

Variance-Covariance Matrix

Assembled by Ben Danforth

June 1, 2009

1 Covariance

Covariance measures the degree to which two variables change or vary together (i.e. co-vary). On the one hand, the covariance of two variables is positive if they vary together in the same direction relative to their expected values (i.e. if one variable moves above its expected value, then the other variable also moves above its expected value). On the other hand, if one variable tends to be above its expected value when the other is below its expected value, then the covariance between the two variables is negative. If there is no linear dependency between the two variables, then the covariance is 0.

The covariance of two random variables, X_i and X_j , can be mathematically represented as

$$\text{cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] \quad (1)$$

where $\mu_i = E(X_i)$ and $\mu_j = E(X_j)$.¹ This relationship can be further generalized to a multivariate situation and for estimated coefficients (i.e. $\hat{\beta}$ s) using matrix notation

$$\text{cov}(\hat{\beta}) = [(\hat{\beta} - \mu)(\hat{\beta} - \mu)^T] = E(\hat{\beta}\hat{\beta}^T) - \mu\mu^T \quad (2)$$

where $\hat{\beta}$ is a vector of estimated coefficients and $\mu = E(\hat{\beta}) = \beta$. This relationship can be more explicitly represented as

$$\text{cov}(\hat{\beta}) = \begin{bmatrix} \text{cov}(\hat{\beta}_1, \hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \text{cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{cov}(\hat{\beta}_2, \hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \dots & \dots & \dots & \dots \\ \text{cov}(\hat{\beta}_k, \hat{\beta}_1) & \text{cov}(\hat{\beta}_k, \hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_k, \hat{\beta}_k) \end{bmatrix}, \quad (3)$$

where the diagonal of this matrix represents the *variance* of the vector $\hat{\beta}$ with k elements. Therefore, the matrix can be rewritten as

$$\text{cov}(\hat{\beta}) = \begin{bmatrix} \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \text{cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{var}(\hat{\beta}_2) & \dots & \text{cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \dots & \dots & \dots & \dots \\ \text{cov}(\hat{\beta}_k, \hat{\beta}_1) & \text{cov}(\hat{\beta}_k, \hat{\beta}_2) & \dots & \text{var}(\hat{\beta}_k) \end{bmatrix}. \quad (4)$$

This essentially represents the *covariance* or *variance-covariance matrix*.

¹Sometimes $\text{cov}(X_i, X_j)$ is denoted as $\Sigma_{i,j}$.

2 Deriving the Variance-Covariance Matrix

First, begin with the OLS estimator in matrix notation

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (5)$$

Next, substitute $y = X\beta + u$ into the preceding equation to arrive at

$$\hat{\beta} = (X^T X)^{-1} X^T (X\beta + u)$$

Then distribute the first set of terms

$$\hat{\beta} = (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T u$$

And simplify, recalling that $(X^T X)^{-1}(X^T X) = 1$

$$\hat{\beta} = \beta + (X^T X)^{-1} X^T u$$

Finally, subtract β from both sides to get

$$\hat{\beta} - \beta = (X^T X)^{-1} X^T u \quad (6)$$

Recalling Eq. 2, with β substituted for μ

$$\text{cov}(\hat{\beta}) = E [(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \quad (7)$$

Insert Eq. 6 into Eq. 7 and then reorder the terms (remember $(AB)^T = B^T A^T$)

$$\begin{aligned} \text{cov}(\hat{\beta}) &= E [((X^T X)^{-1} X^T u)((X^T X)^{-1} X^T u)^T] \\ \text{cov}(\hat{\beta}) &= E [(X^T X)^{-1} X^T u u^T X (X^T X)^{-1}] \end{aligned}$$

Take the expectation (the X s are non-stochastic, and it is assumed that $E(uu^T) = \sigma^2 I$)

$$\begin{aligned} \text{cov}(\hat{\beta}) &= (X^T X)^{-1} X^T E(uu^T) X (X^T X)^{-1} \\ \text{cov}(\hat{\beta}) &= (X^T X)^{-1} X^T \sigma^2 I X (X^T X)^{-1} \end{aligned}$$

And simplify, producing the final equation for the variance-covariance matrix

$$\text{cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \quad (8)$$

In this last equation, σ^2 is the homoskedastic variance of u_i and $(X^T X)^{-1}$ is the inverse matrix appearing in the OLS estimator.

$$\boxed{\text{cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}}$$

3 Estimating the Variance-Covariance Matrix

To estimate the variance-covariance matrix, the variance is needed. Since the true variance is unknown, it must be estimated. An unbiased estimator of σ^2 for the k -th variable case is

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n - k} \quad (9)$$

which is equivalent to

$$\hat{\sigma}^2 = \frac{\hat{u}^T \hat{u}}{n - k} \quad (10)$$

where \hat{u} is the estimated residuals, n is the number of observations, and k the number of parameters being estimated.

Although $\hat{\sigma}^2$ can be computed directly from the estimated residuals, there is a second approach to calculating it. First, recall that

$$TSS = ESS + RSS \quad (11)$$

where TSS is the total sum of squares, ESS is the explained sum of squares, and RSS is the residual sum of squares. This equation can be rewritten as

$$RSS = TSS - ESS \quad (12)$$

Next, remember that

$$RSS = \hat{u}^T \hat{u} \quad (13)$$

$$TSS = y^T y - n\bar{Y}^2 \quad (14)$$

$$ESS = \hat{\beta}^T X^T y - n\hat{Y}^2 \quad (15)$$

where the term $n\bar{Y}^2$ is known as the correction for mean. Therefore, substituting Eqs. 13-15 into Eq. 12 produces

$$\hat{u}^T \hat{u} = y^T y - \hat{\beta}^T X^T y \quad (16)$$

4 Covariance vs. Correlation

Covariance and correlation are related but not equivalent statistical measures. In particular, the correlation of two variables, X_i and X_j , is their normalized covariance, which is defined as

$$\rho_{i,j} = \frac{E[(X_i - \mu_i)(X_j - \mu_j)]}{\sigma_i \sigma_j}, \quad (17)$$

where $\rho_{i,j}$ is the correlation coefficient, $\mu_i = E(X_i)$, $\mu_j = E(X_j)$, σ_i is the standard deviation of X_i , and σ_j is the standard deviation of X_j . Because the correlation is normalized, it is dimensionless (i.e. it is a pure number without a unit of measure). By contrast, covariance does have a unit of measure—the product of the units of two variables.

5 Exercise

You are given the following pieces of information:

$$\begin{aligned}n &= 15 \\ \bar{Y} &= 68.7333 \\ \hat{\beta} &= |5.5027 \quad 4.1535 \quad 3.3787 \quad 8.2903| \\ y^T y &= 259651 \\ X^T y &= \begin{vmatrix} 1031 \\ 11370 \\ 35981 \\ 10176 \end{vmatrix}\end{aligned}$$

1. Compute the variance-covariance matrix. You can also compute R^2 .
2. Use the variance-covariance matrix to make inferences about the estimated coefficients.
3. Is there any evidence of violations of the OLS assumptions given the above information and your calculations?
4. The covariance matrix of the three variables of interest (X_1, X_2, X_3) is

$$\begin{vmatrix} & X_1 & X_2 & X_3 \\ X_1 & 53.5523 & 167.2238 & 4.1667 \\ X_2 & 167.2238 & 522.9238 & 12.5238 \\ X_3 & 4.16667 & 12.5238 & 12.3810 \end{vmatrix}$$

Given that the standard deviations between X_1 and X_2 are 7.3179 and 22.8675, respectively, what is the correlation between these two variables?