Lecture 2

Estimated covariance matrix

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})^T$$

($\underset{\sim}{x}_i$: p vectors)

If $p \geq n$ then S is singular (S^{-1} does not exist)
Why? Need to show S is not positive definite
$\underset{\sim}{a}^T S \underset{\sim}{a} = 0$ for some $\underset{\sim}{a} \neq \underset{\sim}{0}$

$$\begin{pmatrix} \underset{\sim}{x}_1^T \\ \underset{\sim}{x}_2^T \\ \vdots \\ \underset{\sim}{x}_n^T \end{pmatrix}\begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

(n rows, p cols)

$p \geq n \Rightarrow$ there exists $\underset{\sim}{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}$ satisfying the equation

$$\underset{\sim}{x}_i^T \underset{\sim}{a} = 1 \text{ for all } i = 1, \cdots, n \Rightarrow \underset{\sim}{a}^T S \underset{\sim}{a} = 0$$

Simple sol'n (how to fix)

Add a positive definite matrix to S.

$$\tilde{S} = S + \varepsilon I$$

(ε: small)

$$\underset{\sim}{a}^T \tilde{S} \underset{\sim}{a} = \underbrace{\underset{\sim}{a}^T S \underset{\sim}{a}}_{\geq 0} + \underbrace{\varepsilon \underset{\sim}{a}^T \underset{\sim}{a}}_{> 0} > 0 \text{ for } \underset{\sim}{a} \neq \underset{\sim}{0}$$

Correlation Matrix

$$R = \begin{pmatrix} 1 & \frac{\sigma_{12}}{\sigma_1\sigma_2} & \cdots & \frac{\sigma_{1p}}{\sigma_1\sigma_p} \\ \frac{\sigma_{21}}{\sigma_2\sigma_1} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sigma_p\sigma_1} & \cdots & \cdots & 1 \end{pmatrix}$$

$\sigma_{ij} = \text{cov}(x_i, x_j)$

$$\frac{\sigma_{ij}}{\sigma_i\sigma_j} = \text{Corr}(x_i, x_j)$$

$$= D^{-1/2} C D^{-1/2}$$

(D: diagonal matrix with diagonal elements of C)

- contains same information as C but R is dimensionless (no units)
  - independent of the scaling of $X_1, \cdots, X_p$
- R is also the covariance of standardized variables

$$x_j' = \frac{x_j - E(x_j)}{\sqrt{\text{Var}(x_j)}}$$

- very important in multivariate analysis
  - guarantees similar range for all (transformed) variables.
- Taking logs of positive variables often achieves the same goal.
- Estimate of R:

$$\hat{R} = \hat{D}^{-1/2} S \hat{D}^{-1/2}$$

($\hat{D}$: diagonal element of S)

Distance between points (observations)

Idea: Define clusters of observations
- If 2 multivariate observations are "close" then they should belong to the same cluster

How to define distance?

$$\underset{\sim}{x}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix} \qquad \underset{\sim}{x}_j = \begin{pmatrix} x_{j1} \\ \vdots \\ x_{jp} \end{pmatrix}$$

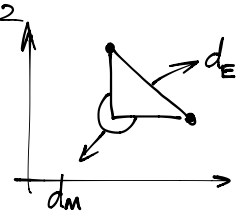
Naive: ① Euclidean distance

$$d_E(\underset{\sim}{x}_i, \underset{\sim}{x}_j) = \left\{ \sum_{k=1}^{p} (x_{ik} - x_{jk})^2 \right\}^{\frac{1}{2}}$$

② Manhatten (L_1) distance:

$$d_M(\underset{\sim}{x}_i, \underset{\sim}{x}_j) = \sum_{k=1}^{p} |x_{ik} - x_{jk}|$$

e.g. $p=2$



What's wrong here?
- adding apples to oranges! (with diff units!)
- to fix this, scale all the variables to be dimensionless.

