \eta = \frac{1}{v}(\Sigma\, V_q - G)$$

$$rEv_q = V_q - \frac{1}{k}\sum_{p=p(q)} B_p - \frac{1}{v}(\Sigma\, V_j - G) + \frac{r}{kb}(\Sigma\, B_p - G).$$

Putting $V_q = 1$, all others $= 0$, we get $\mathrm{Var}\, v_q = \frac{\sigma^2}{rE}(1 - \frac{1}{v})$.

Putting $V_t = 1$, all others $= 0$, we get $\mathrm{Cov}(v_q, v_t) = \frac{\sigma^2}{rE}(-\frac{1}{v})$.

Hence, we can calculate $\mathrm{Var}(v_q - v_t) = \frac{2\sigma^2}{rE}(1 - \frac{1}{v}) + \frac{2\sigma^2}{rE}(\frac{1}{v})$

$$= \frac{2\sigma^2}{rE}.$$

In a randomized block design with the same number of observations and the same error variance, the variance of this difference would be $\frac{2\sigma^2}{r}$, smaller than we get with the incomplete block arrangement. Thus E reflects the loss in precision resulting from the fact that our contrasts come only from the blocks in which the components appear together. E is called the efficiency of the design.

Of course, the point in going to the smaller, incomplete blocks is usually that by so doing, we attain a genuinely smaller error variance, presumably enough smaller to more than compensate for the loss of efficiency.

The sum of squares of the residuals may be calculated according to the following pattern

Blocks (ignoring varieties)	$b-1$	$\frac{1}{k}\Sigma B_p^2 - G^2/bk$
Varieties (eliminating blocks)	$v-1$	$\Sigma v_q Q_q = \Sigma Q_q^2/rE$
Residual	$rv-b-v+1$	by subtraction
Total	$rv-1$	$Sy^2 - G^2/bk$

The Recovery of Interblock Information

This is a poor topic for lecturing. It is well laid out in Kempthorne. See also Rao, J.A.S.A. Dec. 1947, where he gives a rule for getting the equations for the combined analysis from those of the intrablock analysis.

Biological Assay. The Interaction of Quantity and Quality.

Only the case of measured responses will be discussed here. Instances in which the response is a count or the time until death require rather different methods.

Usually, the administration of a graded series of doses yields a monotonic response curve, mostly *S*-shaped. Sometimes the upper part of the curve is non-existent.

Two cases must be distinguished. Sometimes small doses lead to definite responses; this response curve leaves the zero dose with a non-zero slope. Usually, in these instances, the response curve is practically straight over a large enough range to permit an analysis based on straight lines and working on the lower end of the response curve. In other cases, low doses elicit responses so small as not to be useful and the testing is carried out in the middle part of the response curve where, as a rule, there is a range within which the curve is sensibly straight and again the analysis can be conducted in terms of straight lines.

Often assays are carried out to compare an unknown preparation with a standard one, usually supplied by a government. The comparison has two objectives, to decide whether the unknown preparation is qualitatively like the standard and, if so, to measure its potency relative to the standard. The comparison is carried out by comparing the two dose-response curves.

Think of some preparation and its dose response curve. Now think of diluting this preparation with some inert material, so that a given dose now contains less of the active ingredient than the same

dose of the original preparation (dose is measured in mg., cc., tons, etc.). The dose-response curve of this diluted preparation will be the same as that of the original preparation, except that it is plotted on a different dose scale. If doses of the standard are denoted x_1 and of the unknown x_2, a transformation of the form $x_2 = kx_1$, with k properly chosen, should bring the response curves into coincidence. The value of k that accomplishes this is called the relative potency of the unknown (relative, of course, to the standard).

If, in any instance, the curves cannot be brought into coincidence by the transformation $x_2 = kx_1$, the preparations are qualitatively different and we have, in Fisher's words, an interaction of quantity and quality. (Design of Experiments, Chapter 8.)

The Micro-assay

When small doses have appreciable effects and when we can work within a range where the response curves are linear, we will have two response curves radiating from a point on the response axis, whether or not we have observations at the zero level of dose. (If we do, we get into the "dummy treatment" situation.) Assuming straightness, we seek the value of the constant k which will bring the lines into coincidence.

The question may be set up as follows. Let $x_{1\alpha}$, $\alpha = 1,2,\ldots,n_1$ be doses of the standard yielding responses $y_{1\alpha}$ and $x_{2\alpha}$, $\alpha = 1,2,\ldots,n_2$ be doses of the unknown yielding responses $y_{2\alpha}$.

Define a new independent variable ξ, defined by $\xi = x$ in sample 1 and $\xi = kx$ in sample 2. That is, $\xi = x\{1 + (k-1)\delta\}$, where the indicator variable takes value zero in the first sample and unity in the second. Then, fit the line $Y = b_0 + b_1\xi$ to the sample $(y_{i\alpha}, x_{i\alpha}, \delta_{i\alpha})$.

We have

$$\begin{aligned} Y &= b_0 + b_1\xi \\ &= b_0 + b_1 x\ \{1 + (k-1)\delta\} \\ &= b_0 + b_1 x + b_2 \delta x,\ b_2 = b_1(k-1). \end{aligned}$$

The normal equations for this fitting are

$$\begin{aligned} (n_1+n_2)b_0 + b_1 Sx + b_2 S\delta x &= Sy \\ b_0 Sx + b_1 Sx^2 + b_2 S\delta x^2 &= Sxy \\ b_0 S\delta x + b_1 S\delta x^2 + b_2 S\delta^2 x^2 &= S\delta xy \end{aligned}$$

These equations may be solved for b_0, b_1, b_2, the s.s. residuals calculated (to estimate σ^2), and the relative potency calculated from $k-1 = \dfrac{b_2}{b_1}$.

Presumably we would want also to check on the straightness of the response curves and to provide some evidence of the precision of the estimate of relative potency. Testing straightness will be discussed later. The question of precision may be dealt with by calculating confidence limits by the argument first put forward by Fieller.

Let us write $\beta_1 = Eb_1$, $\beta_2 = Eb_2$, $k-1 = \beta_2/\beta_1 = \lambda$ (say). The function $b_2 - \lambda b_1$ is normal (assuming normal errors) with mean $\beta_2 - \lambda\beta_1 = 0$ and variance $\sigma^2(c_{22} - 2\lambda c_{12} + \lambda^2 c_{11})$. The elements of the inverse matrix can be found from the normal equations and σ^2 estimated by s^2 from the s.s. residuals. Then

$$\frac{(b_2 - \lambda b_1)^2}{s^2(c_{22} - 2\lambda c_{12} + \lambda^2 c_{11})} = t^2 \text{ with } n_1+n_2 - 3 \text{ d.f.}$$

According to the usual confidence limits argument, we seek those values of λ for which $t^2 \le t^2_{\alpha/2}$ to provide a $1-\alpha$ confidence range. Thus, we seek the values of λ for which

$$\lambda^2(b_1^2 - t^2_{\alpha/2}s^2c_{11}) - 2\lambda(b_1b_2 - t^2_{\alpha/2}s^2c_{12}) + (b_2^2 - t^2_{\alpha/2}s^2c_{22}) \le 0.$$

The two roots of the quadratic equation provide the limits of this range.

