CRC

HANDBOOK *of tables for* PROBABILITY *and* STATISTICS

SECOND EDITION

EDITOR

WILLIAM H. BEYER, PH.D.

Professor, Department of Mathematics
University of Akron (Ohio)

Published by

THE CHEMICAL RUBBER CO.

18901 Cranwood Parkway, Cleveland, Ohio 44128

Second printing, 1974

Library of Congress Card No. 66-17301

International Standard Book No. 0-87819-692-7

Preface

Statistics is an important component of scientific reasoning, as well as an integral part of academics, business, and technology. As viewed by the late Sir Ronald Fisher, statistics is the key technology of the present day. Practicing statisticians and scientists working in diverse fields need an authoritative reference handbook of statistical tables developed to "aid" in the investigation and solution of many of today's challenging problems. This book has been compiled and arranged to meet the needs of these users of statistics.

This Second Edition of the Handbook of tables for Probability and Statistics brings together in a logically arranged, documented, and readily usable form an extensive collection of relatively standard statistical tables. The general arrangement of the First Edition has been retained. Many of the tables have been expanded and increased in effectiveness. All tables have been corrected of all errors detected. Examples of expanded tables are:

Individual Terms of the Binomial Distribution
Cumulative Terms of the Binomial Distribution
Confidence Limits for Proportions
Tests of Significance in 2 × 2 Contingency Tables
Critical Values for Testing Outliers
Critical Values of U in the Mann-Whitney Test
Distribution of the Total Number-of-Runs Test
Number of Combinations

Included in the expository section of the Handbook (Part I) is a completely rewritten section on descriptive statistics.

Additional tables and graphs which enhance the importance of this Second Edition are:

Summary of Significance Tests
Summary of Confidence Intervals
Table of Signs for Calculating Effects in Factorial Designs up to Six Factors
Operating Characteristic (OC) Curves for Tests on the Mean and Standard Deviation(s) of Normal Distributions
Cochran's Test for the Homogeneity of Variances
Percentage Points of the Maximum F-Ratio
Confidence Limits for σ Based on Mean Range
Critical Values for Duncan's New Multiple Range Test
Critical Values for Rank-Sum Tests for Dispersion
Cumulative Sum Control Charts (CSCC)
Logarithms of the Binomial Coefficients

Preparation of this enlarged Second Edition has been possible only through the participation of recognized authorities who have taken time from their busy schedules to interpret their thoughts into writing. The Editor has been fortunate indeed to secure the aid of a well coordinated and specially selected advisory board. The names of members of the advisory board are presented in the forefront of this handbook. The Editor is most grateful to them for their continued cooperation and for their invaluable contributions.

The Editor gratefully acknowledges the authors, editors, and publishers who gave permission to reproduce these tables. Reference to the sources of material used in this handbook is indicated in the acknowledgment section. It is quite possible that proper credit has not always been given. Regrets and apologies are offered to the authors of such material.

To the many users of the current edition who sent in suggestions for alterations and additions, the Editorial Staff extends a special thanks. It is hoped that those interested will continue to send in suggestions and comments to assist in the continuous improvement of the contents.

William H. Beyer
April, 1968

Advisory Board

Contents

GREEK ALPHABET

Greek letter	Greek name	English equivalent	Greek letter	Greek name	English equivalent
Α α	Alpha	a	Ν ν	Nu	n
Β β	Beta	b	Ξ ξ	Xi	x
Γ γ	Gamma	g	Ο ο	Omicron	ŏ
Δ δ	Delta	d	Π π	Pi	p
Ε ε	Epsilon	ĕ	Ρ ρ	Rho	r
Ζ ζ	Zeta	z	Σ σ ς	Sigma	s
Η η	Eta	ē	Τ τ	Tau	t
Θ θ ϑ	Theta	th	Υ υ	Upsilon	u
Ι ι	Iota	i	Φ φ ϕ	Phi	ph
Κ κ	Kappa	k	Χ χ	Chi	ch
Λ λ	Lambda	l	Ψ ψ	Psi	ps
Μ μ	Mu	m	Ω ω	Omega	ō

I. Probability and Statistics

DESCRIPTIVE STATISTICS

a) Ungrouped Data

The formulas of this section designated as *a)* apply to a random sample of size n, denoted by x_i, $i = 1, 2, \ldots, n$.

b) Grouped Data

The formulas of this section designated as *b)* apply to data grouped into a frequency distribution having class marks x_i, $i = 1, 2, \ldots, k$, and corresponding class frequencies f_i, $i = 1, 2, \ldots, k$. The total number of observations given by

$$n = \sum_{i=1}^{k} f_i$$

In the formulas that follow, c denotes the width of the class interval, x_o denotes one of the class marks taken to be the computing origin, and $u_i = \frac{x_i - x_o}{c}$. Then coded class marks are obtained by replacing the original class marks with the integers $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$ where 0 corresponds to class mark x_o in the original scale.

Mean (Arithmetic Mean)

a) $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

b.1) $$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i = \frac{f_1x_1 + f_2x_2 + \cdots + f_kx_k}{n}$$

If data is coded

b.2) $$\bar{x} = x_o + c\,\frac{\sum_{i=1}^{k} f_i u_i}{n}$$

Weighted Mean (Weighted Arithmetic Mean)

If with each value x_i is associated a weighting factor $w_i \geq 0$, then $\sum_{i=1}^{n} w_i$ is the total weight, and

a) $$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{w_1x_1 + w_2x_2 + \cdots + w_nx_n}{w_1 + w_2 + \cdots + w_n}$$

Geometric Mean

a) $$\text{G.M.} = \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$$

In logarithmic form

$$\log(\text{G.M.}) = \frac{1}{n}\sum_{i=1}^{n} \log x_i = \frac{\log x_1 + \log x_2 + \cdots + \log x_n}{n}$$

b) $$\text{G.M.} = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdots x_k^{f_k}}$$

In logarithmic form

$$\log(\text{G.M.}) = \frac{1}{n}\sum_{i=1}^{k} f_i \log x_i = \frac{f_1 \log x_1 + f_2 \log x_2 + \cdots + f_k \log x_k}{n}$$

Harmonic Mean

a) $$\text{H.M.} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

b) $$\text{H.M.} = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_k}{x_k}}$$

Relation Between Arithmetic, Geometric, and Harmonic Mean

H.M. $\leq$ G.M. $\leq \bar{x}$, (Equality sign holds only if all sample values are identical.)

Mode

a) A mode M_o of a sample of size n is a value which occurs with greatest frequency, i.e., it is the most common value. A mode may not exist, and even if it does exist it may not be unique.

b) $$M_o = L + C\,\frac{\Delta_1}{\Delta_1 + \Delta_2},$$

where L is the lower class boundary of the modal class (class containing the mode),
Δ_1 is the excess of modal frequency over frequency of next lower class,
Δ_2 is the excess of modal frequency over frequency of next higher class.

Median

a) If the sample is arranged in ascending order of magnitude, then the median M_d is given by the $\frac{n+1}{2}$ *nd* value. When n is odd, the median is the middle value of the set of ordered data; when n is even, the median is usually taken as the mean of the two middle values of the set of ordered data.

b) $$M_d = L + c\,\frac{\frac{n}{2} - F_c}{f_m},$$

where L is lower class boundary of median class (class containing the median),
F_c is the sum of the frequencies of all classes lower than the median class,
f_m is the frequency of the median class.

Empirical Relation Between Mean, Median, and Mode

Mean-Mode = 3 (Mean-Median)

Chemical Rubber Co. Cleveland OH 1968.

ANALYSIS OF VARIANCE (ANOVA) TABLES

The analysis of variance (ANOVA) table containing the sum of squares, degrees of freedom, mean square, expectations, etc., present the initial analysis in a compact form. This kind of tabular representation is customarily used to set out the results of analysis of variance calculations. Appropriate ANOVA tables for various experimental design models are presented here. In the tables, the use of "dot notation" indicates a summing over all observations in the population, i.e., when summing over a suffix, that suffix is replaced by a dot. Small letters refer to observations, whereas capital letters refer to observation totals.

ANALYSIS OF VARIANCE AND EXPECTED MEAN SQUARES FOR THE ONE-WAY CLASSIFICATION

Model: $y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n_i)$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Test Statistic
Between groups	$k - 1$	$S_1 = \sum_i n_i(\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_i \left(\frac{Y_{i.}^2}{n_i}\right) - \frac{Y_{..}^2}{n_.}$	$s_1^2 = \frac{S_1}{k - 1}$	$F = \frac{s_1^2}{s_e^2}$
Within groups	$n_. - k$	$S_e = \sum_i \sum_j (\bar{y}_{ij} - \bar{y}_{i.})^2 = \sum_i \sum_j y_{ij}^2 - \sum_i \left(\frac{Y_{i.}^2}{n_i}\right)$	$s_e^2 = \frac{S_e}{n_. - k}$	
Total	$n_. - 1$	$S = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j y_{ij}^2 - \frac{Y_{..}^2}{n_.}$		

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square for: Fixed Model	Expected Mean Square for: Random Model
Between groups	$k - 1$	s_1^2	$\sigma^2 + \frac{\sum_i n_i\alpha_i^2}{k - 1}$	$\sigma^2 + \frac{1}{k - 1}\left(n_. - \frac{\sum n_i^2}{n_.}\right)\sigma_\alpha^2$
Within groups	$n_. - k$	s_e^2	σ^2	σ^2
Total	$n_. - 1$			

Notation:

$$Y_{i.} = \sum_j y_{ij}; \quad Y_{..} = \sum_i \sum_j y_{ij}; \quad \bar{y}_{i.} = \frac{1}{n_i}\sum_j y_{ij} = \frac{1}{n_i} Y_{i.};$$

$$n_. = \sum_i n_i; \quad \bar{y}_{..} = \frac{1}{n_.}\sum_i \sum_j y_{ij} = \frac{Y_{..}}{n_.}$$

σ_a^2 between

σ^2 with

ANALYSIS OF VARIANCE AND EXPECTED MEAN SQUARES FOR THE TWO-WAY CLASSIFICATION WITH ONE OBSERVATION PER CELL

Model: $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad (i = 1, 2, \ldots c;\ j = 1, 2, \ldots, r)$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Test Statistic
Column effects	$c - 1$	$SSC = \frac{\sum_i Y_{i.}^2}{r} - \frac{Y_{..}^2}{cr}$	$s_1^2 = \frac{SSC}{c - 1}$	$\frac{s_1^2}{s_e^2}$
Row effects	$r - 1$	$SSR = \frac{\sum_j Y_{.j}^2}{c} - \frac{Y_{..}^2}{cr}$	$s_2^2 = \frac{SSR}{r - 1}$	$\frac{s_2^2}{s_e^2}$
Error	$(c - 1)(r - 1)$	$SSE = SST - SSC - SSR$	$s_e^2 = \frac{SSE}{(c - 1)(r - 1)}$	
Total	$cr - 1$	$SST = \sum_i \sum_j y_{ij}^2 - \frac{Y_{..}^2}{cr}$		

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square for: Fixed Model	Expected Mean Square for: Mixed Model (α)	Expected Mean Square for: Random Model
Column effects	$c - 1$	s_1^2	$\sigma^2 + r\left(\frac{\sum_i \alpha_i^2}{c - 1}\right)$	$\sigma^2 + r\left(\frac{\sum_i \alpha_i^2}{c - 1}\right)$	$\sigma^2 + r\sigma_\alpha^2$
Row effects	$r - 1$	s_2^2	$\sigma^2 + c\left(\frac{\sum_j \beta_j^2}{r - 1}\right)$	$\sigma^2 + c\sigma_\beta^2$	$\sigma^2 + c\sigma_\beta^2$
Error	$(c - 1)(r - 1)$	s_e^2	σ^2	σ^2	σ^2
Total	$cr - 1$				

$\frac{1}{360}$

ANALYSIS OF VARIANCE AND EXPECTED MEAN SQUARES FOR NESTED CLASSIFICATIONS WITH UNEQUAL SAMPLES

Model: $y_{iju} = \mu + \alpha_i + \delta_{ij} + \epsilon_{iju}$ $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n_i;\ u = 1, 2, \ldots, n_{ij})$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square for Fixed Model (α, δ)
Between main groups	$k - 1$	$S_1 = \sum_i \frac{Y_{i..}^2}{n_{i.}} - \frac{Y_{...}^2}{n_{..}}$	$s_1^2 = \frac{S_1}{k-1}$	$\sigma^2 + \frac{\sum_i n_{i.}\alpha_i^2}{k-1}$
Subgroups within main groups (experimental error)	$\sum_i n_i - k$	$S_2 = \sum_i \sum_j \frac{Y_{ij.}^2}{n_{ij}} - \sum_i \frac{Y_{i..}^2}{n_{i.}}$	$s_2^2 = \frac{S_2}{\sum_i n_i - k}$	$\sigma^2 + \frac{\sum_i \sum_j n_{ij}\delta_{ij}^2}{\sum_i n_i - k}$
Within subgroups (sampling error)	$n_{..} - \sum_i n_i$	$S_e = \sum_i \sum_j \sum_u y_{iju}^2 - \sum_i \sum_j \frac{Y_{ij.}^2}{n_{ij}}$	$s_e^2 = \frac{S_e}{n_{..} - \sum_i n_i}$	σ^2
Total	$n_{..} - 1$	$S = \sum_i \sum_j \sum_u y_{iju}^2 - \frac{Y_{...}^2}{n_{..}}$		

