STA 437: Applied Multivariate Statistics

Al Nosedal.
University of Toronto.

Winter 2015

1. Chapter 3. Characterizing and Displaying Multivariate Data
   - Mean and Variance of a Univariate Random Variable
   - Covariance and Correlation
   - Mean Vectors
   - Covariance Matrices
   - Correlation Matrices
   - Linear combinations of random variables
   - Linear Transformations of random variables

"If you can't explain it simply, you don't understand it well enough"

Albert Einstein.

Mean

The population mean of a random variable Y is defined as the mean of all possible values of Y and is denoted by μ. The mean is also referred to as the *expected value* of Y or $E(Y)$.

Variance

The variance of the population is defined as $Var(Y) = \sigma^2 = E(Y - \mu)^2$. This is the average squared deviation from the mean and is thus an indication of the extent to which the values of Y are spread or scattered. It can be shown that $\sigma^2 = E(Y^2) - \mu^2$.
The square root of the population variance is called the standard deviation.

Recall that if a single random variable, such as Y, is multiplied by a constant c, then

$$E(cY) = cE(Y) = c\mu$$

and

$$V(cY) = E[cY - c\mu]^2 = E[c^2(Y - \mu)^2] = c^2 E[(Y - \mu)^2] = c^2 V(Y)$$

Covariance

The population covariance is defined as

$$Cov(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)],$$

where μ_X and μ_Y are the means of X and Y, respectively.
It can be shown that $\sigma_{XY} = E(XY) - \mu_X \mu_Y$.

Independence between X and Y implies $Cov(X, Y) = 0$, but $Cov(X, Y) = 0$ does not imply independence. It is easy to show that if X and Y are independent, then $Cov(X, Y) = 0$:
$Cov(X, Y) = E(XY) - \mu_X \mu_Y$
$= E(X)E(Y) - \mu_X \mu_Y$ (because X and Y are independent)
$= \mu_X \mu_Y - \mu_X \mu_Y = 0$

Correlation

The population correlation of two random variables X and Y is

$$\rho_{XY} = Corr(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sqrt{E[(X - \mu_X)]^2 E[(Y - \mu_Y)]^2}}.$$

A random vector is a vector whose elements are random variables. The expected value of a random vector is the vector consisting of the expected values of each of its elements.
Let $\mathbf{y}$ represent a random vector of p variables measured on a sampling unit (subject or object). The mean of $\mathbf{y}$ over all possible values in the population is called the population mean vector or expected value of $\mathbf{y}$. It is defined as a vector of expected values of each variable,

$$E(\mathbf{y}) = E\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$$

Example

Consider the random vector $\mathbf{X}' = [X_1, X_2]$. Let the discrete random variable X_1 have the following probability function:

x_1	-1	0	1
$p_1(x_1)$	0.3	0.3	0.4

Similarly, let the discrete random variable X_2 have the probability function:

x_2	0	1
$p_2(x_2)$	0.8	0.2

Find $E(\mathbf{X})$.

If $\mathbf{y}$ is a random vector taking on any possible value in a multivariate population, the population covariance matrix is defined as

$$\mathbf{\Sigma} = cov(\mathbf{y}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$$

The diagonal elements $\sigma_{jj} = \sigma_j^2$ are the population variances of the y's, and the off-diagonal elements σ_{jk} are the population covariances of all possible pairs of y's.
The population covariance matrix can also be found as
$\mathbf{\Sigma} = E[(\mathbf{y} - E(\mathbf{y}))(\mathbf{y} - E(\mathbf{y}))']$

The population correlation matrix is defined as

$$\mathbf{P}_{\rho} = (\rho_{jk}) = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{pmatrix}$$

where $\rho_{jk} = \frac{\sigma_{jk}}{\sigma_j \sigma_k}$.

Let's now derive the variance of a linear combination of random variables. We will begin with a two-dimensional random vector and generalize to the case of an arbitrary dimension p.
Assume that $E(X_1) = \mu_1$, $E(X_2) = \mu_2$, $V(X_1) = \sigma_1^2$, and $V(X_2) = \sigma_2^2$.

