CONTINUOUS MULTIVARIATE ANALYSIS: LECTURE NOTES

Joseph George Caldwell, PhD (Statistics)
1432 N Camino Mateo, Tucson, AZ 85745-3311 USA
Tel. (001)(520)222-3446, E-mail jcaldwell9@yahoo.com

Version without figures

March 21, 2018

Contents

1. OVERVIEW

THIS PRESENTATION SUMMARIZES THE THEORY OF MULTIVARIATE STATISTICAL ANALYSIS FOR CONTINUOUS RANDOM VARIABLES. THE PRESENTATION INCLUDES A SUMMARY DESCRIPTION OF THE MAJOR TOPICS OF THIS FIELD. A BRIEF DESCRIPTION OF MULTIVARIATE ANALYSIS FOR DISCRETE RANDOM VARIABLES IS INCLUDED IN ANOTHER PRESENTATION (ON STATISTICAL INFERENCE).

A MAJOR FOCUS OF THE PRESENTATION IS TO COVER MATERIAL THAT IS NECESSARY TO AN UNDERSTANDING OF MULTIVARIATE TIME SERIES ANALYSIS.

CONTINUOUS MULTIVARIATE ANALYSIS REQUIRES A BASIC KNOWLEDGE OF VECTOR AND MATRIX OPERATIONS, SOME OF WHICH HAS BEEN INCLUDED IN OTHER PRESENTATIONS. THIS PRESENTATION SUMMARIZES THAT MATERIAL, AND THEN BUILDS ON IT.

THE TOPICS COVERED IN THIS PRESENTATION ARE THE FOLLOWING:

> CLASSIFICATION OF MULTIVARIATE ANALYSIS METHODOLOGIES
> VECTOR AND MATRIX ALGEBRA RELATED TO CONTINUOUS MULTIVARIATE ANALYSIS
> SOLUTION OF EQUATIONS

THE GENERAL LINEAR STATISTICAL MODEL
THE GENERALIZED LINEAR STATISTICAL MODEL
OPTIMIZATION AND CONSTRAINED OPTIMIZATION
SOME MULTIVARIATE DISTRIBUTIONS
THE MULTIVARIATE GENERAL LINEAR MODEL
MULTIVARIATE ANALYSIS OF VARIANCE AND ANALYSIS OF COVARIANCE
PRINCIPAL COMPONENTS ANALYSIS
FACTOR ANALYSIS
CANONICAL CORRELATION
DISCRIMINANT ANALYSIS
CLASSIFICATION ANALYSIS
MULTIVARIATE TIME SERIES MODELS (SUMMARY)

THERE ARE A LARGE NUMBER OF TEXTBOOKS COVERING THE SUBJECT OF
MULTIVARIATE ANALYSIS. A SELECTION OF THESE BOOKS IS THE FOLLOWING:

Fienberg, Stephen E., *The Analysis of Cross-Classified Categorical Data*, The
MIT Press, 1977

Agresti, Alan, *An Introduction to Categorical Data Analysis*, Wiley, 1996

Agresti, Alan, *Categorical Data Analysis*, Wiley, 1990

Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, Wiley,
1958

Roy, S. N., *Some Aspects of Multivariate Analysis*, Asia Publishing House,
1958

Lawley, D. N., and A. E. Maxwell, *Factor Analysis as a Statistical Method*,
Butterworths, 1963

Harman, Harry H., *Modern Factor Analysis* 2nd ed. revised, University of
Chicago Press, 1960

Cooley, William W. and Paul R. Lohnes, *Multivariate Data Analysis*, Wiley,
1971

Berridge, Damon M. and Robert Crouchley, *Multivariate Generalized Linear Mixed Models Using R*, CRC Press / Chapman & Hall, 2011

MANY BOOKS CONTAIN SECTIONS DEALING WITH MULTIVARIATE APPLICATIONS WITHIN A LARGER SUBJECT AREA, SUCH AS:

Rao, C. Radhakrishna Rao, *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley, 1965, 1973

McCulloch, Charles E., Shayne R. Searle, and John M. Neuhaus, *Generalized, Linear and Mixed Models*, 2nd ed., Wiley, 2008

MANY BOOKS INCLUDE APPENDICES THAT PRESENT SUMMARY DESCRIPTIONS OF MULTIVARIATE ANALYSIS AND RELATED VECTOR AND MATRIX ALGEBRA. THESE INCLUDE:

Tsay, Ruey S., *Multivariate Time Series Analysis with R and Financial Applications*, Wiley, 2014

Lütkepohl, Helmut, *New Introduction of Multiple Time Series Analysis*, Springer, 2006

Hamilton, James D., *Time Series Analysis*, Princeton University Press, 1994

Greene, William H., *Econometric Analysis* 7th ed., Prentice Hall, 2012

2. CLASSIFICATION OF THE PROBLEMS OF MULTIVARIATE ANALYSIS

FIENBERG (OP. CIT.) PROVIDES A SUCCINCT CLASSIFICATION OF THE MAJOR CATEGORIES OF MULTIVARIATE ANALYSIS, CLASSIFIED BY WHETHER THE EXPLAINED AND EXPLANATORY VARIABLES OF AN APPLICATION ARE CONTINUOUS, DISCRETE, OR BOTH.

	Explanatory Variables		
	Categorical	Continuous	Mixed

Response Variables	Categorical	a	b	c
	Continuous	d	e	f
	Mixed	g	h	i

APPLICATIONS IN CATEGORY a ARE CROSS-CLASSIFIED CATEGORICAL DATA PROBLEMS (MULTIDIMENSIONAL CONTINGENCY TABLES), AND ARE DEALT WITH IN TEXTS SUCH AS FIENBERG AND AGRESTI (OP. CIT.). AN EXAMPLE OF THIS TYPE OF APPLICATION IS THE LOG-LINEAR MODEL, WHICH IS DISCUSSED IN A SEPARATE PRESENTATION (ON STATISTICAL INFERENCE). FOR APPLICATIONS INVOLVING A SINGLE RESPONSE VARIABLE, THOSE IN CATEGORIES b AND c ARE DEALT WITH USING GENERALIZED LINEAR MODELS (SUCH AS LOGISTIC AND PROBIT MODELS).

APPLICATIONS IN THE MIDDLE ROW OF THE TABLE (d, e, AND f) CORRESPOND TO THE STANDARD MODELS OF MULTIVARIATE ANALYSIS, INCLUDING MULTIVARIATE ANALYSIS OF VARIANCE (d), MULTIVARIATE REGRESSION ANALYSIS (e), AND MULTIVARIATE ANALYSIS OF COVARIANCE (OR REGRESSION ANALYSIS WITH CATEGORICAL ("DUMMY") EXPLANATORY VARIABLES (f).

APPLICATIONS IN THE LAST ROW OF THE TABLE ARE HANDLED BY MEANS OF SPECIALIZED MODELS. EXAMPLES OF TECHNIQUES FOR ANALYZING SUCH PROBLEMS ARE PRESENTED IN BERRIDGE AND CROUCHLY (OP. CIT.).

THIS PRESENTATION DEALS MAINLY WITH APPLICATIONS OF TYPE e (I.E., MULTIVARIATE PROBLEMS IN WHICH ALL VARIABLES ARE CONTINUOUS, WHETHER EXPLANATORY OR EXPLAINED).

A COMMENT ON TERMINOLOGY. CONFUSION OFTEN ARISES INVOLVING THE MEANING OF THE TERM "MULTIVARIATE." AN APPLICATION IS A MULTIVARIATE APPLICATION ONLY IF IT INVOLVES A JOINT DISTRIBUTION OF RANDOM VARIABLES THAT MAY NOT BE INDEPENDENT. AN APPLICATION MAY INVOLVE MULTIPLE VARIABLES, EITHER RANDOM OR DETERMINISTIC, AND NOT BE A MULTIVARIATE APPLICATION. FOR EXAMPLE, A REGRESSION ANALYSIS INVOLVING A SINGLE RESPONSE VARIABLE AND A NUMBER OF EXPLANATORY VARIABLES (EITHER RANDOM OR DETERMINISTIC) IS A "MULTIVARIABLE" PROBLEM, BUT IT IS NOT A MULTIVARIATE PROBLEM – IT IS A UNIVARIATE PROBLEM.

3. VECTOR AND MATRIX ALGEBRA RELATED TO CONTINUOUS MULTIVARIATE ANALYSIS

THE SUBJECT OF VECTOR AND MATRIX ALGEBRA IS COVERED IN A COURSE IN LINEAR ALGEBRA. THIS SECTION WILL DESCRIBE CONCEPTS OF VECTOR AND MATRIX ALGEBRA SUFFICIENT TO ENABLE AN UNDERSTANDING OF THE BASIC CONCEPTS OF CONTINUOUS MULTIVARIATE ANALYSIS.

REFERENCE TEXTS ON VECTORS AND MATRICES INCLUDE THE FOLLOWING:

Perlis, Sam, *Theory of Matrices*, Addison-Wesley, 1952 (reprinted by Dover, 1991)

Hadley, G, *Linear Algebra*, Addison-Wesley, 1961

Schneider, Hans and George Phillip Barker, *Matrices and Linear Algebra*, 2nd ed. Holt, Reinhart and Winston, 1968 (Dover edition 1989)

Searle, S. R. and W. H. Hausman, *Matrix Algebra for Business and Economics*, Wiley, 1970

Graybill, Franklin A., *Introduction to Matrices with Applications in Statistics*, Wadsworth, 1969

THE PRECEDING TEXTS ARE COMPREHENSIVE AND DETAILED. FOR THE PURPOSES OF THIS PRESENTATION, THEY INCLUDE MUCH MORE INFORMATION THAN IS NEEDED.THE MATERIAL PRESENTED IN THE APPENDICES LISTED EARLIER ON MULTIVARIATE ANALYSIS IS SUFFICIENT (I.E., TO TSAY, HAMILTON, LÜTKEPOHL, AND GREENE). THIS PRESENTATION IS SIMILAR TO BUT BRIEFER THAN THOSE SUMMARIES, PRESENTING THE MINIMAL AMOUNT OF THEORY REQUIRED FOR THIS PRESENTATION.

AS MENTIONED, SOME BASIC MATERIAL ON VECTORS AND MATRICES IS PRESENTED IN OTHER PRESENTATIONS OF THIS SERIES (IN PARTICULAR, DAY 10 ON SMALL-AREA ESTIMATION). IN THE INTEREST OF CONTINUITY AND COMPLETENESS OF THIS PRESENTATION, SOME OF THAT MATERIAL IS REPEATED HERE.

A *COLUMN VECTOR* IS A VERTICAL ARRAY OF A SEQUENCE OF n ELEMENTS

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

A *ROW VECTOR* IS A HORIZONTAL ARRAY OF A SEQUENCE OF n ELEMENTS

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n).$$

IN THIS APPLICATION, THE ELEMENTS ARE SYMBOLS OR VARIABLES OR NUMBERS.

VECTORS ARE INDICATED BY BOLDFACE OR UNDERLINED FONT. A VECTOR CONSISTING OF ONE ELEMENT IS CALLED A *SCALAR*. FOR EXAMPLE, x MAY DENOTE A SCALAR AND **x** AND x̲ MAY DENOTE VECTORS.

THE ELEMENT x_i IS CALLED THE i-th COMPONENT OF **x**. THE NUMBER OF COMPONENTS IN **x** IS VARIOUSLY CALLED THE DIMENSION OR SIZE OR LENGTH OF **x**. (THE TERMS "LENGTH" AND "SIZE" HAVE DIFFERENT MEANINGS, TO BE DEFINED LATER.)

THE *TRANSPOSE* OF A COLUMN VECTOR **x**, DENOTED BY **x**' or **x**^{T} IS THE ROW VECTOR OF LENGTH n, $\boldsymbol{x}' = \boldsymbol{x}^T = (x_1, x_2, \ldots, x_n)$ OR $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)' = (x_1, x_2, \ldots, x_n)^T$. THE TRANSPOSE OF A ROW VECTOR IS DEFINED SIMILARLY.

THE TERM "VECTOR" MAY REFER TO EITHER A COLUMN VECTOR OR A ROW VECTOR. ABSENT AN EXPLICIT INDICATOR (PRIME OR "T"), AN ARBITRARY VECTOR IS ASSUMED TO BE A COLUMN VECTOR.

A *MATRIX* X OF m ROWS AND n COLUMNS (AN "m by n" MATRIX) IS A RECTANGULAR ARRAY OF ELEMENTS:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$$

WE SHALL DENOTE MATRICES STANDARD (NON-BOLDFACE) FONT. (THIS IS NOT A UNIVERSAL CONVENTION – MOST AUTHORS WRITE VECTORS IN BOLDFACE, BUT MANY AUTHORS WRITE MATRICES IN BOLDFACE, ALSO.)

IF x_{ij} DENOTES THE ELEMENT IN ROW i AND COLUMN j OF MATRIX X, THEN THE MATRIX X MAY BE DENOTED AS X = $[x_{ij}]$. IF WE WISH TO MAKE THE NUMBER OF ROWS AND COLUMNS EXPLICIT, WE WRITE X = $[x_{ij}]_{mxn}$. THE TRANSPOSE X' (OR X^T) IS DEFINED AS THE MATRIX HAVING ELEMENT x_{ji} IN ROW i AND COLUMN j. THE ROWS AND THE COLUMNS OF A MATRIX ARE VECTORS. A MATRIX HAVING JUST ONE ROW OR ONE COLUMN IS A VECTOR (OR, IF JUST ONE ROW AND ONE COLUMN, A SCALAR).

IF A MATRIX HAS THE SAME NUMBER OF ROWS AS COLUMNS, IT IS CALLED *SQUARE*, AND THE NUMBER OF ROWS (OR COLUMNS) IS CALLED THE *SIZE* OR *ORDER* OF THE MATRIX.

BASIC OPERATIONS ON MATRICES ARE THE FOLLOWING. SUPPOSE THAT A = $[a_{ij}]$ IS AN m x n MATRIX AND B = $[b_{ij}]$ IS A p x q MATRIX.

ADDITION: IF m=p AND n=q, THEN A + B = $[a_{ij} + b_{ij}]_{mxn}$.
SUBTRACTION: IF m=p AND n=q, THEN A - B = $[a_{ij} - b_{ij}]_{mxn}$.
SCALAR MULTIPLICATION: IF c IS A SCALAR, THEN cA = $[ca_{ij}]$.
MULTIPLICATION: AB = $[\sum_{k=1}^{n} a_{ik} b_{kj}]$ where n=p.

THE PRODUCT OF AN n BY m MATRIX A AND AN m BY k MATRIX B IS THE n BY k MATRIX WHOSE i,j-th ELEMENT (I.E., ENTRY IN ROW i AND COLUMN j) IS THE VECTOR PRODUCT OF THE i-th ROW OF A AND THE j-th COLUMN OF B. NOTE THAT THE PRODUCT IS DEFINED ONLY IF THE MATRICES ARE *CONFORMABLE*, I.E., THE NUMBER OF COLUMNS OF A IS EQUAL TO THE NUMBER OF ROWS OF B.

SINCE VECTORS ARE MATRICES, THE PRECEDING DEFINITIONS APPLY TO VECTORS. FOR EXAMPLE, THE PRODUCT OF A SCALAR a AND A VECTOR **x** WHOSE i-th COMPONENT IS x_i IS THE VECTOR WHOSE i-th COMPONENT is ax_i: a **x**' = (ax_1, ax_2,...,ax_n).

IF VECTORS **a** AND **b** ARE OF THE SAME LENGTH n, THE *VECTOR PRODUCT* (OR INNER PRODUCT) IS $\boldsymbol{a'b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$. TWO VECTORS WHOSE

INNER PRODUCT IS ZERO ARE SAID TO BE ORTHOGONAL. (THIS DEFINITION REFERS TO GEOMETRIC ORTHOGONALITY; IN STATISTICS, TWO RANDOM VECTORS ARE SAID TO BE ORTHOGONAL IF THEY ARE UNCORRELATED.)

THE *DIAGONAL* (OR MAIN DIAGONAL OR PRINCIPAL DIAGONAL) OF A MATRIX IS THE VECTOR OF ELEMENTS FOR WHICH THE ROW INDEX EQUALS THE COLUMN INDEX.

THE SQUARE MATRIX, I, SUCH THAT ALL OF THE DIAGONAL ELEMENTS EQUAL TO ONE AND ZEROS ELSEWHERE IS CALLED THE *IDENTITY* MATRIX, SINCE IA = A FOR ANY MATRIX A. (IF I IS AN n x n MATRIX, IT IS DENOTED AS I_n IF IT IS DESIRED TO INDICATE ITS SIZE.)

THE *INVERSE* OF A SQUARE MATRIX A IS A MATRIX B = A^{-1} SUCH THAT AB = I, IF IT EXISTS, IT IS UNIQUE, AND A^{-1} A = I. IF THE INVERSE EXISTS, THE MATRIX IS CALLED *NONSINGULAR* OR *INVERTIBLE*.

THE *RANK*, r, OF A MATRIX IS THE NUMBER OF LINEARLY INDEPENDENT ROWS OR COLUMNS (WHICH ARE EQUAL). FOR A SQUARE MATRIX, IF r=n, THE MATRIX IS SAID TO BE OF FULL RANK, AND IT IS INVERTIBLE.

A *SYMMETRIC MATRIX* IS A SQUARE MATRIX FOR WHICH $x_{ij} = x_{ji}$.

EXAMPLES:

THE MATRIX

$$\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

IS OF RANK 2. ITS INVERSE IS THE SAME MATRIX.

THE MATRIX

$$\begin{matrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 0 & -1 & 1 \end{matrix}$$

IS SINGULAR (NON-INVERTIBLE), SINCE THE THIRD ROW IS THE FIRST MINUS THE SECOND. IT IS OF RANK 2.

THE MATRIX

$$\begin{matrix} 1 & .5 & 0 \\ .5 & 1 & 0 \\ 0 & 0 & 0 \end{matrix}$$

IS SINGULAR, AND OF RANK 2.

