# Sample mean and covariance based on Random Sampling

Woojoo Lee

What is the meaning of ”random sampling” in multivariate data analysis ?

This implies that
(1) measurements taken on different items (individuals) are unrelated to one another and

(2) the joint distribution of all $p$ variables remains the same for all items.

Again, in our class, $X$ is a $n \times p$ matrix:

$$\begin{array}{cc}
 & \begin{array}{cccc} \text{Variable1} & \text{Variable2} & \cdots & \text{Variable p} \end{array} \\
\begin{array}{c} \text{Item1} \\ \text{Item2} \\ \vdots \\ \text{Item n} \end{array} & \begin{pmatrix} x_{11} & \cdots & & \\ \vdots & & & \\ & & & \\ & & & \end{pmatrix}
\end{array}$$

Q) Let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be a random sample from a joint distribution with mean $\mu$ and covariance $\Sigma$.
Show that for $\bar{\mathbf{x}} = 1/n \sum_i \mathbf{x}_i$,

$$\begin{aligned} E(\bar{\mathbf{x}}) &= \mu \\ Cov(\bar{\mathbf{x}}) &= \Sigma/n \end{aligned}$$

$\rightarrow$ $\bar{\mathbf{x}}$ can be regarded as a good estimator for $\mu$.

Q) Compute

$$E(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = \frac{n-1}{n}\Sigma.$$

Derive an unbiased estimator $S$ for $\Sigma$.

Q) Find matrix expressions for $\bar{\mathbf{x}}$ and $S$ by using the data matrix $X$.

$$X = \begin{array}{cc} & \begin{array}{cccc} \text{Variable1} & \text{Variable2} & \cdots & \text{Variable p} \end{array} \\ \begin{array}{c} \text{Item1} \\ \text{Item2} \\ \vdots \\ \text{Item n} \end{array} & \begin{pmatrix} x_{11} & \cdots & & x_{1p} \\ \vdots & & & \\ & & & \\ x_{n1} & & & x_{np} \end{pmatrix} \end{array} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix}$$

$$\begin{aligned} \bar{\mathbf{x}} &= \frac{1}{n} X^T \mathbf{1} \\ S &= \frac{1}{n-1} X^T \left(I - \frac{1}{n} \mathbf{1}\mathbf{1}^T\right) X \end{aligned}$$

Consider the following linear combination:

$$\mathbf{c}^T\mathbf{x}_i = c_1x_{i1} = \cdots + c_px_{ip}$$

Q) Find the sample mean and variance of $\mathbf{c}^T\mathbf{x}_i$ $(i = 1, \cdots, n)$.

Q) Find the sample covariance of $\mathbf{b}^T\mathbf{x}_i$ and $\mathbf{c}^T\mathbf{x}_i$ $(i = 1, \cdots, n)$.