Chapter

Multiple and Partial Correlation

MULTIPLE CORRELATION

The degree of relationship existing between three or more variables is called *multiple corre* The fundamental principles involved in problems of multiple correlation are analogous to th simple correlation, as treated in Chapter 14.

SUBSCRIPT NOTATION

To allow for generalizations to large numbers of variables, it is convenient to adopt a n involving subscripts.

We shall let $X_1, X_2, X_3, \ldots$ denote the variables under consideration. Then we $X_{11}, X_{12}, X_{13}, \ldots$ denote the values assumed by the variable X_1, and $X_{21}, X_{22}, X_{23}, \ldots$ deno values assumed by the variable X_2, and so on. With this notation, a sum such as $X_{21} + X_{22} + X$ X_{2N} could be written $\sum_{j=1}^{N} X_{2j}$, $\sum_j X_{2j}$, or simply $\sum X_2$. When no ambiguity can result, we use notation. In such case the mean of X_2 is written $\bar{X}_2 = \sum X_2 / N$.

REGRESSION EQUATIONS AND REGRESSION PLANES

A *regression equation* is an equation for estimating a dependent variable, say X_1, fro independent variables $X_2, X_3, \ldots$ and is called a *regression equation of* X_1 *on* $X_2, X_3, \ldots$ In func notation this is sometimes written briefly as $X_1 = F(X_2, X_3, \ldots)$ (read "X_1 is a function of X_2, X so on").

For the case of three variables, the simplest regression equation of X_1 on X_2 and X_3 has th

$$X_1 = b_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$$

where $b_{1.23}$, $b_{12.3}$, and $b_{13.2}$ are constants. If we keep X_3 constant in equation (*1*), the grap versus X_2 is a straight line with slope $b_{12.3}$. If we keep X_2 constant, the graph of X_1 versus straight line with slope $b_{13.2}$. It is clear that the subscripts after the dot indicate the variab constant in each case.

Due to the fact that X_1 varies partially because of variation in X_2 and partially because of in X_3, we call $b_{12.3}$ and $b_{13.2}$ the *partial regression coefficients* of X_1 on X_2 keeping X_3 constan X_1 on X_3 keeping X_2 constant, respectively.

Equation (*1*) is called a *linear regression equation* of X_1 on X_2 and X_3. In a three-dim rectangular coordinate system it represents a plane called a *regression plane* and is a generaliz the regression line for two variables, as considered in Chapter 13.

NORMAL EQUATIONS FOR THE LEAST-SQUARES REGRESSION PLANE

Just as there exist least-squares regression lines approximating a set of N data points $(X, Y$ two-dimensional scatter diagram, so also there exist *least-squares regression planes* fitting a se data points (X_1, X_2, X_3) in a three-dimensional scatter diagram.

The least-squares regression plane of X_1 on X_2 and X_3 has the equation (*1*) where $b_{1.23}$, $b_{12.3}$, and $b_{13.2}$ are determined by solving simultaneously the *normal equations*

$$\begin{aligned} \sum X_1 &= b_{1.23}N + b_{12.3}\sum X_2 + b_{13.2}\sum X_3 \\ \sum X_1X_2 &= b_{1.23}\sum X_2 + b_{12.3}\sum X_2^2 + b_{13.2}\sum X_2X_3 \\ \sum X_1X_3 &= b_{1.23}\sum X_3 + b_{12.3}\sum X_2X_3 + b_{13.2}\sum X_3^2 \end{aligned} \tag{2}$$

These can be obtained formally by multiplying both sides of equation (*1*) by 1, X_2, and X_3 successively and summing on both sides.

Unless otherwise specified, whenever we refer to a regression equation it will be assumed that the least-squares regression equation is meant.

If $x_1 = X_1 - \bar{X}_1$, $x_2 = X_2 - \bar{X}_2$, and $x_3 = X_3 - \bar{X}_3$, the regression equation of X_1 on X_2 and X_3 can be written more simply as

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3 \tag{3}$$

where $b_{12.3}$ and $b_{13.2}$ are obtained by solving simultaneously the equations

$$\begin{aligned} \sum x_1x_2 &= b_{12.3}\sum x_2^2 + b_{13.2}\sum x_2x_3 \\ \sum x_1x_3 &= b_{12.3}\sum x_2x_3 + b_{13.2}\sum x_3^2 \end{aligned} \tag{4}$$

These equations which are equivalent to the normal equations (*2*) can be obtained formally by multiplying both sides of equation (*3*) by x_2 and x_3 successively and summing on both sides (see Problem 15.8).

REGRESSION PLANES AND CORRELATION COEFFICIENTS

If the linear correlation coefficients between variables X_1 and X_2, X_1 and X_3, and X_2 and X_3, as computed in Chapter 14, are denoted respectively by r_{12}, r_{13}, and r_{23} (sometimes called *zero-order correlation coefficients*), then the least-squares regression plane has the equation

$$\frac{x_1}{s_1} = \left(\frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2}\right)\frac{x_2}{s_2} + \left(\frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2}\right)\frac{x_3}{s_3} \tag{5}$$

where $x_1 = X - \bar{X}_1$, $x_2 = X_2 - \bar{X}_2$, and $x_3 = X_3 - \bar{X}_3$ and where s_1, s_2, and s_3 are the standard deviations of X_1, X_2, and X_3, respectively (see Problem 15.9).

Note that if the variable X_3 is nonexistent and if $X_1 = Y$ and $X_2 = X$, then equation (*5*) reduces to equation (*25*) of Chapter 14.

STANDARD ERROR OF ESTIMATE

By an obvious generalization of equation (*8*) of Chapter 14, we can define the *standard error of estimate* of X_1 on X_2 and X_3 by

$$s_{1.23} = \sqrt{\frac{\sum (X_1 - X_{1,\text{est}})^2}{N}} \tag{6}$$

where $X_{1,\text{est}}$ indicates the estimated values of X_1 as calculated from the regression equations (*1*) or (*5*).

In terms of the correlation coefficients r_{12}, r_{13}, and r_{23}, the standard error of estimate can also be computed from the result

$$s_{1.23} = s_1\sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}} \tag{7}$$

The sampling interpretation of the standard error of estimate for two variables as given on page 296 for the case when N is large can be extended to three dimensions by replacing the lines parallel

to the regression line with planes parallel to the regression plane. A better estimate of the
standard error of estimate is given by $\hat{s}_{1.23} = \sqrt{N/(N-3)}\, s_{1.23}$.

