L02: Basic statistics

1. Sample mean vector and sample covariance matrix

   (1) Data matrix

   Population $\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$. Data matrix $\mathbf{X} = (x_{ij})_{n\times p}$.

   $\mathbf{X}' = (\mathbf{x}_1, ..., \mathbf{x}_n)$ contains n observations from the population. Here $\mathbf{x}_r = \begin{pmatrix} x_{r1} \\ \vdots \\ x_{rp} \end{pmatrix}$, a row in $\mathbf{X}$.

   $\mathbf{X} = (\mathbf{x}_{(1)}, .., \mathbf{x}_{(p)})$ contains p samples from the p population. Here $\mathbf{x}_{(i)} = \begin{pmatrix} x_{1i} \\ \vdots \\ x_{ni} \end{pmatrix}$, a column of $\mathbf{X}$.

   (2) Sample mean vector

   $\overline{\mathbf{x}} = \frac{\mathbf{X}'\mathbf{1}_n}{n} = \frac{\sum_{r=1}^n \mathbf{x}_r}{n} = \begin{pmatrix} \overline{x}_1 \\ \vdots \\ \overline{x}_p \end{pmatrix}$ where $\overline{x}_i = \frac{\mathbf{x}'_{(i)}\mathbf{1}'_n}{n}$ is the mean of the ith sample $\mathbf{x}_{(i)}$.

   $\overline{\mathbf{x}}$ is the center of n observations since $\sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}}) = 0$.

   (3) Scatter matrix

   CSSCP $= \mathbf{X}'\mathbf{H}\mathbf{X} = \mathbf{X}'\left(\mathbf{I} - \frac{\mathbf{1}_n\mathbf{1}'_n}{n}\right)\mathbf{X}$ is also called scatter matrix that can be expressed via observations.

   $$\text{CSSCP} = \sum_{r=1}^n (\mathbf{x}_r - \overline{\mathbf{x}})(\mathbf{x}_r - \overline{\mathbf{x}})' = \sum_{r=1}^n \mathbf{x}_r\mathbf{x}'_r - n\overline{\mathbf{x}}\,\overline{\mathbf{x}}'.$$

   It can also be expressed via samples. For example

   $$(\text{CSSCP})_{ii} = \mathbf{x}'_{(i)}\mathbf{H}\mathbf{x}_{(i)} = \sum_{r=1}^n (x_{ri} - \overline{x}_i)^2 = \sum_{r=1}^n x_{ki}^2 - n\overline{x}_i^2$$

   is from the sample $\mathbf{x}_{(i)}$ that measures the magnitude of the value fluctuation in that sample and

   $$(\text{CSSCP})_{ij} = \mathbf{x}'_{(i)}\mathbf{H}\mathbf{x}_{(j)} = \sum_{k=1}^n (x_{ki} - \overline{x}_i)(x_{kj} - \overline{x}_j) = \sum_{k=1}^n x_{ki}x_{kj} - n\overline{x}_i\overline{x}_j$$

   is from samples $\mathbf{x}_{(i)}$ and $\mathbf{x}_{(j)}$ that measures the correlation of the values in the two samples.

   (4) Two sample covariance matrices

   $\mathbf{S} = \frac{\text{CSSCP}}{n} = (s_{ij})_{p\times p}$ with $s_{ii} > 0$ denoted as s_i^2, $i = 1, ..., p$.

   $\mathbf{S}_u = \frac{\text{SCCSP}}{n=1}$.

2. Sample correlation matrix

   (1) Definitions

   From sample variance-covariance matrix $\mathbf{S} = (s_{ij})_{p\times p}$,
   let $\mathbf{D}^2 = \text{diag}(\mathbf{S}) = \text{diag}(s_1^2, ..., s_p^2)$ be the sample variance matrix.
   Then $\mathbf{D} = \text{diag}(s_1, ..., s_p)$ is the sample standard deviation matrix
   and $\mathbf{D}^{-1} = \text{diag}(1/s_1, ..., 1/s_p)$.
   Define $\mathbf{R} = \mathbf{D}^{-1}\mathbf{S}\mathbf{D}^{-1} \in R^{p\times p}$ and call it sample correlation matrix.
   So $\mathbf{R} = (r_{ij})_{p\times p}$ where $r_{ij} = \frac{s_{ij}}{s_i\, s_j}$.

