

**THE COORDINATE-FREE  
APPROACH TO  
LINEAR MODELS**

**MICHAEL J. WICHURA**

*The University of Chicago  
Chicago, Illinois*

In memoriam  
William H. Kruskal  
1919–2005

## PREFACE

When I was a graduate student in the mid 1960's I finally came to understand the mathematical theory underlying analysis of variance and regression after reading a draft of William Kruskal's monograph on the so-called coordinate-free, or geometric, approach to these subjects. Alas, with Kruskal's demise, this excellent treatise will never be published.

From time to time during the 1970's, 80's and early 90's, I had the good fortune of teaching the coordinate-free approach to linear models. While doing so, I evolved my own set of lecture notes, presented here. With regard to inspiration and content, my debt to Kruskal is clear. However, my notes are intended for a one, rather than three, quarter course, and are aimed at Statistics graduate students who are already fairly well versed in linear algebra. I have also included some of the highlights of the optimality theory for estimation and testing in linear models under the assumption of normality, feeling that the elegant setting provided by the coordinate-free approach is a natural one in which to place these jewels of mathematical statistics. Out of deference to Kruskal, who was my colleague here at the University of Chicago, I have not until now made my notes public. My hope is that readers will find the presentation both instructive and enjoyable.

Various graduate students, in particular Neal Thomas and Nathaniel Schenker have made many comments which have greatly improved these notes. My thanks go to all of them, and also to David Van Dyke and Peter Meyer for suggesting how easy/hard each of the some 200 exercises are. Thanks are also due to Mitzi Nakatsuka for her help in converting the notes to  $\text{\TeX}$ , and to Persi Diaconis for his advice and encouragement.

*Michael J. Wichura*

University of Chicago  
July 2005



# CONTENTS

## INTRODUCTION

|                                                           |    |
|-----------------------------------------------------------|----|
| <b>1. Introduction</b>                                    | 1  |
| 1. Orientation                                            | 1  |
| 2. An illustrative example                                | 2  |
| 3. Overview                                               | 4  |
| <b>2. Topics in linear algebra</b>                        | 6  |
| 1. Orthogonal projections                                 | 6  |
| 2. Properties of orthogonal projections                   | 12 |
| 2A. Characterization of orthogonal projections            | 12 |
| 2B. Differences of orthogonal projections                 | 13 |
| 2C. Sums of orthogonal projections                        | 16 |
| 2D. Products of orthogonal projections                    | 17 |
| 2E. An algebraic form of Cochran's theorem                | 19 |
| 3. Tjur's theorem                                         | 21 |
| 4. Self-adjoint transformations and the spectral theorem  | 32 |
| 5. Representation of linear and bilinear functionals      | 36 |
| 6. Problem set: Cleveland's identity                      | 40 |
| 7. Appendix: Rudiments                                    | 41 |
| 7A. Vector spaces                                         | 41 |
| 7B. Subspaces                                             | 42 |
| 7C. Linear functionals                                    | 43 |
| 7D. Linear transformations                                | 43 |
| <b>3. Random Vectors</b>                                  | 45 |
| 1. Random vectors taking values in an inner product space | 45 |
| 2. Expected values                                        | 46 |
| 3. Covariance operators                                   | 47 |
| 4. Dispersion operators                                   | 49 |
| 5. Weak sphericity                                        | 51 |
| 6. Getting to weak sphericity                             | 52 |
| 7. Normality                                              | 53 |
| 8. The main result                                        | 54 |
| 9. Problem set: Distribution of quadratic forms           | 58 |

|                                                           |     |
|-----------------------------------------------------------|-----|
| <b>4. Gauss-Markov Estimation</b>                         | 60  |
| 1. Linear functionals of $\mu$                            | 60  |
| 2. Estimation of linear functionals of $\mu$              | 62  |
| 3. Estimation of $\mu$ itself                             | 67  |
| 4. Estimation of $\sigma^2$                               | 70  |
| 5. Using the wrong inner product                          | 71  |
| 6. Invariance of GMEs under linear transformations        | 74  |
| 7. Some additional optimality properties of GMEs          | 75  |
| 8. Estimable parametric functionals                       | 77  |
| 9. Problem set: quantifying the Gauss-Markov theorem      | 85  |
| <b>5. Normal Theory: Estimation</b>                       | 88  |
| 1. Maximum likelihood estimation                          | 88  |
| 2. Minimum variance unbiased estimation                   | 89  |
| 3. Minimavity of $P_M Y$                                  | 91  |
| 4. James-Stein estimation                                 | 96  |
| 5. Problem set: Admissible minimax estimation of $\mu$    | 102 |
| <b>6. Normal Theory: Testing</b>                          | 108 |
| 1. The likelihood ratio test                              | 109 |
| 2. The $F$ -test                                          | 110 |
| 3. Monotonicity of the power of the $F$ -test             | 114 |
| 4. An optimal property of the $F$ -test                   | 118 |
| 5. Confidence intervals for linear functionals of $\mu$   | 125 |
| 6. Problem set: Wald's theorem                            | 134 |
| <b>7. Analysis of Covariance</b>                          | 138 |
| 1. Preliminaries on non-orthogonal projections            | 138 |
| 1A. Characterization of projections                       | 139 |
| 1B. The adjoint of a projection                           | 139 |
| 1C. An isomorphism between $J$ and $I^\perp$              | 140 |
| 1D. A formula for $P_{J;I}$ when $J$ is given by a basis  | 140 |
| 1E. A formula for $P'_{J;I}$ when $J$ is given by a basis | 142 |
| 2. The analysis of covariance framework                   | 142 |
| 3. Gauss-Markov estimation                                | 144 |
| 4. Variances and covariances of GMEs                      | 146 |
| 5. Estimation of $\sigma^2$                               | 149 |
| 6. Scheffé intervals for functionals of $\mu_M$           | 150 |
| 7. $F$ -Testing                                           | 152 |
| 8. Problem set: The Latin square design                   | 155 |
| <b>8. Missing Observations</b>                            | 160 |
| 1. Framework; Gauss-Markov estimation                     | 160 |
| 2. Obtaining $\hat{\mu}$                                  | 165 |
| 2A. The consistency equation method                       | 165 |
| 2B. The quadratic function method                         | 168 |
| 2C. The analysis of covariance method                     | 169 |
| 3. Estimation of $\sigma^2$                               | 171 |

CONTENTS

vii

|                                               |            |
|-----------------------------------------------|------------|
| 4. $F$ -testing . . . . .                     | 173        |
| 5. Estimation of linear functionals . . . . . | 177        |
| 6. Problem set: Extra observations . . . . .  | 179        |
| <b>9. References</b> . . . . .                | <b>184</b> |
| <b>Index</b> . . . . .                        | <b>186</b> |





## CHAPTER 1

## INTRODUCTION

In this chapter we introduce and contrast the matricial and geometric formulations of the so-called general linear model and describe the organization of the rest of these notes.

## 1. Orientation

Recall the classical framework of the *general linear model (GLM)*. One is given an  $n$ -dimensional random vector  $\mathbf{Y}^{n \times 1} = (Y_1, \dots, Y_n)^T$ , perhaps multivariate normally distributed, with covariance matrix  $(\text{Cov}(Y_i, Y_j))^{n \times n} = \sigma^2 \mathbf{I}^{n \times n}$  and mean vector  $\boldsymbol{\mu}^{n \times 1} = E(\mathbf{Y}) = (EY_1, \dots, EY_n)^T$  of the form

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta},$$

where  $\mathbf{X}^{n \times p}$  is known and  $\sigma^2$  and  $\boldsymbol{\beta}^{p \times 1} = (\beta_1, \dots, \beta_p)^T$  are unknown; in addition the  $\beta_i$ 's may be subject to linear constraints  $\mathbf{R}\boldsymbol{\beta} = 0$ , where  $\mathbf{R}^{c \times p}$  is known.  $\mathbf{X}$  is called the *design*, or *regression matrix*, and  $\boldsymbol{\beta}$  is called the *parameter vector*.

**1.1 Example.** In the classical *two-sample problem*, one has

$$\mathbf{X}^T = \left( \underbrace{1 \ 1 \ \dots \ 1}_{n_1 \text{ times}} \ \underbrace{0 \ 0 \ \dots \ 0}_{n_2 \text{ times}} \right) \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

i.e.,

$$E(Y_i) = \begin{cases} \mu_1, & \text{if } 1 \leq i \leq n_1, \\ \mu_2, & \text{if } n_1 < i \leq n_1 + n_2 = n \end{cases} \quad \bullet$$

**1.2 Example.** In *simple linear regression*, one has

$$\mathbf{X}^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} a \\ b \end{pmatrix},$$

i.e.,

$$E(Y_i) = a + bx_i \quad \text{for } i = 1, \dots, n. \quad \bullet$$

Typical problems are the estimation of linear combinations of the  $\beta_i$ 's, testing that some such linear combinations are 0 (or some other prescribed value), and estimation of  $\sigma^2$ .

**1.3 Example.** In the two-sample problem, one is often interested in estimating the difference  $\mu_2 - \mu_1$  or in testing the null hypothesis that  $\mu_1 = \mu_2$ . •

**1.4 Example.** In simple linear regression, one seeks estimates of the intercept  $a$  and slope  $b$  and may want to test, e.g., the hypothesis that  $b = 0$ , or the hypothesis that  $b = 1$ . •

If you've had some prior statistical training, you may well have already encountered the resolution of these problems. You may know, for example, that provided  $\mathbf{X}$  is of full rank and no linear constraints are imposed on  $\boldsymbol{\beta}$ , the *best (minimum variance) linear unbiased estimator (BLUE)* of  $\sum_{1 \leq i \leq p} c_i \beta_i$  is  $\sum_{1 \leq i \leq p} c_i \hat{\beta}_i$ , where

$$(\hat{\beta}_1, \dots, \hat{\beta}_p)^T = \mathbf{C}\mathbf{X}^T\mathbf{Y}, \quad \text{with } \mathbf{C} = \mathbf{A}^{-1}, \quad \mathbf{A} = \mathbf{X}^T\mathbf{X};$$

this is called the *Gauss-Markov theorem*.

In these notes we will be studying the GLM from a geometric point of view, using linear algebra in place of matrix algebra. Although we will not reach any conclusions that could not be obtained solely by matrix techniques, the basic ideas will emerge more clearly. With the added intuitive feeling and mathematical insight this provides, one will be better able to understand old results and formulate and prove new ones.

From a geometric perspective, the GLM may be described as follows, using some terms which will be defined in subsequent chapters. One is given a *random vector*  $Y$  taking values in some given *inner product space*  $(V, \langle \cdot, \cdot \rangle)$ . It is assumed that  $Y$  has a *weakly spherical covariance operator* and that the *mean*  $\mu$  of  $Y$  lies in a given *manifold*  $M$  of  $V$ ; for purposes of testing it is further assumed that  $Y$  is *normally distributed*. One desires to estimate  $\mu$  (or *linear functionals* of  $\mu$ ) and to test hypotheses such as  $\mu \in M_0$ , where  $M_0$  is a given *submanifold* of  $M$ . The Gauss-Markov theorem says that the BLUE of the linear functional  $\psi(\mu)$  is  $\psi(\hat{\mu})$ , where  $\hat{\mu}$  is the *orthogonal projection* of  $Y$  onto  $M$ . As we will see, this geometric description of the problem encompasses the matricial formulation of the GLM not only as it is set out above (take, e.g.,  $V = \mathbb{R}^n$ ,  $\langle \cdot, \cdot \rangle = \text{dot-product}$ ,  $Y = \mathbf{Y}$ ,  $\mu = \boldsymbol{\mu}$ , and  $M =$  the subspace of  $\mathbb{R}^n$  spanned by the columns of the design matrix  $\mathbf{X}$ ), but also in cases where  $\mathbf{X}$  is of less than full rank and/or linear constraints are imposed on the  $\beta_i$ 's.

## 2. An illustrative example

To illustrate the differences between the matricial and geometric approach-

es, let's compare the ways in which one establishes the independence of

$$\bar{Y} = \hat{\mu} = \frac{\sum_{1 \leq i \leq n} Y_i}{n} \quad \text{and} \quad s^2 = \hat{\sigma}^2 = \frac{1}{n-1} \sum_{1 \leq i \leq n} (Y_i - \bar{Y})^2$$

in the one-sample problem

$$\mathbf{Y}^{n \times 1} \sim N(\mu \mathbf{e}, \sigma^2 \mathbf{I}^{n \times n}) \quad \text{with} \quad \mathbf{e} = (1, 1, \dots, 1)^T. \quad (2.1)$$

(The vector  $\mathbf{e}$  is called the *equiangular vector*.)

The classical matrix proof, which uses some facts about multivariate normal distributions, runs like this. Let  $\mathbf{B}^{n \times n} = (b_{ij})$  be the matrix

$$\begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{-3}{\sqrt{12}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{1}{\sqrt{n(n-1)}} & \frac{-(n-1)}{\sqrt{n(n-1)}} \end{pmatrix}.$$

Note that the rows (and columns) of  $\mathbf{B}$  are orthonormal ( $\sum_{1 \leq k \leq n} b_{ik} b_{jk} = \delta_{ij} \equiv \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$ ) and that the first row is  $\frac{1}{\sqrt{n}} \mathbf{e}^T$ . Set

$$\mathbf{Z} = \mathbf{B}\mathbf{Y}.$$

Then

$$\mathbf{Z} \sim N(\boldsymbol{\nu}, \boldsymbol{\Sigma})$$

with

$$\boldsymbol{\nu} = \mathbf{B}(\mu \mathbf{e}) = \mu \mathbf{B}\mathbf{e} = (\sqrt{n} \mu, 0, \dots, 0)^T$$

and

$$\boldsymbol{\Sigma} = \mathbf{B}(\sigma^2 \mathbf{I})\mathbf{B}^T = \sigma^2 \mathbf{B}\mathbf{B}^T = \sigma^2 \mathbf{I};$$

that is,  $Z_1, Z_2, \dots, Z_n$  are independent normal random variables, each with variance  $\sigma^2$ ,  $E(Z_1) = \sqrt{n} \mu$ , and  $E(Z_j) = 0$  for  $2 \leq j \leq n$ . Moreover

$$Z_1 = \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} Y_i = \sqrt{n} \bar{Y}, \quad \text{or} \quad \bar{Y} = \frac{Z_1}{\sqrt{n}},$$

while

$$\begin{aligned} (n-1)s^2 &= \sum_{1 \leq i \leq n} (Y_i - \bar{Y})^2 = \sum_{1 \leq i \leq n} Y_i^2 - n\bar{Y}^2 \\ &= \sum_{1 \leq i \leq n} Y_i^2 - Z_1^2 = \sum_{1 \leq i \leq n} Z_i^2 - Z_1^2 = \sum_{2 \leq i \leq n} Z_i^2 \end{aligned} \quad (2.2)$$

because

$$\sum_{1 \leq i \leq n} Z_i^2 = \mathbf{Z}^T \mathbf{Z} = \mathbf{Y}^T \mathbf{B}^T \mathbf{B} \mathbf{Y} = \mathbf{Y}^T \mathbf{Y} = \sum_{1 \leq i \leq n} Y_i^2.$$

This gives the independence of  $\bar{Y}$  and  $s^2$  and it's an easy step to get the marginal distributions:  $\bar{Y} \sim N(\mu, \sigma^2/n)$  and  $(n-1)s^2/\sigma^2 \sim \chi_{n-1}^2$ .

What is the nature of the transformation  $\mathbf{Z} = \mathbf{B}\mathbf{Y}$ ? Let  $\mathbf{b}_1 = \mathbf{e}/\sqrt{n}$ ,  $\mathbf{b}_2, \dots, \mathbf{b}_n$  denote the transposes of the rows of  $\mathbf{B}$ . The coordinates of  $\mathbf{Y} = \sum_{1 \leq j \leq n} C_j \mathbf{b}_j$  with respect to this new orthonormal basis for  $\mathbb{R}^n$  are given by

$$C_i = \mathbf{b}_i^T \mathbf{Y} = Z_i, \quad i = 1, \dots, n.$$

The effect of the change of coordinates  $\mathbf{Y} \rightarrow \mathbf{Z}$  is to split  $\mathbf{Y}$  into its components along, and orthogonal to, the equiangular vector  $\mathbf{e}$ .

Now I'll show you the geometric proof, which uses some properties of (weakly) spherical normal random vectors taking values in an inner product space  $(V, \langle \cdot, \cdot \rangle)$ , here  $(\mathbb{R}^n, \text{dot-product})$ . The assumptions imply that  $\mathbf{Y}$  is spherical normally distributed about its mean  $E(\mathbf{Y})$  and that  $E(\mathbf{Y})$  lies in the manifold  $M$  spanned by  $\mathbf{e}$ . Let  $P_M$  denote orthogonal projection onto  $M$ ,  $Q_M$  orthogonal projection onto the orthogonal complement  $M^\perp$  of  $M$ . Basic distribution theory says that  $P_M \mathbf{Y}$  and  $Q_M \mathbf{Y}$  are independent. But

$$P_M \mathbf{Y} = \frac{\langle \mathbf{e}, \mathbf{Y} \rangle}{\langle \mathbf{e}, \mathbf{e} \rangle} \mathbf{e} = \bar{Y} \mathbf{e} \quad (2.3)$$

and

$$Q_M \mathbf{Y} = \mathbf{Y} - P_M \mathbf{Y} = \mathbf{Y} - \bar{Y} \mathbf{e} = (Y_1 - \bar{Y}, \dots, Y_n - \bar{Y})^T;$$

it follows that  $\bar{Y}$  and  $(n-1)s^2 = \sum_{1 \leq i \leq n} (Y_i - \bar{Y})^2 = \|Q_M \mathbf{Y}\|^2$  are independent. Again it is an easy matter to get the marginal distributions.

To my way of thinking, granted the technical apparatus, the second proof is clearer, being more to the point. The first proof does the same things, but (to the uninitiated) in an obscure manner.

### 3. Overview

Chapter 2 covers fundamental concepts from linear algebra, such as the notion of orthogonal projection. Basic distribution theory for random vectors taking values in inner product spaces is developed in Chapter 3. The “geometric” version of the Gauss-Markov theorem is discussed in Chapter 4, and optimal properties of Gauss-Markov estimation under the assumption of normality are considered in Chapter 5.  $F$ -testing of null hypotheses and the related issue of interval estimation are taken up in Chapter 6. Chapters 7 and 8 deal with the analysis of covariance and missing observations respectively.

From the perspective of mathematical statistics, there are some very elegant results, and some notable surprises, connected with the optimality theory for Gauss-Markov estimation /  $F$ -testing under the assumption of

normality. In the sections that deal with these matters—in particular Sections 5.4, 5.5, 6.3, 6.4, and 6.6—the mathematics is somewhat harder than elsewhere, corresponding to the greater depth of the theory.

These notes are aimed at students who have already had some exposure to the matricial formulation of the GLM, perhaps through a methodology course, and are interested in the underlying theory. Each of the following chapters contains numerous exercises, along with a problem set which develops some topic complementing the material in that chapter. Most of the exercises are easy but, I hope, instructive. I typically devote a substantial amount of class time to having students present solutions to the exercises. Some exercises foreshadow what’s to come, by covering a special case of material that will be presented in full generality later on. Moreover, the assertions of some exercises are appealed to later on the text. If you are working through the book on your own, you should at least read over each exercise, even if you don’t work things out. Each exercise is assigned a difficulty level using the syntax “Exercise [ $d$ ]”, where  $d$  is an integer in the range 1 to 5 — the larger is  $d$ , the harder the exercise. The value of  $d$  depends both on the intrinsic difficulty of the exercise and the length of time needed to write up the solution. The problem sets are somewhat harder than the exercises and require a sustained effort for their completion.

Chapters are organized into sections. Within each section of the current chapter, enumerated items are numbered consecutively in the form (*section number.item number*). References to items in a different chapter take the expanded form (*chapter number.section number.item number*). For example, (2.4) refers to the 4<sup>th</sup> numbered item (which may be an example, exercise, theorem, formula, or whatever) in the 2<sup>nd</sup> section of the current chapter, while (6.1.3) refers to the 3<sup>rd</sup> numbered item in the 1<sup>st</sup> section of the 6<sup>th</sup> chapter.

The end of: a proof is marked by a ■; of an example, by a ●; of an exercise, by a ◇; of a part of problem set, by a ○.

To help distinguish between the matricial and geometric points of view, matrices, including row and column vectors, are written in *italic boldface* type while linear transformations and elements of abstract vector spaces are written simply in *italic* type. We speak, for example, of the design matrix  $\mathbf{X}$ , but of vectors  $v$  and  $w$  in an inner product space  $V$ .

## CHAPTER 2

## TOPICS IN LINEAR ALGEBRA

In this chapter we discuss some topics from linear algebra which play a central role in the geometrical analysis of the GLM. The notion of orthogonal projection in an inner product space is introduced in Section 2.1 and studied in Section 2.2. A class of orthogonal decompositions that are useful in the design of experiments is studied in Section 2.3. The spectral representation of self-adjoint transformations is developed in Section 2.4. Linear and bilinear functionals are discussed in Section 2.5. The chapter closes with a problem set in Section 2.6, followed by an Appendix containing a brief review of the basic definitions and facts from linear algebra with which we presume the reader is already familiar.

## 1. Orthogonal projections

Throughout these notes we operate in the context of a finite-dimensional *inner product space*  $(V, \langle \cdot, \cdot \rangle)$  —  $V$  is a finite-dimensional vector space and  $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$  an *inner product*:

- (i) (*positive definiteness*)  $\langle x, x \rangle \geq 0$  for all  $x \in V$  and  $\langle x, x \rangle = 0$  if and only if  $x = 0$ ,
- (ii) (*symmetry*)  $\langle x, y \rangle = \langle y, x \rangle$  for all  $x, y \in V$ ,
- (iii) (*bilinearity*) For all  $c_1, c_2 \in \mathbb{R}$  and  $x_1, x_2, x, y_1, y_2, y \in V$ , one has

$$\begin{aligned}\langle c_1x_1 + c_2x_2, y \rangle &= c_1\langle x_1, y \rangle + c_2\langle x_2, y \rangle \\ \langle x, c_1y_1 + c_2y_2 \rangle &= c_1\langle x, y_1 \rangle + c_2\langle x, y_2 \rangle.\end{aligned}$$

The canonical example is  $V = \mathbb{R}^n$  endowed with the dot-product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = \sum_{1 \leq i \leq n} x_i y_i = \mathbf{x}^T \mathbf{y}$$

for  $\mathbf{x} = (x_1, \dots, x_n)^T$  and  $\mathbf{y} = (y_1, \dots, y_n)^T$ . Unless specifically stated to the contrary, we always view  $\mathbb{R}^n$  as endowed with the dot-product.

