# CHAPTER 0

# Basic Prerequisite Knowledge

Readers need some of the knowledge contained in a basic course in statistics to tackle regression. We summarize some of the main requirements very briefly in this chapter. Also useful is a pocket calculator capable of getting sums of squares and sums of products easily. Excellent calculators of this type cost about $25–50 in the United States. Buy the most versatile you can afford.

## 0.1. DISTRIBUTIONS: NORMAL, *t*, AND *F*

### Normal Distribution

The normal distribution occurs frequently in the natural world, either for data "as they come" or for transformed data. The heights of a large group of people selected randomly will look normal in general, for example. The distribution is symmetric about its mean $\mu$ and has a standard deviation $\sigma$, which is such that practically all of the distribution (99.73%) lies inside the range $\mu - 3\sigma \leq x \leq \mu + 3\sigma$. The frequency function is

$$f(x) = \frac{1}{\sigma(2\pi)^{1/2}} \exp\left(\frac{(x-\mu)^2}{-2\sigma^2}\right), \qquad -\infty \leq x \leq \infty. \tag{0.1.1}$$

We usually write that $x \sim N(\mu, \sigma^2)$, read as "$x$ is normally distributed with mean $\mu$ and variance $\sigma^2$." Most manipulations are done in terms of the *standard normal* or *unit normal* distribution, $N(0, 1)$, for which $\mu = 0$ and $\sigma = 1$. To move from a general normal variable $x$ to a standard normal variable $z$, we set

$$z = (x - \mu)/\sigma. \tag{0.1.2}$$

A standard normal distribution is shown in Figure 0.1 along with some properties useful in certain regression contexts. All the information shown is obtainable from the normal table in the Tables section. Check that you understand how this is done. Remember to use the fact that the total area under each curve is 1.

### Gamma Function

The gamma function $\Gamma(q)$, which occurs in Eqs. (0.1.3) and (0.1.4), is defined as an integral in general:

$$\Gamma(q) = \int_0^\infty e^{-x} x^{q-1}\, dx.$$

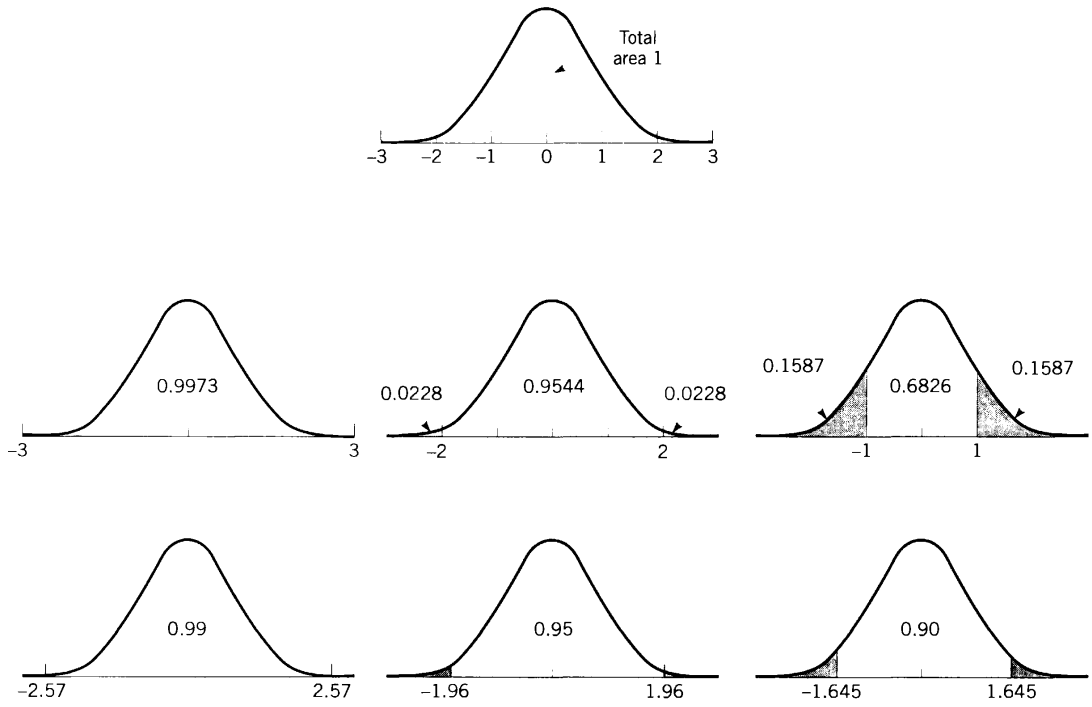


**Figure 0.1.** The standard (or unit) normal distribution $N(0, 1)$ and some of its properties.

However, it is easier to think of it as a generalized factorial with the basic property that, for any $q$,