One curious fact about these limits is the requirement that $b_1^2 - t_{\alpha/2}^2 s^2 c_{11}$ must be positive if the inequality is to be satisfied between the two roots. That is, b_1 must be significantly different from zero at significance level α. The rest of this development concerns the design of the assay to render the statistical treatment as simple as possible.

A first obvious restriction, one that would naturally always be observed, is to use the same doses for both preparations, $x_{1\alpha} = x_{2\alpha} = x_\alpha$ (say) and $n_1 = n_2 = n$ (say). Then, the arrangement is completely orthogonal. A concomitant rearrangement of the regression equation, to take advantage of the orthogonality, is to replace

$$Y = b_0 + b_1 x + b_2 \delta x$$

by

$$Y = B_0 + B_1(x - \bar{x}) + B_2 \delta' x, \quad \delta' = \delta - \frac{1}{2}.$$

Then

$$b_0 = B_0 - B_1 \bar{x}$$

$$b_1 = B_1 - \frac{1}{2} B_2$$

$$b_2 = B_2$$

The normal equations are

$$2nB_0 = Sy = \sum_{\alpha=1}^{n} (y_{1\alpha} + y_{2\alpha}),$$

$$2\Sigma(x_\alpha - \bar{x})^2 B_1 = \Sigma(x_\alpha - \bar{x})(y_{1\alpha} + y_{2\alpha}),$$

$$\frac{1}{2}\Sigma x_\alpha^2 B_2 = \frac{1}{2}\Sigma x_\alpha (y_{2\alpha} - y_{1\alpha}).$$

The first two equations represent an analysis of the sums (or averages), and indeed they constitute the fitting of a regression line to the average of the pairs of y's on the same dose, The third equation is a fitting of a regression to the differences of the same pair of y's, constrained to pass through the origin.

The two questions about the straightness of the two response curves has been changed to one about the straightness of the average curve and one about the straightness of the curve relating differences to dose. These two questions of straightness may be approached in the usual ways.

The analysis thus far can be embedded in an orthogonal transformation.

	y_{11}	y_{21}	$\cdots$	$y_{1\alpha}$	$y_{2\alpha}$	$\cdots$	y_{1n}	y_{2n}	divisor
z_1	1	1		1	1	...	1	1	$\sqrt{2n}$
z_2	$x_1-\bar{x}$	$x_1-\bar{x}$		$x_\alpha-\bar{x}$	$x_\alpha-\bar{x}$		$x_n-\bar{x}$	$x_n-\bar{x}$	$\sqrt{2\Sigma(x_\alpha-\bar{x})^2}$
z_3			orthogonal						
$\vdots$									
z_n	a_1	a_1		a_α	a_α		a_n	a_n	
z_{n+1}	$-x_1$	x_1		$-x_\alpha$	x_α		$-x_n$	x_n	$\sqrt{2\Sigma x_\alpha^2}$
z_{n+2}			orthogonal						
$\vdots$									
z_{2n}	$-b_1$	b_1		$-b_\alpha$	b_α		$-b_n$	b_n	

The analysis of variance table corresponding to this is:

	d.f.	s.s.				
among levels	$n-1$	$z_2^2+\ldots z_n^2 = \frac{\Sigma T_\alpha^2}{2} - \frac{G^2}{2n}$	average slope	1	$z_2^2 = \frac{(\Sigma(x_\alpha-\bar{x})T_\alpha)^2}{2\Sigma(x_\alpha-\bar{x})^2}$
			residuals	$n-2$	by subtraction
within levels	n	$z_{n+1}^2+..z_{2n}^2$ by subtraction	preparations	1	$z_{n+1}^2 = \frac{[\Sigma x_\alpha(y_{2\alpha}-y_{1\alpha})]^2}{2\Sigma x_\alpha^2}$
			levels × preparations	$n-1$	by subtraction
total	$2n-1$	$Sy^2-G^2/2n$			

We note that $z_1 = \sqrt{2n}\, B_0$

$$z_2 = \sqrt{2\Sigma(x_\alpha - \bar{x})^2}\, B_1$$

$$z_{n+1} = \sqrt{\frac{\Sigma x_\alpha^2}{2}}\, B_2$$

The components $z_3, \ldots z_n$ would be chosen to test the nature of the departures from linearity of the average points. Values of polynomials chosen to be orthogonal over the doses used would be most satisfactory. If we define

$$P_0 = \lambda_{00}$$

$$P_1 = \lambda_{10} + \lambda_{11}\, x$$

$$P_2 = \lambda_{20} + \lambda_{21}\, x + \lambda_{22}\, x^2$$

etc.

with the λ's chosen so that $SP_{i\alpha}P_{j\alpha} = 0$, $i \neq j$. We could choose $\lambda_{00} = 1$, $P_1 = x - \bar{x}$, etc. Then $z_3 = \Sigma P_{2\alpha} y_\alpha$ is the quadratic component, and so on. If the doses are arranged to be equally spaced, Fisher's ξ' polynomials can be used.

In the same way, we may wish to choose the coefficients in $z_{n+2}, \ldots, z_{2n}$ to display the nature of the trend in the curve relating the differences to the dose. We could define polynomials

$$Q_1 = \mu_{11}\, x$$

$$Q_2 = \mu_{21}\, x + \mu_{22}\, x^2$$

and so on

with the μ's chosen so that $SQ_{i\alpha}Q_{j\alpha} = 0$, $i \neq j$. These polynomials would be suitable for fitting polynomial regressions constrained to pass through the origin.

We could arrange to have the doses not only equally spaced, but

reducible to 1,2,3 by a change of scale. These Q-polynomials and tables of their values have been determined by Fick (1985).

With $Q_1 \propto x$, a little computation yields

$$Q_2 \propto \frac{n(n+1)}{2} x - \frac{2n+1}{3} x^2$$

If we take $n = 3$, we get:

$Q_1(x) = x$	$Q_2(x)$	$Q_3(x)$
1	11	3
2	8	-3
3	-9	1

	y_{11}	y_{21}	y_{12}	y_{22}	y_{13}	y_{23}	divisor	
z_1	1	1	1	1	1	1	$\sqrt{6}$	B_0
z_2	-1	-1	0	0	1	1	$\sqrt{4}$	B_1
z_3	-1	-1	2	2	-1	-1	$\sqrt{12}$	quadratic
z_4	-1	1	-2	2	-3	3	$\sqrt{28}$	B_2
z_5	-11	11	-8	8	9	-9	$\sqrt{532}$	quadratic
z_6	-3	3	3	-3	-1	1	$\sqrt{38}$	cubic

z_5 and z_6 are interactions of doses and preparations.