   (2) Correlation relation

   Two samples $\mathbf{x}_{(i)}$ and $\mathbf{x}_{(j)}$ are positively correlated $\overset{def}{\Longleftrightarrow}$ $r_{ij} > 0$
   $\Longleftrightarrow$ $s_{ij} > 0 \Longleftrightarrow \text{CSSCP}_{ij} > 0$

   Two samples $\mathbf{x}_{(i)}$ and $\mathbf{x}_{(j)}$ are negatively correlated $\overset{def}{\Longleftrightarrow}$ $r_{ij} < 0$
   $\Longleftrightarrow$ $s_{ij} < 0 \Longleftrightarrow \text{CSSCP}_{ij} < 0$

   Two samples $\mathbf{x}_{(i)}$ and $\mathbf{x}_{(j)}$ are uncorrelated $\overset{def}{\Longleftrightarrow}$ $r_{ij} = 0$
   $\Longleftrightarrow$ $s_{ij} = 0 \Longleftrightarrow \text{CSSCP}_{ij} = 0$

Comparing with s_{ij} and CSSCP_{ij}, r_{ij} is better scaled since $-1 \le r_{ij} \le 1$ and

$$r_{ij} = -1 \Longleftrightarrow \mathbf{x}_{(i)} = a\mathbf{x}_{(j)} + \mathbf{b} \text{ with } a > 0;$$
$$r_{ij} = 1 \Longleftrightarrow \mathbf{x}_{(i)} = a\mathbf{x}_{(j)} + \mathbf{b} \text{ with } a < 0.$$

Ex: Show $r_{ii} = 1$. Method I: $\mathbf{x}_{(i)} = \mathbf{I_n}\mathbf{x}_{(i)} + \mathbf{0}$. So $r_{ii} = 1$. Method II: $r_{ii} = \frac{s_{ii}}{s_i\, s_i} = \frac{s_i^2}{s_i\, s_i} = 1$.

(3) Equivalent expressions

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}\, s_{jj}}} = \frac{\text{CSSCP}_{ij}}{\sqrt{\text{CSSCP}_{ii}\, \text{CSSCP}_{jj}}} = \frac{(s_u)_{ij}}{\sqrt{(s_u)_{ii}\,(s_u)_{jj}}}.$$

Proof
$$\begin{aligned} r_{ij} &= \frac{s_{ij}}{\sqrt{s_{ii}\, s_{jj}}} = \frac{n s_{ij}}{\sqrt{n s_{ii}\, n s_{jj}}} = \frac{\text{CSSCP}_{ij}}{\sqrt{\text{CSSCP}_{ii}\, \text{CSSCP}_{jj}}} \\ &= \frac{\text{CSSCP}_{ij/(n-1)}}{\sqrt{\text{CSSCP}_{ii/(n-1)}\, \text{CSSCP}_{jj/(n-1)}}} = \frac{(s_u)_{ij}}{\sqrt{(s_u)_{ii}\,(s_u)_{jj}}} \end{aligned}.$$

Matrix forms:
For $\text{CSSCP} = (\text{CSSCP}_{ij})_{p\times p}$, let $\mathbf{D}^2_{\text{CSSCP}} = \text{diag}(\text{CSSCP})$.
For $\mathbf{S}_u$ let $\mathbf{D}^2_{\mathbf{S}_u} = \text{diag}(\mathbf{S}_u)$. Then

$$\mathbf{R} = \mathbf{D}^{-1}\mathbf{S}\mathbf{D}^{-1} = \mathbf{D}^{-1}_{\mathbf{S}_u}\mathbf{S}_u\mathbf{D}^{-1}_{\mathbf{S}_u} = \mathbf{D}^{-1}_{\text{CSSCP}}(\mathbf{CSSCP})\mathbf{D}^{-1}_{\text{CSSCP}}.$$

3. Comments on two exercises

(1) 1.4.1: A transformation on population

From population $\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ data matrix $\mathbf{X} \in R^{n\times p}$ is obtained. It in turn produced sample meran vector $\overline{\mathbf{x}} = \frac{\mathbf{X}'\mathbf{1}_n}{n}$ and sample covariance matrix $\mathbf{S} = \mathbf{X}'\frac{\mathbf{H}}{n}\mathbf{X}$.

If transformation on the population $\begin{pmatrix} y_1 \\ \vdots \\ y_q \end{pmatrix} = \mathbf{A}\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} + \mathbf{b}$ is performed, then data matrix is transformed to $\mathbf{Y}$ where $\mathbf{Y}' = \mathbf{A}\mathbf{X}' + \mathbf{b}\mathbf{1}_n'$. By 1.4.1 the new sample mean vector is $\mathbf{A}\overline{\mathbf{x}} + \mathbf{b}$ and the new sample covariance matrix is $\mathbf{A}\mathbf{S}\mathbf{A}'$.