Two vectors  $x$  and  $y$  in  $V$  are said to be *perpendicular*, or *orthogonal* (with respect to  $\langle \cdot, \cdot \rangle$ ), if

$$\langle x, y \rangle = 0;$$

one writes

$$x \perp y.$$

The quantity

$$\|x\| = \sqrt{\langle x, x \rangle}$$

is called the *length*, or *norm* of  $x$ . The squared length of the sum of two vectors is given by

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle = \langle x, x + y \rangle + \langle y, x + y \rangle \\ &= \|x\|^2 + 2\langle x, y \rangle + \|y\|^2, \end{aligned}$$

which reduces to the *Pythagorean theorem*

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 \quad (1.1)$$

when  $x \perp y$ .

**1.2 Exercise** [1]. Let  $v_1$  and  $v_2$  be two non-zero vectors in  $\mathbb{R}^2$  and let  $\theta$  be the angle between them, measured counter-clockwise from  $v_1$  to  $v_2$ . Show that

$$\cos(\theta) = \frac{\langle v_1, v_2 \rangle}{\|v_1\| \|v_2\|}$$

and deduce that  $v_1 \perp v_2$  if and only if  $\theta = 90^\circ$  or  $270^\circ$ .

[Hint: Use the identity  $\cos(\theta_2 - \theta_1) = \cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2)$ .]  $\diamond$

**1.3 Exercise** [3]. Let  $x$ ,  $y$  and  $z$  be the vectors in  $\mathbb{R}^n$  given by

$$x_i = 1, \quad y_i = i - \frac{n+1}{2}, \quad \text{and} \quad z_i = \left(i - \frac{n+1}{2}\right)^2 - \frac{n^2-1}{12} \quad \text{for } 1 \leq i \leq n. \quad (1.4)$$

Show that  $x$ ,  $y$ , and  $z$  are mutually orthogonal and span the same subspace as do  $(1, 1, \dots, 1)^T$ ,  $(1, 2, \dots, n)^T$ , and  $(1, 4, \dots, n^2)^T$ . Exhibit  $x$ ,  $y$ , and  $z$  explicitly for  $n = 5$ .  $\diamond$

**1.5 Exercise** [1]. Let  $\mathcal{O}: V \rightarrow V$  be a linear transformation. Show that  $\mathcal{O}$  preserves lengths:

$$\|\mathcal{O}x\| = \|x\| \quad \text{for all } x \in V$$

if and only if it preserves inner products:

$$\langle \mathcal{O}x, \mathcal{O}y \rangle = \langle x, y \rangle \quad \text{for all } x, y \in V.$$

Such a transformation is said to be *orthogonal*.

[Hint: Observe

$$\langle x, y \rangle = \frac{\|x + y\|^2 - \|x\|^2 - \|y\|^2}{2} \quad (1.6)$$

for  $x, y \in V$ .]  $\diamond$

**1.7 Exercise** [2]. (1) Let  $v_1$  and  $v_2$  be elements of  $V$ . Show that  $v_1 = v_2$  if and only if  $\langle v_1, w \rangle = \langle v_2, w \rangle$  for each  $w \in V$ , or just for each  $w$  in some basis for  $V$ . (2) Let  $T_1$  and  $T_2$  be two linear transformations of  $V$ . Show that  $T_1 = T_2$  if and only if  $\langle v, T_1 w \rangle = \langle v, T_2 w \rangle$  for each  $v$  and  $w$  in  $V$ , or just for each  $v$  and  $w$  in some basis for  $V$ .  $\diamond$

As intimated in the introduction, the notion of orthogonal projection onto subspaces of  $V$  plays a key role in the study of the GLM. We begin our study of projections with the following seminal result.

**1.8 Theorem.** *Suppose that  $M$  is a subspace of  $V$  and that  $x \in V$ . There is exactly one vector  $m \in M$  such that the residual  $x - m$  is orthogonal to  $M$ :*

$$(x - m) \perp y \quad \text{for all } y \in M, \quad (1.9)$$

or, equivalently, such that  $m$  is closest to  $x$ :

$$\|x - m\| = \inf\{\|x - y\| : y \in M\}. \quad (1.10)$$

The proof will be given shortly. The unique  $m \in M$  such that (1.9) and (1.10) hold is called the *orthogonal projection of  $x$  onto  $M$* , written  $P_M x$ , and the mapping  $P_M$  which sends  $x \in V$  to  $P_M x$  is called *orthogonal projection onto  $M$* . In the context of  $\mathbb{R}^n$  with  $\mathbf{x} = (x_i)$  and  $\mathbf{m} = (m_i)$ , (1.9) reads

$$\sum_{1 \leq i \leq n} (x_i - m_i)y_i = 0 \quad \text{for all } \mathbf{y} = (y_i) \in M,$$

while (1.10) is the *least squares characterization* of  $\mathbf{m}$ :

$$\sum_{1 \leq i \leq n} (x_i - m_i)^2 = \inf\left\{\sum_{1 \leq i \leq n} (x_i - y_i)^2 : \mathbf{y} \in M\right\}.$$

**1.11 Exercise** [1]. Let  $M$  be a subspace of  $V$ . (a) Show that

$$\|x - P_M x\|^2 = \|x\|^2 - \|P_M x\|^2 \quad (1.12)$$

for all  $x \in V$ . (b) Deduce that

$$\|P_M x\| \leq \|x\| \quad (1.13)$$

for all  $x \in V$ , with equality holding if and only if  $x \in M$ .  $\diamond$

**Proof of Theorem 1.8.** (1.9) *implies* (1.10): Suppose  $m \in M$  satisfies (1.9). Then for all  $y \in M$  the Pythagorean theorem gives

$$\|x - y\|^2 = \|(x - m) + (m - y)\|^2 = \|x - m\|^2 + \|m - y\|^2,$$

so  $m$  satisfies (1.10).

(1.10) *implies* (1.9): Suppose  $m \in M$  satisfies (1.10). Then for any  $0 \neq y \in M$  and any  $\delta \in \mathbb{R}$ ,

$$\|x - m\|^2 \leq \|x - m + \delta y\|^2 = \|x - m\|^2 + 2\delta \langle x - m, y \rangle + \delta^2 \|y\|^2;$$

this relation forces  $\langle x - m, y \rangle = 0$ .



*Uniqueness:* If

$$m_1 + y_1 = x = m_2 + y_2$$

with  $m_i \in M$  and  $y_i \perp M$  for  $i = 1$  and  $2$ , then the vector

$$m_1 - m_2 = y_2 - y_1$$

lies in  $M$  and is perpendicular to  $M$ , and so is perpendicular to itself:

$$0 = \langle m_1 - m_2, m_1 - m_2 \rangle = \|m_1 - m_2\|^2,$$

whence  $m_1 = m_2$  by the positive-definiteness of  $\langle \cdot, \cdot \rangle$ .

*Existence:* We will show in a moment that  $M$  has a basis  $m_1, \dots, m_k$  consisting of mutually orthogonal vectors. For any such basis the generic  $m = \sum_{1 \leq j \leq k} c_j m_j$  in  $M$  satisfies

$$(x - m) \perp M$$

if and only if  $x - m$  is orthogonal to each  $m_i$ , i.e., if and only if

$$\langle m_i, x \rangle = \langle m_i, \sum_{1 \leq j \leq k} c_j m_j \rangle = \sum_{1 \leq j \leq k} \langle m_i, m_j \rangle c_j = \langle m_i, m_i \rangle c_i$$

for  $1 \leq i \leq k$ . It follows that we can take

$$P_M x = \sum_{1 \leq i \leq k} \frac{\langle m_i, x \rangle}{\langle m_i, m_i \rangle} m_i. \quad (1.14)$$

To produce an orthogonal basis for  $M$ , let  $m_1^*, \dots, m_k^*$  be any basis for  $M$  and inductively define new basis vectors  $m_1, \dots, m_k$  by the recipe  $m_1 = m_1^*$  and

$$\begin{aligned} m_j &= m_j^* - P_{[m_1^*, \dots, m_{j-1}^*]} m_j^* = m_j^* - P_{[m_1, \dots, m_{j-1}]} m_j^* \\ &= m_j^* - \sum_{1 \leq i \leq j-1} \frac{\langle m_j^*, m_i \rangle}{\langle m_i, m_i \rangle} m_i \end{aligned} \quad (1.15)$$

for  $j = 2, \dots, k$ ; here  $[m_1^*, \dots, m_{j-1}^*]$  denotes the span of  $m_1^*, \dots, m_{j-1}^*$  and  $[m_1, \dots, m_{j-1}]$  denotes the (identical) span of  $m_1, \dots, m_{j-1}$  (see Subsection 2.6.1). ■

The recursive scheme for cranking out the  $m_j$ 's above is called *Gram-Schmidt orthogonalization*. As a special case of (1.14) we have the following simple, yet key, formula for projecting onto a one-dimensional space:

$$P_{[m]} x = \frac{\langle x, m \rangle}{\langle m, m \rangle} m \quad \text{for } m \neq 0. \quad (1.16)$$

**1.17 Example.** In the context of  $\mathbb{R}^n$  take  $\mathbf{m} = \mathbf{e} \equiv (1, \dots, 1)^T$ . Formula (1.16) then reads

$$P_{[\mathbf{e}]} \mathbf{x} = \frac{\langle \mathbf{x}, \mathbf{e} \rangle}{\langle \mathbf{e}, \mathbf{e} \rangle} \mathbf{e} = \frac{\sum_i x_i}{\sum_i 1} \mathbf{e} = \bar{x} \mathbf{e} = (\bar{x}, \dots, \bar{x})^T;$$

we used this result in the introduction (see (1.2.3)). Formula (1.12) reads

$$\sum_i (x_i - \bar{x})^2 = \|\mathbf{x} - \bar{x}\mathbf{e}\|^2 = \|\mathbf{x}\|^2 - \|\bar{x}\mathbf{e}\|^2 = \sum_i x_i^2 - n\bar{x}^2;$$

this is just the computing formula for  $(n-1)s^2$  used at (1.2.2). According to (1.10),  $c = \bar{x}$  minimizes the sum of squares  $\sum_i (x_i - c)^2$ . •

**1.18 Exercise** [2]. Use the preceding techniques to compute  $P_M \mathbf{y}$  for  $\mathbf{y} \in \mathbb{R}^n$  and  $M = [\mathbf{e}, (x_1, \dots, x_n)^T]$ , the manifold spanned by the columns of the design matrix for simple linear regression. ◊

**1.19 Exercise** [3]. Show that the transposes of the rows of the matrix  $\mathbf{B}$  of Section 1.2 result from first applying the Gram-Schmidt orthogonalization scheme to the vectors  $\mathbf{e}$ ,  $(1, -1, 0, \dots, 0)^T$ ,  $(0, 1, -1, 0, \dots, 0)^T$ ,  $\dots$ ,  $(0, \dots, 0, 1, -1)^T$  in  $\mathbb{R}^n$  and then normalizing to unit length. ◊

**1.20 Exercise** [2]. Suppose  $x, y \in V$ . Prove the *Cauchy-Schwarz inequality*:

$$|\langle x, y \rangle| \leq \|x\| \|y\|, \quad (1.21)$$

with equality if and only if  $x$  and  $y$  are linearly dependent. Deduce *Minkowski's inequality*:

$$\|x + y\| \leq \|x\| + \|y\|. \quad (1.22)$$

[Hint: For (1.21), take  $m = y$  in (1.16) and use part (b) of Exercise 1.11.] ◊

**1.23 Exercise** [4]. Define  $d: V \times V \rightarrow \mathbb{R}$  by  $d(x, y) = \|y - x\|$ . Show that  $d$  is a metric on  $V$  such that the set  $\{x \in V : \|x\| = 1\}$  is compact. ◊

**1.24 Exercise** [3]. Show that for any two subspaces  $M$  and  $N$  of  $V$ ,

$$\sup\{\|P_M x\| : x \in N \text{ and } \|x\| = 1\} \leq 1, \quad (1.25)$$

with equality holding if and only if  $M$  and  $N$  have a non-zero vector in common. [Hint: A continuous real-valued function on a compact set attains its maximum.] ◊

To close this section we generalize (1.14) to cover the case of an arbitrary basis for  $M$ . Suppose then that the basis vectors  $m_1, \dots, m_k$  are not necessarily orthogonal and let  $x \in V$ . As in the derivation of (1.14),

$$P_M x = \sum_j c_j m_j$$

is determined by the condition

$$m_i \perp (x - P_M x) \quad \text{for } i = 1, \dots, k,$$

i.e., by the so-called *normal equations*

$$\sum_{1 \leq j \leq k} \langle m_i, m_j \rangle c_j = \langle m_i, x \rangle, \quad i = 1, \dots, k. \quad (1.26)$$

In matrix notation (1.26) reads

$$\langle m, m \rangle \mathbf{c} = \langle m, x \rangle$$

so that

$$\mathbf{c} = \langle m, m \rangle^{-1} \langle m, x \rangle,$$

the  $k \times k$  matrix  $\langle m, m \rangle$  and the  $k \times 1$  column vectors  $\langle m, x \rangle$  and  $\mathbf{c}$  being given by

$$(\langle m, m \rangle)_{ij} = \langle m_i, m_j \rangle, \quad (\langle m, x \rangle)_i = \langle m_i, x \rangle, \quad \text{and} \quad (\mathbf{c})_i = c_i$$

for  $i, j = 1, \dots, k$ . If the  $m_i$ 's are mutually orthogonal,  $\langle m, m \rangle$  is diagonal and (1.14) follows immediately.

**1.27 Exercise** [3]. Show that the matrix  $\langle m, m \rangle$  above is in fact nonsingular, so that the inverse  $\langle m, m \rangle^{-1}$  exists.

[Hint: Show that  $\sum_j \langle m_i, m_j \rangle c_j = 0$  for each  $i$  implies  $c_1 = \dots = c_k = 0$ .]  $\diamond$

**1.28 Exercise** [2]. Redo Exercise 1.18 using (1.26). Check that the two ways of calculating  $P_M y$  do in fact give the same result.  $\diamond$

**1.29 Exercise** [2]. Let  $M$  be the subspace of  $\mathbb{R}^n$  spanned by the columns of an  $n \times k$  matrix  $\mathbf{X}$ . Supposing these columns to be linearly independent, use (1.26) to show that the matrix representing  $P_M$  with respect to the usual coordinate basis of  $\mathbb{R}^n$  is  $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ . Exhibit this matrix explicitly in the case  $\mathbf{X} = e$ .  $\diamond$

**1.30 Exercise** [4]. Suppose  $X_0, X_1, \dots, X_k$  are square integrable random variables defined on a common probability space. Consider using  $X_1, \dots, X_k$  to predict  $X_0$ . Show that among predictors of the form

$$\hat{X}_0 = c_0 + \sum_{1 \leq i \leq k} c_i X_i,$$

the  $c$ 's being constants, the one minimizing the *mean square error of prediction*

$$E(\hat{X}_0 - X_0)^2$$

is

$$\hat{X}_0 = \mu_0 + \sum_{1 \leq i \leq k} \left( \sum_{1 \leq j \leq k} \sigma^{ij} \sigma_{j0} \right) (X_i - \mu_i),$$

where

$$\begin{aligned} \mu_i &= E(X_i), & 0 \leq i \leq k, \\ \sigma_{ij} &= \text{Cov}(X_i, X_j), & 0 \leq i, j \leq k, \end{aligned}$$

and the  $k \times k$  matrix  $(\sigma^{ij})_{1 \leq i, j \leq k}$  is the inverse of the matrix  $(\sigma_{ij})_{1 \leq i, j \leq k}$ , the latter assumed to be nonsingular.

[Hint: This is just a matter of projecting  $X_0 - \mu_0$  onto the subspace spanned by  $1, X_1 - \mu_1, \dots, X_k - \mu_k$  in the space  $\mathcal{L}_2$  of square integrable random variables on the given probability space, the inner product between two variables  $A$  and  $B$  being  $E(AB)$ .]  $\diamond$

## 2. Properties of orthogonal projections

This section develops properties of orthogonal projections. The results will prove to be useful in analyzing the GLM.

Let  $(V, \langle \cdot, \cdot \rangle)$  be an inner product space, let  $M$  be a subspace of  $V$ , and let

$$M^\perp = \{x \in V : x \perp M\} \equiv \{x \in V : x \perp y \text{ for all } y \in M\} \quad (2.1)$$

be the *orthogonal complement* of  $M$ . Note that  $M^\perp$  is a subspace of  $V$ . Let  $P_M$  and  $Q_M$  denote orthogonal projection onto  $M$  and  $M^\perp$  respectively and let  $x \in V$ . By Theorem 1.8,

$$x = P_M x + Q_M x$$

is the unique representation of  $x$  as the sum of an element of  $M$  and an element of  $M^\perp$ .

**2.2 Exercise** [2]. Let  $M$  and  $M_1, \dots, M_k$  be subspaces of  $V$ . Show that

$$\begin{aligned} d(M^\perp) &= d(V) - d(M), \\ (M^\perp)^\perp &= M, \\ M_1 \subset M_2 &\iff M_1^\perp \supset M_2^\perp, \\ \left(\sum_{1 \leq i \leq k} M_i\right)^\perp &= \bigcap_{1 \leq i \leq k} M_i^\perp, \\ \left(\bigcap_{1 \leq i \leq k} M_i\right)^\perp &= \sum_{1 \leq i \leq k} M_i^\perp. \quad \diamond \end{aligned}$$

### 2A. Characterization of orthogonal projections

A linear transformation  $T: V \rightarrow V$  is said to be *idempotent* if  $T^2 = T$  and said to be *self-adjoint* if

$$\langle Tx, y \rangle = \langle x, Ty \rangle \quad (2.3)$$

for all  $x, y$  in  $V$ .

**2.4 Proposition.** (i)  $P_M$  is an idempotent, self-adjoint, linear transformation with range  $M$  and null space  $M^\perp$ .

(ii) Conversely, an idempotent, self-adjoint, linear transformation  $T$  mapping  $V$  into  $V$  is the orthogonal projection onto its range.

**Proof.** (i): Write  $P$  for  $P_M$  and  $Q$  for  $Q_M$ . The properties of  $P_M$  are all immediate consequences of the orthogonality of  $M$  and  $M^\perp$  and the uniqueness of the decompositions

$$\begin{aligned} x &= Px + Qx && (Px \in M, \quad Qx \in M^\perp) \\ y &= Py + Qy && (Py \in M, \quad Qy \in M^\perp), \end{aligned}$$

to wit:

(a)  $P$  is linear, since  $x + y = Px + Qx + Py + Qy$  implies  $P(x + y) = Px + Py$  and  $cx = cPx + cQx$  implies  $P(cx) = c(Px)$ ;

(b)  $P$  is idempotent, since  $Px = Px + 0$  implies  $P^2x = Px$ ;

(c)  $P$  is self-adjoint, since  $\langle Px, y \rangle = \langle Px, Py + Qy \rangle = \langle Px, Py \rangle = \langle Px + Qx, Py \rangle = \langle x, Py \rangle$  for all  $x, y \in V$ ;

(d)  $M$  is the range of  $P$ , since on the one hand  $x \in V$  implies  $Px \in M$  and on the other hand  $x \in M$  implies  $x = Px$ ;

(e)  $M^\perp$  is the null space of  $P$ , since  $x \in M^\perp \iff x = Qx \iff Px = 0$ .

(ii): For  $x \in V$  write

$$x = Tx + (x - Tx).$$

Since  $Tx \in \mathcal{R}(T)$ , we need only show  $x - Tx \in (\mathcal{R}(T))^\perp$ . For this observe that for all  $y \in V$ , one has  $\langle x - Tx, Ty \rangle = \langle T(x - Tx), y \rangle = \langle Tx - T^2x, y \rangle = \langle Tx - Tx, y \rangle = 0$ . ■

**2.5 Exercise** [2]. Given  $0 \neq m \in M$ , show that the transformation

$$T: v \rightarrow \frac{\langle v, m \rangle}{\langle m, m \rangle} m$$

has all the properties required to characterize it as orthogonal projection onto  $[m]$ . ◇

**2.6 Exercise** [2]. Let  $M$  be a subspace of  $V$ . Show that the orthogonal linear transformation  $T: V \rightarrow V$  which reflects each  $x \in M$  through the origin and leaves each  $x \in M^\perp$  fixed is  $T = I - 2P_M$ . ◇

**2.7 Exercise** [2]. Let  $T: V \rightarrow V$  be self-adjoint. Show that

$$(\mathcal{R}(T))^\perp = \mathcal{N}(T) \tag{2.8}$$

and deduce

$$\mathcal{R}(T^2) = \mathcal{R}(T). \tag{2.9}$$

[Hint: To get (2.9), first show  $\mathcal{N}(T^2) \subset \mathcal{N}(T)$ .] ◇

## 2B. Differences of orthogonal projections

If  $M$  and  $N$  are two subspaces of  $V$  with  $M \subset N$ , the subspace

$$N - M = \{x \in N : x \perp M\} = N \cap M^\perp \tag{2.10}$$

is called the (*relative*) *orthogonal complement* of  $M$  in  $N$ .

**2.11 Exercise** [2]. Let  $M$  and  $N$  be subspaces of  $V$  with  $M \subset N$ . Show that if the vectors  $v_1, \dots, v_k$  span  $N$ , then the vectors  $Q_M v_1, \dots, Q_M v_k$  span  $N - M$ . ◇

For self-adjoint transformations  $S$  and  $T$  mapping  $V$  to  $V$ , the notation

$$S \leq T \tag{2.12}$$

means  $\langle x, Sx \rangle \leq \langle x, Tx \rangle$  for all  $x \in V$ .

**2.13 Proposition.** *Let  $M$  and  $N$  be subspaces of  $V$ . The following are equivalent:*

- (i)  $P_N - P_M$  is an orthogonal projection,
- (ii)  $M \subset N$ ,
- (iii)  $P_M \leq P_N$ ,
- (iv)  $\|P_N x\|^2 \geq \|P_M x\|^2$  for all  $x \in V$ ,
- (v)  $P_M = P_M P_N$ ,

and

- (vi)  $P_N - P_M = P_{N-M}$ .

**Proof.** (ii) *implies* (i) and (vi): Suppose  $M \subset N$ . We will show

$$P_N = P_M + P_{N-M}$$

so that (i) and (vi) hold.  $P_M + P_{N-M}$  is a self-adjoint, linear transformation which is idempotent ( $(P_M + P_{N-M})^2 = P_M^2 + P_M P_{N-M} + P_{N-M} P_M + P_{N-M}^2 = P_M + 0 + 0 + P_{N-M}$ ); its range is contained in  $N$ , and yet contains  $N$  because it contains both  $M$  and  $N - M$ . According to Proposition 2.4, these properties characterize  $P_M + P_{N-M}$  as  $P_N$ .