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square for		
			Mixed Model (α)	Mixed Model (δ)	Random Model
Between main groups	$k - 1$	s_1^2	$\sigma^2 + b\sigma_\delta^2 + \frac{\sum_i n_{i.}\alpha_i^2}{k-1}$	$\sigma^2 + c\sigma_\alpha^2$	$\sigma^2 + b\sigma_\delta^2 + c\sigma_\alpha^2$
Experimental error	$\sum_i n_i - k$	s_2^2	$\sigma^2 + a\sigma_\delta^2$	$\sigma^2 + \frac{\sum_i \sum_j n_{ij}\delta_{ij}^2}{\sum_i n_i - k}$	$\sigma^2 + a\sigma_\delta^2$
Sampling error	$n_{..} - \sum_i n_i$	s_e^2	σ^2	σ^2	σ^2
Total	$n_{..} - 1$				

where

$$
\begin{cases}
a = \dfrac{n_{..} - \sum_i \dfrac{\sum_j n_{ij}^2}{n_{i.}}}{\sum_i n_i - k} \\[2ex]
b = \dfrac{\sum_i \dfrac{\sum_j n_{ij}^2}{n_{i.}} - \dfrac{\sum_i \sum_j n_{ij}^2}{n_{..}}}{k - 1} \\[2ex]
c = \dfrac{n_{..} - \dfrac{\sum_i n_{i.}^2}{n_{..}}}{k - 1}
\end{cases}
$$

ANALYSIS OF VARIANCE AND EXPECTED MEAN SQUARES FOR NESTED CLASSIFICATIONS WITH EQUAL SAMPLES

Model: $y_{iju} = \mu + \alpha_i + \delta_{ij} + \epsilon_{iju}$ $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n;\ u = 1, 2, \ldots, r)$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square for Fixed Model (α, δ)
Between main groups	$k - 1$	$S_1 = \sum_i \frac{Y_{i..}^2}{nr} - \frac{Y_{...}^2}{knr}$	$s_1^2 = \frac{S_1}{k-1}$	$\sigma^2 + nr\frac{\sum_i \alpha_i^2}{k-1}$
Experimental error	$k(n - 1)$	$S_2 = \frac{\sum_i \sum_j Y_{ij.}^2}{r} - \frac{\sum_i Y_{i..}^2}{nr}$	$s_2^2 = \frac{S_2}{k(n-1)}$	$\sigma^2 + r\frac{\sum_i \sum_j \delta_{ij}^2}{k(n-1)}$
Sampling error	$kn(r - 1)$	$S_e = \sum_i \sum_j \sum_u y_{iju}^2 - \frac{\sum_i \sum_j Y_{ij.}^2}{r}$	$s_e^2 = \frac{S_e}{kn(r-1)}$	σ^2
Total	$knr - 1$	$S = \sum_i \sum_j \sum_u y_{iju}^2 - \frac{Y_{...}^2}{knr}$		

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square for		
			Mixed Model (α)	Mixed Model (δ)	Random Model
Between main groups	$k - 1$	s_1^2	$\sigma^2 + r\sigma_\delta^2 + nr\left(\frac{\sum \alpha_i^2}{k-1}\right)$	$\sigma^2 + nr\sigma_\alpha^2$	$\sigma^2 + r\sigma_\delta^2 + nr\sigma_\alpha^2$
Experimental error	$k(n - 1)$	s_2^2	$\sigma^2 + r\sigma_\delta^2$	$\sigma^2 + \frac{r\sum_i \sum_j \delta_{ij}^2}{k(n-1)}$	$\sigma^2 + r\sigma_\delta^2$
Sampling error	$kn(r - 1)$	s_e^2	σ^2	σ^2	σ^2
Total	$knr - 1$				

where

$$a = b = r$$
$$c = nr$$

ANALYSIS OF VARIANCE AND EXPECTED MEAN SQUARES FOR A FIXED MODEL TWO-FACTOR FACTORIAL EXPERIMENT IN A ONE-WAY CLASSIFICATION DESIGN

Model: $y_{iju} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{iju}$ $(i = 1, 2, \ldots, c;\ j = 1, 2, \ldots, r;\ u = 1, 2, \ldots, n)$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square for Fixed Model $[\alpha, \beta, (\alpha\beta)]$
Treatment combinations	$cr - 1$	$SSTr$	$s_0^2 = \frac{SSTr}{cr - 1}$	$\sigma^2 + n\frac{\sum_i^c \sum_j^r (\mu_{ij} - \mu)^2}{cr - 1}$
Factor A	$c - 1$	SSA	$s_1^2 = \frac{SSA}{c - 1}$	$\sigma^2 + rn\frac{\sum_i^c \alpha_i^2}{c - 1}$
Factor B	$r - 1$	SSB	$s_2^2 = \frac{SSB}{r - 1}$	$\sigma^2 + cn\frac{\sum_j^r \beta_j^2}{r - 1}$
Interaction	$(c - 1)(r - 1)$	$SSAB = SSTr - SSA - SSB$	$s_3^2 = \frac{SSAB}{(c - 1)(r - 1)}$	$\sigma^2 + n\frac{\sum_i^c \sum_j^r (\alpha\beta)_{ij}^2}{(c - 1)(r - 1)}$
Within (error)	$cr(n - 1)$	$SSW = SST - SSTr$	$s_e^2 = \frac{SSW}{cr(n - 1)}$	σ^2
Total	$crn - 1$	SST		

where

$$SSTr = \frac{\sum_i^c \sum_j^r Y_{ij.}^2}{n} - \frac{Y_{...}^2}{crn} \qquad SSA = \frac{\sum_i^c Y_{i..}^2}{rn} - \frac{Y_{...}^2}{crn}$$

$$SSB = \frac{\sum_j^r Y_{.j.}^2}{cn} - \frac{Y_{...}^2}{crn} \qquad SST = \sum_i^c \sum_j^r \sum_u^n y_{iju}^2 - \frac{Y_{...}^2}{crn}$$

Source of Variation	Mean Square	Expected Mean Square for	
		Random Model	Mixed Model (α)
Factor A	$s_1^2 = \frac{SSA}{c - 1}$	$\sigma^2 + n\sigma_{\alpha\beta}^2 + rn\sigma_\alpha^2$	$\sigma^2 + n\sigma_{\alpha\beta}^2 + rn\frac{\sum_i \alpha_i^2}{c - 1}$
Factor B	$s_2^2 = \frac{SSB}{r - 1}$	$\sigma^2 + n\sigma_{\alpha\beta}^2 + cn\sigma_\beta^2$	$\sigma^2 + cn\sigma_\beta^2$
Interaction	$s_3^2 = \frac{SSAB}{(c - 1)(r - 1)}$	$\sigma^2 + n\sigma_{\alpha\beta}^2$	$\sigma^2 + n\sigma_{\alpha\beta}^2$
Within (error)	$s_e^2 = \frac{SSW}{cr(n - 1)}$	σ^2	σ^2
Total	$s_t^2 = \frac{SST}{crn - 1}$		

ANALYSIS OF VARIANCE AND EXPECTED MEAN SQUARES FOR A THREE-FACTOR FACTORIAL EXPERIMENT IN A COMPLETELY RANDOMIZED DESIGN

Model: $y_{ijku} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijku}$
$(i = 1, 2, \ldots, c;\ j = 1, 2, \ldots, r;\ k = 1, 2, \ldots, l;\ u = 1, 2, \ldots, n)$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square for *Fixed Model*
Factor A	$c - 1$	SSA	s_1^2	$\sigma^2 + rln\frac{\sum_i \alpha_i^2}{c - 1}$
Factor B	$r - 1$	SSB	s_2^2	$\sigma^2 + cln\frac{\sum_j \beta_j^2}{r - 1}$
Factor C	$l - 1$	SSC	s_3^2	$\sigma^2 + crn\frac{\sum_k \gamma_k^2}{l - 1}$
Interaction $A \times B$	$(c - 1)(r - 1)$	$SSAB$	s_4^2	$\sigma^2 + ln\frac{\sum_i \sum_j (\alpha\beta)_{ij}^2}{(c - 1)(r - 1)}$
Interaction $A \times C$	$(c - 1)(l - 1)$	$SSAC$	s_5^2	$\sigma^2 + rn\frac{\sum_i \sum_k (\alpha\beta)_{ik}^2}{(c - 1)(l - 1)}$
Interaction $B \times C$	$(r - 1)(l - 1)$	$SSBC$	s_6^2	$\sigma^2 + cn\frac{\sum_j \sum_k (\beta\gamma)_{jk}^2}{(r - 1)(l - 1)}$
Interaction $A \times B \times C$	$(c - 1)(r - 1)(l - 1)$	$SSABC$	s_7^2	$\sigma^2 + n\frac{\sum_i \sum_j \sum_k (\alpha\beta\gamma)_{ijk}^2}{(c - 1)(r - 1)(l - 1)}$
Within (error)	$crl(n - 1)$	SSE	s_e^2	σ^2
Total	$crln - 1$	SST		

where

$$SST = \sum_i \sum_j \sum_k \sum_u y_{ijku}^2 - \frac{Y_{....}^2}{crln}$$

$$SSA = \frac{\sum_i Y_{i...}^2}{rln} - \frac{Y_{....}^2}{crln}$$

$$SSB = \frac{\sum_j Y_{.j..}^2}{cln} - \frac{Y_{....}^2}{crln}$$

$$SSC = \frac{\sum_k Y_{..k.}^2}{crn} - \frac{Y_{....}^2}{crln}$$

$$SSTr(ABC) = \frac{\sum_i \sum_j \sum_k Y_{ijk.}^2}{n} - \frac{Y_{....}^2}{crln}$$

$$SSTr(AB) = \frac{\sum_i \sum_j Y_{ij..}^2}{ln} - \frac{Y_{....}^2}{crln}$$

$$SSTr(AC) = \frac{\sum_i \sum_k Y_{i.k.}^2}{rn} - \frac{Y_{....}^2}{crln}$$

$$SSTr(BC) = \frac{\sum_j \sum_k Y_{.jk.}^2}{cn} - \frac{Y_{....}^2}{crln}$$

$$SSAB = SSTr(AB) - SSA - SSB$$
$$SSAC = SSTr(AC) - SSA - SSC$$
$$SSBC = SSTr(BC) - SSB - SSC$$
$$SSABC = SSTr(ABC) - SSA - SSB - SSC - SSAB - SSAC - SSBC$$
$$SSE = SST - SSTr(ABC)$$

Source of Variation	Mean Square	Expected Mean Square for the		
		Random Model	Mixed Model (α)	Mixed Model (α, β)
Factor A	$s_1^2 = \frac{SSA}{c-1}$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + ln\sigma_{\alpha\beta}^2 + rn\sigma_{\alpha\gamma}^2 + rln\sigma_{\alpha}^2$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + ln\sigma_{\alpha\beta}^2 + rn\sigma_{\alpha\gamma}^2 + rln\frac{\sum \alpha_i^2}{c-1}$	$\sigma^2 + rn\sigma_{\alpha\gamma}^2 + rln\frac{\sum_i \alpha_i^2}{c-1}$
Factor B	$s_2^2 = \frac{SSB}{r-1}$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + ln\sigma_{\alpha\beta}^2 + cn\sigma_{\beta\gamma}^2 + cln\sigma_{\beta}^2$	$\sigma^2 + cn\sigma_{\beta\gamma}^2 + cln\sigma_{\beta}^2$	$\sigma^2 + cn\sigma_{\beta\gamma}^2 + cln\frac{\sum_j \beta_j^2}{r-1}$
Factor C	$s_3^2 = \frac{SSC}{l-1}$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + rn\sigma_{\alpha\gamma}^2 + cn\sigma_{\beta\gamma}^2 + crn\sigma_{\gamma}^2$	$\sigma^2 + cn\sigma_{\beta\gamma}^2 + crn\sigma_{\gamma}^2$	$\sigma^2 + crn\sigma_{\gamma}^2$
$A \times B$	$s_4^2 = \frac{SSAB}{(c-1)(r-1)}$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + ln\sigma_{\alpha\beta}^2$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + ln\sigma_{\alpha\beta}^2$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + ln\frac{\sum_i \sum_j (\alpha\beta)_{ij}^2}{(c-1)(r-1)}$
$A \times C$	$s_5^2 = \frac{SSAC}{(c-1)(l-1)}$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + rn\sigma_{\alpha\gamma}^2$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + rn\sigma_{\alpha\gamma}^2$	$\sigma^2 + rn\sigma_{\alpha\gamma}^2$
$B \times C$	$s_6^2 = \frac{SSBC}{(r-1)(l-1)}$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2 + cn\sigma_{\beta\gamma}^2$	$\sigma^2 + cn\sigma_{\beta\gamma}^2$	$\sigma^2 + cn\sigma_{\beta\gamma}^2$
$A \times B \times C$	$s_7^2 = \frac{SSABC}{(c-1)(r-1)(l-1)}$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2$	$\sigma^2 + n\sigma_{\alpha\beta\gamma}^2$
Within (error)	$s_e^2 = \frac{SSE}{crl(n-1)}$	σ^2	σ^2	σ^2
Total	$s_0^2 = \frac{SST}{crln-1}$			

