CHAPTER I

REVIEW OF ARITHMETIC

1.1. Addition and multiplication.

There are certain facts that hold for the operations of addition, subtraction, multiplication and division. These facts are stated below.

Fact 1.1.1A. The order in which numbers are added does not affect the result.

Illustration: Many two-way tables of numbers occur in statistics. For example, the following table might represent the numbers of flies that hatched on three successive dates from four different crosses.

<table>
<thead>
<tr>
<th>Crosses</th>
<th>Dates</th>
<th>Cross totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Date totals</td>
<td>11</td>
<td>23</td>
</tr>
</tbody>
</table>

We may add the numbers in each row to get four row totals, then add these to obtain the grand total. Alternately, we may add the numbers in each column to get three column totals, then add these to obtain the grand total. In the two cases, the order of addition is different but the final result is the same.

Fact 1.1.1B. The order in which numbers are multiplied does not affect the result.

Illustration: Multiplication of three numbers is not uncommon in statistics. For example, the product of two averages, each based on the same number of observations, by this common number must frequently be computed. For example, if 12 men are measured and their average weight is 160 lbs. and their average height is 70 inches, then

\[ 12 \times 160 \times 70 = 134,400 = 70 \times 160 \times 12 \]
Subtraction is the inverse of addition and Fact 1.1.1A applies also to this operation. Subtraction of a negative number is equivalent to addition of a positive one. Thus, \(7-(-3)=7+3\).

Addition and subtraction together are often called algebraic addition. Fact 1.1.1A applies to algebraic addition. In statistics, we often wish to add a series of mixed positive and negative numbers. Since Fact 1.1.1A holds, we may choose to add all the positive ones, then all the negative ones, and finally obtain the difference. This difference may be positive or negative.

Division is the inverse of multiplication. (Division by zero is not allowed.) When multiplication and division occur in the same problem, Fact 1.1.1B applies. For example,

\[
\frac{3 \times 8}{2} = 3 \times \frac{8}{2} = 3 \times 4 = 12
\]

This is division followed by multiplication. Also

\[
\frac{3 \times 8}{2} = \frac{24}{2} = 12
\]

This is multiplication followed by division. The same result is obtained.

1.2. Order of arithmetic operations.

In arithmetic, there are certain rules of procedure concerning the order in which the operations may be performed. These rules are a matter of general agreement. They permit arithmetic problems to be done in an orderly, consistent manner.

Rule 1.2.1. If both multiplication (division) and addition (subtraction) are included in the same problem, multiplication is done first.

For example,

\[
5+2 \times 3 = 5+6=11
\]
\[
2 \times 8-3 \times 3=16-9= 7
\]
\[
7+6 \div 3 = 7+2= 9
\]
\[
8 \div 2 - 1 = 4-1= 3
\]

Rule 1.2.2. Parentheses, brackets, or other symbols are used to evade Rule 1.2.1. when this is required. For example,
(7+9)\frac{1}{4}=16\frac{1}{4}=4
(5-2)\times 12=3\times 12=36
\frac{15}{3+2}=\frac{15}{5}=3 \quad \text{(The horizontal line is an "other symbol".)}
9+4+25-\frac{10\times 10}{3}=38-\frac{100}{3}

Rule 1.2.2a. Application of a distributive law may be used as an alternative to Rule 1.2.2. Multiplication is said to be distributive with respect to addition. For example,

(5-2)\times 12=5\times 12-2\times 12
=60-24=36

(7+9)\frac{1}{4}=7\frac{1}{4}+9\frac{1}{4}
=1\frac{3}{4}+2\frac{1}{4}=4

Note that the following example does not provide a distributive law.

\frac{15}{3+2} \neq \frac{15}{3} + \frac{15}{2} \quad (\neq \text{means "does not equal".})

1.3. Fractions.

A fraction is a ratio of two positive numbers called the numerator (top) and denominator (bottom), for example \(\frac{3}{8}\), \(\frac{7}{4}\). The denominator tells into how many parts the whole was "fractured"; the numerator tells or "names" how many parts are in the fraction. In our example we had three eighths and seven quarters.

Fractions and division are related, the denominator of the fraction corresponds to the divisor in division.

A decimal fraction is a fraction in which the denominator is always a power of 10. A decimal point (period) is generally used to indicate the denominator of a decimal fraction. For example \(3.3=33/10\), \(3.49=349/100\). The number of zeros in the denominator equals the number of digits after the decimal point.

Decimal fractions are very convenient and one has no trouble in deciding which of two is larger. On the other hand, a decision as to whether \(73/97\) or \(3/4\) is the larger is not reached as easily by most people.
Law 1.3.1. Multiplication (or division) of both numerator and denominator of a fraction by the same number, other than zero, does not alter the value of the fraction.

Application: When fractions are to be added or subtracted, they are first brought to a common denominator.

Fact 1.1.1B and the notion of a fraction can be used to show that the product of fractions is a new fraction with numerator equal to the product of the original numerators and denominator equal to the product of the original denominators.

Law 1.3.1 may be used to show that division by a fraction may be accomplished by inverting the fraction and multiplying.

1.4. Laws for decimal points.

Law 1.4.1. When two decimal fractions are multiplied the number of digits to the right of the decimal point in the product equals the sum of the number of decimals in the two fractions.

Law 1.4.2. When dividing a number by a decimal fraction write the two numbers down as for long division; move the decimal point of the divisor to the right of the last digit; move the decimal point of the dividend an equal number of places to the right, adding zeros if necessary; proceed as in division with the decimal place of the quotient placed directly above the newly placed decimal point in the dividend. Notice that the shifting of decimal points in Law 1.4.2 is an application of Law 1.3.1.

1.5. Integral powers.

Consider the number 5^3. The number 3 placed to the right and above the number 5 is called the power of 5 or the power to which 5 is raised. The power tells how many times the number is to appear in a product. Thus 5^3 = 5x5x5.

In science, we may deal with very large or very small numbers and it is often convenient to make use of powers of 10. For numbers greater than one, positive powers are used. For example, the population of the U.S.A. is approximately 175 million or, more conveniently, 175x10^6. For numbers less than one, negative powers are used. For example, in speaking of six parts per hundred thousand, we write 6x10^{-5}. Notice that

\[
\frac{1}{100,000} = \frac{1}{10^5} = 10^{-5} = 0.00001
\]
The power two is the most common power (other than one, which is rarely shown) used in statistical computations. A number raised to the power two is said to be squared. We also speak of the square of a number.

1.6 Fractional powers.

The most common fractional power in statistical computations is the power one-half. A number raised to the one-half power is commonly called a square root. A square root may be written as $65^{1/2}$ or, more commonly, $\sqrt{65}$. Negative fractional powers are also used. Thus $5^{-1/2}=1/5^{1/2}=1/\sqrt{5}$.

Problems

In the following problems, complete the indicated operations.

1.1 (Algebraic addition)

<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>ii</th>
<th>iii</th>
<th>iv</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>+5</td>
<td>+3</td>
<td>+4</td>
<td>+8</td>
</tr>
<tr>
<td></td>
<td>-3</td>
<td>+7</td>
<td>-7</td>
<td>+6</td>
</tr>
<tr>
<td>ii</td>
<td>+8</td>
<td>-2</td>
<td>+6</td>
<td>+3</td>
</tr>
<tr>
<td></td>
<td>-4</td>
<td>-8</td>
<td>-3</td>
<td>-7</td>
</tr>
<tr>
<td>iii</td>
<td>-9</td>
<td>-3</td>
<td>-5</td>
<td>-4</td>
</tr>
<tr>
<td>iv</td>
<td>+6</td>
<td>+1</td>
<td>+2</td>
<td>-3</td>
</tr>
</tbody>
</table>

1.2 (Multiplication)

i $8 \times 93$; ii $17 \times (-14)$; iii $67 \times 340$; iv $(-29) \times (-33)$.

1.3 (Division)

i $45 \div 5$; ii $261 \div (-3)$; iii $-328 \div (-4)$; iv $-65 \div 13$.

1.4 (Multiple operations)

i $\frac{4 \times 7}{2}$; ii $\frac{5 \times 8}{2 + 6}$; iii $\frac{6 \times 7 - 2}{8 - 4}$; iv $\frac{3 + 7 \times 2 - 5}{3 \times 2}$; v $\frac{6 \times 7 - 4 \times 4}{3 \times 2 + 4 - 2}$;

vi $\frac{3(4+2)-3}{2+3}$; vii $\frac{3(8-2)-5 \times 3}{2(7-4)}$. 
1.5 (Fractions)
   i \( \frac{3}{4} + \frac{4}{5} \); ii \( \frac{2}{3} + \frac{7}{8} - \frac{2}{5} \); iii \( \frac{7}{8} - (-\frac{2}{3}) \); iv \( \frac{6}{5} + \frac{1}{3} \left( \frac{1}{2} + \frac{1}{6} \right) \);
   v \( \frac{3}{4} \times \frac{1}{2} \); vi \( \frac{4}{5} \times (-\frac{2}{3}) \); vii \( \frac{5}{6} \div \frac{3}{8} \); viii \( -\frac{2}{5} \div \frac{7}{16} \);
   ix \( \frac{6}{11} \div (\frac{2}{3} - \frac{7}{8}) \)

1.6 (Decimal fractions)
   i \( 3.6 \times 2.2 \); ii \( 4.12 \times 0.03 \); iii \( 8.11 \times 3.003 \); iv \( 0.081 \times 0.007 \);
   v \( 4.2 \div 0.007 \); vi \( 8.12 \div 0.40 \); vii \( 360 \div 0.09 \); viii \( 0.0076 \div 0.004 \);
   ix \( \frac{4.2 \times 2.2}{0.6} \); x \( \frac{3.1 \times 0.009}{30} \); xi \( \frac{2.73 \times 0.008}{0.003} \); xii \( \frac{0.6 \times 7.5}{450} \)

1.7 (Integral powers)
   Show the result as a product of prime numbers raised to appropriate powers.
   (A prime number is divisible, with no remainder, by itself and one only.
   E.g. \( 2, 3, 5, 7, 11, 13, \ldots \)).
   i \( 5^2 \times 5^3 \); ii \( 6^3 \times 2^{-2} \); iii \( 2^4 \times 10^{-3} \); iv \( 21^2 \times 6^{-1} \); v \( 35 \times 10^2 \).

1.8 Write the following numbers conveniently as powers of 10.
   i \( 670,000 \); ii \( 157,000 \); iii \( 3,800,000 \); iv \( 0.0035 \); v \( 0.000,672 \);
   vi \( 0.0008 \).

1.9 (Powers of 10)
   i \( (25 \times 10^3)(3 \times 10^4) \); ii \( (12 \times 10^{-4})(6 \times 10^{-3}) \); iii \( (7 \times 10^3)(8 \times 10^{-5}) \);
   iv \( (6 \times 10^{-4})(8 \times 10^4) \).

1.10 Find the square root of
   i \( 0.625 \); ii \( 0.0031 \); iii \( 1.44 \); iv \( 22.5 \); v \( 0.016 \).
2.1. Introduction to algebra.

Arithmetic is concerned with operations involving numbers whereas algebra is concerned with operations involving symbols associated with elements, a term we now define.

The individuals in a class or system of undefined entities or objects, finite or infinite in number, are called elements. For example, the positive integers between 1 and 100 are a finite class of elements while all the numbers, positive and negative, are an infinite class of elements.

In algebra, we use symbols to identify elements although the elements themselves are undefined objects. However, since we will be mainly interested in statistical computations, the system of elements of concern to us is the system of numbers.

Having a set of elements is like having a deck of cards but not knowing any card games. In other words, we need a set of rules. In algebra, the set of rules consists of permissible operations. For our purposes, these are addition (+) and multiplication (x). The operations have inverses called subtraction (-) and division (÷). It can be shown that if we permit division by zero then ridiculous results, such as that one equals two, may be obtained. Hence we will not permit division by zero. We will not define our four operations but simply state that they are those to which the student is already accustomed.

With a set of elements and the permissible operations, we may consider certain statements or propositions with a view to proving whether they are true or false. Before we can illustrate this sort of game, we need to introduce some symbols and notation.

2.2. Symbols and notation.

Let X be associated with our system of elements in such a way that \( X_i \) represents the i-th element in a sub-set of elements. For example, let X be associated with the system of elements consisting of all positive numbers. The subset might consist of the birthweights, to the nearest tenth of a pound, of ten children, say 6.2, 5.9, 7.5, 6.5, 3.0, 6.6, 7.5, 3.2, 6.3, and 5.5 pounds. Here \( X_1 = 6.2, X_2 = 5.9, X_3 = 7.5, \ldots, X_{10} = 5.5 \). The three dots between \( X_3 \) and \( X_{10} \) are
notation to indicate the unwritten values.

In more general terms, a subset of elements consisting of an unstated number of elements is written as \( X_1, X_2, \ldots, X_n \).

Notation is also used to indicate what operation is to be performed. Addition is generally indicated by \( \Sigma \) (Greek capital sigma). Thus \( \Sigma \) represents "the sum of" and we have, as a kind of shorthand, definition 2.2.1.

\[
\sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n
\]

Here, \( i \) is called the index of summation and we read \( \sum_{i=1}^{n} X_i \) as "the sum of the \( X_i \)s for \( i \) equal to 1 to \( n \)." The limits of summation, 1 and \( n \), and even the index \( i \) may be omitted when all the numbers are to be added. Thus we write \( \sum_{i=1}^{n} X_i \) or simply \( \sum X_i \) or \( \sum X \).

Similarly, \( \prod \) represents "the product of" and we have definition 2.2.2.

\[
\prod_{i=1}^{n} X_i = X_1 \cdot X_2 \cdot \cdots \cdot X_n
\]

The limits of multiplication may be treated similarly to those of summation when no confusion is likely to arise.

Three other bits of shorthand that are continually used in statistics are presented as definitions 2.2.3 to 2.2.5.

Arithmetic mean: \( \bar{x} = \frac{\sum X_i}{n} \)

Deviation (from mean): \( x_i = X_i - \bar{x} \)

Sum: \( \sum X_i \)

2.2.3

2.2.4

2.2.5

2.3. Some algebraic facts, rules, and laws.

The facts, rules, and laws presented in Chapter 1 for arithmetic have their parallels in algebra. Thus, corresponding to fact 1.1.1A we have

Fact 2.3.1A. The order in which elements are added does not affect the result. For example, \( \Sigma(X_i - Y_i) = \Sigma X_i - \Sigma Y_i \).
Illustration: Paired observations are common in statistics and we are often required to sum their differences. Fact 2.3.1A states that $\sum (X_i - Y_i) = \sum X_i - \sum Y_i$ or that the sum of differences equals the difference of sums, the latter usually being the simpler form to obtain.

Corresponding to fact 1.1.1B we have

**Fact 2.3.1B.** The order in which elements are multiplied does not affect the result.

Illustration: $n \bar{x} \bar{y} = (n \bar{x}) \bar{y} = (n \bar{y}) \bar{x}$. In many situations, $n \bar{x}$ and $n \bar{y}$ are readily available in multiplied form so we may simply multiply either $n \bar{x}$ or $n \bar{y}$ by $\bar{y}$ or $\bar{x}$.

Rules 1.2.1, 1.2.2, and 1.2.2a for arithmetic concern the order of addition and multiplication and the use of parentheses and other symbols; these also hold for the elements with which we are dealing so will not be repeated here.

Illustration for rule 1.2.2a: Applying this rule to definition 2.2.3 gives $\bar{x} = (\sum X_i / n) = \sum (X_i / n)$. Clearly the former expression is the more convenient for computations; however, the latter may be algebraically useful.

**Law 2.3.1.** Since this is simply law 1.3.1 applied to elements, we illustrate it in a useful situation.

Illustration: $\bar{X} \bar{Y} = \bar{X} / \bar{Y} = (\sum X_i / n) / (\sum Y_i / n) = (\sum X_i / n) / (\sum Y_i / n)$

In addition to the above law, we present two others applicable in arithmetic but not presented in Chapter 1.