$$\Gamma(q) = (q-1)\Gamma(q-1)$$
$$= (q-1)(q-2)\Gamma(q-2),$$

and so on. Moreover,

$$\Gamma(\tfrac{1}{2}) = \pi^{1/2} \quad \text{and} \quad \Gamma(1) = 1.$$

So, for the applications of Eqs. (0.1.3) and (0.1.4), where $\nu$, $m$, and $n$ are integers, the gamma functions are either simple factorials or simple products ending in $\pi^{1/2}$.

***Example 1***

$$\Gamma(5) = 4 \times \Gamma(4) = 4 \times 3 \times \Gamma(3) = 4 \times 3 \times 2 \times \Gamma(2)$$
$$= 4 \times 3 \times 2 \times 1 \times \Gamma(1) = 24.$$

***Example 2***

$$\Gamma(\tfrac{5}{2}) = \tfrac{3}{2} \times \Gamma(\tfrac{3}{2}) = \tfrac{3}{2} \times \tfrac{1}{2} \times \Gamma(\tfrac{1}{2}) = 3\pi^{1/2}/4.$$

### *t*-Distribution

There are many $t$-distributions, because the form of the curve, defined by

$$f_\nu(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{(\nu\pi)^{1/2}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} \qquad (-\infty \le t \le \infty), \tag{0.1.3}$$

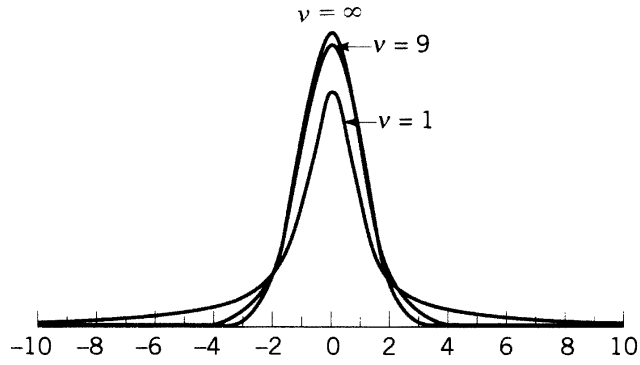


**Figure 0.2.** The $t$-distributions for $\nu = 1, 9, \infty$; $t(\infty) = N(0, 1)$.

depends on $\nu$, the number of degrees of freedom. In general, the $t(\nu)$ distribution looks somewhat like a standard (unit) normal but is "heavier in the tails," and so lower in the middle, because the total area under the curve is 1. As $\nu$ increases, the distribution becomes "more normal." In fact, $t(\infty)$ *is* the $N(0, 1)$ distribution, and, when $\nu$ exceeds about 30, there is so little difference between $t(\nu)$ and $N(0, 1)$ that it has become conventional (but not mandatory) to use the $N(0, 1)$ instead. Figure 0.2 illustrates the situation. A two-tailed table of percentage points is given in the Tables section.

### *F*-Distribution

The $F$-distribution depends on two separate degrees of freedom $m$ and $n$, say. Its curve is defined by

$$f_{m,n}(F) = \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{m/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{F^{m/2-1}}{(1+mF/n)^{(m+n)/2}} \qquad (F \geq 0). \tag{0.1.4}$$

The distribution rises from zero, sometimes quite steeply for certain $m$ and $n$, and reaches a peak, falling off very skewed to the right. See Figure 0.3. Percentage points for the upper tail levels of 10%, 5%, and 1% are in the Tables section.