IN MATRIX NOTATION, A SYSTEM OF SIMULTANEOUS LINEAR EQUATIONS, FOR EXAMPLE,

$$y_1 = a_{11} x_1 + a_{12} x_2$$
$$y_2 = a_{21} x_1 + a_{22} x_2$$

MAY BE REPRESENTED AS

$$\mathbf{y} = A \mathbf{x}$$

WHERE $\mathbf{y}' = (y_1, y_2)$, $\mathbf{x}' = (x_1, x_2)$ AND $A = (\mathbf{a}_1, \mathbf{a}_2)$ WHERE $\mathbf{a}_1' = (a_{11}, a_{21})$ AND $\mathbf{a}_2' = (a_{12}, a_{22})$.

CONSIDER THE SYSTEM OF LINEAR EQUATIONS

$A\mathbf{x} = \mathbf{b}$

WHERE A IS AN n x n MATRIX AND $\mathbf{x}$ AND $\mathbf{b}$ ARE VECTORS OF LENGTH n. IT IS DESIRED TO DETERMINE NONZERO SOLUTIONS, $\mathbf{x}$, TO THIS SYSTEM, WHICH ARE LINEAR COMBINATIONS OF THE COLUMNS OF A. THERE ARE TWO CASED TO CONSIDER, DEPENDING ON WHETHER $\mathbf{b}$ IS ZERO OR NONZERO. IF $\mathbf{b}$ IS ZERO, THE SYSTEM IS CALLED *HOMOGENEOUS,* AND IF $\mathbf{b}$ IS NONZERO THE SYSTEM IS CALLED *NONHOMOGENEOUS*.

IF $\mathbf{b} = \mathbf{0}$, THEN A SOLUTION EXISTS IF AND ONLY IF THERE EXISTS A LINEAR COMBINATION (I.E., A$\mathbf{b}$) OF THE COLUMNS OF A THAT IS EQUAL TO $\mathbf{0}$. THAT IS, THERE IS A NONZERO SOLUTION TO THE SYSTEM IF AND ONLY IF A IS NOT OF FULL RANK, I.E., IS SINGULAR.

SUPPOSE THAT **b** IS NONZERO. IT IS DESIRED TO DETERMINE CONDITIONS UNDER WHICH THE SYSTEM HAS SOLUTIONS FOR *ALL* NONZERO VECTORS **b**, NOT JUST FOR SOME VECTORS **b**. IN THIS CASE, THERE IS A SOLUTION IF A IS INVERTIBLE, SINCE IN THAT CASE WE HAVE (PREMULTIPLYING BOTH SIDES OF THE SIMULTANEOUS EQUATIONS BY A^{-1}),

A^{-1}Ax = A^{-1}b

OR

x = A^{-1}b.

IF A IS NOT INVERTIBLE, THEN THERE IS NO MATRIX A^{-1} FOR WHICH A^{-1}A = I, THAT IS, THERE IS NO MATRIX A^{-1} FOR WHICH A^{-1}A**x** = A^{-1}**b**, I.E., NO MATRIX A^{-1} FOR WHICH **x** = A^{-1}**b**. THAT IS, THERE IS NO LINEAR COMBINATION OF THE COLUMNS OF A WHICH SATISFY THE SYSTEM OF EQUATIONS.

IN SUMMARY, FOR **b** NOT EQUAL TO ZERO, THE SYSTEM HAS A NONZERO SOLUTION THAT MAY BE EXPRESSED AS A LINEAR COMBINATION OF THE COLUMNS OF A IF AND ONLY IF A IS NONSINGULAR. THE SOLUTION IS x = A^{-1}b.

THE TRACE OF A MATRIX IS THE SUM OF ITS DIAGONAL ELEMENTS: IF A IS AN nxn MATRIX, THEN $tr(A) = \sum_{i=1}^{n} a_{ii}$. IT HOLDS THAT tr(A + C) = tr(A) + tr(C), tr(AC) = tr(CA) AND tr(A) = tr(A') (WHERE THE MATRICES ARE ASSUMED CONFORMABLE).

SUPPOSE THAT A IS AN nxn SQUARE MATRIX. A NUMBER λ AND AN nx1 NONZERO VECTOR **b** ARE A RIGHT EIGENVALUE AND EIGENVECTOR PAIR OF A IF A**b**=λ**b**. THERE ARE UP TO n POSSIBLE EIGENVALUES FOR A. THEY MAY BE COMPLEX NUMBERS, IN WHICH CASE THEY OCCUR IN CONJUGATE PAIRS. DENOTE THE n EIGENVALUES AS λ_i FOR i= 1,...,n. THEN tr(A) = $\sum_{i=1}^{n} \lambda_i$.

THE DETERMINANT OF A MATRIX, A, DENOTED BY |A|, IS DEFINED AS $|A| = \prod_{i=1}^{n} \lambda_i$. THE NOTATION det(A) IS ALSO COMMON FOR |A|.

IT CAN BE SHOWN THA T THE DETERMINANT OF A MATRIX A = [a$_{ij}$] MAY BE CALCULATED AS THE SUM OF ALL PRODUCTS

$$(-1)^p a_{1i_1}, \ldots, a_{mi_m}$$

CONSISTING OF ONE ELEMENT FROM EACH ROW AND EACH COLUMN, WHERE p IS THE NUMBER OF INVERSIONS REQUIRED TO TRANSFORM THE PERMUTATION $i_1, \ldots, i_m$ INTO $1, \ldots, m$. FOR EXAMPLE, FOR THE 2x2 MATRIX A = $[a_{ij}]$, THE DETERMINANT IS $a_{11}a_{22} - a_{12}a_{21}$.

SOME FEATURES OF DETERMINANTS ARE THE FOLLOWING, WHERE IT IS ASSUMED THAT ALL MATRICES ARE NONSINGULAR (INVERTIBLE):

1. IF A AND B ARE n x n MATRICES, THEN |AB| = |A||B|.
2. |A′| = |A|.
3. $|A^{-1}|$ = 1/|A|.

A MATRIX IS NONSINGULAR IF AND ONLY IF ALL OF ITS EIGENVALUES ARE NONZERO, I.E., ITS DETERMINANT IS NONZERO.

ALL OF THE EIGENVALUES OF A SYMMETRIC MATRIX ARE REAL.

THE RANK OF A MATRIX, A, IS THE NUMBER OF NONZERO EIGENVALUES OF THE SYMMETRIC MATRIX AA′.

SINCE THE EIGENVALUE / EIGENVECTOR PAIR λ, **b** SATISFY

A**b** = λ**b**,

IT FOLLOWS THAT

A**b** − λI**b** = **0**

OR

(A − λI)**b** = **0**.

FROM THE DISCUSSION EARLIER ABOUT SOLUTIONS TO A SYSTEM OF SIMULTANEOUS EQUATIONS, THIS SYSTEM HAS A SOLUTION ONLY IF THE MATRIX A − ΛI IS SINGULAR.

IF $(A - \lambda I)$ WERE NONSINGULAR, THEN WE COULDPREMULTIPLY THE PRECEDING EXPRESSION BY $(A - \lambda I)^{-1}$ AND OBTAIN **b** = **0**. BUT **b** IS ASSUMED TO BE NONZERO, HENCE $(A - \lambda I)$ MUST BE SINGULAR. HENCE AN EIGENVALUE OF A IS A NUMBER λ SUCH THAT $|A - \lambda I| = 0$.

EIGENVALUES ARE ALSO CALLED CHARACTERISTIC VALUES OR CHARACTERISTIC ROOTS OR LATENT ROOTS, AND THE JUST-PRECEDING EQUATION IS CALLED THE CHARACTERISTIC EQUATION.

EIGENVECTORS SPAN THE VECTOR SPACE SPANNED BY THE ROWS OR COLUMNS OF A MATRIX. EIGENVALUES ARE MEASURES OF LENGTH ALONG EIGENVECTORS, AND THE DETERMINANT IS A MEASURE OF VOLUME. FOR EXAMPLE, CONSIDER THE CASE OF AN n x n IDENTITY MATRIX. THE EIGENVECTORS ARE UNIT VECTORS POINTING ALONG n ORTHOGONAL AXES. THE LENGTH OF EACH EIGENVECTOR IS ONE (THE CORRESPONDING EIGENVALUE), AND THE PRODUCT OF ALL OF THE LENGTHS (EIGENVALUES) IS THE n-DIMENSIONAL VOLUME OF THE n-DIMENSIONAL HYPERCUBE SPANNED BY THE n EIGENVECTORS.

THERE ARE A NUMBER OF NUMERICAL METHODS FOR EVALUATING THE DETERMINANT AND FINDING THE EIGENVALUES AND EIGENVECTORS.

DECOMPOSITION, OR FACTORING, OF SYMMETRIC MATRICES

THE JORDAN CANONICAL FORM; JORDAN DECOMPOSITION

LET A DENOTE AN n x n MATRIX WITH EIGENVALUES $\lambda_1,...,\lambda_n$. THEN THERE EXISTS A NONSINGULAR MATRIX P SUCH THAT

$$P^{-1}AP = \begin{bmatrix} \Lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Lambda_n \end{bmatrix} = \Lambda$$

OR

$$A = P\Lambda P^{-1}$$

WHERE

$$\Lambda_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & & \lambda_i \end{bmatrix}$$

THIS DECOMPOSITION OF A IS CALLED THE JORDAN DECOMPOSITION, OR THE JORDAN CANONICAL FORM. (A DECOMPOSITION IS ALSO CALLED A FACTORING.)

IF A MATRIX A HAS DISTINCT EIGENVALUES, THEN IT MAY BE DECOMPOSED AS

$$A = T\Lambda T^{-1}$$

A SYMMETRIC n x n REAL MATRIX A IS SAID TO BE POSITIVE DEFINITE IF $x^T Ax$ IS POSITIVE FOR EVERY NON-ZERO COLUMN VECTOR x OF n REAL NUMBERS. ALTERNATIVELY, A SQUARE MATRIX A IS POSITIVE DEFINITE IF IT IS SYMMETRIC AND ALL EIGENVALUES ARE POSITIVE. A SYMMETRIX REAL MATRIX A IS SAID TO BE POSITIVE SEMIDEFINITE IF $x^T Ax$ IS NONNEGATIVE FOR EVERY NON-ZERO COLUMN VECTOR x OF n REAL NUMBERS.

A TRIANGULAR MATRIX IS ONE HAVING ALL ZEROS ABOVE THE DIAGONAL (IN WHICH CASE IT IS CALLED A LOWER TRIANGULAR MATRIX) OR ALL ZEROS BELOW THE DIAGONAL (IN WHICH CASE IT IS CALLED AN UPPER TRIANGULAR MATRIX).

THE CHOLESKI DECOMPOSITION

IF A IS A POSITIVE DEFINITE n x n MATRIX, THEN THERE EXISTS A TRIANGULAR MATRIX (EITHER UPPER OR LOWER) P WITH POSITIVE MAIN DIAGONAL SUCH THAT

$$P^{-1}AP'^{-1} = I_n$$

OR

$$A = PP'.$$

IF P IS LOWER TRIANGULAR, THIS DECOMPOSITION IS CALLED A CHOLESKI DECOMPOSITION (OR CHOLESKI FACTORIZATION). (IF P IS LOWER TRIANGULAR, THEN P' IS UPPER TRIANGULAR. IF P IS DENOTED BY L (FOR "LOWER") AND P' IS DENOTED BY U (FOR "UPPER"), THEN A = LU AND THE CHOLESKI DECOMPOSITION IS REFERRED TO AS THE "LU" DECOMPOSITION.)

SINCE ALL OF THE EIGENVALUES ARE ASSUMED POSITIVE, THE SQUARE ROOT MAY BE TAKEN OF EACH. IF WE DEFINE

$$\Lambda^{-1/2} = \begin{bmatrix} \sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\lambda_n} \end{bmatrix}$$

AND

$$Q = P\Lambda^{-1}P'$$

THEN

$$QQ = A$$

AND Q IS CALLED THE SQUARE ROOT OF A AND IS DENOTED AS $A^{-1/2}$.

IF A IS POSITIVE SEMIDEFINITE WITH RANK r < n, THEN THERE EXISTS A NONSINGULAR MATRIX P SUCH THAT

$$P^{-1}AP'^{-1} = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}$$

AND

$$A = QQ'$$

WHERE

$$Q = P\begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}.$$

IF **x** DENOTES AN n-COMPONENT VECTOR, THEN THE *VARIANCE* OF A VECTOR RANDOM VARIABLE **x** (DENOTED BY VAR(**x**) OR V(**x**) OR var(**x**)) IS DEFINED TO BE THE n BY n MATRIX Σ WHOSE ij-th ELEMENT IS COV(x_i, x_j). THE i-th DIAGONAL ELEMENT IS THE VARIANCE OF x_i. (THE VARIANCE OF A VECTOR **x** IS ALSO CALLED THE VARIANCE MATRIX OR THE COVARIANCE MATRIX OR THE VARIANCE-COVARIANCE MATRIX OR THE DISPERSION MATRIX.) NOTE THAT A CORRELATION MATRIX IS SYMMETRIC.CORRELATION MATRICES OF FULL RANK ARE POSITIVE DEFINITE.

IF x AND y DENOTE ANY TWO SCALAR RANDOM VARIABLES AND a DENOTES A SCALAR CONSTANT, THEN VAR(ax) = a^2x AND VAR(x + y) = VAR(x) + VAR(y) + 2 COV(x,y).

IF a IS A SCALAR, WE HAVE V(a**x**) = a^2V(**x**). IF **a** IS A VECTOR OF LENGTH n, WE HAVE V(a**x**) = **a** V(**x**) **a**'. IF A IS A MATRIX HAVING n COLUMNS, THEN V(Ax) = AV(**x**)A'. IF A COVARIANCE MATRIX, Σ, IS OF FULL RANK, IT IS POSITIVE DEFINITE AND HAS A UNIQUE MATRIX SQUARE ROOT, DENOTED BY $Σ^{1/2}$, SUCH THAT $Σ^{1/2}Σ^{1/2}$ = Σ. NOTE THATΣ^{-1} = $Σ^{-1/2}Σ^{-1/2}$.

PARTITION OF A MATRIX

A PARTITIONED MATRIX IS A MATRIX WHOSE ELEMENTS ARE MATRICES. FOR EXAMPLE, IF A DENOTES AN m x n MATRIX, THEN THE MATRIX

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

IS A PARTITION OF MATRIX A INTO FOUR SUBMATRICES, WHERE THE NUMBER OF ROWS OF A SUBMATRIX IS THE SAME AS THAT OF A SUBMATRIX TO THE LEFT OR RIGHT, AND THE NUMBER OF COLUMNS OF A SUBMATRIX IS THE SAME AS THAT OF A SUBMATRIX ABOVE OR BELOW. PARTITIONED MATRICES MAY BE ADDED AND MULTIPLIED AS IF THEY WERE COMPOSED OF SCALARS, AS LONG AS THE COMPONENTS OF THE SUM OR PRODUCT ARE CONFORMABLE.

THE REFERENCED APPENDICES PRESENT A NUMBER OF PROPERTIES OF PARTITIONED MATRICES (I.E., EXPRESSIONS FOR INVERSES AND DETERMINANTS).

THE STACKING (vec, vech) AND KRONECKER PRODUCT OPERATORS

IF A = [a_{ij}] IS AND m x n MATRIX AND B = [b_{ij}]IS A p x q MATRIX, THEN THE KRONECKER PRODUCT (OR DIRECT PRODUCT) IF A AND B IS THE mp X nq MATRIX

$$A\otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{ma}B & \cdots & a_{nm}B \end{bmatrix}.$$

IFA = (a_1,...,a_n) IS AN m x n MATRIX WITH m COLUMNS a_i. THE vec OPERATOR (THE *STACKING OPERATOR*) TRANSFORMS A INTO AN mn x 1 VECTOR BY STACKING THE COLUMNS:

$$vec(A) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

THE vech OPERATOR (THE *HALF-STACKING OPERATOR*) IS A STACKING OPERATOR THAT STACKS ONLY THE ELEMENTS ON OR BELOW THE MAIN DIAGONAL OF A SQUARE MATRIX. FOR EXAMPLE:

$$vech\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{22} \\ a_{32} \\ a_{33} \end{bmatrix}.$$

A SUMMARY OF PROPERTIES OF THE vec, vech AND KRONECKER PRODUCT OPERATORS IS PRESENTED IN TSAY (OP. CIT.). FOR EXAMPLE:

vec(A + B) = vec(A) + vec(B)
vec(AB) = (I $\otimes$ A)vec(B) = (B' $\otimes$ I)vec(A).

VECTOR AND MATRIX DIFFERENTIATION

LET y BE A (REAL-VALUED) FUNCTION OF A (REAL-VALUED) VARIABLE x, DENOTED BY y = y(x) OR y = f(X). WE ASSUME THAT THE FUNCTION f(x) IS CONTINUOUS AND DIFFERENTIABLE, WITH DERIVATIVES DENOTED AS

$$f'(x) = \frac{dy}{dx}, f''(x) = \frac{d^2y}{dx^2}, \dots, f^{(n)}(x) = \frac{d^ny}{dx^n} \dots.$$

A TAYLOR SERIES APPROXIMATION OF THE FUNCTION f ABOUT A POINT X^0 IS DEFINED AS

$$f(x) \approx f(x^0) + \sum_{i=1}^{p} \frac{1}{i!} \frac{d^i f(x^0)}{d(x^0)^i} (x - x^0)^i.$$

FOR p=1 WE HAVE A LINEAR TAYLOR SERIES APPROXIMATION:

$$f(x) \approx [f(x^0) - f'(x^0)x^0] + f'(x^0)x.$$

NOW, LET y = f(**x**) BE A (REAL-VALUED) FUNCTION OF A VECTOR **x**. THE GRADIENT VECTOR, OR GRADIENT OF y IS THE VECTOR OF PARTIAL DERIVATIVES:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \partial y/\partial x_1 \\ \partial y/\partial x_2 \\ \vdots \\ \partial y/\partial x_n \end{bmatrix}.$$

THE VECTOR SHAPE OF A DERIVATIVE (ROW OR COLUMN) IS THE SAME AS THAT OF THE DENOMINATOR OF THE DERIVATIVE.