Fisher's example has doses 0,1,2. With two preparations, the transformation would be

	y_{10}	y_{20}	y_{11}	y_{21}	y_{12}	y_{22}	
z_1	1	1	1	1	1	1	
z_2	-1	-1	0	0	1	1	average slope
z_3	-1	-1	2	2	-1	-1	quadratic
z_4	0	0	-1	1	-2	2	preparations
z_5	0	0	2	-2	-1	1	interaction

The multipliers in z_5 are values of Q_2, found here simply by orthogonality with z_4.

Estimation of Error

When n is small, the assay would presumably be replicated and an estimate of error would be thus provided. If n is large and no replication provided, the first few components (the P's and Q's) would probably exhaust the trends, leaving the rest to represent error only.

Standard errors of b_1 and b_2

We have Var $B_1 = \sigma^2/2\Sigma(x_\alpha - \bar{x})^2$, Var $B_2 = 2\sigma^2/\Sigma x_\alpha^2$, Cov $(B_1, B_2) = 0$. Also, $b_1 = B_1 - \frac{1}{2}B_2$, $b_2 = B_2$. Then,

$$\text{Var } b_1 = \text{Var } B_1 + \frac{1}{4} B_2$$

$$= \frac{\sigma^2}{2}\left(\frac{1}{\Sigma(x_\alpha - \bar{x})^2} + \frac{1}{\Sigma x_\alpha^2} \right)$$

We need also, Cov (b_1, b_2), which may be computed from

Var $(b_1 + b_2)$ = Var b_1 + Var b_2 + 2 Cov (b_1, b_2).

i.e. Var $(B_1 + \frac{1}{2} B_2)$ = Var $(B_1 - \frac{1}{2} B_2)$ + Var B_2 + 2 Cov (b_1, b_2). Thus Cov $(b_1, b_2) = -\frac{1}{2}$ Var $B_2 = -\sigma^2/\Sigma x_\alpha^2$.

The parallel assay

When it is necessary to work in the middle of the response curve, it is customary to plot response against log (dose). The transformation $x_2 = kx_1$ becomes $\log x_2 = \log k + \log x_1$ and the two response curves, if qualitatively alike, become parallel curves, in the sense that they may

be brought into coincidence by a transformation in the direction of the log (dose) axis. The distance between the two curves is log (relative potency).

If we confine the doses to the part of the range where the response curves are sensibly linear, the statistical analysis amounts simply to fitting a pair of straight lines, constrained to be parallel and finding the distance between them. We would, of course, want to check on linearity and parallelism.

If the sample values are $(x_{i\alpha}, y_{i\alpha})$, $\alpha = 1,2,\ldots,n_i$, $i = 1,2$, we can fit a pair of lines, constrained to be parallel, by defining $\xi = x + k\delta$ and fitting

$$Y = b_0 + b_1 \xi$$
$$= b_0 + b_1 x + b_2 \delta, \; b_2 = k\, b_1.$$

Here, the x's stand for log (doses) and k is the log (relative potency). $\delta = 0$ in sample 1, $\delta = 1$ in sample 2. The normal equations are

$$(n_1 + n_2)b_0 + b_1 Sx + b_2 S\delta = Sy$$
$$b_0 Sx + b_1 Sx^2 + b_2 S\delta x = Sxy$$
$$b_0 S\delta + b_1 S\delta x + b_2 S\delta^2 = S\delta y$$

Again, the system becomes simpler if we arrange that $x_{1\alpha} = x_{2\alpha}$, $n_1 = n_2 = n$ and fit

$$Y = B_0 + B_1(x - \bar{x}) + B_2\delta', \; \delta' = \delta - \frac{1}{2}.$$

Then,
$$b_0 = B_0 - B_1\bar{x} - \frac{1}{2} B_2$$
$$b_1 = B_1$$
$$b_2 = B_2$$

The normal equations are

$$2nB_0 = \Sigma(y_{1\alpha} + y_{2\alpha})$$
$$2\Sigma(x_\alpha - \bar{x})^2 B_1 = \Sigma(x_\alpha - \bar{x})(y_{1\alpha} + y_{2\alpha})$$
$$\frac{n}{2} B_2 = \frac{1}{2}\Sigma(y_{2\alpha} - y_{1\alpha})$$

We see that the B's are independent, with

$$\text{Var } B_1 = \frac{\sigma^2}{2\Sigma(x_\alpha - \bar{x})^2} \qquad \text{Var } B_2 = \frac{2\sigma^2}{n} \quad .$$

The first two of the equations represent the fitting of a regression line to the average of the response curves and the third the fitting of the regression $E(y_2 - y_1) = \text{constant}$.

We can construct an orthogonal transformation to exhibit all the features of this regression. In particular, if the doses are chosen to be in geometric progression, so that the log (doses) are equally spaced, we would naturally use the ξ' orthogonal polynomial values.

	y_{11}	y_{21} $\cdots$	$y_{1\alpha}$	$y_{2\alpha}$ $\cdots$	y_{1n}	y_{2n}	divisor
z_1	1	1	1	1	1	1	$\sqrt{2n}$
z_2	$\xi_1'(x_1)$	$\xi_1'(x_1)$	$\xi_1'(x_\alpha)$	$\xi_1'(x_\alpha)$	$\xi_1'(x_n)$	$\xi_1'(x_n)$	$\sqrt{2S\ \xi_1'^2}$
$\vdots$							
z_n	$\xi_{n-1}'(x_1)$	$\xi_{n-1}'(x_1)$	$\xi_{n-1}'(x_\alpha)$	$\xi_{n-1}'(x_\alpha)$	$\xi_{n-1}'(x_n)$	$\xi_{n-1}'(x_n)$	$\sqrt{2S\ \xi_{n-1}'^2}$
z_{n+1}	-1	1	-1	1	-1	1	$\sqrt{2n}$
z_{n+2}	$-\xi_1'$	ξ_1'	$-\xi_1'$	ξ_1'	$-\xi_1'$	ξ_1'	
$\vdots$							
z_{2n}	$-\xi_{n-1}'$	ξ_{n-1}'	$-\xi_{n-1}'$	ξ_{n-1}'	$-\xi_{n-1}'$	ξ_{n-1}'	

$z_1 \propto B_0, z_2 \propto B_1, z_3 \ldots z_n$ are quadratic, cubic etc. components, depicting lack of straightness of the average line. $z_{n+1} \propto B_2, z_{n+1}, \ldots z_{2n}$ reflect lack of straightness of the line showing the dependency of the differences on log dose, i.e. lack of parallelism of the two response curves, i.e. interaction of doses and preparations.

Confidence limits for $k = b_2/b_1$ are obtained as before.