**Law 2.3.2.** If equals be added to (algebraically) equals, the results are equal. Symbolically

if $X = Y$, then $X + A = Y + A$

**Law 2.3.3.** If equals are multiplied by (or divided by) equals, the results are equal. Symbolically

if $X = Y$, then $AX = AY$
Problems

2.3.1. Write in full (\ldots may be used) to show that you understand the notation of section 2.2.

\[ \begin{align*}
\text{i) } & \sum_{i=1}^{n} (X_i - a) ; \\
\text{ii) } & \sum_{i=1}^{n} (X_i - Y_i) ; \\
\text{iii) } & \sum_{\alpha=1}^{n} \alpha X_\alpha ; \\
\text{iv) } & \sum_{i=1}^{n} n X_i ; \\
\text{v) } & \sum_{i=1}^{n} (X_i + a)^2 ; \\
\text{vi) } & \sum_{i=1}^{n} (X_i - Y_i)^2 ; \\
\text{vii) } & \sum_{\alpha=1}^{n} \alpha X_\alpha ; \\
\text{viii) } & \sum_{j=1}^{n} (-1)^j X_j ; \\
\text{ix) } & \sum_{j=1}^{2} \sum_{i=1}^{n} X_{i+j-1}^2 ; \\
\text{x) } & \sum_{i=1}^{n} (\sum_{i=1}^{3} X_i)^2 ; \\
\text{xi) } & \sum_{i=1}^{n} X_i^2 ; \\
\text{xii) } & \sum_{i=1}^{n} X_i (X_i + 2) .
\end{align*} \]

2.3.2. Let \(X_1, \ldots, X_n\) be a set of numbers. Express symbolically, the following:

i) The sum of all the numbers.

ii) The sum of the numbers with even subscripts.

iii) The sum of the numbers with odd subscripts.

iv) The product of the numbers.

v) The sum of the squares of the numbers.

vi) The sum of the products of all pairs of different numbers, sometimes called cross-products, i.e., the sum \(X_1 X_2 + X_1 X_3 + \cdots + X_1 X_n + X_2 X_3 + \cdots + X_{n-1} X_n\).

vii) The sum of the squares of the numbers plus twice the sum of cross-products.

2.4. Some elementary theorems.

Three useful theorems are presented below. These are regularly applied in later chapters. Theorem 2.4.2 is proved so that it may supply a model for the presentation of solutions to algebraic problems in general.

Theorem 2.4.1. \[ \sum_{i=1}^{n} (X_i + Y_i - Z_i) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} Z_i \]

Theorem 2.4.2. \[ \sum_{i=1}^{n} (X_i + a) = \sum_{i=1}^{n} X_i + na , \text{ where } a \text{ is a constant} \]
Theorem 2.4.3: \( \Sigma x_i = c \Sigma x_i \)

Theorem 2.4.2 is now proved.

<table>
<thead>
<tr>
<th>Step</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \Sigma(x_i+a) = \Sigma x_i + na )</td>
</tr>
<tr>
<td>2</td>
<td>Def'n 2.2.1</td>
</tr>
<tr>
<td>3</td>
<td>Fact 2.3.1A</td>
</tr>
<tr>
<td>4</td>
<td>Def'n 2.2.1</td>
</tr>
</tbody>
</table>

The procedure is seen to be as follows:

Step 1: We state the problem in this case a theorem. At this stage, we might even place a question mark at the end. We also draw a vertical line below the equality sign and proceed with one side or/and the other.

Step 2: In this case we proceed with the left side and use definition 2.2.1. This is so-stated at far right.

Step 3: Fact 2.3.1A allows us to reorder the arithmetic.

Step 4: Again we apply definition 2.2.1 and a definition of multiplication though the latter has not been stated here. Since we have now shown that the expression first given on the left equals that on the bottom left, which is also that first given on the right, we have proved the theorem and may write \( \equiv \) through our vertical line to indicate that we have completed the proof.

In some cases it may be convenient to work with both sides, reducing both to the same common value. If this approach is used, equality signs associated with the right side are placed to the right of the vertical sign. Only when the two sides have been reduced to the same expression do we write \( \equiv \) through the vertical line. For example
In this proof, we have used a fact not previously stated, namely

Fact 2.4.1. Things which equal the same thing, equal one another.

Problems

2.4.1. Prove theorems 2.4.1 and 2.4.3.

2.4.2. Prove that $\Sigma(aX_i + bY_i) = a\Sigma X_i + b\Sigma Y_i$.

2.4.3. Prove $\Sigma(X_i - \bar{x}) = 0$.

2.4.4. Prove $\left(\Sigma X_i\right)^2/n = nx^2$.

[Hint: apply law 2.3.1.]

2.4.5. Prove $\Sigma(X_i - \bar{x})^2 = \Sigma X_i^2 - (\Sigma X_i)^2/n$.

[Hint: use one of the identities $(a+b)^2 = a^2 + 2ab + b^2$.]

2.4.6. Prove $\Sigma X_i(X_i - \bar{x}) = \Sigma X_i^2 - (\Sigma X_i)^2/n$.

2.4.7. Prove $\Sigma(X_i - a)^2 = \Sigma(X_i - \bar{x})^2 + n(\bar{x} - a)^2$.

[Hint: use one of the identities given in hint for problem 2.4.4, after rewriting the expression on the left as $\Sigma(X_i - \bar{x} - a)^2$.]

2.4.8. Use problem 2.4.6 to show that $\Sigma(X_i - a)^2$ is a minimum when $a = \bar{x}$.

(It is often stated that $\Sigma(X_i - \bar{x})^2$ is a minimum sum of squares.)

2.5. Coding.

The arithmetic of statistics is often simplified by coding the original data. For example, the numbers $17.037, 17.011, 17.109, 17.310, \cdots$ are a set for which the mean is obviously going to be $17.??\bar{?}$. Hence, why
bother with the 17 in the addition and division processes of finding the mean? The numbers .037, .011, .109, .310, ••• have the same variability, by definition, as the original set. Hence they are not required in the computation of the variance and the standard deviation. As a matter of convenience, it might be still easier to manipulate the numbers 37, 11, 109, 310, •••. These matters can be handled by coding.

Let \( X_1, \ldots, X_n \) be the original data. Code these as \( Y_1, \ldots, Y_n \) by use of the instruction \( Y_i = a + bX_i \). For example 37, 11, ••• are obtained from 17.037, 17.011, ••• when \( b = 1000 \) and \( a = -17,000 \).

Coding is seen to consist of

1. Multiplying each \( X \) by the same constant, \( b \).
2. Adding a constant, \( a \), to the result of step 1.

Decoding is accomplished by using inverse operations in the reverse order. Thus, it consists of

1. Subtracting \( a \) from each \( Y \).
2. Dividing the result by \( b \).

This is obvious since \( X_i = (Y_i - a)/b \).

After coding the \( X \)'s, the computations are carried out with the \( Y \)'s and the final result is decoded to get back the original units in which the observations were made.

Problems

2.5.1. If \( Y_i = a + bX_i \), show that \( \bar{y} = a + b\bar{x} \) and \( \bar{x} = (\bar{y} - a)/b \).

2.5.2. The sample variance is defined as \( s_y^2 = (Y_i - \bar{y})^2/(n-1) \). Show that \( s_y^2 = b^2 s_x^2 \) and \( s_x^2 = s_y^2/b^2 \). Hence \( s_y = b s_x \) and \( s_x = s_y/b \).

2.5.3. Prove that for \( n = 2 \), \( s^2 = (X_1 - X_2)^2/2 \).
3.1. **Expected values.**

By definition, an expected value is a population mean. The population may involve a continuous or discrete variable and, in the case of the discrete variable, sampling may be with or without replacement. When we sample from a population with mean \( \mu \), we are sampling from a population for which the expected value of \( X \) is \( \mu \). This may be written as definition 3.1.1.

\[
E(X) = \mu
\]

**Illustration 1:** Suppose we toss a die for which each face is equally likely to appear. This is sampling with replacement. Symbolically, let \( X_1 = 1, X_2 = 2, \ldots, X_6 = 6 \). Then \( P(X=1) = 1/6, \ldots, P(X=6) = 1/6 \). (Note that subscripts on the \( X \)'s apply to values in the population rather than in the sample.) This may be written more concisely as

\[
P(X=X_i) = 1/6, \; X_i = 1, 2, \ldots, 6.
\]

We read \( P(X=X_i) \) as "the probability that the random variable \( X \) takes the value \( X_i \)."

In order to find the mean of this, or any, population, we must weight each value of \( X \) with its relative frequency or probability. In the illustration all probabilities are equal and we apply Theorem 2.4.3.

\[
E(X) = 1(1/6) + 2(1/6) + \cdots + 6(1/6)
\]

\[
= (1+2+\cdots+6)/6
\]

\[
= 3.5
\]

**Illustration 2:** Probabilities vary with the outcome when two coins are tossed. The outcomes and their probabilities are two heads with probability one-quarter, one head and one tail with probability one-half, and two tails with probability one-quarter. If we let number of heads be the variable, then \( X_1 = 2, X_2 = 1, X_3 = 0 \) and the expected value of \( X \) is

\[
E(X) = 2(1/4) + 1(1/2) + 0(1/4) = 1
\]
A general definition for an expected value in a finite population is given as definition 3.1.2.

\[ E(X) = \sum_{i} x_i p(x_i) \]  

Thus, an expected value is seen to be a weighted average of the individuals in a population. This definition holds also for continuous variables though the \( \sum \) is replaced by a different symbol.

The similarity between \( E \) and \( \sum \) is now apparent. Each letter serves as a computing instruction; each calls for a summation. Such letters, used to give mathematical instructions, are called operators.

**Problems**

Suppose two dice are tossed and the mean of the two observations is recorded. This is sampling with replacement.

3.1.1. What is the distribution of the means? (You will need all possible \( \bar{x} \)'s with their probabilities.)

3.1.2. Find \( E(\bar{x}) \) for the distribution obtained in problem 3.1.1.

3.1.3. Repeat problems 3.1.1 and 3.1.2 for sums of two observations.

3.2. The variance of a finite population.

The variable whose expectation is required is not necessarily that of the parent population. See problem 3.1.2 for example. Also, the expectation of \( x^2 \), of \( (x-\mu)^2 \), and of \( (\sum x_i)^2/n \) are commonly required.

The expected value of \( (x-\mu)^2 \) is defined as the variance and denoted by \( \sigma^2 \), definition 3.2.1, and its computation for sampling a discrete variable with replacement is given by definition 3.2.2.

\[ \sigma^2 = E[(x-E(x))^2] \]  

\[ = E(x-\mu)^2 \]  

\[ = \sum (x_i-\mu)^2 p(x_i=x_i) \]  

Illustration: The variance of the variable in the one die problem is computed as follows:
\[ \sigma^2 = E(X - \mu)^2 \]
\[ = \sum_{i} (X_i - \mu)^2 P(X = X_i) \]
\[ = (1-3.5)^2/6 + (2-3.5)^2/6 + (3-3.5)^2/6 \]
\[ + (4-3.5)^2/6 + (5-3.5)^2/6 + (6-3.5)^2/6 \]
\[ = 35/12 \]

Problems

3.2.1. For the one die problem, find \( E(X_i^2) \).

3.2.2. For the two dice problem, find \( E(X^2) \) and \( E(x_2^2) \).

3.2.3. Show that for sampling with replacement \( \sigma^2 = E(X^2) - \mu^2 \).

3.2.4. Use the result obtained in problem 3.2.3 and the computations from problem 3.2.1 to find the variance of \( X \) for the one die problem.

3.2.5. Use the result obtained in problem 3.2.3 and the computations from problems 3.2.2, 3.1.2, and 3.1.3 to find the variance of the mean and the sum in the two dice problems. (These should be \( \sigma^2/n \) and \( n\sigma^2 \) respectively.)

3.2.6. Compute the sample variances for samples of size 2. (The two dice problem.)

3.2.7. Find \( \mu_s = E(s^2) \) for problem 3.2.6.

3.3. Expected values for continuous variables.

Definitions 3.1.1 and 3.2.1 are applicable to any population. In particular, we will use them to solve many important problems for continuous variables. On the other hand, definitions 3.1.2 and 3.2.2 are intended for computations involving discrete variables. Modification of these definitions to handle continuous variables is possible if we introduce integral calculus. However, this is not necessary for our purposes and we will proceed without it, using definitions 3.1.1 and 3.2.1 and the following properties of the operator \( E \). These properties would normally be stated as propositions or theorems requiring proof. They follow.
Property 3.3.1 states that the average value of a constant is a constant. It is seen to be closely related to that part of Theorem 2.4.2 which deals with the constant.

Property 3.3.2 states that the order of multiplication and taking an expectation or averaging may be interchanged. It is seen to be closely related to Theorem 2.4.3.

Property 3.3.3 states that the order of taking expectations and addition may be interchanged, that is, the expectation of a sum equals the sum of the expectations. It is seen to be closely related to Theorem 2.4.1 since taking expectations is a kind of addition process.

Property 3.3.4 is a general property from which properties 3.3.2 and 3.3.3 could be obtained. However, it is convenient to have all three expressions for the work to follow.

These properties apply for continuous variables and for discrete variables whether sampled with or without replacement.

We now proceed to some useful statistical theorems.

Theorem: \( E(\bar{X}) = \mu \)  

Proof: \[
E(\bar{X}) = E\left( \frac{\sum X_i}{n} \right) \\
= \frac{1}{n} \sum E(X_i) \\
= \frac{1}{n} \sigma \mu 
\]

It seems useful to point out what \( E(X_i) \) means. The expectation, \( E(\bar{X}) \), is taken over the population of \( \bar{X} \)'s. Hence \( E(X_i) \) is the average of the population of \( X_i \)'s for fixed \( i \). In other words, we must have the \( X \)'s from the \( i \)-th position in all possible samples. Since there is nothing to distinguish the \( i \)-th position from any other (there would be if we ordered the sample values by
magnitude), we simply generate the parent population again. Hence $E(X_i) = E(X) = \mu$.

The distributions of $X$ and of $\bar{X}$ are called univariate distributions. For a discrete variable, the distribution can be shown as a series of lines vertical to an axis, each line being of height proportional to the probability of the corresponding $X$ or $\bar{X}$. For a continuous variable, the distribution can be shown as a curve above an axis. A sample of size one ($n=1$) from either type of distribution can be plotted as a point on a line.

The distribution of a sample, not of its mean, of two observations is a bivariate distribution, and the notion may be extended to other multivariate distributions or joint (probability) distributions. A bivariate distribution for a discrete variable can be shown as a series of lines vertical to a plane, each line being of height proportional to the probability of the pair of numbers. For a continuous variable, the distribution can be shown as a curved surface above a plane. A sample of size two ($n=2$) may be considered to be a sample of size one from a bivariate distribution and can be plotted as a point in a plane.

It is often necessary to consider joint distributions in finding expectations. We shall need property 3.3.6 to cover certain common situations. The property states that

$$E(X_i X_j) = E(X_i)E(X_j), \quad i \neq j$$

when sampling is random for a continuous variable or random and with replacement for a discrete variable. The drawing of any observation does not affect and is not affected by the other sample observations. This property is called independence. For tossing a coin or die successively, it is equivalent to saying that the coin or die has no memory.