ANALYSIS OF VARIANCE AND EXPECTED MEAN SQUARES FOR A $t \times t$ LATIN SQUARE

Model: $y_{ij(k)} = \mu + \alpha_i + \beta_j + \gamma_{(k)} + \epsilon_{ij(k)}$ $(i = 1, 2, \ldots, t; j = 1, 2, \ldots, t; k = 1, 2, \ldots, t)$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square for Fixed Model
Columns	$t-1$	$SSC = \frac{\sum_i Y_{i..}^2}{t} - \frac{Y_{...}^2}{t^2}$	$s_1^2 = \frac{SSC}{t-1}$	$\sigma^2 + t\frac{\sum_i \alpha_i^2}{t-1}$
Rows	$t-1$	$SSR = \frac{\sum_j Y_{.j.}^2}{t} - \frac{Y_{...}^2}{t^2}$	$s_2^2 = \frac{SSR}{t-1}$	$\sigma^2 + t\frac{\sum_j \beta_j^2}{t-1}$
Treatments	$t-1$	$SSTr = \frac{\sum_k Y_{..(k)}^2}{t} - \frac{Y_{...}^2}{t^2}$	$s_3^2 = \frac{SSTr}{t-1}$	$\sigma^2 + t\frac{\sum_k \gamma_k^2}{t-1}$
Error	$(t-1)(t-2)$	$SSE = SST - SSC - SSR - SSTr$	$s_e^2 = \frac{SSE}{(t-1)(t-2)}$	σ^2
Total	$t^2 - 1$	$SST = \sum_i \sum_j y_{ij(k)}^2 - \frac{Y_{...}^2}{t^2}$		

Source of Variation	Mean Square	Expected Mean Square for		
		Random Model	Mixed Model (γ)	Mixed Model (α, γ)
Columns	$s_1^2 = \frac{SSC}{t-1}$	$\sigma^2 + t\sigma_{\alpha}^2$	$\sigma^2 + t\sigma_{\alpha}^2$	$\sigma^2 + t\frac{\sum_i \alpha_i^2}{t-1}$
Rows	$s_2^2 = \frac{SSR}{t-1}$	$\sigma^2 + t\sigma_{\beta}^2$	$\sigma^2 + t\sigma_{\beta}^2$	$\sigma^2 + t\sigma_{\beta}^2$
Treatments	$s_3^2 = \frac{SSTr}{t-1}$	$\sigma^2 + t\sigma_{\gamma}^2$	$\sigma^2 + t\frac{\sum_k \gamma_k^2}{t-1}$	$\sigma^2 + t\frac{\sum_k \gamma_k^2}{t-1}$
Error	$s_e^2 = \frac{SSE}{(t-1)(t-2)}$	σ^2	σ^2	σ^2

ANALYSIS OF VARIANCE FOR A GRAECO-LATIN SQUARE

Model: $y_{ijuk} = \mu + \alpha_i + \beta_j + \gamma_u + \delta_k + \epsilon_{ijuk}$ $\quad (i, j, u, k = 1, 2, \ldots, n)$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Factor I (Rows)	$n - 1$	$S_1 = \frac{\sum_i Y_{i...}^2}{n} - \frac{Y_{....}^2}{n^2}$	$s_1^2 = \frac{S_1}{n-1}$
Factor II (Columns)	$n - 1$	$S_2 = \frac{\sum_j Y_{.j..}^2}{n} - \frac{Y_{....}^2}{n^2}$	$s_2^2 = \frac{S_2}{n-1}$
Factor III (Latin Letters)	$n - 1$	$S_3 = \frac{\sum_u Y_{..u.}^2}{n-1} - \frac{Y_{....}^2}{n^2}$	$s_3^2 = \frac{S_3}{n-1}$
Factor IV (Greek Letters)	$n - 1$	$S_4 = \frac{\sum_k Y_{...k}^2}{n} - \frac{Y_{....}^2}{n^2}$	$s_4^2 = \frac{S_4}{n-1}$
Residual	$(n-1)(n-3)$	$S_e =$ difference	$s_e^2 = \frac{S_e}{(n-1)(n-3)}$
Total	$n^2 - 1$	$S = \sum_i \sum_j y_{ijuk}^2 - \frac{Y_{....}^2}{n^2}$	

ANALYSIS OF VARIANCE FOR A YOUDEN SQUARE

Model: $y_{iju} = \mu + \alpha_i + \beta_j + \gamma_u + \epsilon_{iju}$
$(i = 1, 2, \ldots, b; j = 1, 2, \ldots, t(= b); u = 1, 2, \ldots, k(<t))$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Blocks (crude)		$S_1 = \sum_i \frac{Y_{i..}^2}{k} - \frac{Y_{...}^2}{bk}$	
Treatments (adjusted)	$t - 1$	$S_2 = \frac{t-1}{bk^2(k-1)} \sum_j \left(kY_{.j.}^2 - \sum_{i(j)} Y_{i..}^2\right)$	$s_1^2 = \frac{S_2}{t-1}$
Treatments (crude)		$S_3 = \sum_j \frac{Y_{.j.}^2}{r} - \frac{Y_{...}^2}{tr}$	
Blocks (adjusted)	$b - 1$	$S_4 = \frac{b-1}{bk^2(k-1)} \sum_i \left(rY_{i..}^2 - \sum_{j(i)} Y_{.j.}^2\right)$	$s_2^2 = \frac{S_4}{b-1}$
Factor II (γ)	$k - 1$	$S_5 = \frac{\sum_u Y_{..u}^2}{k} - \frac{Y_{...}^2}{bk}$	$s_3^2 = \frac{S_5}{k-1}$
Residual	$bk - t - b - k + 2$	$S_e = S - (S_1 + S_2 + S_5) = S - (S_3 + S_4 + S_5)$	$s_e^2 = \frac{S_e}{bk - t - b - k + 2}$
Total	$bk - 1$	$S = \sum_i \sum_j y_{iju}^2 - \frac{Y_{...}^2}{bk}$	

(Note that $S_1 + S_2 = S_3 + S_4$)

ANALYSIS OF VARIANCE FOR BALANCED INCOMPLETE BLOCK (BIB)

Model: $y_{iju} = \mu + \alpha_i + \beta_j + \epsilon_{iju}$ $\quad (i = 1, 2, \ldots, b; j = 1, 2, \ldots, t; u = n_{ij})$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Blocks	$b - 1$	$S_1 = \frac{\sum_i Y_{i..}^2}{k} - \frac{Y_{...}^2}{bk}$	$s_1^2 = \frac{S_1}{b-1}$
Treatments (adjusted)	$t - 1$	$S_2 = \frac{t-1}{bk^2(k-1)} \sum_j \left[kY_{.j.} - \sum_{i(j)} Y_{i..}\right]^2$	$s_2^2 = \frac{S_2}{t-1}$
Residual	$bk - t - b + 1$	$S_e =$ difference	$s_e^2 = \frac{S_e}{bk - t - b + 1}$
Total	$bk - 1$	$S = \sum_i \sum_j y_{iju}^2 - \frac{Y_{...}^2}{bk}$	

where

t = number of treatment levels
b = number of blocks
k = number of treatment levels per block
r = number of replications of each treatment level
λ = number of blocks in which any given pair of treatment levels appear together
$bk = tr$
$r(k-1) = \lambda(t-1)$

GENERAL LINEAR MODEL

by Dr. Rolf E. Bargmann

1. NOTATION

A matrix will be denoted by bold-face capital letters, e.g., if **A** has m rows and n columns, we may often specify $\mathbf{A} = \mathbf{A}(m \times n)$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

$\mathbf{A}'$ denotes the transpose of **A**.
$(\mathbf{A})_{ij} = a_{ij}$ denotes the element in the i'th row and j'th column of **A**.
I denotes the identity matrix.
$\mathbf{D}_a$ denotes a diagonal matrix. The subscript indicates the terms in the diagonal.
$\tilde{\mathbf{T}}$, $\tilde{\mathbf{U}}$, i.e., any matrix with a tilde (~) above it will denote a lower triangular matrix.

A column vector will be denoted in general, by a lower-case bold-face letter, e.g.,

$$\mathbf{x}, \mathbf{y}, \mathbf{j}, \boldsymbol{\beta}, \boldsymbol{\xi}; \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_p \end{bmatrix}.$$

Occasionally, capital bold-face letters represent column vectors. Examples are as follows:

T (vector of treatment totals)
B (vector of Block totals).

A lower case letter with a prime denotes a row vector, e.g.,

$$\mathbf{x}' = [x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_p] .$$

2. THE GENERAL LINEAR MODEL

2.1. The Simple Regression Model

$$y_i = \alpha + \beta x_i + e_i ,$$

where x_i is a fixed concomitant variable whose values are assumed to be known before an experiment is performed and is not subject to chance. Let E denote the expectation operator, var and cov the (population) variances and covariances, respectively.

$E(e_i) = 0$, var $(e_i) = \sigma^2$, cov $(e_i,e_j) = 0$, i.e., $E(y_i) = \alpha + \beta x_i$.

If we write

$$\begin{aligned} y_1 &= \alpha + \beta x_1 + e_1 \\ y_2 &= \alpha + \beta x_2 + e_2 \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \\ y_n &= \alpha + \beta x_n + e_n , \end{aligned}$$

we can write, in matrix form

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \cdot \\ e_n \end{bmatrix}$$

$$\mathbf{y} = \mathbf{A} \quad \boldsymbol{\xi} + \mathbf{e} ,$$

where **A** is the design matrix and can be also written in the form

$$\mathbf{A} = [\underset{(1)}{\mathbf{j}} , \quad \underset{(1)}{\mathbf{x}}](n).$$

The numbers in parentheses denote the order of the matrix.

j is a column vector containing all ones.
x is a column vector of all concomitant observations.

The simple regression model is frequently written in the form

$$y_i = \mu + \beta(x_i - \bar{x}) + e_i ,$$

where $\mu = \alpha + \beta\bar{x}$.

This, too, can be written in the general linear model form

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & (x_1 - \bar{x}) \\ 1 & (x_2 - \bar{x}) \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & (x_n - \bar{x}) \end{bmatrix} \begin{bmatrix} \mu \\ \beta \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{bmatrix}$$

$$\mathbf{y} = \mathbf{A} \quad \boldsymbol{\xi} + \mathbf{e} ,$$

where $\mathbf{A} = [\mathbf{j}, (\mathbf{x} - \bar{x}\mathbf{j})]$.

2.2. Multiple Regression Model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots \beta_k x_{ki} + e_i .$$

Assumption: $E(e_i) = 0$,

$$\text{var}\ (y_i) = \text{var}\ (e_i) = \sigma^2, \qquad \text{cov}\ (y_i,y_j) = 0 .$$

If we write

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \beta_3 x_{31} + \cdots \beta_k x_{k1} + e_1 \\ y_2 &= \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \beta_3 x_{32} + \cdots \beta_k x_{k2} + e_2 \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \\ y_n &= \beta_0 + \beta_n x_{1n} + \beta_2 x_{2n} + \beta_3 x_{3n} + \cdots \beta_k x_{kn} + e_n , \end{aligned}$$

we can write

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & x_{32} & \cdots & x_{k2} \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & x_{1n} & x_{2n} & x_{3n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \cdot \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \cdot \\ e_n \end{bmatrix}$$

$$\mathbf{y} = \mathbf{A} \quad \boldsymbol{\xi} + \mathbf{e} ,$$

where $\mathbf{A} = [\underset{(1)}{\mathbf{j}} , \quad \underset{(k)}{\mathbf{X}}](n)$, and **X** denotes the matrix of all observations on all concomitant variables.