Transect sampling

Frequently an area is sampled by taking strips, parallel to one another and equally spaced, making complete counts or other assessments and forming an estimate of the total by dividing this count by the sampling ratio. An example is the timber cruise, used to estimate the total stand on some given area. Strips of width d are cruised, recording the total stand (number, volume, or whatever) of each strip and its length, so that the total area sampled may be calculated. If this is a, and the total area is A, the recorded count, C, say is "blown up" to yield an estimate of the total stand, T, by calculating $T = CA/a$.

This estimate may be questioned, but the real difficulty arises because the systematic (i.e. not random) allocation of the samples provides no definition of error and therefore no way of estimating the precision of the estimate.

Practitioners of the art recoil in horror from the suggestion that the strips should be allocated randomly, for the good reason that it would be unmanageable in the field. They have therefore developed a certain mythology about the systematic sampling they practice.

1. The systematic sample yields more precise estimates than a randomly chosen sample. In this they may be correct.
2. The "blow-up" is the correct procedure for estimating the total stand. This would be the case for random sampling, but is not quite correct for the systematic sample.
3. The observations yielded by the systematic sample may be treated as if they arose from random samples for the purpose of calculating the error variance. This, of course, is nonsense.

We can look at the problem in the following way. Let us regard the count on each strip, divided by the width of the strip, d, as an estimate of the (linear) density of stand at the middle of the strip. If we could know this density corresponding to each point of the base line, the stand could be described by the density curve and the total stand would be given by the integral of this density function.

We have, in fact, only a few points on the density curve and they, furthermore, have been observed with error. Our determination of the total stand becomes, in this construction, an exercise in numerical integration when the observed points are subject to error.

The appraoch put forward here is simply to fit a polynomial up to the point where the residuals appear to display no trend and treat them as displaying error only. This is a new and somewhat subjective definition of error.

Since the sampling strips are equally spaced, the fitting can be carried out easily using the tabulated values of the ξ'-polynomials. If, in addition, the integrals of these polynomials are available, the

whole of the arithmetic becomes quite simple. See DeLury D.B. 'Values and Integrals of the Orthogonal Polynomials'.

Capture-recapture methods

Much has been said about the essential role of randomness in the conduct of experiments. It provides a definition of error, and indeed converts what would otherwise enter the comparisons as biases with error.

When the object of sampling is the making of an absolute estimate of some quantity, randomness is still important if we need a sound definition of error, but sometimes the estimate itself may be excellent, even through randomness has not been introduced. We may then seek some substitute for a definition of error based on randomness which, even though lacking the authority of an a priori definition based on randomness, may still prove to be useful. The estimation of stands of timber from a systematic sample appears to be an instance of this. Nothing general can be said here; each device is special to its own particular set of circumstances.

It is obvious, too, that when one is estimating, rather than comparing, biases which can be "designed out", without knowing exactly what they are, in an experiment, may become important when estimation is the object. Indeed, it seems that usually the possibility of bias transcends all other considerations.

One instance of methods that have evolved, in spite of the impossibility of arranging a dependable procedure for randomization, will be discussed here. It has to do with the estimation of the number

of individuals in a population which is not tied to any framework which could be used to provide a basis for selecting samples randomly. An instance of this is the estimation of the number of fish in a lake. Methods of this kind have been used with biological populations of other kinds and in other circumstances as well. It seems likely that they have wider fields of applicability than is generally realized, however they are not widely known.

Mark-recapture; tagging

The first proposal is rather obvious and has been in use for a long time. (Laplace)

Let us suppose, to start with, that we have a population of N individuals that is closed, no recruitment or emigration or depletion through mortality. Let us capture a number X of these individuals, mark them and replace them in the population. Then, select randomly a sample of size n and ascertain the number x of them bearing marks. Then, equating x/n with X/N, we get an estimator $\hat{N} = nX/x$.

To put this in a statistical setting, we must regard x as a statistical variable. Its distribution is obviously hypergeometric, but if, as is likely to be the case, n is small compared to N, x may be treated as binomial (for simplicity) without sensible inaccuracy. The parameters of the binomial are $\pi = X/N$ and n. We have, then,

$$E\,\frac{x}{n} = \frac{X}{N},\ \mathrm{Var}\,\frac{x}{n} = \frac{X}{N}\left(1 - \frac{X}{N}\right)/n$$

When we come to deal with the effects of the binomial error on

the estimator $\hat{N} = nX/x$, we are confronted with the reciprocal of a binomial variable, which therefore becomes infinite with positive probability. It therefore seems best to deal with the reciprocal, $1/\hat{N}$, which is binomially distributed, and invert. This is, of course, what we have done when we write $\hat{N} = nX/x$. We may note, even though it is not important, that this estimator is biased, since its reciprocal is not.

If we wish to exhibit the precision of the estimate (as we should), we can make use of tables or charts giving confidence limits for the binomial distribution to set limits for X/N and invert. Mostly, though, it is sufficient to invoke the normal approximation to the binomial and write

$$\frac{\frac{x}{n} - \frac{X}{N}}{\sqrt{\frac{1}{n}\,\frac{X}{N}\left(1 - \frac{X}{N}\right)}} = N(0,1).$$

Hence, we can evaluate the probability that this quantity will lie in any given range. For example, the probability is about 0.95 that it will lie in the range (-2,2).

Alternatively, the 95% interval is composed of those values of N for which

$$\frac{\left(\frac{x}{n} - \frac{X}{N}\right)^2}{\frac{X}{N}\left(1 - \frac{X}{N}\right)} \leq z,$$

where z is the 5% point of $\chi^2_{(1)}$.

Rearranging this inequality, we get

$$\frac{X^2}{N^2}\left(1 + \frac{z}{n}\right) - \frac{X}{N}\left(2\,\frac{x}{n} + \frac{z}{n}\right) + \frac{x^2}{n^2} \leq 0.$$

We would expect that, replacing this inequality by an equality and solving for $\frac{X}{N}$, we would get the endpoints of the confidence interval covering the mean of x/n, from which we could calculate the range for N. It seems prudent, though, to check this through.

The discriminant of the quadratic must be positive if it is to yield real roots. The discriminant is

$$(2\,\frac{x}{n} + \frac{z}{n})^2 - 4(1 + \frac{z}{n})\,\frac{x^2}{n^2}$$

which may be reassembled in the form

$$\frac{z}{n}\,[1 + \frac{z}{n} - 4(\frac{x}{n} - \frac{1}{2})^2].$$

Now, the greatest value that $(\frac{x}{n} - \frac{1}{2})^2$ can take is $\frac{1}{4}$. The discriminant is therefore positive in all circumstances and, since the quadratic is concave upward, the function is negative between its two zeros.

We may as well, now, rewrite the inequality in terms of N

$$x^2N^2 - nX(2x + z)N + nX^2(n + z) \leq 0.$$

To get 95% confidence limits for N, we put $z = 4$ and solve the quadratic for N.

Remarks

1. The procedure is heavily dependent on randomness in the sampling, which in many applications is unattainable.

2. The estimate is reached under the supposition that the proportion tagged remains constant. This assumption would not be violated by mortality or emigration as long as tagged and not tagged are equally affected.

