1. Background Material from Linear Algebra

It is assumed that you are familiar with the definition of a matrix and how to multiply vectors and matrices.

A vector or matrix is said to be $n \times m$ if it has n rows and m columns.

A *vector* or *column vector* will refer to an $n \times 1$ array:

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

The *transpose* of a vector v, denoted by v' or v^T, is obtained by writing v as a *row vector*:

$$\text{if} \quad v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad \text{then} \quad v' = [v_1 \; v_2 \; \cdots \; v_3]$$

In statistics texts, the prime notation is more common.

The transpose of a matrix A, denoted by A' or A^T, is obtained by interchanging the rows and columns of A. That is, the first column of A becomes the first row of A', the second column of A becomes the second row of A', and so on. For example,

$$\text{if} \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \quad \text{then} \quad A' = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix}$$

The *identity matrix* I or I_n is a square matrix having ones on the main diagonal and zeroes elsewhere:

$$I = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

For any square matrix A,

$$AI = IA = A$$

With each *square* matrix A there is associated a unique scalar called the *determinant* of A, denoted by $|A|$.

The *rank* of a matrix A, denoted by $Rank(A)$ or $r(A)$, is the number of linearly independent rows (or colunms) that A has. The number of linearly independent rows is always the same as the number of linearly independent columns.

The *trace* of an $n \times n$ matrix A, denoted by $Tr(A)$, is the sum of its diagonal entries:

$$Tr(A) = \sum_{i=1}^{n} a_{ii}$$

If the determinant of a matrix A is not zero, there exists a unique *inverse* of A, denoted by A^{-1}, with the property that:

$$AA^{-1} = A^{-1}A = I$$

In general, multiplication of matrices is not commutative:

$$AB \neq BA \quad \text{(in most cases)}$$

An $n \times n$ matrix A with the property that:

$$a_{ij} = a_{ji} \quad \forall i, j \leq n$$

is called *symmetric.*

An $n \times n$ matrix A is said to be *idempotent* if

$$A^2 = A$$

If x and y are vectors with n elements, and A is an $n \times n$ matrix, an expression of the form

$$x'Ax$$

is called a *quadratic form.*

An expression of the form

$$x'Ay$$

is called a *bilinear form.*

2. The Variance-Covariance Matrix

Suppose

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

is a vector of jointly distributed random variables, the *variance-covariance matrix* V of Y is defined as:

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}$$

That is,

$$v_{ii} = \sigma_i^2 = Var(Y_i) \quad \text{and} \quad v_{ij} = \sigma_{ij} = Cov(Y_i, Y_j)$$

where

$$Var(Y_i) = E(Y_i^2) - [E(Y_i)]^2 \quad \text{and} \quad Cov(Y_i, Y_j) = E(Y_iY_j) - E(Y_i)E(Y_j)$$

Note that V is symmetric since $Cov(Y_i, Y_j)$ is always equal to $Cov(Y_j, Y_i)$.

3. Linear Combinations

If

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

is a vector of random variables with expected value

$$E(Y) = \mu = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

and variance-covariance matrix

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}$$

Any linear combination of the elements of Y can be written in the form $t'Y$ for some vector t coefficients. If

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

is a vector of random variables with expected value

$$E(Y) = \mu = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

and variance-covariance matrix

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}$$

The expected value and variance of $t'Y$ is:

$$E(t'Y) = t'E(Y) = t'\mu \quad \text{and} \quad Var(t'Y) = t'Vt$$

Example. *Suppose*

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

is a vector of random variables with expected value

$$E(Y) = \mu = \begin{bmatrix} E(Y_1) \\ E(Y_2) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

and variance-covariance matrix

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

Find the expected value and variance of $U = 3Y_1 - 4Y_2$.

In this case,

$$t = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$$

so

$$E(t'Y) = t'\mu = [3 \ -4] \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = 3\mu_1 - 4\mu_2$$

and

$$Var(t'Y) = t'Vt = [3 \ -4] \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 3 \\ -4 \end{bmatrix}$$

$$= [(3\sigma_1^2 - 4\sigma_{12}) \ (3\sigma_{12} - 4\sigma_2^2)] \begin{bmatrix} 3 \\ -4 \end{bmatrix}$$

$$= 9\sigma_1^2 - 12\sigma_{12} - 12\sigma_{12} + 16\sigma_2^2$$

$$= 9\sigma_1^2 - 24\sigma_{12} + 16\sigma_2^2$$

More generally, a vector of linear combinations of the elements of Y can be written in the form $A'Y$ for a matrix A of coefficients. The expected value and variance-covariance matrix of $A'Y$ is:

$$E(A'Y) = A'E(Y) = A'\mu \quad \text{and} \quad V = A'VA$$

Example. *This time suppose*

$$U = \begin{bmatrix} Y_1 + Y_2 \\ Y_1 - Y_2 \end{bmatrix} = A'Y$$

where

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Then

$$E(U) = E(A'Y) = A'\mu = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_1 + \mu_2 \\ \mu_1 - \mu_2 \end{bmatrix}$$

and the variance-covariance matrix of $A'Y$ is:

$$V_U = A'VA' = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 + \sigma_{12} & \sigma_{12} + \sigma_2^2 \\ \sigma_1^2 - \sigma_{12} & \sigma_{12} - \sigma_2^2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 + 2\sigma_{12} + \sigma_2^2 & \sigma_1^2 - \sigma_2^2 \\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 - 2\sigma_{12} + \sigma_2^2 \end{bmatrix}$$

$$= \begin{bmatrix} Var(Y_1 + Y_2) & Cov(Y_1 + Y_2, Y_1 - Y_2) \\ Cov(Y_1 + Y_2, Y_1 - Y_2) & Var(Y_1 - Y_2) \end{bmatrix}$$