Theorem: For any reasonable population, the variance of a mean $\bar{X}$ of $n$ observations is given by

$$\frac{\sigma^2}{\bar{X}} = \frac{\sigma^2}{n}$$

where $\sigma^2$ is the variance of the parent population.
Proof:

\[ \sigma^2 = E[\bar{x} - E(x)]^2 \]

Definition 3.2.1

\[ = E(\bar{x}^2) - \mu^2 \]

Problem 3.3.1

\[ = E\left[\frac{\sum x_i}{n}\right]^2 - \mu^2 \]

Definition 2.2.3

\[ = E\left[\frac{\sum x_i^2}{n^2} + \frac{\sum x_i x_j}{n^2}\right] - \mu^2 \]

\[ = \frac{n(\sigma^2 + \mu^2)}{n^2} + \frac{n(n-1)\mu^2}{n^2} - \mu^2 \]

Problem 3.3.2

\[ = \frac{\sigma^2}{n} \]

Property 3.3.6

Note the expansion of \((\sum x_i)^2\). By observing \((\sum x_i)^2\) as \((x_1 + \cdots + x_n)(x_1 + \cdots + x_n)\), we readily see that the possible terms are \(x_1^2, x_1 x_2, \ldots, x_1 x_n, x_2 x_1, x_2 x_2\), \ldots, \(x_n x_1, \ldots, x_n^2\). Thus we have \(n^2\) crossproducts. We may write these as \(\sum x_i x_j\) or as \(2 \sum x_i x_j\); it is also clear that in the former representation, \(i \neq j\) and \(i < j\); there are \(n(n-1)\) crossproducts.

Theorem: \(E(s^2) = \sigma^2\) 3.3.0

Proof: Consider \(\Sigma(x_i - \bar{x})^2\), the numerator of \(s^2\).

\[ E[\Sigma(x_i - \bar{x})^2] = E[\Sigma x_i^2 - \frac{(\sum x_i)^2}{n}] \]

Problem 2.4.6

\[ = \sum E(x_i^2) - \frac{1}{n} E[(\sum x_i)^2] \]

Property 3.3.4

\[ = n(\sigma^2 + \mu^2) - \frac{1}{n}[n(\sigma^2 + \mu^2)] + n(n-1)\mu^2 \]

Problem 3.3.2

\[ + n(n-1)\mu^2 \]

Property 3.3.6

\[ = (n-1)\sigma^2 \]
Hence, as a result of property 3.3.2, we have
\[
E(s^2) = \mu s^2 = \sigma^2
\]

The preceding theorem states that \( s^2 \) is an **unbiased** estimate of \( \sigma^2 \). Any statistic intended to estimate a parameter is said to be unbiased if its average value is the parameter being estimated. Thus also, \( \bar{x} \) is an unbiased estimate of \( \mu \).

**Problems**

3.3.1. Show that \( E[(X-E(X))^2] = E(X^2) - \mu^2 \).

3.3.2. Show that \( E(X^2) = \sigma^2 + \mu^2 \). Note that this and theorem 3.3.7 also prove that \( E(\bar{x}^2) = \sigma^2/n + \mu^2 \).

3.3.3. Show that \( E[(\bar{x})^2] = n\sigma^2 + n^2\mu^2 \).

3.3.4. Show that \( E((X_i-\mu)(X_j-\mu)) = 0 \), \( i \neq j \), in random sampling of a continuous variable or of a discrete variable sampled with replacement.

3.3.5. Prove theorem 3.3.7 by using problem 3.3.4. [Hint: Write \( (\bar{x}-\mu) \) as \( \Sigma(X_i-\mu)/n \).]
CHAPTER IV

SAMPLING SEVERAL POPULATIONS

4.1. Notation for two samples.

In much experimental work, two or more populations are sampled simultaneously. When the observations occur in pairs, one from each of two populations, it is possible to consider differences and reduce many problems to one-population problems. The case of paired observations is an exception.

While it is possible to use two letters, say $X$ and $Y$, to distinguish between two sampled populations, the use of an extra subscript leads more readily to a generalization to many samples. Hence, let $X_{ij}$ be the $j$-th observation on the $i$-th sample, $i=1,2$ and $j=1,\ldots,n_1$. It is usual to let $n_1+n_2=n$; however, if $n_1=n_2$, it is customary to let the common value be $n$.

In using $\Sigma$ for summation, we must now be careful to designate the appropriate index or indices of summation. When a grand total is required, a $\Sigma$ may be used for each subscript. However, since the order in which numbers are added does not effect the result, a single $\Sigma$ with multiple indices is just as common. An alternative to $\Sigma$ is to use a "dot-notation". We have

$$
\sum_{j} X_{ij} = X_{i.}, \text{ the total for the } i\text{-th sample}
$$

$$
\sum_{i} X_{ij} = X_{.j}, \text{ the sum of the two } j\text{-th observations}
$$

$$
\sum_{i,j} X_{ij} = X_{..}, \text{ the total of all observations}
$$

Notice that a dot replaces a subscript over which addition has been carried out; the information available in that subscript has gone; the individual observations have been replaced by a summary, the total.

We have no dot-notation for a sum of squares. Thus

$$
X_{i.}^2 = (\sum_{j} X_{ij})^2, \quad X_{.j}^2 = (\sum_{i} X_{ij})^2 \quad \text{and} \quad X_{..}^2 = (\sum_{i,j} X_{ij})^2.
$$

The dot notation carries over to means directly, thus

$$
\bar{x}_{i.} = \frac{X_{i.}}{n_1}, \quad \bar{x}_{.j} = \frac{X_{.j}}{2} \text{ (two sample case), and } \bar{x}_{..} = \frac{X_{..}}{n}.
$$
Problems

4.1.1. Given the sample observations 7, 3, 6, 3 and 3, 2, 7, 9, 5, what is the value of
a) \( k \) \( n_1 \) \( n_2 \)?
b) \( X_{11} \) \( X_{13} \) \( X_{25} \)?
c) \( X_{ij} \) for \( i=2, j=4 \) \( i=1, j=3 \)?

4.1.2. For the above samples,

a) Find \( \sum_{j} x_{ij} \) for \( i=1 \).
b) Find \( \sum_{i} x_{ij} \) for \( j=3, j=4 \).
c) Find \( x_{1\cdot}, x_{2\cdot}, x_{\cdot2} \).
d) Find \( x_{j\cdot} \) for \( j=3 \).
e) How many \( x_{ij} \)'s are there?
f) Find \( \bar{x}_{1\cdot}, \bar{x}_{2\cdot}, \bar{x}_{\cdot2} \).

4.1.3. By definition \( \bar{x}_{ij} = \frac{\sum x_{ij}}{n_1 + n_2} \). Show that \( \bar{x}_{..} = \frac{n_1 \bar{x}_{1\cdot} + n_2 \bar{x}_{2\cdot}}{n_1 + n_2} \). (This is called a weighted mean of means. Note that the divisor is the sum of the weights.)

4.1.4. Show that \( \sum_{i=1}^{2} n_1 (\bar{x}_{1\cdot} - \bar{x}_{..})^2 = (\bar{x}_{1\cdot} - \bar{x}_{2\cdot})^2 n_1 n_2 / (n_1 + n_2) \).

4.1.5. Show that, for \( i=1 \) and \( 2 \) only,
\[
s_p^2 = \frac{\sum x_{ij}^2 - (\sum x_{ij})^2 / n_1 - (\sum x_{2j})^2 / n_2}{(n_1 + n_2) - 2} = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 - 1 + n_2 - 1}
\]
(It is the second equality that is to be proven.)

4.1.6. Show that
\[
t^2 = \left( \frac{\bar{x}_{1\cdot} - \bar{x}_{2\cdot}}{\sqrt{s_p^2 / n_1 + s_p^2 / n_2}} \right)^2 = \frac{x_{1\cdot}^2 / n_1 + x_{2\cdot}^2 / n_2 - x_{..}^2 / (n_1 + n_2)}{s_p^2} = F
\]
(Note that \( t^2 \) and \( F \) are defined by the first and last equalities. It is the center equality that is to be proven.)
4.2. The completely random design.

When experimental material exhibits no variability which requires control, then the only additional source of variation, in observations made during the course of an experiment using this material, is that attributable to treatments. The data arise from a so-called completely random design and are recorded in a one-way classification, the classification being treatments.

For the completely random design, let $X_{ij}$ be the j-th observation in the i-th sample, $i=1,\ldots,k$, $j=1,\ldots,n_i$. The error variance, error mean square, or within-classes variance may be computed by equation 4.2.1 a generalization of the definition given for $k=2$ in Problem 4.1.5.

$$s^2 = \frac{\sum X_{ij}^2 - (\sum X_{ij})^2/n_i}{\sum n_i - k} \quad 4.2.1$$

Means computed from different numbers of observations have different variances, in particular, $\sigma^2 = \sigma^2/n_i$. In computing an estimate of $\sigma^2$ from a set of such means, the most efficient procedure is one which weights the means inversely as their variances. In particular, we define a sum of squares for use in estimating $\sigma^2$ by equation 4.2.2.

$$SS = \sum n_i (\bar{X}_i - \bar{X})^2 \quad 4.2.2$$

This is called the treatment sum of squares or among classes sum of squares. $\bar{X}$ is defined as the general mean but is also a weighted mean of the means as was seen for $k=2$ in Problem 4.1.3. The idea of the weighted mean is more appropriate in the present context.

While equation 4.2.2 may be used for computing, a more satisfactory computing formula is that given by equation 4.2.3.

$$SS = \sum \frac{X_i^2}{n_i} - \frac{\sum X_i^2}{\sum n_i} \quad 4.2.3$$
Problems

4.2.1. Show that \( \bar{x}_{ij} = \frac{\sum X_{ij}}{n_{ij}} \). In other words, show that the weighted mean of the means is also the general mean.

4.2.2. Show that the numerator of equation 4.2.1 equals \( \sum_{i,j} (\bar{x}_{ij} - \bar{x}_{..})^2 \).

4.2.3. Show that equations 4.2.2 and 4.2.3 give the same result.

4.2.4. Show that \( \sum_{i,j} (\bar{x}_{ij} - \bar{x}_{..})^2 = \sum_{i} \sum_{j} X_{ij}^2 - n \bar{x}_{..}^2 \) where \( n = \sum_{i,j} \).

4.2.5. Show that the numerator of equation 4.2.1 plus the sum of squares given by equation 4.2.2 equals the total sum of squares \( \sum_{i,j} (X_{ij} - \bar{x}_{..})^2 \). (This proves that the property of additivity of sums of squares holds for the completely random design. It obviously holds also for degrees of freedom.) In other words, you have proven that equation 4.2.4 is valid.

\[
\sum_{i,j} (X_{ij} - \bar{x}_{..})^2 = \sum_{i,j} (X_{ij} - \bar{x}_{..})^2 + \sum_{i} (\bar{x}_{i.} - \bar{x}_{..})^2 \]  

4.2.4
CHAPTER V
THE RANDOMIZED COMPLETE BLOCK DESIGN

5.1. Paired observations.
Problems dealing with paired observations may be reduced to single sample problems by considering the signed differences of the paired observations. While this is an expeditious solution, it possibly causes the experimenter to lose sight of the reason for pairing and certainly hides some of the arithmetic that would illustrate the difference between the methods of processing paired and unpaired observations. Also, this reduction of the problem has no obvious generalization to the case where the pair becomes a set of \( k > 2 \) observations.

Paired observations, like unpaired observations, may be designated by \( X_{ij} \), \( i=1,2, j=1,\ldots,n \). Now, however, the subscript \( j \) takes on additional meaning since it, like \( i \), implies a variable of classification. Consequently, \( X_{ij} \) is a total of two observations which are alike in one respect which is not possessed by \( X_{1j} \) and \( X_{1j} \), for example; this was not so for unpaired observations. The observations now constitute a two-way classification.

Problems

5.1.1. Let \( D_j = X_{1j} - X_{2j} \). Show that \( \bar{d} = \bar{x}_{1j} - \bar{x}_{2j} \).

5.1.2. Show that \( E_2 = \sum_{j=1}^{n} (X_{1j}^2 + X_{2j}^2 - 2X_{1j}X_{2j}) \). Compare this with the error sum of squares for unpaired observations, \( n \) per treatment.

5.1.3. Show that \( \sum_{j=1}^{n} x_{1j}x_{2j} = \sum_{j=1}^{n} (X_{1j}X_{2j} - (\bar{X}_{1j})(\bar{X}_{2j})/n) = \sum_{j=1}^{n} x_{1j}x_{2j} \).

5.2. The randomized complete block design.
The randomized complete block design is a generalization of the paired comparison design. Here we have \( k \geq 2 \) experimental units which may be expected to respond alike if treated alike. The experiment requires \( n \) sets of \( k \) units, there being no requirement of alikeness from set to set, only within each set.

An observation from a randomized complete block experiment is designated by \( X_{ij} \), \( i=1,\ldots,k \), \( j=1,\ldots,n \). The observations are recorded in a two-way classification, one classification for treatments and one for pairs.

For \( k > 2 \), we no longer compute \( \bar{x}_{1i} - \bar{x}_{2j} \) and compare it with a standard
deviation appropriate to a difference between paired means; instead, we compute
a variation among treatment means, multiply by n (because $\sigma^2 = e^2 / n$), and compare
the result with the error variance. In the process, we compute a so-called
analysis of variance.

The analysis of variance is, computationally, a partition of the total sum
of squares for the $kn$ observations into component sums of squares for blocks,
treatments, and error. The subsequently-found mean squares may be used to test
hypotheses concerning the reality of block and treatment effects or to estimate
components of variance attributable to block, treatment, and random effects.

Sums of squares for block and treatment effects are computed on a per-
observation basis, that is, so that all mean squares are estimates of $\sigma^2$, the
error variance, provided appropriate null hypotheses are true. The computations
for block and treatment sums of squares are given by equations 5.2.1 and 5.2.2.

\[
\text{Block SS} = \frac{\sum X^2_{ij} - X^2_{..}}{k} - \frac{X^2_{..}}{kn} \tag{5.2.1}
\]

\[
\text{Treatment SS} = \frac{\sum X^2_{ij} - X^2_{..}}{n} - \frac{X^2_{..}}{kn} \tag{5.2.2}
\]

It is customary to obtain the error sum of squares by subtracting the block and
treatment sums of squares from the total sum of squares. This relies upon the
property of additivity of sums of squares first mentioned in equation 4.2.4. For
the randomized complete block design, we have equation 5.2.3.

\[
\sum_{i,j} (X_{ij} - \bar{X}_{..})^2 = k \sum_j (\bar{X}_{ij} - \bar{X}_{..})^2 + n \sum_i (\bar{X}_{i.} - \bar{X}_{..})^2 \tag{5.2.3}
\]

\[+ \sum_{i,j} (X_{ij} - \bar{X}_{i.} - \bar{X}_{j.} + \bar{X}_{..})^2 \]

The last summation might serve as a definition of the error sum of squares, but
is not ordinarily used as a computing formula.
Problems

5.2.1. Show that:
\[ \frac{\sum_{j} x_{ij}^2}{k} - \frac{x^2}{kn} = k \sum_{j} (\bar{x}_j - \bar{x})^2. \]
See equation 5.2.1 and the first term on the right of equation 5.2.3.

5.2.2. Show that:
\[ \frac{\sum_{i} x_{i.}^2}{n} - \frac{x^2}{kn} = n \sum_{i} (\bar{x}_i - \bar{x})^2. \]
See equation 5.2.2 and the second term on the right of equation 5.2.3.

5.2.3. For \( k = 2 \), it should be possible to show that the \( t \) test based on differences leads to the same conclusion as the \( F \) test in the analysis of variance. To this end, show that:
\[
\sum_{j} (x_{1j} - x_{2j}) - (\bar{x}_1 - \bar{x}_2)^2
= 2 \left[ \sum_{i,j} (x_{ij} - \bar{x})^2 - 2 \sum_{j} (\bar{x}_j - \bar{x})^2 - n \sum_{i} (\bar{x}_i - \bar{x})^2 \right].
\]
CHAPTER VI
MODELS AND EXPECTATIONS OF MEAN SQUARES


A model, in statistics, is a description of an observation in terms of its make-up; it consists of an algebraic description together with the assumptions concerning the components in the algebraic description. Consideration of the model leads to appropriate computing, testing, and inferential procedures.

Our concern is with the linear additive model, which describes each observation as the sum of a mean and a random component. The mean, in turn, may consist of the sum of components associated with sources of variation.