3. An influx of fish after the marks were released and before the sample

is taken, being necessarily not marked, would do no more than create a new population, which would be the one whose size is estimated.

A sampling of this sort is not likely to be employed much in practice. Rather, there would be a sequence of samples, in each of which records of tagged fish would be kept and perhaps untagged fish would be tagged and released. This sort of thing can be done in a wide variety of patterns. To discuss this kind of procedure, we need some more notation.

N_t and X_t: the number of individuals and of tagged individuals in the population <u>just</u> <u>before</u> the t^{th} sample is taken

n_t and x_t: the number of individuals and of tagged individuals in the t^{th} sample

m: the number of samples taken

$n = \Sigma n_t$ and $x = \Sigma x_t$.

In particular, N_1 and X_1 are the numbers before the sampling starts. It will be convenient to drop the subscripts, $N_1 \equiv N$, $X_1 \equiv X$.

The object of the sampling is to estimate N.

We can develop a general formula that is not overly complicated if we make the simplifying assumption that each sample takes only a small fraction of the population, i.e. n_t/N_t is small. Then, we may regard the sampling as binomial rather than hypergeometric. It is assumed, of course, that the sampling is random.

The probability of obtaining x_t tagged fish in a sample of n_t is

$$\binom{n_t}{x_t}\left(\frac{X_t}{N_t}\right)^{x_t}\left(1 - \frac{X_t}{N_t}\right)^{n_t - x_t} = f_t \quad \text{(say)}.$$

Adopting the method of maximum likelihood, the likelihood function is

$$L \alpha \prod_{t=1}^{m} f_t$$

$$\log L = - \Sigma x_t \log N_t + \Sigma (n_t - x_t) \log \left(1 - \frac{X_t}{N_t}\right) \text{ plus other terms}$$

not depending on N.

$$\frac{\partial}{\partial N} \log L = \left[-\Sigma x_t \frac{1}{N_t} + \Sigma (n_t - x_t) \frac{1}{1 - \frac{X_t}{N_t}} \cdot \frac{X_t}{N_t^2} \right] \frac{\partial N_t}{\partial N} = 0.$$

Now, even though N_t may vary with t, we may reasonably suppose that it varies in a manner expressible by $N_t = Ng(t)$ and therefore $\frac{1}{N_t} \frac{\partial N_t}{\partial N}$ does not vary with t. The estimating equation may therefore be organized to read:

$$\sum_{t=1}^{m} \frac{x_t - n_t \frac{X_t}{N_t}}{1 - \frac{X_t}{N_t}} = 0.$$

Some special cases

1. The size of the population does not change during the sampling period. The population must be closed and all samples returned to the population. Usually, all unmarked fish in the sample are marked and returned. In any event, as long as the records are kept so that the X_t are known, the equation may be set up and solved for N.

In this case, $N_t = N$, for all t. The equation is

$$\sum \frac{X_t - n_t \frac{X_t}{N}}{1 - \frac{X_t}{N}} = 0.$$

This kind of sampling scheme and the above estimating equation are

referred to as Schnabel estimates, after Miss Zoe Schnabel derived the estimating equation in a Master's thesis at Wisconsin (supervisor Mark Ingraham) (Schnabel, Bull. Amer. Math. Soc. 1938). Miss Schnabel suggested writing the equation in the form

$$\sum \left(x_t - \frac{n_t X_t}{N} \right) \left(1 + \frac{X_t}{N} + \frac{X_t^2}{N^2} + \ldots \right) = 0$$

and truncating at some point to form an equation of not too high a degree. This is a dubious procedure and has, in fact, rarely been used, except in the form $\Sigma(x_t - \frac{n_t X_t}{N}) = 0$, which leads to an explicit formula for a first order approximation.

$$\hat{N}_1 = \frac{\Sigma n_t X_t}{\Sigma x_t} = \frac{\Sigma n_t X_t}{x}$$

If more accuracy is needed (as presumably it is), a sensible procedure is to compute "weights",

$$W_t = \frac{1}{1 - \frac{X_t}{\hat{N}_1}}$$

and solve the equation

$$\Sigma(x_t - n_t \frac{X_t}{N}) W_t = 0,$$

yielding the adjusted solution

$$\hat{N}_2 = \frac{\Sigma W_t n_t X_t}{\Sigma W_t x_t} .$$

This procedure may be continued, computing new "weights", $\frac{1}{1 - \frac{X_t}{\hat{N}_2}}$ and so on. This iterative process converged very fast, in most cases.

2. The proportion of tagged individuals in the population remains constant. This kind of sampling is associated with the names Peterson

(fish) and Lincoln (birds). The first example of this discussion is of this type.

To meet the conditions of this type, samples should be returned, but a computation based on the hypergeometric distribution leads to the same estimation formula when samples are not returned.

Writing $\frac{X_t}{N_t} = \frac{X}{N}$, the solution is $N = X \frac{n_t}{x_t} = n \frac{X}{x}$.

Confidence Limits

For the Peterson type, the discussion already given is sufficient, since the sum of binomials variables with the same proportion throughout is binomial. The Schnabel type of estimate requires a bit more discussion. See DeLury, Journal of the Fisheries Research Board of Canada, 8 (4), 1951 and 15 (1), 1958.

Another approach to the Schnabel type of tagging plan is, I think, preferable. Since

$$E \frac{x_t}{n_t} = \frac{X_t}{N}, \quad \text{Var} \frac{x_t}{n_t} = \frac{1}{n_t} \frac{X_t}{N} \left(1 - \frac{X_t}{N}\right),$$

a plot of $\frac{x_t}{n_t}$ against X_t should yield a straight line through the origin, with slope $\frac{1}{N} = P$ (say). We may therefore think of fitting a regression of $\frac{x_t}{n_t}$ on X_t, constrained to pass through the origin.

According to standard statistical theory, the residuals should be weighted inversely as their variances. Thus we minimize

$$\sum W_t \left(\frac{x_t}{n_t} - PX_t\right)^2, \quad W_t = \frac{n_t}{\frac{X_t}{N_t}\left(1 - \frac{X_t}{N_t}\right)}.$$

This leads to the estimating equation

$$\hat{P} \Sigma W_t X_t^2 = \Sigma W_t \frac{x_t}{n_t} X_t \text{ , i.e.}$$

$$\hat{P} \sum \frac{n_t X_t^2}{\frac{X_t}{N}(1 - \frac{X_t}{N})} = \sum \frac{x_t X_t}{\frac{X_t}{N}(1 - \frac{X_t}{N})}$$

Writing $\hat{P} = \frac{1}{\hat{N}}$ and rearranging,

$$\sum \frac{x_t - \frac{n_t X_t}{\hat{N}}}{1 - \frac{X_t}{\hat{N}}} = 0 .$$

This seems to be identical with the maximum likelihood equation.

Thus we see that the maximum likelihood estimator weights the observations according to sample size and also according to the proportions of tagged individuals in the population.

