Math A7800 8/29/17, Tue

Marshak 1026 → NAC 7/313A   6-7:40 pm

office : NAC 4/114B

email : shirchat1@gmail.com

office hour : (after class) 7:45 - 8:45 pm , Tue & Thur (in office)

Grading : Midterms (60%), Final (40%) (cumulative)

Three midterms (lowest will be dropped)

Usually, the first class of Oct, Nov, Dec

Hws : post it at Blackboard.

will Not be collected.

Exam questions are mostly from Homework and problems/theorm covered in class.

Text book : 1. Applied Multivariate Statistical Analysis by Johnson and Wichern.

2. Applied Linear Statistical Models by Kutner, Nachtsheim, Neter, Li. (5th edition)

Topics : Multivariate Analysis including Multivariate Normal distribution,

Linear regression (Matrix approach),

Multiple Linear regression, associated inference,

Some Diagnostic tools, ANOVA.

If time permits, then we'll cover Generalized Linear Model.

- random sampling

(1) measurements taken on different items (or trials) are unrelated to one another.

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

\# of data / dimension

p-different variables taken on the same item or trial

$$X_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{bmatrix}$$

n-observations

"single multivariate observation"

a sample of size n from a p-variate "population".

← 1st multivariate observation

← nth multivariate observation

Random sample

We say that $(X_1, X_2, \ldots, X_n)$ is a random sample of size n from a distribution $f(x)$, when $X_1, \ldots, X_n$ are i.i.d. having common distribution $f(x)$.

Recall The cdf (cumulative distribution function) of a R.V. Y is $F_Y(t) = \mathbb{P}(Y \le t)$

There are mostly two kinds of random variables : Discrete and continuous

For discrete random variables, we have pmf (probability mass function)

For continuous random variables, we have pdf (probability density function)

If Y has pmf $f(y)$, then $f(y) = \mathbb{P}(Y=y)$.

$$\sum_y f(y) = 1, \quad f(y) \ge 0.$$

If Y has pdf $f(y)$, then $\mathbb{P}(Y\in A) = \int_A f(y)\,dy$

we should have $f(y)\ge 0$ and $\int f(y)\,dy = 1$.

Suppose Y takes values $\{y_1, \dots, y_n, \dots\}$ and "discrete"

$\mathbb{P}(Y = y_i) = P_i = f_Y(y_i)$

How to compute $\mathbb{E}[g(Y)] = \sum_i g(y_i)\, P_i$

If Y is continuous with pdf $f(y)$, then

$$\mathbb{E}[g(Y)] = \int g(y) f(y)\, dy$$

- $\operatorname{var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2$

Multiple random variables

$\underset{\sim}{X} = (X_1, X_2, \dots, X_k)$ is a random vector with joint pdf/pmf

$f_{\underset{\sim}{X}}(\underset{\sim}{x})$, $\underset{\sim}{x} = (x_1, \dots, x_k)$

For pmf case,

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k)$$

For pdf case,

$$\mathbb{P}((X_1, \dots, X_k) \in A) = \int\int_A \cdots \int f_{\underset{\sim}{X}}(\underset{\sim}{x})\, d\underset{\sim}{x}$$

Definition Let $\underset{\sim}{X}$ be a random variable assuming the values $x_1, x_2, \dots$ with corresponding probabilities $f(x_1), f(x_2), \dots$. The mean or expected value of X is defined by

$$\mathbb{E}(X) = \sum_k x_k f(x_k)$$

Given joint distribution $f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \mathbb{P}(X_1 = x_1, \dots, X_k = x_k)$

how to obtain marginal dist. of X_1 $= \int\int_A \cdots \int f_{\underset{\sim}{X}}(\underset{\sim}{x})\, d\underset{\sim}{x}$

$$f_{X_1}(x_1) \overset{?}{=} \underset{\text{over } x_2, \dots, x_k}{\int \cdots \int} f_{\underset{\sim}{X}}(x_1, \dots, x_k)$$

Conditional Distribution

Suppose $(\underset{\sim}{X}, \underset{\sim}{Y})$ has joint pdf $f_{(\underset{\sim}{X}, \underset{\sim}{Y})}(\underset{\sim}{x}, \underset{\sim}{y})$

The conditional pdf

$$f_{\underset{\sim}{X}|\underset{\sim}{Y}}(\underset{\sim}{x}|\underset{\sim}{y}) = \frac{f_{(\underset{\sim}{X}, \underset{\sim}{Y})}(\underset{\sim}{x}, \underset{\sim}{y})}{\int f_{\underset{\sim}{X}, \underset{\sim}{Y}}(\underset{\sim}{x}, \underset{\sim}{y})\, d\underset{\sim}{x}} = \mathbb{P}(\underset{\sim}{X} | \underset{\sim}{Y})$$

EX

Suppose (X, Y) has joint pdf $f(x, y) = e^{-y}$, $0 \le x \le y$

Obtain $\mathbb{E}[X^2 | Y = 3]$

Step 1 Find conditional pdf.

$$f_{X|Y}(x|3) = \frac{e^{-y}}{\int_0^3 e^{-y}\, dx} = \frac{1}{3}, \quad 0 \le x \le 3$$

$U \sim \text{uniform}(0, 3)$

$$\mathbb{E}(U^2) = \int_0^3 x^2 \frac{1}{3}\, dx$$

the conditional probability of the event $X = x_k$, given that $Y = y_j$ is

$$\mathbb{P}(X = x_k | Y = y_j) = \frac{P(y_j, x_k)}{f(y_j)}$$

$$\mathbb{E}(X^2 \mid Y=3) = \int x^2 f_{X|Y}(x|3)\,dx$$

$$= \int_0^3 x^2 \frac{1}{3}\,dx$$

$\mathbb{E}(X^2 \mid Y) \longleftrightarrow$ this should be a function of Y.

$$= \int_0^Y x^2 \frac{1}{Y}\,dx = \frac{Y^2}{3}.$$

$$\left.\frac{x^3}{3}\right|_0^Y \frac{1}{Y} \quad \frac{Y^3}{3}$$

$$\mathbb{E}\left[\mathbb{E}(X^2 \mid Y)\right] = \mathbb{E}\left[\frac{Y^2}{3}\right]$$

$$= \mathbb{E}(X^2).$$

For an event A, $\mathbb{1}_A$ is a random variable which take two values,

$$\mathbb{1}_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbb{E}[\mathbb{1}_A] = 1 \cdot P(A) + 0 \cdot P(A^c) = \mathbb{P}(A)$$

$$\mathbb{1}_A \cdot \mathbb{1}_B = \mathbb{1}_{A \cap B}.$$

Independence

$\underset{\sim}{X}, \underset{\sim}{Y}$ are <u>independent</u> if $f_{\underset{\sim}{X},\underset{\sim}{Y}}(\underset{\sim}{x},\underset{\sim}{y}) = f_{\underset{\sim}{X}}(\underset{\sim}{x}) f_{\underset{\sim}{Y}}(\underset{\sim}{y})$ for all $\underset{\sim}{x}, \underset{\sim}{y}$.

In other words,

$$f_{\underset{\sim}{X}|\underset{\sim}{Y}}(\underset{\sim}{x}|\underset{\sim}{y}) = f_X(\underset{\sim}{x}) \text{ and vice-versa.}$$

If $\underset{\sim}{X}, \underset{\sim}{Y}$ are independent, then

$$\mathbb{E}[g_1(\underset{\sim}{X})\, g_2(\underset{\sim}{Y})] = \int g_1(\underset{\sim}{x}) g_2(\underset{\sim}{y}) f_{\underset{\sim}{X},\underset{\sim}{Y}}(\underset{\sim}{x},\underset{\sim}{y})\, d\underset{\sim}{x}\, d\underset{\sim}{y}$$

$$= \int g_1(\underset{\sim}{x}) g_2(\underset{\sim}{y}) f_{\underset{\sim}{X}}(\underset{\sim}{x}) f_{\underset{\sim}{Y}}(\underset{\sim}{y})\, d\underset{\sim}{x}\, d\underset{\sim}{y}$$

$$= \int g_1(\underset{\sim}{x}) f_{\underset{\sim}{X}}(\underset{\sim}{x})\, d\underset{\sim}{x} \int g_2(\underset{\sim}{y}) f_{\underset{\sim}{Y}}(\underset{\sim}{y})\, d\underset{\sim}{y}$$

$$= \mathbb{E}[g_1(\underset{\sim}{X})]\, \mathbb{E}[g_2(\underset{\sim}{Y})].$$

Q. If X, Y are independent, then

$$\mathbb{P}(X \le 2, Y > 3) = \mathbb{P}(X \le 2)\, \mathbb{P}(Y > 3).$$

Define $g_1 = \mathbb{1}_{(-\infty, 2]}$, $g_2 = \mathbb{1}_{(3, \infty)}$

$$\mathbb{E}[g_1(X)] = \mathbb{P}(X \le 2)$$

$$\mathbb{E}[g_2(Y)] = \mathbb{P}(Y > 3).$$

$$\mathbb{E}[g_1(X) g_2(Y)] = \mathbb{P}(X \le 2, Y > 3).$$

Suppose we have n-samples from a p-variate distribution

$\underset{\sim}{X}_1, \underset{\sim}{X}_2, \dots, \underset{\sim}{X}_n$. (each is a $p \times 1$ vector)

They are independent and common distribution $f(\underset{\sim}{x}) = f(x_1, \dots, x_p)$.

$\underset{\sim}{X}_i = (x_1, \dots, x_p)$.

$1 \le i \le n$. $\quad \underset{\sim}{Y} = (Y_1, \dots, Y_p)$

$\mathbb{E}[\underset{\sim}{Y}] = (\mathbb{E}(Y_1), \dots, \mathbb{E}(Y_p))$

If $\underset{\sim}{X}$ is $(p \times 1)$, so is $\mathbb{E}(\underset{\sim}{X})$

$\mathrm{cov}(\underset{\sim}{Y}) = \Big(\quad \Big)_{p \times p}$ $(i,j)^{th}$ entry is $\mathrm{cov}(Y_i, Y_j)$

mqudd0@citymail.cuny.edu

$(\underset{\sim}{x} - \underset{\sim}{\mu})^{\phi} (\underset{\sim}{x} - \underset{\sim}{\mu})'$

the correlation coefficient

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]}{\sigma_X \sigma_Y}.$$

Math A7800 8/31/17, Thur

Def Let X be a random variable with second moment $\mathbb{E}(X^2)$ and let $\mu = \mathbb{E}(X)$ be its mean. $Var(X) = \mathbb{E}[(X-\mu)^2] = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Suppose $\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{pmatrix}$ is a p-variate random vector

Then, $\underset{p \times 1 \text{ vector}}{\mathbb{E}(\underset{\sim}{Y})} = \begin{pmatrix} \mathbb{E}(Y_1) \\ \vdots \\ \mathbb{E}(Y_p) \end{pmatrix} = \underset{\sim}{\mu}$.

covariance : variance of sum

$$\mathrm{cov}(\underset{\sim}{Y}) = \mathbb{E}\Big[\underbrace{\underbrace{(\underset{\sim}{Y} - \underset{\sim}{\mu})}_{p \times 1} \underbrace{(\underset{\sim}{Y} - \underset{\sim}{\mu})^T}_{1 \times p}}_{p \times p \text{ matrix "symmetric"}} \Big]$$

$\shortparallel$ $Var(\underset{\sim}{Y})$

$= \Sigma$ (population variance)

$\Sigma_{ij} = \mathbb{E}[(Y_i - \mu_i)(Y_j - \mu_j)] = \mathrm{cov}(Y_i, Y_j)$

$(i,j)^{th}$ entry of Σ $\quad = \mathbb{E}(Y_i Y_j) - \mu_{Y_i} \mu_{Y_j}$

Def The covariance of X and Y is defined by $\mathrm{cov}(X,Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}(XY) - \mu_X \mu_Y$.

If X, Y are independent, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

Thm If X, Y are independent, then $\mathrm{cov}(X,Y) = 0$.

Note: suppose

$$\underset{\sim}{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}, \quad \underset{\sim}{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix}$$

Then $\underset{\sim}{u}\,\underset{\sim}{v}^T =$

$$\underset{p\times 1}{\begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}} \underset{1\times p}{(v_1, \cdots, v_p)}$$

$$= \begin{array}{c} \\ i \\ \\ \end{array} \begin{array}{c} j \\ \begin{pmatrix} u_1v_1 & \cdots & & u_1v_p \\ \vdots & & \boxed{u_iv_j} & \vdots \\ u_pv_1 & \cdots & & u_pv_p \end{pmatrix}_{p\times p} \end{array}$$

$$\Sigma = \mathbb{E}\Big[\underbrace{(\underset{\sim}{y} - \underset{\sim}{\mu})}_{\underset{\sim}{u}}\underbrace{(\underset{\sim}{y} - \underset{\sim}{\mu})^T}_{\underset{\sim}{u}^T}\Big]$$

The (i,j)th entry is $(y_i - \mu_i)(y_j - \mu_j)$. So,

$$\boxed{\underbrace{\Sigma_{ij}}_{\sigma_{ij}} = \mathbb{E}\big[(y_i - \mu_i)(y_j - \mu_j)\big] = cov(y_i, y_j)}$$

$$= \Sigma_{ji} \quad (\text{as } cov(y_i, y_j) = cov(y_j, y_i))$$

So, Σ is symmetric.

If the joint distribution of X and Y is $\{P(x_j, y_k)\}$, then the expectation of XY is given by $\mathbb{E}(XY) = \sum x_j y_k P(x_j, y_k)$.

suppose $D = Diag(\Sigma) = \begin{pmatrix} \sigma_{11} & & 0 \\ & \sigma_{11} & \\ 0 & & \ddots \; \sigma_{pp} \end{pmatrix}$

$\sigma_{ii} = var(y_i)$, $i = 1, 2, \ldots, p$.

$$D^{1/2} = \begin{pmatrix} \sqrt{\sigma_{11}} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\sigma_{pp}} \end{pmatrix}, \quad D^{-1/2} = \begin{pmatrix} 1/\sqrt{\sigma_{11}} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{\sigma_{pp}} \end{pmatrix}$$

$$= D^{-1/2}\Sigma D^{-1/2} = \begin{pmatrix} 1/\sqrt{\sigma_{11}} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{\sigma_{pp}} \end{pmatrix} \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix} \begin{pmatrix} 1/\sqrt{\sigma_{11}} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{\sigma_{pp}} \end{pmatrix}$$

ex

$$\begin{bmatrix} 1/\sqrt{a_{11}} & 0 \\ 0 & 1/\sqrt{a_{22}} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1/\sqrt{a_{11}} & 0 \\ 0 & 1/\sqrt{a_{22}} \end{bmatrix} = \begin{pmatrix} \frac{a_{11}}{\sqrt{a_{11}}} & \frac{a_{12}}{\sqrt{a_{11}}} \\ \frac{a_{21}}{\sqrt{a_{22}}} & \frac{a_{22}}{\sqrt{a_{22}}} \end{pmatrix} \begin{pmatrix} 1/\sqrt{a_{11}} & 0 \\ 0 & 1/\sqrt{a_{22}} \end{pmatrix} = \begin{pmatrix} a_{11}/a_{11} & \frac{a_{12}}{\sqrt{a_{11}a_{22}}} \\ \frac{a_{21}}{\sqrt{a_{11}a_{22}}} & a_{22}/a_{22} \end{pmatrix} = \begin{pmatrix} 1 & \frac{a_{12}}{\sqrt{a_{11}a_{22}}} \\ \frac{a_{21}}{a_{11}a_{22}} & 1 \end{pmatrix}$$

$$\begin{pmatrix} \frac{a_{11}}{\sqrt{a_{11}}} & \frac{a_{12}}{\sqrt{a_{11}}} \\ \frac{a_{21}}{\sqrt{a_{22}}} & \frac{a_{22}}{\sqrt{a_{22}}} \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} \ddots & & \\ & \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\cdot\sigma_{jj}}} & \\ & & \end{pmatrix}_{p\times p}}_{\text{Correlation matrix}}$$

$= $ Correlation coeff. of (y_i, y_j)

So, $D^{-1/2}\Sigma D^{-1/2}$ gives the correlation matrix.

Also, remember: For two random vectors $\underset{\sim}{y}$ and $\underset{\sim}{z}$

$$\begin{pmatrix} 1 & & \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\cdot\sigma_{jj}}} \\ & 1 & \\ \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} & & \ddots \; 1 \end{pmatrix}$$

$$\underline{\Sigma = D^{1/2} R D^{1/2}}$$

$$var(\underset{\sim}{Y}) = cov(\underset{\sim}{Y}, \underset{\sim}{Y}) = \mathbb{E}\left(\left(\underset{\sim}{Y} - \mu\right)\left(\underset{\sim}{Y} - \mu\right)^T\right].$$

"$cov(\underset{\sim}{Y}, \underset{\sim}{Z}) = \mathbb{E}\left[\left(\underset{\sim}{Y} - \mathbb{E}(\underset{\sim}{Y})\right)\left(\underset{\sim}{Z} - \mathbb{E}(\underset{\sim}{Z})\right)\right]$"

$$= \mathbb{E}(\underset{\sim}{Y}\underset{\sim}{Z}) - \mathbb{E}(\underset{\sim}{Y})\mathbb{E}(\underset{\sim}{Z})$$

Suppose $\underset{\sim}{Y}$ is a random vector and A is a matrix consisting of constants.

"$\boxed{\underset{\sim}{Y} \text{ is } p\times 1}$, $\boxed{A \text{ is } m\times p}$"

$$\underset{m\times 1}{\underset{\sim}{Z}} = \underset{m\times p}{A}\underset{p\times 1}{\underset{\sim}{Y}} \qquad \mathbb{E}(\underset{\sim}{Z}) = \begin{pmatrix} \mathbb{E}(Z_1) \\ \vdots \\ \mathbb{E}(Z_p) \end{pmatrix} = \begin{pmatrix} a_{1\cdot} \\ \vdots \\ a_{p\cdot} \end{pmatrix} \mathbb{E}(\underset{\sim}{Y})$$

$$Z_i = \sum_{j=1}^{p} a_{ij} y_j \qquad = A\,\mathbb{E}(\underset{\sim}{Y})$$

So, $\mathbb{E}(Z_i) = \sum_{j=1}^{p} a_{ij}\, \mathbb{E}(y_j)$.

$$= a_{i\cdot}\, \mathbb{E}(y_j)$$

($a_{i\cdot}$ is the ith row of A. $a_{\cdot j}$ is the jth column of A.)

$$A = \underset{4\times p}{\begin{bmatrix} a_{1\cdot} \to \\ a_{2\cdot} \to \\ a_{3\cdot} \to \\ a_{4\cdot} \to \end{bmatrix}} \underset{p\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}} \quad \underset{\sim}{Y}$$

Remember

$$\mathbb{E}(A\underset{\sim}{X}) = A\,\mathbb{E}(\underset{\sim}{X}).$$

similarly, $\mathbb{E}(\underset{\sim}{Y}B) = \mathbb{E}(\underset{\sim}{Y})B$ for constant matrices A, B.

$$var(A\underset{\sim}{Y}) = \mathbb{E}\left((A\underset{\sim}{Y})^2\right) - \left(\mathbb{E}(A\underset{\sim}{Y})\right)^2$$

$$= A^2\,\mathbb{E}(Y^2) - A^2\left(\mathbb{E}(\underset{\sim}{Y})\right)^2$$

$$= A^2\left(\mathbb{E}(Y^2) - \left(\mathbb{E}(\underset{\sim}{Y})\right)^2\right)$$

$$= A^2\, var(\underset{\sim}{Y}).$$

or,

$$= \mathbb{E}\left[\left(A\underset{\sim}{Y} - \mathbb{E}(A\underset{\sim}{Y})\right)\left(A\underset{\sim}{Y} - \mathbb{E}(A\underset{\sim}{Y})\right)^T\right]$$

$$= \mathbb{E}\left[\left(A\left(\underset{\sim}{Y} - \mathbb{E}(\underset{\sim}{Y})\right)\right)\left(A\left(\underset{\sim}{Y} - \mathbb{E}(\underset{\sim}{Y})\right)\right)^T\right].$$

$$= \mathbb{E}\left[A\left(\underset{\sim}{Y} - \mathbb{E}(\underset{\sim}{Y})\right)\left(\underset{\sim}{Y} - \mathbb{E}(\underset{\sim}{Y})\right)^T A^T\right]$$

$$= A\,\underbrace{\mathbb{E}\left[\left(\underset{\sim}{Y} - \mathbb{E}(\underset{\sim}{Y})\right)\left(\underset{\sim}{Y} - \mathbb{E}(\underset{\sim}{Y})\right)^T\right]}\,A^T$$

$$= A\, var(\underset{\sim}{Y})\, A^T$$

$$= A\,\Sigma\,A^T$$

So, $cov(A\underset{\sim}{Y}, B\underset{\sim}{Z}) = A\, cov(\underset{\sim}{Y}, \underset{\sim}{Z})\, B^T$

$$= A\,\Sigma\,B^T.$$

$$cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}\left[(x - \mathbb{E}(x))(y - \mathbb{E}(y))'\right]$$

Suppose $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n$ are i.i.d. from a p-variate distribution having mean $\underset{\sim}{\mu}$ and var Σ.

Sample mean $\bar{\underset{\sim}{X}} = \frac{1}{n}\sum_{i=1}^{n} \underset{\sim}{X}_i$ (component-wise mean)

Sample variance

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

Note: $\mathbb{E}[\underset{\sim}{X}_1] = \underset{\sim}{\mu}$

$$\Sigma = \mathbb{E}\left[(\underset{\sim}{X}_1 - \underset{\sim}{\mu})(\underset{\sim}{X}_1 - \underset{\sim}{\mu})^T\right]$$

variance-covariance matrix

$$= \mathbb{E}\left[\underset{\sim}{X}_1\underset{\sim}{X}_1^T - \underset{\sim}{X}_1\underset{\sim}{\mu}^T - \underset{\sim}{\mu}\underset{\sim}{X}_1^T + \underset{\sim}{\mu}\underset{\sim}{\mu}^T\right]$$

$$= \mathbb{E}(\underset{\sim}{X}_1\underset{\sim}{X}_1^T) - \underset{\sim}{\mu}\underset{\sim}{\mu}^T - \underset{\sim}{\mu}\underset{\sim}{\mu}^T + \underset{\sim}{\mu}\underset{\sim}{\mu}^T$$

$$= \mathbb{E}(\underset{\sim}{X}_1\underset{\sim}{X}_1^T) - \underset{\sim}{\mu}\underset{\sim}{\mu}^T$$

So, $\mathbb{E}[\underset{\sim}{X}_1\underset{\sim}{X}_1^T] = \Sigma + \underset{\sim}{\mu}\underset{\sim}{\mu}^T$.

$$\mathbb{E}[\bar{\underset{\sim}{X}}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\underset{\sim}{X}_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\underset{\sim}{X}_i) = \underset{\sim}{\mu}$$

$$\text{``}\ \frac{1}{n} n\mu$$

If $u_1, \dots, u_n$ are iid with variance σ^2

$var(\bar{u}) =$

$$var\left(\frac{1}{n}\sum_{i=1}^{n}u_i\right) = \frac{1}{n^2}var\left(\sum_{i=1}^{n}u_i\right)$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n}var(u_i) + \sum_{i\neq j}cov(u_i, u_j)\right]$$

$$= \frac{1}{n^2}\left[n\,var(u_1) + 0\right] = \frac{1}{n}\sigma^2$$

$var(\bar{X})$ $= \mathbb{E}[\bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}^T] - \underset{\sim}{\mu}\underset{\sim}{\mu}^T$

$$= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\underset{\sim}{X}_i \cdot \frac{1}{n}\sum_{j=1}^{n}\underset{\sim}{X}_j\right] - \underset{\sim}{\mu}\underset{\sim}{\mu}^T$$

$$= \frac{1}{n^2}\sum_i\sum_j \mathbb{E}(X_i X_j^T) - \underset{\sim}{\mu}\underset{\sim}{\mu}^T$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n}\mathbb{E}[\underset{\sim}{X}_i\underset{\sim}{X}_i^T] + \sum_{i\neq j}\mathbb{E}(\underset{\sim}{X}_i)\mathbb{E}(\underset{\sim}{X}_j^T)\right] - \underset{\sim}{\mu}\underset{\sim}{\mu}^T$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n}(\Sigma + \underset{\sim}{\mu}\underset{\sim}{\mu}^T) + \sum_{i\neq j}\underset{\sim}{\mu}\underset{\sim}{\mu}^T\right] - \underset{\sim}{\mu}\underset{\sim}{\mu}^T = \frac{1}{n}\Sigma$$

Math A7800

9/5/2011, Tue

Recall

$\underset{\sim}{X}_1, \underset{\sim}{X}_2, \dots, \underset{\sim}{X}_n$ are iid with common p-variate joint distribution $f(\underset{\sim}{x}) = f(x_1, \dots, x_p)$

Define $\boxed{\bar{\underset{\sim}{X}} = \frac{1}{n}\sum_{j=1}^{n} \underset{\sim}{X}_j}$

$\boxed{S_n = \frac{1}{n}\sum_{j=1}^{n} (\underset{\sim}{X}_j - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_j - \bar{\underset{\sim}{X}})^T}$

$\mathrm{cov}(U,V) = \mathbb{E}[(U-\mu_U)(V-\mu_V)]$

$\mu_i = \int_{\mathbb{R}^p} x_i f(\underset{\sim}{x})\, d\underset{\sim}{x}$, $i = 1, 2, \dots, p$.

we saw that

$\mathbb{E}(\bar{\underset{\sim}{X}}) = \underset{\sim}{\mu}$ (population mean)

$\mathrm{Var}(\bar{\underset{\sim}{X}}) = \mathbb{E}[(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu})(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu})^T]$, $\mathbb{E}(\bar{\underset{\sim}{X}}) = \underset{\sim}{\mu}$

$= \boxed{\frac{1}{n}\Sigma}$, where Σ is the population variance.