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}_{2\times 1}$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{2\times 1}$$

$$\mathbf{b}'\mathbf{x} = b_1 x_1 + b_2 x_2.$$

Let us find the expected value of $\mathbf{b}^{'}\mathbf{x} = b_1x_1 + b_2x_2$.
$E(\mathbf{b}^{'}\mathbf{x}) = E(b_1x_1 + b_2x_2) = b_1E(x_1) + b_2E(x_2)$
$E(\mathbf{b}^{'}\mathbf{x}) = b_1\mu_1 + b_2\mu_2$

Now, let us find the variance of $\mathbf{b}'\mathbf{x} = b_1x_1 + b_2x_2$.

$$V(\mathbf{b}'\mathbf{x}) = E[(b_1x_1 + b_2x_2) - (b_1\mu_1 + b_2\mu_2)]^2$$
$$= E[b_1(x_1 - \mu_1) + b_2(x_2 - \mu_2)]^2$$
$$= E[b_1^2(x_1 - \mu_1)^2 + 2b_1b_2(x_1 - \mu_1)(x_2 - \mu_2) + b_2^2(x_2 - \mu_2)^2]$$
$$= b_1^2E[(x_1 - \mu_1)^2] + 2b_1b_2E[(x_1 - \mu_1)(x_2 - \mu_2)] + b_2^2E[(x_2 - \mu_2)^2]$$

$$V(\mathbf{b}'\mathbf{x}) = b_1^2\sigma_1^2 + 2b_1b_2Cov(x_1, x_2) + b_2^2\sigma_2^2$$

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

where $Cov(x_1, x_2) = \sigma_{12}$.

Clearly, $V(\mathbf{b}'\mathbf{x}) = \mathbf{b}'\mathbf{\Sigma}\mathbf{b}$.

As an example, let us assume that we want to construct an index of creativity on the basis of four creativity tests given to a large sample of school children. The means of the four tests are $10, 5, 15$, and 20, and the researcher weights their relative importance as $0.5, 1, 0.1$, and 0.2, respectively. Suppose that our four creativity tests had the following covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 25 & 10 & 30 & 50 \\ 10 & 50 & 40 & 30 \\ 30 & 40 & 100 & 60 \\ 50 & 30 & 60 & 125 \end{pmatrix}$$

Find the mean of this linear combination and its variance.

Mean.

$$E[\mathbf{b}'\mathbf{x}] = \begin{pmatrix} 0.5 & 1 & 0.1 & 0.2 \end{pmatrix} \begin{pmatrix} 10 \\ 5 \\ 15 \\ 20 \end{pmatrix} = 15.5$$

$$V[\mathbf{b}^{'}\mathbf{x}] = \begin{pmatrix} 0.5 & 1 & 0.1 & 0.2 \end{pmatrix} \begin{pmatrix} 25 & 10 & 30 & 50 \\ 10 & 50 & 40 & 30 \\ 30 & 40 & 100 & 60 \\ 50 & 30 & 60 & 125 \end{pmatrix} \begin{pmatrix} 10 \\ 5 \\ 15 \\ 20 \end{pmatrix}$$

$$V[\mathbf{b}^{'}\mathbf{x}] = 107.65$$

If we had two linear combinations, each comprising two variables, then the covariance between $b_1x_1 + b_2x_2$ and $b_3x_3 + b_4x_4$ can be defined as

$$E[b_1x_1 + b_2x_2 - (b_1\mu_1 + b_2\mu_2)][b_3x_3 + b_4x_4 - (b_3\mu_3 + b_4\mu_4)]$$

where $b_1\mu_1 + b_2\mu_2$ is the mean of $b_1x_1 + b_2x_2$ and $b_3\mu_3 + b_4\mu_4$ is the mean of $b_3x_3 + b_4x_4$.