Here our inability to ensure randomness in the sampling becomes embarrassing. Owing to the tendency of fishes to move about in schools, the proportion tagged available at any particular sampling may be vastly different from $\frac{X_t}{N}$ and therefore the weights may be seriously wrong. In these circumstances, it seems prudent to weight by sample size only. This leads to the simple estimator

$$\hat{P} = \frac{\Sigma x_t X_t}{\Sigma n_t X_t^2} ; \quad \hat{N} = \frac{\Sigma n_t X_t^2}{\Sigma x_t X_t} .$$

This is the estimator published somewhat earlier than Schnabel's by Schumacher and Eahmeyer. For the reasons given above, I think it a preferable estimator, even though it does not use the statistically efficient weights.

A simulated sampling study, with randomness assured, showed no persistent differences between the estimators themselves or in the

lengths of the confidence intervals associated with them.

The rest of the analysis may be conducted according to standard regression theory, using the s.s. residuals to estimate error. This is a different definition of error from that used earlier, in that it includes variation from all sources (mechanical, biological) in addition to the binomial sampling error. Of course, no randomness has been arranged, but it seems preferable to include all deviations from expectation than to simply postulate a binomial error distribution which is not even sampled randomly. Another consideration is the fact that this binomial component has a variance that increases as the number of tagged individuals increases, thus violating the condition of constant variance. The effect of this is likely to be small, in as much as the binomial component is usually heavily outweighed by other sources of variation.

The substitution of error, so regarded, may be made also with the Peterson type of estimate. We have $E\,\frac{x_t}{n_t} = \frac{X}{N} = P$ (say) and we may think of estimating P by a weighted least squares fitting, by minimizing $\sum_t n_t \left(\frac{x_t}{n_t} - P\right)^2$. We get, in this way,

$$\Sigma\, n_t \left(\frac{x_t}{n_t} - \hat{P}\right)^2 = 0, \quad \hat{P} = \frac{\Sigma\, x_t}{\Sigma\, n_t} = \frac{x}{n}.$$

The s.s. residuals is $\Sigma\, n_t \left(\frac{x_t}{n_t} - \hat{P}\right)^2$, which reduces to $\Sigma\, \frac{x_t^2}{n_t} - \frac{x^2}{n}$. Then, the estimator of σ^2 is given by $s^2 = \frac{1}{m-1}\left[\Sigma\, \frac{x_t^2}{n_t} - \frac{x^2}{n}\right]$. The estimated variance of $\hat{P}$ is s^2/n.

Confidence limits for P are given by

$$\hat{P} - t\, s/\sqrt{n} \le P \le P + t\, s/\sqrt{n}.$$

Writing $\frac{X}{N}$ for P and $\frac{x}{n}$ for $\hat{P}$, we get, after some rearrangements, confidence limits for N:

$$\frac{nX}{x + t\ s\sqrt{n}} \leq N \leq \frac{nX}{x - t\ s\sqrt{n}}$$

Catch-effort methods

Another approach to the estimation of mobile populations depends on the notion of effort expended in capturing samples — net-night, angler-hour, etc. We then come out of our sampling with a list like the following.

sample number	catch	effort
1	c_1	e_1
2	c_2	e_2
⋮	⋮	⋮
m	c_m	e_m

From these records, we can calculate others, for example, the catch per unit of effort for each sample, $c(t) = \frac{c_t}{e_t}$, total catch up to sample t, $K(t) = \sum_{i=1}^{t-1} c_i$, and total effort expended up to the t^{th} sample, $E(t) = \sum_{i=1}^{t-1} e_i$.

Usually, when appreciable fractions of the population are captured and removed, the depletion shows up in diminished catches per unit of effort. In some circumstances, this change in the catch per unit of effort may be used to estimate the size of the population.

A mathematical formulation

Let $N = N(t)$ be the size of the population at time t, $r(t)$ be the rate of influx into the population and $d(t)$ the rate of depletion. Then,

$$\frac{dN}{dt} = N(r(t) - d(t)),$$

which, of course, says nothing. Now, assume

(1) $r(t) = 0$

(2) $d(t) = k(t)e(t)$

(3) $k(t) = k$, a constant, called the catchability.

(1) and (2) imply that the population is closed and that the rate of depletion through sampling is proportional to the rate of expenditure of effort. (3) says that, throughout the sampling period, each unit of effort captures the same fraction of the population. This in turn implies that the units of effort do not compete with one another.

With these assumptions,

$$\frac{1}{N}\frac{dN}{dt} = -ke(t)$$

$$\log N \Big|_{N(0)}^{N(t)} = -k\int_0^t e(t)dt = -kE(t)$$

$$N = N(0)e^{-kE(t)}$$

Hence $\frac{dN}{dE} = -kN(0)e^{-kE(t)}$ and, because the population is closed, $\frac{dN}{dE} = -c(t)$. Thus, we get

$$C(t) = k\,N(0)e^{-k\,E(t)} \text{ and}$$

$$\log C(t) = \log (k\,N(0)) - k\,E(t).$$

Again, because the population is closed,

$$K(t) = N(0) - N(t) \quad \text{and} \quad C(t) = k\,N(t)$$

by definition. Therefore, eliminating $N(t)$,

$$C(t) = k\,N(0) - k\,K(t).$$

We may, therefore, plot either or both of these straight lines. If they do turn out to be reasonably straight, the assumptions are, in some measure, supported. The fitted lines then yield estimates of $k\,N(0)$ and k.

The question of procedures for a numerical fitting is perhaps not wholly obvious. A simple model may help here.

A bead model

Think of a population of N white beads in a box. A unit of effort may be a dip with a small scoop. The white beads taken by a unit of effort are counted and replaced by an equal number of red beads, to keep the unit of effort constant. Any red beads taken in the sampling are replaced. For simplicity, assume that each unit of effort takes the same number n of beads. Then, $\frac{n}{N}$ is the catchability. Assume also that n is small compared to N, so that the sampling is virtually binomial.

Let the proportion of white beads in the box, just before the t^{th} sample is drawn, be $p(t)$. Then, if $C(t) = c(t)$ is the number of white beads in the t^{th} sample,

$$EC(t) = np(t), \text{ Var } C(t) = np(t)(1-p(t)).$$

If $K(t)$ beads have been removed in the first $t-1$ samples, then

$p(t) = 1 - \frac{K(t)}{N}$ and we have

$$E(C(t)\,|K(t)) = n - \frac{n}{N} K(t) = kN - kK(t).$$

Thus, this relation is of the conditional sort met in regression theory and may be fitted according to regression methods. While a weighted fitting is theoretically preferable, it appears that an unweighted fitting is adequate. Standard regression methods then yield estimates of kN and k and their standard errors, with the estimate of error taken to be the s.s. residuals. Fieller's method supplies confidence limits for N.