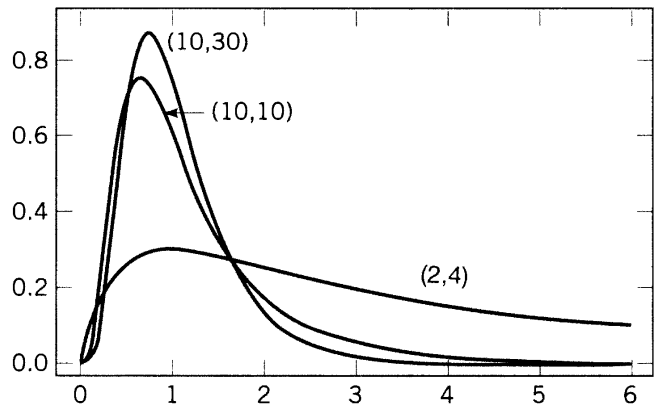


**Figure 0.3.** Some selected $f(m, n)$ distributions.

The $F$-distribution is usually introduced in the context of testing to see whether two variances are equal, that is, the null hypothesis that $H_0$: $\sigma_1^2/\sigma_2^2 = 1$, versus the alternative hypothesis that $H_1$: $\sigma_1^2/\sigma_1^2 \neq 1$. The test uses the statistic $F = s_1^2/s_2^2, s_1^2$ and $s_2^2$ being statistically independent estimates of $\sigma_1^2$ and $\sigma_2^2$, with $\nu_1$ and $\nu_2$ degrees of freedom (df), respectively, and depends on the fact that, if the two samples that give rise to $s_1^2$ and $s_2^2$ are independent and normal, then $(s_1^2/s_2^2)/(\sigma_1^2/\sigma_2^2)$ follows the $F(\nu_1, \nu_2)$ distribution. Thus *if* $\sigma_1^2 = \sigma_2^2$, $F = s_1^2/s_2^2$ follows $F(\nu_1, \nu_2)$. When given in basic statistics courses, this is usually described as a two-tailed test, which it usually is. In regression applications, it is typically a one-tailed, upper-tailed test. This is because regression tests always involve putting the "$s^2$ that could be too big, but cannot be too small" at the top and the "$s^2$ that we think estimates the true $\sigma^2$ well" at the bottom of the $F$-statistic. In other words, we are in the situation where we test $H_0$: $\sigma_1^2 = \sigma_2^2$ versus $H_1$: $\sigma_1^2 > \sigma_2^2$.

## 0.2. CONFIDENCE INTERVALS (OR BANDS) AND *t*-TESTS

Let $\theta$ be a parameter (or "thing") that we want to estimate. Let $\hat{\theta}$ be an estimate of $\theta$ ("estimate of thing"). Typically, $\hat{\theta}$ will follow a normal distribution, either exactly because of the normality of the observations in $\hat{\theta}$, or approximately due to the effect of the Central Limit Theorem. Let $\sigma_{\hat{\theta}}$ be the standard deviation of $\hat{\theta}$ and let $\text{se}(\hat{\theta})$ be the standard error, that is, the estimated standard deviation, of $\hat{\theta}$ ("standard error of thing"), based on $\nu$ degrees of freedom. Typically we get $\text{se}(\hat{\theta})$ by substituting an estimate (based on $\nu$ degrees of freedom) of an unknown standard deviation into the formula for $\sigma_{\hat{\theta}}$.

1. A $100(1 - \alpha)\%$ confidence interval (CI) for the parameter $\theta$ is given by

$$\hat{\theta} \pm t(\nu, 1 - \alpha/2)\text{se}(\hat{\theta}) \tag{0.2.1}$$

where $t_\nu(1 - \alpha/2)$ is the percentage point of a $t$-variable with $\nu$ degrees of freedom (df) that leaves a probability $\alpha/2$ in the upper tail, and so $1 - \alpha/2$ in the lower tail. A two-tailed table where these percentage points are listed under the heading of $2(\alpha/2) = \alpha$ is given in the Tables section. Equation (0.2.1) in words is

$$\left\{\begin{matrix}\text{Estimate}\\ \text{of thing}\end{matrix}\right\} \pm \left\{\begin{matrix}\text{A } t \text{ percentage point}\\ \text{leaving } \alpha/2 \text{ in the}\\ \text{upper tail, based on}\\ \nu \text{ degrees of freedom}\end{matrix}\right\} \left\{\begin{matrix}\text{Standard error}\\ \text{of estimate}\\ \text{of thing}\end{matrix}\right\}. \tag{0.2.2}$$

2. To test $\theta = \theta_0$, where $\theta_0$ is some specified value of $\theta$ that is presumed to be valid (often $\theta_0 = 0$ in tests of regression coefficients) we evaluate the statistic

$$t = \frac{\hat{\theta} - \theta_0}{\text{se}(\hat{\theta})} \tag{0.2.3}$$

or, in words,

$$t = \frac{\left\{\begin{matrix}\text{Estimate}\\ \text{of thing}\end{matrix}\right\} - \left\{\begin{matrix}\text{Postulated or test}\\ \text{value of thing}\end{matrix}\right\}}{\left\{\begin{matrix}\text{Standard error of}\\ \text{estimate of thing}\end{matrix}\right\}}. \tag{0.2.4}$$