From this result we see that, for example,

$$Cov(Y_1 + Y_2, Y_1 - Y_2) = \sigma_1^2 - \sigma_2^2 = Var(Y_1) - Var(Y_2)$$

4. The Multivariate Normal Distribution

If Y has a univariate normal distribution $N(\mu, \sigma)$, the density function of Y is:

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

If

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

has a multivariate normal distribution with mean vector

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

and nonsingular variance-covariance matrix

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}$$

then the joint density function of Y is:

$$f(y_1, y_2, \ldots, y_n) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)'V^{-1}(x-\mu)\right)$$

Note that if all covariances are zero,

$$V = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} \quad \text{and} \quad V^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n}^2 \end{bmatrix}$$

Because the determinant of a diagonal matrix is simply the product of the entries on the diagonal, $|V|^{1/2}$ reduces to:

$$\sqrt{|\Pi_{i=1}^n \sigma_i^2|} = \Pi_{i=1}^n \sigma_i$$

and the exponent reduces to:

$$-\frac{1}{2}(y-\mu)'V^{-1}(y-\mu) = \Pi_{i=1}^n \frac{(y_i-\mu_i)^2}{\sigma_i^2}$$

so the joint density function factors into the product

$$f(y_1,\ldots,y_n) = \Pi_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2}\frac{(y_i-\mu_i)^2}{\sigma_i^2}\right)$$

which is just the product of the n univariate densities for the individual Y_i. This establishes that having all covariances equal to zero is a sufficient condition for the components Y_i of a multivariate normal random vector Y to be independently distributed. This is not true for an arbitrary distribution.

5. Distribution of Quadratic Forms

Suppose Y has a multivariate normal distribution with mean vector μ and variance-covariance matrix V. Very often we are interested in quadratic functions of the Y_i. As it turns out, these can be represented as quadratic forms, that is, expressions of the form

$$Y'AY$$

for some matrix A. The following very general theorem gives necessary and sufficient conditions for a quadratic form to have a *chi-square* (denoted χ^2) distribution:

Theorem (Distribution of a Quadratic Form)**.** *If Y has a multivariate normal distribution with mean vector μ and nonsingular variance-covariance matrix V, and A is an arbitrary matrix, the quadratic form*

$$Y'AY$$

has a χ^2 distribution with degrees of freedom equal to the rank of A if and only if AV is idempotent.

Note: If $\mu = 0$, that is, the vector of means is zero, then $Y'AY$ has a *central* χ^2 distribution, which is the one tabulated in the back of the text. Otherwise, $Y'AY$ has what is called a *noncentral* χ^2 distribution. Many standard hypothesis testing procedures make use of test statistic that has a central χ^2 distribution when a certain hypothesis is true, and a noncentral χ^2 distribution otherwise.

Example. *Suppose Z_i, $i = 1, \cdots, n$ is a collection if IID (independent, identically distributed) random variables having a standard normal distribution $N(0,1)$. Then the vector $Z = (Z_1, \ldots, Z_n)$ has a multivariate normal distribution with $V = I$. Consider the sum of squares:*

$$\sum_{i=1}^{n} Z_i^2 = Y'IY$$

In this case, the matrix A of the quadratic form is $A = I$, so $AV = I^2$ and $(AV)^2 = I^4 = I = AV$. Since AV is idempotent and Rank$(A) = n$, the sum of the squares of the Z_i has a χ^2 distribution with n degrees of freedom.

Example. *Now suppose the Z_i are IID $N(0,\sigma)$, so $V = \sigma^2 I$. We want to consider the distribution of*

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

As we have seen,

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

can be written as

$$\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2$$

which can be written as a quadratic form

$$Y'\left(I - \frac{1}{n}J\right)Y$$

where J is a matrix with every entry equal to 1:

$$J = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

and J^2 is a matrix with every entry equal to n:

$$J^2 = \begin{bmatrix} n & n & \cdots & n \\ n & n & \cdots & n \\ \vdots & \vdots & \cdots & \vdots \\ n & n & \cdots & n \end{bmatrix} \quad = \quad nJ$$

so if

$$A = \frac{1}{\sigma^2}\left(I - \frac{1}{n}J\right) \quad \textit{and} \quad V = \sigma^2 I$$

then

$$AV = I - \frac{1}{n}J \quad \textit{and} \quad (AV)^2 = \left(I - \frac{1}{n}J\right)^2$$

$$= \left(I - \frac{1}{n}J\right)\left(I - \frac{1}{n}J\right) = I^2 - \frac{2}{n}J - \frac{1}{n^2}J^2$$

but $I^2 = I$, and $J^2 = nJ$ so this becomes

$$(AV)^2 = I^2 - \frac{2}{n}J - \frac{1}{n^2}J^2 = I - \frac{2}{n}J + \frac{1}{n}J = I - \frac{1}{n}J = AV$$

If you add the first $n-1$ columns of the matrix $I - 1/nJ$, you obtain the last row. In fact, it can be shown that the rank of this idempotent matrix is $n-1$, so the quadratic form $Y'AY$ has a chi-square distribution with $n-1$ degrees of freedom.

The appearance of $n-1$ in formulas for the sample standard deviation is one of the more mysterious aspects of an introductory statistics course, but considering the above example it should be clear that the $n-1$ factor arises from the rank of a matrix.