(i) *implies* (iii): Write  $P$  for  $P_N - P_M$ . For all  $x \in V$  one has

$$\begin{aligned} \langle P_N x, x \rangle - \langle P_M x, x \rangle &= \langle (P_N - P_M)x, x \rangle \\ &= \langle P x, x \rangle = \langle P^2 x, x \rangle = \langle P x, P x \rangle = \|P x\|^2 \geq 0. \end{aligned}$$

(iii) *implies* (iv): For all  $x \in V$ ,  $\|P_N x\|^2 = \langle P_N x, P_N x \rangle = \langle x, P_N x \rangle \geq \langle x, P_M x \rangle = \|P_M x\|^2$ .

(iv) *implies* (v): For all  $x \in V$ ,

$$\begin{aligned} \|P_M x\|^2 &= \|P_N(P_M x)\|^2 + \|Q_N(P_M x)\|^2 \\ &\geq \|P_N(P_M x)\|^2 \geq \|P_M(P_M x)\|^2 = \|P_M x\|^2; \end{aligned}$$

this implies  $Q_N P_M = 0$ . Thus for all  $x, y \in V$ ,

$$\langle y, P_M Q_N x \rangle = \langle P_M y, Q_N x \rangle = \langle Q_N P_M y, x \rangle = 0$$

so  $P_M Q_N = 0$  also. Now use

$$P_M = P_M(P_N + Q_N) = P_M P_N + P_M Q_N$$

to conclude  $P_M = P_M P_N$ .

(v) *implies* (ii): Since  $P_M = P_M(P_N + Q_N) = P_M P_N + P_M Q_N$ ,  $P_M = P_M P_N$  entails  $P_M Q_N = 0 \implies N^\perp \subset M^\perp \implies M \subset N$ .

(vi) *implies* (i): Trivial. ■

**2.14 Exercise** [3]. Let  $p$  and  $n_1, \dots, n_p$  be given and let  $V \equiv \{ (x_{ij})_{1 \leq j \leq n_i, 1 \leq i \leq p} : x_{ij} \in \mathbb{R} \}$  be endowed with the dot-product

$$\langle (x_{ij}), (y_{ij}) \rangle = \sum_{1 \leq i \leq p} \sum_{1 \leq j \leq n_i} x_{ij} y_{ij}.$$

Put

$$M = \{ (x_{ij}) \in V : \text{for some } \beta_1, \dots, \beta_p \in \mathbb{R}, x_{ij} = \beta_i \text{ for all } i, j \} = [\mathbf{v}_1, \dots, \mathbf{v}_p]$$

and

$$M_0 = \{ (x_{ij}) \in V : \text{for some } \beta \in \mathbb{R}, x_{ij} = \beta \text{ for all } i, j \} = \left[ \sum_{1 \leq i \leq p} \mathbf{v}_i \right],$$

where  $(\mathbf{v}_i)_{i', j'} = \delta_{ii'}$ . Show that for each  $x \in V$ ,

$$\begin{aligned} (P_M x)_{ij} &= \bar{x}_i \equiv \frac{1}{n_i} \sum_{1 \leq j \leq n_i} x_{ij}, \\ (P_{M_0} x)_{ij} &= \bar{x} \equiv \frac{1}{n} \sum_i n_i \bar{x}_i = \frac{1}{n} \sum_{ij} x_{ij} \end{aligned} \quad (2.15)$$

with  $n = \sum_i n_i$ ; deduce that

$$(P_{M-M_0} x)_{ij} = \bar{x}_i - \bar{x}. \quad (2.16)$$

[Hint:  $v_1, \dots, v_p$  is an orthogonal basis for  $M$ .]  $\diamond$

**2.17 Exercise** [1]. Let  $M$  and  $N$  be subspaces of an inner product space  $V$ . Put  $L = M \cap N$ . Show that  $P_{M-L} P_{N-L} = P_M P_N - P_L$ .  $\diamond$

**2.18 Exercise** [2]. Suppose  $L$  and  $M$  are subspaces of  $V$ . Put

$$K = P_L(M) \equiv \{ P_L m : m \in M \}.$$

Show that

$$P_{V-M} - P_{L-K} = P_{(V-M)-(L-K)}. \quad (2.19)$$

Solve this exercise twice: (1) by arguing that  $L - K \subset M^\perp$  (in fact,  $L - K = L \cap M^\perp$ ); and (2) by arguing (independently of (1)) that  $\|Q_M x\| \geq \|P_{L-K} x\|$  for each  $x \in V$ .  $\diamond$

**2.20 Exercise** [1]. Suppose  $x_1, \dots, x_\ell$  is a basis for a subspace  $N$  of  $V$ . Let  $M = [x_1, \dots, x_k]$  be the subspace spanned by the first  $k$  of the  $x_i$ 's, where  $1 \leq k < \ell$ . For an element  $v$  of  $V$  write

$$P_M v = \sum_{i=1}^k a_i x_i \quad \text{and} \quad P_N v = \sum_{j=1}^{\ell} b_j x_j, \quad (2.21)$$

and for  $j = k+1, \dots, \ell$ , write

$$P_M x_j = \sum_{i=1}^k c_{i,j} x_i; \quad (2.22)$$

the coefficients  $a_i$ ,  $b_j$ , and  $c_{i,j}$  in these sums are of course unique. Show that for  $i = 1, \dots, k$ ,

$$a_i = b_i + \sum_{j=k+1}^{\ell} c_{i,j} b_j. \quad (2.23) \quad \diamond$$

### 2C. Sums of orthogonal projections

Subspaces  $M$  and  $N$  of  $V$  are said to be *orthogonal* (or *perpendicular*), written  $M \perp N$ , if  $m \perp n$  for each  $m \in M$  and  $n \in N$ .

**2.24 Proposition.** *Let  $P_1, P_2, \dots, P_k$  be orthogonal projections onto  $M_1, M_2, \dots, M_k$  respectively. Set  $P = P_1 + P_2 + \dots + P_k$ . The following are equivalent:*

- (i)  $P$  is an orthogonal projection,
- (ii)  $P_i P_j = 0$  for all  $i \neq j$ ,
- (iii) the  $M_i$ 's are mutually orthogonal,

and

- (iv)  $P$  is orthogonal projection onto  $\sum_i M_i$ .

**Proof.** (i) *implies* (ii): For each  $x \in V$ , one has

$$\|x\|^2 \geq \|Px\|^2 = \langle Px, x \rangle = \left\langle \sum_i P_i x, x \right\rangle = \sum_i \langle P_i x, x \rangle = \sum_i \|P_i x\|^2.$$

Taking  $x = P_j y$  gives

$$\|P_j y\|^2 \geq \|P_j y\|^2 + \sum_{i \neq j} \|P_i P_j y\|^2.$$

As  $y \in V$  is arbitrary, we get  $P_i P_j = 0$  for  $i \neq j$ .

(ii) *implies* (iii):  $P_i P_j = 0 \implies M_j \subset M_i^\perp \implies M_i \perp M_j$ .

(iii) *implies* (i), (iv): For any two orthogonal subspaces  $K$  and  $L$  of  $V$ , part (vi) of Proposition 2.13 gives  $P_{K+L} = P_K + P_L$ . By induction

$$P = \sum_i P_i = P_{\sum_i M_i}.$$

(iv) *implies* (i): Trivial. ■

The most important thing here is that (iii) implies (iv): given mutually orthogonal subspaces  $M_1, \dots, M_k$ ,

$$P_{\sum_i M_i} = \sum_i P_{M_i}. \tag{2.25}$$

Notice how this generalizes (1.14).

**2.26 Exercise** [3]. Let  $I$  and  $J$  be positive integers and let the space  $V$  of  $I \times J$  matrices  $(x_{ij})_{1 \leq i \leq I, 1 \leq j \leq J}$  be endowed with the dot-product

$$\langle (x_{ij}), (y_{ij}) \rangle = \sum_i \sum_j x_{ij} y_{ij}.$$

Let  $M_R$  be the subspace of row-wise constant matrices ( $x_{ij} = x_{ij'}$ , for all  $i, j, j'$ ),  $M_C$  the subspace of column-wise constant matrices ( $x_{ij} = x_{i'j}$ , for all  $i, i', j$ ),



and  $M_G$  the subspace of constant matrices ( $x_{ij} = x_{i'j'}$ , for all  $i, i', j, j'$ ). Show that

$$(P_{M_R}x)_{ij} = \bar{x}_{i.} = \frac{1}{J} \sum_j x_{ij}, \quad (2.27_R)$$

$$(P_{M_C}x)_{ij} = \bar{x}_{.j} \equiv \frac{1}{I} \sum_i x_{ij}, \quad (2.27_C)$$

$$(P_{M_G}x)_{ij} = \bar{x}_{..} \equiv \frac{1}{IJ} \sum_{ij} x_{ij} \quad (2.27_G)$$

for each  $x \in V$ . Show further that

$$M_G + (M_R - M_G) + (M_C - M_G)$$

is an orthogonal decomposition of

$$M = M_R + M_C,$$

and deduce

$$(P_Mx)_{ij} = \bar{x}_{i.} + \bar{x}_{.j} - \bar{x}_{..} \quad (2.28)$$

[Hint: The  $I \times J$  matrices  $r_1, \dots, r_I$  defined by  $(r_i)_{i'j'} = \delta_{ii'}$  are an orthogonal basis for  $M_R$ .]  $\diamond$

## 2D. Products of orthogonal projections

Two subspaces  $M$  and  $N$  of  $V$  are said to be *book orthogonal*, written  $M \perp_B N$ , if  $(M - L) \perp (N - L)$  with  $L = M \cap N$ . The imagery is that of two consecutive pages of a book which has been opened in such a way that the pages are at right angles to one another. Note that  $M \perp_B N$  if  $M \perp N$ , or if  $M \subset N$ , or if  $M \supset N$ . In particular,  $V$  and the trivial subspace  $0$  are each book orthogonal to every subspace of  $V$ .

**2.29 Proposition.** *Let  $M$  and  $N$  be two subspaces of  $V$ . The following are equivalent:*

- (i)  $P_M P_N$  is an orthogonal projection,
- (ii)  $P_M$  and  $P_N$  commute,
- (iii)  $M \perp_B N$ ,

and

- (iv)  $P_M P_N = P_L = P_N P_M$  with  $L = M \cap N$ .

**Proof.** (iv) *implies* (i): Trivial.

(i) *implies* (ii): Given that  $P_M P_N$  is an orthogonal projection, one has

$$\langle P_M P_N x, y \rangle = \langle x, P_M P_N y \rangle = \langle P_M x, P_N y \rangle = \langle P_N P_M x, y \rangle$$

for all  $x, y \in V$ , so  $P_M P_N = P_N P_M$ .

(ii) *implies* (iv): Set  $P = P_M P_N$ . Using (ii), we have that  $P$  is linear, self-adjoint, and idempotent ( $P^2 = P_M P_N P_M P_N = P_M^2 P_N^2 = P_M P_N = P$ ); moreover  $\mathcal{R}(P) = L$ , so  $P = P_L$  by Proposition 2.4.

(iv) *is equivalent to* (iii): On the one hand, (iv) means  $P_M P_N - P_L = 0$  and on the other hand, (iii) is equivalent to  $P_{M-L} P_{N-L} = 0$  (use (ii)  $\iff$  (iii) in Proposition 2.24). That (iii) and (iv) are equivalent follows from the identity  $P_{M-L} P_{N-L} = P_M P_N - P_L$  of Exercise 2.17. ■

**2.30 Exercise** [1]. Show that the subspaces  $M_R$  and  $M_C$  of Exercise 2.26 are book orthogonal. ◇

**2.31 Exercise** [1]. In  $V = \mathbb{R}^3$ , let  $\mathbf{x} = (1, 0, 0)^T$ ,  $\mathbf{y} = (0, 1, 0)^T$ , and  $\mathbf{z} = (0, 0, 1)^T$ . Put  $L = [\mathbf{x} + \mathbf{y}]$ ,  $M = [\mathbf{x}, \mathbf{y}]$ , and  $N = [\mathbf{y}, \mathbf{z}]$ . Show that  $L \subset M$  and  $M \perp_B N$ , but it is not the case that  $L \perp_B N$ . ◇

**2.32 Exercise** [2]. Suppose that  $L_1, \dots, L_k$  are mutually orthogonal subspaces of  $V$ . Show that the subspaces

$$M_J = \sum_{j \in J} L_j \tag{2.33}$$

of  $V$  with  $J \subset \{1, 2, \dots, k\}$  are mutually book orthogonal. ◇

**2.34 Exercise** [2]. Show that if  $M$  and  $N$  are book orthogonal subspaces of  $V$ , then so are  $M$  and  $N^\perp$ , and so are  $M^\perp$  and  $N^\perp$ . ◇

**2.35 Exercise** [2]. Suppose that  $M_1, \dots, M_k$  are mutually book orthogonal subspaces of  $V$ . Show that  $\prod_{1 \leq j \leq k} P_{M_j}$  is orthogonal projection onto  $\bigcap_{1 \leq j \leq k} M_j$  and that  $\prod_{1 \leq j \leq k} Q_{M_j}$  is orthogonal projection onto  $(\sum_{1 \leq j \leq k} M_j)^\perp$ . ◇

**2.36 Exercise** [2]. Show that if  $M_1, \dots, M_k$  are mutually book orthogonal subspaces of  $V$ , then so are the subspaces of  $V$  of the form  $\bigcap_{j \in J} M_j$ , where  $J \subset \{1, 2, \dots, k\}$ . ◇

**2.37 Exercise** [3]. Prove the following converse to Exercise 2.32: if  $M_1, \dots, M_n$  are mutually book orthogonal subspaces of  $V$ , then there exist mutually orthogonal subspaces  $L_1, \dots, L_k$  of  $V$  such that each  $M_m$  can be written in the form (2.33).

[Hint: Multiply the identity  $I_V = \prod_{1 \leq \ell \leq n} (P_{M_\ell} + Q_{M_\ell})$  by  $P_{M_m}$  after expanding out the product.] ◇

**2.38 Exercise** [4]. Let  $M$  and  $N$  be arbitrary subspaces of  $V$  and put  $L = M \cap N$ . Let the successive products  $P_M, P_N P_M, P_M P_N P_M, P_N P_M P_N P_M, \dots$  of  $P_M$  and  $P_N$  in alternating order be denoted by  $T_1, T_2, T_3, T_4, \dots$ . Show that for each  $x \in V$ ,

$$T_j x \rightarrow P_L x \quad \text{as } j \rightarrow \infty.$$

[Hint: First observe that  $T_j - P_L$  is the  $j$ -fold product of  $P_{M-L}$  and  $P_{N-L}$  in alternating order and then make use of Exercise 1.24.] ◇

### 2E. An algebraic form of Cochran's theorem

We begin with an elementary lemma:

**2.39 Lemma.** *Suppose  $M, M_1, M_2, \dots, M_k$  are subspaces of  $V$  with  $M = M_1 + M_2 + \dots + M_k$ . The following are equivalent:*

(i) *every vector  $x \in M$  has a unique representation of the form*

$$x = \sum_{i=1}^k x_i$$

*with  $x_i \in M_i$  for each  $i$ ,*

(ii) *for  $1 \leq i < k$ ,  $M_i$  and  $\sum_{j>i} M_j$  are disjoint,*

(iii)  $d(M) = \sum_{i=1}^k d(M_i)$  (“ $d$ ”  $\equiv$  dimension).

**Proof.** (i) *is equivalent to* (ii): Each of (i) and (ii) is easily seen to be equivalent to the following property:

$$\sum_{i=1}^k z_i = 0 \text{ with } z_i \in M_i \text{ for each } i \text{ implies } z_i = 0 \text{ for each } i.$$

(ii) *is equivalent to* (iii): The identity

$$d(K + L) = d(K) + d(L) - d(K \cap L),$$

holding for arbitrary subspaces  $K$  and  $L$  of  $V$ , gives

$$d\left(M_i + \left(\sum_{j>i} M_j\right)\right) = d(M_i) + d\left(\sum_{j>i} M_j\right) - d\left(M_i \cap \left(\sum_{j>i} M_j\right)\right)$$

for  $1 \leq i < k$ . Combining these relations, we find

$$d(M) = d\left(\sum_{j=1}^k M_j\right) = \sum_{j=1}^k d(M_j) - \sum_{i=1}^{k-1} d\left(M_i \cap \left(\sum_{j>i} M_j\right)\right).$$

So  $d(M) = \sum_{i=1}^k d(M_i)$  if and only if  $d(M_i \cap (\sum_{j>i} M_j)) = 0$  for  $1 \leq i < k$ . ■

When the equivalent conditions of the lemma are met, one says that  $M$  is the *direct sum* of the  $M_i$ 's, written  $M = \oplus_{i=1}^k M_i$ . The sum  $M$  of mutually orthogonal subspaces  $M_i$  is necessarily direct and in this situation one says that the  $M_i$ 's form an *orthogonal decomposition* of  $M$ .

We are now ready for the following result, which forms the algebraic basis for Cochran's theorem (see Section 3.9) on the distribution of quadratic forms in normally distributed random variables. There is evidently considerable overlap between the result here and Proposition 2.24. Note though that here it is the *sum* of the operators involved that is assumed at the outset to be an orthogonal projection, whereas in Proposition 2.24 it is the individual *summands* that are postulated to be orthogonal projections.

**2.40 Proposition (Algebraic form of Cochran's theorem).** Let  $T_1, \dots, T_k$  be self-adjoint linear transformations of  $V$  into  $V$  and suppose that

$$P = \sum_{1 \leq i \leq k} T_i$$

is an orthogonal projection. Put  $M_i = \mathcal{R}(T_i)$ ,  $1 \leq i \leq k$ , and  $M = \mathcal{R}(P)$ . The following are equivalent:

- (i) each  $T_i$  is idempotent,
- (ii)  $T_i T_j = 0$  for  $i \neq j$ ,
- (iii)  $\sum_{i=1}^k d(M_i) = d(M)$ ,

and

- (iv)  $M_1, \dots, M_k$  form an orthogonal decomposition of  $M$ .

**Proof.** We treat the case where  $P$  is the identity transformation  $I$  on  $V$ , leaving the general case to the reader as an exercise.

(i) *implies* (iv): Each  $T_i$  is an orthogonal projection by (i) and  $\sum_i T_i$  an orthogonal projection by the umbrella assumption  $\sum_i T_i = I$ . Proposition 2.24 says that  $\sum_i T_i$  is orthogonal projection onto the orthogonal direct sum  $\oplus_i M_i$ . But  $\sum_i T_i = I$  is trivially orthogonal projection onto  $V$ . Thus the  $M_i$ 's form an orthogonal decomposition of  $V$ .

(iv) *implies* (iii): Trivial.

(iii) *implies* (ii): For any  $x \in V$ ,

$$x = Ix = \sum_i T_i x$$

so  $V = \sum_{i=1}^k M_i$ . Taking  $x = T_j y$  and using Lemma 2.39 with  $M = V$ , we get

$$T_i T_j y = 0 \quad \text{for } i = 1, \dots, j-1, j+1, \dots, k$$

(as well as  $T_j y = T_j^2 y$  — but we don't need this now). Since  $j$  and  $y$  are arbitrary,  $T_i T_j = 0$  for  $i \neq j$ .

(ii) *implies* (i): One has, using (ii),

$$T_i - T_i^2 = T_i(I - T_i) = T_i\left(\sum_{j \neq i} T_j\right) = \sum_{j \neq i} T_i T_j = 0$$

for all  $i$ . ■

**2.41 Exercise [3].** Prove Proposition 2.40 for the general orthogonal projection  $P$ .

[Hint: The case of general  $P$  can be reduced to the case  $P = I$  treated in the preceding proof by introducing  $T_0 = Q_M$ . Specifically, let conditions (i'), ..., (iv') be defined like (i), ..., (iv), but with the indices ranging from 0 to  $k$  instead of from 1 to  $k$ , and with  $M$  replaced by  $V$ . The idea is to show that condition (c') is equivalent to condition (c), for  $c = i, \dots, iv$ . In arguing that (ii) implies (ii'), make use of Exercise 2.7.] ◇

**2.42 Exercise** [3]. In the context of  $\mathbb{R}^n$ , suppose  $I = T_1 + T_2$  with  $T_1$  being the transformation having matrix

$$(t_{ij}) = \begin{pmatrix} 1/n & 1/n & \dots & 1/n \\ 1/n & 1/n & \dots & 1/n \\ \vdots & \vdots & & \vdots \\ 1/n & 1/n & \dots & 1/n \end{pmatrix}$$

with respect to the usual coordinate basis, so that

$$[T_1((x_j))]_i = \sum_{1 \leq j \leq n} t_{ij} x_j$$

for  $1 \leq i \leq n$ . In regard to Proposition 2.40, show that  $T_1$  and  $T_2$  are self-adjoint and that some one (and therefore all) of conditions (i)–(iv) is satisfied.  $\diamond$

**2.43 Exercise** [2]. In the context of  $\mathbb{R}^2$ , let  $T_1$  and  $T_2$  be the transformations having matrices

$$[T_1] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad [T_2] = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

with respect to the usual coordinate basis. Show that even though  $T_1$  and  $T_2$  are self-adjoint and sum to the identity, some one (and therefore each) of conditions (i)–(iv) of Proposition 2.40 is not satisfied.  $\diamond$

### 3. Tjur's theorem

Let  $V$  be an inner product space. A finite collection  $\mathcal{L}$  of distinct subspaces of  $V$  such that

- (T1)  $L_1, L_2 \in \mathcal{L}$  implies  $L_1$  and  $L_2$  are book orthogonal,
  - (T2)  $L_1, L_2 \in \mathcal{L}$  implies  $L_1 \cap L_2 \in \mathcal{L}$ , and
  - (T3)  $V \in \mathcal{L}$
- (3.1)

is called a *Tjur system* (after the Danish statistician Tue Tjur). For example, the subspaces  $V$ ,  $M$ , and  $M_0$  of Exercise 2.14 constitute a Tjur system, as do the subspaces  $V$ ,  $M_R$ ,  $M_C$ , and  $M_G$  of Exercise 2.26. Since intersections of mutually book orthogonal spaces are themselves mutually book orthogonal (see Exercise 2.36), any finite collection of mutually book orthogonal subspaces of  $V$  can be augmented to a Tjur system.