COEFFICIENT OF MULTIPLE CORRELATION

The coefficient of multiple correlation is defined by an extension of equation (12) of Chapter 14. In the case of two independent variables, for example, the *coefficient of multiple* is given by

$$R_{1.23} = \sqrt{1 - \frac{s_{1.23}^2}{s_1^2}}$$

where s_1 is the standard deviation of the variable X_1 and $s_{1.23}$ is given by equation (6). quantity $R_{1.23}^2$ is called the *coefficient of multiple determination.*

When a linear regression equation is used, the coefficient of multiple correlation is *coefficient of linear multiple correlation.* Unless otherwise specified, whenever we refer correlation, we shall imply linear multiple correlation.

In terms of r_{12}, r_{13}, and r_{23}, equation (8) can also be written

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}$$

A coefficient of multiple correlation, such as $R_{1.23}$, lies between 0 and 1. The closer it is to 1, better is the linear relationship between the variables. The closer it is to 0, the worse is the relationship. If the coefficient of multiple correlation is 1, the correlation is called *perfect.* Altho correlation coefficient of 0 indicates no linear relationship between the variables, it is possible *nonlinear relationship* may exist.

CHANGE OF DEPENDENT VARIABLE

The above results hold when X_1 is considered the dependent variable. However, if we consider X_3 (for example) to be the dependent variable instead of X_1, we would only have the subscripts 1 with 3, and 3 with 1, in the formulas already obtained. For example, the equation of X_3 on X_1 and X_2 would be

$$\frac{x_3}{s_3} = \left(\frac{r_{23} - r_{13}r_{12}}{1 - r_{12}^2}\right)\frac{x_2}{s_2} + \left(\frac{r_{13} - r_{23}r_{12}}{1 - r_{12}^2}\right)\frac{x_1}{s_1}$$

as obtained from equation (5), using the results $r_{[illegible]} = r_{[illegible]}$, $r_{[illegible]} = r_{[illegible]}$, and $r_{[illegible]} = r_{[illegible]}$

GENERALIZATIONS TO MORE THAN THREE VARIABLES

These are obtained by analogy with the above results. For example, the linear regression equation of X_1 on X_2, X_3, and X_4 can be written

$$X_1 = b_{1.234} + b_{12.34}X_2 + b_{13.24}X_3 + b_{14.23}X_4 \tag{11}$$

and represents a *hyperplane in four-dimensional space.* By formally multiplying both sides of equation (11) by 1, X_2, X_3, and X_4 successively and then summing on both sides, we obtain the normal equations for determining $b_{1.234}$, $b_{12.34}$, $b_{13.24}$, and $b_{14.23}$; substituting these in equation (11) then gives us the *least-squares regression equation of X_1 on X_2, X_3, and X_4*. This least-squares regression equation can be written in a form similar to that of equation (5). (See Problem 15.41.)

PARTIAL CORRELATION

It is often important to measure the correlation between a dependent variable and one particul independent variable when all other variables involved are kept constant; that is, when the effects all other variables are removed (often indicated by the phrase "other things being equal"). This c be obtained by defining a *coefficient of partial correlation*, as in equation (*12*) of Chapter 14, exce that we must consider the explained and unexplained variations that arise both with and without t particular independent variable.

If we denote by $r_{12.3}$ the coefficient of partial correlation between X_1 and X_2 keeping X_3 constar we find that

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} \qquad (12)$$

Similarly, if $r_{12.34}$ is the coefficient of partial correlation between X_1 and X_2 keeping X_3 and X_4 consta then

$$r_{12.34} = \frac{r_{12.4} - r_{13.4}r_{23.4}}{\sqrt{(1-r_{13.4}^2)(1-r_{23.4}^2)}} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{\sqrt{(1-r_{14.3}^2)(1-r_{24.3}^2)}} \qquad (13)$$

These results are useful since by means of them any partial correlation coefficient can ultimately b made to depend on the correlation coefficients r_{12}, r_{23}, etc. (i.e., the *zero-order correlation coefficients*)

In the case of two variables, X and Y, if the two regression lines have equations $Y = a_0 + a_1X$ an $X = b_0 + b_1Y$, we have seen that $r^2 = a_1b_1$ (see Problem 14.22). This result can be generalized. Fo example, if

$$X_1 = b_{1.234} + b_{12.34}X_2 + b_{13.24}X_3 + b_{14.23}X_4 \qquad (14)$$

and

$$X_4 = b_{4.123} + b_{41.23}X_1 + b_{42.13}X_2 + b_{43.12}X_3 \qquad (15)$$

are linear regression equations of X_1 on X_2, X_3, and X_4 and of X_4 on X_1, X_2, and X_3, respectively, then

$$r_{14.23}^2 = b_{14.23}b_{41.23} \qquad (16)$$

(see Problem 15.18). This can be taken as the starting point for a definition of linear partial correlation coefficients.

RELATIONSHIPS BETWEEN MULTIPLE AND PARTIAL CORRELATION COEFFICIENTS

Interesting results connecting the multiple correlation coefficients can be found. For example, we find that

$$1 - R_{1.23}^2 = (1-r_{12}^2)(1-r_{13.2}^2) \qquad (17)$$

$$1 - R_{1.234}^2 = (1-r_{12}^2)(1-r_{13.2}^2)(1-r_{14.23}^2) \qquad (18)$$

Generalizations of these results are easily made.

NONLINEAR MULTIPLE REGRESSION

The above results for linear multiple regression can be extended to nonlinear multiple regression. Coefficients of multiple and partial correlation can then be defined by methods similar to those given above.

Solved Problems

REGRESSION EQUATIONS INVOLVING THREE VARIABLES

15.1 Using an appropriate subscript notation, write the regression equations of (*a*) X_2 on X_1 and X_3; (*b*) X_3 on X_1, X_2, and X_4; and (*c*) X_5 on X_1, X_2, X_3, and X_4.

SOLUTION

(*a*) $X_2 = b_{2.13} + b_{21.3}X_1 + b_{23.1}X_3$

(*b*) $X_3 = b_{3.124} + b_{31.24}X_1 + b_{32.14}X_2 + b_{34.12}X_4$

(*c*) $X_5 = b_{5.1234} + b_{51.234}X_1 + b_{52.134}X_2 + b_{53.124}X_3 + b_{54.123}X_4$

15.2 Write the normal equations corresponding to the regression equations (*a*) $X_3 = b_{3.12} + b_{31.2}X_1 + b_{32.1}X_2$ and (*b*) $X_1 = b_{1.234} + b_{12.34}X_2 + b_{13.24}X_3 + b_{14.23}X_4$.