THE SECOND DERIVATIVES MATRIX, OR HESSIAN, IS DEFINED AS

$$H = \begin{bmatrix} \partial^2 y/\partial x_1 \partial x_2 & \cdots & \partial^2 y/\partial x_1 \partial x_n \\ \vdots & \ddots & \vdots \\ \partial^2 y/\partial x_n \partial x_2 & \cdots & \partial^2 y/\partial x_n \partial x_n \end{bmatrix}.$$

IT FOLLOWS THAT

$$H = \frac{\partial^2 y}{\partial \boldsymbol{x} \partial \boldsymbol{x}'}.$$

FORMULAS AND RESULTS FROM ORDINARY (NON-VECTOR) DIFFERENTIATION HAVE ANALOGS IN VECTOR DIFFERENTIATION. THE FOLLOWING ARE SOME EXAMPLES.

IF **a** AND **x** ARE VECTOR (OF THE SAME LENGTH), THEN

$$\frac{\partial \boldsymbol{a}'\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}(note: not\ \boldsymbol{a}').$$

IF **y** AND **x** ARE VECTORS OF LENGTH n AND A IS AN n x n MATRIX, THEN

$$\frac{\partial A\boldsymbol{x}}{\partial \boldsymbol{x}} = A'$$

AND

$$\frac{\partial (\boldsymbol{x}'A\boldsymbol{x})}{\partial \boldsymbol{x}} = (A + A')\boldsymbol{x}.$$

IF A IS SYMMETRIC, THEN THIS LAST EXPRESSION BECOMES

$$\frac{\partial (\boldsymbol{x}'A\boldsymbol{x})}{\partial \boldsymbol{x}} = 2A\boldsymbol{x}.$$

FOR MATRIX A = [a_{ij}], WE DEFINE

$$\frac{\partial f}{\partial A} = \left[\frac{\partial f}{\partial a_{ij}}\right]$$

AND

$$\frac{\partial A}{\partial x} = \left[\frac{\partial a_{ij}}{\partial x}\right] (for\ scalar\ x).$$

THEN WE HAVE

$$\frac{\partial (\boldsymbol{x}'A\boldsymbol{x})}{\partial A} = \boldsymbol{x}\boldsymbol{x}'.$$

CHAIN RULE. LET **y** BE AN n x 1 VECTOR AND **x** BE AN m x 1 VECTOR. SUPPOSE THAT **y** = **g(x)**. SUPPOSE THAT **h**(.) IS A p x 1 VECTOR OF FUNCTIONS g_i(.) OF **x**. THEN

$$\frac{\partial \boldsymbol{h}[g(x)]}{\partial \boldsymbol{x}'} = \frac{\partial \boldsymbol{h}(\boldsymbol{y})}{\partial \boldsymbol{y}'} \frac{\partial \boldsymbol{g}(x)}{\partial \boldsymbol{x}'}.$$

THERE ARE ALSO ANALOGOUS DIFFERENTIATION PRODUCT RULES (PRESENTED IN APPENDIX A OF TSAY (OP. CIT.).

OPTIMIZATION

CONSIDER THE PROBLEM OF FINDING A LOCAL EXTREME VALUE (OPTIMUM; MINIMUM OR MAXIMUM) VALUE OF A (REAL-VALUED) DIFFERENTIABLE FUNCTION f(x) OF A REAL VARIABLE x.

A NECESSARY CONDITION FOR A LOCAL OPTIMUM IS

$$\frac{dy}{dx} = 0.$$

A SUFFICIENT CONDITION FOR A LOCAL OPTIMUM IS

$$\frac{d^2y}{dx^2} < 0 \; for \; a \; maximum$$

$$\frac{d^2y}{dx^2} > 0 \; for \; a \; minimum.$$

FOR A FUNCTION f(x) OF SEVERAL VARIABLES (I.E., OF A VECTOR x), THE NECESSARY CONDITION IS

$$\frac{df(\boldsymbol{x})}{d\boldsymbol{x}} = \mathbf{0},$$

AND THE SUFFICIENT CONDITION IS THAT

$$H = \frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x} \partial \boldsymbol{x}'}$$

BE POSITIVE DEFINITE FOR A MINIMUM AND NEGATIVE DEFINITE FOR A MAXIMUM.

IN GENERAL, IT IS NECESSARY TO DETERMINE THE EIGENVALUES OF H TO DETERMINE WHETHER IT IS POSITIVE DEFINITE OR NEGATIVE DEFINITE. IF, HOWEVER, H HAS THE FORM H = A'A WHERE A IS A KNOWN MATRIX OF FULL RANK, THEN IT IS POSITIVE DEFINITE. (THIS RESULT IS USEFUL IN LEAST-SQUARES ESTIMATION, SUCH AS REGRESSION ANALYSIS.)

CONSTRAINED OPTIMIZATION

CONSIDER THE PROBLEM OF MAXIMIZING THE FUNCTION DIFFERENTIABLE f(x) OF A REAL VARIABLE x, SUBJECT TO THE CONSTRAINT c(x)=0. THIS PROBLEM MAY BE SOLVED BY THE METHOD OF LAGRANGE MULTIPLIERS. A SOLUTION SATISFIES THE CONDITION THAT THE DERIVATIVES OF THE LAGRANGIAN FUNCTION

$$L(x, \lambda) = f(x) - \lambda c(x)$$

WITH RESPECT TO x AND λ BE ZERO. THAT IS:

$$\frac{\partial L(x, \lambda)}{\partial x} = \frac{\partial f}{\partial x} - \lambda \frac{\partial c(x)}{\partial x} = 0$$

$$\frac{\partial L(x, \lambda)}{\partial \lambda} = c(x) = 0.$$

THE PARAMETER λ IS CALLED A LAGRANGE MULTIPLIER.

FOR MULTIPLE CONSTRAINTS, $c_1(x)=0$, ..., $c_p(x)=0$, THERE ARE MULTIPLE LAGRANGE MULTIPLIERS, $\lambda_1,...,\lambda_p$, AND, DENOTING $\boldsymbol{c}(x) = [c_i(x)]$ AND $\boldsymbol{\lambda}= [\lambda_i]$, THE LAGRANGIAN FUNCTION IS

$$L(x, \boldsymbol{\lambda}) = f(x) - \sum_{j=1}^{p} \lambda_j c_j(x) = f(x) - \boldsymbol{\lambda}' \boldsymbol{c}(x).$$

TO FIND THE CONSTRAINED OPTIMUM, THE DERIVATIVES WITH RESPECT TO x AND ALL OF THE λ_i ARE SET EQUAL TO ZERO.

FOR THE MULTIDIMENSIONAL CASE OF A VECTOR **x**, THE CONDITIONS ARE ANALOGOUS. THE FUNCTION TO BE OPTIMIZED IS STILL A REAL-VALUED (SCALAR) FUNCTION – THE ONLY VECTOR IS THE ARGUMENT **x**. THE LAGRANGIAN FUNCTION IS

$$L(x, \lambda) = f(x) - \lambda' c(x).$$

IN THE VECTOR CASE, THE OPTIMAL SOLUTION SATISFIES THE EQUATIONS

$$\frac{\partial L(x, \lambda)}{\partial x} = \frac{\partial f(x)}{\partial x} + \frac{\partial \lambda' c(x)}{\partial x} = 0$$

$$\frac{\partial L(x, \lambda)}{\partial \lambda} = c(x) = 0.$$

THE SECOND TERM IN THE FIRST EQUATION ABOVE IS

$$\frac{\partial \lambda' c(x)}{\partial x} = \frac{\partial c'(x) \lambda}{\partial x} = \left[\frac{\partial c'(x)}{\partial x} \right] \lambda = C' \lambda$$

WHERE C DENOTES THE MATRIX OF DERIVATIVES OF THE CONSTRAINTS WITH RESPECT TO **x**.

LEAST-SQUARES ESTIMATION (THE GENERAL LINEAR STATISTICAL MODEL)

PERHAPS THE MOST WIDELY USED OF ALL STATISTICAL ESTIMATION METHODS IS THE METHOD OF LEAST SQUARES, AND IN PARTICULAR, THE METHOD OF LEAST SQUARES APPLIED TO A LINEAR MODEL.

IN THE METHOD OF LEAST SQUARES, THE MODEL ESTIMATES ARE TAKEN TO BE THOSE VALUES THAT MINIMIZE THE SUM OF SQUARES OF THE DEVIATIONS BETWEEN THE OBSERVED VALUES OF A RANDOM VARIABLE AND THE VALUES PREDICTED BY THE MODEL, I.E., THAT MINIMIZE THE MODEL ERRORS OF ESTIMATION.

UNIVARIATE LINEAR STATISTICAL MODEL

IN THE UNIVARIATE CASE, A LINEAR STATISTICAL MODEL IS DEFINED AS

$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{e}$

WHERE $\mathbf{y} = (y_1, ..., y_n)'$ DENOTES AN n x 1 VECTOR OF n OBSERVED VALUES OF A RESPONSE VARIABLE, $\boldsymbol{\beta} = (\beta_1,...,\beta_p)'$ DENOTES A p x 1 VECTOR OF PARAMETERS (TO BE ESTIMATED), X DENOTES AN n x p MATRIC OF n OBSERVATIONS ON A SET OF p EXPLANATORY VARIABLES,

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{x'}_1 \\ \vdots \\ \mathbf{x'}_n \end{bmatrix}$$

AND $e = (e_1,...,e_n)'$ DENOTES AN n x 1 VECTOR OF MODEL ERROR TERMS. IN THE SIMPLEST CASE IT IS ASSUMED THAT THE MODEL ERROR TERMS ARE UNCORRELATED WITH COMMON VARIANCE σ^2.

FOR THE i-th OBSERVATION, THE MODEL IS

$y_i = \mathbf{x'}_i\boldsymbol{\beta} + e_i.$

LET $\hat{\beta}$ DENOTE AN ESTIMATE OF THE PARAMETER β, AND

$$\hat{y}_i = \hat{y}(\mathbf{x_i}) = \mathbf{x'}_i\widehat{\boldsymbol{\beta}}$$

DENOTE THE ESTIMATE OF y CORRESPONDING THE VALUE $\mathbf{x}_i$ OF THE EXPLANATORY VARIABLES. THEN THE ESTIMATED MODEL ERROR TERM, OR MODEL RESIDUAL, FOR THE i-th OBSERVATION IS $\hat{e} = \hat{y}_i - y_i$. USING THE METHOD OF LEAST SQUARES, THE GOAL IS TO DETERMINE $\hat{\beta}$ SO THAT THE SUM OF SQUARES OF THE RESIDUALS IS MINIMIZED.

THAT IS, MINIMIZE

$$\sum_{i=1}^{n} \hat{e}^2 = (\hat{y}_i - y)^2$$

OR, IN VECTOR NOTATION, MINIMIZE

$$(\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta}) = \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'X'\boldsymbol{y} - y'X\boldsymbol{\beta} + \boldsymbol{\beta}'X'X\boldsymbol{\beta}$$
$$= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'X\boldsymbol{\beta} + \boldsymbol{\beta}'X'X\boldsymbol{\beta}\ .$$

THE DERIVATIVE OF THIS EXPRESSION WITH RESPECT TO $\boldsymbol{\beta}$ IS

$$X'\boldsymbol{y} + 2X'X\boldsymbol{\beta}.$$

SETTING THIS EQUAL TO ZERO AND DENOTING THE SOLUTION BY $\widehat{\boldsymbol{\beta}}$ PRODUCES

$$X'X\widehat{\boldsymbol{\beta}} = X'\boldsymbol{y}.$$

IF XX' IS INVERTIBLE, PREMULTIPLYING BOTH SIDES OF THIS EQUATION BY $(X'X)^{-1}$ YIELDS

$$\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}.$$

THE GENERAL LINEAR MODEL, UNIVARIATE CASE

IN THE PRECEDING MODEL, THE MODEL ERROR TERMS WERE ASSUMED TO BE UNCORRELATED. THIS ASSUMPTION HOLDS, FOR EXAMPLE, IF THE OBSERVATIONS ARE SAMPLED INDEPENDENTLY. IN MANY APPLICATIONS, THE ASSUMPTION OF UNCORRELATEDNESS DOES NOT APPLY. FOR EXAMPLE, IN PANEL SAMPLING OBSERVATIONS TAKEN ON THE SAME HOUSEHOLD AT DIFFERENT TIMES WILL BE CORRELATED; OR OBSERVATIONS IN THE SAME SAMPLE-SURVEY CLUSTER (FIRST STAGE SAMPLE UNIT) OR STRATUM MAY BE CORRELATED.

ALSO, IN MANY APPLICATIONS THE VARIANCE OF THE MODEL ERROR TERMS IS ASSUMED TO BE THE SAME FOR ALL OBSERVATIONS. IN MANY APPLICATIONS THIS ASSUMPTION IS NOT REASONABLE.

SUPPOSE, IN GENERAL, THAT THE VARIANCE MATRIX OF THE MODEL ERROR TERMS IS Σ, AND THAT IT IS OF FULL RANK. THEN, AS DISCUSSED EARLIER, THE SQUARE ROOT MATRIX $\Sigma^{1/2}$ IS DEFINED, SUCH THAT $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$ AND $\Sigma^{-1/2}\Sigma^{1/2} = I$. THE TRANSFORMED VARIATE $Z = \Sigma^{-1/2}Y$ HAS VARIANCE I.

IF THE MODEL IS

$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{e}$

WHERE var($\underline{e}$) = Σ

THEN THE MODEL

$\mathbf{z} = \Sigma^{-1/2}\mathbf{y} = \Sigma^{-1/2}X\boldsymbol{\beta} + \Sigma^{-1/2}\mathbf{e}$

SATISFIES THE CONDITIONS FOR THE LINEAR STATISTICAL MODEL HAVING UNCORRELATED MODEL ERROR TERMS, AND THE LEAST-SQUARES ESTIMATES OF THE PARAMETERS ARE

$$\widehat{\boldsymbol{\beta}} = ((\Sigma^{-1/2}X)'(\Sigma^{-1/2}X))^{-1}(\Sigma^{-1/2}X)'\boldsymbol{\Sigma}^{-1/2}\mathbf{y}$$

OR

$$\widehat{\boldsymbol{\beta}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\boldsymbol{y}.$$

THIS VERSION OF THE LINEAR STATISTICAL MODEL, WITH ARBITRARY VARIANCE MATRIX FOR THE MODEL ERROR TERMS, IS CALLED THE GENERAL LINEAR STATISTICAL MODEL (FOR THE UNIVARIATE CASE), OR SIMPLY THE GENERAL LINEAR MODEL, WITH THE WORD "STATISTICAL" UNDERSTOOD. THIS GENERAL CASE IS ALSO REFERRED TO AS THE METHOD OF WEIGHTED LEAST SQUARES.

MULTIVARIATE LINEAR STATISTICAL MODEL

THE MULTIVARIATE EXTENSION OF THE GENERAL LINEAR MODEL IS STRAIGHTFORWARD. THE MODEL IS

$$Z = X\mathrm{B} + E$$

WHERE Z IS AN n x m MATRIX, X IS AN n x p MATRIX, B IS A p x m MATRIX, AND E IS AN n x m MATRIX. IN THIS FORMAT, THE n MULTIVARIATE OBSERVATIONS ARE THE n ROWS OF Z.

TO CAST THIS MODEL IN THE FORMAT JUST PRESENTED FOR THE UNIVARIATE CASE, THE COLUMNS OF Z ARE STACKED, YIELDING:

$$vec(Z) = (I \otimes X)vec(B) + vec(E).$$

IN THIS FORM, THE MODEL HAS THE SAME STRUCTURE AS THE GENERAL LINEAR MODEL. NOTE THAT THERE ARE TWO TYPES OF COVARIANCES: THOSE AMONG THE m COMPONENT VARIABLES OF Z (CORRESPONDING TO THE COLUMNS OF Z), AND THOSE AMONG THE n SAMPLE OBSERVATIONS (CORRESPONDING TO THE ROWS OF Z). IT IS CLEAR TO SEE THAT THIS MODEL ALLOWS FOR AN EXTREMELY LARGE NUMBER OF PARAMETERS. A MAJOR PROBLEM IN ESTIMATION FOR THIS LINEAR MODEL IS THE IMPOSITION OF RESTRICTIONS TO REDUCE THE NUMBER OF PARAMETERS.

IF THERE IS NO CORRELATION AMONG SAMPLE OBSERVATIONS, THEN THEVARIANCE MATRIX OF vec(E) IS $\Sigma_e \otimes I_n$, WHERE Σ_e DENOTES THE VARIANCE MATRIX AMONG THE COMPONENT VARIABLES OF THE MULTIVARIATE VECTOR. IN THIS CASE IT MAY BE SHOWN THAT THE LEAST-SQUARES ESTIMATES ARE

$$\hat{B} = (X'X)^{-1}(X'Z).$$

NOTE THAT THIS EXPRESSION DOES NOT DEPEND ON Σ_e.

4. SOME PARTICULAR MULTIVARIATE DISTRIBUTIONS

GIVEN TWO DISCRETE RANDOM VARIABLES X AND Y, THE JOINT PROBABILITY FUNCTION IS DEFINED AS

$f_{X,Y}(x,y) = f(x,y) = P(X=x \text{ and } Y=y) = P(X=x,Y=y)$.

GIVEN TWO CONTINUOUS RANDOM VARIABLES X AND Y, A FUNCTION f(x,y) IS A JOINT PROBABILITY DENSITY FUNCTION (PDF) IF

f(x,y)≥0 for all (x,y)

$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = 1$ and

for any set A⊂R x R, P((X,Y)∊A = $\int_{A} f(x,y)dxdy$.

(RECALL THAT R DENOTES THE REAL LINE.)

IN BOTH THE DISCRETE AND CONTINUOUS CASES, THE JOINT CDF IS DEFINED AS

$F_{X,Y}(x,y)$ = P(X≤x, Y≤Y)

THE PRECEDING DEFINITION DEFINES A JOINT DISTRIBUTION IN THE CASE OF TWO RANDOM VARIABLES. SUCH A DISTRIBUTION IS CALLED A BIVARIATE DISTRIBUTION. THIS DEFINITION MAY BE EXTENDED TO THE CASE OF MORE THAN TWO RANDOM VARIABLES. FOR EXAMPLE, GIVEN THREE DISCRETE RANDOM VARIABLES X, Y AND Z, THE JOINT PROBABILITY FUNCTION IS DEFINED AS

$f_{X,Y,Z}(x,y,z)$ = f(x,y,z) = P(X=x and Y=y and Z=z) = P(X=x,Y=y,Z=z).