2.3. One-way Classification Analysis of Variance

Model

$$y_{ij} = \mu + \tau_i + e_{ij}$$

where μ = general effect
τ_i = treatment effects
e_{ij} = experimental error,
v treatments with effects $\tau_1, \tau_2, \ldots, \tau_v$; $j = 1, 2, 3, \ldots n_i$.

$$\begin{array}{llllll} y_{11} & = \mu + \tau_1 & & & & + e_{11} \\ y_{12} & = \mu + \tau_1 & & & & + e_{12} \\ \cdot & \cdot & & & & \cdot \\ y_{1n_1} & = \mu + \tau_1 & & & & + e_{1n_1} \\ \hline y_{21} & = \mu & + \tau_2 & & & + e_{21} \\ \cdot & \cdot & \cdot & & & \cdot \\ y_{2n_2} & = \mu & + \tau_2 & & & + e_{2n_2} \\ \cdot & \cdot & \cdot & & & \cdot \\ \hline y_{v1} & = \mu & & & + \tau_v & + e_{v1} \\ \cdot & \cdot & & & \cdot & \cdot \\ y_{vn_v} & = \mu & & & + \tau_v & + e_{vn_v} . \end{array}$$

We can again write this

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \cdot \\ y_{1n_1} \\ y_{21} \\ \cdot \\ y_{2n_2} \\ \cdot \\ y_{v1} \\ y_{v2} \\ \cdot \\ y_{vn_v} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \hline 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \hline \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \hline 1 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 1 & 0 & 0 & 0 & & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \cdot \\ \tau_v \end{bmatrix} + \mathbf{e}$$

$$\mathbf{y} \quad = \quad \mathbf{A} \quad \xi \quad + \mathbf{e}\,.$$

The design matrix **A** can be written

$$\mathbf{A} = [\underset{(1)}{\mathbf{j}},\ \underset{(v)}{\mathbf{A}_r}](n),$$

where

$$\mathbf{A}_r = \begin{matrix} (n_1) \\ (n_2) \\ \cdot \\ (n_v) \end{matrix} \begin{bmatrix} \mathbf{j} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{j} & \mathbf{0} & \cdots & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{j} \end{bmatrix}$$

and the parameter vector can also be written

$$\xi = \begin{bmatrix} \mu \\ \tau \end{bmatrix}.$$

2.4. Two-way Classification (Two Factors Factorial)

Model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + e_{ijk}\,,$$

where μ = general effect

α_i = factor *A* effects (usually row effects)

β_j = factor *B* effects (usually column effects)

δ_{ij} = interaction effects.

For example:

$$\begin{bmatrix} \mathbf{y}_{11} \\ \mathbf{y}_{12} \\ \mathbf{y}_{13} \\ \mathbf{y}_{21} \\ \mathbf{y}_{22} \\ \mathbf{y}_{23} \end{bmatrix} = \begin{matrix} (n_{11}) \\ (n_{12}) \\ (n_{13}) \\ (n_{21}) \\ (n_{22}) \\ (n_{23}) \end{matrix} \begin{bmatrix} \mathbf{j} & \mathbf{j} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{j} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{j} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{j} & \mathbf{0} & \mathbf{j} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{0} \\ \mathbf{j} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{j} & \mathbf{0} \\ \mathbf{j} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{j} \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \delta_{11} \\ \delta_{12} \\ \delta_{13} \\ \delta_{21} \\ \delta_{22} \\ \delta_{23} \end{bmatrix} + \mathbf{e}$$

$$\mathbf{y} \quad = \quad \mathbf{A} \quad \xi \quad + \mathbf{e}\,.$$

2.5. Analysis of Covariance

Analysis of covariance is equivalent to analysis of variance with one or more concomitant variables added.

For simplicity, let us take one-way classification and one concomitant variable. Model

$$y_{ij} = \mu + \tau_i + \beta x_{ij} + e_{ij}\,.$$

As a vector equation, this model reads

$$\begin{array}{lllll} y_{11} = \mu + \tau_1 & & & + \beta x_{11} & + e_{11} \\ y_{12} = \mu + \tau_1 & & & + \beta x_{12} & + e_{12} \\ \cdot \qquad \cdot \quad \cdot & & & \cdot & \cdot \\ y_{1n_1} = \mu + \tau_1 & & & + \beta x_{1n_1} & + e_{1n_1} \\ \hline y_{21} = \mu & + \tau_2 & & + \beta x_{21} & + e_{21} \\ \cdot \qquad \cdot & \cdot & & \cdot & \cdot \\ y_{2n_2} = \mu & + \tau_2 & & + \beta x_{2n_2} & + e_{2n_2} \\ \hline y_{31} = \mu & & + \tau_3 & + \beta x_{31} & + e_{31} \\ \cdot \qquad \cdot & & \cdot & \cdot & \cdot \\ y_{3n_3} = \mu & & + \tau_3 & + \beta x_{3n_3} & + e_{3n_3} \end{array}$$

Let $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ denote the concomitant observations in each group, then

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \cdot \\ \cdot \\ y_{1n_1} \\ y_{21} \\ \cdot \\ \cdot \\ \cdot \\ y_{2n_2} \\ y_{31} \\ \cdot \\ \cdot \\ y_{3n_3} \end{bmatrix} = \begin{bmatrix} \mathbf{j} & \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{x}_1 \\ \mathbf{j} & \mathbf{0} & \mathbf{j} & \mathbf{0} & \mathbf{x}_2 \\ \mathbf{j} & \mathbf{0} & \mathbf{0} & \mathbf{j} & \mathbf{x}_3 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \beta \end{bmatrix} + \mathbf{e}$$

$$\mathbf{y} \quad = \quad \mathbf{A} \quad \xi \quad + \mathbf{e}\,.$$

With the above illustrations, it is clear that we can write a great variety of models in the general linear model form

$$\mathbf{y} = \mathbf{A}\xi + \mathbf{e}\,.$$

3. SUMMARY OF RULES FOR MATRIX OPERATIONS

3.1. Let $E(\mathbf{y}) = \boldsymbol{\psi}$, var $(\mathbf{y}) = \boldsymbol{\Sigma}$, a symmetric matrix containing all possible variances and covariances. Then, $E(\mathbf{My}) = \mathbf{M}\boldsymbol{\psi}$, var $(\mathbf{My}) = \mathbf{M\Sigma M'}$, for any conforming matrix **M**. $E(\mathbf{y'M}) = \boldsymbol{\psi}'\mathbf{M}$, var $(\mathbf{y'M}) = \mathbf{M'\Sigma M}$.

3.2. Partitioning of Determinants

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}|\,|\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}| \text{ if } \mathbf{A}^{-1} \text{ exists}\,.$$

$$= |\mathbf{D}|\,|\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}| \text{ if } \mathbf{D}^{-1} \text{ exists}\,.$$

3.3. Inverse of a Partitioned Matrix

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{U} \end{bmatrix}$$

$$\begin{aligned} \text{where } \mathbf{X} &= [\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}]^{-1} \\ \mathbf{U} &= [\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}]^{-1} \\ \mathbf{Y} &= -\mathbf{A}^{-1}\mathbf{B}\mathbf{U} \\ \mathbf{Z} &= -\mathbf{D}^{-1}\mathbf{C}\mathbf{X}\ . \end{aligned}$$

3.3.1. *Symmetric Case*

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ (\mathbf{A}^{12})' & \mathbf{A}^{22} \end{bmatrix}$$

$$\begin{aligned} \text{where } \mathbf{A}^{11} &= [\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}]^{-1} \\ \mathbf{A}^{22} &= [\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}]^{-1} \\ \mathbf{A}^{12} &= -\mathbf{A}^{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \text{or } \mathbf{A}^{12} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}^{22}\ . \end{aligned}$$

Computational steps: Order the sets in such a way that $\mathbf{A}_{22}$ is the smaller matrix.

a. Obtain $\mathbf{A}_{22}^{-1}$
b. Multiply $\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$
c. Obtain $\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}$
d. Obtain $\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}$
e. Invert the matrix in *d.*; thus obtain $\mathbf{A}^{11}$
f. Obtain $\mathbf{A}^{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$ by multiplying matrices from steps *e.* and *b.*
g. Change all signs in *f.*; thus obtain $\mathbf{A}^{12}$.
h. Obtain $(\mathbf{A}^{12})'$ by transposing *g.*
i. Obtain $\mathbf{A}'_{12}\mathbf{A}^{12}$, the latter factor from *g.*
j. Obtain $\mathbf{I} - \mathbf{A}'_{12}\mathbf{A}^{12}$, i.e., change all signs in off-diagonal elements of the matrix in step *i.* and complement diagonal elements to 1.
k. Obtain $\mathbf{A}_{22}^{-1}[\mathbf{I} - \mathbf{A}'_{12}\mathbf{A}^{12}]$, i.e., premultiply the matrix in *j.* by $\mathbf{A}_{22}^{-1}$ obtained in step *a.* This is $\mathbf{A}^{22}$.

3.4. Characteristic Roots

a. ch (**AB**) = ch (**BA**) except, possibly, for zero roots.
b. Corollary: tr (**AB**) = tr (**BA**).
c. If ch (**A**) = λ_i, ch ($\mathbf{A}^{-1}$) = $1/\lambda_i$, and ch ($\mathbf{I} \pm \mathbf{A}$) = $1 \pm \lambda_i$.

3.5. Differentiation

3.5.1. *Definitions:*

Let f be a scalar function of $x_1, x_2, \ldots x_p$.
Then $\partial f/\partial \mathbf{x}$ denotes a column vector whose *i*th element is $\partial f/\partial x_i$.

Let f be a scalar function of $x_{11}, x_{12} \ldots x_{1q}, x_{21}, x_{22} \ldots x_{2q} \ldots x_{p1}, x_{p2}, \ldots x_{pq}$. Then $\partial f/\partial \mathbf{X}$ denotes a matrix whose (i, j) element is $\partial f/\partial x_{ij}$. Note that, in this definition, x_{ij} denotes the element in the *i*'th row and *j*'th column of **X**. If there are any functional relations between the elements of **X** (as, for instance, in a symmetric matrix) these relations will be *disregarded* in the above definition. In other words, x_{ij} denotes the variable in the *i*'th row and *j*'th column of **X**, and x_{ji} denotes that in the *j*'th row and *i*'th column. If the two happen to be identical, a new symbol will be in order. For example, if $x_{ij} = x_{ji} = y_{ij}$, say, $\partial f/\partial y_{ij} = \partial f/\partial x_{ij} \cdot \partial x_{ij}/\partial y_{ij} + \partial f/\partial x_{ji} \cdot \partial x_{ji}/\partial y_{ij} = \partial f/\partial x_{ij} + \partial f/\partial x_{ji} = (\partial f/\partial \mathbf{X})_{ij} + (\partial f/\partial \mathbf{X})_{ji}$. Here, y_{ij} is the symbol for that distinct variable which occurs in two places in **X**.

If $y_1, y_2, \ldots y_p$ are functions of x, $\partial \mathbf{y}'/\partial x$ denotes the row vector whose *i*'th element is $\partial y_i/\partial x$.

If $y_{11}, y_{12}, \ldots y_{1q}, y_{21}, \ldots y_{2q}, \ldots y_{p1}, \ldots y_{pq}$ are functions of x, $\partial \mathbf{Y}/\partial x$ denotes the matrix whose (i,j) element is $\partial y_{ij}/\partial x$.

If each of the quantities $y_1, y_2, \ldots y_q$ is a function of the variables $x_1, x_2, \ldots x_p$, $\partial \mathbf{y}'/\partial \mathbf{x}$ denotes a matrix of order (pxq) whose (i,j) element is $\partial y_j/\partial x_i$. *Note the interchange of subscripts.*

3.5.2. *Rules:*

1. $\partial(\mathbf{x}'\mathbf{x})/\partial \mathbf{x} = 2\mathbf{x}$
2. $\partial(\mathbf{x}'\mathbf{Q}\mathbf{x})/\partial \mathbf{x} = \mathbf{Q}\mathbf{x} + \mathbf{Q}'\mathbf{x}$
3. $\partial(\mathbf{x}'\mathbf{Q}\mathbf{x})/\partial \mathbf{x} = 2\mathbf{Q}\mathbf{x}$ if **Q** is symmetric.
4. $\partial(\mathbf{a}'\mathbf{x})/\partial \mathbf{x} = \mathbf{a}$
5. $\partial(\mathbf{a}'\mathbf{Q}\mathbf{x})/\partial \mathbf{x} = \mathbf{Q}'\mathbf{a}$
6. $\partial\ \mathrm{tr}\ (\mathbf{A}\mathbf{X})/\partial \mathbf{X} = \mathbf{A}'$
7. $\partial\ \mathrm{tr}\ (\mathbf{X}\mathbf{A})/\partial \mathbf{X} = \mathbf{A}'$
8. $\partial \log |\mathbf{X}|/\partial \mathbf{X} = (\mathbf{X}')^{-1}$, if **X** is square and nonsingular.
9. "Chain Rule No. 1": $\partial \mathbf{y}'/\partial \mathbf{x} = \partial \mathbf{z}'/\partial \mathbf{x} \cdot \partial \mathbf{y}'/\partial \mathbf{z}$.
10. $\partial(\mathbf{x}'\mathbf{A})/\partial \mathbf{x} = \mathbf{A}$
11. If $\mathbf{e} = \mathbf{b} - \mathbf{A}'\mathbf{x}$, $\partial(\mathbf{e}'\mathbf{e})/\partial \mathbf{x} = \partial \mathbf{e}'/\partial \mathbf{x} \cdot \partial(\mathbf{e}'\mathbf{e})/\partial \mathbf{e}$ (according to rule 9), $= -2\mathbf{A}'\mathbf{e}$ (by rules 10 and 1).
12. "Chain Rule No. 2": If the scalar z is related to a scalar x through variables y_{ij} ($i = 1, 2, \cdots p; j = 1, 2 \cdots q$),

$$\partial z/\partial x = \mathrm{tr}\ [\partial z/\partial \mathbf{Y} \cdot \partial \mathbf{Y}'/\partial x]$$

or

$$\partial z/\partial x = \mathrm{tr}\ [\partial z/\partial \mathbf{Y}' \cdot \partial \mathbf{Y}/\partial x]$$

This chain rule is correct regardless of any functional relationships which may exist between the elements of **Y**.

