1. Introduction

Welcome to the series of e-learning modules on Partial Correlation and Regression in Three Variables. Here we shall discuss about the partial correlation, plane of regression line and residual, its properties and coefficient of partial correlation for 3 variables and 'n' variables.

By the end of this session, you will be able to explain:

- Partial correlation
- Plane of the regression line
- Properties of residuals
- Coefficient of partial correlation

Suppose in a trivariate or multi-variate distribution we are interested in the relationship between two variables only. There are two alternatives namely,

i. We consider only those two members of the observed data in which the other members have specified values or

ii. We may eliminate mathematically the effect of other variates on two variates.

The first method has the disadvantage that it limits the size of the data and also it will be applicable to only the data in which the other variates have assigned values. In the second method it may not be possible to eliminate the entire influence of the variates but the linear effect can be easily eliminated. The correlation and regression between only two variates eliminating the linear effect of other variates in them is called the partial correlation and partial regression.

Let us consider a distribution involving three random variables 'X1' and 'X2' and X3.
Then the equation of the plane of regression of 'X1' on 'X2' and 'X3' is
'X1' is equal to a plus 'b 1-2 point 3' into 'X2' plus 'b 1-3 point 2' into 'X3'.
Without loss of generality we can assume that the variables 'X1', 'X2' and 'X3' have been measured from their respective mean so that
Expectation of 'X1' is equal to Expectation of 'X2' is equal to Expectation of 'X3' is equal to zero
Hence on taking expectation of both sides in above equation we get a is equal to zero.

Thus the plane of regression of 'X1' on 'X2' and 'X3' becomes
'X1' is equal to 'b 1-2 point 3' into 'X2' plus 'b 1-3 point 2' into 'X3'. Name it as (1).
The coefficients 'b 1-2 point 3' and 'b 1-3 point 2' are known as the partial regression coefficients of 'X1' on 'X2' and of 'X1' on 'X3' respectively.
e 1 point 2, 3 is equal to 'b 1-2 point 3' into 'X2' plus 'b 1-3 point 2' into 'X3' is called the estimate of 'X1' as given by the plane of regression (1) and the quantity, 'X1' point 2, 3 is equal to 'X1' minus b 1 2 point 3 into 'X2' minus 'b 1-3 point 2' into 'X3' is equal to 'X1' minus e 1 point 2, 3 is called the error of estimate or residual.

2. Plane of Regression

The equation of the plane of regression of ‘X1’ on X2 and X3 is
X1 is equal to ‘b 1-2 point 3’ into ‘X2’ plus ‘b 1-3 point 2’ into ‘X3’
The constants b in the above equation are determined by the principle of least square, ie., by minimizing the sum of the squares of residuals, namely, S is equal to summation ‘X1’ point 2, 3 square is equal to summation ‘X1’ minus ‘b 1-2 point 3’ into ‘X2’ minus ‘b 1-3 point 2’ into ‘X3’ whole square, the summation being extended to the given values of the variables.
Here we make N observations on each of the variables ‘X1’, X2 and X3.

The normal equations for estimating ‘b 1-2 point 3’ and ‘b 1-3 point 2’ are,
D s by d b1, 2 point 3 is equal to zero is equal to minus 2 into summation X2 into ‘X1’ minus ‘b 1-2 point 3’ into ‘X2’ minus ‘b 1-3 point 2’ into ‘X3’.
D s by d b1, 3 point 2 is equal to zero is equal to minus 2 into summation X3 into ‘X1’ minus ‘b 1-2 point 3’ into ‘X2’ minus ‘b 1-3 point 2’ into ‘X3’.