$\underset{(p\times p)}{\Sigma} = (\sigma_{ij})$, $i,j = 1, 2, \dots, p$

$$\sigma_{ij} = \int_{\mathbb{R}^p} [(x_i - \mu_i)(x_j - \mu_j)]\, f(\underset{\sim}{x})\, d\underset{\sim}{x}$$

☆

Ex If the common distribution is given by the bivariate pdf $f(u,v) = e^{-v}$ for $0 \le u \le v$, then find $\underset{\sim}{\mu}, \Sigma$.

Go back

$\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n$ are n-samples

$$\boxed{S_n = \frac{1}{n}\sum_{j=1}^{n}(\underset{\sim}{X}_j - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_j - \bar{\underset{\sim}{X}})^T}$$

constant

$$\underline{S_n} = \frac{1}{n}\sum_{j=1}^{n}\left[\underset{\sim}{X}_j \underset{\sim}{X}_j^T - \bar{\underset{\sim}{X}}\underset{\sim}{X}_j^T - \underset{\sim}{X}_j\bar{\underset{\sim}{X}}^T + \bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}^T\right]$$

$$= \frac{1}{n}\sum_{j=1}^{n}\underset{\sim}{X}_j\underset{\sim}{X}_j^T - \bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}^T - \bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}^T + \bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}^T$$

$$= \frac{1}{n}\sum_{j=1}^{n}\underset{\sim}{X}_j\underset{\sim}{X}_j^T - \bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}^T$$

$\frac{1}{n}\Sigma + \underset{\sim}{\mu}\underset{\sim}{\mu}^T$

So, $\mathbb{E}(S_n) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[\underset{\sim}{X}_j\underset{\sim}{X}_j^T] - \mathbb{E}[\bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}^T]$

$\boxed{\mathbb{E}(\underset{\sim}{X}\underset{\sim}{X}^T) = var(\underset{\sim}{X}) + \mathbb{E}(\underset{\sim}{X})\mathbb{E}(\underset{\sim}{X})^T}$ ← $var(\underset{\sim}{X}) = \mathbb{E}[\underset{\sim}{X}\underset{\sim}{X}^T] - \mathbb{E}(\underset{\sim}{X})\mathbb{E}($

So, $\mathbb{E}(\underset{\sim}{X}\underset{\sim}{X}^T) = \underline{var(\underset{\sim}{X}) + \mathbb{E}(\underset{\sim}{X})}$

$$= \frac{1}{n}\sum_{j=1}^{n}(\Sigma + \underset{\sim}{\mu}\underset{\sim}{\mu}^T) - \left(\frac{1}{n}\Sigma + \underset{\sim}{\mu}\underset{\sim}{\mu}^T\right)$$

$\Sigma + \mu\mu^T - \frac{1}{n}\Sigma - \cancel{\mu\mu^T}$

$$= \left(1 - \frac{1}{n}\right)\Sigma.$$

$$= \frac{n-1}{n}\Sigma \qquad \mathbb{E}(S_n) = \frac{n-1}{n}\Sigma$$

$$\Sigma = \frac{n}{n-1}\mathbb{E}(S_n) = \frac{n}{n-1}S_n$$

$$= \frac{n}{n-1}\left(\frac{1}{n}\sum_{j=1}^{n}(\underset{\sim}{X}_j - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_j - \bar{\underset{\sim}{X}})^T\right)$$

Define $S = \frac{n}{n-1}S_n = \frac{1}{n-1}\sum_{j=1}^{n}(\underset{\sim}{X}_j - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_j - \bar{\underset{\sim}{X}})^T$

Then $\mathbb{E}(S) = \Sigma$, S is called <u>the sample variance of $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n$</u>.

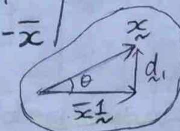
⊙ Suppose $\underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, $\underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$

The <u>deviation</u> corresponding to $\underset{\sim}{x}$ is

$$\underset{\sim}{d}_1 = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} = \underset{\sim}{x} - \bar{x}\underset{\sim}{1}$$

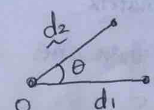
$\underset{\sim}{1}$ is a vector (in this case $n \times 1$) consisting of all 1's

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

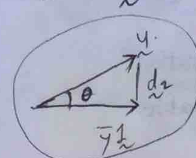


The deviation corresponding to $\underset{\sim}{y}$ is

$\underset{\sim}{d}_2 = \underset{\sim}{y} - \bar{y}\underset{\sim}{1}$, where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.



length of $\underset{\sim}{d}_1$?
length of $\underset{\sim}{d}_2$?
The angle θ ?



$\boxed{\text{length of } \underset{\sim}{d}_1} = \sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \sqrt{n-1}\, S_{xx}$ — sample variance of $\underset{\sim}{x}$.

$\|\underset{\sim}{d}_1\|$ $\qquad \frac{\|d_1\|}{\sqrt{n-1}} = \sqrt{n-1}\,S_{xx}$

$\|\underset{\sim}{d}_2\|$

$\boxed{\text{length of } \underset{\sim}{d}_2} = \sqrt{n-1}\, S_{yy}$ — sample variance of $\underset{\sim}{y}$.

$$\cos\theta = \frac{\underset{\sim}{d}_1 \cdot \underset{\sim}{d}_2}{\|\underset{\sim}{d}_1\| \cdot \|\underset{\sim}{d}_2\|} = \frac{(n-1)S_{xy}}{\sqrt{n-1}S_{xx}\sqrt{n-1}S_{yy}}$$

$$= \frac{S_{xy}}{S_{xx}S_{yy}} = \text{correlation coefficients between } \underset{\sim}{x}, \underset{\sim}{y}.$$

$$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}).$$

Generalized variance

Suppose $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n$ are iid p-variate random vectors.

Then the sample variance $\boxed{S = \frac{1}{n-1}\sum_{i=1}^{n}(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})^T}$

is $p\times p$ matrix.

If we want one value for the variability, then one such measure is $|S|$ = Determinant of S.

Define $\underset{(n\times p)}{X} = \begin{bmatrix} \underset{\sim}{X}_1^T \\ \vdots \\ \underset{\sim}{X}_n^T \end{bmatrix}$, each row $\leftrightarrow$ one observation, each column $\leftrightarrow$ one variable.

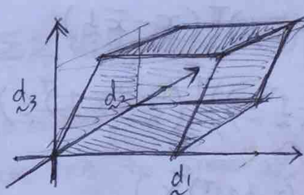
called "data matrix"

Let $\underset{\sim}{y}_1, \dots, \underset{\sim}{y}_p$ be the p-columns of X.

Let $\underset{\sim}{d}_1, \underset{\sim}{d}_2, \dots, \underset{\sim}{d}_p$ be the corresponding deviations.

$|S| = n^{-p} \cdot (\text{volume of the parallelogram formed by } \underset{\sim}{d}_1, \dots, \underset{\sim}{d}_p)^2$

picture for $p=3$.



$\rho^{1/2} = \frac{(S_{xy})^2}{(S_{xx}S_{yy})^2}$

$(S_{xy})^2 = (S_{xx}S_{yy})^2\rho$

proof for $p=2$

$X = [\underset{\sim}{y}_1\ \underset{\sim}{y}_2] = \begin{bmatrix} \underset{\sim}{X}_1^T \\ \vdots \\ \underset{\sim}{X}_n^T \end{bmatrix}$, $\underset{\sim}{X}_j$ is 2×1 for all j.

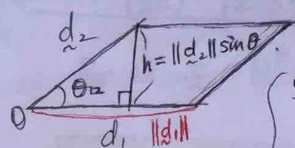
$\boxed{\rho_{12} = \frac{S_{y_1y_2}}{\sqrt{S_{y_1y_1}S_{y_2y_2}}},\ \rho_{12}^2 = \frac{S_{y_1y_2}^2}{S_{y_1y_1}S_{y_2y_2}}}$

Then $S = \begin{bmatrix} S_{y_1y_1} & S_{y_1y_2} \\ S_{y_2y_2} & S_{y_2y_2} \end{bmatrix}$, then $|S| = S_{y_1y_1}S_{y_2y_2} - (S_{y_1y_2})^2$

$= S_{y_1y_1}S_{y_2y_2} - S_{y_1y_1}S_{y_2y_2}\rho_{12}^2$

$= S_{y_1y_1}S_{y_2y_2}(1-\rho_{y_1y_2}^2)$

ρ = corr. coeff.



area $= \|\underset{\sim}{d}_1\| \cdot \|\underset{\sim}{d}_2\| \sin\theta$.

$= \sqrt{(n-1)S_{y_1y_1}}\sqrt{(n-1)S_{y_2y_2}}\sqrt{1-\cos^2\theta}$

$= (n-1)\sqrt{S_{y_1y_1}S_{y_2y_2}(1-\rho_{y_1y_2}^2)}$

$= (n-1)\sqrt{|S|}$.

So, $|S| = \frac{(\text{area})^2}{(n-1)^2}$

$= (n-1)^{-2}(\text{area})^2$

Recall $\underset{\sim}{1}$ is a vector of all 1's.

so, $\underset{\sim}{1}\underset{\sim}{1}^T$ = square matrix of all 1's

$$\underset{\sim}{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}_{n\times 1} \qquad \bar{x} = \frac{1}{n}\underbrace{\underset{\sim}{1}^T}_{(1\times n)}\underbrace{\underset{\sim}{x}}_{(n\times 1)} = \frac{1}{n}\underset{\sim}{x}^T\underset{\sim}{1}$$

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = (\underset{\sim}{x}-\bar{x}\underset{\sim}{1})^T(\underset{\sim}{x}-\bar{x}\underset{\sim}{1})$$

$$= \left(\underset{\sim}{x}-\underset{\sim}{1}\left(\tfrac{1}{n}\underset{\sim}{1}^T\underset{\sim}{x}\right)\right)^T\left(\underset{\sim}{x}-\underset{\sim}{1}\left(\tfrac{1}{n}\underset{\sim}{1}^T\underset{\sim}{x}\right)\right)$$

$$= \left((I-\tfrac{1}{n}\underset{\sim}{1}\underset{\sim}{1}^T)\underset{\sim}{x}\right)^T\left((I-\tfrac{1}{n}\underset{\sim}{1}\underset{\sim}{1}^T)\underset{\sim}{x}\right)$$

$$= \underset{\sim}{x}^T(I-\tfrac{1}{n}\underset{\sim}{1}\underset{\sim}{1}^T)(I-\tfrac{1}{n}\underset{\sim}{1}\underset{\sim}{1}^T)\underset{\sim}{x}$$

?

$$= \underset{\sim}{x}^T\underbrace{(I-\tfrac{1}{n}\underset{\sim}{1}\underset{\sim}{1}^T)}_{\text{projection/idempotent matrix}}\underset{\sim}{x}$$

Suppose $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \cdots, \underset{\sim}{X}_n$ are n - $(p\times 1)$ vectors and

$$\underset{n\times p}{X} = \begin{bmatrix} \underset{\sim}{X}_1^T \\ \vdots \\ \underset{\sim}{X}_n^T \end{bmatrix}$$

$\underset{\sim}{X}_i^T$ is $(1\times p)$ vector, where $1\le i\le n$.

$p\times n$

Then $$\sum_{j=1}^{n}\underset{(p\times 1)}{(\underset{\sim}{X}_j-\bar{\underset{\sim}{X}})}\underset{(1\times p)}{(\underset{\sim}{X}_j-\bar{\underset{\sim}{X}})^T} = \boxed{\underset{p\times n}{X^T}\underbrace{(I-\tfrac{1}{n}\underset{\sim}{1}\underset{\sim}{1}^T)}_{n\times n}\underbrace{X}_{n\times p}} \quad - (*)$$

$$\underset{p\times n}{X^T} = [\underset{(p\times 1)}{\underset{\sim}{X}_1}|\cdots|\underset{\sim}{X}_n](I-\tfrac{1}{n}\underset{\sim}{1}\underset{\sim}{1}^T)\begin{bmatrix}\underset{\sim}{X}_1^T \\ \vdots \\ \underset{\sim}{X}_n^T\end{bmatrix}$$

$$(u_1,\cdots,u_n)(I-\tfrac{1}{n}\underset{\sim}{1}\underset{\sim}{1}^T)\begin{pmatrix}u_1\\ \vdots \\ u_n\end{pmatrix}$$

$$\sum_i\sum_j u_iu_j(I-\tfrac{1}{n}\underset{\sim}{1}\underset{\sim}{1}^T)_{ij}$$

Try to use this to show $(*)$.

Multivariate Normal Dist.

$\underset{\sim}{X} \sim N_p(\underset{p\times 1}{\underset{\sim}{\mu}}, \underset{p\times p}{\Sigma})$ ($\underset{\sim}{X}$ has p-variate normal dist. with mean vector $\underset{\sim}{\mu}$ and variance-covariance matrix Σ)

If the joint density of $(x_1,\ldots,x_p)$ is

$$f_{\underset{\sim}{\mu},\Sigma}(\underset{\sim}{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\tfrac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})^T\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right]$$

for all $\underset{\sim}{x}\in\mathbb{R}^p$.

Math A7800 9/7/17, Thur

⊙ Eigenvalues, eigenvectors, spectral theorem

- For a square matrix A, a number (real/complex) λ is called an eigenvalue if $\exists$ $(n\times 1)$ vector $\underset{\sim}{x}$ s.t. $A\underset{\sim}{x} = \lambda \underset{\sim}{x}$.
- A vector $\underset{\sim}{x}$ is called an eigenvector of A if there is a number λ s.t. $A\underset{\sim}{x} = \lambda\underset{\sim}{x}$.
- If $\underset{\sim}{x}$ is an eigen vector, then so is $c\underset{\sim}{x}$ for any constant c.

$A = A^T$ (A is symmetric)

If λ and μ are two eigenvalues of $A^{\swarrow}$, $\lambda \neq \mu$, and if $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are **associated eigenvectors**, then $\underset{\sim}{x}\cdot\underset{\sim}{y} = 0$.

Pf $\quad \underset{\sim}{y}^T A \underset{\sim}{x} = \underset{\sim}{y}^T(\lambda\underset{\sim}{x}) = \lambda(\underset{\sim}{y}^T\underset{\sim}{x})$ $\quad (\underset{\sim}{y}^T A\underset{\sim}{x})^T = \underset{\sim}{x}^T A \underset{\sim}{y}$

$\underset{\sim}{x}^T A \underset{\sim}{y} = \underset{\sim}{x}^T(\mu\underset{\sim}{y}) = \mu(\underset{\sim}{x}^T\underset{\sim}{y})$

$\underset{\sim}{y}^T A\underset{\sim}{x} = \underset{\sim}{x}^T A\underset{\sim}{y}$ for A is symmetric.

Thus, $\lambda\underset{\sim}{y}^T\underset{\sim}{x} = \mu\underset{\sim}{x}^T\underset{\sim}{y}$.

If $\underset{\sim}{x}^T\underset{\sim}{y} \neq 0$, then $\lambda = \mu$, which is a contradiction.

Thus, $\underset{\sim}{x}^T\underset{\sim}{y} = 0$

For a matrix A, the eigenspace corresponding to λ is

$$\mathrm{Eig}_A(\lambda) = \{ \underset{\sim}{x} : A\underset{\sim}{x} = \lambda \underset{\sim}{x} \}$$ "vector space"

$\mathrm{Eig}_A(\lambda) \neq \{\underset{\sim}{0}\}$, when λ is an eigenvalue of A.

Quadratic Forms

For a square matrix A (n×n symmetric), the function $\underset{\sim}{x}^T A \underset{\sim}{x}$ is called **quadratic forms** corresponding to A.

$$\underset{\sim}{x}^T A \underset{\sim}{x} = \underset{1\times n}{(x_1, \dots, x_n)} \underset{n\times n}{\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}} \underset{n\times 1}{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}$$

$A = (a_{ij})$, $1 \le i, j \le n$

$\underset{\sim}{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij} = Q(\underset{\sim}{x})$$

If A is not symmetric, then symmetrize A by taking

$B = \dfrac{A + A^T}{2}$. Then B is symmetric, and

$$\underset{\sim}{x}^T A \underset{\sim}{x} = \underset{\sim}{x}^T B \underset{\sim}{x} \quad \text{for all } \underset{\sim}{x}.$$

e.g. $Q(\underset{\sim}{x}) = [x_1 \ x_2] \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + 2x_1x_2 + x_2^2$
$= (x_1 + x_2)^2$

$$A = \begin{array}{c|c} & \begin{matrix} & i & j & \end{matrix} \\ \begin{matrix} i \\ j \end{matrix} & \begin{bmatrix} & & & \\ & & \cdot\, a_{ij} & \\ & \cdot\, a_{ji} & & \end{bmatrix} \end{array}$$

In $\underset{\sim}{x}^T A \underset{\sim}{x}$, the coefficient of $x_i x_j$ is $(a_{ij} + a_{ji})$.

$$B = \begin{array}{c|c} & \begin{matrix} & i & j & \end{matrix} \\ \begin{matrix} i \\ j \end{matrix} & \begin{bmatrix} & & & \\ & & \cdot\, \frac{a_{ij}+a_{ji}}{2} & \\ & \cdot\, \frac{a_{ij}+a_{ji}}{2} & & \end{bmatrix} \end{array}$$

So, coefficient of $x_i x_j$ in $\underset{\sim}{x}^T B \underset{\sim}{x}$ is $(a_{ij} + a_{ji})$

Since all coefficients are same in $\underset{\sim}{x}^T A \underset{\sim}{x}$ and $\underset{\sim}{x}^T B \underset{\sim}{x}$, we must have $\underset{\sim}{x}^T A \underset{\sim}{x} = \underset{\sim}{x}^T B \underset{\sim}{x}$

Orthogonal matrix

A square matrix U is called orthogonal if $UU^T = U^T U = I$. So $U^T = U^{-1}$.

If $U = [\underset{\sim}{u}_1 | \cdots | \underset{\sim}{u}_n]$, then $\underset{\sim}{u}_i \cdot \underset{\sim}{u}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$

Also, if

$U = \begin{bmatrix} \underset{\sim}{v}_1^T \\ \vdots \\ \underset{\sim}{v}_n^T \end{bmatrix}$, then $\underset{\sim}{v}_i \cdot \underset{\sim}{v}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$

- **Spectral theorem**

If A is a real symmetric matrix, then A can be written as $\boxed{A = UDU^T}$ for some orthogonal matrix U (real) and real diagonal matrix D.

Moreover, the diagonal entries of D are the eigenvalues of A and the columns of U are the associated eigenvectors.

If $\underset{\sim}{u} = [\underset{\sim}{u}_1 | \cdots | \underset{\sim}{u}_n]$ and $D = \begin{pmatrix} \lambda_{11} & & 0 \\ & \ddots & \\ 0 & & \lambda_{nn} \end{pmatrix}$,

Then $A = \sum_{i=1}^{n} \lambda_i \underset{\sim}{u}_i \cdot \underset{\sim}{u}_i^T$

Def A is symmetric square matrix.

① A is called non-negative definite (nnd) if $\underset{\sim}{x}^T A \underset{\sim}{x} \geq 0$ for all $\underset{\sim}{x}$.

② A is called positive definite (pd) if $\underset{\sim}{x}^T A \underset{\sim}{x} > 0$ for all $\underset{\sim}{x} \neq \underset{\sim}{0}$.

③ Similarly, define non-positive definite and negative definite.

Fact **A is nnd iff all eigenvalues of A are non-negative**

Pf Recall $A = \sum_{i=1}^{n} \lambda_i \underset{\sim}{u}_i \cdot \underset{\sim}{u}_i^T$

If all eigenvalues are non-negative ($\lambda_i \geq 0$), then

$$\underset{\sim}{x}^T A \underset{\sim}{x} = \sum_{i=1}^{n} \underset{\sim}{x}^T (\lambda_i \underset{\sim}{u}_i \cdot \underset{\sim}{u}_i^T) \underset{\sim}{x}$$

$$= \sum_{i=1}^{n} \lambda_i (\underset{\sim}{x}^T \underset{\sim}{u}_i)^2 \geq 0, \text{ for all } \underset{\sim}{x}.$$

So, A is nnd.

Suppose $\lambda_1 < 0$. Then $\underset{\sim}{u}_1^T A \underset{\sim}{u}_1 = \lambda_1 < 0$.

So, A is not nnd. $\lambda_i \geq 0 \Rightarrow A$ is nnd.

Fact A is pd iff all its eigenvalues are strictly positive.

Suppose A has spectral decomposition $A = UDU^T$.

Then $A^2 = UD^2U^T$ and $A^k = UD^kU^T$.

If $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$, then $D^k = \begin{pmatrix} \lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \lambda_n^k \end{pmatrix}$

So, A^k has eigenvalues $\lambda_1^k, \dots, \lambda_n^k$ with same eigenvectors.

If A is pd and $A = UDU^T$.

Then $\underbrace{A^{-1}}_{(UDU^T)^{-1}} = UD^{-1}U^T$, $D^{-1} = \begin{pmatrix} 1/\lambda_1 & & 0 \\ & \ddots & \\ 0 & & 1/\lambda_n \end{pmatrix}$

So, $A^{-k} = UD^{-k}U^T$.

If $B = UD^{1/2}U^T$, where $D^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{pmatrix}$

Then $B^2 = A$.

So, B can be called $A^{1/2}$.

$A^{-1/2} = UD^{-1/2}U^T$, ...

Suppose $\underset{\sim}{X}$ is a random vector with $var(\underset{\sim}{X}) = \Sigma$.

Then Σ is always nnd. "non-negative definite"

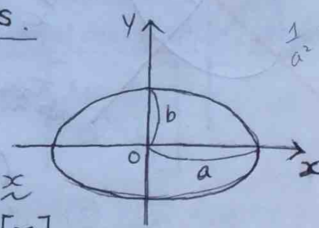
For any vector $\underset{\sim}{a}$,

$\underset{\sim}{a}^T \Sigma \underset{\sim}{a} = var(\underset{\sim}{a}^T \underset{\sim}{X}) \geq 0$ for all $\underset{\sim}{a}$.

$var(\underset{\sim}{a}^T \underset{\sim}{x}) > 0$ unless $\underset{\sim}{a}^T \underset{\sim}{X}$ = constant. ∎

Diagram of ellipsoids.

$$\boxed{\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1}$$



$$\Downarrow$$

$$\underbrace{[x \ \ y]}_{\underset{\sim}{x}^T} \underbrace{\begin{bmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\underset{\sim}{x}} = 1.$$

Generalize

$\underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ Look at $\underset{\sim}{x}^T A \underset{\sim}{x} = 1$, where A is pd.

