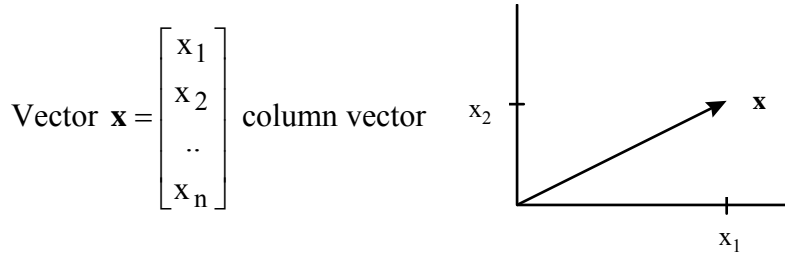
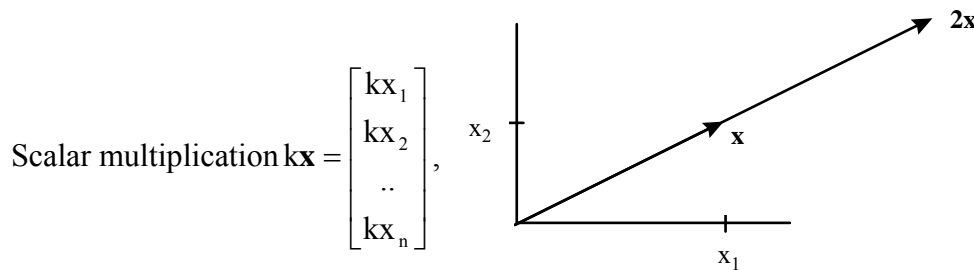
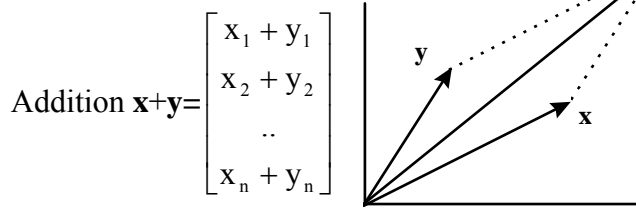


Matrix Manipulation of Data: Matrix Algebra

1. Vectors



Transpose $\mathbf{x}' = [x_1 \ x_2 \ \dots \ x_n]$ row vector,



Definition: \mathbf{x} and \mathbf{y} point in the same direction if for some $k \neq 0$, $\mathbf{y} = k\mathbf{x}$.

Vectors have a direction and length. Euclidian Length $\|\mathbf{x}\| = \sqrt{\sum x_i^2}$

Inner product $\mathbf{x}'\mathbf{y} = \sum x_i y_i$

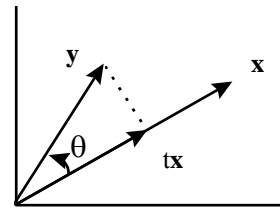
Note: we think of this as a row times a column

Suppose $t\mathbf{x}$ is a projection of \mathbf{y} into \mathbf{x} :

$$\mathbf{y}'\mathbf{y} = t^2 \mathbf{x}'\mathbf{x} + (\mathbf{y} - t\mathbf{x})'(\mathbf{y} - t\mathbf{x}) = t^2 \mathbf{x}'\mathbf{x} + \mathbf{y}'\mathbf{y} - 2t\mathbf{x}'\mathbf{y} + t^2 \mathbf{x}'\mathbf{x} \Rightarrow t = \mathbf{x}'\mathbf{y} / \mathbf{x}'\mathbf{x}.$$

Note: $\cos \theta \equiv t \|\mathbf{x}\| / \|\mathbf{y}\|$, so

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos(\theta).$$



The inner product is the product of lengths correcting for not pointing in the same direction.

If two vectors \mathbf{x} and \mathbf{y} point in the precisely the same direction, then we say that there are **dependent**. More generally, they are dependent if there exists $k_1, k_2 \neq 0$ such that $k_1 \mathbf{x} + k_2 \mathbf{y} = \mathbf{0}$.

That is, a linear combination of \mathbf{x} and \mathbf{y} with weights $\mathbf{k} = (k_1, k_2)'$ equals the zero vector

If \mathbf{x} and \mathbf{y} are not dependent, then they are **independent**: they point in different directions.