The elements of a Tjur system  $\mathcal{L}$  are partially ordered by inclusion. For  $L \in \mathcal{L}$  we write  $K \leq L$  to mean that  $K \in \mathcal{L}$  and  $K$  is a subset of  $L$ , and write  $K < L$  to mean that  $K \leq L$  and  $K \neq L$ . This notation is used in the following key theorem, which shows that the elements of  $\mathcal{L}$  can be represented neatly and simply in terms of an explicit orthogonal decomposition of  $V$ .

**3.2 Theorem (Tjur's theorem).** Let  $\mathcal{L}$  be a Tjur system of subspaces of  $V$ . For each  $L \in \mathcal{L}$ , put

$$L^\circ = L - \sum_{K < L} K. \quad (3.3)$$

Then

- (i) the subspaces  $L^\circ$  for  $L \in \mathcal{L}$  are mutually orthogonal,
- (ii)  $V = \sum_{L \in \mathcal{L}} L^\circ$ ,
- (iii) for each  $L \in \mathcal{L}$ , one has  $L = \sum_{K \leq L} K^\circ$ .

Moreover, the  $L^\circ$ 's are uniquely determined by these conditions.

**Proof.** (i) and (iii) imply (3.3): Suppose that  $\{L^* : L \in \mathcal{L}\}$  is a collection of mutually orthogonal subspaces of  $V$  such that  $L = \sum_{K \leq L} K^*$  for each  $L \in \mathcal{L}$ ; we claim  $L^* = L^\circ$  for each  $L$ . For this note that  $L^* + (\sum_{K < L} K^*)$  is an orthogonal decomposition of  $L$ , so

$$L^* = L - (\sum_{K < L} K^*) = L - (\sum_{K < L} K) = L^\circ.$$

(3.3) implies (i)–(iii): We first use induction to show that

$$L = \sum_{K \leq L} K^\circ \quad (3.4)$$

for each  $L \in \mathcal{L}$ . Condition (T2) implies that  $\mathcal{L}$  has a smallest element, say  $L_0$ ; (3.4) holds for  $L_0$  because  $L_0^\circ = L_0$ . Suppose now

$$K = \sum_{J \leq K} J^\circ \quad \text{for all } K \in \mathcal{L} \text{ with } K < L. \quad (3.5)$$

Then

$$\begin{aligned} L &= L^\circ + \sum_{K < L} K && \text{(by the definition (3.3) of } L^\circ) \\ &= L^\circ + \sum_{K < L} (\sum_{J \leq K} J^\circ) && \text{(by the induction hypothesis (3.5))} \\ &= L^\circ + \sum_{J < L} J^\circ = \sum_{K \leq L} K^\circ, \end{aligned}$$

so (3.4) holds for  $L$ . (iii) now follows from Exercise 3.7 below, and (ii) holds because  $V \in \mathcal{L}$  by (T3).

To complete the proof we show that (i) holds. Let  $L_1, L_2 \in \mathcal{L}$ ; we have to show

$$L_1^\circ \perp L_2^\circ. \quad (3.6)$$

By (T2)  $L = L_1 \cap L_2 \in \mathcal{L}$ . There are three cases to consider: (1)  $L < L_1$  and  $L < L_2$ ; (2)  $L = L_1$ ; and (3)  $L = L_2$ . In case (1),  $L_1^\circ \subset L_1 - L$  and  $L_2^\circ \subset L_2 - L$ , so (3.6) holds because  $L_1$  and  $L_2$  are book orthogonal by (T1). In case (2)  $L_1 < L_2$ , so (3.6) holds because  $L_1^\circ \subset L_1$  and  $L_2^\circ \subset L_2 - L_1$ . Case (3) is like case (2). ■

**3.7 Exercise** (*The principle of induction for a partially ordered set*) [4]. Let  $I$  be a finite set and let " $\leq$ " be a *partial order* on  $I$ :

$$i \leq i \quad \text{for all } i \in I, \quad (3.8_1)$$

$$i \leq j \text{ and } j \leq k \quad \text{implies} \quad i \leq k, \quad (3.8_2)$$

$$i \leq j \text{ and } j \leq i \quad \text{implies} \quad i = j. \quad (3.8_3)$$

For  $i, j \in I$ , write  $i < j$  to mean  $i \leq j$  and  $i \neq j$ . Say that an element  $j$  of  $I$  is *minimal* if there doesn't exist an  $i \in I$  such that  $i < j$ . For each  $i \in I$ , let  $S(i)$  be a statement involving  $i$ . Prove the *principle of induction for  $I$* : if  $S(i)$  is valid for each minimal element  $i$  and if for each non-minimal element  $j$ ,  $S(j)$  is valid whenever  $S(i)$  is valid for all  $i < j$ , then  $S(j)$  is valid for all  $j \in I$ .

[Hint: If  $i < j$  and  $j < k$ , then  $i, j$ , and  $k$  must be distinct elements of  $I$ .]  $\diamond$

**3.9 Exercise** [2]. Show that the subspace  $L^\circ$  defined by (3.3) can be written as  $L \cap (\bigcap_{K < L} (L - K))$ .

[Hint: See Exercise 2.2.]  $\diamond$

**3.10 Exercise** [2]. Show by example that the conclusion (i) of Tjur's theorem would not follow if condition (T2) were dropped from (3.1).  $\diamond$

**3.11 Exercise** [1]. Use Tjur's Theorem to solve Exercise 2.37.  $\diamond$

Tjur's theorem states in part that  $\sum_{L \in \mathcal{L}} L^\circ$  is an orthogonal decomposition of  $V$ . In applications one needs to know how to project onto the subspaces  $L^\circ$ , what dimensions these subspaces have, and how long the components  $P_{L^\circ}v$  of a vector  $v \in V$  are. This information is readily obtained from corresponding information about the original spaces  $L \in \mathcal{L}$ . Indeed, conclusion (iii) in Tjur's theorem implies that

$$P_L = \sum_{K \leq L} P_{K^\circ} \quad \text{for } L \in \mathcal{L}. \quad (3.12)$$

These equations can be solved recursively for the  $P_{L^\circ}$ 's, working upwards from the minimal element  $L_0$  of  $\mathcal{L}$  to the maximal element  $V$ :

$$\begin{aligned} P_{L_0^\circ} &= P_{L_0}, \\ P_{L^\circ} &= P_L - \sum_{K < L} P_{K^\circ} \quad \text{for } L > L_0. \end{aligned} \quad (3.13)$$

Similarly, the dimensions of the  $L^\circ$  spaces can be found by solving the equations

$$d(L) = \sum_{K \leq L} d(K^\circ), \quad L \in \mathcal{L}, \quad (3.14)$$

and for  $v \in V$  the squared lengths

$$\ell_{L^\circ}^2 \equiv \ell_{L^\circ}^2 v = \|P_{L^\circ}v\|^2 \quad (3.15)$$

can be found from the corresponding squared lengths

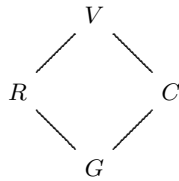
$$\ell_L^2 \equiv \ell_L^2(v) = \|P_L v\|^2 \quad (3.16)$$

by solving the equations

$$\ell_L^2 = \sum_{K \leq L} \ell_{K^\circ}^2, \quad L \in \mathcal{L}. \quad (3.17)$$

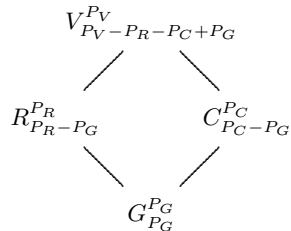
In specific situations the computations are facilitated by referring to a *structure diagram* which displays the ordering of the elements of  $\mathcal{L}$ . The following example illustrates the method.

**3.18 Example.** Consider the Tjur system  $\{V, M_R, M_C, M_G\}$  of Exercise 2.26. For notational simplicity write  $R$  for  $M_R$ ,  $C$  for  $M_C$ , and  $G$  for  $M_G$ . The structure diagram is

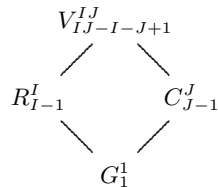


A line between two elements of  $\mathcal{L}$  indicates that the lower subspace is included in the upper one. No line needs to be drawn for the inclusion  $G \subset V$ , since that relation is readily inferred from the diagram as it stands.

Formulas for the projections  $P_L$  were given in Exercise 2.26 (see (2.27)). To calculate the  $P_{L^\circ}$ 's from the  $P_L$ 's using (3.13), begin by superscripting each  $L$  in the structure diagram by  $P_L$ . Then, working upwards from the bottom of the diagram, subscript each  $L$  by  $P_{L^\circ}$ , calculated as the difference between the corresponding superscript and the sum of all subscripts for spaces  $K$  strictly below  $L$  (i.e., for all  $K$  such that  $K < L$ ). For example, the subscript on  $V$  is calculated as  $P_V - (P_{R^\circ} + P_{C^\circ} + P_{G^\circ})$ . The result is



The method of subscripts can also be used to organize the computation of the  $d(L^\circ)$ 's and  $\ell_{L^\circ}^2$ 's. For example, since  $d(V) = IJ$ ,  $d(R) = I$ ,  $d(C) = J$ , and  $d(G) = 1$ , the annotated structure diagram for the calculation of the  $d(L^\circ)$ 's is





The *analysis of variance table* for a Tjur system  $\mathcal{L}$  lists for each  $L \in \mathcal{L}$  the quantities  $L^\circ$ ,  $d(L^\circ)$ , and  $\ell_{L^\circ}^2 = \|P_{L^\circ}v\|^2$  appearing in the decompositions

$$V = \sum_{L \in \mathcal{L}} L^\circ, \quad d(V) = \sum_{L \in \mathcal{L}} d(L^\circ), \quad \|v\|^2 = \sum_{L \in \mathcal{L}} \ell_{L^\circ}^2(v).$$

Following conventional statistical practice, rows for smaller  $L$ 's (used in building simpler models) appear above rows for larger  $L$ 's (used in building more complicated models). The analysis of variance table for the Tjur system in the preceding Example is given in Table 3.19.

**3.19 Table.** *Analysis of variance table for the Tjur system  $\{V, R, C, G\}$  of Example 3.18. The quantities  $\ell_L^2 = \|P_L v\|^2$  appearing in the “squared length” column are given by*

$$\ell_V^2 = \sum_{ij} v_{ij}^2, \quad \ell_R^2 = J \sum_i \bar{v}_i^2, \quad \ell_C^2 = I \sum_j \bar{v}_j^2, \quad \ell_G^2 = IJ \bar{v}^2.$$

| Label | Subspace                    | Dimension        | Squared length                                                                                             |
|-------|-----------------------------|------------------|------------------------------------------------------------------------------------------------------------|
| $G$   | $G^\circ = G$               | 1                | $\ell_G^2 = IJ \bar{v}^2$                                                                                  |
| $R$   | $R^\circ = R - G$           | $I - 1$          | $\ell_R^2 - \ell_G^2 = J \sum_i (\bar{v}_i - \bar{v}.)^2$                                                  |
| $C$   | $C^\circ = C - G$           | $J - 1$          | $\ell_C^2 - \ell_G^2 = I \sum_j (\bar{v}_j - \bar{v}.)^2$                                                  |
| $V$   | $V^\circ = V - (R + C + G)$ | $IJ - I - J + 1$ | $\ell_V^2 - \ell_R^2 - \ell_C^2 + \ell_G^2$<br>$= \sum_{ij} (v_{ij} - \bar{v}_i - \bar{v}_j + \bar{v}.)^2$ |
| Sum   | $V$                         | $IJ$             | $\ell_V^2 = \sum_{ij} v_{ij}^2$                                                                            |

**3.20 Exercise** [2]. Verify the entries in the “squared lengths” column of Table 3.19. To get the expressions to the left of the “=” signs, use the method of  $\text{su}_b^{\text{per}}$  scripts. To get the expressions to the right of the “=” signs, use the formulae for the  $P_{L^\circ}$ 's derived in Example 3.18.  $\diamond$

**3.21 Exercise** [2]. Draw the structure diagram for the Tjur system  $\{V, M, M_0\}$  of Exercise 2.14, and use the method of  $\text{su}_b^{\text{per}}$  scripts to construct the corresponding analysis of variance table.  $\diamond$

**3.22 Exercise** (*The Möbius inversion formula*) [4]. As in Exercise 3.7, let “ $\leq$ ” be a partial order on a finite set  $I$ . (1) The *Möbius function* of  $(I, \leq)$  is the matrix  $\mu = (\mu_{ik})_{i \in I, k \in I}$  such that  $\mu_{ik} = 0$  for  $i \not\leq k$  and such that

$$\sum_{j: i \leq j \leq k} \mu_{ij} = \delta_{ik} \quad \text{for } i \leq k \quad (3.23)$$

or, equivalently,

$$\sum_{j: i \leq j \leq k} \mu_{jk} = \delta_{ik} \quad \text{for } i \leq k. \quad (3.24)$$

Prove that such a matrix  $\mu$  exists, and is inverse to the matrix  $\zeta = (\zeta_{ik})_{i \in I, k \in I}$  with

$$\zeta_{ik} = \begin{cases} 1, & \text{if } i \leq k, \\ 0, & \text{otherwise;} \end{cases}$$

$\zeta$  is called the *zeta function* of  $(I, \leq)$ . (2) Prove of functions  $f$  and  $g$  mapping  $I$  into some vector space  $W$  that

$$g(j) = \sum_{i \leq j} f(i) \quad \text{for all } j \in I \quad (3.25)$$

if and only if

$$f(j) = \sum_{i \leq j} g(i) \mu_{ij} \quad \text{for } j \in I; \quad (3.26)$$

this result is called the *Möbius inversion formula*.

[Hint: For each fixed  $i \in I$ , equation (3.23) can be solved recursively for  $\mu_{ik}$  with  $i \leq k$ .]  $\diamond$

**3.27 Exercise** [3]. Recall that the elements of a Tjur system  $\mathcal{L}$  of subspaces of  $V$  are partially ordered by inclusion (" $\leq$ "). Let  $\mu$  be the Möbius function for  $(\mathcal{L}, \leq)$ . Use the Möbius inversion formula from the preceding exercise to show that

$$P_{L^\circ} = \sum P_K \mu_{KL}, \quad d(L^\circ) = \sum d(K) \mu_{KL}, \quad \ell_{L^\circ}^2 = \sum \ell_K^2 \mu_{KL} \quad (3.28)$$

for all  $L \in \mathcal{L}$ , the summations in each case extending over  $K \leq L$ , or even over all  $K \in \mathcal{L}$ . Check (3.28) against the entries in Table 3.19.  $\diamond$

As Example 3.18 suggests, Tjur systems arise naturally in the theory of the design of experiments. The rest of this section elaborates on this idea. In particular, we will see how the notions of book orthogonality, intersection, and inclusion of subspaces correspond to certain simple relationships among the levels of the treatments assigned to the experimental units.

The design of an experiment often calls for the available experimental units (e.g., plots of land) to be divided into groups, with each unit in a group receiving the same level of some experimental treatment (e.g., type of seed). To abstract this idea, let  $X$  be a finite set whose elements are called *experimental units*. A (treatment) *factor*  $F$  is a partition of  $X$  into nonempty disjoint subsets called *blocks*, or *levels of  $F$* . The factor  $U$  whose blocks are single units is called the *units factor*, while the factor  $T$  having only one block is called the *trivial factor*.

For a subset  $f$  of  $X$  put

$$|f| = \text{cardinality}(f), \quad (3.29)$$

the number of experimental units in  $f$ . A factor  $F$  is said to be *balanced* if each of its levels is assigned the same number of units, i.e., if

$$|f| = |g| \quad \text{for all blocks } f \text{ and } g \text{ of } F. \quad (3.30)$$

For example, if  $X = \{(i, j) : 1 \leq i \leq I, 1 \leq j \leq J\}$ , then the row factor  $R$  with blocks  $r_i = \{(i, j) : 1 \leq j \leq J\}$  for  $1 \leq i \leq I$  is balanced because  $|r_i| = J$  for each  $i$ .

If  $F$  is a factor and  $x$  and  $y$  are experimental units, the notation

$$x \sim_F y \quad (3.31)$$

means that  $x$  and  $y$  are assigned the same level of  $F$ , i.e., belong to the same block of  $F$ . The relation " $\sim_F$ " is an equivalence relation:

$$x \sim_F x \quad \text{for each } x \in X, \quad (3.32_1)$$

$$x \sim_F y \quad \text{implies} \quad y \sim_F x, \quad (3.32_2)$$

$$x \sim_F y \quad \text{and} \quad y \sim_F z \quad \text{implies} \quad x \sim_F z. \quad (3.32_3)$$

Conversely, if " $\sim$ " is an arbitrary equivalence relation on  $X$ , then its equivalence classes constitute a factor. The correspondence between factors and equivalence relations is one-to-one.

Let  $F$  and  $G$  be factors. One says  $F$  is *nested in*  $G$  if each block of  $G$  is a union of blocks of  $F$ , or, equivalently, if

$$\text{for } x, y \in X, \quad x \sim_F y \quad \text{implies} \quad x \sim_G y; \quad (3.33)$$

this situation is written as

$$G \leq F \quad (3.34)$$

because it implies that  $F$  has at least as many levels as  $G$ . The units factor  $U$  is nested in each factor, and each factor is nested in the trivial factor  $T$ . The relation " $\leq$ " is a partial order on the set of factors of  $X$ :

$$F \leq F \quad \text{for all factors } F, \quad (3.35_1)$$

$$H \leq G \quad \text{and} \quad G \leq F \quad \text{implies} \quad H \leq F, \quad (3.35_2)$$

$$G \leq F \quad \text{and} \quad F \leq G \quad \text{implies} \quad F = G. \quad (3.35_3)$$

Interesting factors can be built up from given ones by means of their equivalence relations. For example, suppose  $F$  and  $G$  are factors. Write

$$x \sim_{\times} y \quad (3.36)$$

for  $x, y \in X$  to mean  $x \sim_F y$  and  $x \sim_G y$ . " $\sim_{\times}$ " is an equivalence relation; the factor  $F \times G$  whose blocks are the  $\sim_{\times}$ -equivalence classes is called the *cross-classification induced by  $F$  and  $G$* . It turns out that  $F \times G$  is the least upper bound of  $F$  and  $G$  with respect to the "nested in" ordering, so  $F \times G$  is also called the *maximum of  $F$  and  $G$*  and written as  $F \vee G$ .

**3.37 Exercise** [2]. Let  $F$  and  $G$  be factors. Show that  $F \times G$  is indeed the *least upper bound* of  $F$  and  $G$  with respect to the ordering (3.34):

$$F \leq F \times G \quad \text{and} \quad G \leq F \times G, \quad (3.38_1)$$

$$F \leq H \quad \text{and} \quad G \leq H \quad \text{implies} \quad F \times G \leq H. \quad (3.38_2) \diamond$$

Let again  $F$  and  $G$  be factors. Write

$$x \sim_{\wedge} y \quad (3.39)$$

for  $x, y \in X$  to mean that there exists a finite sequence  $z_0, \dots, z_k$  of experimental units such that

$$z_0 = x \quad \text{and} \quad z_k = y \quad (3.40_1)$$

and

$$z_{j-1} \sim_F z_j \quad \text{and/or} \quad z_{j-1} \sim_G z_j \quad \text{for } 1 \leq j \leq k. \quad (3.40_2)$$

Loosely speaking,  $x \sim_{\wedge} y$  if it is possible to move from  $x$  to  $y$  through  $X$  in a finite number of steps, each of which takes place within a block of  $F$  (an  $F$ -step) or a block of  $G$  (a  $G$ -step); an efficient move would alternate  $F$ -steps with  $G$ -steps, but this is not required. “ $\sim_{\wedge}$ ” is an equivalence relation; the factor whose blocks are the  $\sim_{\wedge}$ -equivalence classes turns out to be the greatest lower bound of  $F$  and  $G$ , and so is called the *minimum of  $F$  and  $G$*  and written as  $F \wedge G$ .

**3.41 Exercise** [3]. Let  $F$  and  $G$  be factors. Verify that the relation “ $\sim_{\wedge}$ ” is indeed an equivalence relation and that the factor  $F \wedge G$  is the *greatest lower bound* of  $F$  and  $G$ :

$$F \wedge G \leq F \quad \text{and} \quad F \wedge G \leq G, \quad (3.42_1)$$

$$H \leq F \quad \text{and} \quad H \leq G \quad \text{implies} \quad H \leq F \wedge G. \quad (3.42_2) \diamond$$

**3.43 Example.** Suppose  $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$  and the experimental units are allocated to the 3 levels  $r_1, r_2$ , and  $r_3$  of a row factor  $R$  and the 3 levels  $c_1, c_2$ , and  $c_3$  of a column factor  $C$  as indicated below:

| Rows  | Columns |       |       |
|-------|---------|-------|-------|
|       | $c_1$   | $c_2$ | $c_3$ |
| $r_1$ | 1,2     | 3     |       |
| $r_2$ | 4,5     | 6     | 9     |
| $r_3$ |         |       | 7,8   |

For example, units 1, 2, and 3 are assigned to level  $r_1$  of  $R$ , and units 3 and 6 are assigned to level  $c_2$  of  $C$ . The blocks of  $R \times C$  are  $\{1, 2\}$ ,  $\{3\}$ ,  $\{4, 5\}$ ,  $\{6\}$ ,  $\{7, 8\}$ , and  $\{9\}$ .  $R \wedge C$  is the trivial factor because each unit can be reached from each other unit by a succession of row and column steps. If the last experimental unit 9 were to be removed from  $X$ , then  $R \wedge C$  would have 2 blocks,  $\{1, 2, 3, 4, 5, 6\}$  and  $\{7, 8\}$ . •

**3.44 Exercise** [4]. Let  $F_1, \dots, F_k$  be factors such that  $f_1 \cap \dots \cap f_k \neq \emptyset$  for all  $f_1 \in F_1, \dots, f_k \in F_k$ . For subsets  $J$  of  $K \equiv \{1, \dots, k\}$ , let

$$F_J = \prod_{j \in J} F_j \quad (3.45)$$

denote the cross-classification induced by factors  $F_j$  with  $j \in J$  —  $x \sim_{F_j} y$  if and only if  $x \sim_{F_j} y$  for all  $j \in J$ . (Take  $F_\emptyset$  to be trivial factor  $T$ .) Show that

$$F_{J_1} \wedge F_{J_2} = F_{J_1 \cap J_2} \quad (3.46)$$

for all subsets  $J_1$  and  $J_2$  of  $K$ .