SOLUTION

(*a*) Multiply the equation successively by 1, X_1, and X_2, and sum on both sides. The normal equations are

[illegible]

$$\sum X_1X_3 = b_{3.12}\sum X_1 + b_{31.2}\sum X_1^2 + b_{32.1}\sum X_1X_2$$
$$\sum X_2X_3 = b_{3.12}\sum X_2 + b_{31.2}\sum X_1X_2 + b_{32.1}\sum X_2^2$$

(*b*) Multiply the equation successively by 1, X_2, X_3, and X_4, and sum on both sides. The normal [illegible] are

$$\begin{aligned}
\sum X_1 &= b_{1.234}N + b_{12.34}\sum X_2 + b_{13.24}\sum X_3 + b_{14.23}\sum X_4 \\
\sum X_1X_2 &= b_{1.234}\sum X_2 + b_{12.34}\sum X_2^2 + b_{13.24}\sum X_2X_3 + b_{14.23}\sum X_2X_4 \\
\sum X_1X_3 &= b_{1.234}\sum X_3 + b_{12.34}\sum X_2X_3 + b_{13.24}\sum X_3^2 + b_{14.23}\sum X_3X_4 \\
\sum X_1X_4 &= b_{1.234}\sum X_4 + b_{12.34}\sum X_2X_4 + b_{13.24}\sum X_3X_4 + b_{14.23}\sum X_4^2
\end{aligned}$$

Note that these are not derivations of the normal equations, but only formal means for [illegible] them.

The number of normal equations is equal to the number of unknown constants.

15.3 The variable X_1 is thought to be a linear function of X_2 and X_3. A sample of 12 pairs of [illegible] (X_2, X_3) produced the values of X_1 shown in Table 15.1.

(*a*) Find the least-squares regression equation of X_1 on X_2 and X_3.

(*b*) Determine the estimated values of X_1 from the given values of X_2 and X_3.

(*c*) Estimate X_1 when $X_2 = 54$ and $X_3 = 9$.

Table 15.1

X_1	64	71	53	67	55	58	77	57	56	51	76	[illegible]
X_2	57	59	49	62	51	50	55	48	52	42	61	[illegible]
X_3	8	10	6	11	8	7	10	9	10	6	12	[illegible]

SOLUTION

(*a*) The linear regression equation of X_1 on X_2 and X_3 can be written

$$X_1 = b_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$$

The normal equations of the least-squares regression equation are

$$\begin{aligned}\sum X_1 &= b_{1.23}N + b_{12.3}\sum X_2 + b_{13.2}\sum X_3\\ \sum X_1X_2 &= b_{1.23}\sum X_2 + b_{12.3}\sum X_2^2 + b_{13.2}\sum X_2X_3\\ \sum X_1X_3 &= b_{1.23}\sum X_3 + b_{12.3}\sum X_2X_3 + b_{13.2}\sum X_3^2\end{aligned} \qquad (19)$$

The work involved in computing the sums can be arranged as in Table 15.2. (Although the column headed X_1^2 is not needed at present, it has been added for future reference.)

Table 15.2

X_1	X_2	X_3	X_1^2	X_2^2	X_3^2	X_1X_2	X_1X_3	X_2X_3
64	57	8	4096	3249	64	3648	512	456
71	59	10	5041	3481	100	4189	710	590
53	49	6	2809	2401	36	2597	318	294
67	62	11	4489	3844	121	4154	737	682
55	51	8	3025	2601	64	2805	440	408
58	50	7	3364	2500	49	2900	406	350
77	55	10	5929	3025	100	4235	770	550
57	48	9	3249	2304	81	2736	513	432
56	52	10	3136	2704	100	2912	560	520
51	42	6	2601	1764	36	2142	306	252
76	61	12	5776	3721	144	4636	912	732
68	57	9	4624	3249	81	3876	612	513
$\sum X_1$ $=753$	$\sum X_2$ $=643$	$\sum X_3$ $=106$	$\sum X_1^2$ $=48{,}139$	$\sum X_2^2$ $=34{,}843$	$\sum X_3^2$ $=976$	$\sum X_1X_2$ $=40{,}830$	$\sum X_1X_3$ $=6796$	$\sum X_2X_3$ $=5779$

Using Table 15.2, the normal equations (19) become

$$\begin{aligned}12b_{1.23} + 643b_{12.3} + 106b_{13.2} &= 753\\ 643b_{1.23} + 34{,}843b_{12.3} + 5{,}779b_{13.2} &= 40{,}830\\ 106b_{1.23} + 5{,}779b_{12.3} + 976b_{13.2} &= 6{,}796\end{aligned} \qquad (20)$$

Solving, $b_{1.23}=3.6512$, $b_{12.3}=0.8546$, and $b_{13.2}=1.5063$, and the required regression equation is

$$X_1 = 3.6512 + 0.8546X_2 + 1.5063X_3 \quad \text{or} \quad X_1 = 3.65 + 0.855X_2 + 1.506X_3 \qquad (21)$$

For another method, which avoids solving simultaneous equations, see Problem 15.6.

(b) Using the regression equation (21), we obtain the estimated values of X_1, denoted by $X_{1,\text{est}}$, by substituting the corresponding values of X_2 and X_3. For example, substituting $X_2=57$ and $X_3=8$ in (21), we find $X_{1,\text{est}}=64.414$.

The other estimated values of X_1 are obtained similarly. They are given in Table 15.3 together with the sample values of X_1.

(c) Putting $X_2=54$ and $X_3=9$ in equation (21), the estimate is $X_{1,\text{est}}=63.356$, or about 63.