Maximal independence is **orthogonal** (perpendicular): $\theta = 90^\circ \Rightarrow \cos \theta = 0 \Rightarrow \mathbf{x}'\mathbf{y} = 0$.

Orthonormal vectors are both orthogonal of unit length: $\mathbf{x}'\mathbf{y} = 0$ and $\|\mathbf{x}\| = 1 = \|\mathbf{y}\|$. They lie on the unit circle.

2. Matrices

n×p matrix:
$$X = \begin{matrix} & \begin{matrix} 1 & j & p \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ i \\ n \end{matrix} & \begin{bmatrix} x_{11} & \dots & x_{1p} \\ x_{21} & \dots & x_{2p} \\ \dots & x_{ij} & \dots \\ x_{n1} & \dots & x_{np} \end{bmatrix} \end{matrix}$$
 n and p are the dimensions of the matrix. In a data matrix,

n is the number of observations and p is the number of variables. $X^{(j)}$ denotes the jth column which holds all the observations of the jth variable, while X_i denotes the ith row which holds the ith observation of all the variables.

Matrix algebra

Equality $X=Y \Leftrightarrow x_{ij}=y_{ij}$ if dimensions conform

Matrix addition $X+Y=Z$ $z_{ij}=x_{ij}+y_{ij}$ if dimensions conform

Scalar multiplication $kX=[kx_{ij}]$

Transpose X' denotes a p×n matrix $=[x_{ji}]$ where rows and columns are interchanged

Symmetric square matrix: $X=X'$.

Multiplying a matrix X by a vector **b**, $X\mathbf{b}$, can be thought of in two different ways depending on whether we focus on the rows or columns of X. Focusing on rows, think of X as rows stacked one upon another. Then the product $X\mathbf{b}$ is a sequence of inner products $X_i\mathbf{b}=\sum x_{ij}b_j$ stacked one upon another. Alternatively, if we think of X as a collection of column vectors, then $X\mathbf{b}$ is a linear combination of the vectors with weights given in **b**: $\sum X^{(j)}b_j$.

Rank(X)=maximum number of independent rows in matrix: $\text{rank}(X)\leq n$. If the rank is below n, then some observations are essentially duplicates of others.

Matrix Multiplication: $Z=XY$ is defined if X n×p and Y p×m then Z is n×m where $z_{ij}=X_iY^{(j)}$ = inner product of ith row of X and jth column Y.

Theorem: $A+B=B+A$, $A(B+C)=AB+AC$, $(A')'=A$, $(A+B)'=A'+B'$, $(AB)'=B'A'$

Note: $AB=AC \not\Rightarrow B=C$

Note: $AB\neq BA$ except in special cases.

Special matrices

A is n×n	Square Matrix	
$I_n = I =$	$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix}$	Identity Matrix $IX=X$

$$O = \begin{bmatrix} 0 & \dots & 0 \\ 0 & \dots & 0 \\ 0 & \dots & 0 \end{bmatrix} \quad \begin{array}{l} \text{Zero} \\ \text{Matrix} \\ OX=O \end{array}$$

$$\mathbf{u}_i = \begin{bmatrix} 0 \\ \dots \\ 1 \\ 0 \\ \dots \\ 0 \end{bmatrix} \leftarrow i^{\text{th}} \text{ row} \quad \begin{array}{l} i^{\text{th}} \text{ Unit} \\ \text{Vector} \end{array}$$

$$\text{Diag}(A) = \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & a_{nn} \end{bmatrix} \quad \begin{array}{l} \text{Diagonal} \\ \text{Matrix} \end{array}$$

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} \quad \begin{array}{l} \text{Unit} \\ \text{Vector} \end{array}$$