[Hint First do the case where  $k = 3$ ,  $J_1 = \{1, 2\}$ , and  $J_2 = \{2, 3\}$ .]  $\diamond$

Let now  $V$  be the vector space of real-valued functions on  $X$ . For each subset  $f$  of  $X$ , let  $I_f \in V$  be the *indicator function of  $f$* :

$$I_f(x) = \begin{cases} 1, & \text{if } x \in f, \\ 0, & \text{if } x \notin f. \end{cases} \quad (3.47)$$

For each factor  $F$ , let

$$L_F = [I_f : f \in F] = \{v \in V : x \sim_F y \text{ implies } v(x) = v(y)\} \quad (3.48)$$

be the subspace of  $V$  consisting of functions that are constant on each block of  $F$ . Note that  $L_F$  has dimension

$$d(L_F) = |F|, \text{ the number of blocks of } F. \quad (3.49)$$

**3.50 Proposition.** *Let  $F$  and  $G$  be factors. Then*

$$G \leq F \text{ if and only if } L_G \subset L_F \quad (3.51)$$

and

$$L_{F \wedge G} = L_F \cap L_G. \quad (3.52)$$

**Proof.**  $G \leq F$  implies  $L_G \subset L_F$ :  $G \leq F$  means that each block of  $G$  is a union of blocks of  $F$ , so constancy on blocks of  $G$  trivially implies constancy on blocks of  $F$ .

$L_G \subset L_F$  implies  $G \leq F$ : Suppose  $L_G \subset L_F$ . Let  $g$  be a block of  $G$  and  $f$  a block of  $F$  such that  $f \cap g \neq \emptyset$ ; we have to show that  $f \subset g$ . For this, note that the function  $I_g \in L_G \subset L_F$  has the value 1 at some point of  $f$  and so is identically 1 on  $f$ .

$L_{F \wedge G} \subset L_F \cap L_G$ : This follows from (3.51) applied to  $F \wedge G \leq F$  and  $F \wedge G \leq G$ .

$L_F \cap L_G \subset L_{F \wedge G}$ : Suppose  $v \in V$  is constant over blocks of both  $F$  and  $G$ ; we have to show  $v$  is constant over blocks of  $F \wedge G$ . For this suppose  $x, y \in X$  with  $x \sim_{F \wedge G} y$ . Choose  $z_0, \dots, z_k$  satisfying (3.40). Then  $v(z_{j-1}) = v(z_j)$  for each  $j$ , so  $v(x) = v(z_0) = v(z_k) = v(y)$ .  $\blacksquare$

Let now  $V$  be endowed with the dot-product

$$\langle u, v \rangle = \sum_{x \in X} u(x)v(x). \quad (3.53)$$

Let  $P_F$  denote orthogonal projection onto  $L_F$ . Since the functions  $I_f$  for  $f \in F$  are an orthogonal basis for  $L_F$ , we have

$$P_F v = \sum_{f \in F} \bar{v}_f I_f \quad (3.54)$$

where

$$\bar{v}_f = \frac{\langle I_f, v \rangle}{\langle I_f, I_f \rangle} = \frac{1}{|f|} \sum_{x \in F} v(x); \quad (3.55)$$

in other words, the value of  $P_F v$  at an experimental unit  $y$  is the average value of  $v \in V$  over all units  $x$  that are in the same block of  $F$  as  $y$ . In situations where the symbol  $v$  is adorned with subscripts and/or superscripts, we will sometimes write  $A_f(v)$  in place of  $\bar{v}_f$ . Note that

$$\ell_{L_F}^2(v) \equiv \|P_F v\|^2 = \sum_{x \in X} ((P_F v)(x))^2 = \sum_{f \in F} |f| \bar{v}_f^2 \equiv SS_F, \quad (3.56)$$

the symbol “ $SS$ ” denoting *sum of squares*.

Factors  $F$  and  $G$  are said to be *orthogonal* if  $L_F$  and  $L_G$  are book orthogonal. By Proposition 2.29, this is equivalent to  $P_F P_G = P_G P_F$ , and also to  $P_F P_G = P_{L_F \cap L_G}$ . In view of (3.52), this last condition is the same as  $P_F P_G = P_{F \wedge G}$ . The following proposition gives a useful characterization of orthogonality in terms of the sizes of the blocks of  $F$ ,  $G$ , and  $F \wedge G$ . If  $f$  and  $h$  are subsets of  $X$ , say that  $f$  is *nested in*  $h$  if  $f$  is a subset of  $h$ .

**3.57 Proposition.** *Factors  $F$  and  $G$  are orthogonal if and only if*

$$\frac{|f \cap g|}{|h|} = \frac{|f|}{|h|} \times \frac{|g|}{|h|} \quad (3.58)$$

for all blocks  $f \in F$ ,  $g \in G$ , and  $h \in H \equiv F \wedge G$  such that  $f$  and  $g$  are nested in  $h$ .

**Proof.** The functions  $\delta_x = I_{\{x\}}$ ,  $x \in X$ , form a basis for  $V$ , so  $F$  and  $G$  are orthogonal if and only if  $\langle \delta_x, P_H \delta_y \rangle = \langle \delta_x, P_F P_G \delta_y \rangle$ , and hence if and only if

$$\langle \delta_x, P_H \delta_y \rangle \equiv \langle P_F \delta_x, P_G \delta_y \rangle \quad (3.59)$$

for all  $x, y \in X$ . The left- and right-hand sides of (3.59) are 0 unless  $x$  and  $y$  both belong to the same block  $h$  of  $H$ , in which case the left-hand side is

$$\langle \delta_x, \sum_{\bar{h} \in H} A_{\bar{h}}(\delta_y) I_{\bar{h}} \rangle = A_h(\delta_y) = 1/|h|$$

and the right-hand side is

$$\langle \sum_{\bar{f} \in F} A_{\bar{f}}(\delta_x) I_{\bar{f}}, \sum_{\bar{g} \in G} A_{\bar{g}}(\delta_y) I_{\bar{g}} \rangle = A_f(\delta_x) A_g(\delta_y) \langle I_f, I_g \rangle = \frac{|f \cap g|}{|f| |g|},$$

where  $f$  is the block of  $F$  containing  $x$  and  $g$  is the block of  $G$  containing  $y$ ; necessarily  $f \subset h$  and  $g \subset h$ . ■

**3.60 Exercise [2].** Show of factors  $F$  and  $G$  that  $G \leq F$  implies that  $F$  and  $G$  are orthogonal. Solve this exercise twice, once using (3.51) and once using (3.58). ◇

The following exercise gives a convenient reformulation of the orthogonality criterion (3.58).

**3.61 Exercise** [3]. Let  $F$  and  $G$  be factors. Show that  $F$  and  $G$  are orthogonal if and only if

$$\begin{aligned} &\text{for each level } h \text{ of } H \equiv F \wedge G \text{ and levels } g_1 \text{ and } g_2 \text{ of } G \text{ nested in } h, \\ &\text{there exists a constant } c = c(h, g_1, g_2) \text{ such that } |f \cap g_1| = c|f \cap g_2| \quad (3.62) \\ &\text{for all levels } f \text{ of } F \text{ nested in } h, \end{aligned}$$

this being the so-called *condition of proportional cell counts*. ◇

**3.63 Exercise** [2]. Each of the four tables below gives the number of experimental units to be allocated to the cells of a  $3 \times 3$  rectangular array. In which cases is the implied row factor  $R$  orthogonal to the implied column factor  $C$ ?

| <i>Table 1</i> | <i>Table 2</i> | <i>Table 3</i> | <i>Table 4</i> |   |
|----------------|----------------|----------------|----------------|---|
| 1 1 1          | 3 3 3          | 2 1 0          | 2 1 0          |   |
| 1 1 1          | 2 2 2          | 2 1 1          | 2 1 0          |   |
| 1 1 1          | 1 1 1          | 0 0 2          | 0 0 2          | ◇ |

**3.64 Exercise** [2]. Consider a  $2 \times 2 \times 2$  design having a row factor  $R$  with levels  $r_1$  and  $r_2$ , a column factor  $C$  with levels  $c_1$  and  $c_2$ , and a height factor  $H$  with levels  $h_1$  and  $h_2$ . Suppose the 10 experimental units are allocated in such a way that

$$|r_i \cap c_j \cap h_k| = \begin{cases} 2, & \text{if } (i, j, k) = (1, 1, 2) \text{ or } (2, 2, 2), \\ 1, & \text{otherwise.} \end{cases}$$

Show that  $R \perp H$  and  $C \perp H$ , but  $(R \times C) \not\perp H$ . Is  $(R \wedge C) \perp H$ ? ◇

A *Tjur design*  $\mathcal{D}$  is a collection of distinct factors such that

- (D1)  $F, G \in \mathcal{D}$  implies  $F$  and  $G$  are orthogonal,
  - (D2)  $F, G \in \mathcal{D}$  implies  $F \wedge G \in \mathcal{D}$ , and
  - (D3) the units factor  $U$  is in  $\mathcal{D}$ .
- (3.65)

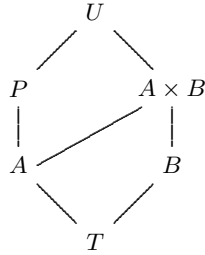
The subspaces  $L_F$  for  $F$  in a Tjur design  $\mathcal{D}$  constitute a Tjur system  $\mathcal{L}$ ; indeed, (D1) trivially implies (T1), (D2) implies (T2) by (3.52), and (D3) implies (T3) since  $L_U = V$ . Moreover by (3.51) the ordering of the factors  $F \in \mathcal{D}$  corresponds to the ordering of the  $L_F$ 's. The quantities  $P_{L_F} \equiv P_F$ ,  $d(L_F)$ , and  $SS_F = \|P_F v\|^2$  are easily obtained from (3.54), (3.49), and (3.56) respectively, and the analysis of variance table can be written down more or less at sight using a factor structure diagram and the method of  $\text{su}_b^{\text{per}}$  scripts.

**3.66 Exercise** (*A split-plot design*) [4]. Suppose that 5 treatments  $a_1, \dots, a_5$  are applied to 15 plots, each treatment being applied to 3 plots. Each plot is subdivided into 2 subplots to which two further treatments  $b_1, b_2$  are applied, one treatment to each subplot. The relevant factors are: the units factor  $U$  (for subplots) with 30 levels; a "plot" factor  $P$  with 15 levels, an " $a$ -treatment" factor  $A$  with 5 levels, a " $b$ -treatment" factor  $B$  with 2 levels, a "combined treatment" factor  $A \times B$  with 10 levels, and the trivial factor  $T$  with 1 level.

Using the following template of  $\bullet$ 's to represent the 30 subplots, show one way of assigning the experimental units to these factors:



Verify that  $\{U, P, A, B, A \times B, T\}$  is a Tjur design with factor structure diagram



and use the method of  $\text{su}_b^{\text{per}}$  scripts to compute the corresponding analysis of variance table.  $\diamond$

#### 4. Self-adjoint transformations and the spectral theorem

Recall that a linear transformation  $T: V \rightarrow V$  such that

$$\langle Tx, y \rangle = \langle x, Ty \rangle \quad \text{for all } x, y \in V \quad (4.1)$$

is said to be *self-adjoint*. In this section we are going to study the geometry of self-adjoint transformations.

A few words are in order about one reason why self-adjoint transformations are of special interest in the study of the GLM. Consider the canonical example:  $(V, \langle \cdot, \cdot \rangle) = (\mathbb{R}^n, \text{dot product})$ . Suppose  $\mathbf{Y}^{n \times 1} = (Y_i)_{1 \leq i \leq n}$  is a random vector in  $\mathbb{R}^n$  with dispersion matrix  $\Sigma^{n \times n} = (\sigma_{ij})_{1 \leq i \leq n, 1 \leq j \leq n}$ :

$$\sigma_{ij} = \text{Cov}(Y_i, Y_j).$$

The covariance between any two linear combinations  $\sum_i c_i Y_i = \mathbf{c}^T \mathbf{Y}$  and  $\sum_j d_j Y_j = \mathbf{d}^T \mathbf{Y}$  of the  $Y$ 's is given by

$$\text{Cov}(\mathbf{c}^T \mathbf{Y}, \mathbf{d}^T \mathbf{Y}) = \sum_{ij} c_i \sigma_{ij} d_j = \mathbf{c}^T \Sigma \mathbf{d}.$$

The probabilistic identity

$$\text{Cov}(\mathbf{c}^T \mathbf{Y}, \mathbf{d}^T \mathbf{Y}) = \text{Cov}(\mathbf{d}^T \mathbf{Y}, \mathbf{c}^T \mathbf{Y})$$

implies the algebraic identity

$$\langle \mathbf{c}, \Sigma \mathbf{d} \rangle = \mathbf{c}^T \Sigma \mathbf{d} = \mathbf{d}^T \Sigma \mathbf{c} = \langle \Sigma \mathbf{c}, \mathbf{d} \rangle.$$

This says that the linear transformation

$$T_\Sigma: \mathbf{x} \rightarrow \Sigma \mathbf{x}$$



is self-adjoint. Notice also that

$$\langle T_{\Sigma} \mathbf{x}, \mathbf{x} \rangle = \langle \Sigma \mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^T \Sigma \mathbf{x} = \text{Var}(\mathbf{x}^T \mathbf{Y}).$$

Thus variance/covariance structures are intimately related to self-adjoint transformations.

**4.2 Exercise [2].** Let  $T: V \rightarrow V$  be a linear transformation and let  $\mathbf{A} = (a_{ik})$  be the matrix of  $T$  with respect to a given orthonormal basis  $b_1, \dots, b_n$  for  $V$ , so that

$$Tb_k = \sum_i a_{ik} b_i \quad \text{for } 1 \leq k \leq n.$$

Show that

$$a_{jk} = \langle b_j, Tb_k \rangle$$

for all  $j, k$  and conclude that  $T$  is self-adjoint if and only if  $\mathbf{A}$  is symmetric.  $\diamond$

We have seen that any orthogonal projection is self-adjoint. And, of course, the action of an orthogonal projection is geometrically obvious:  $P_M$  leaves each  $x \in M$  fixed, kills each  $y \in M^\perp$ , and maps the general  $V$  vector  $z = x + y$  with  $x \in M$  and  $y \in M^\perp$  into  $P_M z = P_M x + P_M y$ . More generally if

$$V = M_1 + \dots + M_k$$

is a decomposition of  $V$  into mutually orthogonal subspaces and if  $\lambda_1, \dots, \lambda_k$  are any scalars, then

$$T = \sum_{1 \leq i \leq k} \lambda_i P_{M_i} \tag{4.3}$$

is self-adjoint, being a linear combination of self-adjoint transformations. Again the action of such a  $T$  is geometrically clear: since

$$T\left(\sum_{1 \leq i \leq k} x_i\right) = \sum_{1 \leq i \leq k} \lambda_i x_i \quad \text{for } x_i \in M_i, \quad 1 \leq i \leq k,$$

$T$  just dilates (or contracts) by a factor of  $\lambda_i$  within  $M_i$ ,  $1 \leq i \leq k$ . The remarkable thing is that (4.3) describes the most general self-adjoint transformation on  $V$ :

**4.4 Theorem (Spectral theorem).** *Suppose  $T: V \rightarrow V$  is linear and self-adjoint. Then there exists an orthogonal decomposition  $V = \bigoplus_{i=1}^k M_i$  of  $V$  into non-trivial subspaces  $M_i$  and distinct scalars  $\lambda_i$  such that*

$$T = \sum_{1 \leq i \leq k} \lambda_i P_{M_i}. \tag{4.5}$$

Suppose  $T$  is self-adjoint and therefore of the form (4.5). One can characterize the  $\lambda_i$ 's and  $M_i$ 's in terms of the eigenvalues and eigenvectors of  $T$ . Recall that for any linear transformation  $S: V \rightarrow V$ ,  $\lambda$  is called an *eigenvalue* of  $S$  if there is a non-zero *eigenvector*  $x \in V$  such that

$$Sx = \lambda x; \tag{4.6}$$

the subspace of all  $x \in V$  such that (4.6) holds is called the *eigenmanifold* of  $\lambda$ , denoted  $E_\lambda$ . From what was said earlier about the geometric action of  $T$  it is clear that the  $\lambda_i$ 's are precisely the distinct eigenvalues of  $T$  and the  $M_i$ 's are the corresponding eigenspaces. Also for any  $x \in V$ ,

$$Q_T(x) \equiv \langle Tx, x \rangle = \left\langle \sum_i \lambda_i x_i, \sum_j x_j \right\rangle = \sum_i \lambda_i \|x_i\|^2, \quad (4.7)$$

where  $x_i = P_{M_i}x$  for  $1 \leq i \leq k$ . (The  $Q$  in  $Q_T$  stands for "quadratic form".) For  $x$  of unit length, i.e., for

$$1 = \|x\|^2 = \sum_i \|x_i\|^2,$$

$Q_T(x)$  is thus a weighted average of the  $\lambda_i$ 's. This observation leads to the following extremal characterization of the  $\lambda_i$ 's. Assuming the  $\lambda_i$ 's to be indexed in decreasing order, so that  $\lambda_1 > \lambda_2 > \cdots > \lambda_k$ , one has

$$\begin{aligned} \lambda_1 &= \sup\{Q_T(x) : \|x\| = 1\} \\ \lambda_2 &= \sup\{Q_T(x) : \|x\| = 1, x \perp E_{\lambda_1}\} \\ &\vdots \\ \lambda_k &= \sup\{Q_T(x) : \|x\| = 1, x \perp (E_{\lambda_1} + \cdots + E_{\lambda_{k-1}})\}. \end{aligned} \quad (4.8)$$

**Proof of the Spectral theorem.** Let  $T$  be self-adjoint and linear. We are going to produce the representation (4.5) using (4.8) to define the  $\lambda_i$ 's. To begin with, consider

$$\lambda_1 \equiv \sup\{Q_T(x) : x \in B\}$$

where

$$B = \{x \in V : \|x\| = 1\}.$$

Since with respect to the distance function

$$d(x, y) = \|y - x\|$$

the closed bounded set  $B$  is compact (see Exercise 1.23) and the mapping

$$x \rightarrow Q_T(x) = \langle Tx, x \rangle$$

is continuous, and since continuous functions on compact sets achieve their suprema, there exists an  $x \in B$  such that

$$Q_T(x) = \lambda_1.$$

We claim that such an  $x$  is an eigenvector of  $T$  with eigenvalue  $\lambda_1$ , i.e.,

$$Tx = \lambda_1 x.$$

Notice that

$$\lambda_1 x = \langle Tx, x \rangle x = \frac{\langle Tx, x \rangle}{\langle x, x \rangle} x = P_{[x]}(Tx).$$

So if

$$Tx \neq \lambda_1 x$$

then the residual vector

$$r = Tx - P_{[x]}(Tx)$$

is non-zero, and the idea is to produce a contradiction by perturbing  $x$  in the direction  $r$  so as to produce a vector  $y \in B$  such that  $Q_T(y) > Q_T(x)$ . Specifically for real  $\delta$  set

$$y_\delta = \frac{x + \delta r}{\|x + \delta r\|} \in B$$

and note that

$$\begin{aligned} Q_T(y_\delta) &= \langle Ty_\delta, y_\delta \rangle \\ &= \frac{\langle Tx + \delta Tr, x + \delta r \rangle}{\|x + \delta r\|^2} \\ &= \frac{\langle Tx, x \rangle + \delta(\langle Tr, x \rangle + \langle Tx, r \rangle) + \delta^2 \langle Tr, r \rangle}{\|x + \delta r\|^2} \\ &= \frac{\langle Tx, x \rangle + 2\delta \langle Tx, r \rangle + \delta^2 \langle Tr, r \rangle}{\|x + \delta r\|^2} && (T \text{ is self-adjoint}) \\ &= \frac{\langle Tx, x \rangle + 2\delta \|r\|^2 + \delta^2 \langle Tr, r \rangle}{\|x\|^2 + \delta^2 \|r\|^2} && (r = Q_{[x]}Tx \perp x). \end{aligned}$$

Since  $\|r\| \neq 0$  by supposition,  $Q_T(y_\delta) > \langle Tx, x \rangle / \|x\|^2 = Q_T(x)$  for sufficiently small positive  $\delta$ .

This shows that  $\lambda_1$  is an eigenvalue of  $T$ . Let  $M_1$  be the associated eigenspace. If  $M_1 = V$ , then  $T = \lambda_1 P_{M_1}$  and we are done. Otherwise note that for  $x \in M_1$  and  $y \in M_1^\perp$ ,

$$\langle x, Ty \rangle = \langle Tx, y \rangle = \lambda_1 \langle x, y \rangle = 0,$$

so  $T$  maps  $M_1^\perp$  into itself. Applying the preceding argument to the restriction  $T_1$  of  $T$  to  $M_1^\perp$  we find that

$$\lambda_2 = \sup\{Q_{T_1}(x) \equiv Q_T(x) : x \in B \text{ and } x \perp M_1\}$$

is an eigenvalue of  $T_1$  (and therefore of  $T$ ) with some non-trivial eigenspace  $M_2 \subset M_1^\perp$ . If  $M_2 = M_1^\perp$  we have  $T = \lambda_1 P_{M_1} + \lambda_2 P_{M_2}$  and are done. Otherwise we recursively define  $\lambda_i$  and  $M_i$  for  $i = 3, \dots$  by

$$\begin{aligned} \lambda_i &= \sup\{Q_T(x) : x \in B \text{ and } x \perp (M_1 + \dots + M_{i-1})\} \\ M_i &= \text{eigenspace of } \lambda_i \text{ for } T \text{ restricted to } (M_1 + \dots + M_{i-1})^\perp, \end{aligned}$$

stopping the process with the first index  $k$  such that  $\sum_{1 \leq i \leq k} M_i$  is  $V$ ; such a  $k$  exists because each new eigenspace adds at least one dimension to the sum. We then have  $T = \sum_{1 \leq i \leq k} \lambda_i P_{M_i}$ . ■

**4.9 Exercise** [3]. Find the spectral representation (4.5) of the transformation  $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$  corresponding to the matrix

$$\begin{pmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ b & b & a & \cdots & b \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ b & b & b & \cdots & a \end{pmatrix}. \quad \diamond$$

**4.10 Exercise** [2]. Use the Spectral Theorem to solve Exercise 2.7.  $\diamond$

**4.11 Exercise** [3]. Use the Spectral Theorem to show that if  $\mathbf{A}^{n \times n}$  is a symmetric matrix, then there exists an orthogonal matrix  $\mathcal{O}^{n \times n}$  such that  $\mathbf{D} \equiv \mathcal{O}^T \mathbf{A} \mathcal{O}$  is diagonal.  $\diamond$

**4.12 Exercise** [3]. Suppose  $T: V \rightarrow V$  is linear self-adjoint and positive semi-definite ( $\langle v, Tv \rangle \geq 0$  for all  $v \in V$ ). Show that there exists a unique linear self-adjoint positive semi-definite transformation  $S: V \rightarrow V$  such that  $T = S^2$ . ( $S$  is called the (positive) *square root* of  $T$ ; one writes  $S = \sqrt{T}$ .)  $\diamond$

## 5. Representation of linear and bilinear functionals

For fixed  $v \in V$ , the mapping

$$x \rightarrow \langle x, v \rangle \quad (x \in V)$$

is a *linear functional* on  $V$ , i.e., a linear transformation from  $V$  to  $\mathbb{R}$ . The following result says every linear functional on  $V$  is of this form:

**5.1 Proposition (Representation theorem for linear functionals).**  
If  $\psi$  is a linear functional on  $V$ , then there exists a unique  $v \in V$  such that

$$\psi(x) = \langle x, v \rangle \quad \text{for all } x \in V. \quad (5.2)$$

The  $v$  of (5.2) is called the *coefficient vector* of  $\psi$ ; one writes  $v = c.v.(\psi)$ , or just  $v = cv(\psi)$ .