 A PROBABILITY DISTRIBUTION OF MORE THAN ONE RANDOM VARIABLE IS CALLED A MULTIVARIATE DISTRIBUTION. WHEN DEALING WITH MULTIVARIATE DISTRIBUTIONS IN GENERAL, IT IS MORE COMMON TO USE SUBSCRIPTED VARIABLE NAMES, SUCH AS x_1, x_2, x_3,... or y_1, y_2, y_3,... THAN SEQUENTIAL NAMES, SUCH AS x, y, z,.... IF SOME OF THE RANDOM VARIABLES ARE RESPONSE VARIABLES AND THE OTHERS ARE EXPLANATORY VARIABLES, DIFFERENT LETTERS MAY BE USED, SUCH AS y_1, y_2, y_3,... FOR THE RESPONSE VARIABLES AND x_1, x_2, x_3,... FOR THE EXPLANATORY VARIABLES.

MULTINOMIAL DISTRIBUTION

THE BINOMIAL DISTRIBUTION IS THE DISTRIBUTION OF THE NUMBER OF OCCURRENCES IN EACH OF TWO CATEGORIES, WHERE THE TOTAL SAMPLE SIZE IS FIXED (n).

THE PROBABILITY MASS FUNCTION FOR THE BINOMIAL DISTRIBUTION IS

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} \\ 0 \ otherwise \end{cases}$$

NOTATION X ~ Binomial(n,p).

THE CUMULATIVE DISTRIBUTION FUNCTION FOR THE BINOMIAL DISTRIBUTION IS:

$$F(x) = \sum_{i=0}^{x} \binom{n}{x} p^x (1-p)^{n-x}$$

THE EXTENSION TO MORE THAN TWO CATEGORIES IS DESCRIBED BY THE MULTINOMIAL DISTRIBUTION.

MULTINOMIAL DISTRIBUTION (CASELLA AND BURGER): LET n AND m BE POSITIVE INTEGERS AND LET $p_1,..., p_n$ BE NUMBERS SATISFYING $0 \le p_i \le 1$, i = 1,...,n AND $\sum p_i$ = 1. A RANDOM VECTOR $(X_1,...,X_n)$ HAS A MULTINOMIAL DISTRIBUTION WITH m TRIALS AND CELL PROBABILITIES $p_1,..., p_n$ IF ITS JOINT PROBABILITY MASS FUNCTION IS

$$f(x_1, ..., x_n) = \frac{m!}{x_1! ... x_n!} p_1^{x_1} ... p_n^{x_n} = m! \prod_{i=1}^{n} \frac{p_i^{x_i}}{x_i!}$$

ON THE SET OF $(x_1,...,x_n)$ SUCH THAT EACH x_i IS A NONNEGATIVE INTEGER AND $\sum x_i$ = m.

AN EXAMPLE OF A MULTINOMIAL DISTRIBUTION IS THE NUMBER OF COUNTS IN EACH CELL OF A CROSSTABULATION, WHERE m IS THE TOTAL NUMBER OF COUNTS.

IT IS NOTED THAT ALTHOUGH THE MULTINOMIAL DISTRIBUTION IS DISCRETE, THE METHODS OF CONTINUOUS DISTRUBUTION THEORY ARE OFTEN APPLIED TO PROBLEMS INVOLVING THIS DISTRIBUTION, SUCH AS IN THE CASE OF USING A GENERALIZED LINEAR STATISTICAL MODEL (IN WHICH CASE THE DISTRIBUTION IS REPRESENTED BY AN UNDERLYING CONTINUOUS MULTIVARIATE DISTRIBUTION AND A LINK FUNCTION).

MULTIVARIATE NORMAL DISTRIBUTION

THE *UNIVARIATE NORMAL (GAUSSIAN) DISTRIBUTION* HAS DENSITY FUNCTION

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, x\epsilon R,, where\ \mu\epsilon R\ and\ \sigma > 0.$$

NOTATION X ~ N(μ, σ²).

AND DISTRIBUTION FUNCTION

$$F(x^*) = \int_{-\infty}^{x^*} \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx.$$

THIS DISTRIBUTION IS DENOTED AS N(μ,σ²) (I.E., IF X HAS A NORMAL DISTRIBUTION WITH PARAMETERS μ AND σ, X~N(μ,σ²)). THERE IS NO CLOSED-FORM EXPRESSION FOR F(x*). THE MEAN OF THE DISTRIBUTION IS μ AND THE VARIANCE IS σ². IF μ=0 AND σ²=1, THE DISTRIBUTION IS CALLED A STANDARD NORMAL DISTRIBUTION. IF x~N(μ,σ²) THEN Z = (X − μ)/σ ~N(0,1).

THE PROBABILITY DENSITY FUNCTION OF A STANDARDIZED NORMAL RANDOM VARIABLE *z* IS DENOTED BY ɸ(z) AND THE CDF IS DENOTED BY Φ(z). TABLES OF ɸ(z) AND Φ(z) ARE INCLUDED IN MOST STATISTICS TEXTS.

LET -∞ <μ_X< ∞, -∞ <μ_Y< ∞, 0 <σ_X, 0 <σ_Y AND -1 < ρ <1 BE FIVE REAL NUMBERS. THE *BIVARIATE NORMAL PDF* WITH MEANS μ_X AND μ_Y, VARIANCES σ_X^2 AND σ_Y^2 AND CORRELATION ρ IS

$$f(x,y) = (2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2})^{-1} exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right.\right.$$
$$\left.\left.- 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right) for -\infty < x$$
$$< \infty\ and -\infty < y < \infty.$$

NOTE THAT IF WE DEFINE THE VECTORS

$$\boldsymbol{x} = (x,y)'$$

AND

$$\boldsymbol{\mu} = (\mu_X, \mu_Y)'$$

AND THE MATRIX

$$\Sigma = \begin{bmatrix} \sigma_{XX} & \sigma_{XY} \\ \sigma_{XY} & \sigma_{YY} \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

WHERE $\sigma_{XY} = \rho\sigma_X\sigma_Y$, THEN WE MAY WRITE THE BIVARIATE NORMAL DENSITY μFUNCTION IN VECTOR/MATRIX NOTATION AS

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-1}|\Sigma|^{-1/2}e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

WHERE

$$|\Sigma| = \sigma_X^2\sigma_Y^2(1 - \rho^2)$$

DENOTES THE DETERMINANT OF THE MATRIX Σ.

THE QUANTITIES μ_X AND μ_Y ARE THE MEANS OF X AND Y; σ_X^2 IS THE VARIANCE OF X; σ_Y^2 IS THE VARIANCE OF Y; AND σ_{XY}^2 IS THE COVARIANCE OF X AND Y.

THE NORMAL DISTRIBUTION MAY BE EXTENDED TO AN ARBITRARY NUMBER OF JOINTLY DISTRIBUTED RVs. FOR MORE THAN TWO RVs, MATRIX NOTATION IS USED.

A VECTOR **X** = (X$_1$,...,X$_k$) HAS A *MULTIVARIATE NORMAL DISTRIBUTION*, DENOTED AS **X** ~ N(**μ**,**Σ**), IF ITS DENSITY FUNCTION IS

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2\pi^{k/2}|\boldsymbol{\Sigma}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}$$

WHERE **μ** IS A VECTOR OF LENGTH k AND **Σ** IS A k x k SYMMETRIC POSITIVE DEFINITE MATRIX. THE VECTOR **μ** IS THE VECTOR OF MEANS, [E(X$_i$)], AND THE MATRIX **Σ** IS THE MATRIX OF COVARIANCES, [E(X$_i$ − E(X$_i$))(X$_j$ − E(X$_j$))]. THAT IS,

$$\boldsymbol{\mu} = E(\boldsymbol{X})$$

AND

$$\Sigma = cov(\boldsymbol{X}, \boldsymbol{X}') = E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})' = \left[E(X_i - \mu_i)(X_j - \mu_j)\right].$$

5. MULTIVARIATE TECHNIQUES WHEN THE POPULATION PARAMETERS ARE KNOWN

THIS SECTION DESCRIBES MAJOR TOPICS IN MULTIVARIATE ANALYSIS, WHEN THE POPULATION DISTRIBUTION IS A KNOWN MULTIVARIATE NORMAL DISTRIBUTION, I.E., THE MEAN $\boldsymbol{\mu}$ AND THE COVARIANCE MATRIX Σ ARE KNOWN.

MANY OF THE PROBAILITY DISTRIBUTIONS THAT ARISE IN NORMAL-BASED MULTIVARIATE ANALYSIS ARE THE SAME ONES (SUCH AS CHI-SQUARED AND FISHER'S F) THAT OCCUR IN UNIVARIATE ANALYSIS. THESE DISTRIBUTIONS HAVE BEEN DISCUSSED IN ANOTHER PRESENTATION, AND THAT DISCUSSION IS NOT REPEATED HERE.

THIS PRESENTATION IS A SURVEY OF MULTIVARIATE ANALYSIS, PRESENTING BASIC INFORMATION IN SUMMARY FORM, AND OMITTING MATHEMATICAL PROOFS. FOR A COMPLETE, DETAILED DISCUSSION OF MULTIVARIATE ANALYSIS, SEE THE BOOK BY ANDERSON (OP. CIT.).

MARGINAL AND CONDITIONAL DISTRIBUTIONS; DISTRIBUTION OF LINEAR COMBINATIONS

SOME OF THE BASIC PROPERTIES OF THE MULTIVARIATE NORMAL DISTRIBUTION ARE THE FOLLOWING:

1. THE MARGINAL DISTRIBUTION OF ONE OR MORE COMPONENTS OF A MULTIVARIATE NORMAL DISTRIBUTION IS ALSO A MULTIVARIATE NORMAL DISTRIBUTION, WITH THE SAME COMPONENT MEANS AND COVARIANCES AS THE FULL DISTRIBUTION.

2. IF **X**~ N(**μ**,Σ), THEN **Y** = C**X**, WHERE C IS A NONSINGULAR MATRIX, HAS DISTRIBUTION N(C**μ**,CΣC'), I.E., Y ~ N(C**μ**,CΣC').
3. SUPPOSE THAT **X**~ N(**μ**,Σ). LET $\Sigma^{1/2}$ DENOTE THE SQUARE ROOT MATRIX OF Σ, AND $\Sigma^{-1/2}$ THE INVERSE OF THE SQUARE ROOT MATRIX. THEN THE DISTRIBUTION OF Y = $\Sigma^{-1/2}$(**X** - **μ**) IS N(0,I).
4. A NECESSARY AND SUFFICIENT CONDITION THAT TWO SUBSETS OF THE RANDOM VARIABLES OF A JOINT MULTIVARIATE VECTOR BE INDEPENDENT IS THAT EACH COVARIANCE OF A VARIABLE FROM ONE SET AND A VARIABLE FROM THE OTHER SET BE ZERO.

REGRESSION AND CORRELATION

LET **X**~ N(**μ**,Σ) AND PARTITION X, HAVING p COMPONENTS, INTO TWO VECTORS HAVING q AND p − q COMPONENTS, RESPECTIVELY:

$$X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix}.$$

LET

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}$$

DENOTE THE CORRESPONDING PARTITION OF THE MEAN VECTOR, AND

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}$$

THE CORRESPONDING PARTITION OF THE COVARIANCE MATRIX.

THEN THE CONDITIONAL DISTRIBUTION OF **X**$^{(1)}$ GIVEN **X**$^{(2)}$ IS A MULTIVARIATE NORMAL DISTRIBUTION f(**x**$^{(1)}$,**x**$^{(2)}$) WITH MEAN

$$\boldsymbol{\mu}^{(1)} + \Sigma_{11}\Sigma_{22}^{-1}(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)})$$

AND COVARIANCE

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

NOTE THAT THE COVARIANCE DOES NOT DEPEND ON MEANS.

THE QUANTITY

$$\boldsymbol{\mu}^{(1)} + \Sigma_{11}\Sigma_{22}^{-1}(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)})$$

(I.E., THE CONDITIONAL MEAN OF $\mathbf{X}^{(1)}$ GIVEN $\mathbf{x}^{(2)}$) IS CALLED THE REGRESSION FUNCTION.

THE q x p-q MATRIX $B = \Sigma_{12}\Sigma_{22}^{-1}$ IS THE MATRIX OF REGRESSION COEFFICIENTS OF $\mathbf{X}^{(1)}$ ON $\mathbf{x}^{(2)}$. (THIS NOTATION IS MISLEADING. THE LETTER B IS INTENDED TO BE AN UPPER-CASE GREEK-LETTER BETA. THAT IS,

$$B = \begin{bmatrix} \boldsymbol{\beta}_1' \\ \vdots \\ \boldsymbol{\beta}_q' \end{bmatrix}$$

WHERE

$$\boldsymbol{\beta}_i' = (\beta_{i1}, \dots, \beta_{p-q}), \quad i = 1, \dots, q.)$$

LET $\sigma_{ij.q+1,\dots,p}$ DENOTE THE i,j-th ELEMENT OF $\Sigma_{11.2}$. THESE ELEMENTS ARE CALLED PARTIAL COVARIANCES. THE QUANTITY

$$\rho_{ij.q+1,\dots,p} = \frac{\sigma_{ij.q+1,\dots,p}}{\sqrt{\sigma_{ii.q+1,\dots,p}}\sqrt{\sigma_{jj.q+1,\dots,p}}}$$

IS CALLED THE *PARTIAL CORRELATION* BETWEEN X_i AND X_j HOLDING $X_{q+1},\dots,X_p$ FIXED.

FOR THE CASE IN WHICH THE MULTIVARIATE DISTRIBUTION CONSISTS OF JUST TWO COMPONENTS, THE PRECEDING FORMULAS REDUCE TO THE FAMILIAR FORMULAS FOR THE REGRESSION OF ONE VARIABLE ON ANOTHER.

MULTIPLE CORRELATION COEFFICIENT

CONTINUING WITH THE PRECEDING DISCUSSION, SUPPOSE THAT A MULTIVARIATE VECTOR **X** IS PARTITIONED AS ABOVE, INTO TWO SUBVECTORS $\mathbf{X}^{(1)}$ AND $\mathbf{X}^{(2)}$, AS BEFORE.

LET X_i DENOTE A COMPONENT (SINGLE RANDOM VARIABLE) OF $\mathbf{X}^{(1)}$. THEN IT CAN BE PROVED THAT THE LINEAR COMBINATION $\alpha \mathbf{X}^{(2)}$ THAT MINIMIZES THE VARIANCE OF $X_i - \alpha \mathbf{X}^{(2)}$ AND THAT MAXIMIZES THE CORRELATION BETWEEN X_i AND $\alpha \mathbf{X}^{(2)}$ IS $\boldsymbol{\beta}'_i \mathbf{X}^{(2)}$.

THE MAXIMUM CORRELATION BETWEEN X_i AND THE LINEAR COMBINATION $\alpha \mathbf{X}^{(2)}$ IS CALLED THE MULTIPLE CORRELATION COEFFICIENT BETWEEN X_i AND $\mathbf{X}^{(2)}$.

DEFINE $\boldsymbol{\sigma}_{(i)}$ AS THE i-th ROW OF Σ_{12}. THEN IT FOLLOWS THAT THE VALUE OF THE MULTIPLE CORRELATION COEFFICIENT IS

$$\bar{R}_{i.q+1,\dots,p} = \frac{\sqrt{\boldsymbol{\sigma}_{(i)}\Sigma_{22}^{-1}\boldsymbol{\sigma}'_{(i)}}}{\sqrt{\sigma_{ii}}},$$

AND THE VALUE OF THE PARTIAL CORRELATION IS GIVEN BY

$$\sigma_{ii.q+1,\dots,p} = \left(1 - \bar{R}^2_{i.q+1,\dots,p}\right)\sigma_{ii}.$$

THE GENERAL LINEAR STATISTICAL MODEL

AS MENTIONED ABOVE, THE CONDITIONAL DISTRIBUTION OF $\mathbf{X}^{(1)}$ GIVEN $\mathbf{X}^{(2)}$ IS A MULTIVARIATE NORMAL DISTRIBUTION $f(\mathbf{x}^{(1)},\mathbf{x}^{(2)})$ WITH MEAN

$$\boldsymbol{\mu}_{1.2}(x^{(2)}) = \boldsymbol{\mu}^{(1)} + \Sigma_{11}\Sigma_{22}^{-1}(\boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)})$$

AND COVARIANCE

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},$$

AND THE CONDITIONAL MEAN OF $\mathbf{X}^{(1)}$ GIVEN $\mathbf{x}^{(2)}$ IS CALLED THE REGRESSION FUNCTION. THAT IS, THE REGRESSION FUNCTION IS

$$E\left(X^{(1)}\big|x^2\right) = \boldsymbol{\mu}^{(1)} + \Sigma_{11}\Sigma_{22}^{-1}\left(x^{(2)} - \boldsymbol{\mu}^{(2)}\right).$$

THE RIGHT-HAND SIDE OF THIS EXPRESSION IS SIMPLY A LINEAR COMBINATION OF $\mathbf{x}^{(2)}$. LET US DENOTE THIS LINEAR COMBINATION AS $\boldsymbol{\beta}_0 + B\mathbf{x}^{(2)}$, WHERE $\boldsymbol{\beta}_0$ IS THE VECTOR OF PARAMETERS

$$\boldsymbol{\beta}_0 = \boldsymbol{\mu}^{(1)} + \Sigma_{11}\Sigma_{22}^{-1}\boldsymbol{\mu}^{(2)}$$

AND B IS A MATRIX OF PARAMETERS

$$B = \Sigma_{11}\Sigma_{22}^{-1}.$$

IN THIS NOTATION, THE REGRESSION FUNCTION BECOMES

$$E\left(X^{(1)}\big|x^2\right) = \boldsymbol{\beta}_0 + \mathbf{B}x^{(2)}.$$

ANY RANDOM VARIABLE, Y, THAT IS CONDITIONAL ON A VARIABLE X MAY BE WRITTEN IN THE FORM

Y = E(Y|X) + U

WHERE E(U|X) = 0. (Y AND X MAY BE SCALARS OR VECTORS OR MATRICES.) THIS FACT IS A DEFINITION OF Y, NOT A PROVED RESULT. IT CAN BE PROVED THAT THE CONDITION E(U|X) = 0 IMPLIES E(U)=0. THIS FACT FOLLOWS FROM THE LAW OF ITERATED EXPECTATIONS:

FOR ANY RANDOM VARIABLES, X, Y AND W, SUPPOSE THAT X IS A FUNCTION OF W, SAY X = f(W). THEN

$$E(Y|X) = E[E(Y|W)|X].$$

(NOTE THAT THIS RESULT IS NOT THE SAME AS THE SIMPLER RESULT CONCERNING EXPECTATIONS, THAT E(Y) = E[E(Y|W)], OR, IN PARTICULAR, E(Y|X) = E[E(Y|X)|W].)