Table 15.3

$X_{1,\text{est}}$	64.414	69.136	54.564	73.206	59.286	56.925	65.717	58.229	63.153	48.582	73.857	65.920
X_1	64	71	53	67	55	58	77	57	56	51	76	68

15.4 Calculate the standard deviations (*a*) s_1, (*b*) s_2, and (*c*) s_3 for the data of Problem 15.3.

SOLUTION

(*a*) The quantity s_1 is the standard deviation of the variable X_1. Then, using Table 15.2 of Problem 15.3(*a*) and the methods of Chapter 4, we find

$$s_1 = \sqrt{\frac{\sum X_1^2}{N} - \left(\frac{\sum X_1}{N}\right)^2} = \sqrt{\frac{48{,}139}{12} - \left(\frac{753}{12}\right)^2} = 8.6035 \quad \text{or} \quad 8.6$$

(*b*)
$$s_2 = \sqrt{\frac{\sum X_2^2}{N} - \left(\frac{\sum X_2}{N}\right)^2} = \sqrt{\frac{34{,}843}{12} - \left(\frac{643}{12}\right)^2} = 5.6930 \quad \text{or} \quad 5.7$$

(*c*)
$$s_3 = \sqrt{\frac{\sum X_3^2}{N} - \left(\frac{\sum X_3}{N}\right)^2} = \sqrt{\frac{976}{12} - \left(\frac{106}{12}\right)^2} = 1.8181 \quad \text{or} \quad 1.8$$

15.5 Compute (*a*) r_{12}, (*b*) r_{13}, and (*c*) r_{23} for the data of Problem 15.3.

SOLUTION

(*a*) The quantity r_{12} is the linear correlation coefficient between the variables X_1 and X_2, ignoring the variable X_3. Then, using the methods of Chapter 14, we have

$$r_{12} = \frac{N\sum X_1X_2 - (\sum X_1)(\sum X_2)}{\sqrt{[N\sum X_1^2 - (\sum X_1)^2][N\sum X_2^2 - (\sum X_2)^2]}}$$

$$= \frac{(12)(40{,}830) - (753)(643)}{\sqrt{[(12)(48{,}139) - (753)^2][(12)(34{,}843) - (643)^2]}} = 0.8196 \quad \text{or} \quad 0.82$$

(*b*) and (*c*) Using corresponding formulas, we obtain $r_{12} = 0.7698$, or 0.77, and $r_{23} = 0.7984$, or 0.80.

15.6 Work Problem 15.3(*a*) by using equation (*5*) of this chapter and the results of Problems 15.4 and 15.5.

SOLUTION

The regression equation of X_1 on X_2 and X_3 is, on multiplying both sides of equation (*5*) by s_1,

$$x_1 = \left(\frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2}\right)\left(\frac{s_1}{s_2}\right)x_2 + \left(\frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2}\right)\left(\frac{s_1}{s_3}\right)x_3 \qquad (22)$$

where $x_1 = X_1 - \bar{X}_1$, $x_2 = X_2 - \bar{X}_2$, and $x_3 = X_3 - \bar{X}_3$. Using the results of Problems 15.4 and 15.5, equation (*22*) becomes

$$x_1 = 0.8546x_2 + 1.5063x_3$$

Since $\bar{X}_1 = \frac{\sum X_1}{N} = \frac{753}{12} = 62.750$ $\quad \bar{X}_2 = \frac{\sum X_2}{N} = 53.583$ $\quad$ and $\quad \bar{X}_3 = 8.833$

(from Table 15.2 of Problem 15.3), the required equation can be written

$$X_1 - 62.750 = 0.8546(X_2 - 53.583) + 1.506(X_3 - 8.833)$$

agreeing with the result of Problem 15.3(*a*).

15.7 For the data of Problem 15.3, determine (*a*) the average increase in X_1 per unit increase in X_2 for constant X_3 and (*b*) the average increase in X_1 per unit increase in X_3 for constant X_2.

SOLUTION

From the regression equation obtained in Problem 15.3(*a*) or 15.6 we see that the answer to (*a*) is 0.8546, or about 0.9, and that the answer to (*b*) is 1.5063, or about 1.5.

15.8 Show that equations (3) and (4) of this chapter follow from equations (1) and (2).

SOLUTION

From the first of equations (2), on dividing both sides by N, we have

$$\bar{X}_1 = b_{1.23} + b_{12.3}\bar{X}_2 + b_{13.2}\bar{X}_3 \tag{23}$$

Subtracting equation (23) from equation (1) gives

$$X_1 - \bar{X}_1 = b_{12.3}(X_2 - \bar{X}_2) + b_{13.2}(X_3 - \bar{X}_3)$$

or

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3 \tag{24}$$

which is equation (3).

Let $X_1 = x_1 + \bar{X}_1$, $X_2 = x_2 + \bar{X}_2$, and $X_3 = x_3 + \bar{X}_3$ in the second and third of equations (2). Then after some algebraic simplifications, using the results $\sum x_1 = \sum x_2 = \sum x_3 = 0$, they become

$$\sum x_1x_2 = b_{12.3}\sum x_2^2 + b_{13.2}\sum x_2x_3 + N\bar{X}_2[b_{1.23} + b_{12.3}\bar{X}_2 + b_{13.2}\bar{X}_3 - \bar{X}_1] \tag{25}$$
$$\sum x_1x_3 = b_{12.3}\sum x_2x_3 + b_{13.2}\sum x_3^2 + N\bar{X}_3[b_{1.23} + b_{12.3}\bar{X}_2 + b_{13.2}\bar{X}_3 - \bar{X}_1] \tag{26}$$

which reduce to equations (4) since the quantities in brackets on the right-hand sides of equations (25) and (26) are zero because of equation (1).

Another method

See Problem 15.30.

15.9 Establish equation (5), repeated here:

$$\frac{x_1}{s_1} = \left(\frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2}\right)\frac{x_2}{s_2} + \left(\frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2}\right)\frac{x_3}{s_3} \tag{5}$$

SOLUTION

From equations (25) and (26)

$$\begin{aligned} b_{12.3}\sum x_2^2 + b_{13.2}\sum x_2x_3 &= \sum x_1x_2 \\ b_{12.3}\sum x_2x_3 + b_{13.2}\sum x_3^2 &= \sum x_1x_3 \end{aligned} \tag{27}$$

Since

$$s_2^2 = \frac{\sum x_2^2}{N} \quad \text{and} \quad s_3^2 = \frac{\sum x_3^2}{N}$$

$\sum x_2^2 = Ns_2^2$ and $\sum x_3^2 = Ns_3^2$. Since

$$r_{23} = \frac{\sum x_2x_3}{\sqrt{(\sum x_2^2)(\sum x_3^2)}} = \frac{\sum x_2x_3}{Ns_2s_3}$$

$\sum x_2x_3 = Ns_2s_3r_{23}$. Similarly, $\sum x_1x_2 = Ns_1s_2r_{12}$ and $\sum x_1x_3 = Ns_1s_3r_{13}$.

