Organisation av observationer

Observationerna i ett stickprov arrangeras i en $n\times p$ matris (datamatris) X, där n är antalet experimentella enheter (storleken på stickprovet) och p är antalet variabler:

$$X=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{pmatrix}$$

Rad j i denna $\boldsymbol{x}_j' = (x_{j1}, x_{j2}, \ldots, x_{jk}, \ldots, x_{jp}$ är en p dimensionell observation.

Vektor av stickprovsmedelvärden

Givet ett stickprovsmedelvärde för en enskild variabel j

$$\bar{x}_j = \frac{1}{n}\sum_{k=1}^{n} x_{jk}$$

definieras

$$\bar{\boldsymbol{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix}$$

$$\bar{\boldsymbol{x}} = \frac{1}{n}\boldsymbol{X}'\boldsymbol{1}$$

Stickprovskovariansmatris

Givet stickprovskovarianser

$$s_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$$

mellan par av variabler i och j, definieras

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix}.$$

Stickprovskorrelationsmatris

Givet stickprovskorrelationer

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}$$

mellan par av variabler i och j, definieras

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}.$$

***** 3.3 Stickprov och väntevärden för medelvärdesvektor och kovariansmatris

n oberoende observationer av $\boldsymbol{X}$: $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n$.

$$\bar{\boldsymbol{X}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i$$

$$E(\bar{\boldsymbol{X}}) = \frac{1}{n} \sum_{i=1}^{n} E(\boldsymbol{X}_i) =$$

3.4 Generaliserad varians

Generaliserad varians

Determinanten av S, $|S|$ är ett vanligt "multivariat" spridningsmått.

För $p = 2$:

$$|S = \begin{vmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{vmatrix}$$

$$= s_{11}s_{12} - s_{12}^2 = s_1^2 s_2^2 - (s_1 s_2 r_{12})^2 = s_1^2 s_2^2 (1 - r_{12}^2)$$

Andra spridningsmått

"Total varians" Ibland används den "totala variansen"

$$\sum_1^p s_{ii} = trace(S)$$

Spåret av S.

Ett annat mått är determinant av korrelationsmatrisen $|R|$. Det gäller att

$$|S| = (s_{11}s_{22}\cdots s_{pp})|R|$$

3.5 Medelvärdesvektor och kovariansmatris som matrisoperationer

Datamatris $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$.

$$\bar{\boldsymbol{x}} = \frac{1}{n}\boldsymbol{X}\mathbf{1}$$

$$(n-1)S = (\boldsymbol{X} - \frac{1}{n}\boldsymbol{X}\mathbf{1}\mathbf{1}')(\boldsymbol{X} - \frac{1}{n}\boldsymbol{X}\mathbf{1}\mathbf{1}')'$$

$$= \boldsymbol{X}(I - \frac{1}{n}\mathbf{1}\mathbf{1}')\boldsymbol{X}'$$

3.6 Linjärkombinationer av variabler i stickprovet

En linjärkombination:

$$Y = \sum_{j=1}^{p} c_j X_j = \boldsymbol{c}'\boldsymbol{X}$$

$$s_Y^2 = \boldsymbol{c}' S \boldsymbol{c}$$

Annan linjärkombination:

$$Z = \sum_{j=1}^{p} b_j X_j = \boldsymbol{b}'\boldsymbol{X}$$

$$s_{YZ} = \boldsymbol{c}' S \boldsymbol{b}$$

Generalisering:

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p$$

$(i = 1, 2, \ldots, q)$

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$$

Populationskovariansmatris Σ

Givet populationsvarianser och kovarianser

$$\sigma_{ij} = E((X_i - \mu_i)(X_j - \mu_j))$$

mellan par av variabler definieras

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}.$$

Skattning av $\boldsymbol{\mu}$

Medelvärdesvektorn $\bar{\boldsymbol{X}}$ är en *väntevärdesriktig skattning* av $\boldsymbol{\mu}$. (JW Thm 3.1)

Skattning av Σ

Det gäller att

$$E(S) = \Sigma$$