IT CAN ALSO BE PROVED THAT THE CONDITION E(U|X) = 0 IMPLIES THAT U IS UNCORRELATED WITH ANY FUNCTION OF X.

THESE RESULTS IMPLY THAT THE FACT THAT A RANDOM VARIABLE $\mathbf{X}^{(1)}|\mathbf{x}^{(2)}$ HAS MULTIVARIATE NORMAL DISTRIBUTION N($\boldsymbol{\mu}_{1.2}(\mathbf{x})$,$\Sigma_{1.2}$) IMPLIES THAT

$$X^{(1)}|x^{(2)} = \boldsymbol{\mu}_{1.2} + \boldsymbol{U},$$

WHERE $\mathbf{U}$ ~ N($\mathbf{0}$,$\Sigma_{1.2}$). USING THE "BETA" NOTATION JUST INTRODUCED, THIS BECOMES

$$X^{(1)}|x^{(2)} = \boldsymbol{\beta_0} + \mathbf{B}x^{(2)} + \boldsymbol{U}.$$

THIS FORM IS A VERSION OF THE GENERAL LINEAR STATISTICAL MODEL, IN WHICH IT IS ASSUMED THAT THE MODEL ERROR TERM ($\mathbf{U}$) HAS A MULTIVARIATE NORMAL DISTRIBUTION (WITH MEAN $\mathbf{0}$ AND VARIANCE $\Sigma_{1.2}$). IN GENERAL, THE DISTRIBUTIONAL ASSUMPTIONS ABOUT THE MODEL ERROR TERM MAY BE RELAXED (E.G., BY DROPPING THE ASSUMPTION THAT IT IS MULTIVARIATE NORMAL, OR THAT THE VARIANCE FUNCTION IS THE SAME FOR ALL OBSERVATIONS). THE ESSENTIAL REQUIREMENT IS THAT $\mathbf{U}$ BE UNCORRELATED WITH $\mathbf{x}^{(2)}$. (THE DESCRIPTOR "LINEAR" IN THE GENERAL LINEAR STATISTICAL MODEL REFERS TO THE FACT THAT THE EXPECTED VALUE IS A LINEAR COMBINATION OF THE PARAMETER B, NOT OF $\mathbf{x}^{(2)}$.)

PRINCIPAL COMPONENTS

A MAJOR DIFFICULTY ASSOCIATED WITH MULTIVARIATE ANALYSIS IS THE VERY LARGE NUMBER OF PARAMETERS INVOLVED. SUPPOSE THAT WE ARE DEALING WITH A MULTIVARIATE VECTOR OF p COMPONENTS. THEN THE NUMBER OF PARAMETERS REQUIRED TO SPECIFY A p-DIMENSIONAL MULTIVARIATE NORMAL DISTRIBUTION IS p MEANS, p VARIANCES AND $(p^2 - p)/2$ COVARIANCES, FOR A TOTAL OF $(p^2 - p)/2 + 2p$ PARAMETERS.

MULTIVARIATE PROBLEMS MAY INVOLVE MANY COMPONENTS. FOR EXAMPLE, A PSYCHOLGICAL QUESTIONNAIRE MAY INVOLVE 100 QUESTIONS. A GENERAL REPRESENTATION OF A 100-COMPONENT MULTIVARIATE NORMAL DISTRIBUTION

INVOLVES $(100^2 - 100)/2 + 2(100) = 5{,}150$ PARAMETERS. THIS NUMBER OF PARAMETERS IS VASTLY TOO LARGE FOR A REASONABLE MODEL. IN ORDER TO DEVELOP A REASONABLE MODEL (USEFUL MODEL OF REALITY), IT IS DESIRABLE TO REDUCE THE NUMBER OF PARAMETERS.

THE NUMBER OF PARAMETERS CAN BE REDUCED IN A NUMBER OF WAYS. THE DISTRIBUTION MAY BE SIMPLIFIED EITHER BY REDUCING THE NUMBER OF COMPONENTS OF THE MULTIVARIATE VECTOR, OR BY REDUCING THE NUMBER OF NONZERO COVARIANCES. PRIOR KNOWLEDGE OR THEORY MAY SUGGEST THAT SOME VARIABLES ARE UNCORRELATED, IN WHICH CASE THE COVARIANCES ASSOCIATED WITH THEM ARE EQUAL TO ZERO. BY EXAMINING THE CORRELATIONS AMONG VARIABLES, IT MAY BE POSSIBLE TO DISCARD SOME VARIABLES, OR TO COMBINE SIMILAR VARIABLES INTO COMPOSITE SCORES. ANOTHER METHOD IS TO TRANSFORM THE DATA IN SUCH A WAY THAT THE NUMBER OF COVARIANCES IS REDUCED.

THIS SECTION DESCRIBES A STANDARD METHODOLOGY FOR SIMPLIFYING A MULTIVARIATE DISTRIBUTION, VIZ., THE METHOD OF PRINCIPAL COMPONENTS. THE METHOD OF PRINCIPAL COMPONENTS TRANSFORMS THE ORIGINAL MULTIVARIATE VECTOR SUCH THAT THE COVARIANCES BETWEEN DIFFERENT COMPONENTS OF THE TRANSFORMED VECTOR ARE ZERO.

A MULTIVARIATE PROBABILITY DISTRIBUTION MAY BE VISUALIZED AS A MULTIDIMENSIONAL ELLIPSOID, WHERE THE ELLIPSOID IS DEFINED BY SURFACES HAVING EQUAL PROBABILITY DENSITY. IF THE COMPONENT VARIABLES ARE UNCORRELATED, THEN THE PRINCIPAL AXES OF THE ELLIPSOID ARE PARALLEL TO THE COMPONENT AXES. IF THE COMPONENT VARIABLES ARE CORRELATED, THEN THE PRINCIPAL AXES OF THE ELLIPSOID ARE OBLIQUE TO THE COMPONENT AXES.

A DISTRIBUTION OF THE LATTER TYPE (PRINCIPAL AXES OBLIQUE TO THE COMPONENT AXES) MAY BE TRANSFORMED TO THE FORMER TYPE (PRINCIPAL AXES PARALLEL TO THE COMPONENT AXES) BY A ROTATION TRANSFORMATION. A ROTATION TRANSFORMATION IS ACCOMPLISHED SIMPLY BY MULTIPLYING THE ORIGINAL (UNTRANSFORMED) MULTIVARIATE VECTOR BY A SUITABLE MATRIX. SUCH A PROCEDURE WAS DISCUSSED EARLIER IN THE SECTION DEALING WITH CONVERTING A MATRIX TO JORDAN CANONICAL FORM, OR WITH APPLYING A CHOLESKI TRANFORMATION. THE TRANSFORMED VARIABLE HAS p MEANS AND p

VARIANCES, BUT ALL OF THE COVARIANCES ARE ZERO. THIS TRANSFORMATION RESULTS IN A TREMENDOUS REDUCTION IN THE NUMBER OF PARAMETERS.

ALTHOUGH THE PRINCIPAL COMPONENTS TRANFORMATION REDUCES THE NUMBER OF PARAMETERS INVOLVED, IT MAY OR MAY NOT BE A REASONABLE APPROACH TO SIMPLIFYING A MULTIVARIATE PROBLEM. IT IS BEST SUITED TO SITUATIONS WHERE THE COMPONENT VARIABLES ARE SIMILAR IN SUBSTANTIVE NATURE, SUCH AS A SET OF PSYCHOLOGICAL TEST SCORES. OTHERWISE, THE COMPONENTS OF THE TRANSFORMED VECTOR MAY BE DIFFICULT TO INTERPRET.

THE METHOD OF PRINCIPAL COMPONENTS FOCUSES ON DESCRIBING A SYSTEM SUCCINCTLY IN TERMS OF VARIANCES OF TRANSFORMED VARIABLES. IT ACCOMPLISHES THIS OBJECTIVE AT THE COST OF ELIMINATING THE INTER-CORRELATIONAL RELATIONSHIPS AMONG THE VARIABLES. THE NEXT SECTION WILL DISCUSS A PROCEDURE (FACTOR ANALYSIS) FOR SIMPLIFYING A MULTIVARIATE PROBLEM IN A WAY THAT FOCUSES ON THE INTER-RELATIONSHIPS AMONG THE ORIGINAL COMPONENT VARIABLES.

IT CAN BE PROVED THAT THE PRINCIPAL COMPONENTS ARE THE CHARACTERISTIC VECTORS OF THE COVARIANCE MATRIX. (SEE ANDERSON OP. CIT. FOR THE PROOF, WHICH INVOLVES LAGRANGIAN OPTIMIZATION OF MATRIX EXPRESSIONS.)

CANONICAL CORRELATIONS AND CANONICAL VARIABLES

IN A PRECEDING SECTION WE PRESENTED THE RESULT THAT THE MULTIPLE CORRELATION COEFFICIENT WAS THE MAXIMUM CORRELATION BETWEEN A VARIABLE X_i AND A LINEAR COMBINATION OF A SET OF COMPONENTS, $\mathbf{X}^{(2)}$. THIS CONCEPT MAY BE EXTENDED. SUPPOSE THAT THE p-COMPONENT VECTOR $\mathbf{X}$ IS PARTITIONED INTO TWO PARTS, $\mathbf{X}^{(1)}$ AND $\mathbf{X}^{(2)}$, CONSISTING OF p_1 AND p_2 COMPONENTS, RESPECTIVELY, WHERE $p_1 + p_2 = n$. (IN THE NOTATION OF THE PREVIOUS SECTION, p_1 WAS DENOTED AS q AND p_2 AS p – q.)

IN THE TOPIC OF CANONICAL CORRELATIONS AND CANONICAL VARIABLES, THE PROBLEM IS TO DETERMINE LINEAR COMBINATIONS OF EACH PARTITION SUCH THAT:

1. THE FIRST LINEAR COMBINATION IN EACH PARTITION HAS MAXIMUM CORRELATION (OF THE SET OF ALL POSSIBLE LINEAR COMBINATIONS);
2. THE SECOND LINEAR COMBINATION IN EACH PARTITION HAS MAXIMUM CORRELATION, OF ALL LINEAR COMBINATIONS THAT ARE UNCORRELATED WITH THE FIRST LINEAR COMBINATIONS;
3. THE i-th SET OF LINEAR COMBINATIONS (ONE FROM EACH PARTITION) HAS MAXIMUM CORRELATION, OF ALL LINEAR COMBINATIONS THAT ARE UNCORRELATED WITH ALL PREVIOUS SETS OF LINEAR COMBINATIONS.

SAID ANOTHER WAY, THE r-th PAIR OF CANONICAL VARIATES ARE THE PAIR OF LINEAR COMBINATIONS $\alpha(r)'X(1)$ and $\gamma(r)'X(2)$, EACH OF UNIT VARIANCE AND UNCORRELATED WITH THE FIRST $r - 1$ PAIRS OF CANONICAL VARIATES AND HAVING MAXIMUM CORRELATION. THE CORRELATION IS THE CANONICAL CORRELATION.

IT CAN BE SHOWN THAT THE r-th CANONICAL CORRELATION IS THE r-th LARGEST ROOT OF THE MATRIX

$$\begin{bmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{bmatrix}.$$

IN THE SPECIAL CASE OF $p_1 = 1$, THE SINGLE CANONICAL CORRELATION IS THE MULTIPLE CORRELATION BETWEEN $\mathbf{X}^{(1)} = X_1$ AND $\mathbf{X}^{(2)}$.

SEE ANDERSON OP. CIT. FOR DERIVATION OF THESE RESULTS.

CLASSIFICATION ANALYSIS

CLASSIFICATION ANALYSIS IS THE PROBLEM OF DECIDING TO WHAT POPULATION (CLASS, GROUP, CATEGORY) AN INDIVIDUAL BELONGS, BASED ON MEASUREMENTS TAKEN ON THE INDIVIDUAL. IN THIS PRESENTATION, THIS PROBLEM WILL BE ADDRESSED FOR THE CASE IN WHICH THE MEASUREMENTS HAVE A MULTIVARIATE NORMAL DISTRIBUTION.

THE PROBLEM WILL BE ADDRESSED BY FIRST CONSIDERING THE CASE IN WHICH THERE EXIST JUST TWO POPULATIONS OF INTEREST. WE SHALL USE THE INDEX

VALUE "1" TO REFER TO THE FIRST POPULATION AND THE VALUE "2" TO REFER TO THE SECOND POPULATION.

THIS PROBLEM IS A TWO-ALTERNATIVE STATISTICAL DECISION PROBLEM. GIVEN A SAMPLE OBSERVATION, THE STATISTICIAN (DECISIONMAKER) MUST DECIDE WHETHER THE INDIVIDUAL BELONGS TO POPULATION 1 OR POPULATION 2.

THE PROBLEM MAY BE REPRESENTED BY A 2 x 2 MATRIX, WHERE THE TWO POPULATIONS ARE REPRESENTED BY THE TWO ROWS, AND THE TWO POSSIBLE DECISIONS ARE REPRESENTED BY THE TWO COLUMNS. THE ENTRIES IN THE TABLE INDICATE WHETHER A CORRECT DECISION IS MADE.

		Decision	
		1	2
Population	1	Correct decision	Error
	2	Error	Correct decision

THIS PROBLEM MAY BE CONSIDERED TO BE A TEST OF THE STATISTICAL HYPOTHESIS THAT THE TRUE POPULATION IS POPULATION 1. IN THAT CASE, THE ERROR OF DECIDING POPULATION 2 WHEN THE INDIVIDUAL IS FROM POPULATION 1 IS CALLED A TYPE 1 ERROR, AND THE ERROR OF DECIDING POPULATION 1 WHEN THE INDIVIDUAL IS FROM POPULATION 2 IS CALLED A TYPE 2 ERROR. THIS FRAMEWORK IS OFTEN ADOPTED IN THE APPLICATION OF MEDICAL DIAGNOSIS, WHERE POPULATION 1 IS THE POPULATION OF INDIVIDUALS HAVING AN ILLNESS OR CONDITION OF INTEREST, AND POPULATION 2 IS THE POPULATION OF INDIVIDUALS NOT HAVE THE ILLNESS OR CONDITION.

THE DECISION ABOUT WHICH POPULATION AN OBSERVED INDIVIDUAL BELONGS TO IS MADE ACCORDING TO A DECISION CRITERION, TAKING INTO ACCOUNT THE OBSERVED MEASUREMENTS. A SPECIFIED DECISION CRITERION WILL BE ASSOCIATED WITH VALUES FOR THE PROBABILITIES OF MAKING THE TWO TYPES OF ERROR. TYPICALLY, IF A DECISION CRITERION HAS A HIGHER PROBABILITY OF ONE TYPE OF ERROR, IT WILL HAVE A LOWER PROBABILITY OF THE OTHER TYPE OF ERROR.

IF THIS (HYPOTHESI-TESTING) FRAMEWORK IS ADOPTED, A STANDARD METHODOLOGY FOR SUMMARIZING THE PERFORMANCE OF ALTERNATIVE

DECISION CRITERIA IS TO CONSTRUCT A GRAPH SHOWING 1 − Pr(Type 2 error) VERSUS Pr(Type 1 error) FOR MEMBERS OF THE SET OF ALTERNATIVE DECISION CRITERIA. SUCH A GRAPH IS CALLED A RECEIVER OPERATING CHARACTERISTIC GRAPH (OR CURVE), OR ROC GRAPH (OR CURVE). A PREFERRED DECISION CRITERIA IS SELECTED CORRESPONDING TO A POINT ON THE ROC CURVE.

(A DETAILED DISCUSSION OF THE USE OF ROC CURVES IS PRESENTED IN *SIGNAL DETECTION THEORY AND PSYCHOPHYSICS* BY DAVID M. GREEN AND JOHN A. SWETS (WILEY, 1966) OR IN *EVALUATION OF DIAGNOSTIC SYSTEMS: METHODS FROM SIGNAL DETECTION THEORY* BY JOHN A. SWETS AND RONALD M. PICKETT (ACADEMIC PRESS, 1982).)

FOR THIS PRESENATION WILL ADOPT A DIFFERENT FRAMEWORK, IN WHICH A COST IS ASSIGNED TO EACH OF THE TWO TYPES OF ERROR, AND THE OBJECTIVE IS TO IDENTIFY A DECISION CRITERION THAT MINIMIZES THE EXPECTED COST. (THIS IS THE APPROACH DESCIRBED IN ANDERSON OP. CIT. THIS APPROACH IS MORE COMMON IN STATISTICAL DECISION THEORY THAN THE ROC APPROACH. IT ALSO GENERALIZES MORE EASILY TO THE CASE IN WHICH THERE ARE MORE THAN TWO POPULATIONS OF INTEREST. THE ROC APPROACH IS APPROPRIATE IF THERE ARE JUST TWO POPULATIONS (E.G., PRESENCE OR ABSENCE OF A CONDITION) AND THE COSTS OF MAKING THE TWO TYPES OF ERRORS ARE NOT COMMENSURABLE (E.G., THE COST OF A FALSE POSITIVE MAY BE A MINOR EXPENSE OR INCONVENIENCE, WHEREAS THE COST OF A FALSE NEGATIVE MAY BE DEATH).

THE METHODOLOGY PRESENTED HERE USES THE SAME NOTATION AS ANDERSON.

THE SOLUTION TO THE CLASSIFICATION PROBLEM INVOLVES TWO PROBABILITY DISTRIBUTIONS: THE A PRIORI PROBABILITY THAT AN OBSERVATION IS DRAWN FROM A PARTICULAR POPULATION, AND THE CONDITIONAL PROBABILITIY DISTRIBUTION OF THE OBSERVED MEASUREMENTS, GIVEN THE POPULATION.

A PRIORI PROBABILITIES OF POPULATION MEMBERSHIP: LET q_1 DENOTE THE PROBABILITY OF DRAWING AN OBSERVATION FROM POPULATION 1 AND q_2 DENOTE THE PROBABILITY OF DRAWING AN OBSERVATION FROM POPULATION 2.