An equation like $\log C(t) = \log (kN(0)) - kE(t)$ may be derived for this model by calculating the unconditional mean of $C(t)$. See DeLury, Biometrics, Vol. 3, No. 4. It turns out to be

$$E\ C(t) = kN(1-k)^{E(t)}, \text{ or}$$

$$\log E\ C(t) = \log (kN) + E(t) \log (1-k)$$

$$= \log (kN) - kE(t) \text{ since } k \text{ is small.}$$

This does not fit into the regression pattern, but extensive simulations indicate that fitting using the usual normal equations yields satisfactory estimates.

These simulations, which used fairly large amounts of effort in each sample, turned out to produce biased estimates and led to the suggestion that $K(t)$ and $E(t)$ be calculated as

$$K(t) = c_1 + c_2 + \ldots + c_{t-1} + \frac{1}{2} c_t,$$

$$E(t) = e_1 + e_2 + \ldots + e_{t-1} + \frac{1}{2} e_t.$$

This correction may be thought of as a continuity correction or as a compensation for using binomial instead of hypergeometric theory.

Catch-effort methods may be used in conjunction with a tagging study, by tagging and returning all captures and disregarding all recaptures. Indeed, there are several advantages to this combination of methods. If the catch-effort and tagging estimates agree reasonably well, depending as they do on randomness in different ways, they give considerable support to each other. Also, each contains information on the assumptions on which the other depends.

1. Catch-effort estimates depend on constant catchability. Tagging records supply a population of known size, so recaptures supply a sequence of direct estimates of catchability, which can be studied for trends.

2. Tagging estimates require that tagged and untagged be equally catchable. This may be studied by applying catch-effort methods to tagged and untagged separately, to detect any difference there may be between them.

1. The rate of a chemical reaction, y, is measured 3 times at each of 5 equally-spaced temperatures, t_1, t_2, t_3, t_4, t_5. The <u>averages</u> of the measured rates are, in appropriate units,

t_1	t_2	t_3	t_4	t_5
1	2	9	28	65

The error sum of squares, with 10 d.f., is found to be 11.3.

Carry out a study of the curve relating rate of reaction to temperature, with a view to deciding the degree of a fitted polynomial that will fit the observations adequately.

Find the values of the coefficients of this fitted polynomial, written in the form $Y = B_0\xi_0' + B_1\xi_1' + \text{etc.}$

2. The fitting of a regression $Y = b_0 + b_1 u + b_2 v$ is carried out in the following arithmetical form. (All numbers are exact, in the sense that they have not been rounded off.)

16	40	200	733	1	0	0
40	120	500	1989	0	1	0
200	500	3000	8285	0	0	1
733	1989	8285	37899			
1	2.5	12.5	45.8125	0.0625	0	0
0	20	0	156.5	-2.5	1	0
0	0	500	-877.5	-12.5	0	1
0	156.5	-877.5	4318.4375			
1	0	12.5	26.25	0.375	-0.125	0
0	1	0	7.825	-0.125	0.05	0
0	0	500	-877.5	-12.5	0	1
0	0	-877.5	3093.825			
1	0	0	48.1875	0.6875	-0.125	-0.025
0	1	0	7.825	-0.125	0.05	0
0	0	1	-1.755	-0.025	0	0.002
0	0	0	1553.8125			

(a) Identify the values of b_0, b_1, b_2 and the sum of squares of residuals.

(b) Calculate the sum of squares attributable to u.

(c) Calculate, in two ways, the sum of squares attributable to v.

(d) Calculate, in three ways, the sum of squares attributable to u and v together.

(e) Do you perceive anything odd, exceptional or unusual in this fitting? Comment and offer an explanation to account for it.

(f) A parameter of interest is defined by $\pi = Eb_1 - Eb_2$. Calculate 95% confidence limits for π.

(g) A parameter of interest is defined by $\lambda = \dfrac{Eb_1}{Eb_2}$. Calculate 95% confidence limits for λ.

3. Observations on a response, y, and a concomitant variable, x, are made in two different sets of circumstances, referred to here as sample 1 and sample 2. Sums of squares and products are as follows.

	d.f.	(xx)	(xy)	(yy)
within sample 1	4	16	30	70
within sample 2	2	6	10	18

(a) Calculate the slope of a linear regression fitted to the observations in sample 1.

(b) Make the same calculation within sample 2.

(c) Carry out a test to decide whether the slopes obtained in (a) and (b) are significantly different.

(d) Whatever the conclusion reached in (c), calculate the slope of two regressions, constrained to be parallel, the sum of squares attributable to regression and the sum of squares of residuals.

4. A randomized block experiment, in which a response, y, is measured and a concomitant variable, x, is recorded, leads to the following list of sums of squares and products.

	d.f.	(xx)	(xy)	(yy)
blocks	9	--	--	--
treatments	1	13	25	48
error	9	7	5	30

Calculate the reduced error sum of squares and the adjusted treatment sum of squares.

What conclusions (or observations or suspicions) would you offer if (a) the variable x cannot reflect treatment differences and, presumably, was included only to establish some control over error; (b) the variable x may be affected by the treatments.

5. D - diameter of weft yarn

   T - amount of twist in weft yarn

   u - number of picks (i.e. weft yarns) per inch

   x - a measure of stiffness of the fabric

	$D_1 = 2.1$		$D_2 = 3.0$		Row Totals	
	u	x	u	x	u	x
$T_1 = 2$	88	84	95	104		
	82	75	88	94		
	74	61	84	75		
Cell totals	244	220	267	273	511	493
$T_2 = 8$	95	71	95	78		
	92	55	92	72		
	84	54	82	56		
Cell totals	271	180	269	206	540	386
Column totals	515	400	536	479	1051	870

(a) Calculate, by the analysis of covariance, adjusted sums of squares of x, corresponding to D, T, $D \times T$ and carry out tests of significance on them.

(b) Calculate a table of x-values, adjusted to the average values of u.

(c) Test the significance of the regression of x on u.

6. Two new designs for a projectile are proposed. These designs differ from the current standard only in the shape of the nose, but are of the same weight and the same diameter. The new designs are intended to reduce air drag, and an experiment is planned to determine which is the better of the new designs, and whether the reduction is sufficient to warrant a change. 10 rounds of each projectile are fired and velocities are measured at the muzzle and at 100 yd. from the muzzle. The average retardation, u, in ft. per sec. per 100 ft., (dv/dx), and the initial velocity, v_0 in ft. per sec., are both recorded. The experiment is carried out over a period of 10 days in order to make comparisons under various weather conditions. The data are given in the following table.

(u, v_0)

u = retardation (ft/sec/100 ft.) v_0 = initial velocity (ft/sec.)

