Univariate and Multivariate Summary Statistics

A Review of Univariate Summary Statistics

Random Sample

Let $x_1, x_2, \cdots, x_N$ be a random sample from some distribution.

Random Sample expressed as a vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

Sample Mean

$$\hat{\mu} = \bar{x} = \frac{1}{N}\sum_{r=1}^{N} x_r = \frac{\mathbf{x}'\mathbf{1}}{N} = \frac{\mathbf{x}'\mathbf{1}}{\mathbf{1}'\mathbf{1}} = \mathbf{x}'\mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}$$

Sample Variance

$$s^2 = \frac{1}{N-1}\left[\sum_{r=1}^{N}\left(x_r - \hat{\mu}\right)^2\right]$$

Sample Variance in matrix notation

The deviation of each observation from the sample mean is

$$\mathbf{x} - \hat{\mu}\mathbf{1} = \mathbf{x} - \mathbf{1}\overline{x} = \mathbf{x} - \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'\mathbf{x}$$

$$= \mathbf{I}_N\mathbf{x} - \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'\mathbf{x}$$

$$= \left(\mathbf{I}_N - \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'\right)\mathbf{x}$$

Sample Variance in matrix notation

Alternatively, the deviation of each observation from the sample mean can be expressed as

$$\mathbf{x} - \hat{\mu}\mathbf{1} = \mathbf{x} - \mathbf{1}\overline{\mathrm{x}} = \left(\mathbf{I}_N - \frac{1}{N}\mathbf{J}_N\right)\mathbf{x}$$

Sample Variance in matrix notation

The sum of the squared deviations
of each observation from the sample mean is

$$\left(\left(\mathbf{I}_N - \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'\right)\mathbf{x}\right)'\left(\left(\mathbf{I}_N - \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'\right)\mathbf{x}\right)$$

$$= \mathbf{x}'\left(\mathbf{I}_N - \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'\right)^2\mathbf{x}$$

$$= \mathbf{x}'\left(\mathbf{I}_N - \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'\right)\mathbf{x}$$

Sample Variance in matrix notation

$$s^2 = \frac{1}{N-1}\mathbf{x}'\left(\mathbf{I}_N - \mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'\right)\mathbf{x}$$

$$= \frac{1}{N-1}\mathbf{x}'\left(\mathbf{I}_N - \frac{1}{N}\mathbf{J}_N\right)\mathbf{x}$$

Multivariate Observations

Let

- p = number of response variables
- N = number of experimental units
- x_{rj} = value of the j^{th} response variable on the r^{th} experimental unit

Data Matrix

$$\mathbf{X}_{N\times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix} = \left(x_{rj} \right)$$

$$= \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_N \end{bmatrix}$$

A Multivariate Observation

$$\mathbf{x}'_r = \begin{pmatrix} x_{r1} & x_{r2} & \cdots & x_{rp} \end{pmatrix}$$

Multivariate Summary Statistics

Random Sample

Let $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N$ be a random sample from a multivariate distribution.

Sample Mean Vector

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \frac{1}{N}\sum_{r=1}^{N}\mathbf{x}_r = \frac{1}{N}\mathbf{X}'\mathbf{1}$$

$$= \mathbf{X}'\mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}$$

Sample Variance-Covariance Matrix

$$\hat{\mathbf{\Sigma}} = \mathbf{S} = \frac{1}{N-1}\left[\sum_{r=1}^{N}(\mathbf{x}_r - \hat{\boldsymbol{\mu}})(\mathbf{x}_r - \hat{\boldsymbol{\mu}})'\right]$$

$$= \frac{1}{N-1}\mathbf{X'}\left(\mathbf{I}_N - \mathbf{1}(\mathbf{1'1})^{-1}\mathbf{1}\right)\mathbf{X}$$

$$= \frac{1}{N-1}\mathbf{X'}\left(\mathbf{I}_N - \frac{1}{N}\mathbf{J}_N\right)\mathbf{X}$$

Sample Correlation Matrix

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{12} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix}$$

The Sample Correlation Matrix

Let $\mathbf{D}_\mathbf{S} = diag\left(\sqrt{s_{11}}, \sqrt{s_{22}}, \cdots, \sqrt{s_{pp}}\right)$

Then the Sample Correlation Matrix $\mathbf{R}$ is given by

$$\mathbf{R} = \left(r_{ij}\right) = \mathbf{D}_\mathbf{S}^{-1}\mathbf{S}\mathbf{D}_\mathbf{S}^{-1}$$