Furthermore,

$$E[b_1x_1 + b_2x_2 - (b_1\mu_1 + b_2\mu_2)][b_3x_3 + b_4x_4 - (b_3\mu_3 + b_4\mu_4)]$$
$$= E[b_1(x_1 - \mu_1) + b_2(x_2 - \mu_2)][b_3(x_3 - \mu_3) + b_4(x_4 - \mu_4)]$$
$$= E[b_1b_3(x_1 - \mu_1)(x_3 - \mu_3) + b_1b_4(x_1 - \mu_1)(x_4 - \mu_4) + b_2b_3(x_2 - \mu_2)(x_3 - \mu_3)] + b_2b_4(x_2 - \mu_2)(x_4 - \mu_4)]$$
$$= b_1b_3E[(x_1 - \mu_1)(x_3 - \mu_3)] + b_1b_4E[(x_1 - \mu_1)(x_4 - \mu_4)] + b_2b_3E[(x_2 - \mu_2)(x_3 - \mu_3)] + b_2b_4E[(x_2 - \mu_2)(x_4 - \mu_4)]$$
$$= b_1b_3\sigma_{13} + b_1b_4\sigma_{14} + b_2b_3\sigma_{23} + b_2b_4\sigma_{24}$$

$$\boldsymbol{\Sigma} = \left(\begin{array}{cc|cc} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \hline \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{array}\right)$$

$$\mathbf{\Sigma}_{12} = \begin{pmatrix} \sigma_{13} & \sigma_{14} \\ \sigma_{23} & \sigma_{24} \end{pmatrix}$$

$$\mathbf{b}_1 = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

$$\mathbf{b}_2 = \begin{pmatrix} b_3 \\ b_4 \end{pmatrix}$$

Clearly, covariance between $b_1x_1 + b_2x_2$ and $b_3x_3 + b_4x_4$ can be expressed as follows

$$\mathbf{b}_1^{'}\boldsymbol{\Sigma}_{12}\mathbf{b}_2.$$

Example. Suppose that a personnel psychologist had a large sample of data available on two tests or measures
X_1 = intelligence (IQ).
X_2 = interaction orientation (IO).
Furthermore, suppose that he also had a large sample of data available on a group of salespersons for two performance ratings (that is, criteria):
X_3 = amount of sales (S).
X_4 = potential for supervisory position (P).

Example (cont.)
The personnel psychologist believes that amount of sales (X_3) is twice as important as potential for supervisory position (X_4). Consequently, he believes that interaction orientation (X_2) is twice as important as intelligence (X_1) in predicting job success. As a result, he wants to calculate the correlation between the two linear combinations $X_1 + 2X_2$ and $2X_3 + X_4$.

$$\mathbf{\Sigma} = \begin{pmatrix} 10 & 4 & 1 & 2 \\ 4 & 10 & 3 & 1 \\ 1 & 3 & 5 & 1 \\ 2 & 1 & 1 & 5 \end{pmatrix}$$

Solution

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 10 & 4 \\ 4 & 10 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_{12} = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}$$

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$\mathbf{b}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Solution

$V(\mathbf{b}_1'\mathbf{x}) = \mathbf{b}_1'\mathbf{\Sigma}_1\mathbf{b}_1$.

$$\begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 10 & 4 \\ 4 & 10 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 66$$

where $\mathbf{\Sigma}_1$ represents the covariance matrix for variables 1 and 2.

Solution

$V(\mathbf{b}_2'\mathbf{x}) = \mathbf{b}_2'\mathbf{\Sigma}_2\mathbf{b}_2.$

$$\begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 29$$

where $\mathbf{\Sigma}_2$ represents the covariance matrix for variables 3 and 4.

Solution

We know that
$Cov(\mathbf{b}_1'\mathbf{x}, \mathbf{b}_2'\mathbf{x}) = \mathbf{b}_1'\mathbf{\Sigma}_{12}\mathbf{b}_2$.

$$\begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 18$$

Solution

Finally,

$$\rho_{X_1+2X_2,\ 2X_3+X_4} = \frac{18}{\sqrt{66}\sqrt{29}} = 0.41.$$

The magnitude of this correlation indicates that the hypothesized relationship is not very large.

A linear combination of a set of p random variables can be considered as a linear transformation from a p-dimensional space to a one-dimensional space. The linear function $y = b_1x_1 + b_2x_2 + ... + b_px_p$ takes the vector $(x_1, x_2, ..., x_p)$ and transforms it into a single number, y. There is no reason we have to be limited to mapping a vector into a single number. We could map this vector into another k-dimensional vector $(y_1, y_2, ..., y_k)$.