Substituting in (27) and simplifying, we find

$$\begin{aligned} b_{12.3}s_2 + b_{13.2}s_3r_{23} &= s_1r_{12} \\ b_{12.3}s_2r_{23} + b_{13.2}s_3 &= s_1r_{13} \end{aligned} \tag{28}$$

Solving equations (28) simultaneously, we have

$$b_{12.3} = \left(\frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2}\right)\left(\frac{s_1}{s_2}\right) \quad \text{and} \quad b_{13.2} = \left(\frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2}\right)\left(\frac{s_1}{s_3}\right)$$

Substituting these in the equation $x_1 = b_{12.3}x_2 + b_{13.2}x_3$ [equation (24)] and dividing by s_1 yields the required result.

STANDARD ERROR OF ESTIMATE

15.10 Compute the standard error of estimate of X_1 on X_2 and X_3 for the data of Problem 15.3.

SOLUTION

From Table 15.3 of Problem 15.3(b) we have

$$s_{1.23} = \sqrt{\frac{\sum (X_1 - X_{1,\text{est}})^2}{N}}$$

$$= \sqrt{\frac{(64-64.414)^2+(71-69.136)^2+\cdots+(68-65.920)^2}{12}} = 4.6447 \qquad \text{or} \qquad 4.6$$

The population standard error of estimate is estimated by $\hat{s}_{1.23} = \sqrt{N/(N-3)}\, s_{1.23} = 5.3$ in this case.

[illegible]

$$s_{1.23} = s_1 \sqrt{\frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}$$

SOLUTION

From Problems 15.4(a) and 15.5 we have

$$s_{1.23} = 8.6035 \sqrt{\frac{1-(0.8196)^2-(0.7698)^2-(0.7984)^2+2(0.8196)(0.7698)(0.7984)}{1-(0.7984)^2}} = 4.6$$

Note that by the method of this problem the standard error of estimate can be found without the regression equation.

COEFFICIENT OF MULTIPLE CORRELATION

15.12 Compute the coefficient of linear multiple correlation of X_1 on X_2 and X_3 for the data of Problem 15.3.

SOLUTION

First method

From the results of Problems 15.4(a) and 15.10 we have

$$R_{1.23} = \sqrt{1 - \frac{s_{1.23}^2}{s_1^2}} = \sqrt{1 - \frac{(4.6447)^2}{(8.6035)^2}} = 0.8418$$

Second method

From the results of Problem 15.5 we have

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}} = \sqrt{\frac{(0.8196)^2+(0.7698)^2-2(0.8196)(0.7698)(0.7984)}{1-(0.7984)^2}} = 0.8418$$

Note that the coefficient of multiple correlation, $R_{1.23}$, is larger than either of the coefficients r_{12} or r_{13} (see Problem 15.5). This is always true and is in fact to be expected, since by taking into account additional relevant independent variables we should arrive at a better relationship between the variables.

15.13 Compute the coefficient of multiple determination of X_1 on X_2 and X_3 for the data of Problem 15.3.

SOLUTION

The coefficient of multiple determination of X_1 on X_2 and X_3 is

$$R_{1.23}^2 = (0.8418)^2 = 0.7086$$

using Problem 15.12. Thus about 71% of the total variation in X_1 is explained by using the regression equation.

5.14 For the data of Problem 15.3, calculate (*a*) $R_{2.13}$ and (*b*) $R_{3.12}$ and compare their values with the value of $R_{1.23}$.

SOLUTION

(*a*) $$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{13}^2}} = \sqrt{\frac{(0.8196)^2 + (0.7984)^2 - 2(0.8196)(0.7698)(0.7984)}{1 - (0.7698)^2}} = 0.8606$$

(*b*) $$R_{3.12} = \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{12}^2}} = \sqrt{\frac{(0.7698)^2 + (0.7984)^2 - 2(0.8196)(0.7698)(0.7984)}{1 - (0.8196)^2}} = 0.8234$$

This problem illustrates the fact that, in general, $R_{2.13}$, $R_{3.12}$, and $R_{1.23}$ are not necessarily equal, as seen by comparison with Problem 15.12.

5 If $R_{1.23} = 1$, prove that (*a*) $R_{2.13} = 1$ and (*b*) $R_{3.12} = 1$.

SOLUTION

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}} \qquad (29)$$

and

$$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{13}^2}} \qquad (30)$$

(*a*) In equation (*29*), setting $R_{1.23} = 1$ and squaring both sides, $r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23} = 1 - r_{23}^2$. Then

$$r_{12}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} = 1 - r_{13}^2 \qquad \text{or} \qquad \frac{r_{12}^2 + r_{23}^2 - 2r_1 r_{13}r_{23}}{1 - r_{13}^2} = 1$$

That is, $R_{2.13}^2 = 1$ or $R_{2.13} = 1$, since the coefficient of multiple correlation is considered nonnegative.

(*b*) $R_{3.12} = 1$ follows from part (*a*) by interchanging subscripts 2 and 3 in the result $R_{2.13} = 1$.

If $R_{1.23} = 0$, does it necessarily follow that $R_{2.13} = 0$?

SOLUTION

From equation (*29*), $R_{1.23} = 0$ if and only if

$$r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23} = 0 \qquad \text{or} \qquad 2r_{12}r_{13}r_{23} = r_{12}^2 + r_{13}^2$$

hen from equation (*30*) we have

$$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - (r_{12}^2 + r_{13}^2)}{1 - r_{13}^2}} = \sqrt{\frac{r_{23}^2 - r_{13}^2}{1 - r_{13}^2}}$$

ich is not necessarily zero.

CORRELATION

r the data of Problem 15.3, compute the coefficients of linear partial correlation (*a*) $r_{12.3}$, (*b*) 2, and (*c*) $r_{23.1}$.

SOLUTION

$$r_{12.3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} \qquad r_{13.2}=\frac{r_{13}-r_{12}r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}} \qquad r_{23.1}=\frac{r_{23}-r_{12}r_{13}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}}$$

Using the results of Problem 15.5, we find that $r_{12.3}=0.5334$, $r_{13.2}=0.3346$, and $r_{23.1}=0.4580$. It follows that for constant X_3, the correlation coefficient between X_1 and X_2 is 0.53; for constant X_2, the correlation coefficient between X_1 and X_3 is only 0.33. Since these results are based on a small sample of only 12 pairs of values, they are of course not as reliable as those which would be obtained from a larger sample.