That is summation ‘X2’ into ‘X1’ point 2, 3 is equal to zero and summation X3 into ‘X1’ point 2, 3 is equal to zero.
Implies summation ‘X1’ into ‘X2’ minus ‘b 1-2 point 3’ into summation ‘X2’ square minus ‘b 1-3 point 2’ into summation ‘X2’ into X3 is equal to zero and
summation ‘X1’ into X3 minus ‘b 1-2 point 3’ into summation ‘X2’ into X3 minus ‘b 1-3 point 2’ into summation X3 square is equal to zero

since Xi’s are measured from their respective means, we have,
sigma I square is equal to 1 by N into summation Xi square, covariance of Xi, Xj is equal to 1 by N into summation Xi into Xj
and r I j is equal to covariance of Xi, Xj divided by sigma I sigma j is equal to summation Xi into Xj divided by N into sigma I into sigma j
hence we get,
r1, 2 into sigma 1 into sigma 2 minus ‘b 1-2 point 3’ into sigma 2 square minus ‘b 1-3 point 2’ into r 2, 3 into sigma 2 into sigma 3 is equal to zero
r1, 3 into sigma 1 into sigma 3 minus ‘b 1-2 point 3’ into r 2, 3 into sigma 2 into sigma 3 minus ‘b 1-3 point 2’ into sigma 3 square is equal to zero.

Solving the above equations for ‘b 1-2 point 3’ and b1, 3 point 2, we get,
B1, 2 point 3 is equal to determinant of r 1,2 into sigma 1 r2, 3 into sigma 3, r1,2 into sigma 1 sigma 3, divided by
Determinant of sigma 2, r2, 3 into sigma 3, r 2,3 into sigma 2, sigma 3 is equal to sigma 1 divided by sigma 2 into determinant of r1,2, r2, 3, r1, 3 1 divided by determinant of 1, r2, 3 r2, 3, 1
Similarly we get,
B1, 3 point 2 is equal to sigma 1 divided by sigma 3 into determinant of 1, r 1, 2, r2, 3, r1, 3 divided by 1, r2, 3, r2, 3 1.

If we write omega is equal to 1, r1, 2, r1, 3, r2, 1. 1 r2, 3, r3, 1, r3,2, 1 and omega I j is the cofactor of the element in the i^{th} row and j^{th} column of omega, we have,
b1, 2 point 3 is equal to sigma 1 into omega 1, 2 divided by sigma 2 into omega 1, 1 and b1, 3 is equal to sigma 1 into omega 1, 3 divided by sigma 3 into omega 1, 1.

Substituting the values in the plane of regression, mentioned in the beginning, we get the required equation of the plane of regression of X1 on X2 and X3 as, X1 is equal to minus sigma 1 divided by sigma 2 into omega 1, 2 divided by omega 1, 1 into 'X2' minus sigma 1 divided by sigma 3 into omega 1, 3 divided by omega 1, 1 into X3
Implies,
X1 by sigma 1 into omega 1, 1 plus X2 by sigma 2 into omega 1, 2 plus X3 divided by sigma 3 into omega 1, 3 is equal to zero.

Properties of residuals.

1. The sum of the product of any residual of order zero with any other residual of higher order is zero, provided the subscript of the former occurs among the secondary subscripts of the later.

2. The sum of the product of any two residuals in which all the secondary subscripts of the first occur among the secondary subscripts of the second is unaltered if we omit any or all of the secondary subscripts of the first. Conversely, the product sum of any residual of order 'p' with a residual of order p plus q, the 'p' subscripts being the same in each case is unaltered by adding to the secondary subscripts of the former any or all the q additional subscripts of the later.

3. The sum of the product of two residuals is zero if all the subscripts (primary as well as secondary) of the one occur among the secondary subscripts of the other.

3. Coefficient of Partial Correlation (Part 1)

Now let us discuss about the Coefficient of Partial Correlation.
Sometimes the correlation between two variables 'X1' and 'X2' may be partly due to the correlation of a third variable, 'X3' with both 'X1' and 'X2'. In such a situation, one may want to know what the correlation between X_1 and X_2 would be if the effect of X_3 on each of X_1 and X_2 were eliminated. This correlation is called the partial correlation and the correlation coefficient between X_1 and X_2 the linear effect of X_3 on each of them has been eliminated is called the partial correlation coefficient.