Invoke the spectral decomposition of A,

$A = UDU^T$, where D is diagonal and U is othogonal.

We have $\underset{\sim}{x}^T (UDU^T) \underset{\sim}{x} = 1$.

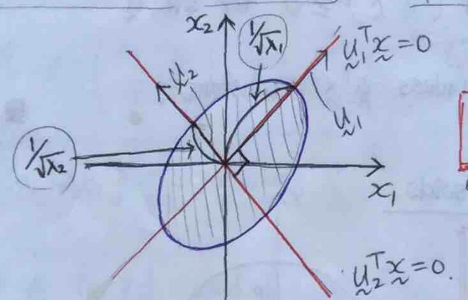
$\Rightarrow \underset{\sim}{y}^T D \underset{\sim}{y} = 1$, where $\underset{\sim}{y} = U^T \underset{\sim}{x}$.

The axes of the ellipse are given by the equations

$\boxed{y_1 = 0}$ and $\boxed{y_2 = 0}$.

If

$\underset{\sim}{U} = [\underset{\sim}{u}_1 \mid \underset{\sim}{u}_2] \iff \boxed{\underset{\sim}{u}_1^T \underset{\sim}{x} = 0} \iff \boxed{\underset{\sim}{u}_2^T \underset{\sim}{x} = 0}$



$$\boxed{\underset{\sim}{x}^T A \underset{\sim}{x} \le 1}$$

"shaded region"

$$(\underset{\sim}{x}-\underset{\sim}{\mu})^T A (\underset{\sim}{x}-\underset{\sim}{\mu}) = 1$$

$$f_{\underset{\sim}{\mu},\Sigma}(\underset{\sim}{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right]$$

Last class: We discussed about eigenvalues and eigenvectors of symmetric matrices, also nnd, pd matrices.

Chapter 4

Finally, we looked at the plots of

$f(\underset{\sim}{x}) = (\underset{\sim}{x}-\underset{\sim}{\mu})' A (\underset{\sim}{x}-\underset{\sim}{\mu})$ for pd matrix A.

The plot for $f(\underset{\sim}{x}) = 1$ is an ellipsoid with center $\underset{\sim}{\mu}$, axes given by ① and half axes length given by ②.

If $\boxed{A = UDU'}$ is the spectral decomposition, where

$$U = [\underset{\sim}{u}_1 | \cdots | \underset{\sim}{u}_p], \quad D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{bmatrix}.$$

$A' = \begin{pmatrix} 5 & 3 \\ 3 & 4 \end{pmatrix}$

the straight line

① $\underset{\sim}{u}_i' \underset{\sim}{x} = 0, \ 1 \le i \le p$. ② $1/\sqrt{\lambda_i}, \ 1 \le i \le p$.

Ex $A = \begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix}$ Find the spectral decomposition.

"solution for quadratic eqn."

$$\boxed{\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}}$$

$$\begin{vmatrix} 5-\lambda & 3 \\ 3 & 4-\lambda \end{vmatrix} = (5-\lambda)(4-\lambda) - 9 = 0$$
$$= \lambda^2 - 4\lambda - 5\lambda + 20 - 9 = 0$$
$$= \lambda^2 - 9\lambda + 11 = 0$$

$b^2 \pm 4ac$

81
44

$$\lambda_1 = \frac{9 + \sqrt{81-44}}{2}, \quad \lambda_2 = \frac{9 - \sqrt{81-44}}{2}$$
$$= \frac{9+\sqrt{37}}{2} \qquad = \frac{9-\sqrt{37}}{2}$$

5 3
3 4

$$A-\lambda_1 I = \begin{pmatrix} 5-\lambda_1 & 3 \\ 3 & 4-\lambda_1 \end{pmatrix} \underset{\sim}{v} = 0, \quad \underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\begin{pmatrix} (5-\lambda_1)x_1 + 3x_2 = 0. \\ 3x_1 + (4-\lambda_1)x_2 = 0 \end{pmatrix}, \quad \begin{pmatrix} 3 \\ 4-\lambda_1 \end{pmatrix} = \frac{3}{5-\lambda_1}\begin{pmatrix} 5-\lambda_1 \\ 3 \end{pmatrix}.$$

So, $\underset{\sim}{v}_1 = \begin{pmatrix} -\frac{3}{5-\lambda_1} \\ 1 \end{pmatrix}$ is an eigenvector for λ_1

normalize

$\underset{\sim}{u}_1 = \begin{pmatrix} v_1^{(1)}/\|v_1\| \\ v_2^{(1)}/\|v_2\| \end{pmatrix}$, similarly, $\underset{\sim}{u}_2 = \begin{pmatrix} v_1^{(2)}/\|v_1\| \\ v_2^{(2)}/\|v_2\| \end{pmatrix}$

normal eigenvector

Recall A random vector

$$\underset{\sim}{X}_{p\times 1} \sim N_p(\underset{\sim}{\mu}, \Sigma)$$

$\underset{\sim}{x}' = (x_1, \dots, x_p)$, where $x_i \sim N(\mu, \sigma^2)$, $1 \le i \le p$.

(p-variate normal distribution with parameters $\underset{\sim}{\mu}_{p\times 1}$ and pd matrix $\Sigma_{p\times p}$)

if the joint pdf of the components of $\underset{\sim}{X}$ is given by

★ $f_{\underset{\sim}{\mu},\Sigma}(\underset{\sim}{x}) = f_{\underset{\sim}{\mu},\Sigma}(x_1, \dots, x_p)$ ☆memo

$$= \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right]$$

, for all $\underset{\sim}{x} \in \mathbb{R}^p$.

$\underset{\sim}{\mu}$ vector Σ

Mean and Covariance matrix of $\underset{\sim}{x}$

Define the linear transformation

$$\underset{\sim}{Y} = \Sigma^{-1/2}(\underset{\sim}{X} - \underset{\sim}{\mu}) = \Sigma^{-1/2}\underset{\sim}{X} - \Sigma^{-1/2}\underset{\sim}{\mu}$$

Find joint pdf of $\underset{\sim}{Y}$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

pdf of $N(\mu, \sigma^2)$

Deviation

(X_1, X_2) has joint pdf $f(x_1, x_2)$.

Suppose $(X_1, X_2) \mapsto (Y_1, Y_2)$ is injective and surjective transformation, where $Y_1 = h_1(x_1, x_2)$

$Y_2 = h_2(x_1, x_2)$

What is the joint pdf of $Y_1 Y_2$?

Since the transformation is one-to-one, onto, we can write x_1 and x_2 as functions of y_1, y_2, say

$X_1 = g_1(y_1, y_2)$ and $X_2 = g_2(y_1, y_2)$.

Then the joint pdf of (Y_1, Y_2)

$$L(y_1, y_2) = f(g_1(y_1, y_2), g_2(y_1, y_2)) \cdot J, \text{ where}$$

$$J = \left| \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} \right| = \left| \frac{\partial \underset{\sim}{x}}{\partial \underset{\sim}{y}} \right| = \frac{1}{\left| \frac{\partial \underset{\sim}{y}}{\partial \underset{\sim}{x}} \right|}$$

absoulte value.

Ex suppose (X_1, X_2) has joint pdf $f(x_1, x_2) = e^{-x/2}$, where $0 \le x_1 \le x_2$.

Obtain the joint pdf of $(\underset{Y_1}{X_1}, \underbrace{X_2 - X_1}_{Y_2})$ $\boxed{f_{Y_1, Y_2}(y_1, y_2)}$

$$\begin{cases} Y_1 = h_1(x_1, x_2) = x_1 \\ Y_2 = h_2(x_1, x_2) = x_2 - x_1 \end{cases}$$

$$\left.\begin{aligned} Y_2 &= x_2 - x_1 \\ &= x_2 - y_1 \end{aligned}\right\} \Rightarrow x_2 = y_1 + y_2$$

$$\begin{cases} x_1 = g_1(y_1, y_2) = y_1 \\ x_2 = g_2(y_1, y_2) = y_1 + y_2 \end{cases}$$

$$\left.\begin{aligned} y_2 &= x_2 - x_1 \\ &= x_2 - y_1 \end{aligned}\right\} \Rightarrow x_2 = y_2 + y_1$$

So, $$J = \left| \begin{vmatrix} \frac{\partial g_1}{\partial y_1} & \frac{\partial g_1}{\partial y_2} \\ \frac{\partial g_2}{\partial y_1} & \frac{\partial g_2}{\partial y_2} \end{vmatrix} \right| = \left| \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} \right| = 1.$$

The joint pdf of (Y_1, Y_2) is

$$L(y_1, y_2) = f_{X_1 X_2}(\underset{x_1 = y_1}{g_1(y_1, y_2)}, \underset{x_2 = y_1 + y_2}{g_2(y_1, y_2)}) \cdot J$$

$$f_{Y_1, Y_2}(y_1, y_2) = e^{-(y_1 + y_2)} \cdot J = e^{-(y_1 + y_2)}, \quad y_1, y_2 \ge 0.$$

Same applies for p-variables.

If $(x_1,\dots,x_p)$ has joint pdf $f(x_1,\dots,x_p)$ and $(x_1,\dots,x_p) \mapsto (y_1,\dots,y_p)$ is an 1-1 and onto transformation, where $y_i = h_i(x_1,\dots,x_p)$, then we obtain the joint pdf of $(y_1,\dots,y_p)$ as follows.

Step1 Express x_i in terms of $y_1,\dots,y_p$, say

$x_i = g_i(y_1,\dots,y_p)$, for $1 \le i \le p$.

Step2 Obtain

$$J = \underbrace{\Bigg|\overbrace{\left|\frac{\partial g}{\partial y}\right|}^{\text{determinant}}\Bigg|}_{\text{absolute value}} = \left|\,\begin{vmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_p}{\partial y_p} \\ \vdots & \ddots & \\ \frac{\partial g_p}{\partial y_1} & \cdots & \frac{\partial g_p}{\partial y_p} \end{vmatrix}\,\right|$$

$$\left|\left|\frac{\partial \underset{\sim}{x}}{\partial \underset{\sim}{y}}\right|\right| = \left|\frac{1}{\left|\frac{\partial h}{\partial x}\right|}\right|$$

Step3

$$\underset{\displaystyle f_{y_1,\dots,y_p}(y_1,\dots,y_p)}{L(y_1,\dots,y_p)} = \underbrace{f(\overset{x_1}{g_1(y_1,\dots,y_p)},\dots,\overset{x_p}{g_p(y_1,\dots,y_p)})}_{f_{x_1,\dots,x_p}(x_1,\dots,x_p)} \cdot J$$

$$f_{\mu,\Sigma}(\underset{\sim}{x}) = f_{\mu,\Sigma}(x_1,\dots,x_p)$$
$$= \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right] \ \forall \underset{\sim}{x} \in \mathbb{R}^p$$

Define the linear transformation.

$$\underset{\sim}{y} = \Sigma^{-1/2}(\underset{\sim}{x}-\underset{\sim}{\mu}) = \Sigma^{-1/2}\underset{\sim}{x} - \Sigma^{-1/2}\underset{\sim}{\mu}$$

$\underset{\sim}{y} = \Sigma^{-1/2}\underset{\sim}{x} - \Sigma^{-1/2}\underset{\sim}{\mu}$

$\underset{\sim}{y} + \Sigma^{-1/2}\underset{\sim}{\mu} = \Sigma^{-1/2}\underset{\sim}{x}$

$\underset{\sim}{x} = \Sigma^{1/2}\underset{\sim}{y} + \underset{\sim}{\mu}$

Find joint pdf of $\underset{\sim}{y}$.

$$\left(\begin{array}{l} \Sigma^{-1/2}\underset{\sim}{x} = \Sigma^{-1/2}\underset{\sim}{\mu} + \underset{\sim}{y} \\ \underset{\sim}{x} = \underset{\sim}{\mu} + \Sigma^{1/2}\underset{\sim}{y} \end{array}\right)$$

Step1 $\underset{\sim}{x} = \underset{\sim}{\mu} + \Sigma^{1/2}\underset{\sim}{y} \sim (1)$

Step2 $\dfrac{\partial \underset{\sim}{x}}{\partial \underset{\sim}{y}} = \Sigma^{1/2}$ so, $J = \Big|\underbrace{\left|\Sigma^{1/2}\right|}_{\left|\frac{\partial x}{\partial y}\right|}\Big| = |\Sigma|^{1/2}$

(as $|A|$ is product of the eigen values of A and eigenvalues of $\Sigma^{1/2}$ are square root of those of Σ.)

Step3

$$L(\underset{\sim}{y}) = f_{\mu,\Sigma}(\overset{\underset{\sim}{x}}{\underset{\sim}{\mu} + \Sigma^{1/2}\underset{\sim}{y}}) \cdot J. \quad \exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right]$$

$$= \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\underset{\sim}{y}'\Sigma^{1/2}\Sigma^{-1}\Sigma^{1/2}\underset{\sim}{y}\right] \cdot |\Sigma|^{1/2}$$

$$= \frac{1}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}\underset{\sim}{y}'\underset{\sim}{y}\right]$$

$$= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} e^{-y_i^2/2} - e\ N(0,1).$$

☆ From (1),

$E(\underset{\sim}{x}) = \underset{\sim}{\mu}$,

$Var(\underset{\sim}{x}) = \Sigma^{1/2}\underset{I}{Var(\underset{\sim}{y})}\Sigma^{1/2} = \Sigma$

Since the joint pdf of $(y_1,\dots,y_p)$ is product of $N(0,1)$ pdf's, we conclude that $y_1,\dots,y_p$ are iid $N(0,1)$.

Then, ☆ $E[\underset{\sim}{y}] = \underset{\sim}{0}$, $Var(\underset{\sim}{y}) = \underset{\sim}{I}$ → identity matrix.

From (1), $\mathbb{E}(\underset{\sim}{X}) = \underset{\sim}{\mu}$, $Var(\underset{\sim}{X}) = \Sigma^{1/2} \underbrace{var(\underset{\sim}{Y})}_{I} \Sigma^{1/2}$

$\underset{\sim}{X} = \underset{\sim}{\mu} + \Sigma^{1/2}\underset{\sim}{Y}$

$Var(\underset{\sim}{X}) = \mathbb{E}(\underset{\sim}{X}^2) - (\mathbb{E}(\underset{\sim}{X}))^2$

$\mathbb{E}(Y^2) = \mu^2 - \mathbb{E}(\Sigma^{1/2}Y^2)$

$$= \Sigma$$

EX $p = 2$, $\underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$, where $\sigma_{11} = Var(X_1)$
$\sigma_{22} = Var(X_2)$
$\sigma_{12} = cov(X_1, X_2)$

If $\rho = \text{corr. coeff}(X_1, X_2)$, then

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}, \quad |\Sigma| = \sigma_{11}\sigma_{22} - \sigma_{12}^2$$

$$\boxed{\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}\sigma_{22}}}$$

$$= \boxed{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)}$$

$$\sigma_{11}\sigma_{22}\left(1 - \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}\right) = \sigma_{11}\sigma_{22}\left(1 - \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}\right) = \rho_{12}$$

$$\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22}(1-\rho^2)} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

$$f_{\mu,\Sigma}(x_1, x_2) = \frac{1}{(2\pi)^{p/2}\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)}} \exp\left[-\frac{1}{2\sigma_{11}\sigma_{22}(1-\rho^2)}\left\{\sigma_{22}(x_1-\mu_1)^2 + \sigma_{11}(x_2-\mu_2)^2 - 2\sigma_{12}(x_1-\mu_1)(x_2-\mu_2)\right\}\right]$$

$$\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right]$$

(by computing $(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})$) — ⊛

$$= \frac{1}{(2\pi)^{p/2}\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right\}\right]$$

$$\boxed{(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})} \text{ — ⊛}$$

$$= [x_1-\mu_1, x_2-\mu_2] \frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)} \begin{bmatrix} \sigma_{22} & \overbrace{-\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}^{\sigma_{12}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

$$= \frac{\sigma_{22}(x_1-\mu_1)^2 + \sigma_{11}(x_2-\mu_2)^2 - 2\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1-\mu_1)(x_2-\mu_2)}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}$$

$$= \frac{1}{1-\rho_{12}^2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]$$

$$corr(X_1, X_2) = \rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$$

so, $\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$.

$$|\Sigma| = \sigma_{11}\sigma_{22} - (\underbrace{\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}_{\sigma_{12}})^2 = \sigma_{11}\sigma_{22} - \rho_{12}^2\sigma_{11}\sigma_{22} = \sigma_{11}\sigma_{22}(1-\rho_{12}^2)$$

$$f_{\mu,\Sigma}(\underset{\sim}{x}) = f_{\mu,\Sigma}(x_1, \ldots, x_p)$$

$$= \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right], \forall \underset{\sim}{x} \in \mathbb{R}^p$$

Alternative definition of Multivariate normal distribution.

A random vector $\underset{\sim}{X}$ has $N_p(\underset{\sim}{\mu}, \Sigma)$ distribution iff every linear combination $\boxed{\underset{\sim}{\ell}'\underset{\sim}{X} = \sum_{i=1}^{p} \ell_i X_i}$ has univariate normal distribution $N(\underset{\sim}{\ell}'\underset{\sim}{\mu}, \underset{\sim}{\ell}'\Sigma\underset{\sim}{\ell})$

Consequence suppose $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$

For any $m \times p$ matrix of constants A,

$$\underset{(m\times 1)}{\underset{\sim}{Y}} = \underset{(m\times p)}{A}\,\underset{(p\times 1)}{\underset{\sim}{X}} \sim N_m(A\underset{\sim}{\mu}, A\Sigma A')$$

Why? Take any linear combination

$$\underset{\sim}{\ell}'\underset{\sim}{Y} = \underset{\sim}{\ell}'A\underset{\sim}{X} = \underset{\sim}{b}'\underset{\sim}{X} \text{ for } A'\underset{\sim}{\ell} = \underset{\sim}{b}$$

So, $\underset{\sim}{\ell}'\underset{\sim}{Y} = \underset{\sim}{b}'\underset{\sim}{X} \sim N(\underset{\sim}{b}'\underset{\sim}{\mu}, \underset{\sim}{b}'\Sigma\underset{\sim}{b})$

$$\sim N\Big(\underbrace{\underset{\sim}{\ell}'(A\underset{\sim}{\mu})}_{\underset{\sim}{\ell}'\underset{\sim}{\nu}}, \underbrace{\underset{\sim}{\ell}'(A\Sigma A')\underset{\sim}{\ell}}_{\underset{\sim}{\ell}'\Gamma\underset{\sim}{\ell}}\Big)$$

This is true for any $\underset{\sim}{\ell}$.

From the previous assertion,

$$\underset{\sim}{Y} \sim N_m(\underset{\sim}{\nu}, \Gamma) = N_m(A\underset{\sim}{\mu}, A\Sigma A')$$

If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then $A\underset{\sim}{X} \sim N_p(A\underset{\sim}{\mu}, A\Sigma A')$.

Consequence

Any subset of $(X_1, \ldots, X_p)$ has multivariate normal distribution.

Ex $\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \right)$, then

$$\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{pmatrix} \right).$$

Why? $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$

$$X_2 \sim N_1(\mu_2, \sigma_{22}).$$

- Important properties:

In case of multivariate normal,

independence $\Longleftrightarrow$ uncorrelatedness.

Not true in general

If (X_i, X_j) are uncorrelated, then they are independent.

- **Range of a sample**

$R = Y_n - Y_1 = \max\{X_1, \dots, X_n\} - \min\{X_1, \dots, X_n\}$
is called the range of the sample $X_1, \dots, X_n$.

Ex $X_1, X_2, X_3 \overset{iid}{\sim} U(0,1)$. Find pdf of $R: Y_3 - Y_1$.

$$g_{13}(s,t) = \begin{cases} 6(t-s) & \text{for } 0 \le s \le t \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$(Y_1, Y_3) \underset{\text{"1-1"}}{\longmapsto} (\underbrace{Y_3 - Y_1}_{Z_1}, \underbrace{Y_3}_{Z_2}).$$

$$\begin{cases} Y_1 = Z_2 - Z_1 \\ Y_3 = Z_2 \end{cases}, \quad f_{Z_1 Z_2}(z_1, z_2) = f_{Y_1 Y_3}(z_2 - z_1, z_2)\,|J| = 6 z_1,$$

for $0 \le z_1 \le z_2 \le 1$.

$$J = \begin{vmatrix} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_1}{\partial z_2} \\ \frac{\partial y_2}{\partial z_1} & \frac{\partial y_2}{\partial z_2} \end{vmatrix} = -1.$$

$$f_{Z_1}(z_1) = \int f_{Z_1 Z_2}(z_1, z_2)\, dz_2$$

$$= \int_{z_1}^{1} 6 z_1 \, dz_2 = 6 z_1 (1 - z_1), \text{ for } 0 \le z_1 \le 1.$$

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \end{pmatrix}$$
A

Ex Suppose $\underset{\sim}{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \sim N_3(\underset{\sim}{\mu}, \Sigma)$, where

$$\underset{\sim}{\mu} = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{array}{c} \sigma_{11} \qquad \sigma_{13} \\ \begin{pmatrix} 3 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \\ \sigma_{31} \qquad \sigma_{33} \end{array}$$

Find the "distribution" of $\underset{\sim}{Y} = \begin{pmatrix} X_1 + X_3 \\ X_1 - X_3 \end{pmatrix}$

$$\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} \sim N_2\left(\begin{pmatrix} -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} \right). \text{ Then}$$

$$\begin{pmatrix} x_1 + x_3 \\ x_1 - x_3 \end{pmatrix} = \underset{A}{\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}} \underset{\underset{\sim}{x}}{\begin{pmatrix} X_1 \\ X_3 \end{pmatrix}}$$

$$\underset{\sim}{Y} \sim N_2\Big(\underset{\begin{pmatrix} 1 \\ -3 \end{pmatrix}}{A\underset{\sim}{\mu}}, \overbrace{\underset{\begin{pmatrix} 6 & 2 \\ 2 & 2 \end{pmatrix}}{A\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}A'}}^{A\Sigma A'}\Big)$$

$$\begin{pmatrix} 3 & 1 & -1 \\ -1 & 0 & 2 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\Sigma = \left(\begin{array}{cc|c} 3 & 1 & -1 \\ 1 & 1 & 0 \\ \hline -1 & 0 & 2 \end{array}\right)$$

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 3+1 & 2 \\ 2 & 0 \end{pmatrix} \quad \begin{pmatrix} 4 & 2 \\ 2 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$\begin{matrix} 6 & 2 \\ 2 & 2 \end{matrix}$$

$$\underset{\sim}{Z} = \begin{pmatrix} x_1 \\ 2x_2 + x_3 \end{pmatrix} = \underset{A}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \end{bmatrix}} \underset{\underset{\sim}{x}}{\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}} \sim N_2\left(A\begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}, A\Sigma A'\right)$$

Independence $\longleftrightarrow$ uncorrelated

Note that for any two random variables U, V,

if U and V are independent, then $\text{cov}(U,V)=0$.

So, U and V are uncorrelated.

But, the converse is not true.

(if $\text{cov}(U,V)=0$, then U, V are independent.)

Ex Suppose U is a random variable.

$$\mathbb{P}(U=1)=\mathbb{P}(U=-1)=\frac{1}{2}, \quad V=U^2$$

$$\text{cov}(U,V)=\mathbb{E}(UV)-\mathbb{E}(U)\mathbb{E}(V)$$

$\mathbb{E}(U^3) \neq \mathbb{E}(U)\mathbb{E}(U^2)$

$$=\mathbb{E}(U^3)-\mathbb{E}(U)\mathbb{E}(V)=0.$$

$U^3=U$. So, $\mathbb{E}(U)=\mathbb{E}(U^3)=0$.

So, U, V are uncorrelated. But they are not independent.

In case of joint normal distribution, the converse is true.

Result If $\begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{pmatrix} = \underset{p\times 1}{\underset{\sim}{X}} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, where $\underset{\sim}{X}_1$: $q\times 1$, $q<p$, $\underset{\sim}{\mu}$: $p\times 1$, $\underset{\sim}{X}_2$: $(p-q)\times 1$

$$\Sigma = \left(\begin{array}{c:c} \underset{q\times q}{\Sigma_{11}} & \underset{q\times(p-q)}{O} \\ \hdashline \underset{(p-q)\times q}{O} & \underset{(p-q)\times(p-q)}{\Sigma_{22}} \end{array}\right) \rightarrow \Sigma_{12}=\underset{\sim}{O},$$

then $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are independent.