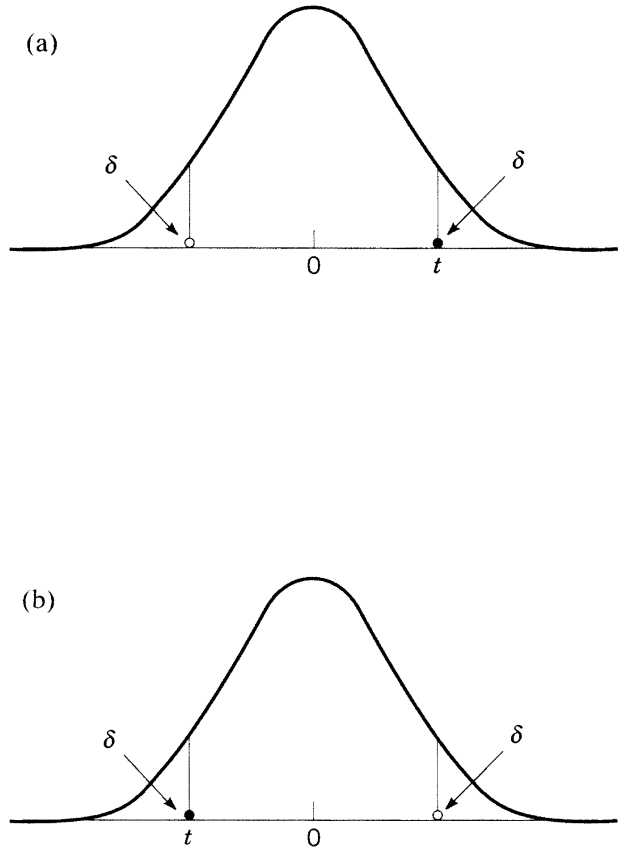


**Figure 0.4.** Two cases for a $t$-test. ($a$) The observed $t$ is positive (black dot) and the upper tail area is $\delta$. A two-tailed test considers that this value could just as well have been negative (open "phantom" dot) and quotes "a two-tailed $t$-probability of $2\delta$." ($b$) The observed $t$ is negative; similar argument, with tails reversed.

This "observed value of $t$" (our "dot") is then placed on a diagram of the $t(\nu)$ distribution. [Recall that $\nu$ is the number of degrees of freedom on which $\text{se}(\hat{\theta})$ is based and that is the number of df in the estimate of $\sigma^2$ that was used.] The tail probability beyond the dot is evaluated and doubled for a two-tail test. See Figure 0.4 for the probability $2\delta$. It is conventional to ask if the $2\delta$ value is "significant" or not by concluding that, if $2\delta < 0.05$, $t$ is significant and the idea (or hypothesis) that $\theta = \theta_0$ is unlikely and so "rejected," whereas if $2\delta > 0.05$, $t$ is nonsignificant and we "do not reject" the hypothesis $\theta = \theta_0$. The alternative hypothesis here is $\theta \neq \theta_0$, a two-sided alternative. Note that the value 0.05 is not handed down in holy writings, although we sometimes talk as though it is. Using an "alpha level" of $\alpha = 0.05$ simply means we are prepared to risk a 1 in 20 chance of making the wrong decision. If we wish to go to $\alpha = 0.10$ (1 in 10) or $\alpha = 0.01$ (1 in 100), that is up to us. Whatever we decide, we should remain consistent about this level throughout our testing.