CONDITIONAL SAMPLING DISTRIBUTION: LET $p_i(\mathbf{x})$ DENOTE THE PROBABILITY DENSITY OF OBSERVATION $\mathbf{x}$, GIVEN THAT IT IS DRAWN FROM POPULATION i.

COSTS OF MISSCLASSIFICATION: LET C(2|1) DENOTE THE COST ASSOCIATED WITH THE ERROR OF DECIDING POPULATION 2 WHEN THE OBSERVATION IS FROM POPULATION 1, AND C(1|2) DENOTE THE COST ASSOCIATED WITH THE ERROR OF DECIDING POPULATION 1 WHEN THE OBSERVATION IS FROM POPULATION 2.

LET R_i DENOTE THE REGION OF CLASSIFICATION FOR POPULATION i, THAT IS, THE DECISION IS MADE THAT THE OBSERVATION IS FROM POPULATION i IF THE OBSERVATION x IS IN REGION R_i.

THEN IT CAN BE PROVED THAT THE REGIONS R_i THAT CORRESPOND TO MINIMUM EXPECTED COST ARE DEFINED BY

$$R_1 : \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} \geq \frac{C(1|2)q_2}{C(2|1)q_1}$$

$$R_2 : \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} < \frac{C(1|2)q_2}{C(2|1)q_1}.$$

THIS CRITERION MAY BE RECOGNIZED AS THE USUAL LIKELIHOOD-RATIO TEST OF THE HYPOTHESIS THAT THE DISTRIBUTION OF THE OBSERVED VALUE $\mathbf{x}$ IS $p_1(\mathbf{x})$ VERSUS THE ALTERNATIVE THAT IT IS $p_2(\mathbf{x})$.

THE PRECEDING RESULT APPLIES TO ANY PROBABILITY DISTRIBUTIONS. IN THE CASE IN WHICH THE TWO DISTRIBUTIONS ARE MULTIVARIATE NORMAL WITH DIFFERENT MEAN VECTORS BUT THE SAME COVARIANCE MATRIX, THE DENSITY FUNCTION IS

$$p_i(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{1}{2}p}|\Sigma|^{\frac{1}{2}}} exp\left[-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}^{(i)})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}^{(i)})\right],$$

AND THE RATIO OF THE DENSITIES BECOMES

$$\frac{p_1(x)}{p_2(x)} = \frac{exp\left[-\frac{1}{2}\left(x - \mu^{(1)}\right)'\Sigma^{-1}\left(x - \mu^{(1)}\right)\right]}{exp\left[-\frac{1}{2}\left(x - \mu^{(2)}\right)'\Sigma^{-1}\left(x - \mu^{(2)}\right)\right]}$$

$$= \exp\left\{-\frac{1}{2}\left[\left(x - \mu^{(1)}\right)'\Sigma^{-1}\left(x - \mu^{(1)}\right) - \left(x - \mu^{(2)}\right)'\Sigma^{-1}\left(x - \mu^{(2)}\right)\right]\right\}.$$

TAKING LOGARITHMS OF THE DECISION CRITERION, IT BECOMES:

$$R_1: -\frac{1}{2}\left[\left(x - \mu^{(1)}\right)'\Sigma^{-1}\left(x - \mu^{(1)}\right) - \left(x - \mu^{(2)}\right)'\Sigma^{-1}\left(x - \mu^{(2)}\right)\right] \geq \log k$$

$$R_2: -\frac{1}{2}\left[\left(x - \mu^{(1)}\right)'\Sigma^{-1}\left(x - \mu^{(1)}\right) - \left(x - \mu^{(2)}\right)'\Sigma^{-1}\left(x - \mu^{(2)}\right)\right] < \log k$$

WHERE

$$k = \frac{q_2 C(1|2)}{q_1 C(2|1)}.$$

THE TERMS IN THE LEFT-HAND-SIDE OF THE INEQUALITY DEFINING THE REGIONS MAY BE REARRANGED TO PRODUCE:

$$R_1: x'\Sigma^{-1}\left(\mu^{(1)} - \mu^{(2)}\right) - \frac{1}{2}\left(\mu^{(1)} + \mu^{(2)}\right)'\Sigma^{-1}\left(\mu^{(1)} - \mu^{(2)}\right) \geq \log k$$

$$R_2: x'\Sigma^{-1}\left(\mu^{(1)} - \mu^{(2)}\right) - \frac{1}{2}\left(\mu^{(1)} + \mu^{(2)}\right)'\Sigma^{-1}\left(\mu^{(1)} - \mu^{(2)}\right) < \log k.$$

THE FIRST TERM OF THE LEFT-HAND SIDE OF THE INEQUALITY,

$$x'\Sigma^{-1}\left(\mu^{(1)} - \mu^{(2)}\right),$$

IS CALLED THE *DISCRIMINANT FUNCTION*. IT IS A LINEAR FUNCTION OF THE COMPONENTS OF THE OBSERVATION VECTOR. THE SECOND TERM,

$$-\frac{1}{2}\left(\mu^{(1)} + \mu^{(2)}\right)'\Sigma^{-1}\left(\mu^{(1)} - \mu^{(2)}\right),$$

IS SIMPLY A CONSTANT (DEPENDENT ON THE DISTRIBUTION PARAMETERS).

THE PRECEDING DISCUSSION HAS DESCRIBED THE TOPIC OF CLASSIFICATION ANALYSIS IN THE CASE OF TWO POPULATIONS. EXTENSION TO THE CASE OF MORE THAN TWO POPULATIONS IS STRAIGHTFORWARD.

THE FOLLOWING RESULT IS QUOTED FROM ANDERSON:

IF qi is the a priori probability of drawing an observation from population π_i = $N(\mu^{(i)},\Sigma)$ (i=1,…,m) and if the costs of misclassification are equal, then the regions of classification $R_1,…,R_m$ that minimize the expected cost are defined by

$$R_j : u_{jk}(\boldsymbol{x}) > log\frac{q_k}{q_j}, k = 1, …, m; k \neq j,$$

and

$$u_{jk}(\boldsymbol{x}) = log\frac{p_j(\boldsymbol{x})}{p_k(\boldsymbol{x})} = \left[\boldsymbol{x} - \tfrac{1}{2}\big(\boldsymbol{\mu}^{(j)} + \boldsymbol{\mu}^{(k)}\big)'\Sigma^{-1}\big(\boldsymbol{\mu}^{(j)} - \boldsymbol{\mu}^{(k)}\big)\right].$$

TIME SERIES

A TIME SERIES IS A SET OF OBSERVATIONS TAKEN AT VARIOUS POINTS IN TIME (OR IN SPACE). IN MANY APPLICATIONS OF INTEREST, NEARBY OBSERVATIONS ARE CORRELATED. IN ORDER TO UNDERSTAND THE STOCHASTIC NATURE OF A SYSTEM, AND TO CONDUCT TESTS OF HYPOTHESES, IT IS HELPFUL TO TRANSFORM THE OBSERVED VARIABLES FROM ONES THAT ARE CORRELATED TO ONES THAT ARE UNCORRELATED (AND CONDUCT THE TESTS ON THE TRANSFORMED VARIABLES). SUCH A TRANSFORMATION CORRESPONDS TO A MATHEMATICAL (STATISTICAL) MODEL OF THE STOCHASTIC PROCESS THAT GENERATES THE OBSERVED DATA. THE DESIRE IS TO IDENTIFY A MODEL THAT HAS A RELATIVELY SMALL NUMBER OF PARAMETERS.

TIME SERIES MAY BE CLASSIFIED IN A NUMBER OF WAYS. ONE CLASSIFICATION IS WHETHER THE OBSERVATIONS ARE UNIVARIATE OR MULTIVARIATE. FOR UNIVARIATE OBSERVATIONS, THE CORRELATION OF INTEREST IS THE CORRELATION BETWEEN NEARBY OBSERVATIONS. FOR MULTIVARIATE OBSERVATIONS, TWO TYPES OF CORRELATIONS MAY OCCUR: CORRELATION BETWEEN NEARBY OBSERVATIONS, AND CORRELATIONS AMONG THE COMPONENT VARIABLES OF THE MULTIVARIATE OBSERVATION VECTOR.

SINCE THE 1960s MUCH WORK HAS BEEN DONE IN THE FIELD OF TIME SERIES ANALYSIS. MOST OF THE METHODOLOGY IN USE TODAY STEMS FROM THE WORK OF G. E. P. BOX AND GWILYM JENKINS, WHO SHOWED IN DETAIL HOW A PARTICULAR CLASS OF TIME-SERIES MODELS – THE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS – MAY BE USED TO DESCRIBE THE STOCHASTIC BEHAVIOR OF A WIDE RANGE OF TIME-VARYING PHENOMENA. THESE MODELS ARE GENERALLY KNOWN AS BOX-JENKINS MODELS.

BOX-JENKINS MODELS MAY BE UNIVARIATE OR MULTIVARIATE. THE CLASSIC TEXT ON BOX-JENKINS MODELS IS

Box, George E. P., Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, *Time Series Analysis, Forecasting and Control*, 5th ed., Wiley, 2016

THERE ARE A LARGE NUMBER OF TEXTS ON UNIVARIATE BOX-JENKINS MODELS. A RECENT ONE IS

Cryer, Jonathan D. and Kung-Sik Chan, *Time Series Analysis with Applications in R*, Springer, 2008.

MAJOR TEXTS ON MULTIVARIATE TIME SERIES ANALYSIS ARE:

Tsay, Ruey S., *Multivariate Time Series Analysis with R and Financial Applications*, Wiley, 2014.

Lütkepohl, Helmut, *New Introduction to Multiple Time Series Analysis*, Springer, 2006.

Hamilton, James D., *Time Series Analysis*, Princeton University Press, 1994.

IN THE UNIVARIATE CASE, A BOX-JENKINS MODEL MAY BE REPRESENTED AS

$$z_t = \phi_0 + \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q},$$

WHERE z_t DENOTES AN OBSERVED VALUE AT TIME t, a_t IS A MODEL ERROR TERM AT TIME t, AND THE ϕs AND θs ARE MODEL PARAMETERS. IT IS ASSUMED THAT

THE a_t's HAVE ZERO MEAN AND CONSTANT VARIANCE σ^2, AND ARE UNCORRELATED. (THE ASSUMPTION OF CONSTANT VARIANCE MAY BE RELAXED.)

IN THE PRECEDING EQUATION, THE "ϕ" TERMS (OF INDEX 1 OR GREATER) ARE REFERRED TO AS THE AUTOREGRESSIVE PART OF THE MODEL, AND THE "θ" TERMS ARE REFERRED TO AS THE MOVING AVERAGE PART OF THE MODEL.

IN THE MULTIVARIATE CASE, THE SCALARS z_t AND a_t ARE REPLACED BY A MULTIVARIATE VECTORS AND THE ϕs AND θs ARE REPLACED BY MATRICES.

THIS PRESENTATION PRESENTS ONLY SUMMARY INFORMATION ABOUT TIME SERIES. THE TOPIC OF TIME SERIES ANALYSIS IS ADDRESSED IN DETAIL IN ANOTHER PRESENTATION. THE GOAL IN TIME SERIES ANALYSIS IS TO DETERMINE THE VALUES OF THE MODEL PARAMETERS, AND TO USE THE ESTIMATED MODEL TO FORECAST AND/OR CONTROL A PROCESS.

THIS SECTION HAS PRESENTED BASIC INFORMATION ABOUT THE NATURE OF A MULTIVARIATE NORMAL DISTRIBUTION, ASSUMING THAT ALL OF THE DISTRIBUTION PARAMETERS (μs and Σs) ARE KNOWN. THE NEXT SECTION ADDRESSES THE ISSUE OF ESTIMATION OF DISTRIBUTION PARAMETERS AND MAKING TESTS OF HYPOTHESIS (OR DECISIONS) ABOUT THEM.)

AS FOR THIS SECTION, THE PRIMARY REFERENCE TEXT IS ANDERSON OP. CIT.

WITH THE UBIQUITY OF GENERAL-PURPOSE STATISTICAL SOFTWARE PACKAGES, MUCH MULTIVARIATE STATISTICAL ANALYSIS IS NOW PERFORMED WITH SUPERFICIAL KNOWLEDGE OF DETAILED DISTRIBUTION THEORY AND ESTIMATION PROCEDURES. ALSO, FOR APPLICATIONS IN WHICH THE DISTRIBUTION THEORY BECOMES COMPLICATED, AS IT DOES IN MULTIVARIATE ANALYSIS, MUCH ANALYSIS IS DONE USING RESAMPLING PROCEDURES ("BOOTSTRAP" METHODS); IN SUCH CASES IT IS NOT NECESSARY TO KNOW THE EXACT FORMULA FOR A SAMPLING DISTRIBUTION. FOR THESE REASONS, THE DISCUSSION THAT FOLLOWS IS SURVEY IN NATURE, WITH MENTION OF MAJOR RESULTS AND SPECIAL CASES BUT LIMITED DISCUSSION OF DETAILS. A PRIMARY CONSIDERATION IN THE LEVEL OF DETAIL OF THE PRESENTATION IS TO PROVIDE SUFFICIENT BACKGROUND TO HAVE A HIGH-LEVEL UNDERSTANDING OF

STANDARD MULTIVARIATE PROCEDURES EMPLOYED BY MAJOR STATISTICAL SOFTWARE PACKAGES.

6. ESTIMATION AND HYPOTHESIS TESTING

ESTIMATION OF THE MEAN VECTOR AND COVARIANCE MATRIX

THE FIRST PROBLEM IN NORMAL-DISTRIBUTION MULTIVARIATE ANALYSIS IS ESTIMATION OF THE PARAMETERS THAT DEFINE THE DISTRIBUTION, VIZ., THE MEAN VECTOR AND THE COVARIANCE MATRIX.

SINCE THE PARAMETRIC FORM OF THE PROBABILITY DISTRIBUTION IS SPECIFIED (I.E., IS MULTIVARIATE NORMAL), THE STANDARD APPROACH TO PARAMETER ESTIMATION IS THE METHOD OF MAXIMUM LIKELIHOOD. THAT IS, THE PARAMETER ESTIMATES ARE THOSE PARAMETER VALUES THAT MAXIMIZE THE LIKELIHOOD FUNCTION. THIS PRESENTATION WILL NOT SHOW DETAILS OF THIS PROCEDURE (WHICH ARE PRESENTED IN DETAIL IN ANDERSON).

LET $x_1,...,x_N$ BE A SAMPLE OF N OBSERVATIONS FROM A MULTIVARIATE NORMAL DISTRIBUTION, $N(\mu,\Sigma)$, WHERE IT IS ASSUMED THAT THE NUMBER OF COMPONENTS OF x_i IS p, AND p < N. THEN THE MAXIMUM LIKELIHOOD ESTIMATES OF μ AND Σ ARE

$$\hat{\mu} = \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

AND

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})'.$$

THE MAXIMUM LIKELIHOOD ESTIMATE OF THE MEAN IS UNBIASED, I.E.,

$$E(\hat{\mu}) = \mu.$$

THE MAXIMUM LIKELIHOOD ESTIMATE OF THE COVARIANCE MATRIX IS NOT UNBIASED. ITS EXPECTED VALUE IS

$$E\left(\widehat{\Sigma}\right) = \frac{N-1}{N}\Sigma.$$

HENCE THE QUANTITY

$$S = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})'.$$

IS AN UNBIASED ESTIMATE OF Σ. S IS CALLED THE *SAMPLE COVARIANCE MATRIX*.
IT IS OFTEN DENOTED AS $\widehat{\Sigma}$.

THE SAMPLING DISTRIBUTION OF $\widehat{\boldsymbol{\mu}}$.

THE DISTRIBUTION OF $\widehat{\boldsymbol{\mu}}$ IS N($\boldsymbol{\mu}$,(1/N)Σ). HENCE THE DISTRIBUTION OF $\sqrt{N}(\bar{X} - \boldsymbol{\mu})$
IS N(**0**, Σ). THE DISTRIBUTION OF

$$N(\bar{X} - \boldsymbol{\mu})'\Sigma^{-1}(\bar{X} - \boldsymbol{\mu})$$

IS A χ^2 (CHI-SQUARED) DISTRIBUTION WITH p DEGREES OF FREEDOM.

HENCE, IF Σ IS KNOWN, THE SAMPLING DISTRIBUTION OF $N(\bar{X} - \boldsymbol{\mu})'\Sigma^{-1}(\bar{X} - \boldsymbol{\mu})$
IS ALSO KNOWN, AND WE CAN THEN CONSTRUCT CONFIDENCE REGIONS FOR **μ**
IN THE USUAL WAY.

IF Σ IS NOT KNOWN, THEN IT CAN BE SHOWN THAT THE DISTRIBUTION OF

$$\frac{T^2}{N-1}\frac{N-p}{p}$$

WHERE

$$T^2 = N(\bar{X} - \boldsymbol{\mu_0})'S^{-1}(\bar{X} - \boldsymbol{\mu_0})$$

IS A NONCENTRAL F DISTRIBUTION WITH p AND N − p DEGREES OF FREEDOM
AND NONCENTRALITY PARAMETER $N(\bar{X} - \boldsymbol{\mu_0})'^{\Sigma^{-1}}(\bar{X} - \boldsymbol{\mu_0})$. IF $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, THEN THE
F DISTRIBUTION IS CENTRAL. THESE RESULTS MAY BE USED TO CONSTRUCT
CONFIDENCE REGIONS FOR **μ** IN THE USUAL WAY.

THE PRECEDING RESULTS ARE ANALOGUES OF THE CORRESPONDING UNIVARIATE RESULTS (z AND t STATISTICS, TESTS, AND CONFIDENCE INTERVALS).

THE SAMPLING DISTRIBUTION OF $\hat{\Sigma}$.