In other word, $\text{cov}(\underset{\sim}{X}_1, \underset{\sim}{X}_2)=O \Rightarrow \underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are independent.

proof For $\Sigma=\begin{pmatrix} \Sigma_{11} & O \\ O & \Sigma_{22} \end{pmatrix}$, $|\Sigma|=|\Sigma_{11}|\cdot|\Sigma_{22}|$

$$\Sigma^{-1}=\begin{pmatrix} \Sigma_{11}^{-1} & O \\ O & \Sigma_{22}^{-1} \end{pmatrix} \qquad \underset{\sim}{x}=\begin{pmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{pmatrix}$$

$$f_{\underset{\sim}{x}_1,\underset{\sim}{x}_2}(\underset{\sim}{x}_1,\underset{\sim}{x}_2)=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\right]$$

$$=\frac{1}{(2\pi)^{p/2}(|\Sigma_{11}|\cdot|\Sigma_{22}|)^{1/2}}\exp\left[-\frac{1}{2}\underbrace{\left((\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)',(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\right)\begin{pmatrix} \Sigma_{11}^{-1} & O \\ O & \Sigma_{22}^{-1} \end{pmatrix}\begin{pmatrix} \underset{\sim}{x}_1-\underset{\sim}{\mu}_1 \\ \underset{\sim}{x}_2-\underset{\sim}{\mu}_2 \end{pmatrix}}_{*}\right]$$

$\Rightarrow$

$$*\; -\left((\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)', (\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\right)\begin{pmatrix}\Sigma_{11}^{-1} & 0\\ 0 & \Sigma_{22}^{-1}\end{pmatrix}\begin{pmatrix}\underset{\sim}{x}_1-\mu_1\\ \underset{\sim}{x}_2-\mu_2\end{pmatrix}$$

$$= (\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)'\Sigma_{11}^{-1}(\underset{\sim}{x}_1-\underset{\sim}{\mu}_1) + (\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Sigma_{22}^{-1}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)$$

return,

$$= \frac{1}{(2\pi)^{q/2}|\Sigma_{11}|^{1/2}}\exp\left[-\frac{1}{2}(\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)'\Sigma_{11}^{-1}(\underset{\sim}{x}_1-\underset{\sim}{\mu}_1)\right]\cdot$$

$$\frac{1}{(2\pi)^{(p-q)/2}|\Sigma_{22}|^{1/2}}\exp\left[-\frac{1}{2}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)'\Sigma_{22}^{-1}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)\right]$$

$$= f_{\underset{\sim}{X}_1}(\underset{\sim}{x}_1)\cdot f_{\underset{\sim}{X}_2}(\underset{\sim}{x}_2).$$

Hence, $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ must be independent.

$$\underset{\sim}{X}_1 \sim N_q(\underset{\sim}{\mu}_1, \Sigma_{11}),\quad \underset{\sim}{X}_2 \sim N_{p-q}(\underset{\sim}{\mu}_2, \Sigma_{22})$$

* Note that in case of joint normal distribution it's sufficient to check zero covariance in order to conclude independence.

EX Suppose $\begin{pmatrix}X_1\\X_2\end{pmatrix} \sim N_2\left(\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix}, \underset{\Sigma}{\begin{pmatrix}\sigma^2 & \rho\sigma^2\\ \rho\sigma^2 & \sigma^2\end{pmatrix}}\right)$.

Then show that $U = X_1+X_2$, $V = X_1-X_2$ are independent.

proof $\begin{pmatrix}U\\V\end{pmatrix} = \underset{A}{\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}}\underset{\underset{\sim}{X}}{\begin{pmatrix}X_1\\X_2\end{pmatrix}} = A\underset{\sim}{X}$ $\quad \begin{pmatrix}X_1+X_2\\X_1-X_2\end{pmatrix} \to \underset{A}{\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}}\underset{\underset{\sim}{X}}{\begin{pmatrix}X_1\\X_2\end{pmatrix}}$

since $\underset{\sim}{X}$ is normally distributed, so is $A\underset{\sim}{x} = \begin{pmatrix}U\\V\end{pmatrix}$.

$$\mathrm{cov}(U,V) = \mathrm{cov}(X_1+X_2, X_1-X_2)$$

$$= \mathrm{cov}(X_1,X_1) - \mathrm{cov}(X_1,X_2) + \mathrm{cov}(X_2,X_1) - \mathrm{cov}(X_2,X_2)$$

$\Sigma_{12} = 0$

$$= \sigma^2 - \rho\sigma^2 + \rho\sigma^2 - \sigma^2 = 0.$$

$$\Sigma = \underset{A}{\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}}\underset{\Sigma}{\begin{pmatrix}\sigma^2 & \rho\sigma^2\\ \rho\sigma^2 & \sigma^2\end{pmatrix}}\underset{A^{-1}}{\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}} = \begin{pmatrix}\sigma^2+\rho\sigma^2+\rho\sigma^2+\sigma^2 & \underset{=0}{\sigma^2-\rho\sigma^2+\rho\sigma^2-\sigma^2}\\ \underset{=0}{\sigma^2+\rho\sigma^2-\rho\sigma^2-\sigma^2} & \sigma^2-\rho\sigma^2-\rho\sigma^2+\sigma^2\end{pmatrix}$$

$$\cancel{\mathrm{cov}(x,y)} = \begin{pmatrix}\cancel{\mathrm{cov}(x,y)} & \\ 1 & -1\end{pmatrix}\begin{pmatrix}\sigma^2+\rho\sigma^2 & \sigma^2-\rho\sigma^2\\ \rho\sigma^2+\sigma^2 & \rho\sigma^2-\sigma^2\end{pmatrix}$$

$$\boxed{\mathrm{cov}(x,y) = \mathbb{E}(xy) - \mathbb{E}(x)\mathbb{E}(y)}$$

Ex If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \sigma^2 I)$ and A is $p\times p$ orthogonal matrix, then $\underset{\sim}{Y} = A\underset{\sim}{X}$ has independent components.

Why? $\underset{\sim}{Y} \sim N_p(A\underset{\sim}{\mu}, \sigma^2 I)$, since A is orthogonal (e.g. $AA' = I$).

So, the component must be independent.

Conditional Distribution

Suppose $\underset{p\times 1}{\underset{\sim}{X}} = \begin{pmatrix} \underset{\sim}{X_1} \\ \underset{\sim}{X_2} \end{pmatrix} \begin{array}{l} q\times 1 \\ (p-q)\times 1 \end{array} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, where

$$\underset{p\times 1}{\underset{\sim}{\mu}} = \begin{pmatrix} \underset{\sim}{\mu_1} \\ \underset{\sim}{\mu_2} \end{pmatrix} \begin{array}{l} q\times 1 \\ (p-q)\times 1 \end{array}, \quad \Sigma = \begin{array}{c} \\ q \\ p-q \end{array} \begin{array}{c} \begin{array}{cc} q & p-q \end{array} \\ \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right) \end{array}.$$

Given $\underset{\sim}{X_2} = \underset{\sim}{x_2}$, the conditional distribution of $\underset{\sim}{X_1}$ is

$$N_q(\underset{\sim}{\mu_1} + \underbrace{\Sigma_{12}}_{q\times(p-q)}\underbrace{\Sigma_{22}^{-1}}_{(p-q)\times(p-q)}(\underset{\sim}{x_2} - \underset{\sim}{\mu_2}), \underbrace{\Sigma_{11}}_{q\times q} - \underbrace{\Sigma_{12}}_{q\times(p-q)}\underbrace{\Sigma_{22}^{-1}}_{(p-q)\times(p-q)}\underbrace{\Sigma_{21}}_{(p-q)\times q})$$

Given $\underset{\sim}{X_1} = \underset{\sim}{x_1}$, then conditional distribution of $\underset{\sim}{X_2}$ is

$$N_{p-q}(\underset{\sim}{\mu_2} + \Sigma_{21}\Sigma_{11}^{-1}(\underset{\sim}{x_1} - \underset{\sim}{\mu_1}), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}).$$

Suppose $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \left(\begin{array}{c|c} \sigma_{11} & \rho\sqrt{\sigma_{11}\sigma_{22}} \\ \hline \rho\sqrt{\sigma_{11}\sigma_{22}} & \sigma_{22} \end{array}\right)\right)$

The conditional dist. of X_1 given $X_2 = 3$ is

$$N\Big(\underbrace{\mu_1 + \frac{\rho\sqrt{\sigma_{11}\sigma_{22}}}{\sigma_{22}}(\underset{X_2=3}{3} - \mu_2)}_{\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X_2} - \underset{\sim}{\mu_2})}, \underbrace{\sigma_{11} - \rho\sqrt{\sigma_{11}\sigma_{22}}\frac{\rho\sqrt{\sigma_{11}\sigma_{22}}}{\sigma_{22}}}_{\substack{\sigma_{11} - \frac{\rho^2\sigma_{11}\cancel{\sigma_{22}}}{\cancel{\sigma_{22}}} \\ = \sigma_{11}(1-\rho^2)}}\Big)$$

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Ex $\begin{pmatrix}X_1\\X_2\\X_3\end{pmatrix} \sim N_3\left(\begin{pmatrix}-1\\0\\2\end{pmatrix}, \begin{pmatrix}3&-1&1\\-1&2&0\\1&0&1\end{pmatrix}\right)$ $\qquad \begin{pmatrix}3&-1&1\\1&0&1\\-1&2&0\end{pmatrix} \to \begin{pmatrix}3&1&-1\\1&1&0\\-1&0&2\end{pmatrix}$

Find conditional distribution of

(a) $\begin{pmatrix}X_1\\X_3\end{pmatrix}$ given $X_2=5$

(b) X_1 given $X_2=5,\ X_3=2$

(a)
$$\begin{pmatrix}X_1\\X_3\\\hline X_2\end{pmatrix} \sim N_3\left(\begin{pmatrix}-1\\2\\\hline 0\end{pmatrix}, \begin{array}{c} \Sigma_{11} \quad \Sigma_{12} \\ \left(\begin{array}{cc|c}3&1&-1\\1&1&0\\\hline -1&0&2\end{array}\right) \\ \Sigma_{21} \quad \Sigma_{22}\end{array}\right)$$

$$\begin{pmatrix}X_1\\X_3\end{pmatrix} \sim N_2\left(\underbrace{\begin{pmatrix}-1\\2\end{pmatrix}+\begin{pmatrix}-1\\0\end{pmatrix}\frac{1}{2}(5-0)}_{\boxed{\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)}},\ \underbrace{\begin{pmatrix}3&1\\1&1\end{pmatrix}-\begin{pmatrix}-1\\0\end{pmatrix}\frac{1}{2}(-1,0)}_{\boxed{\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}}\right)$$

$$= N_2\left(\begin{pmatrix}-7/2\\2\end{pmatrix}, \begin{pmatrix}5/2&1\\1&1\end{pmatrix}\right)$$

(b)
$$\begin{pmatrix}X_1\\\hline X_2\\X_3\end{pmatrix} \sim N_3\left(\begin{pmatrix}-1\\\hline 0\\2\end{pmatrix}, \begin{array}{c} \Sigma_{11} \quad \Sigma_{12} \\ \left(\begin{array}{c|cc}3&-1&1\\\hline -1&2&0\\1&0&1\end{array}\right) \\ \Sigma_{21} \quad \Sigma_{22}\end{array}\right)$$

$$X_1 \sim N_1\left(\underset{\mu_1}{-1}+\underset{\Sigma_{12}}{(-1,1)}\underset{\Sigma_{22}^{-1}}{\begin{pmatrix}2&0\\0&1\end{pmatrix}^{-1}}\underset{\substack{x_2-\mu_2\\x_3-\mu_3}}{\begin{pmatrix}5-0\\2-2\end{pmatrix}},\ 3-(-1,1)\begin{pmatrix}2&0\\0&1\end{pmatrix}^{-1}\begin{pmatrix}-1\\1\end{pmatrix}\right)$$

$$= N_1\left(-\frac{15}{2}, \frac{3}{2}\right).$$

proof of conditional dist.

$$\underbrace{\begin{pmatrix}\underset{\sim}{X}_1\\\underset{\sim}{X}_2\end{pmatrix}}_{=\underset{\sim}{X}} \sim N_p\left(\underbrace{\begin{pmatrix}\underset{\sim}{\mu}_1\\\underset{\sim}{\mu}_2\end{pmatrix}}_{=\underset{\sim}{\mu}}, \underbrace{\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix}}_{=\Sigma}\right)$$

Define $A = \begin{pmatrix} I_{q\times q} & (-\Sigma_{12}\Sigma_{22}^{-1})_{q\times(p-q)} \\ O_{(p-q)\times q} & I_{(p-q)\times(p-q)}\end{pmatrix}$

Define $\boxed{\underset{\sim}{Y} = A(\underset{\sim}{X}-\underset{\sim}{\mu})}$

Find dist. of $\underset{\sim}{Y}$? Note: $\boxed{\underset{\sim}{X}-\underset{\sim}{\mu} \sim N_p(\underset{\sim}{0}, \Sigma)}$

so, $\underset{\sim}{Y} \sim N_p(0, A\Sigma A')$.

$$A\Sigma A' = \underbrace{\begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ O & I\end{pmatrix}}_{A}\overbrace{\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix}}^{\Sigma}\overbrace{\begin{pmatrix} I & O \\ -\Sigma_{22}^{-1}\Sigma_{21} & I\end{pmatrix}}^{A^T}$$

$$= \begin{pmatrix}\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, & \Sigma_{12}-\Sigma_{12} \\ \Sigma_{21}, & \Sigma_{22}\end{pmatrix}\begin{pmatrix} I & O \\ -\Sigma_{22}^{-1}\Sigma_{21} & I\end{pmatrix}$$

$$= \begin{pmatrix}\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}-(\Sigma_{12}-\Sigma_{12})(\Sigma_{22}^{-1}\Sigma_{21}), & \underset{0}{\Sigma_{12}-\Sigma_{12}} \\ \Sigma_{21}-\Sigma_{21}=0 & \Sigma_{22}\end{pmatrix}$$

$$= \begin{pmatrix}\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22}\end{pmatrix}$$

so, if $\underset{\sim}{Y} = \begin{pmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{Y}_2 \end{pmatrix} \begin{matrix} \in q\times 1 \\ \in (p-q)\times 1 \end{matrix}$, then $\underset{\sim}{Y}_1$ and $\underset{\sim}{Y}_2$ are independent

conditional dist. of Y_1 given $\underset{\sim}{Y}_2 = \underset{\sim}{y}_2$ is

$$N_q\left(\underset{\sim}{0}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

$$\begin{pmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{Y}_2 \end{pmatrix} = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix}\begin{pmatrix} \underset{\sim}{X}_1 - \underset{\sim}{\mu}_1 \\ \underset{\sim}{X}_2 - \underset{\sim}{\mu}_2 \end{pmatrix}$$

so, $\underset{\sim}{Y}_1 = (\underset{\sim}{X}_1 - \underset{\sim}{\mu}_1) - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X}_2 - \underset{\sim}{\mu}_2)$

$\underset{\sim}{Y}_2 = \underset{\sim}{X}_2 - \underset{\sim}{\mu}_2$

Conditional dist. of $\underset{\sim}{X}_1 - [\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{y}_2]$ given $\underset{\sim}{X}_2 \sim N_q(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$

Taking the vector to the other side, we see that conditional dist. of $\underset{\sim}{X}_1$ given $\underset{\sim}{x}_2$ is

$$N_q\left(\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

Math A7800 — 9/19/17, Tue

First exam : first Tuesday on October (10/3).

- Suppose $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$.

Normalized squared distance of $\underset{\sim}{X}$ from $\underset{\sim}{\mu}$ is

$$\boxed{(\underset{\sim}{X} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})}$$

- $\underset{\sim}{x}, \underset{\sim}{y} \in \mathbb{R}^p$

$dist(\underset{\sim}{x}, \underset{\sim}{y}) = \|\underset{\sim}{x} - \underset{\sim}{y}\|_2 = \sqrt{(\underset{\sim}{x} - \underset{\sim}{y})'(\underset{\sim}{x} - \underset{\sim}{y})}$.

$\|\underset{\sim}{x}\|_q = \left(\sum_{i=1}^{p} |x_i|^q\right)^{1/q}$

(q-norm of $\underset{\sim}{x}$) for $q \geq 1$.

e.g.

$\|\underset{\sim}{x}\|_1 = \sum_{i=1}^{p} |x_i|$

$\|\underset{\sim}{x}\|_\infty = \max_{1 \leq i \leq p} |x_i|$

Ex Suppose $X_1, X_2 \overset{iid}{\sim} N(\mu, \sigma^2)$

Normalize X_1 : $\frac{X_1 - \mu}{\sigma}$ ← standard deviation

Normalize X_2 : $\frac{X_2 - \mu}{\sigma}$

Normalized squared Distance between $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $\begin{pmatrix} \mu \\ \mu \end{pmatrix}$ is $\left(\frac{(X_1 - \mu)}{\sigma}\right)^2 + \left(\frac{(X_2 - \mu)}{\sigma}\right)^2$

$$= [X_1 - \mu, X_2 - \mu]\begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 1/\sigma^2 \end{bmatrix}\begin{bmatrix} X_1 - \mu \\ X_2 - \mu \end{bmatrix}$$

$$= \underbrace{(\underset{\sim}{X} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})}_{UDU'}, \text{ where } \underset{\sim}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \underset{\sim}{\mu} = \begin{pmatrix} \mu \\ \mu \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

Q: Distribution of normalized squared distance?

If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then $\boxed{(\underset{\sim}{X}-\underset{\sim}{\mu})'\underset{UDU'}{\Sigma^{-1}}(\underset{\sim}{X}-\underset{\sim}{\mu}) \sim \chi^2_p} = C^2$

Recall: If $Z_1, \dots, Z_K \overset{iid}{\sim} N(0,1)$, then

$\sum_{i=1}^{k} Z_i^2 = Y = Z_1^2 + Z_2^2 + \cdots + Z_K^2$ is said to follows χ^2_K distribution, K is referred as the degree of freedom of the chi-square distribution.

Pf Define $\underset{\sim}{Y} = \Sigma^{-1/2}(\underset{\sim}{X}-\underset{\sim}{\mu})$. Then $\underset{\sim}{Y} \sim N_p(\underset{\sim}{0}, I)$.

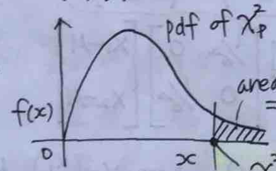
$\mathbb{E}(\underset{\sim}{Y}) = \underset{\sim}{0}$

$Var(\underset{\sim}{Y}) = \Sigma^{-1/2}\Sigma\Sigma^{-1/2} = I$

Express $(\underset{\sim}{X}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{X}-\underset{\sim}{\mu})$ in terms of $\underset{\sim}{Y}$.

$(\underset{\sim}{X}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{X}-\underset{\sim}{\mu}) = \underset{\sim}{Y}'\underset{\sim}{Y} = \sum_{i=1}^{p} Y_i^2 \sim \chi^2_p$, as $Y_1, \dots, Y_p \overset{iid}{\sim} N(0,1)$.

Notation $\chi^2_{p,\alpha}$ for $0 < \alpha < 1$



$\chi^2_{p,\alpha}$ is called α^{th} upper quantile for χ^2_p distribution.

Note: $\mathbb{P}(\chi^2_p \le \chi^2_{p,\alpha}) = 1-\alpha$.

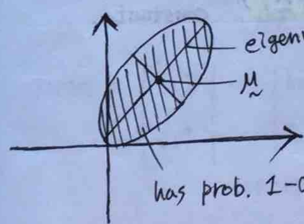
If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then

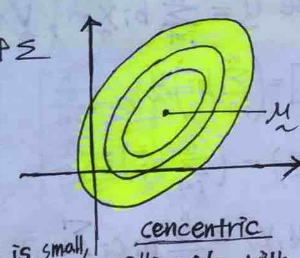
$\mathbb{P}((\underset{\sim}{X}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{X}-\underset{\sim}{\mu}) \le \chi^2_{p,\alpha}) = 1-\alpha$.

$\{\underset{\sim}{x} : (\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}) \le \chi^2_{p,\alpha}\}$ is an ellipsoid with center $\underset{\sim}{\mu}$ axes along the eigenvecters of Σ and half axes length given by $\sqrt{\lambda_i \chi^2_{p,\alpha}}$ (assuming $\lambda_1, \dots, \lambda_p$ are the eigenvalues of Σ).

$(\underset{\sim}{X}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}) \le 1$
$\underbrace{\chi^2_{p,\alpha}}$

$A = \frac{1}{\chi^2_{p,\alpha}}\Sigma^{-1}$, A has eigenvalues $\frac{1}{\lambda_i \chi^2_{p,\alpha}}$, $i = 1, \dots, p$





if α is small, the area of ellipsoid becomes larger.

concentric ellipsoids with different coverage prob.

Contour lines (level sets)

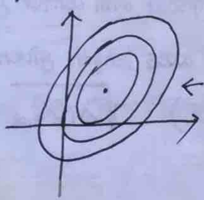
suppose $h: \mathbb{R}^p \to \mathbb{R}^1$. $h(x_1, \dots, x_p)$ is a real number.

Then the k-th level set is $\{\underset{\sim}{x} : h(\underset{\sim}{x}) = k\}$, $k \in \mathbb{R}$.

$X \sim N_p(\mu, \Sigma)$ with joint pdf

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right]$$

$$\{x : f(x) = k\} = \{x : (x-\mu)'\Sigma^{-1}(x-\mu) = -\log\left(k(2\pi)^{p/2}|\Sigma|^{1/2}\right)\}$$



← depends on "some constant."

Suppose $X_1, \dots, X_n \overset{iid}{\sim} N_p(\mu, \Sigma)$.

Suppose $U = \sum_{i=1}^n b_i X_i$, $V = \sum_{i=1}^n c_i X_i$, b_i, c_i are constant.