3.6. Some Additional Definitions and Rules

j denotes a column vector, each element of which is 1. Hence **j'A** is a row vector whose elements are the column sums of **A**, and **Aj** denotes a column vector whose elements are the row sums of **A**. **j'Aj** denotes the sum of all elements in the matrix **A**.

I denotes the identity matrix. If the order must be stated it will be added in parentheses. Hence $\mathbf{I}(p)$ denotes a $(p \times p)$ identity matrix.

If a tilde (~) is placed above a matrix, the matrix is assumed to be triangular. For definiteness, the untransposed matrix

$$\tilde{\mathbf{T}} = \begin{bmatrix} t_{11} & 0 & 0 & \cdots & 0 \\ t_{21} & t_{22} & 0 & \cdots & 0 \\ t_{31} & t_{32} & t_{33} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ t_{p1} & t_{p2} & t_{p3} & \cdots & t_{pp} \end{bmatrix}$$

is a "lower" triangular matrix, whereas the transposed matrix

$$\tilde{\mathbf{T}}' = \begin{bmatrix} t_{11} & t_{21} & t_{31} & \cdots & t_{p1} \\ 0 & t_{22} & t_{32} & \cdots & t_{p2} \\ 0 & 0 & t_{33} & \cdots & t_{p3} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & t_{pp} \end{bmatrix}$$

is an "upper" triangular matrix.

If $\tilde{\mathbf{T}}$ is lower triangular, so is $\tilde{\mathbf{T}}^{-1}$.

If $\mathbf{Q}$ is a symmetric, positive-definite matrix, we can find, uniquely, a real matrix $\tilde{\mathbf{T}}$, such that $\mathbf{Q} = \tilde{\mathbf{T}}\tilde{\mathbf{T}}'$, provided we let the diagonal elements of $\tilde{\mathbf{T}}$ be positive. This matrix and its inverse can be readily obtained from the forward Doolittle solution. If, in each cycle, we divide each element of the next-to-last row (the row which is immediately above the one beginning with unity) by the square-root of the "leading" (first) element, we obtain $\tilde{\mathbf{T}}'$ on the left and $\tilde{\mathbf{T}}^{-1}$ on the right-hand side.

If $\mathbf{Q}$ is a $(p \times p)$, symmetric, positive-semidefinite matrix of rank r, the matrix $\tilde{\mathbf{T}}$ obtained in the above manner will have zeros to the right of the r'th column. $\mathbf{Q}$ can then be represented as

$$\mathbf{Q} = \begin{matrix} (r) \\ (p-r) \end{matrix} \overset{(r)}{\begin{bmatrix} \tilde{\mathbf{T}}_1 \\ \mathbf{T}_2 \end{bmatrix}} [\underset{(r)}{\tilde{\mathbf{T}}_1'} \; , \; \underset{(p-r)}{\mathbf{T}_2'}],$$

where, of course, only $\tilde{\mathbf{T}}_1$ is triangular. This is an important computational device in connection with rule 3.4.a. on characteristic roots. For, if the largest root of **AB** is desired, where both **A** and **B** are symmetric, but **A** is of low rank, we can obtain the representation

$$\mathbf{A} = \begin{matrix} (r) \\ (p-r) \end{matrix} \underset{(r)}{\begin{bmatrix} \tilde{\mathbf{T}}_1 \\ \mathbf{T}_2 \end{bmatrix}} [\tilde{\mathbf{T}}_1' \; , \; \mathbf{T}_2']$$

by the forward Doolittle solution. Then, by 3.4.a., $\text{ch}\,(\mathbf{AB}) = \text{ch}\left([\tilde{\mathbf{T}}_1' \, , \, \mathbf{T}_2']\mathbf{B}\begin{bmatrix} \tilde{\mathbf{T}}_1 \\ \mathbf{T}_2 \end{bmatrix}\right)$, and the matrix in parentheses is of small order and symmetric.

$\mathbf{D}_u$ denotes a diagonal matrix whose non-zero elements are $u_1, u_2 \cdots u_p$.

4. PRINCIPLE OF MINIMIZING QUADRATIC FORMS AND GAUSS MARKOV THEOREM

4.1. Some Remarks on Multivariate Distributions

In univariate situation, suppose we have a random variable x, such that

$$E(x) = \mu$$
$$\text{var}\,(x) = \sigma^2 \, .$$

If we want to find a random variable y, such that

$$E(y) = 0 \quad \text{and} \quad \text{var}\,(y) = 1 \, ,$$

i.e., we are to find y such that it has mean 0 and variance 1, we perform the "standardization"

$$y = \frac{x - \mu}{\sigma} \, .$$

We also recall that, if y is normally distributed,

$$y^2 = \frac{(x-\mu)^2}{\sigma^2} = \chi^2 \text{ with 1 d.f .}$$

In multivariate situations, we have random variables $\mathbf{x}$ such that

$$E(\mathbf{x}) = \boldsymbol{\mu} \quad \text{and} \quad \text{var}\,(\mathbf{x}) = \Sigma$$

and we wish to find $\mathbf{y}$ such that

$$E(\mathbf{y}) = \mathbf{0} \quad \text{and} \quad \text{var}\,(\mathbf{y}) = \mathbf{I} \, .$$

To obtain this, we will proceed as follows: Let

$$\boldsymbol{\Sigma} = \tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\Gamma}}',$$

where $\tilde{\boldsymbol{\Gamma}}$ is a lower triangular matrix, which, given $\boldsymbol{\Sigma}$, can be obtained conveniently as a by-product of the forward Doolittle analysis. Then,

$$\boldsymbol{\Sigma}^{-1} = (\tilde{\boldsymbol{\Gamma}}')^{-1}\tilde{\boldsymbol{\Gamma}}^{-1} \, .$$

Now, let

$$y = \tilde{\boldsymbol{\Gamma}}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$
$$E(\mathbf{y}) = \tilde{\boldsymbol{\Gamma}}^{-1}E(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{0}$$

and

$$\begin{aligned} \text{var}\,(\mathbf{y}) &= \tilde{\boldsymbol{\Gamma}}^{-1}\,\text{var}\,(\mathbf{x} - \boldsymbol{\mu})(\tilde{\boldsymbol{\Gamma}}')^{-1} \\ &= \tilde{\boldsymbol{\Gamma}}^{-1}\,\text{var}\,(\mathbf{x})(\tilde{\boldsymbol{\Gamma}}')^{-1} \\ &= \tilde{\boldsymbol{\Gamma}}^{-1}\Sigma(\tilde{\boldsymbol{\Gamma}}')^{-1} \\ &= \tilde{\boldsymbol{\Gamma}}^{-1}\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\Gamma}}'(\tilde{\boldsymbol{\Gamma}}')^{-1} = \mathbf{I} \, . \end{aligned}$$

Hence $\mathbf{y}$ is of the desired standard form. Then

$$\begin{aligned} \mathbf{y}'\mathbf{y} &= (\mathbf{x}' - \boldsymbol{\mu}')(\Gamma^{-1})'\Gamma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ &= (\mathbf{x}' - \boldsymbol{\mu}')\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \, . \end{aligned}$$

This is called the "*Standard Quadratic Form*".

Since it is equal to the sum-of-squares of p standard variables, it will be distributed as χ^2 with p degrees of freedom, if $\mathbf{x}$ has the multivariate normal distribution.

4.2. The Principle of Least Squares

Recall that the General Linear Model is

$$\mathbf{y} = \mathbf{A}\boldsymbol{\xi} + \mathbf{e} \, .$$

On the assumption, for the time being that **A** is of full rank, the "least squares" approach tells us to estimate $\boldsymbol{\xi}$ in such a way that the sum of squares of errors is minimized. Then, $\mathbf{e}'\mathbf{e}$ is the desired sum of squares and

$$\begin{aligned} \mathbf{e} &= \mathbf{y} - \mathbf{A}\boldsymbol{\xi} \\ \mathbf{e}' &= \mathbf{y}' - \boldsymbol{\xi}'\mathbf{A}' \\ \mathbf{e}'\mathbf{e} &= (\mathbf{y}' - \boldsymbol{\xi}'\mathbf{A}')(\mathbf{y} - \mathbf{A}\boldsymbol{\xi}) \\ \frac{\partial(\mathbf{e}'\mathbf{e})}{\partial\boldsymbol{\xi}} &= -2\mathbf{A}'(\mathbf{y} - \mathbf{A}\boldsymbol{\xi}) \, . \end{aligned}$$

Setting this equal to zero, we obtain

$$\begin{aligned} \mathbf{A}'(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\xi}}) &= \mathbf{0} \\ (\mathbf{A}'\mathbf{A})\hat{\boldsymbol{\xi}} &= \mathbf{A}'\mathbf{y} \, . \end{aligned}$$

These are called the *Normal Equations* for the estimation of $\boldsymbol{\xi}$.

4.3. Minimum Variance Unbiased Estimates

The minimum variance, unbiased, linear estimate of $\boldsymbol{\xi}$ is obtained by the application of a very general form of the *Gauss Markov Theorem*:

Let

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\boldsymbol{\xi} + \mathbf{e} \\ E(\mathbf{e}) &= \mathbf{0} \\ \text{var}\,(\mathbf{y}) &= \text{var}\,(\mathbf{e}) = \sigma^2\mathbf{V} \end{aligned}$$

where $\mathbf{V}$ is a matrix (square, symmetric, non-singular) of order $(n \times n)$ with known elements. That is to say that variances of y_i (regardless of i) and covariances between y_i and y_j are known except for an arbitrary scalar multiplier applied to all of them. Then the best linear estimate of an arbitrary linear function $\mathbf{l}'\xi$ is equal to $\mathbf{l}'\hat{\xi}$ where ξ minimizes the quadratic form

$$\mathbf{e}'\mathbf{V}^{-1}\mathbf{e} \ .$$

Since

$$E(\mathbf{e}) = 0$$

and

$$\text{var}(\mathbf{e}) = \text{var}(\mathbf{y}) = \sigma^2\mathbf{V} \ ,$$

the standard quadratic form of $\mathbf{e}$ would be

$$(\mathbf{e}' - [E(\mathbf{e})'])[\text{var}(\mathbf{e})]^{-1}(\mathbf{e} - [E(\mathbf{e})]) = (\mathbf{e}' - \mathbf{0}')[\sigma^2\mathbf{V}]^{-1}(\mathbf{e} - \mathbf{0})$$
$$= \frac{1}{\sigma^2}\mathbf{e}'\mathbf{V}^{-1}\mathbf{e} \ .$$

Minimizing this expression is equivalent to minimizing

$$\mathbf{e}'\mathbf{V}^{-1}\mathbf{e} \ .$$

Hence, the statement, "The best linear estimate of an arbitrary function $\mathbf{l}'\xi$ is equal to $\mathbf{l}'\hat{\xi}$ where $\hat{\xi}$ is obtained by minimizing the quadratic form $\mathbf{e}'\mathbf{V}^{-1}\mathbf{e}$", as made in the Gauss-Markov Theorem is, in fact, equivalent to the statement . . . where $\hat{\xi}$ is obtained by minimizing the standard quadratic form due to error. The Normal Equations in this general case are $\mathbf{A}'\mathbf{V}^{-1}\mathbf{A}\hat{\xi} = \mathbf{A}'\mathbf{V}^{-1}\mathbf{y}$.

5. GENERAL LINEAR HYPOTHESIS OF FULL RANK

In this section, we shall discuss, with illustrations, the problem of testing hypotheses about certain parameters and also derive some necessary distribution in connection with testing hypotheses.