Models are further described according to the assumptions made about the components. Thus, among possible models, we have

Model I, Fixed Effects: which calls for the components in any mean to be fixed parameters,

Model II, Random Effects: which calls for the components in any mean to be obtained by random sampling from populations with zero means and unknown variances; here, the variances are the parameters, and the

Mixed Model: which calls for at least one set of effects to be in each class.

In a well-planned and conducted experiment, procedures for estimating the values of the parameters are relatively simple. The computations for the analysis of variance, usually the first step in the estimation procedure, are also simple and possess the property of additivity of sums of squares, mentioned in Chapters IV and V. This property allows us to compute error sums of squares by subtraction rather than directly, the latter procedure generally being lengthy and inconvenient.

6.2. The completely random design.

For the completely random design, equation 6.2.1 describes the make-up of an observation.

\[ X_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i=1, \cdots, k; \quad j=1, \cdots, n_i \]

\[ = \mu + \tau_i + \epsilon_{ij} \]
Notice that the mean requires only one subscript since this is a one-way classification.

It is fairly obvious that equations 6.2.2 supply estimates, distinguished from the parameters by the use of \( \hat{\cdot} \) over the parameter symbols, of the components of the mean.

\[
\hat{\mu} = \bar{x}_{..} \\
\hat{\tau}_i = \bar{x}_{i..} - \bar{x}_{..}
\]

6.2.2

In fact, these estimates are such that equation 6.2.3 holds.

\[
\sum_{i,j} \hat{\varepsilon}_{ij}^2 = \sum_{i,j} [\bar{x}_{ij} - \bar{x}_{..} - (\bar{x}_{i..} - \bar{x}_{..})]^2 = \text{minimum}\]

6.2.3

In other words, the residual or error sum of squares is a minimum; there is no other way to estimate \( \mu \) and the \( \tau \)'s so that \( \sum_{i,j} \hat{\varepsilon}_{ij}^2 \) will be smaller. The estimates are termed minimum sum of squares estimates and their deviation is discussed in Chapter .

Using equations 6.2.1 and 6.2.2, we are led to write equation 6.2.4.

\[
X_{ij} = \bar{x}_{..} + (\bar{x}_{i..} - \bar{x}_{..}) + (\bar{x}_{ij} - \bar{x}_{i..})
\]

or

\[
X_{ij} - \bar{x}_{..} = (\bar{x}_{i..} - \bar{x}_{..}) + (X_{ij} - \bar{x}_{i..})
\]

6.2.4

Consequently, we have equation 6.2.5, from which the property of additivity of sums of squares may be shown. See equation 4.2.4.

\[
\sum_{i,j} (X_{ij} - \bar{x}_{..})^2 = \sum_{i,j} (\bar{x}_{i..} - \bar{x}_{..})^2 + 2 \sum_{i,j} (\bar{x}_{i..} - \bar{x}_{..})(X_{ij} - \bar{x}_{i..}) + \sum_{i,j} (X_{ij} - \bar{x}_{i..})^2
\]

6.2.5

Problems

6.2.1. Show that \( \sum_{i,j} (\bar{x}_{i..} - \bar{x}_{..})(X_{ij} - \bar{x}_{i..}) = 0 \)

6.2.2. Show that \( \sum_{i,j} (\bar{x}_{i..} - \bar{x}_{..})^2 = \sum_{i} n_i (\bar{x}_{i..} - \bar{x}_{..})^2 = \sum_{i} \frac{x_{i..}^2}{n_i} - \frac{\bar{x}_{..}^2}{n} \) where \( n = \sum_{i} n_i \).
6.3. Expectations of mean squares in the C.R.D.

When it is desired to draw conclusions from a completed analysis of variance, it is necessary to describe fully the model. We have the following assumptions.

Model I:

1. Treatment effects are fixed, i.e., \( E(\tau_i) = \tau_i \). It then follows that \( E(\tau^2_i) = \tau^2_i \). We also set \( \Sigma \tau_i = 0 \) for equal sized samples or \( \Sigma \tau^2_i = 0 \) for unequal sized samples.

2. The random components have zero mean and common variance, i.e., \( E(\varepsilon_{ij}) = 0 \) and \( E(\varepsilon^2_{ij}) = \sigma^2 \).

Model II:

1. Treatment effects are random independent variates with zero mean and common but unknown variance, i.e., \( E(\tau_i) = 0 \) and \( E(\tau^2_i) = \sigma^2 \).

   The first assumption, \( E(\tau_i) = 0 \), corresponds to the restriction, \( \Sigma \tau_i = 0 \), for Model I.

2. The random components are as in Model I.

It now becomes informative to consider the average value in repeated experimentation, i.e., the expectation, of the various mean squares. (For the completely random design, we shall consider only the case of equal numbers of observations in each sample.) We proceed by finding expectations of the terms that appear in the computing formulas.

First consider the correction term, \( (X_{..})^2/nk \).

Model I: \[ X_{..} = nk\mu + \varepsilon \] since \( \Sigma \tau_i = 0 \)

\[
E\left(\frac{X^2_{..}}{nk}\right) = \frac{1}{nk} E(nk\mu + \varepsilon)^2
\]

\[= nk\mu^2 + \sigma^2 \]

See problem 3.3.3

Model II: \[ X_{..} = nk\mu + n\Sigma \tau_i + \varepsilon \]
\[ E\left( \frac{\Sigma X^2_{i*}}{n} \right) = \frac{1}{n} \Sigma E\left( \eta\mu + n\tau_i + \epsilon_i \right)^2 \]

\[ = nk\mu^2 + \frac{n}{k} E\left( \Sigma \tau_i \right)^2 + \sigma^2 \]

\[ = nk\mu^2 + n\sigma^2 + \sigma^2 \]

Now consider the treatment sum of squares, beginning with \( E(\sum X^2_{i*}/n) \).

**Model I:** \( X_{i*} = \eta\mu + n\tau_i + \epsilon_i \).

\[ E\left( \frac{\Sigma X^2_{i*}}{n} \right) = \frac{1}{n} \Sigma E\left( \eta\mu + n\tau_i + \epsilon_i \right)^2 \]

\[ = \frac{1}{n} \Sigma (n^2\mu^2 + n^2\sigma^2 + 2n^2\mu\tau_i) \]

\[ = nk\mu^2 + n\Sigma \tau_i^2 + n\sigma^2 \]

Hence,

\[ E\left( \Sigma X^2_{i*} - \frac{X_{i*}^2}{nk} \right) = n\Sigma \tau_i^2 + (k-1)\sigma^2 \]

and

\[ E(\text{Trt MS}) = \sigma^2 + \frac{n\Sigma \tau_i^2}{k-1} \]

**Model II:** \( X_{i*} = \eta\mu + n\tau_i + \epsilon_i \).

\[ E\left( \frac{\Sigma X^2_{i*}}{n} \right) = \frac{1}{n} \Sigma E\left( \eta\mu + n\tau_i + \epsilon_i \right)^2 \]

\[ = \frac{1}{n} \Sigma (n^2\mu^2 + n^2\sigma^2 + n\sigma^2) \]

\[ = nk\mu^2 + n\sigma^2 + nk\sigma^2 \]

\[ E\left( \Sigma X^2_{i*} - \frac{X_{i*}^2}{nk} \right) = n(k-1)\sigma^2 + (k-1)\sigma^2 \]

\[ E(\text{Trt MS}) = \sigma^2 + n\sigma^2 \]
Finally, consider the error sum of squares, beginning with $E(\Sigma x_{ij}^2)$.

**Model I:**  
\[ x_{ij} = \mu + \tau_i + \epsilon_{ij} \]

\[ E(\Sigma x_{ij}^2) = E(\mu^2 + \tau_i^2 + \epsilon_{ij}^2) \]

\[ = nk\mu^2 + n\tau_i^2 + nk\sigma^2 \]

Hence,

\[ E\left\{ \Sigma x_{ij}^2 - \frac{\Sigma x_{ij}^2}{n} \right\} = k(n-1)\sigma^2 \]

and

\[ E(\text{Error MS}) = \sigma^2 \]

**Model II:**  
\[ x_{ij} = \mu + \tau_i + \epsilon_{ij} \]

\[ E(\Sigma x_{ij}^2) = E(\mu^2 + \sigma^2_{\tau_i} + \sigma^2) \]

\[ = nk\mu^2 + nk\sigma^2_{\tau_i} + nk\sigma^2 \]

Hence,

\[ E(\Sigma x_{ij}^2 - \frac{\Sigma x_{ij}^2}{n}) = k(n-1)\sigma^2 \]

and

\[ E(\text{Error MS}) = \sigma^2 \]

**Problems**

6.3.1. Why is there only one cross-product term in equation 6.3.1? Why is there none in equation 6.3.2?

6.3.2. Find expectations for error and treatment mean squares in the analysis of a completely random design with unequal sample sizes.

6.3.3. In equation 6.3.2, we have written $E(\tau_i^2) = \sigma^2_{\tau_i}$ although $i$ is considered to be fixed since the summation is carried out at the next step. Discuss this.
6.4. The randomized complete block design.

For the randomized complete block design, as an rxt array of numbers, equation 6.4.1 describes an observation.

\[ X_{ij} = \mu_{ij} + \varepsilon_{ij} \quad i=1,\ldots,r, \quad j=1,\ldots,t \]

\[ = \mu + \tau_i + \varepsilon_{ij} \quad 6.4.1 \]

The first expression indicates, by the use of the double subscript, that there is one mean, \( \mu_{ij} \), per cell. In the second expression, this mean is written to show the treatment effect. However, there is a further source of variation, that of blocks, not present in the completely random design so that it is necessary to retain use of the subscript \( i \). Finally, we write equation 6.4.1, showing all sources of variation.

It is fairly obvious that equations 6.4.2 provide estimates of components of the cell means.

\[ \hat{\mu} = \bar{X} \quad 6.4.2 \]

\[ \hat{\rho}_i = \bar{X}_i - \bar{X} \]

\[ \hat{\tau}_j = \bar{X}_j - \bar{X} \]

These are minimum sum of squares estimates so that equation 6.4.3 holds.

\[ \sum_{ij}^2 = \sum [(X_{ij} - \bar{X}) + (\bar{X}_i - \bar{X}) + (\bar{X}_j - \bar{X})]^2 \quad 6.4.3 \]

Equation 6.4.4 may be developed as was equation 6.2.4.

\[ X_{ij} - \bar{X} = (\bar{X}_i - \bar{X}) + (\bar{X}_j - \bar{X}) + (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}) \quad 6.4.4 \]

Consequently, we have equation 6.4.5.

\[ \sum (X_{ij} - \bar{X})^2 = \sum (X_{ij} - \bar{X}_i)^2 + \sum (X_{ij} - \bar{X}_j)^2 + \sum (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X})^2 \quad 6.4.5 \]
This equation demonstrates the additivity of sums of squares and, with equation 6.4.4, shows the nature of the error term in the analysis of variance.

Problems

6.4.1. Consider each of the three cross-product terms one would expect to appear in equation 6.4.5 and show why it equals zero.

6.4.2. Show that \( \sum_{i,j} (\bar{x}_{ij} - \bar{x})^2 = \sum_j (\bar{x}_{.j} - \bar{x})^2 = N \sum_j \bar{x}_{.j}^2 / r - \bar{x}^2 / n \).

6.5. Expectations of mean squares in the R.C.B. design.

Equation 6.4.5 is an algebraic identity and requires no assumptions. Interpretation of the results, usually presented in an analysis of variance table, requires valid assumptions about the components of equation 6.4.1. Thus we have, among possible models, the following.

Model I, Fixed Effects:

1. Replicate and treatment effects are fixed, that is, \( E(p_i) = r_i \) and \( E(\tau_j) = \tau_j \). Consequently \( E(p_i^2) = r_i^2 \) and \( E(\tau_j^2) = \tau_j^2 \).

2. The \( \epsilon_{ij} \) are random, independent variables with \( E(\epsilon_{ij}) = 0 \) and \( E(\epsilon_{ij}^2) = \sigma^2 \).

Model II, Random Effects:

1. Replicate and treatment effects are random with \( E(p_i) = 0 = E(\tau_j) \) and \( E(p_i^2) = \sigma_p^2 \) and \( E(\tau_j^2) = \sigma_\tau^2 \).

2. The \( \epsilon_{ij} \)'s are as in Model I.

The Mixed Model:

1. Replicate effects are random with \( E(p_i) = 0 \) and \( E(p_i^2) = \sigma_p^2 \).

2. Treatment effects are fixed with \( E(\tau_j) = \tau_j \).

3. The \( \epsilon_{ij} \)'s are as in Models I and II.

Once again we consider the average values of the mean squares in the analysis of variance. First we consider the correction term.
Model I:  \( X_{..} = rtu + \varepsilon_{..} \) since \( \Sigma_i = 0 = \Sigma_j \)

\[
E(\frac{X_{..}^2}{rt}) = \frac{1}{rt} E((rtu + \varepsilon_{..})^2) \\
= rtu^2 + \sigma^2
\]

Model II:  \( X_{..} = rtu + t\Sigma_i + r\Sigma_j + \varepsilon_{..} \)

\[
E(\frac{X_{..}^2}{rt}) = \frac{1}{rt} E((rtu + t\Sigma_i + r\Sigma_j + \varepsilon_{..})^2) \\
= \frac{1}{rt} (rtu^2 + rt^2\rho^2 + r^2\sigma^2 + rt\sigma^2 + r^2\sigma^2 + \sigma^2) \\
= rtu^2 + t\sigma^2 + r\sigma^2 + \sigma^2
\]

Mixed Model:  \( X_{..} = rtu + t\Sigma_i + \varepsilon_{..} \) since \( \Sigma_i = 0 \)

\[
E(\frac{X_{..}^2}{rt}) = \frac{1}{rt} E((rtu + t\Sigma_i + \varepsilon_{..})^2) \\
= \frac{1}{rt} (rtu^2 + rt^2\rho^2 + r^2\sigma^2 + t\sigma^2 + \sigma^2) \\
= rtu^2 + t\rho^2 + \sigma^2
\]

Now consider the replicate (or treatment) sum of squares.

Model I:  \( X_{1..} = tu + \rho_i + \varepsilon_{1..} \)

\[
E(\frac{\Sigma_{1..}^2}{t}) = rtu^2 + t\rho_i^2 + \sigma^2
\]

Model II:  \( X_{1..} = tu + \rho_i + \Sigma_j + \varepsilon_{1..} \)

\[
E(\frac{\Sigma_{1..}^2}{t}) = rtu^2 + rt\rho_i^2 + r^2\sigma^2 + \sigma^2
\]
Finally, consider the error sum of squares.

\[ X_{ij} = \mu + \rho_i + \tau_j + \epsilon_{ij} \]

**Model I:**

\[ E(\Sigma X^2_{ij}) = \tau \mu^2 + t \Sigma \rho_i^2 + t \Sigma \tau_j^2 + r \sigma^2 \]

**Model II:**

\[ E(\Sigma X^2_{ij}) = \tau \mu^2 + r \sigma^2 \rho + t \sigma^2 + r \sigma^2 \]

**Mixed Model:**

\[ E(\Sigma X^2_{ij}) = \tau \mu^2 + r \sigma^2 \rho + t \sigma^2 \tau_j + r \sigma^2 \]

**Problems**

6.5.1. From the text material, prepare a table showing the expected values of the mean squares in the analysis of variance of a R.C.B experiment, according to model.