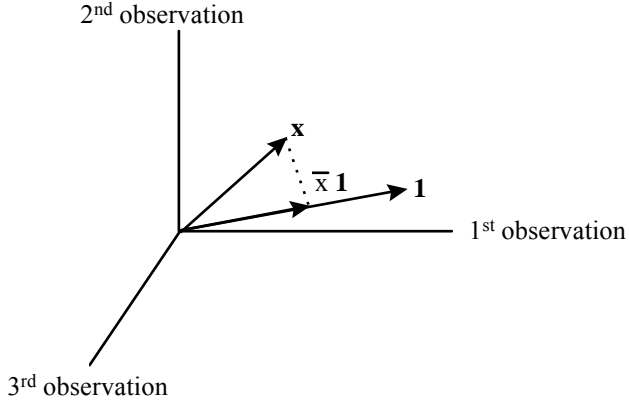
$$J = \mathbf{1}\mathbf{1}' = \begin{bmatrix} 1 & \dots & 1 \\ 1 & \dots & 1 \\ \dots & \dots & \dots \\ 1 & \dots & 1 \end{bmatrix} \quad \begin{array}{l} \text{Unit} \\ \text{Matrix} \end{array}$$

$$H = I - J/n = \begin{bmatrix} \frac{n-1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \dots & \frac{n-1}{n} \end{bmatrix} \quad \begin{array}{l} \text{Centering} \\ \text{Matrix} \end{array}$$

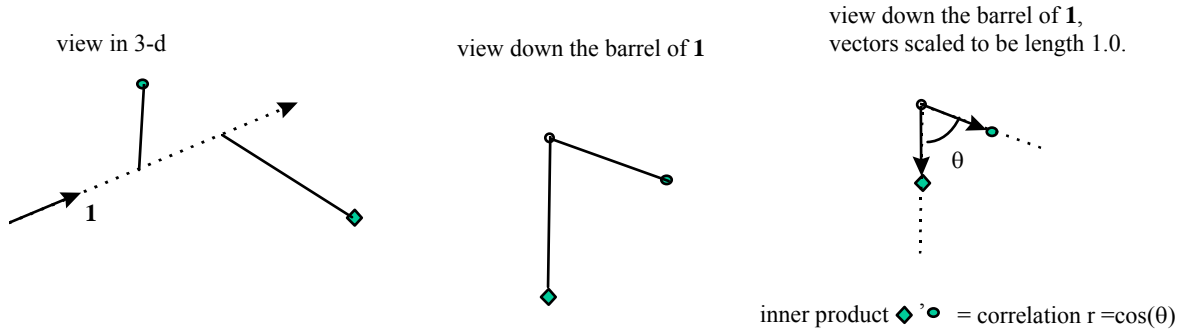
Theorem: The centering matrix H is idempotent, $HH=H$.

3. Descriptive Statistics Using Matrix Notation

Mean vector: $\bar{\mathbf{x}} = \mathbf{X}'\mathbf{1}/n$ is a p -vector of means. Note that the interpretation of inner product as a projection implies that the mean is a projection into the unit vector in an n -dimensional space. See below for 3 observations of a single variable (this is not a scatterplot).



The mean p -vector $\bar{\mathbf{x}}$ tells us where the center of the scatterplot is located. We are also interested in how the observations differ from the mean. Data as deviations from mean $\mathbf{X}_d = \mathbf{HX} = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}'$.



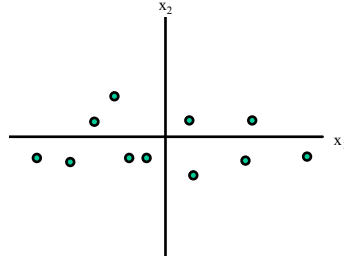
Covariance matrix $\mathbf{S}_n = \mathbf{X}_d' \mathbf{X}_d / n = (\mathbf{HX})' \mathbf{HX} / n = \mathbf{X}' \mathbf{HX} / n$ is a $p \times p$ matrix with variance along diagonal and covariances off-diagonal. If we calculate covariance matrix dividing by $n-1$ rather than n , we write it as \mathbf{S} . $\mathbf{M} = \mathbf{X}' \mathbf{X}$ is the matrix of sums of squares and products.