**Proof of the Representation theorem.** *Uniqueness:* If  $v$  and  $w$  are both coefficient vectors for  $\psi$ , then their difference is orthogonal to every vector in  $V$  and so is 0.

*Existence:* If the desired representation holds, then the null space of  $\psi$  is the orthogonal complement of  $[v]$  in  $V$ . This observation suggests how to proceed. To avoid trivialities, suppose  $\psi$  is not identically 0 and let  $N$  be its null space. Since

$$d(N) = d(V) - d(\mathcal{R}(\psi)) = d(V) - 1,$$

$N^\perp$  is 1-dimensional and so of the form  $[w]$  for some  $0 \neq w \in V$ . The idea is to exhibit the desired coefficient vector as a scalar multiple of  $w$ , say  $dw$ .

Since the general vector

$$z = y + cw \quad (y \in N, c \in \mathbb{R})$$

of  $V$  is mapped by  $\psi$  into

$$\psi(z) = \psi(y) + c\psi(w) = c\psi(w)$$

and by  $\langle \cdot, dw \rangle$  into

$$\langle z, dw \rangle = \langle y + cw, dw \rangle = cd\langle w, w \rangle,$$

the appropriate choice of  $d$  is

$$d = \frac{\psi(w)}{\|w\|^2}. \quad \blacksquare$$

There are two common ways to produce the coefficient vector  $v$  of a given linear functional. One way is the *direct-construction method* which constructs  $v$  as in the existence part of the proof. The other way is the *confirmation method* which checks that a conjectured  $v$  satisfies (5.2) and then appeals to uniqueness.

**5.3 Exercise [3].** Suppose  $x_1, x_2, \dots, x_n$  is a basis for  $V$ , so that each  $x \in V$  has a unique representation of the form

$$x = x_1\beta_1 + x_2\beta_2 + \dots + x_n\beta_n.$$

Show that the coefficient vector of the so-called  $j^{\text{th}}$  *coordinate functional*

$$\psi_j: x \rightarrow \beta_j$$

is

$$v_j \equiv \left( \frac{Q_j x_j}{\|Q_j x_j\|} \right) \frac{1}{\|Q_j x_j\|} \quad (5.4)$$

where  $Q_j = I - P_j$ ,  $P_j$  being projection onto the manifold spanned by the  $x_i$ 's for  $i \neq j$ . Solve this exercise twice, once using the direct-construction method and once using the confirmation method.  $\diamond$

Suppose now  $F: V \times V \rightarrow \mathbb{R}$  is a *bilinear functional*, i.e., linear in each component, the other component being held fixed. For each  $z \in V$ , the map

$$y \rightarrow F(y, z)$$

is linear and so of the form

$$F(y, z) = \langle y, Cz \rangle$$

for a unique vector  $Cz \in V$ . Since

$$\begin{aligned} \langle y, C(c_1 z_1 + c_2 z_2) \rangle &= F(y, c_1 z_1 + c_2 z_2) = c_1 F(y, z_1) + c_2 F(y, z_2) \\ &= \langle y, c_1 C z_1 + c_2 C z_2 \rangle \end{aligned}$$

for all  $y \in V$ , the mapping  $C: V \rightarrow V$  is linear. We have established:

**5.5 Proposition (Representation theorem for bilinear functionals).** *If  $F: V \times V \rightarrow \mathbb{R}$  is bilinear, then there exists a unique linear transformation  $C: V \rightarrow V$  such that*

$$F(x, y) = \langle x, Cy \rangle \quad \text{for all } x \text{ and } y \in V. \quad (5.6)$$

This result has several important consequences. First, if  $T: V \rightarrow V$  is linear, then

$$F(x, y) \equiv \langle Tx, y \rangle$$

defines a bilinear functional on  $V$  which is necessarily of the form

$$\langle Tx, y \rangle = \langle x, T'y \rangle$$

for some unique linear transformation  $T': V \rightarrow V$ .  $T'$  is called the *adjoint* of  $T$ . One has

$$(T')' = T, \quad (ST)' = T'S', \quad (5.7)$$

and  $T' = T$  if and only if  $T$  is self-adjoint in the sense of (4.1). Moreover,  $T$  is orthogonal if and only if  $T' = T^{-1}$ , because the identity

$$\langle Tx, Ty \rangle - \langle x, y \rangle = \langle (T'T - I)x, y \rangle$$

for  $x, y \in V$  implies that  $T$  preserves inner products if and only if  $T'T = I$ .

**5.8 Exercise [1].** Let  $T$  be a linear transformation from  $V$  to  $V$ , and let  $\mathbf{A} = (a_{ij})$  be the matrix of  $T$  with respect to a given orthonormal basis for  $V$ . Show that the matrix of  $T'$  is the transpose  $(a_{ji})$  of  $\mathbf{A}$ .  $\diamond$

**5.9 Exercise [3].** Let  $T$  be a linear transformation from  $V$  to  $V$ . Show that  $\mathcal{N}(T') = (\mathcal{R}(T))^\perp$  and deduce that  $T'$  is nonsingular if and only if  $T$  is.  $\diamond$

**5.10 Exercise [3].** Let  $M$  be a subspace of  $V$  and let  $\mathcal{O}: V \rightarrow V$  be an orthogonal linear transformation. Show that  $\mathcal{O}P_M\mathcal{O}'$  is orthogonal projection onto  $\mathcal{O}(M)$ .  $\diamond$

**5.11 Exercise [3].** Locate at least two places where use of the relation  $(ST)' = T'S'$  would simplify the calculations in Section 2.2.  $\diamond$

Next suppose that  $F: V \times V \rightarrow \mathbb{R}$  is both bilinear and symmetric. The representing transformation  $C$  is then self-adjoint and one can find an orthonormal basis for  $V$  consisting of eigenvectors  $b_1, \dots, b_n$  of  $C$  with associated eigenvalues  $\lambda_1, \dots, \lambda_n$ .  $F$  then takes the form

$$F(x, y) = \langle x, Cy \rangle = \left\langle \sum_i c_i b_i, C \sum_j d_j b_j \right\rangle = \sum_i \lambda_i c_i d_i, \quad (5.12)$$

where the  $c_i = \langle x, b_i \rangle$ 's and  $d_j = \langle y, b_j \rangle$ 's are the coefficients of  $x$  and  $y$  with respect to the  $b$ -basis. Thus when viewed from the right perspective, the general symmetric bilinear functional on  $V$  is just a weighted dot-product (the weights can be zero or negative).

**5.13 Exercise** [3]. A quadratic form  $Q$  on  $V$  is a functional of the form

$$Q(v) = \langle Cv, v \rangle$$

for some linear transformation  $C: V \rightarrow V$ . Show that there is no loss of generality in supposing that  $C$  is self-adjoint, in which case  $C$  is uniquely determined by  $Q$ . [Hint: For uniqueness, generalize (1.6) to

$$\langle Cv, w \rangle = \frac{Q(v+w) - Q(v) - Q(w)}{2} .] \quad (5.14) \diamond$$

Finally, suppose  $F = [\cdot, \cdot]$  is another inner product on  $V$ , so that  $F$  is positive-definite in addition to being symmetric and bilinear. We then have

$$[x, y] = \langle x, Cy \rangle \quad \text{for all } x \text{ and } y \text{ in } V \quad (5.15)$$

for a unique self-adjoint linear transformation  $C$  on  $V$  which is positive definite in the sense that

$$\langle x, Cx \rangle \geq 0 \text{ for all } x \in V \quad \text{and} \quad \langle x, Cx \rangle = 0 \text{ only for } x = 0.$$

In other words, with respect to an appropriate orthonormal basis  $b_1, \dots, b_n$ ,  $[\cdot, \cdot]$  takes the form

$$[x, y] = \sum_i \lambda_i c_i d_i, \quad \text{for } c_i = \langle x, b_i \rangle, \quad d_i = \langle y, b_i \rangle \quad (5.16)$$

with the weights  $\lambda_1, \lambda_2, \dots, \lambda_n$  being strictly positive.

**5.17 Exercise** [4]. Suppose  $V$  and  $W$  are both inner product spaces. Formulate and prove a representation theorem for bilinear functionals on  $V \times W$ . Show that if  $T: V \rightarrow W$  is linear, then there exists a unique linear transformation  $T': W \rightarrow V$  such that

$$\langle Tv, w \rangle_W = \langle v, T'w \rangle_V \quad \text{for all } v \in V \text{ and } w \in W; \quad (5.18)$$

$T'$  is called the *adjoint of  $T$* . Formulate and prove extensions of (5.7) and the assertion of Exercise 5.8.  $\diamond$

**5.19 Exercise** [2]. Suppose  $M$  is a subspace of  $V$ . Up to now we have considered  $P_M$  as a linear transformation from  $V$  to  $V$ , and as such it is self-adjoint. One can, however, think of  $P_M$  as being a linear transformation from  $V$  to  $M$ . If one takes this point of view, what then is the adjoint of  $P_M$ ? (The inner product of two elements of  $M$  is defined to be their inner product in  $V$ .)  $\diamond$

**5.20 Exercise** [3]. Suppose  $V$  and  $W$  are both inner product spaces and that  $T: V \rightarrow W$  is linear. Show that

$$\mathcal{R}(T') = \mathcal{R}(T'T) \quad \text{and} \quad \mathcal{N}(T) = \mathcal{N}(T'T). \quad (5.21)$$

Deduce that  $T'T$  is nonsingular if and only if  $T$  is.  $\diamond$

**5.22 Exercise** [3]. Suppose  $B$  and  $V$  are both inner product spaces and that  $T: B \rightarrow V$  is linear. Put  $M = T(B) = \{Tb : b \in B\}$ . Show that  $M$  is a subspace of  $V$  and that for each  $v \in V$ , one has  $P_M v = Tb$  if and only if  $b \in B$  satisfies the so-called *normal equations*  $T'Tb = Tv$ . Do not assume that  $T$  is nonsingular.  $\diamond$

**5.23 Exercise** [3]. Suppose  $\langle \cdot, \cdot \rangle$  and  $[\cdot, \cdot]$  are both inner products on  $V$ . Express the unique  $[\cdot, \cdot]$ -self-adjoint, positive definite linear transformation  $D$  such that  $\langle x, y \rangle = [x, Dy]$  for all  $x, y \in V$  in terms of the  $C$  of equation (5.15).  $\diamond$

### 6. Problem set: Cleveland's identity

Let  $(V, \langle \cdot, \cdot \rangle)$  be an inner product space. Let  $\langle \cdot, \cdot \rangle^* = \langle \cdot, \Delta \cdot \rangle$  be another inner product on  $V$ :  $\Delta$  is a positive definite self-adjoint linear transformation from  $V$  to  $V$ . For any subspace  $M$  of  $V$ , let  $P_M$  denote projection onto  $M$  with respect to  $\langle \cdot, \cdot \rangle$ , and let  $P_M^*$  denote projection onto  $M$  with respect to  $\langle \cdot, \cdot \rangle^*$ . Denote the corresponding residual projections by  $Q_M$  and  $Q_M^*$  respectively. For any vector  $z$  in  $V$ , one has of course

$$\|Q_M^* z\|^2 \geq \|Q_M z\|^2.$$

The question of how much larger  $\|Q_M^* z\|^2$  can be than  $\|Q_M z\|^2$  is addressed by *Cleveland's identity*, which asserts

$$\sup_M \sup_{z \in V, z \notin M} \frac{\|Q_M^* z\|^2}{\|Q_M z\|^2} = \frac{(1 + \tau)^2}{4\tau}, \quad (6.1)$$

where

$$\tau = \frac{\lambda_n}{\lambda_1} \quad (6.2)$$

is the ratio of the largest eigenvalue  $\lambda_n$  of  $\Delta$  to the smallest eigenvalue  $\lambda_1$  of  $\Delta$ . By the extremal characterization of the eigenvalues of  $\Delta$ , one has

$$\lambda_1 = \inf_{v \in V, v \neq 0} \left( \frac{\|v\|^*}{\|v\|} \right)^2 \quad \text{and} \quad \lambda_n = \sup_{v \in V, v \neq 0} \left( \frac{\|v\|^*}{\|v\|} \right)^2. \quad (6.3)$$

In this problem set you are asked to establish (6.1). Here are the steps in the argument. To begin with, suppose  $M$  and  $z \notin M$  are given; one needs to show

$$\frac{\|Q_M^* z\|^2}{\|Q_M z\|^2} \leq \frac{(1 + \tau)^2}{4\tau}. \quad (6.4)$$

**A.** Show that without loss of generality, one can consider just the case where  $z \perp M$  (with respect to  $\langle \cdot, \cdot \rangle$ ) with  $\|z\| = 1$  and where  $M$  is one-dimensional, say  $M = [y]$  with  $\|y\| = 1$ .

[Hint: Given  $z$  and  $M$  such that  $P_M z \neq P_M^* z$ , find a one-dimensional subspace  $L$  of  $M$  such that

$$\|Q_M^* z\| = \|Q_L^* u\| \quad \text{and} \quad \|Q_M z\| = \|Q_L u\|$$

for  $u = Q_M z$ .]

$\circ$



**B.** Let  $y$  and  $z$  be as in **(A)**. Put  $W = [y, z]$ . Show that there exists a basis for  $W$ , say  $\{b_1, b_2\}$ , orthonormal with respect to  $\langle \cdot, \cdot \rangle$ , such that for  $v, w \in W$

$$\langle v, w \rangle^* = \beta_1 v_1 w_1 + \beta_2 v_2 w_2$$

where  $v_i = \langle v, b_i \rangle$ ,  $w_i = \langle w, b_i \rangle$ , and

$$\beta_1 = \inf_{u \in W, u \neq 0} \left( \frac{\|u\|^*}{\|u\|} \right)^2 \quad \text{and} \quad \beta_2 = \sup_{u \in W, u \neq 0} \left( \frac{\|u\|^*}{\|u\|} \right)^2.$$

[Hint: Restrict  $\langle \cdot, \cdot \rangle$  and  $\langle \cdot, \cdot \rangle^*$  to  $W \times W$ .]

**C.** Let  $y$  and  $z$  be as in **(A)**,  $b_1, b_2, \beta_1, \beta_2$  as in **(B)**. Show that

$$\frac{\|Q_M^* z\|^2}{\|Q_M z\|^2} = \frac{\beta_1^2 y_1^2 + \beta_2^2 y_2^2}{(\beta_1 y_1^2 + \beta_2 y_2^2)^2},$$

where  $y_i = \langle y, b_i \rangle$ .

[Hint: Compute! Note that the matrix  $\begin{pmatrix} y_1 & z_1 \\ y_2 & z_2 \end{pmatrix}$  is orthogonal.]

**D.** Show that

$$\sup_{p_1, p_2 \geq 0; p_1 + p_2 = 1} \frac{\beta_1^2 p_1 + \beta_2^2 p_2}{(\beta_1 p_1 + \beta_2 p_2)^2} = \frac{(\beta_1 + \beta_2)^2}{4\beta_1\beta_2}$$

and deduce (6.4).

**E.** Show that equality holds in (6.1) by exhibiting  $M$  and  $z$  such that

$$\frac{\|Q_M^* z\|^2}{\|Q_M z\|^2} = \frac{(1 + \tau)^2}{4\tau}.$$

## 7. Appendix: Rudiments

This appendix reviews basic definitions and facts about linear algebra with which we presume the reader is acquainted.

### 7A. Vector spaces

A (real) *vector space* is a triple  $(V, +, \cdot)$  consisting of a set  $V$  of objects called vectors and two operations, vector addition, “+”, which associates to each pair  $v_1, v_2$  of vectors in  $V$  a vector  $v_1 + v_2$  in  $V$ , and scalar multiplication, “ $\cdot$ ”, which associates to each vector  $v \in V$  and each scalar  $c \in \mathbb{R}$  a vector  $c \cdot v \equiv cv$  in  $V$ . The operations are assumed to have the following properties:

- (1) For all  $v_1, v_2, v_3 \in V$ ,  $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$ .
- (2) For all  $v_1, v_2 \in V$ ,  $v_1 + v_2 = v_2 + v_1$ .
- (3) There exists a vector  $0 \in V$  such that  $v + 0 = v$  for all  $v \in V$ .

- (4) For each  $v \in V$ , there exists a vector  $-v \in V$  such that  $-v + v = 0$ .  
 For  $v_1 + (-v_2)$  we write also  $v_1 - v_2$ .
- (5) For all  $c \in \mathbb{R}$  and  $v_1, v_2 \in V$ ,  $c(v_1 + v_2) = cv_1 + cv_2$ .
- (6) For all  $c_1, c_2 \in \mathbb{R}$  and  $v \in V$ ,  $(c_1 + c_2)v = c_1v + c_2v$ .
- (7) For all  $c_1, c_2 \in \mathbb{R}$  and  $v \in V$ ,  $(c_1c_2)v = c_1(c_2v)$ .
- (8) For all  $v \in V$ ,  $1 \cdot v = v$ .

An important example of a vector space is  $\mathbb{R}^n$ , the set of  $n$ -tuples (or  $n \times 1$  column vectors) of real numbers, for which vector addition and multiplication by scalars are defined componentwise as addition and multiplication of real numbers. The *transpose* of a column vector  $\mathbf{x} \in \mathbb{R}^n$  with components  $x_1, \dots, x_n$  is the row vector  $\mathbf{x}^T = (x_1, \dots, x_n)$ .

Vectors  $v_1, \dots, v_n$  in a vector space  $V$  are said to be (*linearly*) *independent* if

$$\sum_{1 \leq i \leq n} c_i v_i = 0$$

implies that the  $c_i$ 's are all zero; otherwise the  $v_i$ 's are said to be (*linearly*) *dependent*. The subset of  $M$  of vectors in  $V$  of the form  $\sum_{1 \leq i \leq n} c_i v_i$ , where each  $c_i$  varies over  $\mathbb{R}$ , is called the *span* of  $v_1, \dots, v_n$ , denoted  $[v_1, \dots, v_n]$ . Vectors  $v_1, \dots, v_n$  are said to form a *basis* for  $V$  if and only if they are linearly independent and  $V$  is their span, i.e., if and only if each vector in  $V$  has a unique representation of the form  $\sum_i c_i v_i$ . When  $v_1, \dots, v_n$  is a basis for  $V$ ,  $c_i$  is called the *coordinate* of  $\sum_{1 \leq j \leq n} c_j v_j$  with respect to  $v_i$ .

$V$  is said to be *finite dimensional* if it has a basis consisting of finitely many elements. Every basis for a finite dimensional vector space has the same number of elements; this number is called the *dimension* of  $V$ , denoted  $\dim(V)$ , or just  $d(V)$ . If  $V$  is an  $n$ -dimensional vector space, then  $v_1, \dots, v_n$  is a basis for  $V$  if and only if the  $v_i$ 's are linearly independent, or, equivalently, if and only if they span  $V$ . The *usual coordinate basis* for  $\mathbb{R}^n$  is comprised of the vectors  $\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(n)}$  defined by  $\mathbf{e}_j^{(i)} = \delta_{ij}$ .

We assume henceforth that all vector spaces we deal with are finite dimensional.

## 7B. Subspaces

A nonempty subset  $M$  of a vector space  $V$  which is closed under vector addition and multiplication by scalars is called a *subspace*, or (*linear*) *manifold*, of  $V$ . For example, the span  $[v_1, \dots, v_m]$  of given vectors  $v_1, \dots, v_m$  is a subspace. We write the trivial subspace  $[0]$  simply as  $0$ . A subset  $F$  of  $V$  of the form  $F = v_0 + M = \{v_0 + m : m \in M\}$ , where  $M$  is a subspace and  $v_0 \in V$ , is called a *flat*, or *shifted manifold*;  $F$  is a flat if and only if  $c_1 f_1 + c_2 f_2 \in F$  whenever  $f_1, f_2 \in F$  and  $c_1 + c_2 = 1$ . The dimension of  $F$  is defined to be the dimension of  $M$ .

Two subspaces of  $V$  are said to be *disjoint* if they have only the zero vector in common (this notion should not be confused with the set-theoretic

one). The *intersection*,  $M_1 \cap M_2$ , of two subspaces is defined to be the subspace of  $V$  consisting of all vectors common to  $M_1$  and to  $M_2$ . To say  $M_1 \cap M_2 = 0$  is to say  $M_1$  and  $M_2$  are disjoint. The *sum*  $M_1 + M_2$  of two subspaces of  $V$  is defined to be the subspace of  $V$  consisting of all vectors of the form  $m_1 + m_2$  where  $m_1 \in M_1$  and  $m_2 \in M_2$ . A sum  $M_1 + M_2$  is said to be a *direct sum*, denoted  $M_1 \oplus M_2$ , if  $M_1$  and  $M_2$  are disjoint, i.e., if the representation of  $m \in M_1 + M_2$  as  $m = m_1 + m_2$  with  $m_i \in M_i$  is unique. As regards to dimension, one has

$$d(M_1 + M_2) + d(M_1 \cap M_2) = d(M_1) + d(M_2); \quad (7.1)$$

in particular,  $d(M_1 \oplus M_2) = d(M_1) + d(M_2)$ .

### 7C. Linear functionals

A mapping  $\psi$  of  $V$  into  $\mathbb{R}$  is called a *linear functional* on  $V$  provided it preserves addition and scalar multiplication:  $\psi(c_1v_1 + c_2v_2) = c_1\psi(v_1) + c_2\psi(v_2)$  for all  $c_1, c_2 \in \mathbb{R}$  and  $v_1, v_2 \in V$ . Operations of addition and scalar multiplication on the set  $V^*$  of all linear functionals on  $V$  are defined pointwise, so that  $(\psi_1 + \psi_2)(v) = \psi_1(v) + \psi_2(v)$  and  $(c\psi)(v) = c\psi(v)$ . Under these operations,  $V^*$  is itself a vector space, called the *dual* of  $V$ .  $V^*$  has the same dimension as  $V$ . Examples of linear functionals on  $V$  are the *coordinate functionals*  $\psi_i$  defined relative to a given basis  $v_1, \dots, v_n$  for  $V$  by  $\psi_i(\sum_{1 \leq j \leq n} c_j v_j) = c_i$  for  $i = 1, \dots, n$ .