The linear transformation from a p-dimensional vector to a k-dimensional vector would be expressed as

$$y_1 = b_{11}x_1 + b_{12}x_2 + \ldots + b_{1p}x_p$$
$$y_2 = b_{12}x_1 + b_{22}x_2 + \ldots + b_{2p}x_p$$
$$\vdots$$
$$y_k = b_{k1}x_1 + b_{k2}x_2 + \ldots + b_{kp}x_p.$$

We can express the linear transformation of a p-dimensional vector into a k-dimensional vector as $\mathbf{y} = \mathbf{Bx}$, where $\mathbf{y}' = (y_1, y_2, \ldots, y_k)$

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1p} \\ b_{21} & b_{22} & \ldots & b_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ b_{k1} & b_{k2} & \ldots & b_{kp} \end{pmatrix}$$

and $\mathbf{x}' = (x_1, x_2, \ldots, x_p)$.

For the case of linearly transforming a three-dimensional vector into a two-dimensional vector, we would have
$\mathbf{y}' = (y_1, y_2)$

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix}$$

and $\mathbf{x}' = (x_1, x_2, x_3)$.

We can apply this procedure to our personnel psychology problem discussed earlier and generate the variance-covariance matrix of the two combos described there directly in one matrix expression as follows: Let $y_1 = x_1 + 2x_2 + 0x_3 + 0x_4$ and $y_2 = 0x_1 + 0x_2 + 2x_3 + x_4$; then the covariance matrix of $\mathbf{y}^{'} = (y_1, y_2)$ is

$$\mathbf{B} = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 10 & 4 & 1 & 2 \\ 4 & 10 & 3 & 1 \\ 1 & 3 & 5 & 1 \\ 2 & 1 & 1 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 66 & 18 \\ 18 & 29 \end{pmatrix}$$

Next, we might want to convert this covariance matrix into a correlation matrix. If we define a diagonal matrix with the inverse of the standard deviations of the combos y_1 and y_2, then the correlation matrix is

$$\mathbf{B} = \begin{pmatrix} \frac{1}{\sqrt{66}} & 0 \\ 0 & \frac{1}{\sqrt{29}} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{66}} & 0 \\ 0 & \frac{1}{\sqrt{29}} \end{pmatrix} = \begin{pmatrix} 1 & 0.41 \\ 0.41 & 1 \end{pmatrix}$$

Let $z = \mathbf{a}'\mathbf{y}$, where $\mathbf{a}$ is a vector of constants. Then the population mean of z is

$$E(z) = E(\mathbf{a}'\mathbf{y}) = \mathbf{a}'E(\mathbf{y})$$

and the population variance is

$$\sigma_z^2 = V(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\mathbf{\Sigma}\mathbf{a}$$

Let $w = \mathbf{b}^{'}\mathbf{y}$, where $\mathbf{b}$ is a vector of constants different from $\mathbf{a}$. The population covariance of $z = \mathbf{a}^{'}\mathbf{y}$ and $w = \mathbf{b}^{'}\mathbf{y}$ is

$$cov(z, w) = \sigma_{zw} = \mathbf{a}^{'}\mathbf{\Sigma}\mathbf{b}.$$

The population correlation of z and w is

$$\rho_{zw} = corr(z, w) = corr(\mathbf{a}^{'}\mathbf{y}, \mathbf{b}^{'}\mathbf{y}) = \frac{\sigma_{zw}}{\sigma_z \sigma_w} = \frac{\mathbf{a}^{'}\mathbf{\Sigma}\mathbf{b}}{\sqrt{\mathbf{a}^{'}\mathbf{\Sigma}\mathbf{a}\mathbf{b}^{'}\mathbf{\Sigma}\mathbf{b}}}$$

If $\mathbf{Ay}$ represents several linear combinations, the population mean vector and covariance matrix are given by

$$E(\mathbf{Ay}) = \mathbf{A}E(\mathbf{y}),$$

$$cov(\mathbf{Ay}) = \mathbf{A\Sigma A}'.$$