However, it is pointless to agonize too much about $\alpha$. A journal editor who will publish a paper describing an experiment if $2\delta = 0.049$, but will not publish it if $2\delta = 0.051$ is placing a purely arbitrary standard on the work. (Of course, it is the editor's right to do that.) Such an attitude should not necessarily be imposed by experimenters on themselves, because it is too restrictive a posture in general. Promis-

**TABLE 0.1. Example Applications for Formulas (0.2.1)–(0.2.4)**

| Situation | $\theta$ | $\hat{\theta}$ | $\text{se}(\hat{\theta})$ |
|---|---|---|---|
| Straight line fit<br>$Y = \beta_0 + \beta_1 X + \varepsilon$ | $\beta_1$ | $b_1$ | $\frac{s}{S_{XX}^{1/2}}, S_{XX} = \Sigma(X_i - \overline{X})^2$ |
| | $\beta_0$ | $b_0$ | $s\left\{\frac{\Sigma X_i^2}{nS_{XX}}\right\}^{1/2}$ |
| Predicted response<br>$\hat{Y}_0 = b_0 + b_1 X_0$<br>at $X = X_0$ | $E(Y)$ at $X_0$ | $\hat{Y}_0$ | $s\left\{\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{S_{XX}}\right\}^{1/2}$ |

ing experimental leads need to be followed up, even if the arbitrary $\alpha$ standard has not been attained. For example, an $\alpha = 0.05$ person might be perfectly justified in following up a $2\delta = 0.06$ experiment by performing more experimental runs to further elucidate the results attained by the first set of runs. To give up an avenue of investigation merely because one experiment did not provide a significant result may be a mistake. The $\alpha$ value should be thought of as a guidepost, not a boundary.

In every application of formulas (0.2.1)–(0.2.4), we have to ask what $\theta$, $\hat{\theta}$, $\theta_0$, $\text{se}(\hat{\theta})$, and the $t$ percentage point are. The actual formulas we use are always the same. Table 0.1 contains some examples of $\theta$, $\hat{\theta}$, and $\text{se}(\hat{\theta})$ we shall subsequently meet. (The symbol $s$ replaces the $\sigma$ of the corresponding standard deviation formulas.)

## 0.3. ELEMENTS OF MATRIX ALGEBRA

### Matrix, Vector, Scalar

A $p \times q$ matrix $\mathbf{M}$ is a rectangular array of numbers containing $p$ rows and $q$ columns written

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1q} \\ m_{21} & m_{22} & \dots & m_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ m_{p1} & m_{p2} & \dots & m_{pq} \end{bmatrix}.$$

For example,

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & 3 & 7 \\ -1 & 0 & 2 & 2 \\ 6 & 5 & -2 & 1 \end{bmatrix}$$

is a $3 \times 4$ matrix. The plural of *matrix* is *matrices*. A "matrix" with only one row is called a *row vector*: a "matrix" with only one column is called a *column vector*. For example, if

$$\mathbf{a}' = [1, 6, 3, 2, 1] \qquad \text{and} \qquad \mathbf{b} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix},$$

then $\mathbf{a}'$ is a row vector of length five and $\mathbf{b}$ is a column vector of length three. A $1 \times 1$ "vector" is an ordinary number or *scalar*.

### Equality

Two matrices are equal if and only if their dimensions are identical and they have exactly the same entries in the same positions. Thus a matrix equality implies as many individual equalities as there are terms in the matrices set equal.

### Sum and Difference

The sum (or difference) of two matrices is the matrix each of whose elements is the sum (or difference) of the corresponding elements of the matrices added (or subtracted). For example,

$$\begin{bmatrix} 7 & 6 & 9 \\ 4 & 2 & 1 \\ 6 & 5 & 3 \\ 2 & 1 & 4 \end{bmatrix} - \begin{bmatrix} 1 & 2 & 4 \\ -1 & 3 & -2 \\ 6 & 2 & 1 \\ 7 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 6 & 4 & 5 \\ 5 & -1 & 3 \\ 0 & 3 & 2 \\ -5 & 1 & 2 \end{bmatrix}.$$

The matrices must be of exactly the same dimensions for addition or subtraction to be carried out. Otherwise the operations are not defined.

### Transpose

The transpose of a matrix $\mathbf{M}$ is a matrix $\mathbf{M}'$ whose rows are the columns of $\mathbf{M}$ and whose columns are the rows of $\mathbf{M}$ in the same original order. Thus for $\mathbf{M}$ and $\mathbf{A}$ as defined above,

$$\mathbf{M}' = \begin{bmatrix} m_{11} & m_{21} & \dots & m_{p1} \\ m_{12} & m_{22} & \dots & m_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ m_{1q} & m_{2q} & \dots & m_{pq} \end{bmatrix},$$

$$\mathbf{A}' = \begin{bmatrix} 4 & -1 & 6 \\ 1 & 0 & 5 \\ 3 & 2 & -2 \\ 7 & 2 & 1 \end{bmatrix}.$$

Note that the transpose notation enables us to write, for example,

$$\mathbf{b}' = (-1, 0, 1) \quad \text{or alternatively} \quad \mathbf{b} = (-1, 0, 1)'.$$

*Note:* The parentheses around a matrix or vector can be square-ended or curved. Often, capital letters are used to denote matrices and lowercase letters to denote vectors. Boldface print is often used, but this is not universal.