6.5.2. For Model II, show that \( E(\Sigma \rho_i^2) = r \sigma^2 \rho \).

6.5.3. For model II, show that \( E(\Sigma \rho_i^2) = r \sigma^2 \rho \).

6.5.4. For Model II, show that \( E(\Sigma \rho_i \tau_j) = 0 \).
7.1. Regression.

Suppose we have a population of pairs, viz. \((X,Y)\)'s. The regression of \(Y\) on \(X\) is defined as those points in the \(X,Y\)-plane which are expected values, that is population means, of \(Y\)'s, a population being determined by choosing an \(X\). Clearly, a regression will ordinarily be a line. We shall be concerned, at this point, with straight lines or linear regression. When the regression of \(Y\) on \(X\) is a linear function of \(X\), we have equation 7.1.1.

\[
\mu_{Y|X} = \alpha + \beta X
\]

Equation 7.1.1 states that the mean of a population of \(Y\)'s having a particular \(X\) value is obtained by substituting the particular \(X\) in the formula \(\alpha + \beta X\), where \(\alpha\) and \(\beta\) are non-observable parameters and \(X\) is an observable parameter.

For bivariate normal distributions, regression is linear; in many practical situations, linear regression is adequate because it is a reasonable approximation to the true regression for the portion of the curve under consideration.

From the definition of regression, it is apparent that \(X\) must be measured without error. If \(X\) were measured with error, then the \(Y\) obtained would not come from the stated population of \(Y\)'s because this population is determined by \(X\).

In sampling, \(Y\) must be random whereas \(X\) may be chosen at the convenience of the investigator.

The sample regression is written as equation 7.1.2 or 7.1.3, the latter form anticipating a part of the computational procedure.

\[
Y = a + bX
\]

\[
= \bar{y} + b(x - \bar{x})
\]

Unfortunately, the notation of these equations completely obscures the fact that their intent is to estimate a population mean. This is sometimes remedied, to some degree, by replacing \(Y\) by \(\hat{Y}\).

The estimate of \(\beta\), written \(\hat{\beta}\) or \(b\) or \(b_{yx}\), is given by equation 7.1.4.

\[
b = \frac{\Sigma xy}{\Sigma x^2}
\]
Then $\hat{a} = \bar{y} - b\bar{x}$ from equation 7.1.3.

In estimating a population mean, it is customary to choose that value for which the sum of the squares of the deviations is minimized; see problem 2.4.8. This least squares estimation procedure was also applied to means in experimental designs; see equations 6.2.3 and 6.4.3. Again, it is applied to regression and equation 7.1.5 holds.

$$\sum (Y_i - \hat{Y}_i)^2 = \text{min}$$  \hspace{1cm} 7.1.5

In addition, equation 7.1.6, similar to that of problem 2.4.3 and to others that will be given in Chapter for experimental designs, holds.

$$\sum (Y_i - \hat{Y}_i) = 0$$  \hspace{1cm} 7.1.6

The quantity $\sum (Y-\hat{Y})^2/(n-2)$ is the residual or unexplained variance, is used as error variance and denoted by $s^2_{Y\cdot X}$. It can be shown that equation 7.1.7 is valid.

$$\Sigma y^2 = b^2 \Sigma x^2 + \Sigma (Y-\hat{Y})^2$$  \hspace{1cm} 7.1.7

From this equation, we obtain equation 7.1.8 which is a computing procedure for $\Sigma (Y-\hat{Y})^2$.

$$\Sigma (Y-\hat{Y})^2 = \Sigma y^2 - b^2 \Sigma x^2$$  \hspace{1cm} 7.1.8

The latter expression on the right of this equation is the sum of squares of $Y$ attributable to the regression of $Y$ on $X$, that is, attributable to variation in $X$.

The regression coefficient $b$ may be tested for significance using Student's $t$ as test criterion, as in equation 7.1.9.

$$t = \frac{b - \beta}{s_{Y\cdot X}/\sqrt{\Sigma x^2}}$$  \hspace{1cm} 7.1.9
Problems

7.1.1. Show that Exy=ΣXY-(ΣX)(ΣY)/n.

7.1.2. Show that Exy=ExY=Σxy.

7.1.3. Show that equation 7.1.6 is valid.

7.1.4. Show that equation 7.1.8 is valid.

(Hint: Make use of equation 7.1.3 with Y written as \( \hat{Y} \).)

7.1.5. Show that \( b^2Σx^2=bΣxy=(Σxy)^2/Σx^2 \).

7.1.6. Show that \( t^2=\frac{\text{Reduction in } Σy^2/\text{Error MS}}{\text{when the hypothesis to be tested is that } β=0} \). This quantity is distributed as F(1,n-2).

7.2. Correlation.

When the (X,Y) pairs are random, it is appropriate to compute the sample coefficient of linear correlation, more generally called the correlation coefficient. This is defined by equation 7.2.1.

\[
r = \frac{Σxy}{\sqrt{Σx^2 Σy^2}}
\]

7.2.1

The population counterpart is denoted by \( ρ \).

Problems

7.2.1. Show that \( r=\text{cov } xy/(s_x s_y) \) where \( \text{cov } xy=Σxy/(n-1) \).

7.2.2. Show that 100\( r^2 \) equals the percent reduction in \( Σy^2 \) attributable to X when regression is being considered.

7.2.3. Show that \( r^2Σy^2=(Σxy)^2/Σx^2 \).

7.2.4. Show that \( r^2=b_{yx}b_{xy} \).
7.3. Expected value of $a$ and of $b$.

The algebraic expression describing an observation in a regression problem is given by equation 7.3.1.

\[ Y_i = \alpha + \beta X_i + \epsilon_i \]  

Least squares estimates of $\alpha$ and $\beta$ have been given as

\[ \hat{\alpha} = \bar{y} - b \bar{x} \quad \text{and} \quad \hat{\beta} = \frac{\sum xy}{\sum x^2} \]

Since $a$ requires $b$ in its computation, we first find $E(b)$ assuming that the $X$'s are constants.

**Theorem:** $E(b) = \beta$  

**Proof:**

\[
E(b) = E\left\{ \frac{\sum xy}{\sum x^2} \right\} \\
= \frac{1}{\sum x^2} \sum x_i E(Y_i - \bar{y}) \\
= \frac{1}{\sum x^2} \left\{ \sum x_i E(Y_i) - \sum x_i E(\bar{y}) \right\} \\
= \frac{1}{\sum x_i^2} \sum x_i (\alpha + \beta x_i) \quad \text{Eq. 7.1.1 and problem 2.4.3} \\
= \frac{1}{\sum x_i^2} \beta \sum x^2 = \beta \quad \text{Problems 2.4.3 and 5.1.3}
\]

It has now been shown that $b$ is an unbiased estimator of $\beta$.

**Problem**

7.3.1. Show that $E(a) = \alpha$.

7.4. The variance of $a$ and of $b$.

The variance of $b$ was implied in equation 7.1.9. We now obtain it directly.

**Theorem:**

\[ \sigma_b^2 = \frac{\sigma^2}{\sum x^2} \]  

\[ 7.4.1 \]
Proof: \[ \sigma_a^2 = \mathbb{E} \left\{ (a - \mathbb{E}(a))^2 \right\} \]

Eq. 3.2.1 applies

Problem 7.4.1

We proceed to find \( \mathbb{E}(b^2) \)

\[ \mathbb{E}(b^2) = \mathbb{E} \left( \frac{\sum XY}{\Sigma x^2} \right)^2 \]

Definition

\[ = \frac{1}{(\Sigma x^2)^2} \mathbb{E}(\sum XY)^2 \]

Eq. 3.3.2 and Problem 7.1.2

\[ = \frac{1}{(\Sigma x^2)^2} \left\{ \sum x_i x_j \mathbb{E}(Y_i Y_j) \right\} \]

Note after Theorem 3.3.7; Eq. 3.3.4

\[ = \frac{1}{(\Sigma x^2)^2} \left\{ \sum x_i x_j \mathbb{E}(Y_i Y_j) \right\} \]

Problem 7.4.2; Eq. 3.3.6

\[ = \frac{\sum x_i Y_i X_i}{\Sigma x^2} + \frac{\left( \sum x_i Y_i X_i \right)^2}{(\Sigma x^2)^2} \]

Problem 7.4.3

\[ = \frac{\sum x_i Y_i X_i}{\Sigma x^2} + \frac{\left( \sum x_i Y_i X_i \right)^2}{(\Sigma x^2)^2} \]

Problem 2.4.3

Hence we have proved equation 7.4.1.

We proceed to find the variance of \( a \).

Theorem: \[ \sigma_a^2 = \sigma_{Y \cdot X}^2 \frac{\Sigma x^2}{n \Sigma x^2} \] 7.4.2

Proof: \[ \sigma_a^2 = \mathbb{E} \left\{ (a - \mathbb{E}(a))^2 \right\} \]

\[ = \mathbb{E}(\tau^2) - \sigma^2 \]
We now find \( E(a^2) \).

\[
E(a^2) = E(\overline{y} - b\overline{x})^2 \\
= E(\overline{y}^2) - 2\overline{x}E(\overline{y}) + \overline{x}^2E(b^2)
\]

Consider these terms separately. First,

\[
E(\overline{y}^2) = \frac{\sum_{Y•X} a^2_{Y•X}}{n} + \mu_{Y•X}^2
\]

Second

\[
E(by) = E(\frac{\sum_{Y•X} y}{\sum_{X} x}) = \frac{1}{\sum_{X} x} \sum_{X} E(Y_i \overline{y})
\]

Now, for fixed \( i \),

\[
E(Y_i \overline{y}) = E(Y_i \frac{\sum_{Y•X} y}{n})
\]

\[
= \frac{1}{n} \left\{ E(Y_i^2) + E(Y_i) \sum_{j \neq i} E(Y_j) \right\}
\]

\[
= \frac{1}{n} \left\{ \sum_{Y•X} a^2_{Y•X} + \mu_{Y•X}^2 \sum_{j \neq i} \mu_{Y•X} \right\}
\]

\[
= \frac{1}{n} \left\{ \sum_{Y•X} a^2_{Y•X} + \mu_{Y•X} \sum_{j \neq i} \mu_{Y•X} \right\}
\]

\[
= \frac{1}{n} \left\{ \sum_{Y•X} a^2_{Y•X} + \mu_{Y•X} \sum_{j \neq i} \mu_{Y•X} \right\}
\]

Hence,

\[
E(by) = \frac{1}{\sum_{X} x} \sum_{X} \left\{ \frac{\sum_{Y•X} a^2_{Y•X}}{n} + \mu_{Y•X} \mu_{Y•X} \right\}
\]

\[
= \frac{1}{\sum_{X} x} \left\{ \mu_{Y•X} \sum_{X} \mu_{Y•X} \right\}
\]

\[
= \frac{1}{\sum_{X} x} \mu_{Y•X} \sum_{X} x
\]

\[
= \beta \mu_{Y•X}
\]
E(b^2) has already been found so that finally

\[ E(a^2) - E^2(a) = \frac{\sigma^2_{YX}}{n} + \mu^2_{YX} - 2x_0\beta_1Y_1 + \bar{X}^2 \left( \frac{\sigma^2_{YX}}{\Sigma x^2} + x^2\beta^2 - (\alpha + \beta x_0)^2 \right) \]

\[ = \frac{\sigma^2_{YX}}{n} + \mu^2_{YX} - 2\bar{Y}_X\mu_{YX} + \bar{X}^2 \left( \frac{\sigma^2_{YX}}{\Sigma x^2} + \beta^2 \bar{x}^2 - \mu^2_{YX} + 2\mu_{YX}x - \beta^2 x^2 \right) \]

\[ = \sigma^2_{YX} \left( \frac{1}{n} + \frac{\bar{x}^2}{\Sigma x^2} \right) \]

\[ = \sigma^2_{YX} \frac{\bar{X}^2}{n\Sigma x^2} \]

\[ \text{Problem 7.4.6} \]

**Problems**

7.4.1. Prove that \( \sigma_b^2 = E(b^2) - \beta^2 \).

7.4.2. Prove that \( E(Y_1^2) = \sigma_{YX}^2 + E^2(Y_1) \) given that \( \sigma_{YX}^2 \) is independent of \( X \).

7.4.3. Prove that \( \Sigma x_1 \mu_{YX} = \Sigma x_1 \{ \mu_{YX} + \beta(X_1 - \bar{x}) \} = \beta \Sigma x_1^2 \).

7.4.4. The variance of \( b \) can be found more simply from the fact that \( b \) is a linear function of the \( Y \)'s and equation 7.4.3 holds for linear functions

\[ V(\Sigma a_i Y_i) = \Sigma a_i^2 \sigma_i^2 + \Sigma a_i a_j \sigma_{ij} \]

where \( V \) stands for variance and \( \sigma_{ij} = E(Y_i - \mu_i)(Y_j - \mu_j) \). Of course, when \( Y_i \) and \( Y_j \) are independent, as in random sampling, \( \sigma_{ij} = 0 \). Prove equation 7.4.3 and apply it to finding \( \sigma_b^2 \).

7.4.5. Prove that \( \Sigma x_1 \mu_{YX} = n \mu_{YX} \).

7.4.6. Prove that \( \left\{ \frac{1}{n} + \frac{\bar{x}^2}{\Sigma x^2} \right\} = \frac{\bar{X}^2}{n\Sigma x^2} \).

7.5. The covariance of \( a \) and \( b \).

By definition, the covariance of \( a \) and \( b \) is given by equation 7.5.1.

\[ \sigma_{ab} = \text{cov}(ab) = E(ab) - E(a)E(b) \]

\[ \text{Theorem:} \quad \sigma_{ab} = \frac{-x_0 \sigma_{YX}^2}{\Sigma x^2} \]

7.5.2
Proof: \[ E(ab) = E \left\{ (\bar{y} - b\bar{x})b \right\} \]

\[ = E \left\{ b\bar{y} - b^2\bar{x} \right\} \]

\[ = \beta \mu_{Y|X} - \bar{x} \left\{ \frac{\sigma_{Y|X}^2}{\Sigma x^2} + \beta^2 \right\} \]

See Theorems 7.4.2 and 7.4.1

\[ = (\mu_{Y|X} - \beta \bar{x})\beta - \bar{x} \frac{\sigma_{Y|X}^2}{\Sigma x^2} \]

\[ = \alpha \beta - \bar{x} \frac{\sigma_{Y|X}^2}{\Sigma x^2} \]

Equation 7.5.2 follows.
CHAPTER VIII
THE CHI-SQUARE TEST CRITERION

3.1. Chi-Square.

Chi-square is defined as the sum of squares of normally and independently distributed quantities with zero means and unit variances; thus we have equation 8.1.1.

\[ \chi^2 = \sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma_i^2}, \quad n \text{ df} \quad 8.1.1 \]

In random sampling from a single normal distribution, we have an interesting special case given by equation 8.1.2.

\[ \chi^2 = \frac{\sum(X_i - \bar{x})^2}{\sigma^2} \]
\[ = \frac{(n-1)s^2}{\sigma^2}, \quad n-1 \text{ df} \quad 8.1.2 \]

This special case is of interest in that it provides a means for computing a confidence interval for \( \sigma^2 \). We begin with the probability statement (for \( 1-\alpha=0.95 \)) given by equation 8.1.3.

\[ P(\chi^2_{0.975} < \chi^2 < \chi^2_{0.025}) = 0.95 \quad 8.1.3 \]

Next, using equation 8.1.2, we arrive at equation 8.1.4.

\[ P\left(\frac{(n-1)s^2}{\chi^2_{0.025}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{0.975}}\right) = 0.95 \quad 8.1.4 \]

Equation 8.1.4 provides a simple method for computing equal-tailed confidence intervals. However, such intervals are not the shortest attainable.

Problems

8.1.1. Derive equation 8.1.4 from equations 8.1.2 and 8.1.3.
8.2. The chi-square test of goodness-of-fit.

The chi-square test of goodness-of-fit is defined by equation 8.2.1.

\[ \chi^2 = \sum \frac{(O-E)^2}{E} \]  

8.2.1

where \( O \) = observed and \( E \) = expected or theoretical. The appropriate degrees of freedom depend upon the particular test.

Equation 8.2.1 may be shown to be equivalent to equation 8.2.2.

\[ \chi^2 = \sum \frac{O^2}{E} - N \]  

8.2.2

where \( N \) is the total number, i.e. \( N = \Sigma O \).