$E[U] = \left(\sum_{i=1}^n b_i\right)\mu$,

$Var(U) = \left(\sum_{i=1}^n b_i^2\right)\Sigma$

$cov(U, V) = \sum_{i=1}^n b_i c_i \Sigma$

Also, $$\begin{pmatrix} U \\ V \end{pmatrix} \sim N_{2p}\left(\begin{pmatrix} \sum_{i=1}^n b_i\mu \\ \sum_{i=1}^n c_i\mu \end{pmatrix}, \begin{pmatrix} \sum_{i=1}^n b_i^2\Sigma & \sum_{i=1}^n b_i c_i\Sigma \\ \sum_{i=1}^n b_i c_i\Sigma & \sum_{i=1}^n c_i^2\Sigma \end{pmatrix}\right).$$

consequence: If b, c are such that $\sum_{i=1}^n b_i c_i = 0$, then U and V are independent.

proof Define $$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}_{np\times 1} \sim N_{np}\left(\begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}, \begin{pmatrix} \Sigma & & 0 \\ & \ddots & \\ 0 & & \Sigma \end{pmatrix}_{np\times np}\right)$$

$$\begin{pmatrix} U \\ V \end{pmatrix} = \overbrace{\underbrace{\left[\begin{array}{c|c|c} b_1 I & \cdots & b_n I \\ \hline c_1 I & \cdots & c_n I \end{array}\right]}_{2p\times np}}^{A} \underbrace{\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}}_{X}$$, each Σ is $p\times p$. , $I_{(p\times p)}$

$(p\times p)$

$U = b_1 X_1 + \cdots + b_n X_n$

$= b_1 I X_1 + \cdots + b_n I X_n$

Since $X \sim N_{np}\left(\begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}, \begin{pmatrix} \Sigma & & 0 \\ & \ddots & \\ 0 & & \Sigma \end{pmatrix}\right)$, then

$$AX \sim N_{2p}\left(A\begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}, A\begin{pmatrix} \Sigma & & 0 \\ & \ddots & \\ 0 & & \Sigma \end{pmatrix}A'\right).$$

$$\text{mean} = \underbrace{\left[\begin{array}{c|c|c} b_1 I & \cdots & b_n I \\ c_1 I & \cdots & c_n I \end{array}\right]}_{2p\times np} \underbrace{\begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}}_{(np\times 1)} = \begin{bmatrix} \sum_{i=1}^n b_i\mu \\ \sum_{i=1}^n c_i\mu \end{bmatrix}_{2p\times 1}.$$

$$\Sigma = \left[\begin{array}{c|c|c} b_1 I & \cdots & b_n I \\ c_1 I & \cdots & c_n I \end{array}\right] \begin{bmatrix} \Sigma & & 0 \\ & \ddots & \\ 0 & & \Sigma \end{bmatrix} \left[\begin{array}{c|c} b_1 I & c_1 I \\ \vdots & \vdots \\ b_n I & c_n I \end{array}\right]$$

$$= \left[\begin{array}{c|c} \sum_{i=1}^n b_i^2\Sigma & \sum_{i=1}^n b_i c_i\Sigma \\ \hline \sum_{i=1}^n b_i c_i\Sigma & \sum_{i=1}^n c_i^2\Sigma \end{array}\right]$$

If $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n$ are independent and $\underset{\sim}{X}_i \sim N_p(\underset{\sim}{\mu}_i, \Sigma_i)$, then suppose $\underset{\sim}{U} = \sum_{i=1}^n b_i \underset{\sim}{X}_i$, $\underset{\sim}{V} = \sum_{i=1}^n c_i \underset{\sim}{X}_i$.

$$\begin{bmatrix} \underset{\sim}{U} \\ \underset{\sim}{V} \end{bmatrix} \sim N_{2p}\left(\begin{bmatrix} \sum_{i=1}^n b_i \mu_i \\ \sum_{i=1}^n c_i \mu_i \end{bmatrix}, \begin{bmatrix} \sum_{i=1}^n b_i^2 \Sigma_i & \sum_{i=1}^n b_i c_i \Sigma_i \\ \sum_{i=1}^n b_i c_i \Sigma_i & \sum_{i=1}^n c_i^2 \Sigma_i \end{bmatrix} \right).$$

Suppose $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n \overset{iid}{\sim} N_p(\underset{\sim}{\mu}, \Sigma)$. Find MLE of $\underset{\sim}{\mu}, \Sigma$.

Linear Algebra fact

For two matrices A, B,

$$\text{trace}(AB) = \text{trace}(BA).$$

$$\sum_i (AB)_{ii} = \sum_i \sum_j A_{ij} B_{ji} = \sum_j \sum_i B_{ji} A_{ij} = \sum_j (BA)_{jj}$$

Corollary For a vector $\underset{\sim}{x}$ and matrix A,

$$\underbrace{\underset{\sim}{x}'}_{B} \underbrace{A \underset{\sim}{x}}_{A} = \text{trace}(\underbrace{A \underset{\sim}{x}}_{A} \underbrace{\underset{\sim}{x}'}_{B})$$

$$= \text{trace}(\underbrace{\underset{\sim}{x}'}_{B} \underbrace{A \underset{\sim}{x}}_{A})$$

$$= \underset{\sim}{x}' A \underset{\sim}{x}.$$

Math A7800 9/26/17, Tue

Exam 01 Tue $\longrightarrow$ Thur. (10/4)

cover by this thursday and HW 01, 02, 03.

one-proof in Lectures.

Fact 1. For two matrices A and B square symmetric matrices

$$tr(AB) = tr(BA) \Rightarrow \boxed{\begin{aligned} \underset{\sim}{x}' A \underset{\sim}{x} &= tr(\underset{\sim}{x}' A \underset{\sim}{x}) \\ &= tr(A \underset{\sim}{x} \underset{\sim}{x}'). \end{aligned}}$$

2. For a square symmetric matrix A,

$$tr(A) = \sum_{i=1}^n \lambda_i, \text{ where } \lambda_i\text{'s are the eigenvalues of } A.$$

proof (2) Suppose $A = PDP'$ is the spectral decomposition.

so, P is orthogonal and D is diagonal $\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$.

$$tr(\overset{A}{P}\overset{B}{DP'}) \overset{\text{by (1) } tr(AB) = tr(BA)}{=} tr(\overset{B}{D}\underbrace{\overset{A}{P'P}}_{I}) = tr(D) = \sum_{i=1}^n \lambda_i$$

Goal: Find MLE

Suppose $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n \overset{iid}{\sim} N_p(\mu, \Sigma)$, μ, Σ are unknown.

Find MLE of $\underset{\sim}{\mu}$ and Σ.

Likelihood function.

$L(\underset{\sim}{\mu}, \Sigma)$ = joint density of $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n$

$L(\underset{\sim}{\mu}, \Sigma | \underset{\sim}{x}_1, \dots, \underset{\sim}{x}_n)$ = product of the marginal pdf (as they are ind.)

$$= \prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underset{\sim}{x}_i - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_i - \underset{\sim}{\mu})\right] = \prod_{i=1}^{n} f_{\underset{\sim}{\mu},\Sigma}(\underset{\sim}{x}_i)$$

$$\underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & & \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}$$

"p-parameters"

number of parameters of Σ

$$= \frac{p(p+1)}{2} = p + (p-1) + \cdots + 1$$

Total number of parameters $= \dfrac{p(p+3)}{2}$.

so, $L : \mathbb{R}^{p(p+3)/2} \longrightarrow \mathbb{R}^1$.

First, simplify L.

$\underset{\sim}{x}'A\underset{\sim}{x} = tr(A\underset{\sim}{x}\underset{\sim}{x}')$
$= tr(\underset{\sim}{x}'A\underset{\sim}{x})$

$$L(\underset{\sim}{\mu}, \Sigma) = (2\pi)^{-\frac{np}{2}}|\Sigma|^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_i - \underset{\sim}{\mu})\right]$$

$$= (2\pi)^{-\frac{np}{2}}|\Sigma|^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n} tr\left(\Sigma^{-1}(\underset{\sim}{x}_i - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \underset{\sim}{\mu})'\right)\right]$$

$$= (2\pi)^{-\frac{np}{2}}|\Sigma|^{-n/2} \exp\left[-\frac{1}{2} tr\left(\Sigma^{-1}\underbrace{\sum_{i=1}^{n}(\underset{\sim}{x}_i - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \underset{\sim}{\mu})'}_{*}\right)\right]$$

$$(\underset{\sim}{x}_i - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_i - \underset{\sim}{\mu}) = tr\left((\underset{\sim}{x}_i - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}_i - \underset{\sim}{\mu})\right)$$
$$= tr\left(\Sigma^{-1}(\underset{\sim}{x}_i - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \underset{\sim}{\mu})'\right)$$

$\underset{\sim}{x}'A\underset{\sim}{x} = tr(\underset{\sim}{x}'A\underset{\sim}{x})$
$= tr(A\underset{\sim}{x}\underset{\sim}{x}')$ if A is $(k\times k)$ square symmetric matrix and $\underset{\sim}{x}$ is $(k\times 1)$ vector.

Look at

$$\sum_{i=1}^{n}(\underset{\sim}{x}_i - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \underset{\sim}{\mu})' = \sum_{i=1}^{n}(\underbrace{\underset{\sim}{x}_i - \bar{\underset{\sim}{x}}} + \underbrace{\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}})(\underbrace{\underset{\sim}{x}_i - \bar{\underset{\sim}{x}}} + \underbrace{\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}})'$$

$$= \boxed{\sum_{i=1}^{n}(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})'} + \underbrace{\boxed{\sum_{i=1}^{n}(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})'}}_{=\underset{\sim}{0}} + \underbrace{\boxed{\sum_{i=1}^{n}(\underset{\sim}{x}_i - \underset{\sim}{\mu})(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})'}}_{=\underset{\sim}{0}}$$

$$+ \underbrace{\boxed{\sum_{i=1}^{n}(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})'}}_{= n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})'}$$

Remember $\bar{\underset{\sim}{x}} = \frac{1}{n}\sum_{i=1}^{n}\underset{\sim}{x}_i$, so $\sum_{i=1}^{n}(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}}) = \underset{\sim}{0}$.
$n\bar{\underset{\sim}{x}} - n\bar{\underset{\sim}{x}}$

Thus,

$$\underbrace{\sum_{i=1}^{n}(\underset{\sim}{x}_i - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \underset{\sim}{\mu})'}_{*} = \sum_{i=1}^{n}(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})' + n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})'.$$

so, $\boxed{tr\left(\Sigma^{-1}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \underset{\sim}{\mu})'\right)}$

$$= tr\left(\Sigma^{-1}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})'\right) + \underbrace{n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})'\Sigma^{-1}(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})}_{\text{positive}}$$

$$\geq \boxed{tr\left(\Sigma^{-1}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})'\right)}$$

and equality holds iff $\bar{\underset{\sim}{x}} = \underset{\sim}{\mu}$.

$$\sum_{i=1}^{n}(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})'$$
$$= (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})\left(\sum_{i=1}^{n} x_i - n\bar{\underset{\sim}{x}}\right)'$$
$$= (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})(n\bar{\underset{\sim}{x}} - n\bar{\underset{\sim}{x}})'$$
$$= \underset{\sim}{0}$$

$\bar{\underset{\sim}{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i$
$\Rightarrow \boxed{\sum_{i=1}^{n} x_i = n\bar{\underset{\sim}{x}}}$

This implies

$$L(\underset{\sim}{\mu}, \Sigma) \leq L(\bar{\underset{\sim}{x}}, \Sigma).$$

Hence, $\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{x}}$

$L(\bar{\underset{\sim}{x}}, \Sigma) = (2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}} \exp\left[-\frac{1}{2} tr(\Sigma^{-1}B)\right]$, where $B = \sum_{i=1}^{n}(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})'$.

$$\boxed{L(\bar{x},\Sigma)} = (2\pi)^{-\frac{np}{2}}|\Sigma|^{-n/2}\exp\left[-\frac{1}{2}tr(\Sigma^{-1}B)\right] \quad (*)$$

, where $B=\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'$.

Linear Algebra Fact

For pd matrices Σ and B.

$$\frac{\exp(-\frac{1}{2}tr(\Sigma^{-1}B))}{|\Sigma|^a} \le \frac{\exp(-ap)(2a)^{ap}}{|B|^a}$$

upper bound

Equality holds, for $\boxed{\Sigma = \frac{1}{2a}B}$.

Applying this fact in our case:

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})' = \boxed{\frac{n-1}{n}S}$$

so, $L(\bar{x},\Sigma) \le \boxed{L(\bar{x},\frac{1}{n}B)}$, $\boxed{S=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'}$

$\ge$

$L(\mu,\Sigma)$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})' = \frac{n-1}{n}S$$

$= \frac{1}{n}B.$

$$tr(\Sigma^{-1}B) = tr(\Sigma^{-1}B^{1/2}B^{1/2}) = tr(B^{1/2}\Sigma^{-1}B^{1/2})$$

symmetric pd matrix

$=\sum_{i=1}^{n}\lambda_i$, where λ_i's are the eigenvalues of $B^{1/2}\Sigma^{-1}B^{1/2}$.

(λ_i's are our new variables).

$$\prod_{i=1}^{p}\lambda_i = |B^{1/2}\Sigma^{-1}B^{1/2}| = |B^{1/2}||\Sigma^{-1}||B^{1/2}|$$

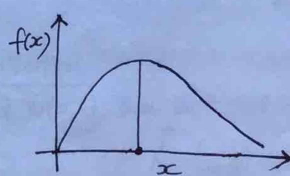
$$= |B|^{1/2}|\Sigma|^{-1}|B|^{1/2} = \frac{|B|}{|\Sigma|} \qquad |\Sigma| = \frac{|B|}{\prod_{i=1}^{p}\lambda_i}$$

so, $|\Sigma| = \dfrac{|B|}{\prod_{i=1}^{p}\lambda_i}$.

Thus, $$\frac{\exp\left[-\frac{1}{2}tr(\Sigma^{-1}B)\right]}{|\Sigma|^a} = \frac{\prod_{i=1}^{p}\lambda_i^a}{|B|^a}\exp\left[-\frac{1}{2}\sum_{i=1}^{p}\lambda_i\right]$$

$$= \frac{1}{|B|^a}\prod_{i=1}^{p}\left(\lambda_i^a e^{-\lambda_i/2}\right)$$

$$\boxed{f(x) = x^a e^{-x/2}}$$



maximize $f(x)$

$$f'(x) = ax^{a-1}e^{-x/2} - \frac{1}{2}x^a e^{-x/2} = 0$$

$x = 2a$.

So, $\prod_{i=1}^{p} \lambda_i^{a} e^{-\lambda_i/2}$ is maximized, when $\lambda_i = 2a$, for all i.

For these values of λ_i,

$$B^{1/2}\Sigma^{-1}B^{1/2} = 2a I$$

> **Fact**: If a symmetric matrix A has all eigenvalues equal to α, then $A = \alpha I$.
>
> *Why?* look at spectral decom.
>
> $$A = PDP' = P(\alpha I)P' = \alpha I$$

$$B^{-1/2}(B^{1/2}\Sigma^{-1}B^{1/2})B^{-1/2} = B^{-1/2}(2aI)B^{-1/2} = 2aB^{-1}$$

$$\overset{\shortparallel}{\Sigma^{-1}} \Rightarrow \underline{\Sigma = (2aB^{-1})^{-1} = \frac{1}{2a}B}.$$

Distribution of MLE

Set up: $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n \overset{iid}{\sim} N_p(\underset{\sim}{\mu}, \Sigma)$

$$\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{X}} = \frac{1}{n}\sum_{i=1}^{n} \underset{\sim}{X}_i, \quad \hat{\Sigma} = \frac{n-1}{n} S = \frac{1}{n}\sum_{i=1}^{n}(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})', \text{ where}$$

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})'.$$

Result

(1) $\bar{\underset{\sim}{X}} \sim N_p(\underset{\sim}{\mu}, \frac{1}{n}\Sigma)$

(2) $(n-1)S \sim W_{n-1}(\Sigma)$ (Wishart distribution with $n-1$ d.f. and parameter Σ)

(3) $\bar{\underset{\sim}{X}}$ and S are independent.

(see that $\bar{\underset{\sim}{X}}$ is a random vector and S is a random matrix)

Def **Wishart Distribution** (multivariate version of χ^2_p)

If $\underset{\sim}{Z}_1, \underset{\sim}{Z}_2, \ldots, \underset{\sim}{Z}_m \overset{iid}{\sim} N_p(0, \Sigma)$, then

$\underbrace{\sum_{i=1}^{m} \underset{\sim}{Z}_i \underset{\sim}{Z}_i'}_{p \times p} = W$ is said to have Wishart dist. with m d.f. and parameter Σ.

Analogy with χ^2_m distribution

If $U_1, \ldots, U_m$ are i.i.d $N(0,1)$, then $U_1^2 + \cdots + U_m^2 \sim \chi^2_m$.

$$\underset{\sim}{Z}_1\underset{\sim}{Z}_1' + \cdots + \underset{\sim}{Z}_m\underset{\sim}{Z}_m' \sim W_m(\Sigma)$$

Q–Q Plot (quantile vs quantile plot)

A diagonostic test to find out whether a given set of numbers is comming from a normal dist. or not.

setup

Suppose we obtained $x_1, \dots, x_n$ (n-samples).

Q: Are they observations from an underlying normal dist.?

Step 1 Obtain sample quantiles:

Order the samples in increasing order,

say $\boxed{x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}}$ → corresponding

$x_{(j)}$ is the jth order statistics. probability levels (values)

$$\frac{(1-\frac{1}{2})}{n} \cdots \frac{(n-\frac{1}{2})}{n}$$

↕

$\left(\frac{j}{n}\right)^{th}$ sample quantile

meaning that the proportion of sample values less or equal to $x_{(j)}$ is $\frac{j}{n}$.

Step 2 Obtain population quantiles (Standard Normal quantiles)

Given α between 0 and 1, the α^{th} quantile of $\underset{"Z"}{N(0,1)}$ is the value $q_{(j)}$ such that $\alpha = \frac{j-\frac{1}{2}}{n}$

☆

$$\boxed{\mathbb{P}(\underset{N(0,1)}{Z} \le q_{(j)}) = \int_{-\infty}^{q_{(j)}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = P_{(j)} = \underbrace{\frac{j-\frac{1}{2}}{n}}_{\alpha}}$$

"$q_{(1)}, \dots, q_{(n)}$"

Let $q_{(j)}$ denote the $\frac{j-\frac{1}{2}}{n}$ th population quantile.

or $\frac{j}{n+1}$

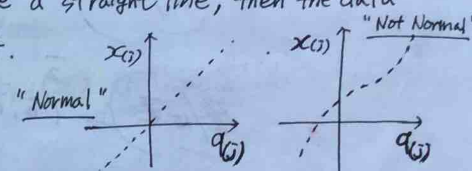
"Standard Normal Quantiles"

$q_{(1)}, q_{(2)}, \dots, q_{(n)}$

Step 3 Plot $(q_{(j)}, x_{(j)})$ $j = 1, 2, \dots, n$

This plot is called the Q-Q plot for sample.

If the plot looks like a straight line, then the data is from Normal Dist.



Test for normality using correlation Coeff.

Obtain correlation coeff. between $\begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix}$ and $\begin{pmatrix} x_{(1)} \\ \vdots \\ x_{(n)} \end{pmatrix}$.

Let it be $r_Q = \dfrac{\sum_{j=1}^{n} (x_{(j)} - \bar{x})(q_j - \overset{0}{\bar{q}})}{\sqrt{\sum_{j=1}^{n} (x_{(j)} - \bar{x})^2} \sqrt{\sum_{j=1}^{n} q_{(j)}^2}}$

If r_Q is smaller than critical points for the Q-Q plot, then we reject the hypothesis of normality.

If r_Q is larger that C.P., then we accept the normality.

is "cyclic"?

$\mathbb{Z}/n\mathbb{Z} = (\bar{0}, \ldots, \overline{n-1})$

$\bar{2}$

$\mathbb{Z}/15\mathbb{Z} = (\bar{0}, \bar{1}, \ldots, \overline{14})$

Distributions of MLE

$\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n \overset{iid}{\sim} N_p(\underset{\sim}{\mu}, \Sigma)$. Then

(1) $\bar{\underset{\sim}{X}} \sim N_p(\underset{\sim}{\mu}, \frac{1}{n}\Sigma)$

(2) $S = \frac{1}{n-1}\sum_{i=1}^{n}(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})'$

and $\bar{\underset{\sim}{X}}$ are independent.

(3) $(n-1)S \sim \text{Wishart}_{n-1}(\Sigma)$, $(n-1)S = \sum_{i=1}^{n}(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})'$

Recall Projection matrix.

For a matrix Y (full column rank), the projection matrix onto $C(Y)$ (column space) is $\boxed{P_Y = Y(Y'Y)^{-1}Y'}$.

$= col(Y)$

$P_{\underset{\sim}{1}} = \underset{\sim}{1}(\underset{\sim}{1}'\underset{\sim}{1})^{-1}\underset{\sim}{1}'$, $\underset{\sim}{1}_{n\times 1} = \begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}$

$(n\times n)$

$= \frac{1}{n}\underset{\sim}{1}\,\underset{\sim}{1}'$

Suppose $X = \begin{bmatrix}\underset{\sim}{X}_1' \\ \vdots \\ \underset{\sim}{X}_n'\end{bmatrix}_{n\times p}$ Data matrix

$\left(n\bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}' = nX'\frac{1}{n}\underset{\sim}{1}\frac{1}{n}\underset{\sim}{1}'X\right)$

$= X'P_1X$

$\bar{\underset{\sim}{X}} = X'\underset{\sim}{1}\cdot\frac{1}{n}$, $\boxed{(n-1)S = \sum_{i=1}^{n}(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})'}$

$\bar{\underset{\sim}{X}}' = \frac{1}{n}\underset{\sim}{1}'X$

$= \sum_{i=1}^{n}\underset{\sim}{X}_i\underset{\sim}{X}_i' - n\bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}'$

$= X'X - X'P_{\underset{\sim}{1}}X = X'(I - P_{\underset{\sim}{1}})X$

Note that $(I-P_{\underset{\sim}{1}})\underset{\sim}{1} = \underset{\sim}{0}$.

$$\underset{\sim}{1} - P_{\underset{\sim}{1}}\underset{\sim}{1} = \underset{\sim}{1} - \frac{1}{n}\underset{\sim}{1}\,\underset{\sim}{1}'\,\underset{\sim}{1}$$

Fact $P_{\underset{\sim}{1}}$ has rank 1.

$(I-P_{\underset{\sim}{1}})$ has rank $n-1$.

The eigenvalues of $P_{\underset{\sim}{1}}$ are $1, 0, 0, \ldots, 0$.

The eigenvalues of $I-P_{\underset{\sim}{1}}$ are $0, 1, 1, \ldots, 1$.

Spectral decomposition

$$I-P_{\underset{\sim}{1}} = U\begin{pmatrix} 0 & & 0 \\ & 1 & \\ 0 & & \ddots\ 1 \end{pmatrix}U', \text{ where } U \text{ is orthogonal.}$$

$$= VV' \qquad V = \left[U_{\cdot 2} \mid U_{\cdot 3} \mid \cdots \mid U_{\cdot n}\right]_{n\times(n-1)}$$

Let $W = \begin{bmatrix} \frac{1}{\sqrt{n}}\underset{\sim}{1}' \\ V' \end{bmatrix}_{n\times n}$ is an orthogonal matrix.

$$WW' = \begin{pmatrix} \frac{1}{\sqrt{n}}\underset{\sim}{1}' \\ V' \end{pmatrix}\left(\frac{1}{\sqrt{n}}\underset{\sim}{1} \,\middle|\, V\right) = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & I_{n-1} \end{array}\right) \text{ as } V'\underset{\sim}{1} = \underset{\sim}{0}$$

orthogonal matrix

$U = [U_{\cdot 1} \mid V]$, $U' = \begin{bmatrix} U_{\cdot 1}' \\ V' \end{bmatrix}$

$$I_n = U'U = \begin{bmatrix} U_{\cdot 1}' \\ V' \end{bmatrix}[U_{\cdot 1} \mid V]$$

$$= \begin{bmatrix} * & * \\ * & V'V \end{bmatrix}$$

$$(I-P_1)\underset{\sim}{1} = VV'\underset{\sim}{1} \Rightarrow V'\underbrace{(I-P_{\underset{\sim}{1}})\underset{\sim}{1}}_{\underset{\sim}{0}} = V'\underset{\sim}{1} = \underset{\sim}{0}$$

Thus, W is an orthogonal matrix.

Look at $\underset{n\times n}{W}\;\underset{n\times p}{X} = Y_{n\times p} = \begin{bmatrix} \underset{\sim}{Y}_1' \\ \vdots \\ \underset{\sim}{Y}_n' \end{bmatrix}$

$$\underset{\sim}{Y}_1' = \frac{1}{\sqrt{n}}\underset{\sim}{1}'X = \sqrt{n}\,\bar{\underset{\sim}{X}}'$$

$\bar{X}$ is a function of (Y_1).

$$\sum_{i=2}^{n} \underset{\sim}{Y}_i \cdot \underset{\sim}{Y}_i' = X'VV'X \text{ as } WX = \begin{pmatrix} \frac{1}{\sqrt{n}}\underset{\sim}{1}'X \\ V'X \end{pmatrix}$$

So, $V'X = \begin{pmatrix} \underset{\sim}{Y}_2' \\ \vdots \\ \underset{\sim}{Y}_n' \end{pmatrix}$

Since $VV' = (I-P_{\underset{\sim}{1}})$,

$$\boxed{\sum_{i=2}^{n} \underset{\sim}{Y}_i \cdot \underset{\sim}{Y}_i'} = X'(I-P_{\underset{\sim}{1}})X = \boxed{(n-1)S}$$

So, if we show that $\underset{\sim}{Y}_1$ is independent with $\underset{\sim}{Y}_2, \ldots, \underset{\sim}{Y}_n$, then $\bar{\underset{\sim}{X}}$ and S must also be independent.