(2) 1.4.2: Minimized $\mathbf{S}(\alpha)$
Let $\mathbf{S}(\alpha) = \sum_{r=1}^n (\mathbf{x}_r - \alpha)(\mathbf{x}_r - \alpha)'$. In 1.4.2 we see that $\mathbf{S}(\alpha) = \mathbf{S} + (\overline{\mathbf{x}} - \alpha)(\overline{\mathbf{x}} - \alpha)'$.
We claim that $\mathbf{S}$ is minimized $\mathbf{S}(\alpha)$ in the following sense.
(i) $|\mathbf{S}(\alpha)| \ge |\mathbf{S}|$. So $|\mathbf{S}| = \min_\alpha |\mathbf{S}(\alpha)|$.
(ii) $\text{tr}[\mathbf{S}(\alpha)] \ge \text{tr}(\mathbf{S})$. So $\text{tr}(\mathbf{S}) = \min_\alpha \text{tr}(\mathbf{S}(\alpha))$.
(iii) $\mathbf{S}(\alpha) \ge \mathbf{S}$ defined as $\mathbf{S}(\alpha) - \mathbf{S}$ is a non-negative definite matrix. So $\mathbf{S} = \min_\alpha(\mathbf{S}(\alpha))$.

(i) For $\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, let $A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ and $A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$.
Then $|A| = |A_{11}| \cdot |A_{22.1}| = |A_{11.2}|\,|A_{22}|$ (covered in Stat701).
So with $A = \begin{pmatrix} 1 & -(\overline{\mathbf{x}} - \alpha)' \\ \overline{\mathbf{x}} - \alpha & \mathbf{S} \end{pmatrix}$, $|A| = 1 \cdot |\mathbf{S} + (\overline{\mathbf{x}} - \alpha)(\overline{\mathbf{x}} - \alpha)'| = [1 + (\overline{\mathbf{x}} - \alpha)'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \alpha)]\,|\mathbf{S}|$.
Thus $|\mathbf{S}(\alpha)| = [1 + (\overline{\mathbf{x}} - \alpha)'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \alpha)]\,|\mathbf{S}| \ge |\mathbf{S}|$.
(ii) See HW01
(iii) $\mathbf{S}(\alpha) - \mathbf{S} = (\overline{\mathbf{x}} - \alpha)(\overline{\mathbf{x}} - \alpha)'$ is a non-negative definite matrix.

L03: Distribution of multivariate population

0. Computation for statistics

(1) Entering data matrix $\mathbf{X} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$.

```
data a;
  input x1 x2;
  datalines;
  1 2
  3 4
  5 6
  ;
```

```
data a;
  infile "D:\myStat776\MyData.txt";
  put x1 x2;
```

(2) Requesting statistics

```
proc corr;
  var x1 x2;
  run;
```

```
proc corr SSCP CSSCP COV;
  var x1 x2;
  run;
```

1. Population distributions

(1) Probability density function (pdf)

Distribution of continuous population $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ is given by its pdf $f(\mathbf{x}) = f(x_1, .., x_p) \geq 0$ such that $P(\mathbf{X} \in A) = \iint_A f(x_1, .., x_p)\, dx_1, .., dx_p$.
A function $f(\mathbf{x})$ can be used as a pdf if (i) $f(\mathbf{x}) \geq 0$ and (ii) $\iint_{R^p} f(x_1, ..., x_p) dx_1, ..dx_p = 1$.

(2) Cumulative distribution function (cdf)

Distribution of $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ can be given by its cdf

$$F(\mathbf{x}) = F(x_1, .., x_p) = P(\mathbf{x}_1 \leq x_1, ..., \mathbf{x}_p \leq x_p) = P(\mathbf{X} \leq \mathbf{x}).$$

(3) Relations
For continuous $\mathbf{X} \in R^p$ with pdf $f(\mathbf{x})$, the cdf
$F(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}) = \int_{-\infty}^{x_1} dx_1 \cdots \int_{-\infty}^{x_p} f(x_1, .., x_p) dx_p$. If cdf $F(\mathbf{x})$ is given, then the pdf is $f(x_1, .., x_p) = \frac{\partial^p F(x_1, .., x_p)}{\partial x_1, .., x_p}$.

Ex1: $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ has pdf $f(x_1, x_2) = 1$ on $0 \leq x_1 \leq 1$ and $0 \leq x_2 \leq 1$. Then its cdf

$$F(x_1, x_2) = P(\mathbf{X}_1 \leq x_1, \mathbf{X}_2 \leq x_2) = \begin{cases} 0, & x_1 < 0 \text{ or } x_2 < 0 \\ \min(1, x_1) \cdot \min(1, x_2), & x_1 \geq 0 \text{ and } x_2 \geq 0 \end{cases}.$$

2. Marginal distributions

(1) Marginal pdf

$\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}$ where $\mathbf{X} \in R^p$ and $\mathbf{Y} \in R^q$ has joint pdf $f(\mathbf{x}, \mathbf{y}) = f(x_1, .., x_p, y_1, .., y_q)$. Then the marginal pdf of $\mathbf{X}$ is $f_X(\mathbf{x}) = \iint_{R^q} f(\mathbf{x}, \mathbf{y}) dy_1, .., dy_q$.