5.1. Notation

In general, a null hypothesis will be stated as

$$\mathbf{C}\xi = \mathbf{k} \ ,$$

where $\mathbf{C} = \mathbf{C}(n_h \times m)$, $(n_h \leq m)$, is called the hypothesis matrix and is of rank n_h.
ξ is an $(m \times 1)$ column vector of parameters as defined in the general linear model.
$\mathbf{k}$ is a vector of n_h known elements, usually equal to 0.
n_h is called *degrees of freedom due to hypothesis*. Actually it is the number of rows in the hypothesis matrix $\mathbf{C}$. In other words, it is the number of nonredundant statements embodied in the null hypothesis.
n_e is called the *degrees of freedom due to error* and is equal to the number of observations minus the effective number of parameters.

It is important to keep in mind that in stating a composite hypothesis, we should *never* make:

(1) contradicting statements such as,

$$H_0\colon \beta_1 = \beta_2 \quad \text{and} \quad \beta_1 = 2\beta_2 \text{ simultaneously}$$

(2) redundant statements such as,

$$H_0\colon \tau_1 = \tau_2 \quad \text{and} \quad 3\tau_1 = 3\tau_2 \ .$$

5.2. Simple Linear Regression

Model

$$y_i = \mu + \beta x_i + e_i$$
$$\text{parameter vector } \xi = \begin{bmatrix}\mu\\ \beta\end{bmatrix}$$

EXAMPLE 1

$$H_0\colon \mu = 0$$
$$\text{Alt.}\colon \mu \neq 0$$

General linear hypothesis

$$[1,0]\begin{bmatrix}\mu\\ \beta\end{bmatrix} = 0$$
$$\mathbf{C} \quad \xi \ = 0 \ , \qquad n_h = 1$$

EXAMPLE 2

$$H_0\colon \beta = 0$$
$$\text{Alt.}\colon \beta \neq 0$$
$$[0,1]\begin{bmatrix}\mu\\ \beta\end{bmatrix} = 0$$
$$\mathbf{C} \quad \xi \ = 0 \ , \qquad n_h = 1$$

EXAMPLE 3

$$H_0\colon \mu = 0,\ \beta = 0 \text{ simultaneously}$$
$$\text{Alt.: At least one of the } \mu \text{ and } \beta \neq 0$$
$$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\mu\\ \beta\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$$
$$\mathbf{C} \quad \xi \ = \ \mathbf{0} \ , \qquad n_h = 2$$

EXAMPLE 4

$$H_0\colon \mu = \beta$$
$$\text{Alt.}\colon \mu \neq \beta$$
$$[1,-1]\begin{bmatrix}\mu\\ \beta\end{bmatrix} = 0 \ , \qquad n_h = 1$$

5.3. Analysis of Variance, One-way Classification

$$y_{ij} = \mu + \tau_i + e_{ij} \qquad (i = 1, 2, 3 \cdots v)$$
$$\text{Parameter vector } \xi' = [\mu, \tau_1, \tau_2, \tau_3, \ \ldots \ \tau_v]$$

EXAMPLE 1

$$H_0\colon \tau_1 = \tau_2 = \tau_3 = \cdots = \tau_v$$
$$\text{Alt.}\colon \tau_r \neq \tau_s \qquad \text{for at least one pair}$$

Keep in mind that we must not make redundant statements. Here we have $(v - 1)$ rows in the hypothesis matrix. i.e. $n_h = v - 1$

$$(v-1)\begin{bmatrix}0 & 1 & -1 & 0 & 0 & \cdots & 0\\ 0 & 1 & 0 & -1 & 0 & \cdots & 0\\ 0 & 1 & 0 & 0 & -1 & \cdots & 0\\ \cdot & & \cdot & & \cdot & & \cdot\\ 0 & 1 & 0 & 0 & 0 & \cdots & -1\end{bmatrix}\begin{bmatrix}\mu\\ \tau_1\\ \tau_2\\ \tau_3\\ \cdot\\ \cdot\\ \cdot\\ \tau_v\end{bmatrix} = \begin{bmatrix}0\\ 0\\ 0\\ 0\\ \cdot\\ \cdot\\ \cdot\\ 0\end{bmatrix}(v-1)$$
$$(v+1)$$
$$\mathbf{C} \qquad\qquad \xi \ = \ \mathbf{0}$$

EXAMPLE 2

$$H_0: \tau_1 = \tau_2 = \tau_3 = \cdots = \tau_v = 0$$
Alt.: At least one $\tau \neq 0$

$$(v)\underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}}_{(v+1)} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \cdot \\ \cdot \\ \cdot \\ \tau_v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} (v)$$

$$\mathbf{C} \qquad \xi = \mathbf{0}$$

n_h = number of rows in C and is equal to v.

EXAMPLE 3. For simplicity, let us take $i = 1, 2, 3, 4$

H_0: $-\tau_1 + 2\tau_2 - \tau_3 = 0$ (Quadratic contrast of three effects)
Alt.: Quadratic contrast $\neq 0$

$$[0 \quad -1 \quad +2 \quad -1 \quad 0] \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{bmatrix} = 0$$

$$\mathbf{C} \qquad \xi = \mathbf{0}$$

5.4. Multiple Linear Regression

$$y_i = \mu + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots \beta_k x_{ki} + e_i$$
parameter vector $\xi' = [\mu, \beta_1, \beta_2, \beta_3, \ldots \beta_k]$

EXAMPLE 1

H_0: $\beta_1 = 0$
Alt.: $\beta_1 \neq 0$

$$[0,1,0,0 \cdots 0] \begin{bmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \cdot \\ \beta_k \end{bmatrix} = 0$$

$$\mathbf{C} \qquad \xi = \mathbf{0}, \qquad n_h = 1.$$

EXAMPLE 2

H_0: $\beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k = 0$
Alt.: At least one $\beta \neq 0$

$$(k)\underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}}_{(k+1)} \begin{bmatrix} \mu_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} = \mathbf{0}$$

$$\mathbf{C} \qquad \xi = \mathbf{0}, \qquad n_h = k.$$

EXAMPLE 3

H_0: $\beta_1 = 0$, $\beta_2 = 0$ simultaneously
Alt.: At least one of β_1 and $\beta_2 \neq 0$

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{C} \qquad \xi = \mathbf{0}; \qquad n_h = 2.$$

5.5. Randomized Blocks

$$y_{ij} = \mu + \tau_i + \beta_j + e_{ij}$$

where μ is the general effect
τ_i are the treatment effects $(i = 1, 2, 3)$
β_j are the block effects $(j = 1, 2, 3, 4)$
e_{ij} is the experimental error

parameter vector $\xi' = [\mu, \tau_1, \tau_2, \tau_3, \beta_1, \beta_2, \beta_3, \beta_4]$

EXAMPLE 1

H_0: $\tau_1 = \tau_2 = \tau_3$ (all treatments effects are equal)
Alt.: At least one pair $\tau_r \neq \tau_s$

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{C} \qquad \xi = \mathbf{0}; \qquad n_h = 2.$$

EXAMPLE 2

H_0: $-\tau_1 + 2\tau_2 - \tau_3 = 0$ (quadratic contrast)
Alt.: Quadratic contrast $\neq 0$

$$[0 \quad -1 \quad 2 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0] \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = 0$$

$$\mathbf{C} \qquad \xi = \mathbf{0}; \qquad n_h = 1.$$

From the above illustrations, it can be seen that we can write a great variety of tests in the form of the General Linear Hypothesis provided that we make no redundant hypothesis statements.

5.6. Quadratic Form due to Hypothesis

So far we have discussed only the model of full rank. i.e., in the normal equations,

$$A'A\hat{\xi} = A'y \ ,$$

(A'A) has an inverse. We shall continue to assume this model throughout this chapter. Recall the General Linear Model

$$y = A\xi + e$$
$$E(y) = A\xi \quad \text{and} \quad \text{var}\ (y) = \sigma^2 I \ .$$

We then have the normal equations,

$$A'A\hat{\xi} = A'y \ .$$

The estimate of ξ is

$$\hat{\xi} = (A'A)^{-1}A'y \ ,$$

and the variance of the estimate is

$$\begin{aligned} \text{var}\ \hat{\xi} &= (A'A)^{-1}\ \text{var}\ (A'y)(A'A)^{-1} \\ &= (A'A)^{-1}A'\ \text{var}\ (y)A(A'A)^{-1} \\ &= \sigma^2(A'A)^{-1}A'A(A'A)^{-1} \\ &= \sigma^2(A'A)^{-1} \ . \end{aligned}$$

This is the expression for the **variance-covariance matrix** of the estimates of ξ. Now suppose that we have a null hypothesis

$$H_0\text{: } C\xi = 0 \ .$$

We have an unbiased estimate of $C\xi$ namely $C\hat{\xi}$, i.e., under the null hypothesis,

$$\begin{aligned} E(C\hat{\xi}) &= C\xi = 0 \\ \text{var}\ (C\hat{\xi}) &= C\ \text{var}\ (\hat{\xi})C' \\ &= \sigma^2C(A'A)^{-1}C', \\ [\text{var}\ (C\hat{\xi})]^{-1} &= \frac{1}{\sigma^2}[C(A'A)^{-1}C']^{-1} \ . \end{aligned}$$

Thus, under the null hypothesis, the standard quadratic form is

$$\frac{1}{\sigma^2}\hat{\xi}'C'[C(A'A)^{-1}C']^{-1}C\hat{\xi} \ .$$

The expression, $\hat{\xi}'C'[C(A'A)^{-1}C']^{-1}C\hat{\xi}$, is called the *sum of squares due to hypothesis*, usually denoted by SSH. If y has the multivariate normal distribution, SSH/σ^2 is distributed as χ^2 with n_h degrees of freedom, since it is a standard quadratic form.

5.7. Sum of Squares due to Error

Recall the general linear model

$$y = A\xi + e \ .$$

Let us define $\hat{e} = y - A\hat{\xi}$, the error of estimation. Then, $\sum_{i=1}^{n} \hat{e}_i^2 = \hat{e}'\hat{e}$ is called the sum of squares of errors of estimation. It is customarily denoted by SSE

$$\begin{aligned} SSE = \hat{e}'\hat{e} &= (y' - \hat{\xi}'A')(y - A\hat{\xi}) \\ &= y'y - \hat{\xi}'A'y - y'A\hat{\xi} + \hat{\xi}'A'A\xi \\ &= y'y - \hat{\xi}'A'y - y'A\hat{\xi} + \hat{\xi}'A'y \\ &= y'y - y'A\hat{\xi} \ , \end{aligned}$$

where y'y is the sum of squares over all observations.
A'y is the column vector on the right hand side of the normal equations.
$\hat{\xi}$ is a column vector whose elements are the estimates of ξ.

In words, SSE is obtained by subtracting from the sum of squares of all observations, the scalar product of the vector of estimates of ξ and the vector on the right-hand side of the normal equations.

It should be noted that SSE can depend only on the model, and is determined once the model is stated; it is entirely independent of any hypothesis which may be stated or tested.

If y is normally distributed, SSE/σ^2 has the χ^2 distribution with n_e degrees of freedom. It is independent of any SSH.

5.8. Summary

We have the general linear model

$$y = A\xi + e$$
$$E(y) = A\xi \ .$$

We assume that the model is of full rank, that is, A'A is non-singular and thus has an inverse. If we further assume that

$$\text{var}\ (y) = \sigma^2 I \ ,$$

that is, homoscedasticity plus independence, we will have the normal equations

$$(A'A)\hat{\xi} = A'y$$

and we can obtain the estimate of ξ by

$$\hat{\xi} = (A'A)^{-1}A'y \ .$$

Again, if we further assume that the elements of y are normally distributed, we may test the following hypothesis:

$$H_0\text{: } C\xi = 0$$
$$\text{Alt.: } C\xi = \mathbf{a} \qquad (\neq 0)$$

This hypothesis matrix has n_h rows and, if we avoid inconsistency and redundancies in the statement of the hypothesis, n_h will be the "degrees of freedom due to hypothesis".

5.9. Computational Procedure for Testing a Hypothesis

In testing a hypothesis, proceed as follows:

(1) Obtain SSH, the so called "sum of squares due to hypothesis" from the formula

$$SSH = \hat{\xi}'C'[C(A'A)^{-1}C']^{-1}C\hat{\xi} \ .$$

(2) Obtain SSE, the "sum of squares due to error" from

$$SSE = \sum_{\text{all}} y_i^2 - y'A\hat{\xi} \ .$$

(3) Introduce n_e, the "degrees of freedom due to error" which equals n (sample size) minus effective number of parameters in the model.