### 7D. Linear transformations

A mapping  $T$  of a vector space  $V$  into a vector space  $W$  is called a *linear transformation* if it preserves linear structure:  $T(c_1v_1 + c_2v_2) = c_1T(v_1) + c_2T(v_2)$  for all  $c_1, c_2 \in \mathbb{R}$  and  $v_1, v_2 \in V$ . Addition and scalar multiplication of linear transformations from  $V$  to  $W$  are defined pointwise. A linear functional is simply a linear transformation from  $V$  into  $\mathbb{R}^1$ . A linear transformation  $T: V \rightarrow W$  is called a *isomorphism* if it is one-to-one and onto.  $V$  and  $W$  are said to be *isomorphic* if there is an isomorphism mapping one onto the other; isomorphic vector spaces are identical so far as their linear structure is concerned. The *composition* of the linear transformations  $S: U \rightarrow V$  and  $T: V \rightarrow W$  is defined to be the linear transformation  $TS: U \rightarrow W$  which sends  $u \in U$  to  $T(S(u)) \in W$ . A linear transformation  $T: V \rightarrow V$  such that  $T = T^2$  ( $\equiv TT$ ) is said to be *idempotent*. The linear transformation which leaves each  $v \in V$  fixed is called the *identity* transformation and is commonly denoted by  $I$ , or  $I_V$ .

The *range* of a linear transformation  $T$  mapping  $V$  to  $W$  is the subspace of  $W$ , denoted  $\mathcal{R}(T)$  or  $\mathcal{R}(T)$ , consisting of all vectors of the form  $Tv$  for  $v \in V$ . The *rank*,  $\rho(T)$ , of  $T$ , is the dimension of  $\mathcal{R}(T)$ . The *null space* of  $T$  is the subspace of  $V$ , denoted  $\mathcal{N}(T)$  or  $\mathcal{N}(T)$ , consisting of all vectors in  $V$  which are mapped by  $T$  into zero in  $W$ .

The formula

$$d(\mathcal{R}(T)) + d(\mathcal{N}(T)) = d(V), \quad (7.2)$$

holding for any linear transformation  $T$  from  $V$  to  $W$ , has many applications. In particular, if  $V$  and  $W$  have the same dimension and  $\mathcal{N}(T) = 0$ , then  $T$  is an isomorphism of  $V$  and  $W$ . A transformation  $T$  such that  $\mathcal{N}(T) = 0$  is said to be *nonsingular*. The *inverse*,  $T^{-1}$ , of a nonsingular transformation exists on  $\mathcal{R}(T)$  and is linear.

The *matrix* of a linear transformation  $T: V \rightarrow W$  with respect to bases  $v_1, \dots, v_m$  for  $V$  and  $w_1, \dots, w_n$  for  $W$  is the  $n \times m$  array  $[T] \equiv (t_{ij})$  defined by  $T(v_j) = \sum_{1 \leq i \leq n} t_{ij} w_i$  for  $1 \leq j \leq m$ , or, equivalently, by the condition  $T(\sum_j c_j v_j) = \sum_i (\sum_j t_{ij} c_j) w_i$  for  $c_1, \dots, c_m \in V$ . If  $S: U \rightarrow V$  and  $T: V \rightarrow W$  are both linear, then the identity  $[TS] = [T][S]$  holds relative to fixed bases for  $U$ ,  $V$ , and  $W$ .

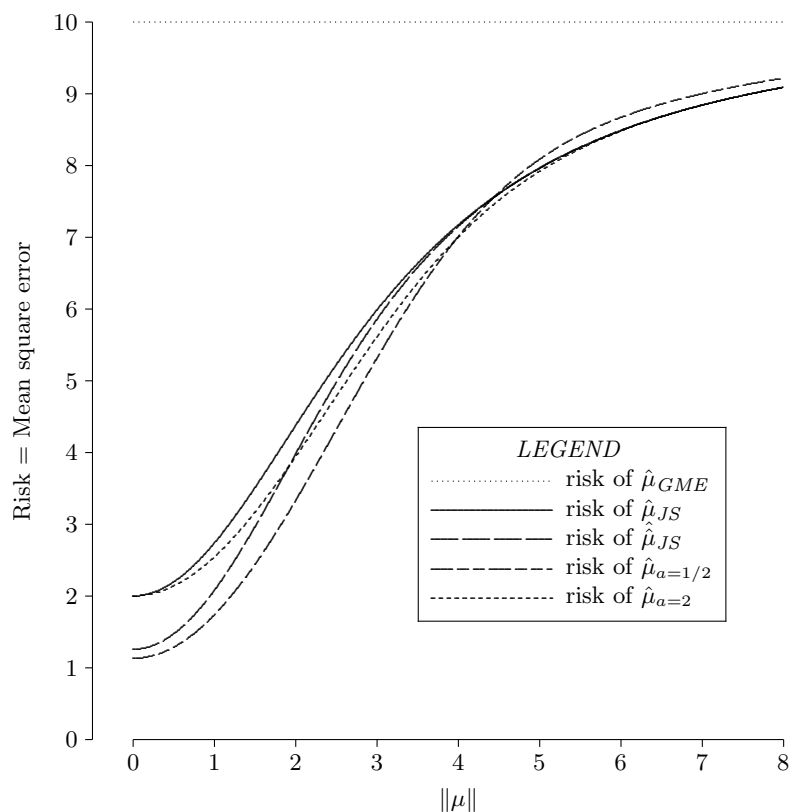
The *transpose* of an  $m \times n$  matrix  $\mathbf{A} = (a_{ij})_{i=1, \dots, m, j=1, \dots, n}$  is the  $n \times m$  matrix  $\mathbf{A}^T = (b_{ji})_{j=1, \dots, n, i=1, \dots, m}$  where  $b_{ji} = a_{ij}$ .

(3)  $\hat{\mu}_{JS}$  is minimax, since it dominates the minimax estimator  $X = P_M Y$ .  $\hat{\mu}_{JS}$  is not, however, admissible, since by Exercise 4.20 it is itself dominated by the so called *positive-part estimator*

$$\hat{\mu}_{JS} \equiv \left(1 - \frac{p-2}{S}\right)^+ X = \max\left(0, 1 - \frac{p-2}{S}\right) X, \quad (4.25)$$

which unlike  $\hat{\mu}_{JS}$  never reverses the sign of  $X$  in estimating  $\mu$ . It is known that every admissible rule in the problem at hand must be a smooth function of the data, so  $\hat{\mu}_{JS}$  is also inadmissible; a dominating rule has yet to be found.

**4.26 Figure.** For the case  $p = 10$  the following diagram displays the risk functions of the the GME  $\hat{\mu}_{GME} = X = P_M Y$ , the James-Stein estimator  $\hat{\mu}_{JS}$  (4.24), the positive part estimator  $\hat{\mu}_{JS}$  (4.25), and the generalized Bayes estimator  $\hat{\mu}_a$  of the next section for the cases  $a = 1/2$  and  $a = 2$  (see (5.13)). The risk functions depend on  $\mu$  only through  $\|\mu\|$ , which is given on the horizontal axis. Evidently all four of the alternatives to  $\hat{\mu}_{GME}$  have substantially smaller risk than  $\hat{\mu}_{GME}$  for  $\mu$ 's near 0, but there is not much difference in the risks of the alternative estimators. The Bayes estimator  $\hat{\mu}_{a=1/2}$  is somewhat preferable to  $\hat{\mu}_{JS}$  and  $\hat{\mu}_{JS}$  in the region where those estimators provide a considerable reduction in risk.  $\hat{\mu}_{a=2}$  dominates  $\hat{\mu}_{JS}$ .



## CHAPTER 9

## REFERENCES

The references given here comprise only those works cited in the notes, and papers from which the problem sets were derived.

- [1] Baranchik, A. J. (1970). A family of minimax estimator of the mean of a multivariate normal distribution. *Ann. Math. Statist.*, **41**, 642–645.
- [2] Berger, J. O. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.*, **4**, 223–226.
- [3] Bickel, P. J., and Doksum, K. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden Day, San Francisco.
- [4] Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.*, **42**, 855–903.
- [5] Cleveland, W. S. (1971). Projecting with the wrong inner product and its application to regression with correlated errors and linear filtering of time series. *Ann. Math. Statist.*, **42**, 616–624.
- [6] Efron, B., and Morris, C. (1973). Stein’s estimation rule and its competitors — An empirical Bayes approach. *J. Amer. Statist. Assoc.*, **68**, 117–130.
- [7] Efron, B., and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.*, **4**, 11–21.
- [8] Haberman, S. J. (1975). How much do Gauss-Markov and least squares estimates differ? — A coordinate-free approach. *Ann. Statist.*, **3**, 982–990.
- [9] Kruskal, W. K. (1960). Untitled. Unpublished monograph on the coordinate-free approach to linear models.
- [10] Kruskal, W. K. (1961). The coordinate-free approach to Gauss-Markov estimation, and its application to missing and extra observations. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1**, 435–451.
- [11] Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.
- [12] Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*. Wiley, New York.

- [13] Stein, C. (1966). An approach to the recovery of inter-block information in balanced incomplete designs, in *Festschrift for J. Neyman*, 351–366, Wiley, New York.
- [14] Stein, C. (1973). Estimation of the mean of a multivariate normal distribution. *Proc. Prague. Symp. Asymptotic Statist.*, 345–381.
- [15] Strawderman, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.*, **42**, 385–388.
- [16] Wald, A. (1942). On the power function of the analysis of variance test. *Ann. Math. Statist.*, **13**, 434–439.

## APPENDIX 1

## INDEX

- acceptance region
  - similar of type  $\alpha$ , 131
- adjoint identities, 72
- adjoint of a linear transformation
  - definition of, 38, 39
  - dependence on the inner product, 72
  - for a non-orthogonal projection, 139
  - formula for, given a basis for the range space, 142
  - “reverse and perp” rule, 139
  - for an orthogonal projection, 39
  - for an orthogonal transformation, 38
  - properties of, 38–39
- admissible estimator, 96
- affine estimator, 76
- analysis of covariance
  - $A$  and  $B$  transformations, 143
  - estimation of  $\sigma^2$ , 149
  - $F$ -test of  $H: \mu \in M_0 + N$ , 152–153
    - degrees of freedom for, 153
    - noncentrality parameter for, 154
    - power of, 154
  - $F$ -test of  $H: P_M \mu \in M_0$ , 154
  - formulation of, 143
  - Gauss-Markov estimators, 144–145
  - one-way layout with one covariate
    - estimation of  $\sigma^2$ , 150
    - $F$ -test of the hypothesis of equal adjusted group means, 154, 155
    - $F$ -test of the hypothesis of equal unadjusted group means, 154
  - Gauss-Markov estimators, 146
  - introduction to, 142–143
  - simultaneous confidence intervals
    - for all contrasts in the adjusted group means, 151
  - variances of GMEs, 148–149
- Scheffé confidence intervals
  - for linear functionals of  $\mu_M$ , 150–151
- analysis of covariance (*cont'd*)
  - Scheffé confidence intervals (*cont'd*)
    - for linear functionals of  $\mu_N$ , 151
  - variances and covariances of GMEs, 146–148
- analysis of variance
  - Latin square design, 155–159
  - one-way layout
    - ANOVA table for, 25
    - estimable parametric functionals, 79
    - estimation of  $\sigma^2$ , 70
    - Gauss-Markov estimation in, 63
    - Gauss-Markov estimator of an estimable parametric functional, 81
    - projections in, 15
    - simultaneous confidence intervals
      - for all contrasts in the group means, 132
    - testing the hypothesis that the group means are equal, 109, 114
    - with a covariate, *see* analysis of covariance, one-way layout with one covariate
  - two-way additive layout
    - ANOVA table for, 25
    - differential effect, 65
    - estimable parametric functionals, 79–80
    - estimation of  $\sigma^2$ , 71
    - Gauss-Markov estimation in, 65, 69
    - Gauss-Markov estimator of an estimable parametric functional, 81
    - grand mean, 65
    - projections in, 17, 24
    - with a missing observation, *see* incomplete observations model,



- analysis of variance (*cont'd*)
  - two-way additive layout (*cont'd*)
    - with a missing observation (*cont'd*)
      - two-way additive layout with a missing observation
    - with replicated observations, 183
  - analysis of variance table
    - for a Latin square design, 158
    - for a Tjur system, 25
    - for the two-way layout, 25
- ANCOVA, *see* analysis of covariance
- ANOVA, *see* analysis of variance
  
- Baranchik's theorem, 105
- bases
  - canonical, *see* canonical bases for a pair of subspaces
- basis, 42
  - coordinates with respect to, 42
  - orthogonal, 9
  - usual coordinate basis for  $\mathbb{R}^n$ , 42
- Bayes estimator
  - as the mean of the posterior distribution of the parameter given the data, 92
  - definition of, 91
  - extended, 91
  - formal, 106
  - generalized, 106
  - improper, 106
  - proper, 106
  - within  $\epsilon$ , 91
- best linear unbiased estimator, 64
- bilinear functionals
  - definition of, 37
  - positive definite, 39
  - representation theorem for, 38
  - symmetric, 38
- BLUE, *see* best linear unbiased estimator
- book orthogonal, *see* subspaces, book orthogonal
- Borel  $\sigma$ -field, 46
  
- canonical bases for a pair of subspaces, 85
- canonical transformation, 75
- Cauchy-Schwarz inequality, 10
- characteristic function
  - of a normal random vector, 53
  - of a random vector, 53
- chisquare distribution
  - central, 54
  - moment generating function of, 57
  - noncentral, 54, 55
  - Poisson representation of, 56–57
- Cleveland's identity, 40
- closed form of a residual squared length, 70
  
- Cochran's theorem
  - algebraic form of, 20–21
  - for quadratic forms in normally distributed random variables, 59
- coefficient vector, 36
  - confirmation method for, 37
  - direct-construction method for, 37
  - effect of covariates on, 144
  - effect of missing observations on, 177
  - for a coordinate functional, 37
  - in the case of an orthogonal basis, 62
  - obtained by partial orthogonalization, 61–62
  - properties of, 62
  - obtained by projection, 60
- collinearity
  - effect on the variance of a GME, 64
- complete statistic
  - condition for, 90
  - definition of, 89
  - in the GLM, 90
- composition of linear transformations, 43
- confidence intervals
  - for the value of a linear functional  $\psi$  of  $\mu$ , 126
  - relation to the  $F$ -test of  $H: \psi(\mu) = 0$ , 126
  - simultaneous, *see* simultaneous confidence intervals
- confidence region procedure, 131
- contrasts, 132
  - spanning set for a space of contrasts, 133
- coordinate functional, 37, 43, 61
- covariance matrix, 47
- covariance operator, 47–48
  
- degrees of freedom
  - of a chisquare distribution, 54
  - of a squared length, 25
  - of a weakly spherical random vector, 51
  - of an  $\mathcal{F}$  distribution, 57
- design matrix, 1
- design of experiments, 26
- differential effect
  - in a Latin square design, 156
  - in the two-way additive layout, 65
- dimension
  - of a vector space, 42
- direct sum, *see* subspaces, direct sum of
- disjoint subspaces, 42
- dispersion matrix, 32, 49
- dispersion operator ( $\Sigma$ ), 49–50
  - of the projection of a weakly spherical random vector onto a subspace, 51

- dot-product, 6
  - convention about re  $\mathbb{R}^n$ , 6, 45
  - weighted, 38
    - used with uncorrelated random variables, 52
- dual space, 43
- Efron-Morris estimator of a mean vector
  - assumed to lie in the space spanned by the equiangular line, 102
- eigenmanifold, 34
- eigenvalue, 33
- eigenvector, 33
- equi-correlated, 73
- equiangular vector ( $e$ ), 3
- equivalence relation, 27
- $\epsilon$ -Bayes estimator, 91
- estimable parametric functional, 77–84
  - and the covariance of Gauss-Markov estimators, 81
  - characterization of, 78–79
    - subject to constraints, 84
  - Gauss-Markov estimator of, 80
    - subject to constraints, 84
  - in a Tjur design, 82–83
- estimation
  - of a mean, *see* estimation assuming normality, Gauss-Markov estimator, James-Stein estimator of  $\mu$
  - of a variance, *see* estimation of  $\sigma^2$
- estimation assuming normality, 88–107
  - Bayes estimator of  $\mu$  versus the prior  $N_M(0, \lambda I_M)$ , 94
    - with a hyper-prior on  $\lambda$ , 106
  - complete sufficient statistic for  $\mu$  and  $\sigma^2$ , 90
  - conditional distribution of  $\Theta$  given  $Y = y$ , when the marginal distribution of  $\Theta$  is  $N_M(0, \lambda I_M)$  and the conditional distribution of  $Y$  given  $\Theta = \mu$  is  $N_V(\mu, I_V)$ , 94
  - desirable properties of the GME of  $\mu$ 
    - extended Bayes, 95
    - maximum likelihood estimator, 88
    - minimax for mean square error, 95
    - minimum dispersion in the class of all unbiased estimators, 89
  - Efron-Morris estimator of  $\mu$ , 102
  - maximum likelihood estimators of  $\mu$  and  $\sigma^2$ , 88–89
    - joint distribution thereof, 89
  - minimum dispersion unbiased estimators of  $\mu$  and  $\sigma^2$ , 89
  - smoothness with respect to  $\mu$  and  $\sigma^2$  of the expected value of a function of  $Y$ , 90
- estimation assuming normality (*cont'd*)
  - undesirable properties of the GME of  $\mu$ 
    - inadmissibility with respect to mean square error when  $\dim(M) \geq 3$ , 100
  - see also* James-Stein estimator of  $\mu$ , James-Stein type estimators of  $\mu$  of the form  $\phi(S)X$ , Scheffé confidence intervals
- estimation of  $\sigma^2$ 
  - effect of covariates, 149
  - effect of extra observations, 182
  - effect of missing observations, 171–173
  - in a Tjur system of subspaces, 71
  - unbiased, 70–71
    - potential bias in, 71
  - under normality, *see* estimation assuming normality
  - see also* specific models, such as Latin square designs and simple linear regression
- expectation
  - as a nondecreasing function of a parameter, 117
  - of a random vector, 46–47
  - of the squared length of a weakly spherical random vector, 52
- experimental units, 26
- exponential family
  - and completeness, 90
  - and MLR, 115
- extended Bayes estimator, 91
- extra observations model, 179–183
- $\mathcal{F}$  distribution, 57
  - central, 57
  - noncentral, 57
    - MLR property of, 116
    - Poisson representation of, 57
  - unnormalized ( $\mathcal{F}^*$ ), 57
- $F$ -test of  $H: \mu \in M_0$ 
  - as a likelihood ratio test, 110–111
  - as a similar test with maximum average power over the surface of certain spheres, 134
  - as a uniformly most powerful invariant test, 125
    - when  $\sigma^2$  is known, 125
  - as a uniformly most powerful test whose power depends only on the distance from  $\mu$  to  $M_0$  in  $\sigma$  units, 137
- definition of, 111
- effect of covariates, *see* analysis of covariance
- effect of extra observations, 182

- F*-test of  $H: \mu \in M_0$  (*cont'd*)
  - effect of missing observations, *see* incomplete observations model
  - noncentrality parameter for, 113
  - power of, 113
    - as an increasing function of the noncentrality parameter, 114, 117
  - simple interpretations of, 111–112
  - sum of squares for error ( $SS_e$ ), 111
  - sum of squares for testing  $H$  ( $SS_H$ ), 111
  - unbiasedness of, 117
  - when  $M_0$  is the null space of a collection of linear functionals of  $\mu$ 
    - related confidence intervals, 130–131
  - when  $M_0$  is the null space of a linear functional of  $\mu$ , 114
    - related confidence intervals, 126
  - when  $M_0$  is the null space of a set of contrasts, 133
- factor, 26
  - balanced, 26
  - blocks of, 26
  - equivalence relation for, 27
  - levels of, 26
  - orthogonal projection associated with, 29
  - subspace associated with, 29
  - sum of squares associated with, 30
  - trivial, 26
  - units, 26
- factors
  - cross-classification of ( $\times$ ), 27, 29
  - maximum of ( $\vee$ ), 27
  - minimum of ( $\wedge$ ), 28
  - nestedness of, 27
  - orthogonal, 30
    - condition of proportional cell counts for, 31
  - partial ordering ( $\leq$ ) of, 27
  - Tjur design of, *see* Tjur design
- fitted values, 67
- flat, 42
- four penny problem, 65–66
- Gauss-Markov estimator
  - of a linear functional of  $\mu$ 
    - center of a confidence interval, 126
    - closure under linear combinations, 63
    - covariance between two such estimators, 64
    - definition of, 62
    - effect of covariates, *see* analysis of covariance
    - effect of extra observations, 182
- Gauss-Markov estimator (*cont'd*)
  - of a linear functional of  $\mu$  (*cont'd*)
    - effect of missing observations, *see* incomplete observations model
    - estimated standard deviation of, 70
    - guessing method for, 63
    - invariance under linear transformations, 74
    - mean and variance of, 63–64
    - minimum variance among linear unbiased estimators, 64
    - significantly different from zero, 131
    - used to estimate a mean vector, 69
  - of an estimable parametric functional with constraints, 84
  - without constraints, 80
  - of  $\mu$ 
    - as a unique invariant estimator, 77
    - as possibly distinct from the least squares estimator, 72
    - definition of, 67
    - effect of covariates, *see* analysis of covariance
    - effect of extra observations, 181
    - effect of missing observations, *see* incomplete observations model
    - invariance under linear transformations, 74–75
    - methods for finding, 68, 74–75
    - minimum dispersion among linear unbiased estimators, 67
    - minimum MSE among affine estimators having bounded MSE, 76
    - risk under squared error, 91
    - used to estimate linear functionals, 69
    - see also* estimation assuming normality, James-Stein estimator of  $\mu$ , using the wrong inner product, as well as specific models, such as the Latin square design and simple linear regression
- Gauss-Markov theorem, 2
  - for estimable parametric functionals, 80, 84
  - for linear functionals of  $\mu$ , 64
  - for linear transformations of  $\mu$ , 67
  - for  $\mu$ , 67
    - when there are missing observations, 163
  - quantification of, 85–87
- general linear model
  - basis case, *see* multiple linear regression
  - geometric formulation of, 2, 60
  - matricial formulation of, 1
  - reduction to canonical form, 75