15.18 If $X_1=b_{1.23}+b_{12.3}X_2+b_{13.2}X_3$ and $X_3=b_{3.12}+b_{32.1}X_2+b_{31.2}X_1$ are the regression equations of X_1 on X_2 and X_3 and of X_3 on X_2 and X_1, respectively, prove that $r_{13.2}^2=b_{13.2}b_{31.2}$.

SOLUTION

The regression equation of X_1 on X_2 and X_3 can be written [see equation (5) of this chapter]

$$X_1=\bar{X}_1=\left(\frac{r_{12}-r_{13}r_{23}}{1-r_{23}^2}\right)\left(\frac{s_1}{s_2}\right)(X_2-\bar{X}_2)+\left(\frac{r_{13}-r_{12}r_{23}}{1-r_{23}^2}\right)\left(\frac{s_1}{s_3}\right)(X_3-\bar{X}_3)$$

The regression equation of X_3 on X_2 and X_1 can be written [see equation (10)]

$$X_3-\bar{X}_3=\left(\frac{r_{23}-r_{13}r_{12}}{1-r_{12}^2}\right)\left(\frac{s_3}{s_2}\right)(X_2-\bar{X}_2)+\left(\frac{r_{13}-r_{23}r_{12}}{1-r_{12}^2}\right)\left(\frac{s_3}{s_1}\right)(X_1-\bar{X}_1)$$

From equations (31) and (32) the coefficients of X_3 and X_1 are, respectively,

$$b_{13.2}=\left(\frac{r_{13}-r_{12}r_{23}}{1-r_{23}^2}\right)\left(\frac{s_1}{s_3}\right) \qquad \text{and} \qquad b_{31.2}=\left(\frac{r_{13}-r_{23}r_{12}}{1-r_{12}^2}\right)\left(\frac{s_3}{s_1}\right)$$

Thus

$$b_{13.2}b_{31.2}=\frac{(r_{13}-r_{12}r_{23})^2}{(1-r_{23}^2)(1-r_{12}^2)}=r_{13.2}^2$$

15.19 If $r_{12.3}=0$, prove that

$$(a)\quad r_{13.2}=r_{13}\sqrt{\frac{1-r_{23}^2}{1-r_{12}^2}} \qquad\qquad (b)\quad r_{23.1}=r_{23}\sqrt{\frac{1-r_{13}^2}{1-r_{12}^2}}$$

SOLUTION

If

$$r_{12.3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}=0$$

we have $r_{12}=r_{13}r_{23}$.

(a)

$$r_{13.2}=\frac{r_{13}-r_{12}r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}}=\frac{r_{13}-(r_{13}r_{23})r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}}=\frac{r_{13}(1-r_{23}^2)}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}}=r_{13}\sqrt{\frac{1-r_{23}^2}{1-r_{12}^2}}$$

(b) Interchange the subscripts 1 and 2 in the result of part (a).

MULTIPLE AND PARTIAL CORRELATION INVOLVING FOUR OR MORE VARIABLES

15.20 A college entrance examination consisted of three tests: in mathematics, English, and general knowledge. To test the ability of the examination to predict performance in a statistics course, data concerning a sample of 200 students were gathered and analyzed. Letting

$X_1=$ grade in statistics course $\qquad X_3=$ score on English test

$X_2=$ score on mathematics test $\qquad X_4=$ score on general knowledge test the following calculations were obtained:

$$\bar{X}_1 = 75 \qquad s_1 = 10 \qquad \bar{X}_2 = 24 \qquad s_2 = 5$$
$$\bar{X}_3 = 15 \qquad s_3 = 3 \qquad \bar{X}_4 = 36 \qquad s_4 = 6$$
$$r_{12} = 0.90 \qquad r_{13} = 0.75 \qquad r_{14} = 0.80 \qquad r_{23} = 0.70 \qquad r_{24} = 0.70 \qquad r_{34} = 0.85$$

Find the least-squares regression equation of X_1 on X_2, X_3, and X_4.

SOLUTION

Generalizing the result of Problem 15.8, we can write the least-squares regression equation of X_1 on X_2, X_3, and X_4 in the form

$$x_1 = b_{12.34}x_2 + b_{13.24}x_3 + b_{14.23}x_4 \tag{33}$$

where $b_{12.34}$, $b_{13.24}$, and $b_{14.23}$ can be obtained from the normal equations

$$\begin{aligned}
\sum x_1x_2 &= b_{12.34}\sum x_2^2 + b_{13.24}\sum x_2x_3 + b_{14.23}\sum x_2x_4\\
\sum x_1x_3 &= b_{12.34}\sum x_2x_3 + b_{13.24}\sum x_3^2 + b_{14.23}\sum x_3x_4\\
\sum x_1x_4 &= b_{12.34}\sum x_2x_4 + b_{13.24}\sum x_3x_4 + b_{14.23}\sum x_4^2
\end{aligned} \tag{34}$$

and where $x_1 = X_1 - \bar{X}_1$, $x_2 = X_2 - \bar{X}_2$, $x_3 = X_3 - \bar{X}_3$, and $x_4 = X_4 - \bar{X}_4$.

From the given data, we find

$$\begin{aligned}
&\sum x_2^2 = Ns_2^2 = 5000 \qquad &&\sum x_1x_2 = Ns_1s_2r_{12} = 9000 \qquad &&\sum x_2x_3 = Ns_1s_3r_{23} = 2100\\
&\sum x_3^2 = Ns_3^2 = 1800 &&\sum x_1x_3 = Ns_1s_3r_{13} = 4500 &&\sum x_2x_4 = Ns_2s_4r_{24} = 4200\\
&\sum x_4^2 = Ns_4^2 = 7200 &&\sum x_1x_4 = Ns_1s_4r_{14} = 9600 &&\sum x_3x_4 = Ns_3s_4r_{34} = 3060
\end{aligned}$$

Putting these results into equations (*34*) and solving, we obtain

$$b_{12.34} = 1.3333 \qquad b_{13.24} = 0.0000 \qquad b_{14.23} = 0.5556 \tag{35}$$

which, when substituted in equation (*33*), yield the required regression equation

$$x_1 = 1.3333x_2 + 0.0000x_3 + 0.5556x_4$$

or

$$X_1 - 75 = 1.3333(X_2 - 24) + 0.5556(X_4 - 27)$$

or

$$X_1 = 22.9999 + 1.3333X_2 + 0.5556X_4 \tag{36}$$

An exact solution of equations (*34*) yields $b_{12.34} = \frac{4}{3}$, $b_{13.24} = 0$, and $b_{14.23} = \frac{5}{9}$, so that the regression equation can also be written

$$X_1 = 23 + \tfrac{4}{3}X_2 + \tfrac{5}{9}X_4 \tag{37}$$

It is interesting to note that the regression equation does not involve the score in English, namely, X_3. This does not mean that one's knowledge of English has no bearing on proficiency in statistics. Instead, it means that the need for English, insofar as prediction of the statistics grade is concerned, is amply evidenced by the scores achieved on the other tests.