(4) Then, if H_0 is true

$$\frac{SSH/n_h}{SSE/n_e} = F_{(n_h, n_e)} \ .$$

5.10. Regression Significance Test

Suppose that we have the general linear model

$$\mathbf{y} = \mathbf{A}\xi + \mathbf{e}$$
$$E(\mathbf{y}) = \mathbf{A}\xi \quad \text{and} \quad \text{var}(\mathbf{y}) = \sigma^2\mathbf{I} \; .$$

Under this model, we have the normal equations

$$\mathbf{A}'\mathbf{A}\hat{\xi} = \mathbf{A}'\mathbf{y}$$
$$\text{SSE} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{A}\hat{\xi}$$
$$\text{where } \hat{\xi} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} \; .$$

Now, suppose that our hypothesis is of such a nature that we can easily write the reduced model under the assumption that H_0 is true.

$$y = \mathbf{A}\xi + \mathbf{e}^*$$
$$\text{subject to the condition } \mathbf{C}\xi = \mathbf{0} \; .$$

Analogously, after estimating ξ in the above model (reduced model) we may write

$$\text{SSE (reduced)} = y'y - y'\mathbf{A}\hat{\xi}$$

where $\hat{\xi}$ is the estimate of ξ in the reduced model. We can then find the sum of squares due to that hypothesis by obtaining

SSE (reduced) and subtracting SSE (the original or general model), i.e.,

$$\text{SSH} = \text{SSE (reduced)} - \text{SSE}.$$

5.11. Alternate Form of the Distribution

$\dfrac{\text{SSE}}{\text{SSE} + \text{SSH}}$ has the Beta distribution with parameters $(n_e/2, n_h/2)$.

The beta tests are *lower-tail* tests, i.e., we reject H_0 if the value of the observed ratio is *smaller* than the tabulated one, i.e.,

$$\text{rejection region } \beta < \text{constant}.$$

Actually in the tables, percentage points of β are stated as

$$\beta(a,b),$$

where $a = 2$ (second parameter)
$b = 2$ (first parameter).

Hence, read those tables simply as $\beta(n_h, n_e)$

$$\frac{\text{SSE}}{\text{SSE (reduced)}} = \beta\left(\frac{n_e}{2}, \frac{n_h}{2}\right) \quad \text{(usual notation)}$$
$$= \beta^*(n_h, n_e) \quad \text{(Tables for Beta percentage points)}$$
$$= I\left(\frac{n_e}{2}, \frac{n_h}{2}\right) \quad \text{(Tables of the Incomplete Beta Function).}$$

6. GENERAL LINEAR MODEL OF LESS THAN FULL RANK

So far we have restricted our discussion to models of full rank in the General Linear Model. In practice, many design models are not initially of this form. Models not of full rank are sometimes called singular models.

If, in the General Linear Model

$$\mathbf{y} = \mathbf{A}\xi + \mathbf{e} \; ,$$

with normal equations

$$\mathbf{A}'\mathbf{A}\hat{\xi} = \mathbf{A}'\mathbf{y}$$

the rank of the design matrix $\mathbf{A}$ is less than m $(r < m)$ then $(\mathbf{A}'\mathbf{A})$ would be singular and has no inverse. We must examine the system to see whether a solution exists. We wish to find functions of the ξ_i's for which unbiased estimates exist.

6.1. Estimable Function and Estimability

Let us estimate a function $\mathbf{l}'\xi$, i.e., find $\mathbf{c}'\mathbf{y} = \mathbf{l}'\hat{\xi}$ such that the expectation

$$E(\mathbf{c}'\mathbf{y}) = \mathbf{l}'\xi \text{ for all } \xi \; ,$$

and var $(\mathbf{c}'\mathbf{y})$ = minimum, i.e., we would like to find a linear function of the y_i's such that

$$E(\mathbf{c}'\mathbf{y}) = \mathbf{l}'\xi \; ,$$

where $\mathbf{l}'$ is a given vector of "weights". The constraints of unbiasedness are

$$E(\mathbf{c}'\mathbf{y}) = \mathbf{l}'\xi$$
$$\mathbf{c}'E(\mathbf{y}) = \mathbf{l}'\xi$$
$$\mathbf{c}'\mathbf{A}\xi = \mathbf{l}'\xi \qquad \text{for all } \xi \text{, hence,}$$
$$\mathbf{c}'\mathbf{A} = \mathbf{l}'$$
$$\mathbf{c}'\mathbf{A} - \mathbf{l}' = \mathbf{0} \; .$$

Hence, we are minimizing

$$\text{var}(\mathbf{c}'\mathbf{y}) \text{ subject to the constraints } \mathbf{c}'\mathbf{A} = \mathbf{l}'$$

where var $(\mathbf{c}'\mathbf{y}) = \sigma^2\mathbf{c}'\mathbf{c}$.

The criterion function Φ is then

$$\Phi = \tfrac{1}{2}\mathbf{c}'\mathbf{c} - [\mathbf{c}'\mathbf{A} - \mathbf{l}']\lambda$$
$$\frac{\partial\Phi}{\partial\mathbf{c}} = \mathbf{c} - \mathbf{A}\lambda \; .$$

Setting the derivative equal to zero, we obtain

(6.1.1)
$$\mathbf{A}\lambda = \hat{\mathbf{c}}$$

Premultiplying by $\mathbf{A}'$, we have

$$\mathbf{A}'\mathbf{A}\lambda = \mathbf{A}'\hat{\mathbf{c}} \; ,$$

which is equal to $\mathbf{l}$ under our constraints.

Hence,

(6.1.2)
$$\mathbf{A}'\mathbf{A}\lambda = \mathbf{l} \; .$$

(6.1.1) and (6.1.2) are called "conjugate normal equations". If $\mathbf{A}$ has rank r $(<m)$ we can always select r columns which form a "basis" and take the remaining $(m - r)$ columns as an extension. The latter columns are linear combinations of the former.

In the model

$$\mathbf{y} = \mathbf{A}\xi + \mathbf{e}$$

let us order the elements in ξ as well as the columns in $\mathbf{A}$ in such a way that

$$\xi' = [\underset{(r)}{\xi_1'} \;,\; \underset{(m-r)}{\xi_2'}]$$

and

$$\mathbf{A} = [\underset{(r)}{\mathbf{A}_1} \;,\; \underset{(m-r)}{\mathbf{A}_2}](r)$$

and that $\mathbf{A}_1$ is a basis of $\mathbf{A}$. The columns of $\mathbf{A}_2$ must then be linear combinations of those in $\mathbf{A}_1$. We may express this fact formally by saying there exists

$$\mathbf{Q}(r \times \overline{m-r}) \qquad \text{such that} \qquad \mathbf{A}_2 = \mathbf{A}_1\mathbf{Q}\ .$$

Suppose

$$\begin{aligned} \mathbf{A}_2 &= \mathbf{A}_1 Q \\ \mathbf{A}_1'\mathbf{A}_2 &= \mathbf{A}_1'\mathbf{A}_1 Q \\ \mathbf{Q} &= (\mathbf{A}_1'\mathbf{A}_1)^{-1}\mathbf{A}_1'\mathbf{A}_2\ . \end{aligned}$$

This is one of the ways to determine $\mathbf{Q}$ when $\mathbf{A}_1$ and $\mathbf{A}_2$ are given. Usually, however, we would try to find $\mathbf{Q}$ by inspection.

Now,

$$\begin{aligned} \mathbf{A} &= \underset{(r)}{[\mathbf{A}_1}\ ,\ \underset{(m-r)}{\mathbf{A}_2]}(r) \\ &= [\mathbf{A}_1, \mathbf{A}_1\mathbf{Q}] = \mathbf{A}_1\underset{(r)}{[\mathbf{I}} \quad \underset{(m-r)}{\mathbf{Q}]}(r) \end{aligned}$$

(6.1.2) can be written as

$$\begin{bmatrix} \mathbf{A}_1' \\ \mathbf{Q}'\mathbf{A}_1' \end{bmatrix} [\mathbf{A}_1\ ,\ \mathbf{A}_1\mathbf{Q}]\boldsymbol{\lambda} = \begin{bmatrix} \mathbf{A}_1'\mathbf{A}_1 & \mathbf{A}_1'\mathbf{A}_1\mathbf{Q} \\ \mathbf{Q}'\mathbf{A}_1'\mathbf{A}_1 & \mathbf{Q}'\mathbf{A}_1'\mathbf{A}_1\mathbf{Q} \end{bmatrix} \boldsymbol{\lambda} = \begin{bmatrix} \mathbf{l}_1 \\ \mathbf{l}_2 \end{bmatrix} \begin{matrix} (r) \\ (m-r) \end{matrix}\ .$$

Expanding we have

$$[\mathbf{A}_1'\mathbf{A}_1, \mathbf{A}_1'\mathbf{A}_1\mathbf{Q}]\boldsymbol{\lambda} = \mathbf{l}_1 \tag{6.1.3}$$

$$[\mathbf{Q}'\mathbf{A}_1'\mathbf{A}_1, \mathbf{Q}'\mathbf{A}_1'\mathbf{A}_1\mathbf{Q}]\boldsymbol{\lambda} = \mathbf{l}_2\ . \tag{6.1.4}$$

Premultiply (6.1.3) by $\mathbf{Q}'$ and obtain

$$[\mathbf{Q}'\mathbf{A}_1'\mathbf{A}_1, \mathbf{Q}'\mathbf{A}_1'\mathbf{A}_1\mathbf{Q}]\boldsymbol{\lambda} = \mathbf{Q}'\mathbf{l}_1\ .$$

For consistency of the equation system, the condition

$$\mathbf{l}_2 = \mathbf{Q}'\mathbf{l}_1$$

must be met.

That is to say, in the function $\mathbf{l}'\boldsymbol{\xi}$, $\mathbf{l}'$ cannot be chosen arbitrarily but must be of the form

$$\mathbf{l}' = \underset{(r)}{[\mathbf{l}_1'}\ ,\ \underset{(m-r)}{\mathbf{l}_2']}\ , \quad \text{where}$$

$$\mathbf{l}_2' = \mathbf{l}_1'\mathbf{Q}\ . \tag{6.1.5}$$

Only an $\mathbf{l}$ satisfying this relation can be used in the construction of a function which admits of a linear unbiased (and mathematically consistent) estimate.

(6.1.5) is called the condition of "estimability" of a linear function. Hence, we will call a function $\mathbf{l}'\boldsymbol{\xi}$ *estimable* if $\mathbf{l}'$ can be written as

$$[\mathbf{l}_1', \mathbf{l}_2']\ ,$$

where $\mathbf{l}_2'$ is related to $\mathbf{l}_1'$ in the same way as $\mathbf{A}_2$ to $\mathbf{A}_1$.

We may then define that a parametric function is said to be linearly *estimable* if there exists a linear combination of the observations whose expected value is equal to the function, i.e., if there exists an unbiased estimate.

Now, if the function $\mathbf{l}'\boldsymbol{\xi}$ is estimable, (6.1.3) can be written as

$$(\mathbf{A}_1'\mathbf{A}_1)[\mathbf{I},\mathbf{Q}]\boldsymbol{\lambda} = \mathbf{l}_1\ .$$

Notice that $\mathbf{l}_2$ may be disregarded since it is determined by the relation $\mathbf{l}_2 = \mathbf{Q}'\mathbf{l}_1$, hence,

$$[\mathbf{I},\mathbf{Q}]\boldsymbol{\lambda} = (\mathbf{A}_1'\mathbf{A}_1)^{-1}\mathbf{l}_1\ . \tag{6.1.6}$$

The first conjugate normal equations (6.1.1) stated

$$\mathbf{A}\boldsymbol{\lambda} = \hat{\mathbf{c}}$$

or,

$$\begin{aligned} [\mathbf{A}_1,\mathbf{A}_2]\boldsymbol{\lambda} &= \hat{\mathbf{c}} \\ \mathbf{A}_1[\mathbf{I},\mathbf{Q}]\boldsymbol{\lambda} &= \hat{\mathbf{c}}\ . \end{aligned}$$

Inserting (6.1.6), we obtain

$$\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^{-1}\mathbf{l}_1 = \hat{\mathbf{c}}\ .$$

Hence,

$$\hat{\mathbf{c}}'\mathbf{y} = \widehat{\mathbf{l}'\boldsymbol{\xi}} = \mathbf{l}_1'(\mathbf{A}_1'\mathbf{A}_1)^{-1}\mathbf{A}_1'\mathbf{y}\ ,$$

which is of the same form as in the non-singular case, except that $\mathbf{A}$ has been replaced by its basis $\mathbf{A}_1$ and in $\mathbf{l}$ we consider only the first r elements, i.e., $\mathbf{l}_1$.