Standard deviation matrix $\mathbf{D}^{1/2} = \begin{bmatrix} \sqrt{S_{11}} & 0 & 0 & 0 \\ 0 & \sqrt{S_{22}} & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \sqrt{S_{pp}} \end{bmatrix} = \text{sqrt}(\text{diag}(\mathbf{S}))$ SPSS, N.B. text uses \mathbf{D}

Correlation matrix $\mathbf{R} = \mathbf{D}^{-1/2} \mathbf{S} \mathbf{D}^{-1/2} = \begin{bmatrix} 1 & \frac{S_{12}}{\sqrt{S_{11}} \sqrt{S_{22}}} & \dots & \frac{S_{1p}}{\sqrt{S_{11}} \sqrt{S_{pp}}} \\ \frac{S_{21}}{\sqrt{S_{22}} \sqrt{S_{11}}} & 1 & \dots & \frac{S_{2p}}{\sqrt{S_{22}} \sqrt{S_{pp}}} \\ \vdots & \dots & \ddots & \vdots \\ \frac{S_{p1}}{\sqrt{S_{pp}} \sqrt{S_{11}}} & \dots & \dots & 1 \end{bmatrix}$

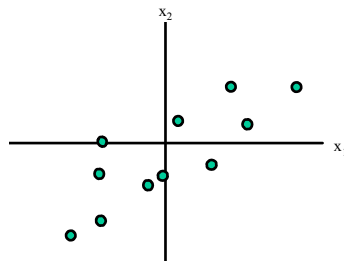
A linear combination of variables is defined by a vector \mathbf{a} : \mathbf{Xa} . Sample mean of linear combination is $\bar{x}'\mathbf{a}$. Sample variance of a linear combination is $\mathbf{a}'\mathbf{S}\mathbf{a}$.

In a p-dimensional scatter plot, we have the n observations of the p variables.




For each observation (row X_i of data matrix X), we could calculate the Euclidian distance $\sqrt{X_i X_i'} = \sqrt{\sum_{j=1}^p x_{ij}^2}$. However, it is possible that one variable fluctuates much more than another and yet all variables are treated equally in the summation. To adjust for this, we often take a scaling transformation: $\mathbf{z}_i = \mathbf{D}^{-1/2}(\mathbf{x}_i - \bar{\mathbf{x}})$; this subtracts the mean and divides by the standard deviation. The resulting distance does not put extra emphasis on the variables with greater variance.

The variables can be correlated with one another so that the scatter diagram is tilted.



To account for this, we take the Mahalanobis transformation $\mathbf{zm}_i = \mathbf{S}^{-1/2}(\mathbf{x}_i - \bar{\mathbf{x}})$, where $\mathbf{S}^{1/2}$ is the square root matrix which has the property it is symmetric and $\mathbf{S}^{1/2} \mathbf{S}^{1/2} = \mathbf{S}$ (more on this later). The square root matrix is not a diagonal matrix like the standard deviation matrix, and takes into account the covariance between variables. The statistical length of an observation is the length of the Mahalanobis vector.

The $p \times p$ covariance matrix \mathbf{S} summarizes all that we know about the shape of the scatter of points around the mean. When there are many variables (p is large), \mathbf{S} has so many numbers in it that it too needs to be summarized. If we think of our scatterplot as a football  in a p-dimensional space, we would like to know how long the football compared to how round and in which direction it points. This leads to eigenvalues and eigenvectors of \mathbf{S} . See below.

4. Data Decompositions

Y	X_1	X_2
6	1	2
4	2	5
2	6	-3

Can we express Y in terms of X_1 and X_2 ?

$$1\beta_1 + 2\beta_2 = 6$$

Equations with “row focus”: $2\beta_1 + 5\beta_2 = 4$

$$6\beta_1 - 3\beta_2 = 2$$

Equations with “column focus”:
$$\begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} \beta_1 + \begin{bmatrix} 2 \\ 5 \\ -3 \end{bmatrix} \beta_2 = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}$$

Numbers can be decomposed into products in multiple ways: a positive number can be decomposed into the product of its square roots: $X = \sqrt{X} \sqrt{X}$, but it can also be decomposed into the product of a number and one plus that number: $Y(Y+1)=X \Rightarrow Y^2+2 \frac{1}{2} Y + \frac{1}{2}^2 - \frac{1}{2}^2 X \Rightarrow (Y+ \frac{1}{2})^2=X + \frac{1}{4} \Rightarrow Y=\sqrt{X+1/4} - \frac{1}{2}$. For example, $0.1 = .316 \times .316 = .09 \times 1.09$. Similarly, a matrix can be decomposed into products of matrices.