### Symmetry

A matrix **M** is said to be *symmetric* if $\mathbf{M}' = \mathbf{M}$.

### Multiplication

Suppose we have two matrices, **A,** which is $p \times q$, and **B,** which is $r \times s$. They are *conformable* for the product $\mathbf{C} = \mathbf{AB}$ only if $q = r$. The resulting product is then a $p \times s$ matrix, the multiplication procedure being defined as follows: If

$$\mathbf{A} = \underset{p \times q}{\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1q} \\ a_{21} & a_{22} & \dots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pq} \end{bmatrix}}, \qquad \mathbf{B} = \underset{q \times s}{\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1s} \\ b_{21} & b_{22} & \dots & b_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ b_{q1} & b_{q2} & \dots & b_{qs} \end{bmatrix}},$$

then the product

$$\mathbf{AB} = \mathbf{C} = \underset{p \times s}{\begin{bmatrix} c_{11} & c_{12} & \dots & c_{1s} \\ c_{21} & c_{22} & \dots & c_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \dots & c_{ps} \end{bmatrix}}$$

is such that

$$c_{ij} = \sum_{l=1}^{q} a_{il} b_{lj};$$

that is, the entry in the $i$th row and $j$th column of **C** is the *inner product* (the element by element cross-product) of the $i$th row of **A** with the $j$th column of **B**. For example,

$$\underset{2 \times 3}{\begin{bmatrix} 1 & 2 & 1 \\ -1 & 3 & 0 \end{bmatrix}} \underset{3 \times 3}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 0 & -1 \\ -2 & 1 & 3 \end{bmatrix}}$$

$$= \begin{bmatrix} 1(1) + 2(4) + 1(-2) & 1(2) + 2(0) + 1(1) & 1(3) + 2(-1) + 1(3) \\ -1(1) + 3(4) + 0(-2) & -1(2) + 3(0) + 0(1) & -1(3) + 3(-1) + 0(3) \end{bmatrix}$$

$$= \underset{2 \times 3}{\begin{bmatrix} 7 & 3 & 4 \\ 11 & -2 & -6 \end{bmatrix}}.$$

We say that, in the product **AB,** we have *premultiplied* **B** by **A** or we have *postmultiplied* **A** by **B.** Note that, in general, **AB** and **BA,** even if both products are permissible (conformable), do not lead to the same result. In a matrix multiplication, the order in which the matrices are arranged is crucially important, whereas the order of the numbers in a scalar product is irrelevant.

When several matrices and/or vectors are multiplied together, the product should be carried out in the way that leads to the least work. For example, the product

$$\begin{matrix} \mathbf{W} & \mathbf{Z}' & \mathbf{y} \\ p \times p & p \times n & n \times 1 \end{matrix}$$

could be carried out as $(\mathbf{WZ}')\mathbf{y}$, or as $\mathbf{W}(\mathbf{Z}'\mathbf{y})$, where the parenthesized product is evaluated first. In the first case we would have to carry out $pn$ $p$-length cross-products and $p$ $n$-length cross-products; in the second case $p$ $p$-length and $p$ $n$-length, clearly a saving in effort.

## Special Matrices and Vectors

We define

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix},$$

a square $n \times n$ matrix with 1's on the diagonal and 0's elsewhere, as the *unit matrix* or *identity matrix*. This fulfills the same role as the number 1 in ordinary arithmetic. If the size of $\mathbf{I}_n$ is clear from the context, the subscript $n$ is often omitted. We further use $\mathbf{0}$ to denote a vector

$$\mathbf{0} = (0, 0, \dots, 0)'$$

or a matrix

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix},$$

all of whose values are zeros; the actual size of $\mathbf{0}$ is usually clear from the context. We also define

$$\mathbf{1} = (1, 1, \dots, 1)'$$

a vector of all 1's; the size of $\mathbf{1}$ is either specified or is clear in context. Note that $\mathbf{1}'\mathbf{1}$ = the squared length of the vector $\mathbf{1}$, but that $\mathbf{11}'$ is a square matrix, each entry of which is 1, with the number of rows and columns each equal to the length of $\mathbf{1}$.