Hence, the normal equations in the Least Squares approach, i.e.,

$$\mathbf{A}'\mathbf{A}\hat{\boldsymbol{\xi}} = \mathbf{A}'\mathbf{y}$$

can be used formally in the reduced statement

$$(\mathbf{A}_1'\mathbf{A}_1)\hat{\boldsymbol{\xi}}_1 = \mathbf{A}_1'\mathbf{y}\ .$$

6.2. General Linear Hypothesis Model of Less Than Full Rank

We have the general linear model

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\boldsymbol{\xi} + \mathbf{e} \\ &= [\mathbf{A}_1,\mathbf{A}_2]\begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix} + \mathbf{e} \\ &= [\mathbf{A}_1,\mathbf{A}_1\mathbf{Q}]\begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix} + \mathbf{e} \\ &= \mathbf{A}_1\boldsymbol{\xi}_1 + \mathbf{A}_1\mathbf{Q}\boldsymbol{\xi}_2 + \mathbf{e} \\ &= \mathbf{A}_1(\boldsymbol{\xi}_1 + \mathbf{Q}\boldsymbol{\xi}_2) + \mathbf{e}\ . \end{aligned}$$

Hence, we may write the general linear model in the form $\mathbf{y} = \mathbf{A}_1\boldsymbol{\xi}^* + \mathbf{e}$, where $\boldsymbol{\xi}^* = \boldsymbol{\xi}_1 + \mathbf{Q}\boldsymbol{\xi}_2$.

6.2.1. *Sum of Squares Due to Error and Its Distribution*

Notice that $\mathbf{e}$ has not been changed in this model, hence we can set up the normal equations

$$\begin{aligned} (\mathbf{A}_1'\mathbf{A}_1)\hat{\boldsymbol{\xi}}^* &= \mathbf{A}_1'\mathbf{y}. \\ \text{SSE} = \mathbf{e}'\mathbf{e} &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{A}_1\hat{\boldsymbol{\xi}}^* \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{A}_1(\mathbf{A}_1'\mathbf{A}_1)^{-1}\mathbf{A}_1'\mathbf{y} \end{aligned}$$

Then, as before

$$\frac{\text{SSE}}{\sigma^2} = \chi^2(n-r) \qquad \text{where } r \text{ is the rank of } \mathbf{A},$$

and replaces m in the non-singular model. The "effective" number of parameters in the singular model is only r, the remaining $(m-r)$ parameters are determined in terms of the first r by the estimability condition.

6.2.2. *Sum of Squares Due to Hypothesis and Its Distribution*

Suppose that we wish to test

$$H_0: \mathbf{C}\xi = \mathbf{0}\ ,$$

where $C = [\underset{(r)}{\mathbf{C}_1}\ ,\ \underset{(m-r)}{\mathbf{C}_2}]$.

Then $\mathbf{C}\xi = \mathbf{0}$ implies that

$$[\mathbf{C}_1, \mathbf{C}_2]\begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}.$$

Each row on the left-hand side must represent an estimable function, hence, we must have

$$\mathbf{C}_2 = \mathbf{C}_1\mathbf{Q}\ .$$

This is called the condition of "testability", i.e., if

$$\mathbf{C} = \begin{bmatrix}\mathbf{c}_2'\\ \mathbf{c}_1'\\ \cdot\\ \cdot\\ \cdot\\ \mathbf{c}_{n_h}'\end{bmatrix},$$

where $(\mathbf{c}_i'\xi)$ is an estimable function ($i = 1, 2, \ldots, n_h$). Then the null hypothesis

$$H_0: \mathbf{C}_1\xi_1 + \mathbf{C}_2\xi_2 = \mathbf{0}$$

can be written as

$$\mathbf{C}_1\xi_1 + \mathbf{C}_1\mathbf{Q}\xi_2 = \mathbf{0}$$

or simply $\mathbf{C}_1\xi^* = \mathbf{0}$, where $\xi^* = (\xi_1 + \mathbf{Q}\xi_2)$. Hence, we can formally state that a null hypothesis

$$H_0: \mathbf{C}\xi = \mathbf{0}$$

is "*testable*" if $\mathbf{C}\xi$ consists of n_h estimable functions, i.e., if $\mathbf{C}_2 = \mathbf{C}_1\mathbf{Q}$, where $\mathbf{C} = [\mathbf{C}_1,\mathbf{C}_2]$. Consequently,

$$\mathrm{SSH} = \hat{\xi}^{*\prime}\mathbf{C}_1'[\mathbf{C}_1(\mathbf{A}_1'\mathbf{A}_1)^{-1}\mathbf{C}_1']^{-1}\mathbf{C}_1\hat{\xi}^*\ ,$$

where $\hat{\xi}^* = (\mathbf{A}_1'\mathbf{A}_1)^{-1}\mathbf{A}_1\mathbf{y}$.

As before

$$\frac{\mathrm{SSH}}{\sigma^2} = \chi^2_{(n_h)}\ .$$

Again, if the null-hypothesis is true, we have the test statistic F

$$\frac{\mathrm{MSH}/n_h}{\mathrm{MSE}/n_e} = \mathrm{F}_{(n_h, n_e)}\ .$$

6.3. Constraints and Conditions

If the model is singular, of rank $r < m$, $(m - r)$ constraints on the $\hat{\xi}_i$'s (the estimates) may be arbitrarily introduced, for example:

$$\hat{\xi}_{r+1} = 0, \ldots, \hat{\xi}_m = 0 \qquad (6.3.1)$$

or

$$\sum_{i=1}^{m} \hat{\xi}_i = 0, \sum_{i=1}^{m} n_i\hat{\xi}_i = 0\ . \qquad (6.3.2)$$

This is called *reparametrizing* the model. The constraining functions are fairly arbitrary, but they *must not be estimable* functions, otherwise the resulting model will still be singular.

In effect, this is done by deletion of the last $(m - r)$ rows and columns of $\mathbf{A}'\mathbf{A}$ and the last $(m - r)$ elements of $\mathbf{A}'\mathbf{y}$, for constraints of the type (6.3.1), or by adding a constant to all elements of $\mathbf{A}'\mathbf{A}$, for constraints of the type (6.3.2). This has no effect on the value of estimable functions, or test statistics.

An entirely different situation prevails if we place conditions on the *parameters* of a model, especially on interactions. In the two-way classification model

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \delta_{ij}$$

one usually specifies

$$\sum_i n_{ij}\delta_{ij} = 0 \qquad \text{for all } j\text{'s}$$

and

$$\sum_j n_{ij}\delta_{ij} = 0 \qquad \text{for all } i\text{'s}\ ,$$

where n_{ij} denotes the number of observations in the (i,j) cell. These are sometimes called *natural constraints* (they are neither *natural* nor *constraints*). They simply represent a set of *conditions* or *assumptions* on the interactions, minimizing this effect (making SSH for interaction a minimum). After introducing these conditions, one still has a singular model, which can be made nonsingular by introduction of the arbitrary constraints

$$\sum_i \hat{\alpha}_i = 0, \sum_j \hat{\beta}_j = 0\ .$$

(Note the carets, for estimates). One could introduce the different assumptions,

$$\text{All } \alpha_i\text{'s} = 0$$
$$\text{All } \beta_j\text{'s} = 0$$

and would have a simple one-way classification model, quite different from the previous one. A classical example is the following. Suppose some organic substance is attacked by sulphuric acid or by sodium hydroxide.

		NaOH −	NaOH +
H_2SO_4	−	0	4
	+	6	0

Using, formally, the minimizing interaction conditions, one would obtain effect estimates as means of rows and columns

H_2SO_4	absent: 2	present: 3
NaOH	absent: 3	present: 2

and make the ridiculous inference that sodium hydroxide, by itself, has an inhibiting effect. The correct parametric model in this case would be

μ	$\mu + \beta$
$\mu + \alpha$	$\mu + \alpha + \beta + \delta$

,

i.e., interaction occurs only if both substances are present. This leads to the estimation

$$\hat{\mu} = 0$$
$$\hat{\alpha} = 6$$
$$\hat{\beta} = 4$$
$$\hat{\delta} = -10 \ ,$$

which is the appropriate neutralization model.

It is usually quite easy to decide whether a constraint or a condition is involved. The (model-changing) *conditions* are required whenever a hierarchy of effects is present (main effects, interactions, higher-order interactions), while constraints (with no effect on the model) can be introduced within the same kinds of effects (row effect estimates adding to zero, column effect estimates adding to zero). The sum of squares due to a given hypothesis is a good indicator of the situation. If it changes by the introduction of two different sets of combinations, they are *conditions*, and must be determined in accordance with plausibility of the physical model. If it stays the same, they are usually *constraints on the estimates*, and thus arbitrary, without effect on the model.

SIMPLIFIED COMPUTATIONS FOR MULTIPLE REGRESSION

by Dr. Clyde Y. Kramer

The following method is especially suited for a research problem which has several dependent variables with one set of independent variables. It allows the worker to decide which dependent variables are explained by regression with the least time and work. Regression coefficients and their variances are not usually wanted unless the regression is significant. This procedure eliminates the need of calculating these quantities when prediction is not good enough to be useful.

The main advantages of this method are:

(a) the numbers of digits to the left of the decimal points of the elements of the sums of squares and sums of products matrix are adjusted to be one or zero which permits the use of a uniform number of decimal places in the calculations;
(b) the multiple correlation coefficient and entries for the analysis of variance table for multiple regression can be found without computing the regression coefficients and the inverse of the sums of squares and sums of products matrix;
(c) the additional reduction due to any regression variable over that obtained for previous ones is obtainable for every regression variable;
(d) the research worker can fit only those regression variables that add a significant additional reduction if he so desires;
(e) time or work is not lost if one wishes to obtain the regression coefficients and their variances; and
(f) numerous checks are employed on the calculations that are required.

Algebraic Procedure

For simplicity, this method will be illustrated by considering four independent variables (x_1, x_2, x_3, x_4), and one dependent variable, (y). First compute and record the sums of squares and sums of products in the following manner:

$$(1) \qquad \begin{array}{ccccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{1y} \\ & a_{22} & a_{23} & a_{24} & a_{2y} \\ & & a_{33} & a_{34} & a_{3y} \\ & & & a_{44} & a_{4y} \\ & & & & a_{yy} \end{array}$$

where

$$a_{ii} = \sum_{\alpha=1}^{n} x_{i\alpha}^2 - \frac{\left(\sum_{\alpha=1}^{n} x_{i\alpha}\right)^2}{n} ,$$

$$a_{ij} = \sum_{\alpha=1}^{n} x_{i\alpha}x_{j\alpha} - \frac{\left(\sum_{\alpha=1}^{n} x_{i\alpha}\right)\left(\sum_{\alpha=1}^{n} x_{j\alpha}\right)}{n} ,$$

$$a_{iy} = \sum_{\alpha=1}^{n} x_{i\alpha}y_{\alpha} - \frac{\left(\sum_{\alpha=1}^{n} x_{i\alpha}\right)\left(\sum_{\alpha=1}^{n} y_{\alpha}\right)}{n} ,$$

$$a_{yy} = \sum_{\alpha=1}^{n} y_{\alpha}^2 - \frac{\left(\sum_{\alpha=1}^{n} y_{\alpha}\right)^2}{n} ,$$

$i = j = 1, 2, 3, 4$, and n is the number of observations.

The first feature of this method is that the sum of squares for the dependent variable, a_{yy}, is recorded as the last entry of the column containing the sums of products of the dependent variable with the independent variables. The addition of a_{yy} to the last column results in a square matrix. This feature will be utilized to adjust the number of digits preceding the decimal points in the elements of the above matrix. The residual sum of squares is also obtained directly by adding the term a_{yy} to the last column of (1).

Then, in order to simplify the calculations, make the diagonal terms (a_{11}, a_{22}, a_{33}, a_{44}, a_{yy}) lie between 0.1 and 10 by pre- and post-multiplying (1) by a diagonal matrix of powers of ten which is as follows:

$$(2) \qquad \begin{bmatrix} 10^{q_1} & 0 & 0 & 0 & 0 \\ 0 & 10^{q_2} & 0 & 0 & 0 \\ 0 & 0 & 10^{q_3} & 0 & 0 \\ 0 & 0 & 0 & 10^{q_4} & 0 \\ 0 & 0 & 0 & 0 & 10^{p} \end{bmatrix} .$$

This will also result in the sums of products having at most one digit before the decimal points, thus allowing a uniform number of decimal places in all future calculations. The values of the q_i's and p are determined as follows:

Consider only a diagonal term and in it the largest *even* number of places through which the decimal point must be shifted (left or right) to make that term be a number between 0.1 and 10. Then divide this even number by two to get the applicable value of q_1 or p. For example, if $a_{11} = 8{,}238.93$, q_1 would be -2; if $a_{33} = 2{,}213{,}922.00$, q_3 would be -3; and if $a_{yy} = 5{,}098{,}35$, p would be -2.

After pre- and post-multiplying (1) by (2), which in effect is accomplished by adding the q_i's and p according to the term we are adjusting and shifting the decimal point the number of places indicated by the sum, we obtain a matrix of a^*'s. If, as in the above paragraph, $q_1 = -2$ and $p = -2$, we would shift the decimal point of a_{11} four places to the left and the decimal point of a_{1y} four places to the left, etc.