Elimination: subtract a multiple of row i from row k with the objective of putting 0’s everywhere below a pivot element.

By the process of elimination, we can decompose a data matrix X into the product of a lower triangular matrix and an upper triangular matrix.

We will work on the data matrix $X = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 6 & -3 \end{bmatrix}$ using elimination. First, subtract 2

times the first row from the second. Record both a matrix that represents what we have done and

the consequences of doing it: $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 6 & -3 \end{bmatrix}$.

The second step of elimination is to change 6 to 0: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 6 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 0 & -15 \end{bmatrix}$.

At this point we have zeroes below the pivot in row 1, so we switch to row 2 and begin with the pivot 1.

The third step of elimination is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -15 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$. At this point the adjusted data matrix

is an upper triangular matrix - elements on or above the diagonal are non-zero while all the elements below the diagonal are zero.

If we string together the matrices representing the elimination steps we get

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 6 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -15 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 6 & -15 & 1 \end{bmatrix}.$$

The product of $L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 6 & -15 & 1 \end{bmatrix}$ and $U = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ is $X = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 6 & -3 \end{bmatrix}$. That is, $X=LU$ where L is a lower triangular matrix and U is an upper triangular matrix.

Theorem: Every matrix X can be factored into the product of lower and upper triangular matrices: $X=LU$.

This is most basic of all decompositions, but there are others that are more important in statistics:

Spectral Decomposition: If S is a symmetric matrix, then it can be decomposed into $P\Lambda P'$, where Λ is a diagonal matrix and P is an orthogonal matrix. If S is positive definite, then Λ is positive.

Singular Value Decomposition: every X $n \times p$ matrix can be decomposed into QDV' where Q is a $n \times n$ orthogonal matrix, D is a $n \times p$ positive diagonal matrix, and V is a $p \times p$ orthogonal matrix.

Cholesky Decomposition: If S is a symmetric, positive definite matrix, then it can be decomposed into LL' where L is a lower triangular matrix with positive elements along the diagonal.

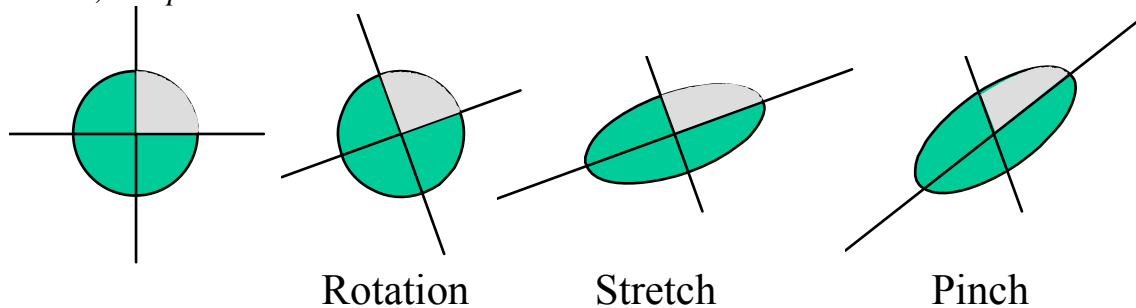
5. Algebra of Square Matrices

From this point forward assume that A is square. You could think of A as a covariance matrix.

Trace $\text{tr}(A) = \sum_i a_{ii}$ sum along diagonal of square matrix

Theorem: $\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$, $\text{tr}(kA) = k\text{tr}(A)$, $\text{tr}(XY) = \text{tr}(YX)$ if X is $n \times p$ and Y is $p \times n$, $\text{tr}(x'y) = \text{tr}(yx')$.

A causes a transformation of a vector x into y : $Ax=y$. This can be thought of as a rotation, stretch, and pinch.



Is there a reverse rotation, stretch and pinch that returns you to where you began?

Inverse: B is the inverse matrix of a square matrix A if and only if $AB=I=BA$. Notationally, we write the inverse A^{-1} .

Example: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $A^{-1} = \frac{1}{(ad-bc)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Theorem: If A^{-1} exists, then it is unique.

Theorem: If A,B are square, then $(AB)^{-1}=B^{-1}A^{-1}$.