" $\underset{\sim}{Y}_1$ is a function of $\bar{\underset{\sim}{X}}$. "

y_i' and y_j' are independent for any $i \neq j$, as $W_{i\cdot} \cdot W_{j\cdot} = 0$

$y_i \sim N_p(0, \Sigma)$ check!

for $i = 2, \dots, n$

so, $(n-1)S = \sum_{i=2}^{n} y_i y_i' \sim \text{Wishart}(\Sigma)_{n-1 \text{ d.f.}}$

. Asymptotic Distribution (large small case)

If $X_1, X_2, \dots, X_n$ are iid p-variate random variables with mean μ and variance Σ., then

WLLN: (weak law of large number)

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[\text{as } n \to \infty]{\text{"converges"}} \mu \text{ in probability}$$

For any $\varepsilon > 0$, $\boxed{\lim_{n\to\infty} \mathbb{P}(\| \bar{X}_n - \mu \| > \varepsilon) = 0}$

$\| \bar{X}_n - \mu \| < \varepsilon$ as $n \to \infty$

so, when n is large, then $\bar{X}$ is close to μ.

. CLT (Central limit Thm)

$\sqrt{n}(\bar{X} - \mu)$ converges in distribution to $N_p(0, \Sigma)$

so, approx. dist. of $\sqrt{n}(\bar{X} - \mu)$ is $N_p(0, \Sigma)$ and approx. dist. of $\bar{X}$ is $N_p(\mu, \frac{1}{n}\Sigma)$

so, $\boxed{n(\bar{X} - \mu)'\Sigma^{-1}(\bar{X} - \mu) \overset{\text{approx.}}{\sim} \chi^2_p}$

$$\underbrace{\sqrt{n}\,\Sigma^{-\frac{1}{2}}(\bar{X} - \mu)}_{y} \overset{\text{approx}}{\sim} N_p(0, I)$$

$$y'y = n(\bar{X} - \mu)'\Sigma^{-1}(\bar{X} - \mu) \sim \chi^2_p$$

MLE fact

Suppose $\hat{\theta}$ is MLE of θ. Then what is an MLE for $\eta = g(\theta)$ (g is a function)

$\hat{\eta} = g(\hat{\theta})$

EX $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n \overset{iid}{\sim} N_p(\underset{\sim}{\mu}, \Sigma)$ $\quad \Sigma_{p\times p}$

Find MLE of ρ_{12} = corr coef. between 1st and 2nd components.

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$$

$$S = \frac{1}{(n-1)}\sum_{i=1}^{n}(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})'.$$

$$\hat{\Sigma}_{MLE} = \frac{n-1}{n}S = \frac{1}{n}\sum_{i=1}^{n}(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})'$$

$$\hat{\rho}_{12\,MLE} = \frac{\frac{n-1}{n}S_{12}}{\sqrt{\frac{n-1}{n}S_{11}\cdot\frac{n-1}{n}S_{22}}} = \frac{S_{12}}{\sqrt{S_{11}\cdot S_{22}}}$$

Checking for multivariate normality.

$p=1$ case $\longleftrightarrow$ Q-Q plots.

we are given $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n$ ($n \times p$ vectors)

The data matrix $X = \begin{pmatrix} \underset{\sim}{X}_1' \\ \vdots \\ \underset{\sim}{X}_n' \end{pmatrix}$

Goal: Check whether they are coming from some multivariate normal dist. or not.

Step 1 obtain p many Q-Q plots for the components.
One Q-Q plot for each column of X.
If one of the plots is not approx. straight line, then stop! and conclude non-normality.
otherwise, go to step 2.

Step 2 check for bivariate normality for the pairs.
For the first two components $\underset{\substack{\uparrow \\ 1^{st}\text{ column}}}{[X_{\cdot 1},} \underset{\substack{\uparrow \\ 2^{nd}\text{ column}}}{X_{\cdot 2}]}$.

Compute the proportion of the sample that lies within the ellipse $\{\underset{\sim}{y} \in \mathbb{R}^2, (\underset{\sim}{y} - \hat{\mu})'S^{-1}(\underset{\sim}{y} - \hat{\mu}) \le \chi_2^2(0.5)\}$,
where $\hat{\mu}$ is average of the rows of $[X_{\cdot 1}, X_{\cdot 2}]$ and S is the corresponding sample variance matrix.

$$\boxed{(\underset{\sim}{x} - \bar{\underset{\sim}{x}})'S^{-1}(\underset{\sim}{x} - \bar{\underset{\sim}{x}}) \le \chi_2^2(0.5)}$$

If **the proportion** is much different from $\frac{1}{2}$, then conclude non-normality.

check for all pairs of components.

If it fails for one pairs, then conclude non-normality.

Otherwise, normality.

(practically if all pairs pass, then mult. normality holds),

$$S_n = \frac{1}{n}\left(\sum_{j=1}^{n}(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})'\right)$$

$$= \frac{1}{n}\left(\sum_{j=1}^{n} \underset{\sim}{x}_j\underset{\sim}{x}_j' - \bar{\underset{\sim}{x}}\underset{\sim}{x}_j' - \underset{\sim}{x}_j\bar{\underset{\sim}{x}}' + \bar{\underset{\sim}{x}}\bar{\underset{\sim}{x}}'\right)$$

$$= \frac{1}{n}\sum_{j=1}^{n}\underset{\sim}{x}_j\underset{\sim}{x}_j' - \bar{\underset{\sim}{x}}\bar{\underset{\sim}{x}}' - \cancel{\bar{\underset{\sim}{x}}\bar{\underset{\sim}{x}}'} + \cancel{\bar{\underset{\sim}{x}}\bar{\underset{\sim}{x}}'}$$

$\mathbb{E}(S_n)$ $\frac{1}{n}\cancel{\Sigma - \underset{\sim}{\mu}\underset{\sim}{\mu}'}$ $\frac{1}{n}(\Sigma - \underset{\sim}{\mu}\underset{\sim}{\mu}')$ $\frac{1}{n}\sum_{j=1}^{n}\underset{\sim}{x}_j$ $\frac{1}{n^2}\sum_{j=1}^{n}\underset{\sim}{x}_j\underset{\sim}{x}_j'$ $\frac{1}{n^2}\sum_{j=1}^{n}(\Sigma - \underset{\sim}{\mu}\underset{\sim}{\mu}')$

$$\mathbb{E}(S_n) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(\underset{\sim}{x}_j\underset{\sim}{x}_j') - \mathbb{E}(\bar{\underset{\sim}{x}}\bar{\underset{\sim}{x}}')$$

$$= \frac{1}{n}\sum_{j=1}^{n}(\Sigma - \underset{\sim}{\mu}\underset{\sim}{\mu}') - \frac{1}{n}\mathbb{E}\left(\sum_{j=1}^{n}\underset{\sim}{x}_j\underset{\sim}{x}_j'\right) - \sum_{j=1}^{n}\mathbb{E}(\underset{\sim}{x}_j\underset{\sim}{x}_j')$$

$$= \frac{1}{n}\sum_{j=1}^{n}(\Sigma - \underset{\sim}{\mu}\underset{\sim}{\mu}') - \sum_{j=1}^{n}(\Sigma - \underset{\sim}{\mu}\underset{\sim}{\mu}')$$

$$= \frac{1}{n}\sum_{j=1}^{n}(\Sigma - \cancel{\underset{\sim}{\mu}\underset{\sim}{\mu}'} - \Sigma + \cancel{\underset{\sim}{\mu}\underset{\sim}{\mu}'})$$

$$= \frac{1}{n}\sum_{j=1}^{n}(\Sigma - \underset{\sim}{\mu}\underset{\sim}{\mu}') - \frac{1}{n^2}\sum_{j=1}^{n}(\Sigma - \underset{\sim}{\mu}\underset{\sim}{\mu}')$$

$$= \frac{1}{n}\sum_{j=1}^{n}(\Sigma - \underset{\sim}{\mu}\underset{\sim}{\mu}') - \frac{1}{n}\Sigma + \cancel{\underset{\sim}{\mu}\underset{\sim}{\mu}'}$$

$$= \Sigma - \underset{\sim}{\mu}\underset{\sim}{\mu}' - \frac{1}{n}\Sigma + \underset{\sim}{\mu}\underset{\sim}{\mu}'$$

$$= \left(1 - \frac{1}{n}\right)\Sigma = \frac{n-1}{n}\Sigma$$

$(\underset{\sim}{x} - \bar{x})' S^{-1} (\underset{\sim}{x} - \bar{x})$

$(\underset{\sim}{x} - \bar{x})' \left(S^{-1}\right) \begin{pmatrix} 2-3 \\ 4-3 \end{pmatrix}$

$\begin{pmatrix} 2-3 \\ 4-3 \end{pmatrix}' \left(S^{-1}\right) \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

S_n $\left(\frac{n}{n-1}S_n\right)$ $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

$(-1, 1)$ $\le \chi_2^2(0.5)$

Chi-Square plots

Given $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n$ (p-variate), Compute

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

sample variance matrix.

Compute $\mathbb{E}(S_n)$

$$d_i^2 = (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})' S^{-1} (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}}) \quad \text{for } i = 1, 2, \ldots, n.$$

Check whether they are from χ_p^2 distribution or not.

Obtain $d_{(1)}^2 \le \cdots \le d_{(n)}^2$ "order statistics"

Obtain $q_1, \ldots, q_n$, where q_j is $\frac{j - \frac{1}{2}}{n} \times 100$ percentile of χ_p^2

plot $(q_j, d_{(j)}^2)$, $1 \le j \le n$. $\mathbb{P}(\chi_p^2 \le q_j) = \frac{j - \frac{1}{2}}{n}$ find this.

The plot is supposed to be approx. straight line if normality assumption holds.

Ex (checking bivariate normality)

x_1 = sales, x_2 = profits for the 10 largest companies.

$$\bar{x} = \begin{pmatrix} 155.60 \\ 14.70 \end{pmatrix}, \quad S = \begin{pmatrix} 7476.45 & 303.62 \\ 303.62 & 26.19 \end{pmatrix}$$

$$S^{-1} = \frac{1}{103{,}623.12} \begin{pmatrix} 26.19 & -303.62 \\ -303.62 & 7476.45 \end{pmatrix} = \begin{pmatrix} 0.000253 & -0.002930 \\ -0.002930 & 0.072148 \end{pmatrix}$$

$$\chi^2_2(0.5) = 1.39.$$

$$(\underset{\sim}{x} - \underset{\sim}{\bar{x}})' S^{-1} (\underset{\sim}{x} - \underset{\sim}{\bar{x}}) = \begin{pmatrix} x_1 - \bar{x}_1 \\ x_2 - \bar{x}_2 \end{pmatrix}' \begin{pmatrix} 0.000253 & -0.002930 \\ -0.002930 & 0.072148 \end{pmatrix} \begin{pmatrix} x_1 - \bar{x}_1 \\ x_2 - \bar{x}_2 \end{pmatrix}$$

$$= \begin{pmatrix} 108.28 - 155.60 \\ 17.05 - 14.70 \end{pmatrix}' \left(S^{-1} \right) \begin{pmatrix} 108.28 - 155.60 \\ 17.05 - 14.70 \end{pmatrix} = \underline{1.61 > 1.39.}$$

this point falls outside the 50% contour.

Check all other points.

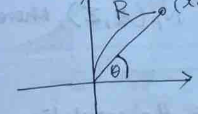
If it fails for one pairs, then conclude non-normality.
otherwise, normality.

Practically, if all pairs pass, then mult. normality holds.

Exam-06 $X, Y \overset{iid}{\sim} N(0,1)$

$R = \sqrt{x^2+y^2}$. Find pdf of R.

Hint: find joint pdf of R and $\theta = \tan^{-1}\frac{y}{x}$.



then $R = \sqrt{x^2+y^2}$

$\theta = \tan^{-1}\frac{y}{x}$

$X = R\cos\theta$

$Y = R\sin\theta$.

Joint pdf of x and y

$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

Joint pdf of R and θ

$$g(r,\theta) = f(r\cos\theta, r\sin\theta) \cdot J, \text{ where } J = \left| \frac{\partial(x,y)}{\partial(r,\theta)} \right|$$

$$\text{so, } g(r,\theta) = \frac{1}{2\pi} e^{-\frac{1}{2}r^2} \cdot r \qquad = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & \sin\theta \\ -\sin\theta & r\cos\theta \end{vmatrix}$$

$r \geq 0,$

$\theta \in [0, 2\pi]$

$= r.$

So, marginal of R is

$$h(r) = \int g(r,\theta)\,d\theta = \int_0^{2\pi} \frac{1}{2\pi} r e^{-\frac{1}{2}r^2} d\theta = r e^{-\frac{1}{2}r^2}$$

for $r \geq 0$.

Chapter 5 (Inference about Population mean)

Question Suppose $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \dots, \underset{\sim}{X}_n$ are iid from $N_p(\underset{\sim}{\mu}, \Sigma)$, where both $\underset{\sim}{\mu}$ and Σ are unknown.

Based on the samples we need to infer about the population mean and covariance matrix.

Hypothesis testing

H_0: (Null hypothesis), $\underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ "specified value".

H_1: (Alternative hypothesis), $\underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$.

This is a two-sided multivariate testing.

Look at the univariate case first:

Suppose $Y_1, Y_2, \dots, Y_n$ are iid $N(\mu, \sigma^2)$ (both μ, σ^2 are unknown)

test the hypothesis $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$.

We use the test statistic

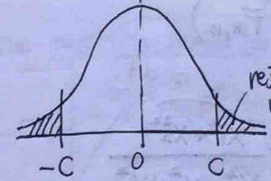
$$t = \frac{\bar{y} - \mu_0}{\sqrt{\frac{S^2}{n}}}, \text{ where } S^2 = \frac{1}{n-1}\sum_i^n (y_i - \bar{y})^2 \text{ (sample variance).}$$

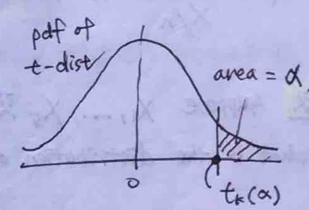
$$= \frac{n(\bar{y}-\mu_0)^2}{S^2}$$

The rejection region at level of significance α is of the form $\{|t| > c\}$ for some c.

$t \sim t_{n-1}$ (t-distribution with d.f. $n-1$).

The pdf of t-dist





Suppose $t_k(\alpha)$ denotes the $(100\alpha)^{th}$ percentile of the t-distribution with d.f. k, where $k = n-1$.

If total area $= \alpha$, then $\boxed{c = t_{n-1}(\alpha/2)}$

So, we will reject H_0 at level of significance α if $|t| \geq t_{n-1}(\alpha/2)$, otherwise accept H_0.

See that

$$t^2 = \frac{n(\bar{y}-\mu_0)^2}{S^2} \overset{H_0}{\sim} F_{1,n-1}$$

If $U \sim N(0,1)$ and $V \sim \chi^2_k$ (Chi-square dist. with k d.f. and U, V are independent), then $\frac{U}{\sqrt{V/k}} \sim t_k$. "t-distribution"

$$\left(\frac{U}{\sqrt{V/k}}\right)^2 = \frac{U^2}{V/k} \sim F_{1,k}.$$

F-dist

If $W \sim \chi^2_m$ and $X \sim \chi^2_n$ and they are independent, then

$$F = \frac{W/m}{X/n} \text{ is said to have F-dist with d.f. } (m, n).$$

(notation $F_{m,n}$)

Ex Suppose $X_1, \dots, X_5 \overset{iid}{\sim} N(0,1)$.

What is the distribution of

$$Y = \frac{3}{2}\,\frac{\overbrace{X_1^2 + X_2^2}^{2 \text{ d.f.}}}{\underbrace{X_3^2 + X_4^2 + X_5^2}_{3 \text{ d.f.}}}$$

$$\sim F_{2,3}.$$

See that

$$t^2 = \frac{n(\bar{y} - \mu_0)^2}{S^2} \overset{H_0}{\sim} F_{1,n-1} \qquad t = \frac{(\bar{y} - \mu_0)}{\sqrt{S^2/n}}$$

$$= n(\bar{y} - \mu_0)(S^2)^{-1}(\bar{y} - \mu_0) \overset{H_0}{\sim} F_{1,n-1}.$$

Go back to original question.

$\underset{\sim}{X}_1, \underset{\sim}{X}_2, \dots, \underset{\sim}{X}_n \overset{iid}{\sim} N(\underset{\sim}{\mu}, \Sigma)$

$H_0: \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ vs $H_1: \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$

the **test statistic** is

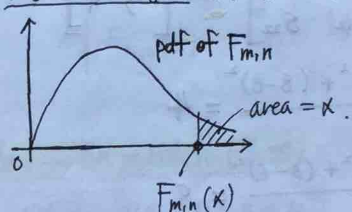
$$T^2 = n(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0)' S^{-1} (\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0), \text{ where } \bar{\underset{\sim}{X}} = \frac{1}{n}\sum_{i=1}^{n} \underset{\sim}{X}_i,$$

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

Under H_0, $\underbrace{\frac{n-p}{(n-1)p} T^2 \sim F_{p,n-p}}_{\text{Need to remember}}$

Equivalently, $\boxed{T^2 \sim \frac{(n-1)p}{n-p} \cdot F_{p,n-p}}$

The rejection region has the form $\{T^2 \geq c\}$ for some c.



If we want the test to have level α, then we reject H_0 if and only if $T^2 \geq F_{m,n}(\alpha) \cdot \frac{(n-1)p}{n-p}$.

The T^2 is called **Hotelling T^2** after the name of Hotelling.

Ex suppose three samples from a bivariate $N_2(\underset{\sim}{\mu}, \Sigma)$ distribution.

$X = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}_{n \times p}$ 3×2 $\quad H_0 : \underset{\sim}{\mu}_0 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$

$\bar{X} = \begin{bmatrix} 8 \\ 6 \end{bmatrix} \quad S = \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix}$

$$S_{11} = \frac{(6-8)^2 + (10-8)^2 + (8-8)^2}{2} = 4$$

$$S_{22} = \frac{(9-6)^2 + (6-6)^2 + (3-6)^2}{2} = 9$$

$$S_{12} = \frac{(6-8)(9-6) + (10-8)(6-6) + (8-8)(3-6)}{2} = -3$$

$$S^{-1} = \frac{1}{36-9}\begin{pmatrix} 9 & 3 \\ 3 & 4 \end{pmatrix} = \frac{1}{27}\begin{pmatrix} 9 & 3 \\ 3 & 4 \end{pmatrix}$$

$$= \begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix}$$

$\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$

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

$$\boxed{T^2 = n(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0)' S^{-1} (\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0)}$$

$$= 3[4\ 3]\begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix}\begin{pmatrix} 4 \\ 3 \end{pmatrix} = \left(5\ \tfrac{24}{9}\right)\begin{pmatrix} 4 \\ 3 \end{pmatrix} =$$

$$20 + 8 = 28.$$

$$= \underset{1\times 2}{[12, 9]}\begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix}\begin{pmatrix} 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 4+1 & 12/9 + \frac{36}{27} \\ 5 & 12/9 + \frac{12}{9} \end{pmatrix}$$

$24/9$

Threshold at 5% level of significance.

Check whether $\boxed{T^2 \geq \frac{(n-1)p}{(n-p)} F_{2,1}(\alpha)}$ ($F_{p,\,n-p}$, $\alpha = 0.05$)

28

$$= 4 F_{2,1}(\alpha) \quad \alpha = 0.05$$

Table

$$n \overset{m}{\overline{\Big|\, F_{m,n}(\alpha)}} \quad = 4 \cdot 199.5 = 798.$$

so, We can not reject H_0.

$\underset{\sim}{X}_1, \underset{\sim}{X}_2, \dots, \underset{\sim}{X}_n \sim iid\ N(\underset{\sim}{\mu}, \Sigma)$

$\boxed{H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0 \quad VS \quad H_1 : \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0}$

suppose $\underset{\sim}{Y} = C\underset{\sim}{X} + \underset{\sim}{\nu}$, C is $p \times p$ nonsingular matrix.

Then $\underset{\sim}{Y}_1, \dots, \underset{\sim}{Y}_n \overset{iid}{\sim} N_p(\overset{\beta}{\overbrace{C\underset{\sim}{\mu} + \underset{\sim}{\nu}}}, C\Sigma C')$.

corresponding $H_0 : \underset{\sim}{\beta} = C\underset{\sim}{\mu}_0 + \underset{\sim}{\nu}$ vs $H_1 : \underset{\sim}{\beta} \neq C\underset{\sim}{\mu}_0 + \underset{\sim}{\nu}$

$\boxed{T^2 = n(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0)' S_x^{-1} (\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0)}$, where $S_x = \frac{1}{n-1}\sum_{i=1}^{n}(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})'$

$\boxed{T^2 = n(\bar{\underset{\sim}{Y}} - \underset{\sim}{\beta}_0)' S_y^{-1} (\bar{\underset{\sim}{Y}} - \underset{\sim}{\beta}_0)}$

$$\bar{\underset{\sim}{Y}} = \frac{1}{n}\sum_{i=1}^{n} Y_i = C\bar{\underset{\sim}{X}} + \underset{\sim}{\nu}$$

$$\bar{\underset{\sim}{Y}} - \underset{\sim}{\beta}_0 = C\bar{\underset{\sim}{X}} + \underset{\sim}{\nu} - C\underset{\sim}{\mu}_0 - \underset{\sim}{\nu} = C(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0)$$

so, $\bar{\underset{\sim}{Y}} - \underset{\sim}{\beta}_0 = C(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0)$.

$$S_y = \frac{1}{n-1}\sum_{i=1}^{n}(\underset{\sim}{y}_i - \bar{\underset{\sim}{y}})(\underset{\sim}{y}_i - \bar{\underset{\sim}{y}})' = C S_x C'$$

$$T^2 = n(\bar{\underset{\sim}{y}} - \underset{\sim}{\beta}_0)' S_y^{-1}(\bar{\underset{\sim}{y}} - \underset{\sim}{\beta}_0)$$
$$= n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)' C'(C S_x C')^{-1} C(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)$$
$$= n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)' S_x^{-1}(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0).$$

So, the value of T^2 does not depend on change of location or change of scale.

Confidence region

In our case, $\underset{\sim}{x}_1, \dots, \underset{\sim}{x}_n \overset{iid}{\sim} N(\underset{\sim}{\mu}, \Sigma)$ a $100(1-\alpha)$ confidence region is :

$$C = \left\{ \underset{\sim}{x} : n(\underset{\sim}{x} - \bar{\underset{\sim}{x}})' S^{-1}(\underset{\sim}{x} - \bar{\underset{\sim}{x}}) \le \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) \right\}$$

$$= \left\{ \underset{\sim}{x} : \underbrace{(\underset{\sim}{x} - \bar{\underset{\sim}{x}})' S^{-1}(\underset{\sim}{x} - \bar{\underset{\sim}{x}})}_{\text{"ellipsoid"}} \le \underbrace{\frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha)}_{C^2} \right\}.$$

center at $\bar{\underset{\sim}{x}}$

axes are eigenvectors of S

major/minor length for axes are

$c\sqrt{\lambda_i}\, \underset{\sim}{e}_i$.

Exam 2 $\overset{?}{\sim}$ 11/7

Hypothesis testing for normal population mean.

$\underset{\sim}{x}_1, \dots, \underset{\sim}{x}_n \sim N_p(\underset{\sim}{\mu}, \Sigma)$. $H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ vs $H_1 : \underset{\sim}{\mu} \ne \underset{\sim}{\mu}_0$.

Rejection rule : Reject H_0 at level of significance α,

if $\underbrace{n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)' S^{-1}(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)}_{\text{"}T^2\text{"}} > \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$.

Likelihood Ratio test

In general, suppose $X_1, \dots, X_n \overset{iid}{\sim} f(y, \theta)$, where θ is a parameter. Θ is a set of possible values of θ. True value of θ is unknown.

Consider two Hypothesis :

$$\boxed{H_0 : \theta \in \Theta_0 \subseteq \Theta \quad \text{vs} \quad H_1 : \theta \notin \Theta_0}$$

Define the likelihood function on

$L(\theta)$ = joint distribution of $Y_1, \dots, Y_n$

function of parameter $= \prod_{i=1}^{n} f(y_i, \theta)$

Likelihood Ratio $\quad \Lambda = \dfrac{\max_{H_0} L(\theta)}{\max L(\theta)} < 1$.

$\Lambda < 1$.

$$\Lambda = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} < 1$$

We expect Λ to be large when H_0 is true.