Proof Let $f_X(\mathbf{x}) = \iint_{R^q} f(\mathbf{x}, \mathbf{y}) dy_1, .., dy_q$. Then $f_X(\mathbf{x}) \geq 0$ since $f(\mathbf{x}, \mathbf{y}) \geq 0$.

For $A \subset R^p$,

$$\begin{aligned} P(\mathbf{X} \in A) &= P(\mathbf{X} \in A, \mathbf{Y} \in R^q) = \iint_{\mathbf{x} \in A, \mathbf{y} \in R^q} f(\mathbf{x}, \mathbf{y}) dx_1, ..dx_p, dy_1, ..., dy_q \\ &= \iint_A dx_1, .., dx_p \iint_{R^q} f(\mathbf{x}, \mathbf{y}) dy_1, .., dy_q = \iint_A f_X(\mathbf{x}) dx_1, ..dx_p. \end{aligned}$$

So $f_X(\mathbf{x})$ is the pdf for $\mathbf{X}$.

(2) Ex2

From joint pdf $f(x_1, x_2) = 2$ on $0 \leq x_1 \leq 1$ and $0 \leq x_2 \leq 1 - x_1$,

$$\begin{aligned} f_1(x_1) &= \int_R f(x_1, x_2) dx_2 = \begin{cases} 0, & x_1 < 0 \text{ or } x_1 > 1 \\ \int_0^{1-x_2} 2dx_2, & 0 \leq x_1 \leq 1 \end{cases} \\ &= \begin{cases} 0, & x_1 < 0 \text{ or } x_1 > 1 \\ 2(1 - x_1), & 0 \leq x_1 \leq 1 \end{cases} \end{aligned}$$

Similarly, $f_2(x_2) = \begin{cases} 0, & x_x < 0 \text{ or } x_2 > 1 \\ 2(1 - x_2), & 0 \leq x_2 \leq 1 \end{cases}$.

3. Conditional distribution and independence

(1) Conditional pdf

$f(\mathbf{x}, \mathbf{y})$ is joint pdf for $\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}$ where $\mathbf{X} \in R^p$ and $\mathbf{Y} \in R^q$. $f_Y(\mathbf{y})$ is the marginal pdf for $\mathbf{Y}$. As a function of $\mathbf{x}$, $f(\mathbf{x}|\mathbf{y}) = \frac{f(\mathbf{x}, \mathbf{y})}{f_y(\mathbf{y})} \geq 0$ and

$$\iint_{R^p} f(\mathbf{x}|\mathbf{y}) \, dx_1, .., dx_p = \iint_{R^p} \frac{f(\mathbf{x}, \mathbf{y})}{f_y(\mathbf{y})} dx_1, .., dx_p = \frac{\iint_{R^p} f(\mathbf{x}, \mathbf{y}) dx_1, .., dx_p}{f_y(\mathbf{y})} = 1.$$

So $f(\mathbf{x}|\mathbf{y})$ is a class of pdf function of $\mathbf{x}$ with index $\mathbf{y}$. This pdf is the conditional pdf of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$.

(2) Independence

$$\begin{aligned} &\mathbf{X} \text{ and } \mathbf{Y} \text{ are independent} \overset{def}{\Longleftrightarrow} f(\mathbf{x}|\mathbf{y}) = f_X(\mathbf{x}) \Longleftrightarrow \tfrac{f(\mathbf{x}, \mathbf{y})}{f_y(\mathbf{y})} = f_X(\mathbf{x}) \\ \Longleftrightarrow \; &f(\mathbf{x}, \mathbf{y}) = f_X(\mathbf{x}) \, f_Y(\mathbf{y}) \Longleftrightarrow \tfrac{f(\mathbf{x},\mathbf{y})}{f_X(\mathbf{x})} = f_y(\mathbf{y}) \Longleftrightarrow f(\mathbf{y}|\mathbf{x}) = f_Y(\mathbf{y}) \end{aligned}$$

Ex3: In Ex2: $f(x_1, x_2) = 2$ on $0 \leq x_1 \leq 1$ and $0 \leq x_2 \leq 1 - x_1$, $f_1(x_1) = 2(1 - x_1)$ on $0 \leq x_1 \leq 1$. So $f(x_2|x_1) = \frac{1}{1-x_1}$ on $0 \leq x_2 \leq 1 - x_1$, i.e., $X_2|x_1 \sim U(0, 1 - x_1)$.

Ex4: In Ex1: $f(x_1, x_2) = 1$ on $0 \leq x_1 \leq 1$. We can derive that $f_1(x_1) = 1$ on $0 \leq x_1 \leq 1$ and $f_2(x_2) = 1$ on $0 \leq x_2 \leq 1$. So X_1 and X_2 are independent since $f(x_1, x_2) = f_1(x_1) \, f_2(x_2)$.