For example, define $s_k^2 = \underbrace{\frac{1}{n-1} \sum_{i=1}^{n} (x_{ik} - \bar{x}_k)^2}_{k\text{-th diagonal of } S}$

Take $d_E(\underset{\sim}{x}_i, \underset{\sim}{x}_j) = \left\{ \sum_{k=1}^{p} \left(\frac{x_{ik} - x_{jk}}{s_k} \right)^2 \right\}^{\frac{1}{2}}$

$$d_M(\underset{\sim}{x}_i, \underset{\sim}{x}_j) = \sum_{k=1}^{p} \left| \frac{x_{ik} - x_{jk}}{s_k} \right|$$

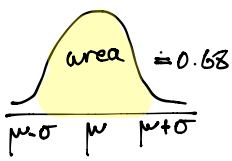
Distance matrix: Symmetric $n \times n$ matrix
- diagonal elements are 0
- (i,j) element $d_{\cdot}(\underset{\sim}{x}_i, \underset{\sim}{x}_j)$ ↘ M or E

Applications (Later)
① *Cluster analysis* (find graphs of similar observations)
② *Multidimensional scaling*: obtain a 2 dimensional representation of data with similar distance structure.

Multivariate Normal Distn

Review: (Univariate) Normal Distribution

$$X \sim N(\mu, \sigma^2), \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2} (x-\mu)^2 \right]$$



Given data $x_1, \cdots, x_n$, a simple graphical check for normality is a normal probability (normal quantile-quantile plot)

Ordered data: $x_{(1)} \le x_{(2)} \le x_{(3)} \le \cdots \le x_{(n)}$

Normal scores: $E[Z_{(i)}] \doteq \Phi^{-1}\left(\frac{i-\frac{1}{2}}{n}\right)$

$Z_{(i)}$: $N(0,1)$ order statistics

Φ^{-1}: quantile of $N(0,1)$

Plot $x_{(i)}$ vs $\Phi^{-1}\left(\frac{i-\frac{1}{2}}{n}\right)$ for $i=1,\cdots,n$

- If data are normal, points should lie close to a straight line
- R function: qqnorm
- Shapiro-Wilk test (R function: shapiro.test)

How to extend to multivariate case?

- random vector $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$

- want all one-dimensional projections $\underset{\sim}{a}^T \underset{\sim}{X}$ to be univariate normal.

Multivariate normal

$\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, C)$: assume C is positive definite

$\underset{\sim}{\mu}$: mean vector; C: covariance matrix

$$f(\underset{\sim}{x}) = f(x_1, \cdots, x_p) = \frac{1}{(2\pi)^{p/2}|C|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})^T C^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right]$$

$|C|$: determinants

Example: $p=2$

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \qquad |C| = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1-\rho^2)$$

$$C^{-1} = \frac{1}{|C|}\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$$

so $$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right]$$

Conditional dist'n of X_1 given $X_2 = x_2$

$$f(x_1|x_2) = \frac{f(x_1,x_2)}{f(x_2)}$$

$\Rightarrow X_1|X_2=x_2 \sim N(\mu_1 + \rho\left(\frac{x_2-\mu_2}{\sigma_2}\right), \sigma_1^2(1-\rho^2))$ conditional dist'n is still a Normal dist'n

Properties: $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, C)$

① : $\underset{\sim}{a}^T \underset{\sim}{X} \sim N(\underset{\sim}{a}^T \underset{\sim}{\mu}, \underset{\sim}{a}^T C \underset{\sim}{a})$

② If $\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X_1} \leftarrow \text{length } r \\ \underset{\sim}{X_2} \leftarrow \text{length } (p-r) \end{pmatrix}$

then the conditional dist'n $\underset{\sim}{X_1}$ given $\underset{\sim}{X_2} = \underset{\sim}{x_2}$ is also multivariate normal
(mean & covariance revealed later!)

③ $(\underset{\sim}{X}-\underset{\sim}{\mu})^T C^{-1} (\underset{\sim}{X}-\underset{\sim}{\mu}) \sim \chi^2_{(p)}$ ← chi-square with $df = p$

Application:

Checking multivariate normality

Data: $\underset{\sim}{x_1}, \cdots, \underset{\sim}{x_n}$ → *is this multivariate normal?*

- estimate $\underset{\sim}{\mu}$, C by $\bar{\underset{\sim}{x}}$ & S (as before)

Two approaches

① Mahalanobis distance (see textbook)

Define $d_i^2 = (\underset{\sim}{x_i} - \bar{\underset{\sim}{x}})^T S^{-1} (\underset{\sim}{x_i} - \bar{\underset{\sim}{x}})$ (↓ est. of $\underset{\sim}{\mu}$; ↘ est. of C)