So, we will reject H_0 when $\Lambda \leq c$ for some c.

To get the value of c, we simplify the expression of Λ and find a statistic $T(y_1, \dots, y_n)$ such that the distribution of $\underbrace{T}_{\text{under } H_0}$ is known and $\{\Lambda \leq c\} \Leftrightarrow \{T \geq c'\}$ for some c'

Then obtain the value of c' using the distribution of T and ensuring $\mathbb{P}(T > c') = \alpha$

Ex (univariate) $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = f(x \mid \mu, \sigma^2)$

Suppose $y_1, \dots, y_n \overset{iid}{\sim} N(\mu, \sigma^2)$, μ, σ^2 are both unknown.

$\boxed{H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0}$

$\theta = (\mu, \sigma^2)$, $\Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$

$\Theta_0 = \{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\}$.

$$\boxed{L(\mu, \sigma^2) = (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right)}$$

$\ell(\mu, \sigma^2) = \log L(\mu, \sigma^2)$, called log-likelihood

$= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2$

Then, maximize $\ell(\mu_0, \sigma^2)$ with respect to σ^2

$$\ell(\mu, \sigma^2) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2.$$

$$\frac{d\ell(\mu_0, \sigma^2)}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu_0)^2 \quad — \circledast$$

Set the $\circledast$ to 0.

$\frac{n}{2\sigma^2} = \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu_0)^2$

$\hat{\sigma}^2 = \cancel{\frac{2\sigma^4}{2\sigma}} = \frac{1}{n} \sum_{i=1}^n (y_i - \mu_0)^2$

$$\boxed{\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mu_0)^2}$$

"check second derivative"

if $d^2\ell < 0$, ℓ has local max

$$L(\mu_0, \hat{\sigma}_0^2) = (\hat{\sigma}_0^2)^{-n/2} e^{-n/2} (2\pi)^{-n/2} = \max_{\theta \in \Theta_0} L(\theta).$$

$\exp\left(-\frac{1}{\frac{2}{n}\sum_{i=1}^n (y_i - \mu_0)^2} \sum (y_i - \mu)^2\right) = e^{-\frac{n}{2}}$

Maximize $L(\mu, \sigma^2)$ with respect to both.

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) = 0$$

$\rightarrow \sum_{i=1}^n (y_i - \mu) = 0.$

$$\hat{\mu} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 = 0.$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

$$L(\hat{\mu}, \hat{\sigma}^2) = (2\pi)^{-n/2} (\hat{\sigma}^2)^{-n/2} e^{-n/2} = \max_{\theta \in \Theta} L(\theta).$$

$$\Lambda = \frac{\max_{\theta \in \Theta_0} L(\mu_0, \sigma^2)}{\max_{\theta \in \Theta} L(\mu, \sigma^2)}.$$

$$\Lambda = \frac{L(\mu_0, \hat{\sigma}_0^2)}{L(\hat{\mu}, \hat{\sigma}^2)} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2}$$

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\mu_0)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y}+\bar{y}-\mu_0)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 + (\bar{y}-\mu_0)^2 + 0.$$

$$= \boxed{\hat{\sigma}^2 + (\bar{y}-\mu_0)^2}$$

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2 = \frac{n}{n-1}\hat{\sigma}^2$$

$S^2 = \frac{n}{n-1}\hat{\sigma}^2$

$\hat{\sigma}^2 = \frac{n-1}{n}S^2$

$$\Lambda \le c \iff \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \le c^{2/n} \iff \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} \ge c^{-2/n}$$

$\frac{\hat{\sigma}^2 + (\bar{y}-\mu_0)^2}{\hat{\sigma}^2}$

$$\iff 1 + \frac{(\bar{y}-\mu_0)^2}{\hat{\sigma}^2} \ge c^{-2/n}$$

$$\iff \underbrace{\frac{n(\bar{y}-\mu_0)^2}{(n-1)S^2}}_{T} \ge \underbrace{\left(c^{-2/n}-1\right)(n-1)}_{c'}$$

$\hat{\sigma}^2 = \frac{n-1}{n}S^2$

$S^2 = \frac{n}{n-1}\hat{\sigma}^2$

$$T = \frac{(\sqrt{n}(\bar{y}-\mu_0))^2}{S^2} = \frac{\left(\frac{\sqrt{n}(\bar{y}-\mu_0)}{\sigma}\right)^2}{(n-1)S^2/\sigma^2}$$

$\frac{(n-1)S^2}{\sigma^2}$

under H_0

- $\sqrt{n}(\bar{y}-\mu_0) \overset{\text{central limit thm}}{\sim} N(0,\sigma^2) = \sigma N(0,1)$
- $\bar{y} \sim N(\mu, \frac{1}{n}\sigma^2)$

$\frac{\sqrt{n}(\bar{y}-\mu_0)}{\sigma} \sim N(0,1).$

$$T = \frac{\left(\frac{\sqrt{n}(\bar{y}-\mu_0)}{\sigma}\right)^2 \sim \chi^2_1}{(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}}$$

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}(y_i-\bar{y})^2 \sim \chi^2_{n-1}.$$

Also, **Num.** and **denom.** are independent.

So, $\boxed{T \sim F_{1,n-1}}$ under H_0.

Reject region: Reject H_0 if $T \ge F_{1,n-1}(\alpha)$.
(at level α)

$\frac{n(\bar{y}-\mu_0)^2}{S^2}$

$\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right)^{n/2} \le c$

$\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \le c^{2/n}$ $\quad \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2} \le c \quad \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \le c^{2/n}$

$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} \ge c^{-2/n}$

- Student's thm

If $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ and $\bar{X}$ (sample mean) and S^2 (sample variance) $= \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$, then

1) $\bar{X} \sim N(\mu, \frac{1}{n}\sigma^2)$
2) $\bar{X}$ and S^2 are independent.
3) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
4) $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}.$

$\hat{\sigma}^2 = \frac{n-1}{n}S^2$

LRT (Multivariate case)

$f_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right]$

$x_1, \dots, x_n \overset{iid}{\sim} N_p(\mu, \Sigma)$. Both μ, Σ are unknown.

$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$

Find Λ and perform the LRT.

$L(\mu, \Sigma) = \prod_{i=1}^{n} f(x_i, \mu, \Sigma)$ $\qquad x'S^{-1}x = tr(S^{-1}xx')$

$= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left[-\frac{1}{2}\underbrace{\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)}_{B}\right]$

$\geq tr\left(\hat\Sigma^{-1}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'\right)$

$= tr(\hat\Sigma^{-1}(n\hat\Sigma)) = np$

MLE $\hat\mu = \bar{x}$,

$\frac{n-1}{n}S = \underset{p\times p}{\hat\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'$.

$\boxed{L(\hat\mu, \hat\Sigma) = (2\pi)^{-\frac{np}{2}} |\hat\Sigma|^{-n/2} e^{-np/2}} = \max_{\theta\in\Theta} L(\hat\mu, \hat\Sigma)$

Find $\hat\Sigma_0$, which maximizes

$\frac{1}{|\Sigma|^{n/2}} \exp\left[-\frac{1}{2} tr(\Sigma^{-1}B)\right]$, $B = \sum_{i=1}^{n}(x_i-\mu_0)(x_i-\mu_0)'$

$\hat\Sigma_0 = \frac{1}{n}B$ (check notes).

$= \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)(x_i-\mu_0)'$

$L(\mu_0, \hat\Sigma_0) = (2\pi)^{\frac{-np}{2}} |\hat\Sigma_0|^{-n/2} e^{-np/2} = \max_{\theta\in\Theta_0} L(\mu_0, \hat\Sigma_0)$

$\Lambda = \frac{L(\mu_0, \hat\Sigma_0)}{L(\hat\mu, \hat\Sigma)} = \left(\frac{|\Sigma|}{|\Sigma_0|}\right)^{n/2} \leq C$

Result: $\Lambda = \left(1 + \frac{T^2}{n-1}\right)^{-2/n}$, where T^2 is the Hotelling T^2.

If we show the above, then $\Lambda \leq C \Leftrightarrow T^2 \geq c'$.

$\frac{P(n-1)}{n-p} F_{p,n-p}(\alpha)$

Matrix fact

If $A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]$, then

$|A| = |A_{11}| \cdot |A_{22} - A_{21}A_{11}^{-1}A_{12}|$.

$= |A_{22}| \cdot |A_{11} - A_{12}A_{22}^{-1}A_{21}|$

check

Exercise from HW.

$\begin{pmatrix} I & 0 \\ C & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} I & B \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{12}A_{11}^{-1}A_{21} \end{pmatrix}$

$(A_{11}^{-1} - A_{21})$

$\begin{pmatrix} A_{11} & A_{12} \\ BA_{11} + A_{21} & BA_{12} + A_{22} \end{pmatrix} \begin{pmatrix} I & B \\ 0 & I \end{pmatrix}$

$A_{11}B + A_{12} = 0$

$B = A_{11}^{-1} - A_{12}$.

$$A = \left(\begin{array}{c|c} \sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})' & \sqrt{n}(\bar{x}-\mu_0) \\ \hline \sqrt{n}(\bar{x}-\mu_0)' & -1 \end{array}\right)$$

$$\quad |A_{11}| \qquad \cdot \qquad |A_{22}-A_{12}A_{11}^{-1}A_{21}|$$

$$|A| = \left|\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'\right| \cdot \left|-1-n(\bar{x}-\mu_0)(A_{11})^{-1}(\bar{x}-\mu_0)'\right|$$

$$= -1\left|\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'+n(\bar{x}-\mu_0)(\bar{x}-\mu_0)'\right|$$

by ⊛

$$= -\underbrace{\left|\sum_{i=1}^{n}(x_i-\mu_0)(x_i-\mu_0)'\right|}_{|n\hat{\Sigma}_0|} = -\underbrace{\left|\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'\right|}_{|n\hat{\Sigma}|}\left(1+\frac{T^2}{n-1}\right)$$

$$\left(\begin{array}{l} T^2 = n(\bar{x}-\mu_0)'S^{-1}(\bar{x}-\mu_0) \\ \hat{\Sigma} = \frac{n-1}{n}S = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})' \end{array}\right.$$

$$\Rightarrow \left(1+\frac{T^2}{n-1}\right) = \frac{|\hat{\Sigma}_0|}{|\hat{\Sigma}|} = \Lambda^{-2/n}$$

$$\Lambda = \left(1+\frac{T^2}{n-1}\right)^{-2/n}$$

$$⊛ \quad \sum_{i=1}^{n}(x_i-\mu_0)(x_i-\mu_0)' = \sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu_0)(x_i-\bar{x}+\bar{x}-\mu_0)'$$
$$= \sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'+n(\bar{x}-\mu_0)(\bar{x}-\mu_0)'$$

Result 5.1 Let $X_1, \dots, X_n$ be a random sample from an $N_p(\mu, \Sigma)$ population. Then the test in (5-7) $n(\bar{X}-\mu_0)'S^{-1}(\bar{X}-\mu_0) > \frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)$ based on T^2 is equivalent to the likelihood ratio test of $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$.

because $\Lambda^{2/n} = \left(1+\frac{T^2}{n-1}\right)^{-1} \iff \Lambda = \left(1+\frac{T^2}{n-1}\right)^{-2/n}$

proof Let $(p+1)\times(p+1)$ matrix

$$A = \left[\begin{array}{c|c} \sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})' & \sqrt{n}(\bar{X}-\mu_0) \\ \hline \sqrt{n}(\bar{X}-\mu_0)' & -1 \end{array}\right] = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]$$

By $\boxed{|A| = |A_{22}||A_{11}-A_{12}A_{22}^{-1}A_{21}| = |A_{11}||A_{22}-A_{21}A_{11}^{-1}A_{12}|}$.

$$|A| = \left|-\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})'-n(\bar{X}-\mu_0)(\bar{X}-\mu_0)'\right|$$
$$= (-1)\left|\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})'+n(\bar{X}-\mu_0)(\bar{X}-\mu_0)'\right|$$
$$= \underbrace{\left|\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})'\right|}_{|A_{11}|}\underbrace{\left|-1-n(\bar{X}-\mu_0)\left(\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})'\right)^{-1}(\bar{X}-\mu_0)'\right|}_{|A_{22}-A_{12}A_{11}^{-1}A_{21}|}$$

Since by (4-14)

$$\sum_{i=1}^{n}(X_i-\mu_0)(X_i-\mu_0)' = \sum_{i=1}^{n}(X_i-\bar{X}+\bar{X}-\mu_0)(X_i-\bar{X}+\bar{X}-\mu_0)'$$
$$= \sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})'+n(\bar{X}-\mu_0)(\bar{X}-\mu_0)'$$

So, $|A| = (-1)\left|\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})'+n(\bar{X}-\mu_0)(\bar{X}-\mu_0)'\right|$

$$= (-1)\left|\sum_{i=1}^{n}(X_i-\mu_0)(X_i-\mu_0)'\right| = \left|\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})'\right|(-1)\left(1+\frac{T^2}{n-1}\right)$$

$$\Rightarrow \quad |n\hat{\Sigma}_0| = |n\hat{\Sigma}|\left(1+\frac{T^2}{n-1}\right) \Rightarrow \Lambda^{-2/n} = \frac{|\hat{\Sigma}_0|}{|\hat{\Sigma}|} = \left(1+\frac{T^2}{n-1}\right)$$

Hence, $\Lambda = \left(1+\frac{T^2}{n-1}\right)^{-2/n}$. $\qquad \Lambda \leq c \iff T^2 \geq c$.

Here, H_0 is rejected for small value of Λ or, equivalently, large values of T^2. The critical values of T^2 are determined by (5-6)

$$\alpha = \mathbb{P}\left[\underbrace{n(\bar{X}-\mu)'S^{-1}(\bar{X}-\mu)}_{T^2} > \frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)\right]$$

Math A7800 10/31/17

$\boxed{\text{NAC } 5/143}$ (Tuesday) real analysis I

$\boxed{\text{Nov}/07,17}$ by Today , $\boxed{\text{Dec}/19/17}$ ~ Final Calculator

Last Time

Hypothesis testing

$\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n \overset{iid}{\sim} N_p(\underset{\sim}{\mu}, \Sigma)$ "both parameters unknown"

$H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ vs $H_1 : \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$

Hotelling T^2 $= n(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0)' S^{-1} (\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}_0)$

Null distribution $T^2 \sim \frac{p(n-1)}{n-p} F_{p,n-p}$ under H_0

So, we would reject H_0 at α level of significance,

if $T^2 > \frac{p(n-1)}{n-p} F_{p,n-p}(\alpha)$.

Also, the LRT

LRT : reject H_0 when LR is small (Likelihood Ratio Test)

$$\Lambda = \frac{\max_{H_0} L(\theta)}{\max L(\theta)} = LR$$

We saw that in our case LRT $\Leftrightarrow$ Hotelling T^2 test

It turned out that

$$\Lambda^{2/n} = \left(1 + \frac{T^2}{n-1}\right)^{-1} = \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}$$

So, $\Lambda \leq c \Leftrightarrow T^2 \geq c'$ for some c'.

Note: LRT is popular, because even if we do not have the exact distribution of Λ (in general case).

under 'mild' condition

$-2\log\Lambda$ approx $\sim \chi^2_{V-V_0}$ under H_0, where

$\begin{cases} V = \#\text{ independent parameters in } \theta. \quad \text{dimention} \\ V_0 = \#\text{ independent parameters in } \theta \text{ under } H_0. \end{cases}$

Ex $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n$ iid with mean $\underset{\sim}{\mu}$, covariance matrix Σ (but not normal)

$H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ vs $H_1 : \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$.

Then, we use the approximate LRT which rejects H_0, when $\boxed{-2\log\Lambda > \chi^2_{V-V_0}(\alpha)}$

$\underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$, $\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ & \ddots & \\ & & \sigma_{pp} \end{pmatrix}$ "numbers of unknown" parameters in Σ.

$V = p + \frac{p(p+1)}{2}$, $V_0 = \frac{p(p+1)}{2}$

So, $V - V_0 = p$

Result When the sample size n is large, under the null hypothesis H_0,

$$-2\ln\Lambda = -2\ln\left(\frac{\max_{\theta\in\Theta_0} L(\theta)}{\max_{\theta\in\Theta} L(\theta)}\right)$$ is, approximately, a $\chi^2_{V-V_0}$ random variable.

Here, the degrees of freedom are $V - V_0 = (\text{dimension of } \Theta) - (\text{dimension of } \Theta_0)$.

Confidence Region for $\underset{\sim}{\mu}$

(*) $- \underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n \overset{iid}{\sim} N_p(\underset{\sim}{\mu}, \Sigma)$, $\underset{\sim}{\mu}, \Sigma$ are unknown.

I want a region which will contain the unknown $\underset{\sim}{\mu}$ with high prob.

A region $R(X)$ is called a $100(1-\alpha)\%$ confidence region of $\underset{\sim}{\mu}$ if

$$P_{\underset{\sim}{\mu}}(\underset{\sim}{\mu} \in R(X)) = 1-\alpha.$$

In case of (*), $R(X)$ can be taken to be

$$\left\{ \underset{\sim}{x} : n(\underset{\sim}{x} - \bar{\underset{\sim}{x}})' S^{-1} (\underset{\sim}{x} - \bar{\underset{\sim}{x}}) \le \frac{p(n-1)}{n-p} F_{p,n-p}(\alpha) \right\}.$$

This is a $100(1-\alpha)\%$ confidence region for $\underset{\sim}{\mu}$.

Note that $R(X) = \{ \mu_0 : \text{the test } H_0 : \underset{\sim}{\mu} = \underset{\sim}{\mu}_0 \text{ vs } H_1 : \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0 \text{ will not be rejected at } \alpha \text{ level} \}$.

Ex $X = \begin{bmatrix} 3 & 6 \\ 2 & 6.5 \\ 3.5 & 5.5 \\ 1.9 & 7.1 \\ 3.1 & 6.3 \end{bmatrix}$ $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_5 \sim N_2(\underset{\sim}{\mu}, \Sigma)$

Find 95% CR for $\underset{\sim}{\mu}$.

$p = 2$, $n = 5$.

$\bar{X} = \begin{bmatrix} 2.7 \\ 6.28 \end{bmatrix}$ $S = \frac{1}{4} X'(I - P_1)X$ $\left(P_1 = \frac{1}{2} \underset{\sim}{1}\underset{\sim}{1}' \right)$

$= \frac{1}{4} \begin{bmatrix} 3 & 2 & 3.5 & 1.9 & 3.1 \\ 6 & 6.5 & 5.5 & 7.1 & 6.3 \end{bmatrix} \Big|^{1/2}$ $F_{2,3}(0.05) = 9.55$

$$S = \frac{1}{4}\left[\begin{pmatrix} 3-2.7 \\ 6-6.28 \end{pmatrix}(3-2.7, 6-6.28)' + \begin{pmatrix} 2-2.7 \\ 6.5-6.28 \end{pmatrix}(2-2.7, 6.5-6.28)' + \begin{pmatrix} 3.5-2.7 \\ 5.5-6.28 \end{pmatrix}\begin{pmatrix} 3.5-2.7 \\ 5.5-6.28 \end{pmatrix}' + \begin{pmatrix} 1.9-2.7 \\ 7.1-6.28 \end{pmatrix}(1.9-2.7, 7.1-6.28)' + \begin{pmatrix} 3.1-2.7 \\ 6.3-6.28 \end{pmatrix}(3.1-2.7, 6.3-6.28)' \right]$$

$$R = \{ \underset{\sim}{x} : 5(\underset{\uparrow\, X_1}{x_1-2.7},\ \underset{\uparrow\, X_2}{x_2-6.28})' S^{-1} (x_1-2.7, x_2-6.28) \le \underbrace{\frac{2\cdot 4}{3}\cdot 9.55}_{C^2} \}.$$

$R(x)$ is an ellipsoid with center $\bar{\underset{\sim}{X}} = \begin{pmatrix} 2.7 \\ 6.28 \end{pmatrix}$

- axes are the eigen vectors of S, and half axes length

$\sqrt{\lambda_i \frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)}$, $i=1,2$, where λ_1, λ_2 are the eigenvalues of S.

$C\sqrt{\lambda_i}\, \underset{\sim}{e}_i$ $\quad C \quad$ $C^2 = \frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)$, $C = \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)}$

- **Simultaneous confidence interval / multiple hypothesis testing.**

Suppose $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n \overset{iid}{\sim} N_p(\underset{\sim}{\mu}, \Sigma)$

$\underset{\sim}{X}$ has p-components $X_1, \dots, X_p$.

Consider linear combinations

$Y = a_1 X_1 + \cdots + a_p X_p$ (a_i's are constants)

$\underset{\sim}{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$, $Y = \underset{\sim}{a}' \underset{\sim}{X} \sim N(\underset{\sim}{a}'\underset{\sim}{\mu}, \underset{\sim}{a}' \Sigma \underset{\sim}{a})$.

Test $H_0: \underset{\sim}{a}'\underset{\sim}{\mu} = \underset{\sim}{a}'\underset{\sim}{\mu}_0$ vs $H_1: \underset{\sim}{a}'\underset{\sim}{\mu} \ne \underset{\sim}{a}'\underset{\sim}{\mu}_0$.

Extend

take k-linear combinations

$$\left.\begin{array}{l} H_0: \underset{\sim}{a}_i\underset{\sim}{\mu} = \underset{\sim}{a}_i\underset{\sim}{\mu}_0 \text{ for } i=1,2,\dots,k \\ H_1: \underset{\sim}{a}_i\underset{\sim}{\mu} \ne \underset{\sim}{a}_i\underset{\sim}{\mu}_0 \end{array}\right\} \text{"multivariates hypotesis test."}$$

corresponding problem in confidence interval.

Find CI for each $\underset{\sim}{a}_i'\underset{\sim}{\mu}$ with total confidence $(1-\alpha)100\%$.

If we have one linear combination $\underset{\sim}{a}'\underset{\sim}{\mu}$, then $100(1-\alpha)\%$ CI is ?

$\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n \overset{iid}{\sim} N_p(\underset{\sim}{\mu}, \Sigma)$

$\underbrace{\underset{\sim}{a}'\underset{\sim}{X}_1}_{Y_1}, \dots, \underbrace{\underset{\sim}{a}'\underset{\sim}{X}_n}_{Y_n} \overset{iid}{\sim} N(\underset{\sim}{a}'\underset{\sim}{\mu}, \underset{\sim}{a}'\Sigma\underset{\sim}{a})$

CI for $\underset{\sim}{a}'\underset{\sim}{\mu}$ $\quad \boxed{\bar{Y} \pm \frac{S_Y}{\sqrt{n}} t_n\left(\frac{\alpha}{2}\right)}$

$\bar{Y} = \underset{\sim}{a}'\bar{\underset{\sim}{X}}$, $S_Y = \sqrt{\underset{\sim}{a}' S \underset{\sim}{a}} \leftarrow S_Y^2 = \underset{\sim}{a}' S \underset{\sim}{a}$

The CI is $\boxed{\underset{\sim}{a}'\bar{\underset{\sim}{X}} \pm \frac{\sqrt{\underset{\sim}{a}'S\underset{\sim}{a}}}{\sqrt{n}} t_{n-1}\left(\frac{\alpha}{2}\right)}$

This is obtained by finding c such that

$$P\left(\left| \frac{n\, \underset{\sim}{a}'(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu})}{\sqrt{\underset{\sim}{a}'S\underset{\sim}{a}}} \right| \le c \right) = 1-\alpha.$$

If we have k-linear combinations $\underset{\sim}{a}_1\underset{\sim}{\mu}_1, \dots, \underset{\sim}{a}_k\underset{\sim}{\mu}$, then two main methods:

- Bonferroni : [we get a rectangle]
- Scheffe : [ellipsoid]

Bonferroni Divided into k-parts ($\alpha_1, \dots, \alpha_k$: usually all equal to $\frac{\alpha}{k}$)

s.t. $\alpha_1 + \dots + \alpha_k = \alpha$.

construct $100(1-\alpha_i)\%$ CI for $\underset{\sim}{a}_i' \underset{\sim}{\mu}$.