The chi-square test criterion is much used with discrete data. For example, for two-cell tables we often hypothesize a constant probability, say \( p \), that an individual will fall into a certain cell. Thus, the expected values for the two cells become \( np \) and \( n(1-p) \), respectively. For such tables, we may use equation 8.2.3 as a test criterion.

\[ \chi^2 = \frac{(X-np)^2}{np(1-p)} \], 1 df  

8.2.3

(For two-cell tables, it is customary to replace \( N \) by \( n \).)

If we choose to use a ratio, say \( r_1 : r_2 \), in place of a proportion and record the corresponding observed numbers as \( n_1 \) and \( n_2 \), then we may use equation 8.2.4 in place of equation 8.2.1.

\[ \chi^2 = \frac{(r_2n_1 - r_1n_2)^2}{r_1r_2(n_1+n_2)} \], 1 df  

8.2.4

If we express the ratio as \( r : 1 \), with \( r \geq 1 \), and the corresponding observed numbers as \( n_1 \) and \( n_2 \), then equation 8.2.5 may be used in place of equation 8.2.1.

\[ \chi^2 = \frac{(n_1 - rn_2)^2}{r(n_1+n_2)} \]  

8.2.5

Finally for two-cell tables, we may use the normal approximation directly. This is given as equation 8.2.6 and its square gives equation 8.2.1.
8.2.6

\[ z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \]

where \( \hat{p} = n_1/n \).

For 2x2 tables, we test for homogeneity or heterogeneity, independence or interaction. If we represent such tables as follows:

\[
\begin{array}{cc}
  n_{11} & n_{12} \\
  n_{21} & n_{22}
\end{array}
\]

then equation 8.2.7 is a commonly used criterion equivalent to equation 8.2.1.

\[ \chi^2 = \frac{(n_{11}n_{22} - n_{12}n_{21})^2n_{1.}n_{2.}}{n_{1.}n_{2.}n_{1.}n_{2.}} \], 1 df \quad 8.2.7

This test may be thought of as the comparison of two proportions. These may be compared, using a normal approximation, by equation 8.2.8 which is again equivalent to equation 8.2.1.

\[ z = \frac{n_{11} - n_{21}}{\sqrt{n_{1.}n_{2.}(\frac{1}{n_{1.}} + \frac{1}{n_{2.}})}} \quad 8.2.8 \]

For r x c tables, Skory has developed a convenient computational procedure for testing independence. The procedure is based on equation 8.2.9 which is equivalent to equation 8.2.1.

\[ \chi^2 = (\sum \frac{n_{i.j}^2}{n_{i.}n_{.j}} - 1)n_{..}, \ (r-1)(c-1) \ df \quad 8.2.9 \]

Many other computational procedures for special cases have been developed.

Problems

8.2.1. Show that equations 8.2.2 and 8.2.1 are equivalent.

8.2.2. Show that for two cell tables, the deviations O-E are equal but opposite in sign.
8.2.3. Show that equations 8.2.3 and 8.2.1 are equivalent.

8.2.4. Show that equations 8.2.4 and 8.2.1 are equivalent.

8.2.5. Show that equations 8.2.5 and 8.2.1 are equivalent.

8.2.6. Show that \( z^2 \) (equation 8.2.6) = \( x^2 \) (equation 8.2.1).

8.2.7. Show that equations 8.2.7 and 8.2.1 are equivalent.

8.2.8. Show that \( z^2 \) (equation 8.2.8) = \( x^2 \) (equation 8.2.7).

8.2.9. Show that equations 8.2.9 and 8.2.1 are equivalent.

8.2.10. When two trials, each resulting in success or failure, are made on the same subject, the data may be recorded in a four-fold table. If the corner elements, \( n_{11} \) and \( n_{22} \), are considered to yield no relevant information, then it may be sufficient to test \( H_0 : p = 1/2 \) for the remaining cells, \( n_{12} \) and \( n_{21} \). Show that equation 8.2.1 may be written as equation 8.2.10 to test \( H_0 \).

\[
\chi^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}, \quad 1 \text{ df} \quad 8.2.10
\]
CHAPTER IX
THE BINOMIAL AND POISSON DISTRIBUTIONS

9.1. The binomial distribution.

The binomial distribution is defined by equation 9.1.1 which gives the probabilities of the two possible, mutually exclusive, outcomes from a single trial.

\[ f(X) = p^X(1-p)^{1-X}, \quad X=0,1. \]  

9.1.1

If we think of the possible outcomes as "success" and "failure," occurring with probabilities \( p \) and \( 1-p \) respectively, then it is clear from equation 9.1.1 that we have quantified these outcomes as 1 and 0 respectively.

The mean and variance of the variable are given by equations 9.1.2.

\[ \mu = p \quad \text{and} \quad \sigma^2 = p(1-p) \]  

9.1.2

As for so many parent distributions, it is clear that they hold little direct interest for us; we are interested in their offspring, the derived distributions of the sample total, a number between 0 and \( n \), and of the sample mean, a number between 0 and 1. The sample total is clearly the number of successes while the sample mean is the proportion of successes. It is the distribution of the sample total that is generally referred to as the binomial distribution.

We now develop the probability distribution of the sample total.

Assume that we draw independent samples from the binomial distribution defined by equation 9.1.1; the parameter \( p \) remains constant from trial to trial.

**Theorem:** The distribution of the sum of \( n \) trials is given by equation 9.1.3.

\[ f(X) = \left( \frac{n!}{X!(n-X)!} \right) p^X(1-p)^{n-X}, \quad X=0,1,\ldots,n \]  

9.1.3

where \( \left( \frac{n!}{X!} \right) = n! / X!(n-X)! \) for \( n! = n(n-1)\ldots2(1) \). [Definition: \( 0! = 1 \).]

**Proof:** If we consider an event consisting of obtaining \( X \) successes and \( n-X \) failures in a stated order, then \( \text{P (this event)} = p^X(1-p)^{n-X} \) because of independent sampling. This probability holds for each specific event consisting of \( X \) successes and \( n-X \) failures so that we now need only find the number of ways such an event may occur.

For the latter part of the problem, we assume all orderings of \( X \) successes and \( n-X \) failures are equally likely. The validity of this assumption depends upon
independent sampling and a constant $p$. If we consider that we have $n$ boxes into which we must place $X$ cards, none or one to the box, then we have $n$ choices for the first card, $n-1$ for the second, $\ldots$, $n-X+1$ for the $X$-th. Altogether, assignment may be made in $n(n-1)\cdots(n-X+1)$ ways. It is clear that we have proceeded so far as though the cards are distinguishable, which they aren't. Thus if we consider the assignment consisting of the $X$ cards in the first $X$ boxes, we have recorded it $X(X-1)\cdots 1$ times. The same holds for every other particular assignment, hence the appropriate "number of ways" is given by equation 9.1.4.

$$\frac{n(n-1)\cdots(n-X+1)}{X(X-1)\cdots 1} = \binom{n}{X}$$ \hspace{1cm} 9.1.4

Therefore, equation 9.1.3 is valid.

**Theorem:** The mean of the number of successes in $n$ binomial trials is given by equation 9.1.5.

$$\mu = np$$ \hspace{1cm} 9.1.5

**Proof:**

$$E(X) = \sum_0^n \binom{n}{X} p^X (1-p)^{n-X}$$

$$= \sum_0^n \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

$$= \sum_1^n \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

$$= np \sum_{X=1}^{n-1} \frac{(n-1)!}{(X-1)!(n-X)!} p^{X-1} (1-p)^{n-X}$$

$$= np \sum_{X=0}^{n-1} \frac{(n-1)!}{(X-1)!(n-X)!} p^{X-1} (1-p)^{n-X}$$

$$= np (p+[1-p])^{n-1} = np$$

**Theorem:** The variance of the sum of $n$ trials is given by equation 9.1.6.

$$\sigma^2_X = np(1-p)$$ \hspace{1cm} 9.1.6
Equations 9.1.5 and 9.1.6 arise naturally from equations 9.1.2 and are particular examples of more general theorems. (E.g., see problem 3.2.5.)

The mean, that is the proportion of successes, has mean and variance $p$ and $p(-\cdot)/n$ respectively. This information has already been used, in equation 8.2.6, when computing a confidence interval for a proportion.

Problems

9.1.1. Show that equations 9.1.2 are valid.
[Hint: see equations 3.1.2 and 3.2.1.]

9.1.2. Show that equation 9.1.4 is valid.

9.1.3. Show that equation 9.1.6 is valid.
[Hint: Find $E[X(X-1)]$ to begin with.]

9.1.4. Find the ratio $f(X+1)/f(X)$. (This ratio may be used to find probabilities successively. How?)

9.2. The Poisson distribution.

The probability distribution for a Poisson variable is given by equation 9.2.1.

$$f(X) = \frac{e^{-\mu} \mu^X}{X!}, \quad X=0,1,\ldots$$

9.2.1

where $\mu$ is a parameter. This distribution is generally associated with rare events.

Theorem: The mean of a Poisson distribution is the parameter $\mu$; i.e., equation 9.2.2 is valid.

$$E(X) = \mu$$

9.2.2

(Note: this formula has been used as a definition of notation, equation 3.1.1. This is not the case here as $\mu$ represents only the parameter of the Poisson distribution until we have shown it to be the mean.)
Proof:\n\[ E(X) = \sum_{X=0}^{\infty} \frac{X e^{-\mu} \mu^X}{X!} = \mu \sum_{X=1}^{\infty} \frac{e^{-\mu} \mu^{X-1}}{(X-1)!} = \mu \sum_{X=1}^{\infty} \frac{e^{-\mu} \mu^{X-1}}{(X-1)!} = \mu \]

Theorem: The variance of a Poisson distribution is the parameter \( \mu \); i.e., equation 9.2.3 holds.

\[ \text{Var}(X) = E[X - E(X)]^2 = \mu \quad 9.2.3 \]

Problems

9.2.1. Show that equation 9.2.3 holds.

[Hint: Find \( E\{X(X-1)\} \) to begin with.]

9.2.2. Find the ratio \( f(X+1)/f(X) \).
10.1. Introduction.

In the usual methods course, the testing of an hypothesis is generally handled by presenting a test criterion and then justifying it, if justification seems desirable. Actually, there are general theories concerning tests of hypotheses and these theories give rise to test criteria. The much-used test criteria $z$, $t$, $X^2$ and $F$ can arise by use of the maximum likelihood theory. When it is assumed that the underlying distributions are normal, an intermediate result of application of the theory requires that we minimize a residual sum of squares. Thus in the case of the normal distribution, the maximum likelihood theory becomes equivalent to that arising from another theory, the least-squares theory.

The aforementioned theories give rise to a set or system of linear equations (for the problems which concern us) which must be solved. In this chapter, we shall present systems of equations and consider their solution, introducing some elementary notions concerning matrices and their application to the problem of obtaining a solution. In chapter XI we will consider the derivation of least squares equations and their application to statistical problems.

10.2. Two equations in two unknowns.

Consider equation 10.2.1.

$$a_1X + b_1Y + c_1 = 0 \tag{10.2.1}$$

where $a_1$, $b_1$ and $c_1$ are real numbers. This is a nonhomogeneous linear equation in two unknowns, $X$ and $Y$. If $c_1 = 0$, the equation is homogeneous. Equation 10.2.1 can be interpreted as a straight line in the $X,Y$-plane. The line is the locus of all points $(X_i,Y_i)$ which satisfy the equation. Any pair of values which satisfies the equation is said to be a solution of the equation.

Suppose we also have equation 10.2.2.

$$a_2X + b_2Y + c_2 = 0 \tag{10.2.2}$$

Now we have a second line in the $X,Y$-plane. If these two lines intersect, the coordinates of the point of intersection satisfy both equations. The common solution of equations 10.2.1 and 10.2.2 is said to be a solution of the system of equations.
If the second line is parallel to the first, then equations 10.2.1 and 10.2.2 have no common solution and the system has no solution.

If we have a third equation, the three corresponding lines may or may not have a point in common. When there is a common point, two of the equations are sufficient to obtain a solution of the system, the third equation is simply a linear function of the other two and, as such, contains no real additional information. For example, consider the following system of 3 equations in 2 unknowns.

\[
\begin{align*}
X &= 2 \\
Y &= 3 \\
X + 2Y &= 8
\end{align*}
\]

Here, \(1(10.2.3) + 2(10.2.4) = (10.2.5)\). Any two of the equations are sufficient to obtain a solution of the system and, in fact, equations 10.2.3 and 10.2.4 are the solution.

Had equation 10.2.5 been \(2Y = 6\), then equations 10.2.3 and one of 10.2.4 and the new 10.2.5 would have provided a solution. Equations 10.2.4 and the new 10.2.5 would not be sufficient to provide a solution for the system. Thus we notice that some care is required in choosing the equations to be retained for solving the system.

In general, to solve a system of equations with \(n\) unknowns, we require exactly \(n\) equations of the system. However, from the preceding discussion, it is clear that these equations must be subject to certain conditions, i.e., the system must possess certain properties before a unique solution exists. We now proceed to solve the system consisting of equations 10.2.1 and 10.2.2.

An obvious method of solving equations 10.2.1 and 10.2.2 is to obtain \(X\) in terms of \(Y\) from one equation, then substitute the result in the other equation and proceed in the obvious way. Thus equation 10.2.1 become 10.2.6.

\[
X = \frac{-b_1 Y - c_1}{a_1}, \quad \text{provided } a_1 \neq 0
\]

Substituting this in equation 10.2.2 and solving for \(Y\), we obtain equation 10.2.7.

\[
Y = \frac{-\left(a_1 c_2 - a_2 c_1\right)}{a_1 b_2 - a_2 b_1}
\]
Substituting this back into equation 10.2.6, we get equation 10.2.8.

\[ x = \frac{-(c_1b_2 - c_2b_1)}{a_1b_2 - a_2b_1} \]

10.2.8

Obviously, neither equation has meaning unless the common denominator is different from zero, i.e., unless condition 10.2.9 holds.

\[ a_1b_2 - a_2b_1 \neq 0 \]

10.2.9

The pattern exhibited by the numerators and denominators of equations 10.2.7 and 10.2.8 lends itself to the use of a new symbol defining a second-order determinant. This is given in equation 10.2.10.

\[ \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = \frac{a_1a_2}{a_1b_2 - a_2b_1} \]

10.2.10

Using this symbol, we may now write equations 10.2.7 and 10.2.8 as equations 10.2.11.

\[ x = \frac{c_1}{a_1a_2} \quad , \quad y = \frac{c_1}{a_1a_2} \]

10.2.11

In equations 10.2.11, note the common denominator, the minus signs of the numerator, and that the numerator determinants consist of the denominator with the \( a_i \)'s, the coefficients of \( X \) in the system, replaced by the \( c_i \)'s, and the \( b_i \)'s, the coefficients of \( Y \) in the system, replaced by the \( c_i \)'s, in solving for \( X \) and \( Y \) respectively.
10.2.1. Evaluate the following determinants:

\[
\begin{vmatrix}
2 & 3 \\
5 & 7
\end{vmatrix}, \begin{vmatrix}
2 & 5 \\
3 & 7
\end{vmatrix}, \begin{vmatrix}
2+3 & 5 \\
5+3 & 7
\end{vmatrix}
\]

10.2.2. Evaluate:

\[
\begin{vmatrix}
1 & -9 \\
3 & 4
\end{vmatrix}, \begin{vmatrix}
-1 & -9 \\
-3 & 4
\end{vmatrix}, \begin{vmatrix}
-9 & 1 \\
4 & 3
\end{vmatrix}, \begin{vmatrix}
1 & -9 \\
2(3) & 2(4)
\end{vmatrix}
\]

10.2.3. Evaluate:

\[
\begin{vmatrix}
3 & -2 \\
-2 & 4
\end{vmatrix}, \begin{vmatrix}
3 & -2 \\
-2+2(3) & 4+2(-2)
\end{vmatrix}
\]

(Note: the results of problems 10.2.1, 10.2.2, and 10.2.3 suggest generalizations which may be checked in almost any elementary mathematics text which includes a chapter on determinants.)