If $\underset{\sim}{x_1}, \cdots, \underset{\sim}{x_n}$ are multivariate normal then $d_1^2, \cdots, d_n^2$ should be approaching $\chi^2_{(p)}$

- order $d_1^2, \cdots, d_n^2$ from smallest to largest
  $d_{(1)}^2 \leq d_{(2)}^2 \leq \cdots d_{(n)}^2$
- compare $d_{(1)}^2, \cdots, d_{(n)}^2$ to quantiles of $\chi^2_{(p)}$ dist'n.

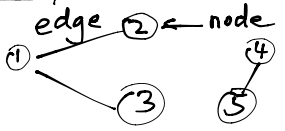
② Ad hoc: Use fact that all 1-dim projections of multivar normal are normal

- take (random) projections $\underset{\sim}{a_1}, \cdots, \underset{\sim}{a_m}$ and look at qqplots of $\{\underset{\sim}{a_k}^T \underset{\sim}{x_i}\}$ for $k = 1, \cdots, m$

Application: Graphical models

Idea: Describe dependence structure of $\underset{\sim}{X}$ using an (undirected) graph.

Example: $p = 5$



$$\underset{\sim}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_5 \end{pmatrix} \sim N_5(\underset{\sim}{\mu}, C)$$

X_1 depends on X_2 & X_3

X_2 & X_3 are conditionally independent given X_1

X_4 & X_5 are dependent

(X_4, X_5) & (X_1, X_2, X_3) are independent

Question:

Given C, how to extract this graph structure?

Partition: $\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X_1} \leftarrow r \\ \underset{\sim}{X_2} \leftarrow p-r \end{pmatrix}$ $\underset{\sim}{\mu} = \begin{pmatrix} \underset{\sim}{\mu_1} \\ \underset{\sim}{\mu_2} \end{pmatrix}$

$$C = \begin{pmatrix} C_{11} \leftarrow r \times r & C_{12} \leftarrow r \times (p-r) \\ C_{21} & C_{22} \rightarrow (p-r) \times (p-r) \end{pmatrix}$$

C_{12}^T

and $K = C^{-1} = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$

↳ called "concentration matrix"

$$\underset{\sim}{X}_1 | \underset{\sim}{X}_2 = \underset{\sim}{x}_2 \sim N_r(\underset{\sim}{\mu}_{1|2}, C_{1|2})$$

where $\underset{\sim}{\mu}_{1|2} = \underset{\sim}{\mu}_1 + C_{12} C_{22}^{-1} (\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2)$ — conditional mean
$= \underset{\sim}{\mu}_1 - K_{11}^{-1} K_{12} (\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2)$

$C_{1|2} = C_{11} - C_{12} C_{22}^{-1} C_{21}$ — conditional covariance
$= K_{11}^{-1}$

Take $r=2$, look at dist'n of 2 variables conditional on remaining $p-2$.
covariance $= K_{11}^{-1}$ → 2×2 matrix

What do we need for 2 variables to be independent given the remaining $p-2$?
$\Rightarrow$ [off-diagonal elements of $K_{11}^{-1} = 0$.] $\Leftrightarrow$ [off-diagonal elements of $K_{11} = 0$]

How to draw graph?
- all dependence information necessary to draw graph contained in $K = C^{-1}$.

$$K = \begin{pmatrix} k_{11} & \cdots & k_{1p} \\ \vdots & \ddots & \vdots \\ k_{p1} & \cdots & k_{pp} \end{pmatrix}$$

- If $k_{ij} \neq 0$ then there is an edge between vars i & j
- If $k_{ij} = 0$ ··· no edge

Example: Exam marks in 5 subjects
- mechanics
- vectors
- algebra — 88 students
- analysis
- statistics

$$\hat{R} = \begin{pmatrix} 1 & 0.55 & 0.55 & 0.41 & 0.39 \\ & 1 & 0.61 & 0.49 & 0.44 \\ & & 1 & 0.71 & 0.66 \\ & & & 1 & 0.61 \\ & & & & 1 \end{pmatrix}$$

Look at $\hat{K} = \hat{R}^{-1}$