Quadratic form in x: $\mathbf{x}'A\mathbf{x}$ where A is a square symmetric matrix.

A is positive definite if all non-trivial quadratic forms are positive: $\mathbf{x}'A\mathbf{x}>0$ for all $\mathbf{x}\neq 0$.

Example 2x2: Let $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$, then $\mathbf{x}'A\mathbf{x}=ax_1^2+2bx_1x_2+cx_2^2$. Complete the square to write this

$\mathbf{x}'A\mathbf{x}=a(x_1^2+2b/ax_1x_2)+dx_2^2= a(x_1^2+2b/ax_1x_2+(b/ax_2)^2)-b^2/ax_2^2+dx_2^2=a(x_1+b/ax_2)^2+(ad-b^2)/a x_2^2$.
Hence if A is positive definite $a>0$ and $ad-b^2>0$. By completing the square with x_2 , it is also true that $d>0$.

Theorem: $\mathbf{x}'S\mathbf{x}=\mathbf{x}'(HX)'(HX)\mathbf{x}/(n-1)=\mathbf{y}'\mathbf{y}/(n-1)\geq 0$ where $\mathbf{y}=HX\mathbf{x}$. Hence the covariance matrix is positive-semidefinite.

Theorem: $\text{tr}\left(\sum_i \mathbf{x}_i' A \mathbf{x}_i\right) = \text{tr}\left(A \sum_i \mathbf{x}_i \mathbf{x}_i'\right)$.

Determinants $|x_{11}|=x_{11}$, $\begin{vmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{vmatrix} = x_{11}x_{22} - x_{12}x_{21}$

Minor $|M_{ij}|$ = determinant of X with ith row and jth column deleted.

Cofactor $|C_{ij}|=(-1)^{i+j}|M_{ij}|$

Determinant via Laplace expansion along ith row: $|X|=x_{i1}|C_{i1}|+x_{i2}|C_{i2}|+\dots+x_{im}|C_{im}|$

Theorem: $|AB|=|A| |B|$ if A and B are square.

A square $p \times p$ matrix X is nonsingular if its rows are independent vectors: $\mathbf{a}'X=0$ implies that $\mathbf{a}=0$, or equivalently that $\text{rank}(X)=p$.

Three Elementary Row Transformations:

1) T_{ij} Interchange rows i and j, e.g. $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} X$

2) $T_i(k)$ Multiply row i by scalar k, e.g. $\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix} X$

3) $T_{ij}(k)$ Add row j times k to row i e.g. $\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} X$

X is equivalent to Y, denoted $X \sim Y$, if X can be transformed into Y by elementary row transformations. Elementary row transformations do not change the rank of a matrix. By elementary row transformations X can be transformed into one of the following four matrices:

$I_m, [I_m \ 0], \begin{bmatrix} I_m \\ 0 \end{bmatrix}$, or $\begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}$. If $X \sim \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}$ then the $\text{rank}(X)=m$.

Theorem: T_{ij} changes sign of determinant; $T_i(k)$ multiples determinant by k, and $T_{ij}(k)$ does not change the determinant of a square matrix.

If $X \sim I$ then there exists a sequence of elementary row transformations: $T_q T_{q-1} \dots T_2 T_1 X = I$. Since I has rank p, X has rank p and is nonsingular. Multiply both sides by X^{-1} to get $X^{-1} = T_q T_{q-1} \dots T_2 T_1 I$. Since $|I|=+1$, $|X|$ cannot be zero.

Theorem: The following are equivalent: $|X| \neq 0$, X is nonsingular, and X^{-1} exists.

Eigenvalues and Eigenvectors

A $p \times p$ matrix S transforms \mathbf{x} into a stretched version of \mathbf{x} (with no rotation) when $S\mathbf{x} = \lambda\mathbf{x}$, for some scalar λ . A nontrivial solution exists if $(S - \lambda I)\mathbf{x} = 0$ for $\mathbf{x} \neq 0$. This says the columns of $S - \lambda I$ are dependent but then $|S - \lambda I| = 0$. This is a polynomial equation that defines the eigenvalue λ .