### Orthogonality

A vector $\mathbf{a} = (a_1, a_2, \ldots, a_n)'$ is said to be *orthogonal* to a vector $\mathbf{b} = (b_1, b_2, \ldots, b_n)'$ if the sum of products of their elements is zero, that is, if

$$\sum_{i=1}^{n} a_i b_i = \mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a} = 0.$$

### Inverse Matrix

The inverse $\mathbf{M}^{-1}$ of a square matrix $\mathbf{M}$ is the unique matrix such that

$$\mathbf{M}^{-1}\mathbf{M} = \mathbf{I} = \mathbf{M}\mathbf{M}^{-1}.$$

The columns $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_n$ of an $n \times n$ matrix are *linearly dependent* if there exist constants $c_1, c_2, \ldots, c_n$, not all zero, such that

$$c_1\mathbf{m}_1 + c_2\mathbf{m}_2 + \cdots + c_n\mathbf{m}_n = \mathbf{0}$$

and similarly for rows. A square matrix, some of whose rows (or some of whose columns) are linearly dependent, is said to be *singular* and does not possess an inverse. A square matrix that is not singular is said to be *nonsingular* and can be inverted.

If $\mathbf{M}$ is symmetric, so is $\mathbf{M}^{-1}$.

### Obtaining an Inverse

The process of matrix inversion is a relatively complicated one and is best appreciated by considering an example. Suppose we wish to obtain the inverse $\mathbf{M}^{-1}$ of the matrix

$$\mathbf{M} = \begin{bmatrix} 3 & 4 & 5 \\ 1 & 2 & 6 \\ 7 & 1 & 9 \end{bmatrix}.$$

Let

$$\mathbf{M}^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}.$$

Then we must find $(a, b, c, \ldots, h, k)$ so that

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}\begin{bmatrix} 3 & 4 & 5 \\ 1 & 2 & 6 \\ 7 & 1 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

that is, so that

$$\begin{aligned} &3a + b + 7c = 1, \quad 3d + e + 7f = 0, \quad 3g + h + 7k = 0, \\ &4a + 2b + c = 0, \quad 4d + 2e + f = 1, \quad 4g + 2h + k = 0, \\ &5a + 6b + 9c = 0, \quad 5d + 6e + 9f = 0, \quad 5g + 6h + 9k = 1. \end{aligned}$$

Solving these three sets of three linear simultaneous equations yields

$$\mathbf{M}^{-1} = \begin{bmatrix} \frac{12}{103} & -\frac{31}{103} & \frac{14}{103} \\ \frac{33}{103} & -\frac{8}{103} & -\frac{13}{103} \\ -\frac{13}{103} & \frac{25}{103} & \frac{2}{103} \end{bmatrix} = \frac{1}{103} \begin{bmatrix} 12 & -31 & 14 \\ 33 & -8 & -13 \\ -13 & 25 & 2 \end{bmatrix}.$$

(Note the removal of a common factor, explained below.) In general, for an $n \times n$ matrix there will be $n$ sets of $n$ simultaneous linear equations. Accelerated methods for inverting matrices adapted specifically for use with electronic computers permit inverses to be obtained with great speed, even for large matrices. Working out inverse matrices "by hand" is obsolete, nowadays, except in simple cases (see below).

### Determinants

An important quantity associated with a square matrix is its *determinant*. Determinants occur naturally in the solution of linear simultaneous equations and in the inversion of matrices. For a $2 \times 2$ matrix

$$\mathbf{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

the determinant is defined as

$$\det \mathbf{M} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

For a $3 \times 3$ matrix

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix},$$

it is

$$a\begin{vmatrix} e & f \\ h & k \end{vmatrix} - b\begin{vmatrix} d & f \\ g & k \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix} = aek - afh - bdk + bfg + cdh - ceg.$$

Note that we expand by the first row, multiplying $a$ by the determinant of the matrix left when we cross out the row and column containing $a$, multiplying $b$ by the determinant of the matrix left when we cross out the row and column containing $b$, multiplying $c$ by the determinant of the matrix left when we cross out the row and column containing $c$. We also attach alternate signs $+$, $-$, $+$ to these three terms, counting from the top left-hand corner element: $+$ to $a$, $-$ to $b$, $+$ to $c$, and so on, alternately, if there were more elements in the first row.