The residual 'X1' point 3 is equal to X_1 minus b13 into X_3, may be regarded as that part of the variable X_1 which remains after the linear effect of X_3 has been eliminated. Similarly, the residual 'X2' point 3 may be interpreted as the part of the variable X_2 obtained after eliminating the linear effect of X_3. Thus the partial correlation coefficient between X_1 and X_2, usually denoted by
R 1, 2 point 3 is equal to covariance of 'X1' point 3 'X2' point 3 divided by square root of variance of 'X1' point 3 into variance of 'X2' point 3.

We can obtain the above expression as follows.
We have, covariance of 'X1' point 3 'X2' point 3 is equal to 1 by N into summation 'X1' point 3 into 'X2' point 3 is equal to 1 by N into summation 'X1' into 'X2' point 3 is equal to 1 by N into summation 'X1' into 'X2' minus b 2, 3 into 'X3'
Is equal to 1 by N into summation 'X1' into 'X2' minus b 2, 3 into 1 by N into summation 'X1' 'X3'.
Is equal to r 1,2 into sigma 1 into sigma 2 minus r 2, 3 into sigma 2 by sigma 3 into r 1, 3 into sigma 1 into sigma 3
Is equal to sigma 1 into sigma 2 into r 1, 2 minus r 1, 3 into r 2, 3.

Variance of 'X1' point 3 is equal to 1 by N into summation 'X1' point 3 square
Is equal to 1 by N into summation 'X1' point 3 into X point 3
Is equal to summation 'X1' into 'X1' point 3
Is equal to summation 'X1' into 'X1' minus b 1, 3 into 'X3'
Is equal to 1 by n into summation 'X1' square minus b 1, 3 into 1 by N into summation 'X1' into 'X3'.
Is equal to sigma 1 square minus r 1, 3 into sigma 1 divided by sigma 3 into r 1, 3 into sigma 1 into sigma 3.
Is equal to sigma 1 square into 1 minus r 1, 3 square.
Similarly we get, variance of 'X2' point 3 is equal to sigma 2 square into 1 minus r 2, 3 square.

Substituting in r 1, 2 point 3, for covariance and variance we get, R 1, 2 point 3 is equal to sigma 1 into sigma 2 into r 1, 2 minus r 1, 3 into r 2,3 divided by square root of sigma 1 square into 1 minus r 1, 3 square into sigma 2 square into 1 minus r 2, 3 square is equal to r 1, 2 minus 4 1,3 into r 2, 3 divided by square root of 1 minus r 1, 3 square into 1 minus r 2, 3 square.

4. Coefficient of Partial Correlation (Part 2)

We can also obtain the expression for r 1, 2 point 3 using regression coefficients.
We have zero is equal to summation ‘X2’ point 3 into ‘X1’ point 2, 3 is equal to summation 2 point 3 into ‘X1’ minus ‘b 1-2 point 3’ into ‘X2’ minus ‘b 1-3 point 2’ into X3.
Is equal to summation ‘X1’ into ‘X2’ point 3 minus ‘b 1-2 point 3’ into summation ‘X2’ point 3 into ‘X2’ minus ‘b 1-3 point 2’ into summation ‘X2’ point 3 into ‘X3’
Is equal to summation ‘X1’ point 3 into ‘X2’ point 3 minus ‘b 1-2 point 3’ into summation ‘X2’ point 3 into ‘X2’ point 3
Implies ‘b 1-2 point 3’ is equal to summation ‘X1’ point 3 into ‘X2’ point 3 divided by summation ‘X2’ point 3 square.
Form this it follows that ‘b 1-2 point 3’ is coefficient of regression of ‘X1’ point 3 on ‘X2’ point 3.