15.21 Two students taking the college entrance examination of Problem 15.20 receive respective scores of (*a*) 30 in mathematics, 18 in English, and 32 in general knowledge; and (*b*) 18 in mathematics, 20 in English, and 36 in general knowledge. What would be their predicted grades in statistics?

SOLUTION

(*a*) Substituting $X_2 = 30$, $X_3 = 18$, and $X_4 = 32$ in equation (*37*), the predicted grade in statistics is $X_1 = 81$.

(*b*) Proceeding as in part (*a*) with $X_2 = 18$, $X_3 = 20$, and $X_4 = 36$, we find $X_1 = 67$.

15.22 For the data of Problem 15.20, find the partial correlation coefficients (*a*) $r_{12.34}$, (*b*) $r_{13.24}$, and (*c*) $r_{14.23}$.

SOLUTION

(*a*) and (*b*) $r_{12.4}=\dfrac{r_{12}-r_{14}r_{24}}{\sqrt{(1-r_{14}^2)(1-r_{24}^2)}}$ $\quad r_{13.4}=\dfrac{r_{13}-r_{14}r_{34}}{\sqrt{(1-r_{14}^2)(1-r_{34}^2)}}$ $\quad r_{23.4}=\dfrac{r_{23}-r_{24}r_{34}}{\sqrt{(1-r_{24}^2)(1-r_{34}^2)}}$

Substituting the values from Problem 15.20, we obtain $r_{12.4}=0.7935$, $r_{13.4}=0.2215$, and $r_{23.4}=0.2791$. Thus

$$r_{12.34}=\frac{r_{12.4}-r_{13.4}r_{23.4}}{\sqrt{(1-r_{13.4}^2)(1-r_{23.4}^2)}}=0.7814 \qquad \text{and} \qquad r_{13.24}=\frac{r_{13.4}-r_{12.4}r_{23.4}}{\sqrt{(1-r_{12.4}^2)(1-r_{23.4}^2)}}=0.0000$$

(*c*) $r_{14.3}=\dfrac{r_{14}-r_{13}r_{34}}{\sqrt{(1-r_{13}^2)(1-r_{34}^2)}}$ $\quad r_{12.3}=\dfrac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$ $\quad r_{24.3}=\dfrac{r_{24}-r_{23}r_{34}}{\sqrt{(1-r_{23}^2)(1-r_{34}^2)}}$

Substituting the values from Problem 15.20, we obtain $r_{14.3}=0.4664$, $r_{12.3}=0.7939$, and $r_{24.3}=$ [illegible] Thus

$$r_{14.23}=\frac{r_{14.3}-r_{12.3}r_{24.3}}{\sqrt{(1-r_{12.3}^2)(1-r_{24.3}^2)}}=0.4193$$

15.23 Interpret the partial correlation coefficients (*a*) $r_{12.4}$, (*b*) $r_{13.4}$, (*c*) $r_{12.34}$, (*d*) $r_{14.3}$, and (*e*) [illegible] obtained in Problem 15.22.

SOLUTION

(*a*) $r_{12.4}=0.7935$ represents the (linear) correlation coefficient between statistics grades and mathem[illegible] scores for students having the same general knowledge scores. In obtaining this coefficient, sco[illegible] English (as well as other factors that have not been taken into account) are not considered, [illegible] evidenced by the fact that the subscript 3 is omitted.

(*b*) $r_{13.4}=0.2215$ represents the correlation coefficient between statistics grades and English scor[illegible] students having the same general knowledge scores. Here, scores in mathematics have no[illegible] considered.

(*c*) $r_{12.34}=0.7814$ represents the correlation coefficient between statistics grades and mathematics scor[illegible] for students having both the same English scores and general knowledge scores.

(*d*) $r_{14.3}=0.4664$ represents the correlation coefficient between statistics grades and general knowled[illegible] scores for students having the same English scores.

(*e*) $r_{14.23}=0.4193$ represents the correlation coefficient between statistics grades and general knowle[illegible] scores for students having both the same mathematics scores and English scores.

15.24 (*a*) For the data of Problem 15.20, show that

$$\frac{r_{12.4}-r_{13.4}r_{23.4}}{\sqrt{(1-r_{13.4}^2)(1-r_{23.4}^2)}}=\frac{r_{12.3}-r_{14.3}r_{24.3}}{\sqrt{(1-r_{14.3}^2)(1-r_{24.3}^2)}}$$

(*b*) Explain the significance of the equality in part (*a*).

SOLUTION

(*a*) The left-hand side of equation (*38*) is evaluated in Problem 15.22(*a*), yielding the result 0.7814. [illegible] evaluate the right-hand side of equation (*38*), use the results of Problem 15.22(*c*); again, the re[illegible] is 0.7814. Thus the equality holds in this special case. It can be shown by direct algebraic proce[illegible] that the equality holds in general.

(*b*) The left side of equation (*38*) is $r_{12.34}$, and the right side is $r_{12.43}$. Since $r_{12.34}$ is the correlation betwe[illegible] variables X_1 and X_2 keeping X_3 and X_4 constant, while $r_{12.43}$ is the correlation between X_1 and X_2 keeping X_4 and X_3 constant, it is at once evident why the equality should hold.

15.25 For the data of Problem 15.20, find (a) the multiple correlation coefficient $R_{1.234}$ and (b) the standard error of estimate $s_{1.234}$.

SOLUTION

(a) $$1 - R^2_{1.234} = (1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23}) \qquad \text{or} \qquad R_{1.234} = 0.9310$$

since $r_{12} = 0.90$ from Problem 15.20, $r_{14.23} = 0.4193$ from Problem 15.22(c), and

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r^2_{12})(1 - r^2_{23})}} = \frac{0.75 - (0.90)(0.70)}{\sqrt{[1 - (0.90)^2][1 - (0.70)^2]}} = 0.3855 .$$

Another method

Interchanging subscripts 2 and 4 in the first equation yields

$$1 - R^2_{1.234} = (1 - r^2_{14})(1 - r^2_{13.4})(1 - r^2_{12.34}) \qquad \text{or} \qquad R_{1.234} = 0.9310$$

where the results of Problem 15.22(a) are used directly.

