A Brief Discussion of Covariance

Charles B. Moss

January 11, 2018

- In AEB 3550 you were introduced to the normal distribution. In general there are two characteristics of the normal which completely define the distribution – the mean and the variance.
  - The mean is simply defined as
$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{1}$$
  - The variance is defined as
$$S_x = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x}) = \frac{1}{N} \left(\sum_{i=1}^{N} x_i^2 \right) - \bar{x}^2. \tag{2}$$

    I want to focus on the first of these terms which basically says that the variance is the average squared departure from the mean.
- These basic relationships are characteristics of the "univariate" normal – the statistical distribution of a single random variable.
- In this class we need to define the statistical relationship between two or more random variables. In the case of the normal distribution, the covariance between two random variables defines the relationship.
- Building on Equation 2 we define the covariance between random variable x and random variable y as
$$S_{xy} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \tag{3}$$

Table 1: Payments on Loans

Period	Loan 1	Loan 2	Loan 3	Loan 4
1	0.15	0.16	0.13	0.13
2	0.15	0.16	0.13	0.13
3	0.15	0.16	0.13	0.13
4	0.00	0.16	0.00	0.13
5	0.00	0.00	0.00	0.00
6	0.15	0.16	0.13	0.13
7	0.15	0.16	0.13	0.13
8	0.15	0.16	0.13	0.13
9	0.00	0.16	0.13	0.13
⋮	⋮	⋮	⋮	⋮
22	0.15	0.16	0.13	0.13
Average	0.1023	0.1309	0.0945	0.1123

- Consider the outcomes in Table 1. These values represent the return on four loans. The loan has an interest rate of 15 percent, the second, 16 percent, and the third and fourth 13 percent. Each of these borrowers pays the interest depending on the outcome of their production decisions. However, whether or not the borrower pays is a function of a common random event (e.g., growth in the economy).
- Table 2 presents the results of $(x_{it} - \bar{x}_i)$ for each observation where i is the loan (i.e., $i = 1, 2, 3, 4$), t is the time period, and $\bar{x}_i$ is the average return for loan i.
- Table 3 then presents the first set of cross-products

$$\begin{array}{c} \epsilon_{1i,t} = (x_{1t} - \bar{x}_1)(x_{it} - \bar{x}_t) \\ S_{1i} = \frac{1}{N}\sum_{t=1}^{N} \epsilon_{1i,t} \end{array} \quad (4)$$

- The resulting variance matrix is presented in Table 4.
- The values in Table 4 are useful, but a little difficult to interpret. One alternative is to normalize these values by the variance of each individual variable. One such normalization is referred to as the correlation cofficient

Table 2: Error – Observed Value minus Average

Period	Loan 1	Loan 2	Loan 3	Loan 4
1	0.0477	0.0291	0.0355	0.0177
2	0.0477	0.0291	0.0355	0.0177
3	0.0477	0.0291	0.0355	0.0177
4	-0.1023	0.0291	-0.0945	0.0177
5	-0.1023	-0.1309	-0.0945	-0.1123
6	0.0477	0.0291	0.0355	0.0177
7	0.0477	0.0291	0.0355	0.0177
8	0.0477	0.0291	0.0355	0.0177
9	-0.1023	0.0291	0.0355	0.0177
⋮	⋮	⋮	⋮	⋮
22	0.0477	0.0291	0.0355	0.0177

Table 3: Cross Products of Errors

Period	Loan 1	Loan 2	Loan 3	Loan 4
1	0.0023	0.0014	0.0017	0.0008
2	0.0023	0.0014	0.0017	0.0008
3	0.0023	0.0014	0.0017	0.0008
4	0.0105	-0.003	0.0097	-0.0018
5	0.0105	0.0134	0.0097	0.0115
6	0.0023	0.0014	0.0017	0.0008
7	0.0023	0.0014	0.0017	0.0008
8	0.0023	0.0014	0.0017	0.0008
9	0.0105	-0.003	-0.0036	-0.0018
⋮	⋮	⋮	⋮	⋮
22	0.0023	0.0014	0.0017	0.0008
Average	0.0049	0.003	0.0036	0.0018

Table 4: Covariance Matrix

	Loan			
Loan	1	2	3	4
1	0.0049	0.0030	0.0036	0.0018
2	0.0030	0.0038	0.0028	0.0023
3	0.0036	0.0028	0.0034	0.0017
4	0.0018	0.0023	0.0017	0.0020

Table 5: Covariance Matrix

	Loan			
Loan	1	2	3	4
1.0000	0.6901	0.8964	0.5817	
0.6901	1.0000	0.7698	0.8429	
0.8964	0.7698	1.0000	0.6489	
0.5817	0.8429	0.6489	1.0000	

$$\rho_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}}\sqrt{S_{jj}}}. \tag{5}$$

The correlation coefficients for the loan portfolio are presented in Table 5. In general the correlation coefficient is bounded between -1.0 and 1.0. If the value of the correlation is greater than zero, the deviations of the random variable tend to be positive at the same time (i.e., if the return on Loan 1 is above its mean the the return on Loan 2 tends to be above its mean). Notice that the correlation matrix (as well as the covariance matrix) is symmetric – $S_{12} = S_{21}$.

- The data for the forgoing example are given in "LoanPort01-11-18.xlsx" – I would suggest verifying the covariance and correlation matrix because some form of these computations will be on the first assignment.