Similarly, b 2, 1 point 3 is the coefficient of regression of ‘X2’ point 3 on ‘X1’ point 3.
Since correlation coefficient is the geometric mean between regression coefficients, we have,
r 1, 2 point 3 square is equal to ‘b 1-2 point 3’ into b 2, 1 point 3.
But by definition,
b1, 2 point 3 is equal to minus sigma 1 divided by sigma 2 into omega 1, 2 divided by omega 1, 1 and b2, 1 point 3 is equal to minus sigma 2 divided by sigma 1 into omega , 1 divided by omega 2, 2.
Therefore, r 1, 2 point 3 square is equal to minus sigma 1 divided by sigma 2 into omega 1, 2 divided by omega 1, 1 into minus sigma 2 divided by sigma 1 into omega 2, 1 divided by omega 2, 2
Is equal to omega 1, 2 square divided by omega 1, 1 into omega 2, 2.

Implies r 1, 2 point 3 is equal to minus omega 1, 2 divided by square root of omega 1, 1 into omega 2, 2

The negative sign being taken since the sign of regression coefficients is the same as that of minus omega 1, 2.
Substituting the values of omega 1, 2, omega 1, 1 and omega 2, 2 in the above expression, we get,
r1, 2, point 3 is equal to r 1, 2 minus r 1, 3 into r 2, 3 divided by square root of 1 minus r 1, 3 square into 1 minus r 2, 3 square.

5. Remarks and Generalization on Partial Correlation Coefficient

Now consider some remarks on partial correlation coefficient.

1. The expressions for r 1, 3 point 2 and r 2, 3 point 1 can be similarly obtained are as follows. r 1, 3 point 2 is equal to r 1, 3 minus 4 1, 2 into r 2, 3 divided by square root of 1 minus r 1, 2 square into 1 minus r 3, 2 square and r2, 3 point 1 is equal to r 2, 3 minus r 2, 1 into r 3, 1 divided by square root of 1 minus r 2, 1 square into 1 minus r 3, 1 square.

2. If r1, 2 point 3 is equal to zero, we have then r 1, 2 is equal to r1, 3 into r 2, 3. It means that r 1, 2 will not be zero if X3 is correlated with both 'X1' and X2. Thus, although X1 and X2 may be uncorrelated when effect of X3 is eliminated, yet X1 and X2 may appear to be correlated because they carry the effect of X3 on them.

3. Partial correlation coefficient helps in deciding whether to include or not an additional independent variable in regression analysis.

Now let us generalize the partial correlation coefficient for n variables.
In the case of n variables, X1, X2, etc., X_n the partial correlation coefficient r1, 2 point 3, 4 etc n between X1 and X2 (after the linear effect of X3, X4, etc, Xn on them has been eliminated), is given by,
r1, 2 point 3, 4 etc n square is equal to b1, 2 point 3, 4 etc., n into b 2, 1 point 3, 4 etc., n but we have b1, 2 point 3, 4 etc. n is equal to minus sigma 1 divided by sigma 2 into omega 1, 2 divided by omega 1, 1 and
b2, 1 point 3, 4 etc. b is equal to minus sigma 2 divided by sigma 1 into omega 2, 1 divided by sigma 2, 2.
Therefore r1, 2 point 3, 4 etc n square is equal to minus sigma 1 divided by sigma 2 into omega 1, 2 divided by omega 1, 1 into b2, 1 point 3, 4 etc. b is equal to minus sigma 2 divided by sigma 1 into omega 2, 1 divided by sigma 2, 2

Implies r1, 2 point t 3, 4 etc n is equal to minus omega 1, 2 divided by square root of omega 1, 1 into omega 2, 2.

The negative sign being taken since the sign of the regression coefficient is the same as that of minus omega 1, 2.

Here's a summary of our learning in this session:

- Partial correlation.
- Plane of the regression line
- Properties of residuals
- Coefficient of Partial Correlation