In fact, the determinant can be written down as an expansion of *any* row or column by the same technique. The signs to be attached are counted $+ - + -$, and so on, from the top left corner element alternating either along row or column (but *not* diagonally). In other words the signs

$$\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$$

are attached and any row or column is used to write down the determinant. For example, using the second row we have

$$-d\begin{vmatrix} b & c \\ h & k \end{vmatrix} + e\begin{vmatrix} a & c \\ g & k \end{vmatrix} - f\begin{vmatrix} a & b \\ g & h \end{vmatrix}$$

to obtain the same result as before.

The same principle is used to get the determinant of any matrix. Any row or column is used for the expansion and we multiply each element of the row or column by:

1. Its appropriate sign, counted as above.
2. The determinant of the submatrix obtained by deletion of the row and column in which the element of the original matrix stands.

Determinants arise in the inversion of matrices as follows. The inverse $\mathbf{M}^{-1}$ may be obtained by first replacing each element $m_{ij}$ of the original matrix $\mathbf{M}$ by an element calculated as follows:

1. Find the determinant of the submatrix obtained by crossing out the row and column of $\mathbf{M}$ in which $m_{ij}$ stands.
2. Attach a sign from the $+ \; - \; + \; -$ count, as above.
3. Divide by the determinant of $\mathbf{M}$.

When all elements of $\mathbf{M}$ have been replaced, *transpose the resulting matrix*. The transpose will be $\mathbf{M}^{-1}$.

The reader might like to check these rules by showing that

$$\mathbf{M}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} d/D & -b/D \\ -c/D & a/D \end{bmatrix},$$

where $D = ad - bc$ is the determinant of $\mathbf{M}$; and that

$$\mathbf{Q}^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}^{-1} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & K \end{bmatrix},$$

where

$$A = (ek - fh)/Z, \quad B = -(bk - ch)/Z, \quad C = (bf - ce)/Z,$$
$$D = -(dk - fg)/Z, \quad E = (ak - cg)/Z, \quad F = -(af - cd)/Z,$$
$$G = (dh - eg)/Z, \quad H = -(ah - bg)/Z, \quad K = (ae - bd)/Z,$$

and where

$$Z = aek + bfg + cdh - afh - bdk - ceg$$

is the determinant of $\mathbf{Q}$. Note that, if $\mathbf{M}$ is symmetric (so that $b = c$), $\mathbf{M}^{-1}$ is also symmetric. Also, if $\mathbf{Q}$ is symmetric (so that $b = d$, $c = g$, $f = h$), then $\mathbf{Q}^{-1}$ is also symmetric because then $B = D$, $C = G$, and $F = H$.

The determinant is essentially a measure of the volume contained in a parallelepiped defined by the vectors in the rows (or columns) of the square matrix. See, for example, Schneider and Barker (1973, pp. 161–169).

If the square matrix is of the form $\mathbf{X}'\mathbf{X}$, dimension $p \times p$, say, the equation

$$\mathbf{u}'\mathbf{X}'\mathbf{X}\mathbf{u} = \text{constant}$$

defines an ellipsoid in the space of the variables $(u_1, u_2, \ldots, u_p) = \mathbf{u}'$ and $\det(\mathbf{X}'\mathbf{X}) = |\mathbf{X}'\mathbf{X}|$ is proportional to the volume contained by the ellipse. The exact area depends on the choice of the constant. This result has application in the construction of joint confidence regions for regression parameters in Section 5.4.

### Common Factors

If *every* element of a matrix has a common factor, it can be taken outside the matrix. Conversely, if a matrix is multiplied by a constant $c$, every element of the matrix is multiplied by $c$. For example,

$$\begin{bmatrix} 4 & 6 & -2 \\ 8 & 6 & 2 \end{bmatrix} = 2\begin{bmatrix} 2 & 3 & -1 \\ 4 & 3 & 1 \end{bmatrix}.$$

Note that, if a matrix is square and of size $p \times p$, and if $c$ is a common factor, then the determinant of the matrix has a factor $c^p$, not just $c$. For example,

$$\begin{vmatrix} 4 & 6 \\ 8 & 6 \end{vmatrix} = 2^2\begin{vmatrix} 2 & 3 \\ 4 & 3 \end{vmatrix} = 2^2(6 - 12) = -24.$$

Additional information on matrices is given where needed in the text. Also see Appendix 5A.