10.2.4. The following equations arise from the linear regression problem. Write the solution in determinantal form.

\[
\hat{\alpha} + \hat{\beta} = \bar{y}
\]

\[(\Sigma x_i)^{\hat{\alpha}} + (\Sigma x_i^2)^{\hat{\beta}} = \Sigma x_i y_i
\]

10.2.5. Show that the solution of the system of equations given in problem 10.2.3 is:

\[
\hat{\alpha} = \bar{y} - \frac{\Sigma x_i y_i}{\Sigma x_i^2} \bar{x}, \quad \hat{\beta} = \frac{\Sigma x_i y_i}{\Sigma x_i^2}
\]

10.2.6. Show that the solution of equations 10.2.1 and 10.2.2 may be written as follows:

\[
\frac{X}{c_1 b_1} = \frac{Y}{a_1 c_1} = \frac{-1}{a_1 b_1}
\]

\[
\frac{a_2 c_2}{a_2 b_2}
\]
10.2.7. Let \( x' = \frac{X - \bar{X}}{s_x} \) and \( y' = \frac{Y - \bar{Y}}{s_y} \). Show that the regression equations of \( X \) on \( Y \) and of \( Y \) on \( X \) are given by

\[
\begin{vmatrix}
x' & y' \\
r & 1
\end{vmatrix} = 0 \quad \text{and} \quad 
\begin{vmatrix}
1 & r \\
x' & y'
\end{vmatrix} = 0
\]

10.3. Matrices and determinants.

A matrix is a rectangular array of numbers. No operation is required as for a determinant. Thus

\[(X, Y), \quad (X, Y) \quad \text{and} \quad \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}\]

are matrices. The first two matrices could determine a point in a plane, the third matrix is the matrix of coefficients of \( X \) and \( Y \) in equations 10.2.1 and 10.2.2.

The usual arithmetic operations are defined for matrices.

Addition is defined as the addition of elements in corresponding positions. (This assumes the matrices will be of the same dimension.) For example, equation 10.3.1 defined addition of two 2x2 matrices,

\[
\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} + \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix} = \begin{pmatrix} a_1+b_1 & a_2+b_2 \\ a_3+b_3 & a_4+b_4 \end{pmatrix} \tag{10.3.1}
\]

If matrices to be added are identical, then it is convenient to write the result simply as a multiple of the original matrix. Thus, we have equation 10.3.2.

\[
k \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} + \cdots + \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \tag{10.3.2}
\]

k times

Definition 10.3.1 now leads us from equation 10.3.2 to a definition of scalar multiplication, equation 10.3.3 for 2x2 matrices.

\[
k \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} = \begin{pmatrix} ka_1 & ka_2 \\ ka_3 & ka_4 \end{pmatrix} \tag{10.3.3}
\]
Equation 10.3.3 may be used when \( k \) is not an integer and regardless of its sign. For an \( n \times n \) matrix, every element is, then, multiplied by the coefficient \( k \).

**Subtraction** is now defined in terms of addition and scalar multiplication. Equations 10.3.1 and 10.3.3 may be used to illustrate subtraction for 2x2 matrices.

**Multiplication** of matrices is defined in a row-by-column manner. In particular, the element at the intersection of the \( i \)-th row and \( j \)-th column of the product matrix is the sum of products of similarly located elements from the \( i \)-th row of the left and \( j \)-th row of the right matrix. E.g., for multiplication of 2x2 matrices, we have equations 10.3.4 and 10.3.5.

\[
\begin{pmatrix}
  a & b \\
  c & d \\
\end{pmatrix}
\begin{pmatrix}
  e & f \\
  g & h \\
\end{pmatrix}
= 
\begin{pmatrix}
  ae+bg & af+bh \\
  ce+dg & cf+dh \\
\end{pmatrix}
\tag{10.3.4}
\]

\[
\begin{pmatrix}
  a_1 & b_1 \\
  a_2 & b_2 \\
\end{pmatrix}
\begin{pmatrix}
  X \\
  Y \\
\end{pmatrix}
= 
\begin{pmatrix}
  a_1X+b_1Y \\
  a_2X+b_2Y \\
\end{pmatrix}
\tag{10.3.5}
\]

It is clear that we cannot necessarily multiply any two matrices; the number of columns in the left matrix must equal the number of rows in the right matrices. Such matrices are said to be **conformable** for multiplication. The product has as many rows as the left matrix and as many columns as the right. It is now apparent that interchanging the order of the matrices in a product may change the product. Indeed, this second product may not exist as is the case if we change the order of the matrices in equation 10.3.5.

**Division**, if it is defined, is done so in terms of multiplication using inverses. The **inverse** of a matrix is like a reciprocal in that a matrix multiplied by its inverse results in a matrix with ones down the **principal diagonal** (top left to bottom right) and zeros elsewhere the so-called **identity matrix**. An inverse is defined for square matrices only. The superscript \(-1\) is used to denote an inverse. For 2x2 matrices, we have equation 10.3.6.

\[
\begin{pmatrix}
  a & b \\
  c & d \\
\end{pmatrix}
\begin{pmatrix}
  a & b^{-1} \\
  c & d \\
\end{pmatrix}
= 
\begin{pmatrix}
  1 & 0 \\
  0 & 1 \\
\end{pmatrix}
\tag{10.3.6}
\]
An inverse matrix exists provided the determinant of the given matrix is not zero. We may proceed to find the inverse matrix of equation 10.3.6 as follows:

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} ax+dz & cy+dw \\ cx+dz & cy+dw \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
\]

Hence \(ax+dz=1\), \(cy+dw=0\), \(cx+dz=0\) and \(cy+dw=1\). Since the equations are in two pairs, they are easily solved and we find equations 10.3.7.

\[
\begin{align*}
x &= \frac{-1 \ b}{0 \ d}, \\
z &= \frac{a \ -1}{c \ 0}
\end{align*}
\]

\[
\begin{align*}
y &= \frac{-1 \ d}{0 \ b}, \\
w &= \frac{a \ 0}{c \ -1}
\end{align*}
\]

Hence,

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}
\]

Note the pattern of the solution. The numerators may be found, apart from sign, by crossing out the row and column containing the corresponding position in the matrix to be inverted. The common denominator is the determinant of the matrix to be inverted.

Systems of simultaneous equations may be written in matrix form. Note equation 10.3.5. This equation need only be equated to a matrix of constants. Thus, equations 10.2.1 and 10.2.2 may be written as the matrix equation 10.3.8.

\[
\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} -c_1 \\ -c_2 \end{pmatrix}
\]

\[10.3.8\]
Problems

10.3.1. Add the following matrices, completing scalar multiplication as required.

\[
\begin{align*}
(2 & 7) + (9 & 9) \\
(3 & 0) + (-8 & 6) \\
(4 & 8) + (3 & -2) \\
(3 & 8) - 2(4 & -6) \\
(-8 & 6) - 3(-3 & 0) \\
(3 & -2) - 3(4 & -2)
\end{align*}
\]

10.3.2. Multiply the following matrices.

\[
\begin{align*}
\begin{pmatrix} 3 & 0 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 4 & 5 \\ 3 & 8 \end{pmatrix}, & \begin{pmatrix} 4 & 5 \\ 3 & 8 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 1 & 5 \end{pmatrix}, \\
\begin{pmatrix} 4 & 3 \\ 5 & 8 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 0 & 5 \end{pmatrix}, & \begin{pmatrix} 4 & 7 & 2 \\ 3 & 9 & 1 \end{pmatrix} \begin{pmatrix} 6 \\ -7 \\ 2 \end{pmatrix}
\end{align*}
\]

10.3.3. Find inverses for the following matrices.

\[
\begin{align*}
\begin{pmatrix} 3 & 5 \\ 5 & 9 \end{pmatrix}, & \begin{pmatrix} 4 & -3 \\ -3 & 7 \end{pmatrix}
\end{align*}
\]

10.3.4. Show that the inverse of \( A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \) is both a right and left inverse. That is, show that \( AA^{-1} = I = A^{-1}A \) for 2x2 matrices.

10.3.5. Write the equations of problem 10.2.4 as a matrix equation.

10.3.6. Show that a solution of the equations of problem 10.2.4 is given by the following matrix equation.

\[
\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = A^{-1} \begin{pmatrix} \bar{y} \\ 0X_1Y_1 \end{pmatrix}
\]
10.4. Three equations in three unknowns.

The solution of systems of linear equations by matrix methods, as in
problem 10.3.6, is a general one. The number of equations must equal the number
of unknowns and the determinant of the coefficients of the unknowns must be non-
zero. The latter condition is equivalent to saying that no equation may be a
linear combination of the others; such an equation supplies no information not
contained in the others. A system of three equations in three unknowns may be
written as the matrix equation 10.4.1.

\[
\begin{pmatrix}
X_1 \\
X_2 \\
X_3
\end{pmatrix}
= \begin{pmatrix}
c_1 \\
c_2 \\
c_3
\end{pmatrix}
\text{ where } A = \begin{pmatrix}
a & b & c \\
d & e & f \\
g & h & i
\end{pmatrix}.
\]

10.4.1

In turn, the solution of equation 10.4.1 is equation 10.4.2.

\[
\begin{pmatrix}
X_1 \\
X_2 \\
X_3
\end{pmatrix}
= A^{-1} \begin{pmatrix}
c_1 \\
c_2 \\
c_3
\end{pmatrix}
\]

10.4.2

In the process of finding \( A^{-1} \), we require the determinant of \( A \), denoted by \( |A| \). This is given by equation 10.4.3.

\[
|A| = \begin{vmatrix}
a & b & c \\
d & e & f \\
g & h & i
\end{vmatrix} = a \begin{vmatrix}
e & f \\
h & i
\end{vmatrix} - b \begin{vmatrix}
d & f \\
g & i
\end{vmatrix} + c \begin{vmatrix}
d & e \\
g & h
\end{vmatrix}
\]

10.4.3

\[
= aei - afh - bdi + bfg + cdh - ceg
\]

Since 2x2 determinants are already defined by equation 10.2.11, we complete the
computation as shown. This definition can be generalized to \( n \times n \) matrices, the
generalization being to expand by use of the first row of elements (alternating
signs) and corresponding \((n-1)(n-1)\) determinants, the definition of a determinant
always being given in terms of determinants with one less row and column.
The 3x3 determinant may be evaluated as indicated below, a procedure which cannot be generalized further.

\[
\begin{vmatrix}
  a & b & c \\
  d & e & f \\
  g & h & i \\
\end{vmatrix}
\]

Note the repetition of the first two columns.

These products with + signs.

These products with - signs.

The 2x2 determinants of equation 10.4.3 are called the minors of a, b and c respectively. With the appropriate sign, minors become cofactors; the appropriate sign for the cofactor of the element a_{ij} in |A| is \((-1)^{i+j}\).

The inverse of A is the matrix of cofactors, each divided by |A|.

While A\(^{-1}\) may be used directly to solve a system of linear equations (see equation 10.4.2), an alternative method of writing the solution is available. We begin with the general system given by equation 10.4.4.

\[
A \begin{pmatrix}
  x_1 \\
  \vdots \\
  x_n \\
\end{pmatrix} = \begin{pmatrix}
  c_1 \\
  \vdots \\
  c_n \\
\end{pmatrix}
\text{ or } AX = C
\]

10.4.4

(With reference to equations 10.2.1 and 10.2.2, the \(c_i\)'s in equation 10.4.4 are now changed in sign.) The solution may be written as \(X = A^{-1}C\). In addition, we have equation 10.4.5.

\[
(X_1, \ldots, X_n) = \frac{1}{|A_i|} (|A_1|, \ldots, |A_n|)
\]

10.4.5

|A_i| is defined as follows. In A, replace the i-th column by the coefficients in C. This type of solution was used for 2 equations in 2 unknowns as equations 10.3.7. Note that 10.4.5 makes use of scalar multiplication. (See equation 10.3.3). Equation 10.4.5 is known as Cramer's rule.
Problems

10.4.1. Evaluate the following determinants.

\[
\begin{vmatrix}
1 & 1 & 1 \\
3 & 0 & 3 \\
2 & 8 & 9
\end{vmatrix},
\begin{vmatrix}
1 & 2 & 3 \\
0 & 4 & 6 \\
0 & 1 & 2
\end{vmatrix},
\begin{vmatrix}
1 & 3 & 2 \\
4 & 0 & 4 \\
2 & 6 & 4
\end{vmatrix}
\]

10.4.2. Show that the following equation is valid.

\[
\begin{pmatrix}
\begin{vmatrix}
e & f \\
h & i
\end{vmatrix} & -\begin{vmatrix}
b & c \\
h & i
\end{vmatrix} & \begin{vmatrix}
b & c \\
e & f
\end{vmatrix} \\
\begin{vmatrix}
d & f \\
g & i
\end{vmatrix} & -\begin{vmatrix}
a & c \\
g & i
\end{vmatrix} & \begin{vmatrix}
a & c \\
d & f
\end{vmatrix} \\
\begin{vmatrix}
d & e \\
g & h
\end{vmatrix} & -\begin{vmatrix}
a & b \\
g & h
\end{vmatrix} & \begin{vmatrix}
a & b \\
d & e
\end{vmatrix}
\end{pmatrix} = \begin{pmatrix}
\begin{vmatrix}
A \\
0
\end{vmatrix} & 0 \\
0 & \begin{vmatrix}
A \\
0
\end{vmatrix}
\end{pmatrix}
\]

where

\[
A = \begin{pmatrix}
a & b & c \\
d & e & f \\
g & h & i
\end{pmatrix}
\]

Satisfy yourself that the signed 2x2 determinants are cofactors of A.

10.4.3. Use Cramer's rule to obtain a solution for the system of equations

\[a_1X + b_1Y + c_1Z + d_1 = 0, \ i = 1, 2, 3.\]
11.1. Rules for obtaining least squares equations.

In Chapters VI and VII in connection with various linear models used to describe observations, it is implied that the basic procedure for estimating $\sigma^2$ is to first estimate the various parameters of the model, then replace the parameters in the model by their estimates, and finally compute a so-called residual sum of squares or sum of squares of the deviations; the estimation procedure is to be one such that the residual sum of squares is a minimum. A step in the procedure requires us to set the sum of the residuals or deviations equal to zero (equation for $\hat{\mu}$).

The algebraic description used with the general linear model may be written as equation 11.1.1, the sum of a mean and a random component.

$$X_i = \mu_i + \epsilon_i$$  \hspace{1cm} 11.1.1

The mean $\mu_i$ is, in turn, usually written as a sum of components, each of which is described in the model. For multi-way classifications, it is invariably convenient to use multiple subscripts. If we let $\hat{\mu}_i$ represent an estimate of $\mu_i$, then equation 11.1.2 describes our choice of the set of $\hat{\mu}_i$'s.

$$\Sigma(X_i - \hat{\mu}_i)^2 = \text{minimum}$$  \hspace{1cm} 11.1.2

To obtain the $\hat{\mu}_i$'s for which equation 11.1.2 holds, we must solve a set of equations equal in number to the number of parameters to be estimated. These may be obtained by the following rules:

1. Write the expression for the general residual as many times as there are parameters to be estimated.

2. Multiply these expressions successively by the coefficients of the parameters to be estimated.

3. Precede each of the resulting expressions by a summation sign, put $^\wedge$'s on the parameters to be estimated, and equate each expression to zero. [In any summation, remember that you are estimating a particular parameter and, on this account, may have fixed one or more subscripts. Hence, summation is not over these subscripts.]
Example 11.1.1.

Model: $X_1 = \mu + \epsilon_1$

Rule 1: $X_1 - \mu$, only one parameter

Rule 2: $-1(X_1 - \mu)$, coefficient is $-1$

Rule 3: $\Sigma(X_1 - \mu) = 0$, since obviously $-1 \neq 0$.