- generalized Bayes estimator, 106  
 GLM, *see* general linear model  
 GME, *see* Gauss-Markov estimator  
 goodness of fit statistic  
   asymptotic distribution of, 56  
 Gram-Schmidt orthogonalization, 9  
   in the one-sample problem, 10  
 grand mean  
   in a Latin square design, 156  
   in the two-way additive layout, 65
- Hsu's theorem, 137  
 hypothesis test  
   admissible, 137  
   invariant, 120  
   power function of, 123, 134  
   randomized, 134  
   similar of size  $\alpha$ , 134  
   unbiased of level  $\alpha$ , 134  
   uniformly most powerful, 123  
   uniformly most powerful invariant,  
     123  
   *see also*  $F$ -test of  $H: \mu \in M_0$   
 hypothesis testing assuming normality  
    $F$ -test, *see*  $F$ -test of  $H: \mu \in M_0$   
   general formulation of the problem,  
     108–109, 131  
   likelihood ratio test of  $H: \mu \in M_0$ ,  
     110  
   *see also* confidence intervals, simul-  
     taneous confidence intervals,  
     estimation assuming normality,  
     invariance in hypothesis testing,  
     as well as specific models, such  
     as the Latin square design and  
     simple linear regression
- idempotent, *see* linear transformations,  
   idempotent  
 identity transformation, 43  
 improper Bayes estimator, 106  
 inadmissible estimator, 96  
 incomplete observations model  
    $A$  and  $B$  transformations, 164  
   diagrams for, 164–165  
   estimation of  $\sigma^2$ , 171–173  
    $F$ -test of  $H: \mu \in M_0$ , 173–174  
   degrees of freedom for, 174  
   noncentrality parameter for, 174  
   Yates procedure, 174, 176  
   formulation of, 162–165  
   Gauss-Markov estimator of a linear  
     functional of  $\mu$ , 177–178  
   Gauss-Markov estimator of  $\mu$   
     ANCOVA method for, 169–170  
     consistency equation for, 165–167  
     definition of, 163  
     quadratic function method for, 168  
   Gauss-Markov theorem for, 163, 177
- incomplete observations model (*cont'd*)  
   Scheffé confidence intervals, 179  
   two-way additive layout with a  
     missing observation  
      $F$ -test of the null hypothesis of no  
     row effects, 174–176  
   Gauss-Markov estimator of  $\mu$ , 167,  
     169–170  
   GMEs of  $\nu$ ,  $\alpha_i$ , and  $\beta_j$ , 167  
   introduction to, 160–161  
   variances of the GMEs of  $\nu$ ,  $\alpha_i$ ,  
     and  $\beta_j$ , 178–179
- indicator function of a set, 29  
 inner product  
   between two sets of vectors, 11  
   of linear functionals, 61  
   of vectors, 6  
 inner product space, 6  
 integrability of random variables  
   facts about, 105  
 intersection of subspaces, 43  
   and projections, 10  
 invariance in hypothesis testing, 118–125  
   and the  $F$ -test, 125  
   constancy on orbits, 120  
   distinguish between orbits, 120  
   group of transformations leaving  
     the testing problem invariant,  
     119–120  
   invariance reduction, 122  
     and UMP tests, 123  
   invariant test, 120  
   maximal invariant, 120  
   orbits, 120  
   principle of invariance, 120  
   reduced problem, 122  
 inverse of a linear transformation, 44  
 isomorphism of vector spaces, 43
- James-Stein estimator of  $\mu$   
   Bayesian motivation for, 96–97  
   definition of, 97  
   improvement in Bayes risk, 100  
   inadmissibility of, 101  
   positive part thereof, 101  
   regression motivation for, 97  
   risk of, 97  
     compared to the risk of other  
       estimators, 101  
     upper bound on, 100  
   shrinkage towards a subspace, 102  
   when  $\sigma^2$  is unknown, 102  
 James-Stein theorem, 97  
 James-Stein type estimators of  $\mu$  of the  
   form  $\phi(S)X$ , 96–102  
   admissible minimax generalized  
     Bayes, 107  
   as spherically symmetric estimators,  
     103

- James-Stein type estimators of  $\mu$  of the form  $\phi(S)X$  (*cont'd*)
  - condition to be minimax, 105
  - motivation for, 103
  - risk of, 99
    - compared to the risk of  $\phi^+(S)X$ , 99
  - unbiased estimator for, 104
- Latin square, 156
- Latin square design, 155–159
- least squares, 8
- least squares estimator of  $\mu$ , 72
- Lebesgue measure, 54
- Lehmann-Scheffé theorem
  - for complete sufficient statistics, 90
- length of a vector, 7
- likelihood ratio test of  $H: \mu \in M_0$ , 110
- linear algebra, review of, 41–44
  - basis, 42
  - composition of two linear transformations, 43
  - coordinate functional, 43
  - dimension, 42
  - direct sum of subspaces, 43
  - disjoint subspaces, 42
  - dual space, 43
  - flat, 42
  - idempotent linear transformation, 43
  - identity transformation, 43
  - intersection of subspaces, 43
  - inverse of a linear transformation, 44
  - isomorphism, 43
  - linear dependence, 42
  - linear functional, 43
  - linear independence, 42
  - linear transformation, 43
  - manifold, 42
  - matrix of a linear transformation, 44
  - nonsingular linear transformation, 44
  - null space of a linear transformation, 43
  - range of a linear transformation, 43
  - rank of a linear transformation, 43
  - span, 42
  - subspaces, 42
  - sum of subspaces, 43
  - transpose
    - of a matrix, 44
    - of a vector, 42
  - vector space, 41
- linear dependence of vectors, 42
- linear functional
  - coefficient vector of, *see* coefficient vector
  - confidence interval for, *see* confidence interval for the value of a linear functional  $\psi$  of  $\mu$
  - coordinate functional, 37, 43
  - linear functional (*cont'd*)
    - definition of, 36, 43
    - Gauss-Markov estimator of, *see* Gauss-Markov estimator of a linear functional of  $\mu$
    - inner product for, 61
    - norm of, 61
    - representation theorem for, 36
    - testing an hypothesis about, 114
- linear independence of vectors, 42
- linear transformation
  - adjoint of, *see* adjoint of a linear transformation
  - composition of, 43
  - definition of, 43
  - eigenmanifolds of, 34
  - eigenvalues of, 33
  - eigenvectors of, 33
  - idempotent, 12, 43
  - identity, 43
  - inverse of, 44
  - matrix of, 33, 44
    - notation for, 44
  - nonsingular, 44
  - null space of ( $\mathcal{N}$ ), 43
  - orthogonal, *see* orthogonal, transformation
  - orthogonal projection, *see* projections, orthogonal
  - range of ( $\mathcal{R}$ ), 43
  - rank of ( $\rho$ ), 43
  - self-adjoint, *see* self-adjoint linear transformation
  - uniqueness of
    - in terms of inner products, 8
- LSD, *see* Latin square design
- manifold
  - eigen, 34
  - regression, 60
  - see also* subspaces
- matrix
  - of a linear transformation, 33
  - of a quadratic form, 58
  - of the adjoint of a linear transformation, 38
  - symmetric
    - diagonalization of, 36
- maximal invariant, 120
- maximum likelihood estimators of  $\mu$  and  $\sigma^2$ , assuming normality, 88–89
- mean
  - grand
    - in the two-way additive layout, 65
  - of a random vector, 46–47

- mean square error
  - definition of, 75
  - formula for, 76, 91
  - normalized, 95
- method of superscripts, 24
- metric induced by an inner product, 10
- minimax estimator
  - as a constant risk, extended-Bayes estimator, 91
  - definition of, 91
- minimum mean square error predictor, 11
- Minkowski's inequality, 10
- missing observations model, 162
  - see also* incomplete observations model
- mixed-up observations model, 162
  - see also* incomplete observations model
- MLR, *see* monotone likelihood ratio
- Möbius function, 25
- Möbius inversion formula, 25
- models
  - analysis of covariance
    - one-way layout with one covariate, *see* analysis of covariance
  - analysis of variance
    - one-way layout, *see* analysis of variance
    - two-way layout, *see* analysis of variance
  - extra observations, *see* extra observations model
  - four penny problem, 65–66
  - general linear, *see* general linear model
  - incomplete observations, *see* incomplete observations model
  - Latin square design, 155–159
  - missing observations, *see* incomplete observations model
  - mixed-up observations, *see* incomplete observations model
  - multiple linear regression, *see* multiple linear regression
  - one-sample problem, *see* one-sample problem
  - replicated observations model, *see* extra observations model
  - simple linear regression, *see* simple linear regression
  - split plot design, *see* split plot design
  - triangle problem, *see* triangle problem
  - two-sample problem, *see* two-sample problem
- monotone likelihood ratio
  - and exponential families, 115
  - and hypothesis testing, 124
  - monotone likelihood ratio (*cont'd*)
    - and the expectation of a nondecreasing function, 117
    - closure under composition, 116
    - definition of, 115
    - examples of
      - exponential families, 115
      - noncentral  $\mathcal{F}$  family, 116
      - Poisson family, 115
      - unnormalized  $\mathcal{F}$  family, 115
      - unnormalized  $t$  family, 118
    - stochastic monotonicity, 117
- MSE, *see* mean square error
- multiple linear regression
  - coordinate functionals in, 61
  - estimable parametric functionals, 79
  - estimation of  $\sigma^2$ , 71
  - expected response at the mean values of the independent variables
    - Gauss-Markov estimation of, 65
  - regression coefficients
    - covariances of, 66
    - expectations of, 66
    - formula for, 66
    - normal equations for, 63
    - variances of, 64
- noncentrality parameter
  - of a chisquare distribution, 54
  - of an  $\mathcal{F}$  distribution, 57
- nonestimable parametric functional, 79
  - completely undetermined by  $\mu$ , 79
  - no unbiased estimator for, 80
- nonsingular distribution, 50–51
- nonsingular linear transformation, 44
- norm
  - of a linear functional, 61
  - of a vector, 7
- normal distribution, 53
  - and conditional distributions, 93
    - conditional distribution of  $\Theta$  given  $Y = y$ , when the marginal distribution of  $\Theta$  is  $N_M(0, \lambda I_M)$  and the conditional distribution of  $Y$  given  $\Theta = \mu$  is  $N_V(\mu, I_V)$ , 94
  - and marginal distributions, 93
  - and quadratic forms, 58–59
  - characteristic function for, 53
  - density of, 54
  - in a lower dimensional subspace, 54
- normal equations
  - abstract formulation of, 40
  - and estimable parametric functionals, 80–81, 84
  - and Gauss-Markov estimation, 68
  - basis formulation of, 10, 69
  - obtained via Gauss-Markov estimation, 63
- normalized mean square error, 95

- notational conventions
  - difficulty level of exercises, 5
  - inner product between
    - a single vector and a set of vectors, 11
    - two sets of vectors, 11
  - markers for the end of a proof,
    - example, exercise, and part of a problem set, 5
  - matrices versus linear transformations, 5
  - numbering scheme, 5
  - summation over non-missing subscripts ( $\oplus$ ), 167
- null space of a linear transformation, 43
- observed values, 67
- one-sample problem
  - and Gram-Schmidt orthogonalization, 10
  - condition for the least squares and Gauss-Markov estimators to be the same, 73
  - equi-correlated case, 73
  - heteroscedastic case, 52
  - independence of the sample mean and variance, 3
  - projections in, 9–10
  - re Cochran's theorem, 59
- one-way layout, *see* analysis of variance, analysis of covariance
- open form of a residual squared length, 70
- orbit, 120
- orthogonal
  - basis, 9
  - and Gauss-Markov estimation, 68
  - complement of a subspace, 12
  - properties of, 12
  - complement of one subspace within another, 13
  - and spanning vectors, 13
  - decompositions, 19
  - factors, *see* factors, orthogonal
  - polynomials
    - derived from 1,  $i$ , and  $i^2$ , 7
  - projection, *see* projections, orthogonal
  - subspaces, 16
  - transformation, 7
    - adjoint of, 38
    - inverse of, 38
    - reflection through a subspace, 13
    - used to characterize strict sphericity, 53
    - used to characterize weak sphericity, 51
  - vectors, 7
- parameter vector, 1
- parametric functional, 78
  - estimable, *see* estimable parametric functional
  - nonestimable, *see* nonestimable parametric functional
- partial order relation ( $\leq$ ) on
  - factors, 27
  - self-adjoint linear transformations, 13
- partial orthogonalization
  - in the analysis of covariance, 145
- partially ordered set, 23
  - greatest lower bound ( $\wedge$ ), 28
  - least upper bound ( $\vee$ ), 27
  - minimal element, 23
  - Möbius function for, 25
  - Möbius inversion formula for, 25
  - principle of induction for, 23
  - zeta function for, 26
- perpendicularity
  - of subspaces, 16
  - of vectors, 7
- positive definiteness, 6, 39
- power function, 123
  - of the  $F$ -test, 113
    - effect of covariates, 154
    - effect of extra observations, 183
    - effect of missing observations, 174
- prediction
  - mean square error of, 11
- principle of invariance
  - in estimation, 77
  - in hypothesis testing, 120
- prior, 91
- projections
  - non-orthogonal
    - adjoints of, 139
    - as idempotent linear transformations, 139
    - definition of, 138
    - formula for the adjoint, given a basis for the range space, 142
    - formula for, given a basis for the range space, 141
    - partial inverses to, 140
  - orthogonal
    - adjoints of, 39
    - and minimum distance, 8, *see also* Cleveland's identity
    - and residuals, 8
    - as idempotent self-adjoint linear transformations, 12–13
    - definition of, 8
    - differences of, 14–15
    - matrix for, 11
    - obtained by making a linear transformation of the problem, 74

- projections (*cont'd*)
  - orthogonal (*cont'd*)
    - onto  $\mathcal{OM}$  for an orthogonal transformation  $\mathcal{O}$ , 38
    - onto a one-dimensional space, 9, 13
    - onto a subspace of a subspace, 14–15
    - onto a subspace that is an element of a Tjur system, 23
    - onto a sum of mutually orthogonal subspaces, 16
    - onto arbitrary subspaces, 10
    - onto the intersection of two subspaces, 18
    - onto the orthogonal complement of the null space of a set of contrasts, 133
    - products of, 17–18
    - sums of, 16–17
    - used in establishing the distributions of quadratic forms, 58
    - used to find Gauss-Markov estimators, 68
- proper Bayes estimator, 106
- proportional cell counts
  - condition for orthogonal factors, 31
- Pythagorean theorem, 7
  
- quadratic form
  - definition of, 39
  - of a normal random vector
    - distribution of, 58–59
  - on an inner product space, 58
  - representation theorem for, 39
  
- random vectors
  - characteristic function for, 53
  - covariance operators for, 47–48
  - definition of, 45
  - dispersion operators of, 49–50
    - spectral representation of, 50
  - distribution of projections of
    - under first and second moment conditions, 54–55
    - under normality assumptions, 55
  - expectation of, 46–47
  - mean of, 46–47
  - normally distributed, 53
    - conditional distribution, 93
    - marginal distribution, 93
    - quadratic forms in, 58–59
  - spherically distributed, 53
  - uncorrelated, 48
  - weakly spherical, 51–52
    - variance parameter of, 51
  - with nonsingular distributions, 50–51
  - with singular distributions, 50–51
- range of a linear transformation, 43
  
- rank of a linear transformation, 43
- reduction to canonical form, 75
- regression, *see* multiple linear regression, simple linear regression
- regression coefficients, 66
  - dependence on the number of explanatory variables, 67
- regression manifold, 60
- regression matrix, 1
- relative orthogonal complement, 13
  - and spanning vectors, 13
- replicated observations model, 180
- representation theorem
  - for bilinear functionals, 38
  - for linear functionals, 36
  - for quadratic forms, 39
- residual squared length
  - closed form of, 70
  - open form of, 70
- residuals, 8
  - inner products of, 141
  - length of, 8
  - uncorrelated with fitted values, 68
- risk, 91
  - of the GME of  $\mu$  relative to mean square error, 91
- see also* James-Stein estimator of  $\mu$ , James-Stein type estimators of  $\mu$  of the form  $\phi(S)X$
  
- $R^2$ 
  - as the ratio of explained variation to total variation, 68
  - as the square of the correlation coefficient between the observed and fitted values, 68
  
- Scheffé confidence intervals
  - for all contrasts in the group means in one-way ANOVA, 132
  - for linear functionals of  $\mu$ , 127–129
    - effect of covariates, *see* analysis of covariance
    - effect of missing observations, *see* incomplete observations model
  - for linear functionals of the component of  $\mu$  orthogonal to  $M_0$ , 132
    - relation to the  $F$ -test, 130–131
- Scheffé multiplier, 128
- $SDFZ$ , *see* significantly different from zero
- self-adjoint linear transformation
  - as a covariance operator, 33
  - definition of, 12
  - extremal characterization of the eigenvalues thereof, 34
  - matrix for, 33
  - partial ordering ( $\leq$ ) of, 13
  - positive definite, 39



- self-adjoint linear transformation
  - (*cont'd*)
  - positive semi-definite, 36
  - square root of, 36
  - range and null space of, 13
  - spectral theorem for, 33–36
  - trace of, 58
    - and preservation of order ( $\leq$ ), 76
- shrinkage estimator, 102
  - see also* James-Stein estimator of  $\mu$ , James-Stein type estimators of  $\mu$  of the form  $\phi(S)X$
- significantly different from zero, 131
- simple linear regression
  - coefficient vector for the slope functional, 62
  - confidence interval(s)
    - for a single predicted value, 126
    - for all predicted values, 129
    - for predicted values when the predictand ranges over an interval, 129
  - definition of, 1
  - estimated SD of
    - the response at the mean of the predictor variable, 70
    - the slope coefficient, 70
  - estimator of  $\sigma^2$ , 70
  - Gauss-Markov estimate of the slope, 63
  - projections in, 10, 11
  - testing for zero slope, 108, 113–114
  - variance of the GME of the slope coefficient, 64
- simultaneous confidence intervals
  - for a line segment of linear functionals, 129
  - for the values of  $\psi(\mu)$  for a linear space of linear functionals  $\psi$ , *see* Scheffé confidence intervals
- singular distribution, 50–51
- span of vectors, 42
- spectral theorem, 33
- spherical distribution, 53
- spherically symmetric estimator, 103
- split plot design, 31
  - estimable parametric functionals, 84
- Stein's lemma, 103
- Stein-Efron-Morris theorem, 104
- stochastic monotonicity, 117
- Strawderman-Berger theorem, 107
- subspaces
  - book orthogonal, 17–18
    - see also* Tjur system
  - definition of, 42
  - direct sum ( $\oplus$ ) of, 19, 43
  - disjoint, 42
  - eigenmanifolds, 34
  - intersection ( $\cap$ ) of, 43
  - subspaces (*cont'd*)
    - non-orthogonal
      - measuring the degree of, 85
    - orthogonal ( $\perp$ ), 16
    - orthogonal complements ( $\cdot^\perp$ ) of, 12
      - properties of, 12
    - orthogonal decompositions of, 19
    - perpendicular ( $\perp$ ), 16
    - relative orthogonal complements of, 13
      - and spanning vectors, 13
    - shifted, 42
    - sum ( $+$ ) of, 43
    - Tjur system of, *see* Tjur system of subspaces
- sufficient statistic
  - definition of, 89
  - factorization criterion for, 90
  - for  $\mu$  and  $\sigma^2$ , 90
    - having minimum dispersion in the class of all unbiased estimators, 90
- sum of squares for error ( $SS_e$ ), 111
- sum of squares for testing  $H$  ( $SS_H$ ), 111
- summation over non-missing subscripts ( $\oplus$ ), 167
- $t$  distribution
  - unnormalized noncentral
    - definition of, 118
    - MLR of, 118
- test, *see*  $F$ -test of  $H: \mu \in M_0$ , hypothesis test
- Tjur design, 31
  - estimable parametric functional, 82–83
  - factor structure diagram for, 32
  - Gauss-Markov estimator of an estimable parametric functional, 82–83
- Tjur system of subspaces, 21–26
  - analysis of variance table for, 25
  - and estimation of  $\sigma^2$ , 71
  - and Gauss-Markov estimation, 69
  - method of  $\text{sup}_b^{\text{per}}$  scripts for, 24
  - Möbius inversion formulas for, 26
  - partial order on, 21
  - representation of, 22
  - structure diagram for, 24
- Tjur's theorem, 22–23
- Tjur, Tue, 21
- trace
  - of a linear transformation, 76
  - of a self-adjoint linear transformation, 58, 76
- transformation, *see* linear transformation

- triangle problem
  - coefficient vectors in, 61
  - covariances of GMEs, 64
  - Gauss-Markov estimation in, 63
  - testing the equilateral hypothesis, 108, 113
- trivial factor, 26
- two-sample problem, 1–2
- two-way layout, *see* analysis of variance
  
- UMP, *see* uniformly most powerful
- UMPI, *see* uniformly most powerful invariant
- unbiased
  - estimator, 63
  - hypothesis test, 134
- uncorrelated random vectors, 48
- uniformly most powerful, 123
- uniformly most powerful invariant, 123
- units factor, 26
- using the wrong inner product
  - conditions for false and true estimates to be the same
    - for linear functionals of  $\mu$ , 74
    - for  $\sigma^2$ , 73
    - for the mean vector  $\mu$ , 72–73
  - difference between the false and true GMEs of  $\mu$ , 86
  - effect on residual squared length, *see* Cleveland's identity
  - ratio of the variances of the false and true GMEs of a linear functional of  $\mu$ , 87
- value of a statistical game, 91
- variance parameter
  - estimation of, *see* estimation of  $\sigma^2$
  - of a weakly spherical random vector
    - assumptions about, 60
    - definition of, 51
- variance-covariance matrix, 49
- variation
  - addition rule for, 68
  - explained, 68
  - residual, 68
  - total, 68
- vector space, 41
  - basis for, 42
  - dimension of, 42
  - dual of, 43
  - finite dimensional, 42
  - isomorphism of, 43
  - manifolds of, 42
  - subspaces of, 42
- vectors
  - angle between, 7
  - equiangular ( $e$ ), 3
  - length of, 7
  - linear dependence of, 42
  - linear independence of, 42
  - norm of, 7
  - orthogonal, 7
  - orthogonal projection of, 8
  - span of, 42
  - uniqueness of
    - in terms of inner products, 8
- Wald's theorem, 134
  - corollaries to, 137
  - reformulation of, 135
- weak sphericity, 51–52
  - achieved by changing the inner product, 52
  - achieved by transformation, 52
- weighing design, 66
  
- Yates procedure, 174
  
- zeta function, 26