Day	New design I	New design II	Standard
1	(86,1985)	(108,2016)	(105,1999)
2	(87,1983)	(90,1981)	(98,1975)
3	(96,2006)	(107,2003)	(102,1992)
4	(105,2020)	(110,2005)	(105,2019)
5	(99,2008)	(88,1984)	(116,2021)
6	(98,2002)	(107,2017)	(112,2001)
7	(117,2024)	(118,2008)	(114,2014)
8	(86,1971)	(113,2014)	(118,2015)
9	(104,2010)	(113,2020)	(110,1996)
10	(94,2002)	(115,2018)	(115,2018)

Using the analysis of covariance technique, and assuming that retardation is a linear function of initial velocity, obtain answers to the following questions. (Note: It is known that initial velocity is not dependent on the shape of the projectile.)

.../cont'd.

Question #6 cont'd.

(a) Do the new designs reduce the air drag? If so, which is the better of these designs? Estimate the reduction in retardation by means of a confidence interval.

(b) Do day-to-day differences in conditions affect the initial velocity? The retardation?

(c) If the analysis had been conducted on the observed retardations without the adjustments for initial velocity, would the difference in designs have shown up as significant? What would have been the residual error in this case? What effect would a 25 ft./sec. increase in initial velocity have on the retardation?

7. A feeding trial to compare three diets is carried out using rats. A randomized block pattern is used, with 4 replications. The final weight of each rat (y), its initial weight (u) and the weight of food it consumed (v) are recorded below.

Carry out an analysis of covariance in which adjustment is made for (1) initial weight only; (2) initial weight and amount of food consumed.

In particular, study the contrasts diet 1 vs. diet 2 and diet 2 vs. diet 3. Calculate the adjusted diet averages in both (1) and (2).

Write a short statement of conclusions indicated by the analysis.

	Diet 1			Diet 2			Diet 3		
	u	v	y	u	v	y	u	v	y
Rep 1	209	301	280	204	315	291	189	269	273
Rep 2	190	284	267	175	280	266	199	275	290
Rep 3	192	296	281	179	298	283	187	266	284
Rep 4	178	271	265	186	307	286	181	258	276
Totals	769	1152	1093	744	1200	1126	756	1068	1123

Sums of Squares and Products

	d.f.	(uu)	(uv)	(vv)	(uy)	(vy)	(yy)
Between Reps	3	599.59	490.33	514.00	136.50	226.33	152.33
Between Diets	2	78.17	-141.00	2232.00	-104.25	- 45.00	166.50
Reps x Diets (Error)	6	551.16	634.67	852.00	561.25	667.67	592.17

.../cont'd.

Question #7 cont'd.

Calculation of Regression in Error Row

551.16	634.67	561.25	1747.08
634.67	652.00	667.67	2154.34
561.25	667.67	592.17	1821.09
1	1.15151680	1.01830684	3.16982364
0	121.16683254	21.38119786	142.54803040
0	21.38119600	20.64528605	42.02648205
1	0	.81510925	1.81510925
0	1	.17646081	1.17646081
0	0	16.87234289	16.87234289

Calculation of Regression in Diets + Error Row

629.33	493.67	457.00	1580.00
493.67	3084.00	622.67	4200.34
457.00	622.67	758.67	1838.34
1	.78443742	.72616910	2.51060651
0	2696.74677887	264.18210040	2960.92888421
0	264.18209906	426.81072130	690.99282493
1	0	.64932305	1.64932304
0	1	.09796326	1.09796326
0	0	400.93058164	400.93058614

.../cont'd.

Question #7 cont'd.

Sums of Squares for Regression and Deviations

	Regression		Deviations		
	d.f.	s.s.	d.f.	s.s.	m.s.
Error	1	Not needed and not calculated	5	20.6453	4.13
Diets + Error	1		7	426.8107	
Adjusted Between Diets			2	406.1654	203.08

Sums of Squares and Products

	d.f.	(uu)	(uv)	(vv)	(uy)	(vy)	(yy)
Diet 3 vs. Diet 2	1	18.00	-198.00	2178.00	- 4.50	49.50	1.12
Error	6	551.16	634.67	852.00	561.25	667.67	592.17
Sum	7	569.16	436.67	3030.00	556.75	717.17	593.29

Calculation of Regression in "Sum" Row

569.16	436.67	556.75	1562.58
436.67	3030.00	717.17	4183.84
556.75	717.17	593.29	1867.21
1	.76721836	.97819594	2.74541430
0	2694.97876311	290.02117888	2984.99993762
0	290.02117807	48.67941041	338.70059404
1	0	.89563144	1.89563144
0	1	.10761539	1.10761538
0	0	17.46866821	17.46867371

Sums of Squares of Deviations from Regression

	d.f.	s.s.	m.s.
Error	5	20.6453	4.13
Diet 3 vs. Diet 2 + Error	6	48.6794	
Adjusted Diet 3 vs. Diet 2	1	28.0341	28.03

.../cont'd.

Question #7 cont'd.

Sums of Squares of Deviations from Regression

	d.f.	s.s.	m.s.
Error	4	16.8723	4.22
Diet 3 vs. Diet 2 + Error	5	17.4687	
Adjusted Diet 3 vs. Diet 2	1	0.5964	0.60

Sums of Squares of Deviations from Regression

	d.f.	s.s.	m.s.
Error	4	16.8723	4.22
Diets + Error	6	400.9306	
Adjusted Between Diets	2	384.0583	192.03

8. A randomized block experiment compared three treatments, t_1, t_2, t_3, in three replications r_1, r_2, r_3. One of the measured responses was lost.

The reported values are given in the following table.

	t_1	t_2	t_3
r_1	4	3	6
r_2	3	6	5
r_3	6	7	

Use any "missing value" technique you wish to obtain the numbers indicated by the asterisks in the following analysis of variance table.

	d.f.	s.s.
replications		
treatments	*	*
error	*	*

Are the sums of squares you obtain theoretically correct, or are they approximations to the correct values?

9. The following data represents the gain in weight in lb. of pigs from 5 litters fed on 4 different diets. During the course of the experiment two of the pigs escaped from their pens into the corn bin thus invalidating two of the observations. The remaining observations are:

	Diet	1	2	3	4
Litter	1	175	139	167	138
	2	136	148	157	126
	3	136	163	141	
	4		169	152	132
	5	166	159	156	156

(a) Assuming an additive model, estimate the missing observations.

(b) Given unbiased estimates of diet and litter effects.

(c) Set up an analysis of variance table and test at the 5 percent level the hypothesis that diet effects are all equal. Use both the approximate and the exact test.

10. Two preparations are known to be qualitatively alike, but may differ in potency. The relative potency is to be estimated by administering doses $x_{11}, x_{12}, \ldots, x_{1n}$ of preparation I and doses $x_{21}, x_{22}, \ldots, x_{2n}$ of preparation II, obtaining measured responses

$$y_{11}, y_{12}, \ldots, y_{1n}$$

and

$$y_{21}, y_{22}, \ldots, y_{2n} .$$

It is known that, at zero doses, the response is zero and that, in the neighborhood of zero dose envisaged in this test, regressions of response on dose are acceptably linear.

Develop algebraic expressions for all the quantities needed to estimate the relative potency and to calculate confidence limits for its value.