AS MENTIONED ABOVE, THE SAMPLE COVARIANCE MATRIX IS

$$\hat{\Sigma} = S = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'.$$

IN THE UNIVARIATE CASE, THIS QUANTITY HAS A χ^2 DISTRIBUTION. IN THE MULTIVARIATE CASE, THIS QUANTITY HAS A GENERALIZATION OF THE χ^2 DISTRIBUTION. SPECIFICALLY, THE DENSITY OF

$$A = (N-1)S = \sum_{i=1}^{N}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'$$

IS

$$\frac{|A|^{-\frac{1}{2}(n-p-1)}\exp\left(-\frac{1}{2}tr\Sigma^{-1}A\right)}{2^{\frac{1}{2}np}\pi^{p(p-1)/4}|\Sigma|^{\frac{1}{2}n}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}$$

FOR A POSITIVE DEFINITE AND 0 OTHERWISE. THIS DENSITY FUNCTION IS DENOTED BY w(A|Σ,n) AND THE CORRESPONDING DISTRIBUTION FUNCTION IS DENOTED BY W(Σ,n). THIS DISTRIBUTION IS CALLED THE WISHART DISTRIBUTION WITH COVARIANCE MATRIX Σ AND n DEGREES OF FREEDOM. THE DISTRIBUTION OF S IS W((1/(N-1))Σ,N-1).

THE WISHART DISTRIBUTION ARISES OFTEN IN MULTIVARIATE ANALYSIS. IN MULTIVARIATE ANALYSIS OF VARIANCE (TO BE DISCUSSED LATER), IT IS THE DISTRIBUTION OF THE COMPONENTS OF A PARTITIONED SUM OF SQUARES.

THE DETERMINANT OF Σ, |Σ| IS CALLED THE *GENERALIZED VARIANCE*, AND THE DETERMINANT OF S, |S| IS CALLED THE SAMPLE GENERALIZED VARIANCE. AS DISCUSSED EARLIER, THE DETERMINANT OF A SYMMETRIC MATRIX IS A MEASURE OF VOLUME OF THE PARALLELOTROPE (HYPER-PARALLELOGRAM) SPANNED BY

THE EIGENVECTORS OF THE MATRIX. ANDERSON DISCUSSES THE GENERALIZED VARIANCE AND ITS DISTRIBUTION.

<u>SAMPLING DISTRIBUTIONS OF SAMPLE CORRELATION COEFFICIENTS, PARTIAL CORRELATION COEFFICIENTS AND MULTIPLE CORRELATION COEFFICIENTS</u>

ANDERSON DESCRIBES ESTIMATION AND THE SAMPLING DISTRIBUTIONS OF SAMPLE CORRELATION COEFFICIENTS, PARTIAL CORRELATION COEFFICIENTS, AND MULTIPLE CORRELATION COEFFICIENTS. THOSE RESULTS ARE NOT DISCUSSED HERE.

TESTS OF HYPOTHESES ABOUT ONE OR TWO MEAN VECTORS

FOR A SINGLE SAMPLE, THE TEST OF WHETHER THE POPULATION MEAN EQUALS A SPECIFIC VALUE IS STRAIGHTFORWARD. IF THE COVARIANCE MATRIX, Σ, IS KNOWN, THEN THE SAMPLING DISTRIBUTION OF THE SAMPLE MEAN IS N($\boldsymbol{\mu}$,Σ/n), AND THIS DISTRIBUTION MAY BE USED AS THE BASIS FOR THE TEST AND FOR CONSTRUCTION OF CONFIDENCE REGIONS. IF Σ IS UNKNOWN AND ESTIMATED BY S, THEN THE DISTRIBUTION OF T^2 IS F WITH p AND n-p+1 DEGREES OF FREEDOM, AND THIS DISTRIBUTION MAY BE USED AS THE BASIS FOR THE TEST AND FOR CONSTRUCTION OF CONFIDENCE REGIONS.

FOR TWO SAMPLES (I.E., TESTING THE HYPOTHESIS OF EQUIVALENCE OF TWO POPULATION MEANS), THE SITUATION IS MORE COMPLICATED. LET $\boldsymbol{\mu}^{(1)}$ AND $\boldsymbol{\mu}^{(2)}$ DENOTE THE TWO POPULATION MEANS, AND Σ_1 AND Σ_2 DENOTE THE TWO POPULATION COVARIANCE MATRICES. IF Σ_1 AND Σ_2 ARE KNOWN, THEN THERE IS NO PROBLEM: THE DIFFERENCE IN THE SAMPLE MEANS,

$$\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)},$$

IS MULTIVARIATE NORMAL WITH MEAN

$$\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$$

AND COVARIANCE MATRIX

$$\frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2.$$

SINCE THE SAMPLING DISTRIBUTION IS KNOWN, IT MAY BE USED TO MAKE TESTS ABOUT THE VALUE OF $\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$, AND TO CONSTRUCT CONFIDENCE REGIONS.

IF Σ_1 AND Σ_2 ARE UNKNOWN BUT EQUAL, SAY $\Sigma_1 = \Sigma_2 = \Sigma$, AND THE SAMPLE SIZE IS THE SAME FOR THE TWO SAMPLES IS THE SAME, $n_1 = n_2 = n$, THEN THE DIFFERENCES IN THE SAMPLE OBSERVATIONS,

$$\boldsymbol{d}_i = \boldsymbol{x}_i^{(1)} - \boldsymbol{x}_i^{(2)}$$

HAVE A MULTIVARIATE NORMAL DISTRIBUTION WITH MEAN

$$\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$$

AND COVARIANCE MATRIX 2Σ. THIS IS EXACTLY THE SITUATION FOR THE ONE-SAMPLE CASE (I.E., A SINGLE SAMPLE OF n OBSERVATIONS HAVING THE SAME MEAN AND VARIANCE).

FOR THE TWO-SAMPLE CASE IN WHICH THE SAMPLE SIZES ARE DIFFERENT AND THE COVARIANCE MATRIX IS UNKNOWN, HOWEVER, THIS APPROACH DOES NOT WORK. SINCE THE ESTIMATED COMMON SAMPLE VARIANCE DOES NOT HAVE A WISHART DISTRIBUTION. IN THIS CASE, THE STANDARD APPROACH IS TO ESTIMATE THE COVARIANCE MATRIX USING THE SAME NUMBER OF OBSERVATIONS FROM EACH SAMPLE, IN WHICH CASE THE ESTIMATED COMMON SAMPLE VARIANCE DOES HAVE A WISHART DISTRIBUTION.

TESTS OF HYPOTHESES ABOUT COVARIANCE MATRICES

TESTS ABOUT COVARIANCE MATRICES (E.G., WHETHER A COVARIANCE MATRIX IS EQUAL TO OR PROPORTIONAL TO A PARTICULAR MATRIX, OR WHETHER A SET OF COVARIANCE MATRICES ARE EQUAL) ARE COMPLICATED. THEY ARE BASED ON LIKELIHOOD RATIO TESTS, AND TYPICALLY INVOLVE THE CHARACTERISTIC ROOTS OF THE COVARIANCE MATRIX (OR MATRICES). THEY ARE COMPLICATED, AND NOT DISCUSSED HERE. THEY ARE DISCUSSED IN ANDERSON FOR SPECIAL CASES (E.G., TESTING WHETHER A SET OF COVARIANCES ARE EQUAL). (SINCE THE

DISTRIBUTION THEORY IS COMPLICATED, SUCH TESTS MAY BE BASED ON NONPARAMETRIC METHODS, OR ON RESAMPLING.)

THE MULTIVARIATE GENERAL LINEAR STATISTICAL MODEL

ESTIMATION

IN THE UNIVARIATE CASE, THE GENERAL LINEAR STATISTICAL MODEL HAS THE FORM, FOR THE i-th OBSERVATION OF A SAMPLE,

$$y_i = \boldsymbol{x_i}'\boldsymbol{\beta} + e_i$$

WHERE y_i IS THE OBSERVED RESPONSE, $\boldsymbol{x_i}' = (x_{i1},...,x_{ip})$ ARE EXPLANATORY VARIABLES, $\boldsymbol{\beta} = (\beta_1,...,\beta_p)'$ IS A VECTOR OF PARAMETERS, AND e_i IS A MODEL ERROR TERM HAVING MEAN ZERO AND VARIANCE σ^2. FOR THE MOMENT, LET US ASSUME THAT THE SAMPLE CONSISTS OF n OBSERVATIONS, AND THAT THE e_i ARE UNCORRELATED (i = 1,…, n).

IN MATRIX FORM, THE SAMPLE OF OBSERVATIONS IS WRITTEN AS

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}$$

WHERE $\boldsymbol{y} = (y_1,...,y_n)'$ IS THE VECTOR OF OBSERVED RESPONSES, X IS A MATRIX WHOSE ROWS ARE THE SAMPLE OF EXPLANATORY VARIABLES, $\boldsymbol{x_i}'$, (i = 1,…,n), AND $\boldsymbol{e} = (e_1,...,e_n)'$ IS A VECTOR OF MODEL ERROR TERMS.

THE MEAN OF $\boldsymbol{e}$ IS E($\boldsymbol{e}$) = $\boldsymbol{0}$. UNDER THE ASSUMPTION THAT THE e_i ARE UNCORRELATED, THE COVARIANCE MATRIX OF $\boldsymbol{e}$ IS $I_n\sigma^2$ (WHERE I_n DENOTES THE n x n IDENTITY MATRIX).

THE LEAST-SQUARES ESTIMATOR OF THE PARAMETER $\boldsymbol{\beta}$ IS

$$\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}$$

WHERE IT IS ASSUMED THAT THE MATRIX X'X IS INVERTIBLE.

IN MANY APPLICATIONS, THE ASSUMPTION THAT THE MODEL ERROR TERMS ARE UNCORRELATED IS REASONABLE, BUT IN MANY OTHERS, SUCH AS IN MULTIVARIATE ANALYSIS AND TIME-SERIES ANALYSIS, IT IS NOT. IF IT IS ASSUMED THAT THE COVARIANCE MATRIX OF THE MODEL ERROR TERMS IS A GENERAL (SYMMETRIC, POSITIVE DEFINITE) MATRIX Σ, THEN THE LEAST SQUARES ESTIMATOR HAS THE FORM

$$\widehat{\boldsymbol{\beta}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\boldsymbol{y}.$$

(THIS RESULT IS OBTAINED DIRECTLY FROM THE FIRST ONE (COVARIANCE MATRIX $I_n\sigma^2$) BY MAKING THE TRANSFORMATION $\boldsymbol{z} = \Sigma^{-1/2}\boldsymbol{y}$, WHICH HAS COVARIANCE MATRIX I, AND USING THE EARLIER ESTIMATOR.)

THE MULTIVARIATE GENERAL LINEAR STATISTICAL MODEL IS A DIRECT EXTENSION OF THE UNIVARITE ONE. INSTEAD OF A UNIVARIATE OBSERVATION y_i, WE NOW HAVE A MULTIVARIATE OBSERVATION $\boldsymbol{y}_i$ COMPOSED OF k-COMPONENT VECTORS. THE PARAMETER VECTOR $\boldsymbol{\beta}$ MAY DIFFER FOR EACH COMPONENT (I.E., IS $\boldsymbol{\beta}_i$ FOR THE i-th COMPONENT). WE SHALL ASSUME THAT THE ERROR TERMS FOR A PARTICULAR COMPONENT HAVE ZERO MEAND AND ARE UNCORRELATED BETWEEN DIFFERENT SAMPLE OBSERVATIONS, BUT THAT THE COVARIANCE MATRIX WITHIN SAMPLE OBSERVATIONS IS Σ_e, A POSITIVE DEFINITE MATRIX.

(NOTE ON NOTATION: EARLIER, WE USED p TO DENOTE THE NUMBER OF COMPONENTS OF A MULTIVARIATE VECTOR, AND WE ALSO USED p TO DENOTE THE NUMBER OF PARAMETERS IN A LINEAR MODEL, I.E., THE NUMBER OF COMPONENTS OF $\boldsymbol{\beta}$. IN WHAT FOLLOWS, WE SHALL USE k TO DENOTE THE NUMBER OF COMPONENTS IN A MULTIVARIATE VECTOR, AND p TO DENOTE THE NUMBER OF COMPONENTS IN $\boldsymbol{\beta}$.)

IN COMPACT FORM, THE FULL SAMPLE OF n OBSERVATIONS FOR THE MULTIVARIATE MODEL MAY BE WRITTEN AS

$$Y = X\mathrm{B} + E$$

WHERE Y IS AN n x k MATRIX WITH ROW $y_i = (y_{i1},...,y_{ik})'$, X IS AN n x kp MATRIX (DESIGN MATRIX OR DATA MATRIX) OF EXPLANATORY VARIABLES, AND B IS A kp x k MATRIX OF PARAMETERS, AND E IS AN n x k MATRIX OF MODEL ERROR TERMS.

THIS COMPACT MULTIVARIATE FORM IS NOT THE FORM USED IN THE JUST-DESCRIBED UNIVARIATE CASE. THAT THEORY MAY BE APPLIED TO THE MULTIVARIATE CASE BY CONCATENATING THE SAMPLE VECTORS INTO A SINGLE COLUMN VECTOR. THAT IS, THE PRECEDING MODEL MAY BE WRITTEN IN "STACKED" FORM AS

$$vec(Y) = (I_k \otimes X)vec(\mathrm{B}) + vec(E).$$

DETERMINATION OF THE LEAST-SQUARES ESTIMATE FOR THE PARAMETER PROCEEDS IN THE SAME WAY AS USUAL, VIZ., BY MINIMIZING THE ERROR SUM OF SQUARES (BY SETTING PARTIAL DERIVATIVES OF THE ERROR SUM OF SQUARES EQUAL TO ZERO).

SOME MATRIX MANIPULATIONS SHOW THAT THE LEAST-SQUARES ESTIMATOR FOR THIS MODEL IS

$$\widehat{\mathrm{B}} = (X'X)^{-1}(X'Y).$$

THIS IS THE SAME FORM AS FOR THE UNIVARIATE CASE, WHERE THE n x 1 VECTOR y HAS BEEN REPLACED BY THE n x k MATRIX Y AND THE DATA/DESIGN MATRIX X IS NOW AN n x kp MATRIX INSTEAD OF AN n x p MATRIX.

NOTE THAT THE SOLUTION DOES NOT INVOLVE THE MATRIX Σ_e. THE REASON FOR THIS IS THE ASSUMPTION THAT THERE IS NO CORRELATION BETWEEN VARIABLES IN DIFFERENT OBSERVATIONS. THE ESTIMATES MAY BE DETERMINED SEPARATELY FOR EACH COMPONENT VARIABLE OF THE MULTIVARIATE VECTOR IN THE USUAL FASHION, USING EXACTLY THE SAME COMPUTATIONAL PROCEDURES AS IN THE UNIVARIATE CASE. THIS SITUATION IS REFERRED TO AS "SEEMINGLY UNRELATED REGRESSIONS."

AS BEFORE, DEFINE THE MODEL RESIDUALS $\widehat{E}$ TO BE THE DIFFERENCES BETWEEN THE OBSERVED VALUES AND THE VALUES PREDICTED BY THE MODEL BY SUBSTITUTING THE LEAST-SQUARES ESTIMATES IN THE MODEL EQUATION. LET $\widetilde{\Sigma}_e$ DENOTE THE SAMPLE COVARIANCE MATRIX DEFINED BY

$$\tilde{\Sigma}_e = \frac{1}{n - (k+1)p - 1} \hat{E}'\hat{E}.$$

THEN IT CAN BE SHOWN THAT THE LEAST-SQUARES ESTIMATE $\hat{B}$ IS UNBIASED ($E(\hat{B}) = B$); THE ESTIMATE $\tilde{\Sigma}_e$ IS UNBIASED ($E(\tilde{\Sigma}_e) = \Sigma_e$; $\hat{B}$ AND $\tilde{\Sigma}_e$ ARE UNCORRELATED; AND THE COVARIANCE MATRIX OF THE (STACKED) PARAMETER ESTIMATES IS

$$Cov[vec(\hat{B})] = \tilde{\Sigma}_e \otimes (X'X)^{-1}.$$

THE LEAST-SQUARES ESTIMATES DO NOT DEPEND ON ANY DISTRIBUTIONAL ASSUMPTIONS ABOUT THE MODEL ERROR TERMS, OTHER THAN THEIR HAVING ZERO MEAN, NO CORRELATION BETWEEN OBSERVATIONS, AND COVARIANCE MATRIX Σ_e WITHIN OBSERVATIONS. IF IT IS ALSO ASSUMED THAT THE ERROR TERMS HAVE A MULTIVARIATE NORMAL DISTRIBUTION, THEN IT CAN BE SHOWN THAT THE MAXIMUM LIKELIHOOD ESTIMATE $\hat{B}$ OF B IS THE SAME AS THE LEAST-SQUARES ESTIMATE; THAT THE MAXIMUM LIKELIHOOD ESTIMATE OF Σ_e IS $\tilde{\Sigma}_e$; THAT (n-p) $\tilde{\Sigma}_e$ HAS A WISHART DISTRIBUTION W$_{k,n-(k+1)p-1}$; THAT $vec(\hat{B})$ IS NORMALLY DISTRIBUTED WITH MEAN $vec(B)$ AND COVARIANCE MATRIX $\Sigma_e \otimes (X'X)^{-1}$; $vec(\hat{B})$ IS INDEPENDENT OF $\tilde{\Sigma}_e$. FURTHERMORE, THE ESTIMATES ARE ASYMPTOTICALLY NORMALLY DISTRIBUTED (SEE TSAY FOR MORE DETAILS.) THESE RESULTS MAY BE USED TO PERFORM TESTS OF HYPOTHESIS AND CONSTRUCT CONFIDENCE INTERVALS.

TESTS OF HYPOTHESIS AND CONFIDENCE REGIONS

FOR THE GENERAL LINEAR MODEL, TESTS OF HYPOTHESIS (AND CONFIDENCE REGIONS) ARE CONSTRUCTED USING THE LIKELIHOOD RATIO CRITERION. MANY OF THE TESTS IN THE MULTIVARIATE CASE ARE ANALOGOUS TO THOSE OF THE UNIVARIATE CASE, SUCH AS TESTING WHETHER REGRESSION PARAMETERS ARE EQUAL TO ZERO. THE ADDITIONAL COMPLEXITY OF THE MULTIVARIATE MODEL LEADS TO ADDITIONAL KINDS OF TEST, SUCH AS TESTS ABOUT MATRICES OF REGRESSION COEFFICIENTS.

SEE ANDERSON FOR DISCUSSION OF THE TOPIC OF TESTS OF HYPOTHESIS AND CONFIDENCE REGIONS IN THE MULTIVARIATE CASE.