Example: eigenvalues of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$: $\begin{vmatrix} a & b \\ c & d \end{vmatrix} - \lambda \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = \begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = (a - \lambda)(d - \lambda) - bc =$

$\lambda^2 - (a + d)\lambda + ad - bc = 0 = \lambda^2 - \text{tr}(S)\lambda + |S| = 0$. There will be two eigenvalues (in general, p). The sum of the eigenvalues equals the trace of the matrix $\text{tr}(S)$ and the product of the eigenvalues equals the determinant of the matrix $|S|$. This is true in general.

For each eigenvalue, λ_i , there is a direction, eigenvector \mathbf{e}_i , such that $S\mathbf{e}_i = \lambda_i\mathbf{e}_i$ and we standardize these to have unit length, $\mathbf{e}_i'\mathbf{e}_i = 1$. Put these eigenvectors side-by-side into a matrix P .

Theorem: if $S = S'$ then $P'P = I$. That is, the eigenvectors are orthonormal.

Proof: $S\mathbf{e}_i = \lambda_i\mathbf{e}_i$ implies $\mathbf{e}_j'S\mathbf{e}_i = \lambda_i\mathbf{e}_j'\mathbf{e}_i$. By similar reasoning and using symmetry, $\lambda_i\mathbf{e}_j'\mathbf{e}_i = \lambda_j\mathbf{e}_j'\mathbf{e}_i$. Since $\lambda_i \neq \lambda_j$, this implies that $\mathbf{e}_j'\mathbf{e}_i = 0$. The length of the eigenvectors equals 1 so that takes care of the diagonal of I . QED

Spectral Decomposition of symmetric square matrix S

$S\mathbf{e}_i = \lambda_i\mathbf{e}_i$ is equivalent to $SP = P\Lambda$ where Λ is a diagonal matrix with eigenvalues along the diagonal. Multiply on the right by P' to get $SPP' = P\Lambda P'$ or (using the above theorem), $S = P\Lambda P' = \sum_i \lambda_i \mathbf{e}_i \mathbf{e}_i'$. That is, a symmetric square matrix S can be expressed as a combination of elements built from the orthonormal eigenvectors. If we order the eigenvalues from largest (λ_1) to smallest (λ_m), then if we used just the first few elements in the sum $\sum_i \lambda_i \mathbf{e}_i \mathbf{e}_i'$ we get the best approximation of S possible.

Recall that the $p \times p$ covariance matrix S can be pretty big in a multivariate study; a survey with 20 questions has a covariance matrix with 210 unique numbers. We might try to capture the essence of what is going on by looking at the 4 or 5 directions suggested by the football-shaped scatterplot that contain the most information. The 4 or 5 largest eigenvalues tells us the ones with the most information and their eigenvectors tell us the direction the football points.

Theorem: S is positive definite if and only if the eigenvalues of S are all positive.

Proof: $\mathbf{x}'S\mathbf{x} = \mathbf{x}'P\Lambda P'\mathbf{x}$ by spectral decomposition. Let $\mathbf{y} = P'\mathbf{x}$, then $\mathbf{x}'S\mathbf{x} = \sum \lambda_i y_i^2 > 0$ if all eigenvalues are positive. QED

Singular-Value Decomposition

Spectral decomposition applies only to symmetric square matrices. Suppose that from the data matrix X we compute the $p \times p$ square matrix $X'X$ and the $n \times n$ square matrix XX' . Use the spectral decomposition to write $X'X = V\Lambda V'$ where V is $p \times p$ and write $XX' = U\Lambda_0 U'$ where U is $n \times n$. Define the matrix $\Lambda^{1/2}$ to be the diagonal matrix with elements equal to the square root of

the eigenvalues in Λ . Compute $Y=U\Lambda^{1/2}V'$. $Y'Y=(U\Lambda^{1/2}V')'U\Lambda^{1/2}V'=V\Lambda^{1/2}U'U\Lambda^{1/2}V'=V\Lambda V'=X'X$ hence $X=U\Lambda^{1/2}V'$. Thus any data matrix is really composed of a collection of uncorrelated variables with unit variances (U), a selection of p of these stretched or shrunk ($\Lambda^{1/2}$) and rotation (V'). See below. This is called the Singular-Value Decomposition.