Example 11.1.2.

Model: $Y_1 = \alpha + \beta x_1 + \epsilon_1$

Rule 1: $Y_1 - \alpha - \beta x_1$, (2 times, $x_1$ are observable parameters)

Rule 2: $-1(Y_1 - \alpha - \beta x_1)$, for the $\alpha$-equation

$-x_1(Y_1 - \alpha - \beta x_1)$, for the $\beta$-equation

Rule 3: $\Sigma(Y_1 - \alpha - \beta x_1) = 0$

Ex. $(Y_1 - \alpha - \beta x_1) = 0$

Example 11.1.3.

Model: $X_{ij} = \mu + \rho_i + \tau_j + \epsilon_{ij}$

Rule 1: $X_{ij} - \mu - \rho_i - \tau_j$, for $\hat{\mu}$

$X_{ij} - \mu - \rho_i - \tau_j$ for $\hat{\rho}_1$

$X_{ij} - \mu - \rho_i - \tau_j$ for $\hat{\rho}_r$

$X_{ij} - \mu - \rho_i - \tau_j$ for $\hat{\tau}_1$

$X_{ij} - \mu - \rho_i - \tau_j$ for $\hat{\tau}_t$

or

$X_{ij} - \mu - \rho_i - \tau_j$ for $\hat{\mu}$

$X_{ij} - \mu - \rho_i - \tau_j$ for $\hat{\rho}_i$, $i = 1, \ldots, r$

$X_{ij} - \mu - \rho_i - \tau_j$ for $\hat{\tau}_j$, $j = 1, \ldots, t$
Rules 2 and 3:
\[
\sum_{i,j} (X_{ij} - \hat{\mu} - \hat{\beta}_i - \hat{\tau}_j) = 0, \quad \mu \text{ is in every } X_{ij}
\]
\[
\sum_{j} (X_{1j} - \hat{\mu} - \hat{\beta}_1 - \hat{\tau}_j) = 0, \quad i \text{ is fixed at 1}
\]
\[
\sum_{j} (X_{rj} - \hat{\mu} - \hat{\beta}_r - \hat{\tau}_j) = 0, \quad i \text{ is fixed at } r
\]
\[
\sum_{i} (X_{1i} - \hat{\mu} - \hat{\beta}_1 - \hat{\tau}_i) = 0, \quad j \text{ is fixed at 1}
\]
\[
\sum_{i} (X_{ti} - \hat{\mu} - \hat{\beta}_i - \hat{\tau}_t) = 0, \quad j \text{ is fixed at } t
\]

Problems

11.1.1. For example 11.1.1, show that \(\hat{\mu} = \bar{x}\).

11.1.2. For example 11.1.2, show that the least squares equations may be written as equations 11.1.3.

\[
\sum_{i} y_i - \hat{\alpha} = 0 \tag{11.1.3}
\]
\[
\sum_{i} X_{ij} y_i - \hat{\beta} \sum_{i} X_{ij} = 0
\]

Solve equations 11.1.3 for \(\hat{\alpha}\) and \(\hat{\beta}\). Compare your results with those found for problem 10.2.5 and comment.

11.1.3. Write out the least squares equations for the completely random design with \(n_i\) observations on the \(i\)-th treatment. Reduce to simple form by application of Theorems 2.4.1, 2.4.2, and 2.4.3.

11.1.4. Write out the least squares equations for the latin-square design. Reduce to simple form.

11.1.5. Consider the model \(Y_i = \mu + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik1} + \epsilon_i\). Write the least squares equations for this model. Reduce to simple form and write the equations for the \(\hat{\beta}\)'s in matrix notation. [HINT: \(\hat{\mu}\) must appear on the right hand side of each equation, in the constant term.]
11.1.6. Consider the model \( Y = \mu + \beta_1 x_1 + \beta_2 (x_1^2 - \Sigma x^2/n) + \epsilon \). Write the least squares equations. Write the equations for \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) in matrix notation and solve by determinantal methods. [Note that the estimate of \( \beta_1 \) is different from the estimate of \( \beta \) in problem 11.1.2.]

11.2. Restrictions for solutions to exist.

Equations obtained by the procedure of section 11.2 will be found to be insoluble in the case of experimental designs. For example, let equation 11.2.1 hold.

\[
X_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i=1,2 \quad \text{and} \quad j=1,\ldots,n \quad 11.2.1
\]

Following the rules previously set out, we obtain equations 11.2.2.

\[
\begin{align*}
\Sigma (x_{ij} - \mu - \hat{\tau}_1) &= 0 \quad \text{or} \quad 2n\hat{\mu} + n\hat{\tau}_1 + n\hat{\tau}_2 = X_1. \\
\Sigma (x_{1j} - \mu - \hat{\tau}_1) &= 0 \quad \text{or} \quad n\hat{\mu} + n\hat{\tau}_1 = X_1. \\
\end{align*} \quad 11.2.2
\]

In matrix notation, equations 11.2.2 are 11.2.3.

\[
\begin{bmatrix} 2n & n & n \\ n & n & 0 \\ n & 0 & n \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad 11.2.3
\]

For a unique solution to exist, the matrix of coefficients must have an inverse. For an inverse to exist, the determinant of this matrix must be non-zero. However,

\[
\begin{vmatrix} 2n & n & n \\ n & n & 0 \\ n & 0 & n \end{vmatrix} = 2n^3 - n^3 - n^3 = 0,
\]

and there is no unique solution.
The preceding conclusion might well have been anticipated since it is clear that the first equation of the set is simply the sum of the other two so contributes no additional information. These two equations may be written as equations 11.2.4.

\[ \hat{\mu} + \hat{\tau}_1 = \bar{x}_1. \]
\[ \hat{\mu} + \hat{\tau}_2 = \bar{x}_2. \]

These equations can be solved provided we impose some condition on our estimates. For example, if we set \( \hat{\tau}_1 + \hat{\tau}_2 = 0 \), then \( \hat{\mu} = \bar{x}, \hat{\tau}_1 = \bar{x}_1 - \bar{x}, \) and \( \hat{\tau}_2 = \bar{x}_2 - \bar{x} \). The chosen restriction seems reasonable in that, for the fixed model, we set \( E(\tau_i) = 0 \) and, for the random model, we set \( E(\tau_i) = 0 \). However, any restriction on the estimates will permit a solution of equations 11.2.2.

Problems

11.2.1. For example 11.1.3, the usual restrictions are \( \sum \hat{\tau}_i = 0 = \sum \hat{\tau}_j \). Use these restrictions and obtain estimates of the parameters.

11.2.2. For example 11.1.3, impose the restrictions that \( \hat{\tau}_1 = 0 = \hat{\tau}_1 \) and obtain estimates of the parameters. [Note: the analyses of variance will be the same regardless of which set of restrictions is used.]

11.2.3. Solve the equations found in problem 11.1.3.

11.2.4. Solve the equations found in problem 11.1.4.

11.3. The residual sum of squares.

A residual sum of squares is most often found by subtracting sums of squares attributable to the appropriate sources of variation from the total sum of squares. The reductions are also said to be attributable to fitting constants in regression or design. Appropriate formulas for the reductions and residuals have been discussed in several instances, for example, equations 4.2.4, 5.2.3, 6.4.5, and 7.1.8, in connection with which, it was simply stated that the residual sum of squares was a minimum. However, the various sums of squares were seen to be independent, a property which does not hold for all
analysis of variance.

We shall now consider a general procedure for computing a residual sum of squares by examining a particular application of it. Consider the randomized complete block design. The residual sum of squares is given by equation 11.3.1.

\[
\text{Residual SS} = \sum_{i,j} (\hat{Y}_{\cdot ij} - \hat{\mu} - \hat{\rho}_{\cdot i} - \hat{\tau}_{\cdot j})^2
\]

In equation 11.3.1, note that the coefficient of \( \hat{\mu} \) is \( \sum_{i,j} X_{\cdot ij} \), the same quantity that appears in the \( \hat{\mu} \) equation of example 11.1.3, the coefficient of \( \hat{\rho}_i \) is \( \sum_j X_{\cdot ij} \), the same quantity that appears in the \( \hat{\rho}_i \) equation, and so on. Estimates of all parameters appear in equation 11.3.1.

Equation 11.3.1 generalizes in the obvious way to give the residual sum of squares for any least squares analysis, regardless of whether or not an experimental design was used.

To evaluate equation 11.3.1, it is only necessary to provide the estimates. To this end, take the equations of example 11.1.3 plus the equations \( \sum \hat{\rho} = 0 = \sum \hat{\tau} \). It is now easy to show that equations 11.3.2 are a solution.

\[
\hat{\mu} = \bar{X}, \quad \hat{\rho}_i = \bar{X}_{\cdot i} - \bar{X}, \quad \hat{\tau}_j = \bar{X}_{\cdot j} - \bar{X}
\]

Problems

11.3.1. Use equations 11.3.2 to evaluate equation 11.3.1, relating each term in your final answer to a source of variation.

11.3.2. Write out the formula for the residual sum of squares in the completely random design. Use the solutions obtained in equation 11.2.3 to evaluate the expression, relating each term in your final answer to a source of variation.

11.3.3. Write out the formula for the residual sum of squares for a latin-square design. Evaluate.

11.3.4. Write out the formula for the residual sum of squares for the linear regression problem with one independent variable. Evaluate, using the results of problem 11.1.2.
11.3.5. Write out the formula for the residual sum of squares where the model of problem 11.1.6 holds. Compare your result with that obtained for problem 11.3.4; in particular note that the product of $\hat{\beta}_1$ and its coefficient differs from that of $\hat{\beta}$ and its coefficient.

11.4. The "minimum" or "least" property.

We now have rules for obtaining least squares estimates of parameters, section 11.1, and an easily-generalized equation for finding the least sum of squares, equation 11.3.1. We have agreed to believe the solution of these equations is such that equation 11.4.1 holds.

$$\Sigma \hat{\epsilon}^2 = \min$$  \hspace{1cm} 11.4.1

where $\hat{\epsilon}$ is the observed value minus the least squares estimate of its expected value. We now show that the residual sum of squares as computed by equation 11.3.1 for the randomized complete block design is a minimum sum of squares. (It is, in fact, the minimum sum of squares.)

Equation 11.4.2 is the estimate of a deviation, or random component, for a randomized complete block design.

$$X_{ij} - \hat{\beta}_i - \hat{\gamma}_j = \hat{\epsilon}_{ij}$$  \hspace{1cm} 11.4.2

Multiply equation 11.4.2 by $X_{ij}$ and sum to get equation 11.4.3.

$$\Sigma X_i^2 - \Sigma \hat{\gamma}X_i - \Sigma \hat{\beta}_r i_i - \Sigma X_{ij} = \Sigma \hat{\epsilon}_{ij}$$

11.4.3

This reduces to equation 11.4.4.

$$\Sigma X_{ij}^2 - \hat{\gamma} \Sigma X_{ij} - \Sigma \hat{\beta}_r i_i X_{ij} = \Sigma \hat{\epsilon}_{ij} X_{ij}$$

11.4.4

The left side of this equation is the Residual SS of equation 11.3.1. Assuming that equation 11.4.1 is valid, we need only prove that equation 11.4.5 is valid in order that 11.3.1 be valid.

$$\Sigma \hat{\epsilon}_{ij} X_{ij} = \Sigma \hat{\epsilon}^2_{ij}$$

11.4.5
Multiply equation 11.4.2 by $\hat{e}_{ij}$ and sum. This gives equation 11.4.6.

$$\sum_{i,j} X_{ij} \hat{e}_{ij} = \sum_{i,j} \hat{e}_{ij} - \sum_{i,j} \hat{e}_{ij} = \sum_{i,j} \hat{e}_{ij}^2$$  \hspace{1cm} 11.4.6

But, from the least squares equations of example 11.3.3, we have equations 11.4.7.

$$\sum_{i,j} \hat{e}_{ij} = \sum_{i,j} (X_{ij} - \hat{\mu} - \hat{\beta}_i - \hat{\gamma}_j) = 0$$  \hspace{1cm} 11.4.7

From this, equation 11.4.5 follows and 11.3.1 is seen to be valid.

Let us also consider the general linear regression problem. It is apparent that, corresponding to equation 11.3.1, we have equation 11.4.8.

Residual $SS = \sum_{i,j} (Y_{ij} - \hat{\mu} - \hat{\beta}_i - \hat{\gamma}_j)^2 \sum_{i} \alpha_{\alpha 1} \alpha_{\alpha 1} Y_{i1}$  \hspace{1cm} 11.4.8

Also, the $\hat{\mu}$ equation is given by equation 11.4.9 and states that the sum of the deviations is zero.

$$\sum_{\alpha} \hat{\beta}_\alpha Y_{i1} = \sum_{\alpha} \hat{\beta}_\alpha X_{i1} \alpha_{\alpha 1} \alpha_{\alpha 1} Y_{i1}$$  \hspace{1cm} 11.4.9

Note that this also states that $\sum \hat{Y}_i = 0$ and that, as a consequence, $\hat{\mu} = \overline{Y}$.

Multiplying $\hat{\mu}$ by $Y_i$ and summing, we obtain equation 11.4.10.

$$\sum_{i} \hat{\beta}_\alpha Y_{i1} = \sum_{i} \hat{\beta}_\alpha X_{i1} \alpha_{\alpha 1} \alpha_{\alpha 1} Y_{i1}$$  \hspace{1cm} 11.4.10

This is the right side of equation 11.4.8.

Multiplying $\hat{\mu}$ by $\hat{\alpha} = Y_i - \hat{\mu} - \hat{\beta}_\alpha X_{i1}$ and summing, we obtain equation 11.4.11.

$$\sum_{i} \hat{\alpha} Y_{i1} = \sum_{i} \hat{\alpha} X_{i1} \alpha_{\alpha 1} \alpha_{\alpha 1} \hat{\beta} Y_{i1}$$  \hspace{1cm} 11.4.11

Since only the first term on the right is non-zero, we have shown that equation 11.4.8 is the residual sum of squares and, as the result of using a minimum sum of squares procedure, is the minimum sum of squares.
Problems

11.4.1. Show that the terms in equation 11.4.11 which are preceded by
negative signs, are all zero. [Hint: Use the results obtained in equation
11.1.5.]

11.5. Experimental design and multiple regression.

The general multiple regression problem covers design, regression and
covariance as a single problem. We will now illustrate the relation for a
very simple case. The general regression model calls for equation 11.5.1.

\[ Y_i = \mu + \sum \alpha X_{\alpha} \]

[In general, it is more convenient to use \( x_{\alpha i} \).]

For a randomized complete block design with two blocks and two treatments, we
write equation 11.5.2.

\[ X_{ij} = \mu + \rho_i + \tau_j + \epsilon_{ij}, \quad i=1,2, \quad j=1,2 \]

However, equation 11.5.2 could be written as equation 11.5.3.

\[ Y_i = \mu + \rho_i X_{1i} + \rho_2 X_{2i} + \tau_1 X_{3i} + \tau_2 X_{4i}, \quad i=1,\ldots,4 \]

where

\[ X_{11} = 1 = X_{12}, \quad X_{13} = 0 = X_{14}, \]
\[ X_{21} = 0 = X_{22}, \quad X_{23} = 1 = X_{24}, \]
\[ X_{31} = 1 = X_{33}, \quad X_{32} = 0 = X_{34}, \]
\[ X_{41} = 0 = X_{43}, \quad X_{42} = 1 = X_{44}. \]

In equation 11.5.3, we have essentially equation 11.5.1. The remaining \( x_{\alpha i} \)
equations tell precisely what values the \( x_{\alpha i} \)'s are to have

Problem

11.5.1. Let \( i=1,2,3 \) and 4 and relate the results to equation 11.5.2. Note
how much simpler and more helpful equation 11.5.2 is, as compared with equation
11.5.3.