(b) $$R_{1.234} = \sqrt{\frac{1 - s^2_{1.234}}{s^2_1}} \qquad \text{or} \qquad s_{1.234} = s_1\sqrt{1 - R^2_{1.234}} = 10\sqrt{1 - (0.9310)^2} = 3.650$$

Compare with equation (8) of this chapter.

Supplementary Problems

REGRESSION EQUATIONS INVOLVING THREE VARIABLES

15.26 Using an appropriate subscript notation, write the regression equations of (a) X_3 on X_1 and X_2 and (b) X_4 on X_1, X_2, X_3, and X_5.

15.27 Write the normal equations corresponding to the regression equations of (a) X_2 on X_1 and X_3 and (b) X_5 on X_1, X_2, X_3, and X_4.

15.28 Table 15.4 shows the corresponding values of three variables: X_1, X_2, and X_3.

(a) Find the least-squares regression equation of X_3 on X_1 and X_2.

(b) Estimate X_3 when $X_1 = 10$ and $X_2 = 6$.

Table 15.4

X_1	3	5	6	8	12	14
X_2	16	10	7	4	3	2
X_3	90	72	54	42	30	12

15.29 An instructor of mathematics wished to determine the relationship of grades on a final examination to grades on two quizzes given during the semester. Calling X_1, X_2, and X_3 the grades of a student on the first quiz, second quiz, and final examination, respectively, he made the following computations for a total of 120 students:

$$\bar{X}_1 = 6.8 \qquad \bar{X}_2 = 7.0 \qquad \bar{X}_3 = 74$$
$$s_1 = 1.0 \qquad s_2 = 0.80 \qquad s_3 = 9.0$$
$$r_{12} = 0.60 \qquad r_{13} = 0.70 \qquad r_{23} = 0.65$$

(a) Find the least-squares regression equation of X_3 on X_1 and X_2.
(b) Estimate the final grades of two students whose respective scores on the two quizzes were (1) [illegible] 7 and (2) 4 and 8.

15.30 Work Problem 15.8 by choosing the variables X_2 and X_3 so that $\sum X_2 = \sum X_3 = 0$.

STANDARD ERROR OF ESTIMATE

15.31 For the data of Problem 15.28, find the standard error of estimate of X_3 on X_1 and X_2.

15.32 For the data of Problem 15.29, find the standard error of estimate of (a) X_3 on X_1 and X_2 and (b) X_1 on X_2 and X_3.

COEFFICIENT OF MULTIPLE CORRELATION

15.33 For the data of Problem 15.28, compute the coefficient of linear multiple correlation of X_3 on X_1 and X_2.

15.34 For the data of Problem 15.29, compute (a) $R_{3.12}$, (b) $R_{1.23}$, and (c) $R_{2.13}$.

15.35 (a) If $r_{12} = r_{13} = r_{23} = r \neq 1$, show that

$$R_{1.23} = R_{2.31} = R_{3.12} = \frac{r\sqrt{2}}{\sqrt{1+r}}$$

(b) Discuss the case $r = 1$.

15.36 If $R_{1.23} = 0$, prove that $|r_{23}| \geq |r_{12}|$ and $|r_{23}| \geq |r_{13}|$ and interpret.

PARTIAL CORRELATION

15.37 Compute the coefficients of linear partial correlation (a) $r_{12.3}$, (b) $r_{13.2}$, and (c) $r_{23.1}$ for the data of Problem 15.28 and interpret your answers.

15.38 Work Problem 15.37 for the data of Problem 15.29.

15.39 If $r_{12} = r_{13} = r_{23} = r \neq 1$, show that $r_{12.3} = r_{13.2} = r_{23.1} = r/(1+r)$. Discuss the case $r = 1$.

15.40 If $r_{12.3} = 1$, show that (a) $|r_{13.2}| = 1$, (b) $|r_{23.1}| = 1$, (c) $R_{1.23} = 1$, and (d) $s_{1.23} = 0$.

MULTIPLE AND PARTIAL CORRELATION INVOLVING FOUR OR MORE VARIABLES

15.41 Show that the regression equation of X_4 on X_1, X_2, and X_3 can be written

$$\frac{x_4}{s_4} = a_1\left(\frac{x_1}{s_1}\right) + a_2\left(\frac{x_2}{s_2}\right) + a_3\left(\frac{x_3}{s_3}\right)$$

where a_1, a_2, and a_3 are determined by solving simultaneously the equations

$$\begin{aligned} a_1r_{11} + a_2r_{12} + a_3r_{13} &= r_{14} \\ a_1r_{21} + a_2r_{22} + a_3r_{23} &= r_{24} \\ a_1r_{31} + a_2r_{32} + a_3r_{33} &= r_{34} \end{aligned}$$

and where $x_j = X_j - \bar{X}_j$, $r_{jj} = 1$, and $j = 1, 2, 3$, and 4. Generalize to the case of more than four variables.

15.42 Given $\bar{X}_1 = 20$, $\bar{X}_2 = 36$, $\bar{X}_3 = 12$, $\bar{X}_4 = 80$, $s_1 = 1.0$, $s_2 = 2.0$, $s_3 = 1.5$, $s_4 = 6.0$, $r_{12} = -0.20$, $r_{13} = 0.40$, $r_{23} =$ [illegible], $r_{14} = 0.40$, $r_{24} = 0.30$, and $r_{34} = -0.10$, (a) find the regression equation of X_4 on X_1, X_2, and X_3, and (b) estimate X_4 when $X_1 = 15$, $X_2 = 40$, and $X_3 = 14$.

Find (*a*) $r_{41.23}$, (*b*) $r_{42.13}$, and (*c*) $r_{43.12}$ for the data of Problem 15.42 and interpret your results.

For the data of Problem 15.42, find (*a*) $R_{4.123}$ and (*b*) $s_{4.123}$.

A scientist collected data concerning four variables: T, U, V, and W. She believed that an equation of the form $W = aT^bU^cV^d$, where a, b, c, and d are unknown constants, could be found from which she could determine W by knowing T, U, and V. Outline clearly a procedure by means of which this aim may be accomplished. [*Hint:* Take logarithms of both sides of the equation.]