MULTIVARIATE ANALYSIS OF VARIANCE AND ANALYSIS OF COVARIANCE

THE GENERAL LINEAR MODEL INCLUDES THE METHODOLOGIES OF ANALYSIS OF VARIANCE AND ANALYSIS OF COVARIANCE AS SPECIAL CASES. ANALYSIS OF VARIANCE ADDRESSES LINEAR MODELS IN WHICH EXPLANATORY VARIABLES ARE CATEGORICAL. ANALYSIS OF COVARIANCE ADDRESSES LINEAR MODELS IN WHICH MAJOR EXPLANATORY VARIABLES ARE CATEGORICAL, AND A FEW VARIABLES ARE CONTINUOUS. THESE METHODOLOGIES ARE USED IN EXPERIMENTAL DESIGNS, WHERE THE EXPERIMENTER HAS CONTROL OVER THE MAIN EXPLANATORY VARIABLES. FOR THESE METHODOLOGIES, RESULTS ARE PRESENTED IN SPECIAL TABLES (VIZ., ANALYSIS OF VARIANCE TABLES AND ANALYSIS OF COVARIANCE TABLES).

THESE TOPICS ARE NOT ADDRESSED IN THIS PRESENTATION (EITHER IN THE UNIVARIATE CASE OR THE MULTIVARIATE CASE), BUT MAY BE THE SUBJECT OF A FUTURE PRESENTATION (ON EXPERIMENTAL DESIGN).

TESTS FOR INDEPENDENCE

A PROBLEM THAT ARISES IN MULTIVARIATE IS THAT OF DIVIDING A MULTIVARIATE VECTOR INTO SETS OF VARIATES, AND TESTING WHETHER THE SETS ARE MUTUALLY INDEPENDENT, I.E., WHETHER EACH MEMBER OF ONE SET IS UNCORRERATED WITH EACH VARIABLE OF ANOTHER SET.

THE METHODOLOGY FOR MULTIVARIATE TESTS FOR INDEPENDENCE IS SIMILAR TO THAT FOR THE TESTS OF HYPOTHESIS FOR THE MULTIVARIATE GENERAL LINEAR MODEL. FOR THE MULTIVARIATE NORMAL DISTRIBUTION, A TEST OF INDEPENDENCE MAY BE FORMULATED AS A TEST ABOUT THE STRUCTURE OF THE COVARIANCE MATRIX, I.E., WHETHER CERTAIN SUBMATRICES ARE ZERO. THE LIKELIHOOD RATIO PROCEDURE IS USED TO CONSTRUCT A TEST. THE TESTS MAY BE FORMULATED IN TERMS OF GENERALIZED VARIANCES.

SEE ANDERSON FOR DETAILS.

PRINCIPAL COMPONENTS ANALYSIS

A FUNDAMENTAL DIFFICULTY WITH MULTIVARIATE ANALYSIS IS THE LARGE NUMBER OF PARAMETERS INVOLVED IN MULTIVARIATE MODELS. AS MENTIONED EARLIER, THE NUMBER OF PARAMETERS IN A BASIC k-DIMENSIONAL MULTIVARIATE NORMAL DISTRIBUTION IS k MEANS, k VARIANCES AND $(k^2 - k)/2$ COVARIANCES, FOR A TOTAL OF $(k^2 -k)/2 + 2k$ PARAMETERS. A MULTIVARIATE GENERAL LINEAR MODEL THAT INCLUDES p EXPLANATORY VARIABLES INCLUDES pk PARAMETERS FOR THE MEAN. IF THERE IS NO INTERCORRELATION AMONG SAMPLE OBSERVATIONS, THEN THE NUMBER OF VARIANCES AND COVARIANCES IS THE SAME AS FOR THE BASIC MULTIVARIATE DISTRIBUTION, BUT IF THERE ARE CORRELATIONS AMONG SAMPLE OBSERVATIONS (AS IS THE CASE IN TIME SERIES ANALYSIS), THE NUMBER OF PARMATERS MAY BE MUCH LARGER.

MODELS HAVING LARGE NUMBER OF PARAMETERS POSSESS AN INHERENT DANGER FOR MANY APPLICATIONS (SUCH AS FORECASTING), IN THAT THE MODEL MAY FIT THE DATA WELL AND YET FORECAST POORLY. IT IS USEFUL, THEREFORE, TO APPLY METHODS TO SIMPLIFY THE MODELS. THIS CAN BE DONE IN A NUMBER OF WAYS, SUCH AS BY MAKING SIMPLIFYING ASSUMPTIONS (JUSTIFIED BY SUBSTANTIVE THEORY) OR BY COMBINING VARIABLES OR ELIMINATING VARIABLES.

WITH RESPECT TO COMBINING VARIABLES AND DROPPING VARIABLES, A POWERFUL TOOL IS THE METHOD OF PRINCIPAL COMPONENTS. AS DISCUSSED EARLIER, THE METHOD OF PRINCIPAL COMPONENTS IS THE APPLICATION OF A LINEAR TRANSFORMATION OF THE ORIGINAL MULTIVARIATE VECTOR TO A NEW VECTOR HAVING USEFUL STATISTICAL PROPERTIES. SPECIFICALLY, THE ORIGINAL COORDINATE SYSTEM IS ROTATED IN SUCH A WAY THAT THE FIRST VARIABLE OF THE NEW COORDINATE SYSTEM HAS MAXIMUM VARIANCE, THE SECOND VARIABLE HAS MAXIMUM VARIANCE OUT OF ALL VARIABLES ORTHOGONAL TO THE FIRST VARIABLE, THE THIRD VARIABLE HAS MAXIMUM VARIANCE OUT OF ALL VARIABLES ORTHOGONAL TO THE FIRST TWO, AND SO ON.

THE EARLIER DISCUSSION ADDRESSED THE ISSUE OF PRINCIPAL COMPONENTS IN THE POPULATION, I.E., IT WAS ASSUMED THAT THE PARAMETERS OF THE MULTIVARIATE DISTRIBUTION WERE KNOWN. THIS SECTION ADDRESSES THE SITUATION IN WHICH THE DISTRIBUTION PARAMETERS ARE ESTIMATED.

THE PRINCIPAL COMPONENTS ARE THE EIGENVECTORS OF THE COVARIANCE MATRIX, WHERE THE LENGTH OF EACH EIGENVECTOR IS THE CORRESPONDING EIGENVALUE. THE EIGENVALUES ARE THE SOLUTIONS (λ_1,…,λ_p) TO THE EQUATION

$$|\Sigma - \lambda I| = 0,$$

AND THE EIGENVECTORS ARE THE SOLUTIONS ($\mathbf{b}_1$,…,$\mathbf{b}_p$) TO THE EQUATION

$$(\Sigma - \lambda I)\mathbf{b} = \mathbf{0}.$$

LET US DEFINE THE MATRIX Λ AS

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix}.$$

THE PROBLEM IS TO ESTIMATE ALL OF THE PARAMETERS $\boldsymbol{\mu}$, Σ, Λ AND B = ($\mathbf{b}_1$,…,$\mathbf{b}_p$). THIS IS DONE, AS USUAL, BY THE METHOD OF MAXIMUM LIKELIHOOD.

IT IS A FACT THAT THE MAXIMUM LIKELIHOOD ESTIMATOR OF A FUNCTION OF A PARAMETER IS THE FUNCTION OF THE MAXIMUM LIKELIHOOD ESTIMATOR OF THE PARAMETER. SINCE Λ AND B ARE FUNCTIONS OF $\boldsymbol{\mu}$ AND Σ, IT FOLLOWS THAT THE MAXIMUM LIKELIHOOD ESTIMATORS OF Λ AND B ARE OBTAINED USING THE SAME EQUATIONS GIVEN ABOVE FOR DETERMINING Λ AND B, SIMPLY BY REPLACING $\boldsymbol{\mu}$ AND Σ BY THEIR MAXIMUM LIKELIHOOD ESTIMATORS ($\hat{\boldsymbol{\mu}}$ AND $\hat{\Sigma}$).

AS DISCUSSED EARLIER, THE METHOD OF PRINCIPAL COMPONENTS IS A METHOD OF ROTATING THE AXES OF A MULTIVARIATE DISTRIBUTION (I.E., TRANSFORMING THE COMPONENT VARIABLES OF A MULTIVARIATE VECTOR) SUCH THAT THE FIRST PRINCIPAL COMPONENT HAS MAXIMUM VARIANCE, THE SECOND PRINCIPAL COMPONENT HAS MAXIMUM VARIANCE, OUT OF ALL VARIABLES ORTHOGONAL TO THE FIRST, AND SO ON.

SINCE THE MATRIX TRANSFORMATIONS INVOLVED IN THE METHOD OF PRINCIPAL COMPONENTS ARE ORTHONORMAL, THE DETERMINANT OF THE COVARIANCE MATRICES OF THE ORIGINAL AND TRANSFORMED VARIABLES HAS THE SAME VALUE. ALSO, THE TRACE OF THE COVARIANCE (SUM OF THE DIAGONAL ENTRIES,

AND ALSO THE SUM OF THE CHARACTERISTIC ROOTS) HAS THE SAME VALUE. THE METHOD OF PRINCIPAL COMPONENTS IS SIMPLY A METHOD OF TRANSFORMING VARIABLES TO NEW VARIABLES IN SUCH A WAY THAT THE TOTAL VARIANCE (AS REPRESENTED BY THE TRACE) IS DECOMPOSED.

FROM AN INFORMATION-THEORETIC VIEWPOINT, THE AMOUNT OF INFORMATION ASSOCIATED WITH A VARIABLE IS INVERSELY PROPORTIONAL TO THE VARIANCE OF THE VARIABLE.

A DIFFICULTY WITH TRANSFORMED VARIABLES IS THEIR INTERPRETATION. THE METHOD OF PRINCIPAL COMPONENTS IS APPROPRIATE FOR TRANSFORMING VARIABLES OF A COMMON METRIC, FOR WHICH THE LINEAR COMBINATIONS OF THE VARIABLES HAVE A REASONABLE MEANING (INTERPRETATION). FOR EXAMPLE, A SET OF 15 MEASURMENTS MAY BE TAKEN ON ANIMAL SKULLS. USING THE METHOD OF PRINCIPAL COMPONENTS, IT MAY BE TURN OUT THAT THE FIRST PRINCIPAL COMPONENT IS APPROXIMATELY EQUAL TO THE SUM OF THE HEIGHT, WIDTH AND DEPTH OF THE SKULL, AND THAT THE OTHER PRINCIPAL COMPONENTS ARE VERY SMALL. THAT IS, MOST OF THE VARIATION IN SKULL DIMENSIONS IS SUMMARIZED IN THIS SINGLE VARIABLE. IN SOME APPLICATIONS, THE ORIGINAL 15 MEASUREMENTS MAY BE REPLACED BY A SINGLE VARIABLE, WHICH CONTAINS MOST OF THE INFORMATION INHERENT IN THE ORIGINAL 15 VARIABLES.

FACTOR ANALYSIS

IN THE METHOD OF PRINCIPAL COMPONENTS, INTEREST FOCUSSES ON THE LARGER CHARACTERISTIC ROOTS. IF THE GOAL IS TO SIMPLIFY A PROBLEM BY REDUCING THE NUMBER OF COMPONENTS OF THE MULTIVARIATE VECTOR, THEN ONE APPROACH IS TO DISCARD VARIABLES ASSOCIATED WITH THE SMALLER CHARACTERISTIC ROOTS. THIS MAY BE DONE IN AN AD HOC MANNER, BUT A MORE STANDARD PROCEDURE IS THE METHOD OF FACTOR ANALYSIS.

IN FACTOR ANALYSIS IT IS ASSUMED THAT THE MULTIVARIATE VECTOR X MAY BE REPRESENTED AS

$$X = \Lambda \boldsymbol{f} + \boldsymbol{\mu} + \boldsymbol{U}$$

WHERE **f** IS AN m-COMPONENT VECTOR OF NONOBSERVABLE FACTOR SCORES, **μ** IS A VECTOR OF MEANS, AND **U** IS A VECTOR OF NONOBSERVABLE ERRORS. THE p x m MATRIX Λ (m < p) IS CALLED A MATRIX OF FACTOR LOADINGS. IT IS ASSUMED THAT **f** IS A RANDOM VARIABLE WITH E(**f**) = **0** AND E**ff′** = M; IT IS ALSO ASSUMED THAT E(**U**) = **0** AND E(**UU′**) = Ψ, A DIAGONAL MATRIX, AND E(**fU′**) = **0**.

UNDER THESE ASSUMPTIONS E(**X**) = **μ** AND

$$E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})' = \Lambda M \Lambda' + \Psi.$$

A PROBLEM WITH THIS MODEL IS THAT IT IS NOT KNOWN WHAT COVARIANCE MATRICES CAN BE REPRESENTED BY THE PRECEDING EXPRESSION OF GIVEN m, AND IF SUCH MATRICES EXIST, WHAT RESTRICTIONS ON Λ AND M MAKE THEM UNIQUE. IN OTHER WORDS, THERE MAY NOT EXIST A MULTIVARIATE NORMAL RANDOM VARIABLE X SATISFYING THE PRECEDING MODEL.

TO ADDRESS THIS ISSUE, THE ASSUMPTION OF NORMALITY OF X MUST BE DROPPED. A VARIETY OF METHODS MAY BE USED TO ESTIMATE THE PARAMETERS OF THE FACTOR MODEL (Λ, M AND Ψ). THE METHOD OF MAXIMUM LIKELIHOOD MAY BE USED, BUT PROBLEMS ARE ENCOUNTERED SINCE THE MODEL IS NOT UNIQUE (I.E., THERE ARE MANY ALTERNATIVE REPRESENTATIONS, OR SELECTIONS OF FACTORS **f**). ALTERNATIVE ESTIMATION METHODS INCLUDE THE CENTROID METHOD AND THE METHOD OF PRINCIPAL COMPONENTS.

DESPITE THESE PROBLEMS, CONSIDERABLE USE IS MADE OF THE PRECEDING MODEL. FOR THE PROBLEM OF DECIDING WHICH VARIABLES TO DISCARD FROM A MULTIVARIATE VECTOR WITH RELATIVELY LITTLE REDUCTION IN THE TOTAL VARIANCE, IT IS USED TO DETERMINE A REASONABLE VALUE FOR m. A PRIMARY CONSIDERATION IN FACTOR ANALYSIS IS THE IDENTIFICATION OF FACTORS (LINEAR COMBINATIONS OF THE ORIGINAL VARIABLES) THAT HAVE A REASONABLE MEANING. IN THE SEARCH FOR SUCH REPRESENTATIONS, A VARIETY OF AXIS ROTATIONS MAY BE CONSIDERED.

THE REFERENCES PROVIDED EARLIER BY LAWLY AND MAXWELL, AND BY HARMON, MAY BE CONSULTED FOR ADDITIONAL DETAILS ON THIS METHOD.

CANONICAL CORRELATION

ESTIMATION OF CANONICAL CORRELATIONS AND CANONICAL VARIATES PROCEEDS BY MAKING THE SAME CALCULATIONS ON THE SAMPLE CORRELATION MATRIX, $\hat{\Sigma}$, AS ARE USED TO OBTAIN THE POPULATION CANONICAL CORRELATIONS.

CLASSIFICATION ANALYSIS

IN THE PRECEDING DISCUSSION OF CLASSIFICATION ANALYSIS, A PROCEDURE WAS PRESENTED THAT MINIMIZES THE EXPECTED COST OF MISCLASSIFICATION. THE CLASSIFICATION CRITERION WAS BASED ON KNOWN PROBABILITIES. IT CANNOT BE PROVED, HOWEVER, THAT THE DESCRIBED PROCEDURE MINIMIZES THE EXPECTED COST IN THE CASE WHERE THE PROBABILITIES ARE ESTIMATED. THE CRITERION APPEARS REASONABLE, HOWEVER, AND IS USED.

MULTIVARIATE TIME SERIES MODELS (SUMMARY)

A PRECEDING SECTION DESCRIBED THE CLASS OF BOX-JENKINS MODELS, WHICH ARE USED AS A BASIS FOR FORECASTING AND CONTROL IN BOTH THE UNIVARIATE AND MULTIVARIATE CASES. DETAILED DISCUSSION OF THE PROCEDURES FOR DEVELOPING BOX-JENKINS MODELS IS THE SUBJECT OF A SEPARATE PRESENTATION.

THE METHODOLOGY FOR ESTIMATING A MULTIVARIATE BOX-JENKINS MODEL FROM SAMPLE DATA PARALLELS THAT FOR A UNIVARIATE MODEL. IT MAKES HEAVY USE OF LEAST-SQUARES THEORY, AND ITERATIVE APPLICATION OF THE STEPS OF MODEL IDENTIFICATION, ESTIMATION AND TESTING.

WHILE MUCH OF THE METHODOLOGY IS SIMILAR IN THE MULTIVARIATE AND UNIVARIATE CASES, THERE ARE SOME DIFFERENCES. ONE AREA OF DIFFERENCE INVOLVES THE FORECASTING ONCE COMPONENT OF A MULTIVARIATE VECTOR, CONDITIONAL ON THE VALUES OF OTHER COMPONENTS. THE PROBLEM THAT ARISES IS THAT IF A FUTURE VALUE OF ONE COMPONENT IS SPECIFIED, IT IS NOT REASONABLE TO BELIEVE THAT THE OTHER VARIABLES MAY BE HELD CONSTANT, IF THE VARIABLES ARE CORRELATED. AN APPROACH TO HANDLING THIS

PROBLEM IS TO APPLY THE MULTIVARIATE PROCEDURE OF PRINCIPAL COMPONENTS TO ORTHOGONALIZE THE VECTOR SO THAT THE COMPONENTS ARE ORTHOGONAL. IN THAT CASE, IT IS REASONABLE TO CONDITION ON ONE OF THEM, HOLDING THE OTHERS FIXED. AS USUAL, THIS APPROACH IS USEFUL IF A MEANINGFUL INTERPRETATION CAN BE GIVEN TO THE TRANSFORMED VARIABLES.

FndID(206)
FndTitle(CONTINUOUS MULTIVARIATE ANALYSIS: LECTURE NOTES)
FndDescription(CONTINUOUS MULTIVARIATE ANALYSIS: LECTURE NOTES)
FndKeywords(multivariate analysis; statistics course; short course)