Why does it work?

$$I_i = \left[\underset{\sim}{a}_i' \bar{\underset{\sim}{x}} \pm \frac{1}{\sqrt{n}} \sqrt{\underset{\sim}{a}_i' S \underset{\sim}{a}_i}\; t_{n-1}\left(\frac{\alpha}{2}\right) \right]$$

$$\mathbb{P}[\underset{\sim}{a}_i' \underset{\sim}{\mu} \in I_i \text{ for all } i = 1, 2, \dots, k]$$

$$= 1 - \mathbb{P}[\underset{\sim}{a}_i' \underset{\sim}{\mu} \notin I_i \text{ for at least one } i] \geq 1 - \sum_{i=1}^{k} \mathbb{P}(\underset{\sim}{a}_i' \underset{\sim}{\mu} \notin I_i)$$

$$= 1 - \sum_{i=1}^{k} \alpha_i = 1 - \alpha.$$

Scheffe $\Leftrightarrow$ Simultaneous CI.

look at $\max_{\underset{\sim}{a}} \left| \frac{\underset{\sim}{a}'(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})}{\sqrt{\underset{\sim}{a}' S \underset{\sim}{a}}} \right|$ and find c such that

$$\mathbb{P}\left(\max_{\underset{\sim}{a}} \left| \frac{\underset{\sim}{a}'(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})}{\sqrt{\underset{\sim}{a}' S \underset{\sim}{a}}} \right| \leq c \right) = 1 - \alpha.$$

Maximize $\dfrac{(\underset{\sim}{a}'(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}))^2}{\underbrace{\underset{\sim}{a}^2 S \underset{\sim}{a}}_{\underset{\sim}{b}'\underset{\sim}{b}}}$ S is positive definite

Transform $\underset{\sim}{b} = S^{1/2} \underset{\sim}{a}$

$\underset{\sim}{a} = S^{-1/2} \underset{\sim}{b}$

$$\frac{\underset{\sim}{a}'(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})'\underset{\sim}{a}}{\underset{\sim}{b}'\underset{\sim}{b}} = \frac{\underset{\sim}{b}' S^{-1/2} (\bar{x} - \mu)(\bar{x} - \mu)' S^{-1/2} \underset{\sim}{b}}{\underset{\sim}{b}'\underset{\sim}{b}}$$

$$\max_{\underset{\sim}{b}} \frac{\underset{\sim}{b}' A \underset{\sim}{b}}{\underset{\sim}{b}'\underset{\sim}{b}} = \text{largest eigen value of } A.$$

$$A = S^{-1/2} (\bar{x} - \mu)(\bar{x} - \mu)' S^{-1/2}$$

The largest eigen value is same as that of $(\bar{x} - \mu)' S^{-1} (\bar{x} - \mu)$.

So, Sheffe CI for $\underset{\sim}{a}' \underset{\sim}{\mu}$ is $\left| \frac{\underset{\sim}{a}'(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})}{\sqrt{\underset{\sim}{a}' S \underset{\sim}{a}}} \right| \leq \sqrt{(\bar{x} - \mu)' S^{-1} (\bar{x} - \mu)} *$

$$\sqrt{\frac{p(n-1)}{n(n-p)} F_{p, n-p}(\alpha)}.$$

$$\textbf{CI}: \; \underset{\sim}{a}' \bar{\underset{\sim}{x}} \pm \sqrt{\underset{\sim}{a}' S \underset{\sim}{a}} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p, n-p}(\alpha)}$$

Result Let $X_1, X_2, \dots, X_n$ be a random sample from an $N_p(\underset{\sim}{\mu}, \Sigma)$ population with Σ positive definite. Then, simultaneously for all $\underset{\sim}{a}$, the interval

$$\left(\underset{\sim}{a}' \bar{\underset{\sim}{x}} - \sqrt{\frac{p(n-1)}{n(n-p)} F_{p, n-p}(\alpha)\, \underset{\sim}{a}' S \underset{\sim}{a}} \, , \; \underset{\sim}{a}' \bar{\underset{\sim}{x}} + \sqrt{\frac{p(n-1)}{n(n-p)} F_{p, n-p}(\alpha)\, \underset{\sim}{a}' S \underset{\sim}{a}} \right)$$

will contain $\underset{\sim}{a}' \underset{\sim}{\mu}$ with probability $1 - \alpha$.

Proof $T^2 = n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})' S^{-1} (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}) \leq c^2 \Rightarrow \dfrac{n(\underset{\sim}{a}' \bar{\underset{\sim}{x}} - \underset{\sim}{a}' \underset{\sim}{\mu})^2}{\underset{\sim}{a}' S \underset{\sim}{a}} \leq c^2 \quad \forall \underset{\sim}{a}$

or $\underset{\sim}{a}' \bar{\underset{\sim}{x}} - c\sqrt{\dfrac{\underset{\sim}{a}' S \underset{\sim}{a}}{n}} \leq \underset{\sim}{a}' \underset{\sim}{\mu} \leq \underset{\sim}{a}' \bar{\underset{\sim}{x}} + c\sqrt{\dfrac{\underset{\sim}{a}' S \underset{\sim}{a}}{n}} \quad \forall \underset{\sim}{a}.$

Choosing $c^2 = \dfrac{p(n-1)}{n-p} F_{p, n-p}(\alpha)$ gives intervals that will contain $\underset{\sim}{a}' \underset{\sim}{\mu}$ for all $\underset{\sim}{a}$, with probability $1 - \alpha = \mathbb{P}(T^2 \leq c^2)$.

Practice Midterm 2

1. Problem 5.1
2. Find 95% and 99% confidence region for the mean vector for the data of problem 1.
3. Find the 95% simultaneous (Bonferoni and Sceffe) confidence intervals for μ_1 and μ_2 from the data of problem 1.
4. Show that the Hotelling T^2 statistic in invariant under change of scale, that is, if $Y = CX$ for some constant matrix C, then the Hotelling T^2 statistic for both variables X and Y will be the same.
5. problem 5.13.

$-2\log\Lambda \sim \chi^2_{\frac{v-v_0}{p}}$

$p + \frac{p(p+1)}{2} - \left(\frac{p(p+1)}{2}\right)$

$\underset{\sim}{x}_0$

$(\underset{\sim}{\mu} - \bar{\underset{\sim}{x}})' S^{-1} (\underset{\sim}{\mu} - \bar{\underset{\sim}{x}}) \le \frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)$

p, n, C^2, α.

Math A7800 — 11/02/17, Thur

Last classes — Today's Lecture will not be asked in Exam 2 (11/07)

We saw that $\max_{\underset{\sim}{a} \ne \underset{\sim}{0}} \frac{(\underset{\sim}{a}'(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}))^2}{\underset{\sim}{a}' S \underset{\sim}{a}}$ (S was assumped to be p.d)

$$= (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})' S^{-1} (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})$$

Using this, we get $(1-\alpha)100\%$ CI for $\underset{\sim}{a}'\underset{\sim}{\mu}$ for all possible $\underset{\sim}{a}$.

Result

$$\mathbb{P}\Big[\, \underset{\sim}{a}'\bar{\underset{\sim}{x}} - \sqrt{\tfrac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)} \sqrt{\underset{\sim}{a}' S \underset{\sim}{a}} \le \underset{\sim}{a}'\underset{\sim}{\mu} \le \underset{\sim}{a}'\bar{\underset{\sim}{x}} + \sqrt{\tfrac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)} \sqrt{\underset{\sim}{a}' S \underset{\sim}{a}} \quad \forall \underset{\sim}{a} \ne \underset{\sim}{0} \Big]$$

$$= 1 - \alpha.$$

$\left| \frac{\underset{\sim}{a}'(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})}{\sqrt{\underset{\sim}{a}' S \underset{\sim}{a}}} \right| \le$ Threshold for $(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})' S^{-1} (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu})$ while obtaining $100(1-\alpha)\%$ CR

$$\sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)}$$

Format of two sided CI

$$\underset{\sim}{a}'\underset{\sim}{\mu} \in \underset{\sim}{a}'\bar{\underset{\sim}{x}} \pm \frac{\sqrt{\underset{\sim}{a}' S \underset{\sim}{a}}}{\sqrt{n}} \left(\text{distribution of term}\right).$$

Today

If we want CR for $(\underset{\sim}{a}'\underset{\sim}{\mu}, \underset{\sim}{b}'\underset{\sim}{\mu})$ for all possible $\underset{\sim}{a}$ and $\underset{\sim}{b}$ with confidence $100(1-\alpha)\%$

For one parameter $\underset{\sim}{a}'\underset{\sim}{\mu}$

$$\frac{(\underset{\sim}{a}'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}))^2}{\underset{\sim}{a}'S\underset{\sim}{a}} \le \frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)$$

$$\underset{\sim}{a}'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})\,[\underset{\sim}{a}'S\underset{\sim}{a}]^{-1}\,\underset{\sim}{a}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})$$

$$(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'\underset{\sim}{a}\,[\underset{\sim}{a}'S\underset{\sim}{a}]^{-1}\,\underset{\sim}{a}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})$$

Two parameters $(\underset{\sim}{a}'\underset{\sim}{\mu}, \underset{\sim}{b}'\underset{\sim}{\mu})$

$$\underbrace{(\underset{\sim}{a}'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}),\ \underset{\sim}{b}'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}))\begin{pmatrix} \underset{\sim}{a}'S\underset{\sim}{a} & \underset{\sim}{a}'S\underset{\sim}{b} \\ \underset{\sim}{b}'S\underset{\sim}{a} & \underset{\sim}{b}'S\underset{\sim}{b}\end{pmatrix}\begin{pmatrix}\underset{\sim}{a}'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) \\ \underset{\sim}{b}'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})\end{pmatrix}}_{(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'A(A'SA)^{-1}A'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})} \le \frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha).$$

$A = [\underset{\sim}{a}\ \underset{\sim}{b}]$

claim

$$\mathbb{P}\Big[(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'A(A'SA)^{-1}A'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) \le \frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha) \text{ for all } A_{p\times 2} \ne 0\Big] = 1-\alpha.$$

It is enough to show:

$$\max_{A\ne 0}\ (\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'A(A'SA)^{-1}A'(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}) = (\bar{\underset{\sim}{x}}-\underset{\sim}{\mu})'S^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu}). \quad -\circledast$$

call $S^{1/2}A = B$. so $A = S^{-1/2}B$.

$$\underset{\sim}{z}'A(A'SA)^{-1}A'\underset{\sim}{z} = \underset{\sim}{z}'S^{-1/2}B(B'B)^{-1}B'S^{-1/2}\underset{\sim}{z} = \|P_B S^{-1/2}\underset{\sim}{z}\|^2 \le \underbrace{\|S^{-1/2}\underset{\sim}{z}\|^2}_{\underset{\sim}{z}'S^{-1}\underset{\sim}{z}}$$

(Here, $\|\underset{\sim}{u}\|^2 = \underset{\sim}{u}'\underset{\sim}{u}$)

Linear algebra fact

For a matrix $B \in \mathbb{R}^{p\times q}$ having linearly independent columns, define $P_B = B(B'B)^{-1}B'$

It is the orthogonal projection matrix onto column space of B.

For any vector $\underset{\sim}{y}$,

$$\underset{\sim}{y} = \underbrace{P_B\cdot\underset{\sim}{y}}_{\in \text{ column space of } B} + \underbrace{(I-P_B)\underset{\sim}{y}}_{\text{orthogonal complement of column space of } B}.$$

has 0 dot product.

$P_B^2 = P_B$. so, $P_B(I-P_B) = 0$.

$$\|y\|^2 = \|P_B y\|^2 + \underbrace{\|(I-P_B)y\|^2}_{=0, \text{ if } y = Bu \text{ for some } u.}$$

Thus, $\max_{A \neq 0} (\bar{x}-\mu)'A(A'SA)^{-1}A'(\bar{x}-\mu) = (\bar{x}-\mu)'S^{-1}(\bar{x}-\mu)$.

- **Find half-axes length of an ellipse**

Try to write the equation of an ellipse in the form

$(y-a)'B^{-1}(y-a) \leq 1$, where a is a constant vector and B is a constant p.d matrix.

- This ellipse has center a;
- axes along eigenvectors of B;
- half axes length : $\sqrt{\lambda_i}$, where $\lambda_i, \dots, \lambda_p$ are eigenvalues of B.

- **$100(1-\alpha)\%$ CR of μ.**

$$(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu) \leq \underbrace{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)}_{\frac{C^2}{n}}$$

$$\left(\frac{C^2}{n}\right)^{-1} S^{-1}$$

$$(\mu-\bar{x})'\left(\frac{C^2}{n}S\right)^{-1}(\mu-\bar{x}) \leq 1.$$

- center : $\bar{x}$
- axes : along the eigen vectors of S.
- axes length (half) : $\sqrt{\lambda_i}\,\frac{c}{\sqrt{n}}$, where $\lambda_1, \dots, \lambda_p$ are the eigen values of S.

Ex (Microwave oven Radiation)

$X_1 = \sqrt[4]{\text{radiation measurement when door is closed}}$

$X_2 = \sqrt[4]{\text{radiation measurement when door is open}}$.

$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(\mu, \Sigma)$

Based on 42 sample, $\bar{x} = \begin{pmatrix} 0.564 \\ 0.603 \end{pmatrix}$, $S = \begin{pmatrix} 0.0144 & 0.0117 \\ 0.0117 & 0.0146 \end{pmatrix}$

1. Find 95% CR for μ.
2. Test $H_0 : \mu' = (0.562, 0.589) = \mu_0'$ vs $H_1 : \mu' \neq \mu_0'$ at 5% level.
3. Find 95% simultaneous CI for μ_1, μ_2 using
   (a) Scheffe (or T^2) method
   (b) Bonferroni method.

$[A|I] \rightarrow [I|A^{-1}]$

Check $S^{-1}_{3\times 3}$ or $S^{-1}_{n\times n}$

Sol-1 $n=42$, $p=2$, $\bar{x}$ is given, S is given

$$S^{-1} = \frac{1}{0.0144 \times 0.0146 - 0.0117^2} \begin{bmatrix} 0.0146 & -0.0117 \\ -0.0117 & 0.0144 \end{bmatrix}$$

$$= \begin{pmatrix} 199.0457 & -159.5092 \\ -159.5092 & 196.319 \end{pmatrix}, \quad F_{2,40}(0.05) = 3.23$$

$$\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha) = 0.157$$

CR : (μ_1, μ_2)

$$(\mu_1 - 0.564,\ \mu_2 - 0.603) \begin{pmatrix} 199.0457 & -159.5092 \\ -159.5092 & 196.319 \end{pmatrix} \begin{pmatrix} \mu_1 - 0.564 \\ \mu_2 - 0.603 \end{pmatrix} \leq 0.157$$

$$\left\{ \mu : (\mu-\bar{x})'S^{-1}(\mu-\bar{x}) \leq \frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha) \right\}$$

Sol-2 Plug in $\mu_1 = 0.562$, $\mu_2 = 0.581 = \underset{\sim}{\mu}_0$

and compute the matrix multiplication.

$$0.03927 < 0.157$$

So, H_0 can not be rejected at 5% level.

Sol-3 $\mu_1 = [1,0]\underset{\sim}{\mu}$, $\mu_2 = [0,1]\underset{\sim}{\mu}$

Scheff CI for μ_1

$$\underbrace{0.564}_{\bar{x}_1} \pm \sqrt{\underbrace{0.0144}_{S_{11}}}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)} = [0.516, 0.611]$$

Scheff CI for μ_2

$$\underbrace{0.603}_{\bar{x}_2} \pm \sqrt{\underbrace{0.0146}_{S_{22}}}\sqrt{0.157}$$

$$\underset{\sim}{\bar{x}}_i \pm \sqrt{S_{ii}}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}$$

Bonferroni CI for μ_1

$$\underset{\bar{x}_1}{0.564} \pm \frac{\sqrt{0.0144}\,S_{11}}{\sqrt{n}}\, t_{\underset{n-1}{41}}\left(\overset{\frac{\alpha}{4}}{0.0125}\right)$$

$$t_{n-1}\left(\frac{\alpha}{2p}\right)$$

Bonferroni CI for μ_2

$$\underset{\bar{x}_2}{0.603} \pm \frac{\sqrt{0.0146}\,S_{22}}{\sqrt{n}}\, t_{41}\left(\overset{\frac{\alpha}{4}}{0.0125}\right)$$

$$\underset{\sim}{\bar{x}}_i \pm \frac{\sqrt{S_{ii}}}{\sqrt{n}}\, t_{n-1}\left(\frac{\alpha}{2p}\right)$$

$V = p + \frac{p(p+1)}{2}$ $\quad V_0 = \frac{p(p+1)}{2}$

Math A7800 11/9/17

Last time

We covered "large sample" test for the mean vector.

Set up $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ are iid from a p-variate distribution (not neccessary normal), with mean vector $\underset{\sim}{\mu} = (\mu_1, \ldots, \mu_p)'$ and covariance matrix Σ.

(1) Test $H_0: \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ vs $H_1: \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$ at level α.

(2) Find confidence interval for μ_i with confidence $100(1-\alpha)\%$.

(3) Find simultaneous CI for k linear combinations $\underset{\sim}{a}_i'\underset{\sim}{\mu}$, $i = 1, \ldots, k$, with total confidence $100(1-\alpha)\%$

(4) Find $100(1-\alpha)\%$ CR for $\underset{\sim}{\mu}$.

We answer (1) – (4) under the assumption "n−p is large".

CLT $\boxed{\bar{X} \approx N(\mu, \frac{\sigma^2}{n})}$ $\boxed{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)}$

P $\quad \boxed{\sqrt{n}\,\sigma^{-1/2}(\bar{X}-\mu)} \sim N(0,1)$

In general,

$$\sqrt{n}\,\Sigma^{-1/2}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}) \approx N_p(\underset{\sim}{0}, I_p), \quad \text{(convergence in Distribution)}.$$

Suppose $\underset{\sim}{Y} \sim N_p(\underset{\sim}{0}, I_p)$, then $\underset{\sim}{Y}'\underset{\sim}{Y} \sim \chi^2_p$.

Similarly, $\boxed{n(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}) \approx \chi^2_p}$

$$\left(\sqrt{n}\,\Sigma^{-1/2}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})\right)'\left(\sqrt{n}\,\Sigma^{-1/2}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})\right)$$

Recall If $Y_1, \ldots, Y_k \overset{iid}{\sim} N(0,1)$, then $\sum_{i=1}^{k} Y_i^2 = Y_1^2 + \cdots + Y_k^2$ is said to follows χ^2_k distribution, k is reffered as the degree of freedom.

By Slutsky's thm,

$n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu) \approx \chi^2_p$. ($\frac{p(n-1)}{n-p} F_{p,n-p}$ is for Normal case)

So, $T^2 \sim \chi^2_p$ (approx) when $n-p$ is large.

(1) We will reject $H_0: \mu = \mu_0$ if $T^2(\mu_0) = n(\bar{x}-\mu_0)'S^{-1}(\bar{x}-\mu_0) > \chi^2_p(\alpha)$.

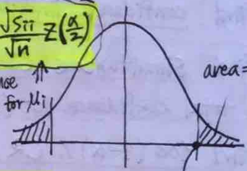
(4) $100(1-\alpha)\%$ CR for μ is

$$\{\mu : (\bar{x}-\mu)'S^{-1}(\bar{x}-\mu) \le \frac{1}{n}\chi^2_p(\alpha)\}$$

(2) μ_i is target estimator $\bar{x}_i$, $\boxed{\bar{x}_i \pm \frac{\sqrt{s_{ii}}}{\sqrt{n}} Z(\frac{\alpha}{2})}$ Interval

Confidence Interval for μ_i

$$\frac{\sqrt{n}(\bar{x}_i - \mu_i)}{\sqrt{\Sigma_{ii}}} \approx N(0,1)$$



By Slutsky, $\frac{\sqrt{n}(\bar{x}_i - \mu_i)}{\sqrt{s_{ii}}} \approx N(0,1)$

(3) — (a) Bonferroni C.I.
(b) Scheffe C.I.

"Bonferroni CI"

(3)-(a) $a_i'\mu$, $\boxed{a_i'\bar{x} \pm \frac{\sqrt{a_i' S a_i}}{\sqrt{n}} Z\left(\frac{\alpha}{2k}\right)}$ for $i = 1, 2, \dots, k$

(3)-(b). Recall $\max_{a \ne 0} \frac{(a'(\bar{x}-\mu))^2}{a' S a} = (\bar{x}-\mu)'S^{-1}(\bar{x}-\mu)$.

"Scheffe CI"

So, Simultaneously for all a, $a'\mu$ lies in $\boxed{a'\bar{x} \pm \sqrt{a'Sa}\sqrt{\frac{1}{n}\chi^2_p(\alpha)}}$ with probability $1-\alpha$.

In particular, $a_i'\mu$ lies in $\boxed{a_i'\bar{x} \pm \sqrt{a_i'Sa_i}\sqrt{\frac{1}{n}\chi^2_p(\alpha)}}$.
(3)-(b).

Ex A music teacher interviewed 96 students and recorded data on the following features.

Variables	$\bar{x}$: sample mean	$\sqrt{s_{ii}}$: Sample standard deviation
Melody	28.3	5.3
Harmony	26.2	3.1
Tempo	35.9	5.4
Meter	34.5	3.9
Phrasing	23.6	3.2
Balance	22.4	4.5
Style	22.8	3.8

(1) Find 95% CR for μ.

(2) Find 95% Simultaneous CI (a) Bonferroni (b) Sheffe for μ_i.

(3) Find CI for $3\mu_1 + 4\mu_3 - 7\mu_7$

(1) $\{\mu : (\bar{x}-\mu)'S^{-1}(\bar{x}-\mu) \le \frac{1}{7}\chi^2_7(0.05)\}$

(3) $a = \begin{bmatrix} 3 \\ 0 \\ 4 \\ 0 \\ 0 \\ 0 \\ -7 \end{bmatrix}$ $\quad a'\bar{x} \pm \frac{\sqrt{a'Sa}}{\sqrt{n}} Z\left(\frac{\alpha}{2}\right)$ 1.96. 0.025.

1,2,3,4

$\frac{4(5)}{2}$

$\underline{(2)-(a)}$ For μ_i, we use $\bar{x}_i \pm \frac{\sqrt{S_{ii}}}{\sqrt{n}} \underbrace{Z\left(\frac{\alpha}{14}\right)}_{2.69}$, $n=96$, 2×7

"Bonferroni"

$\underline{(2)-(b)}$ For μ_i, we use

"Scheffe" $\bar{x}_i \pm \frac{\sqrt{S_{ii}}}{\sqrt{n}} \sqrt{\underbrace{\chi^2_7(0.05)}_{14.07}}$

$n=96$

Missing Data

Suppose $\underset{\sim}{X}_1, \dots, \underset{\sim}{X}_n$ are from (iid) $N_p(\underset{\sim}{\mu}, \Sigma)$.

But the data matrix X has some missing values.

$$X = \begin{bmatrix} ? & 0 & 3 \\ 7 & 2 & 6 \\ 5 & 1 & 2 \\ ? & ? & 5 \end{bmatrix}$$

Goal: Find MLE for $\underset{\sim}{\mu}, \Sigma$.

We use an iterative Prediction – Estimation method, which converges to the MLE.

Step1 start with a naive estimator of $\underset{\sim}{\mu}, \Sigma$.

Step2 Predict the missing values based on the current $\hat{\underset{\sim}{\mu}}, \hat{\Sigma}$

Stop when old $(\hat{\mu}, \hat{\Sigma}) \approx$ new $(\hat{\mu}, \hat{\Sigma})$

Step3 Use the predicted missing values to recompute $\hat{\mu}, \hat{\Sigma}$

Go back to Step2.

Step1 $n=4$, $p=3$. $\underset{\sim}{\mu}' = (\mu_1, \mu_2, \mu_3)'$, $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{pmatrix}$

$$\begin{cases} \hat{\mu}_1 = 6 \\ \hat{\mu}_2 = 1 \\ \hat{\mu}_3 = 4 \end{cases}$$

For $\hat{\Sigma}$, replace the missing values by corresponding sample mean.

$$\hat{X} = \begin{pmatrix} 6 & 0 & 3 \\ 7 & 2 & 6 \\ 5 & 1 & 2 \\ 6 & 1 & 5 \end{pmatrix}$$

$\begin{pmatrix} 6 & 0 \\ 7 & 2 \end{pmatrix}$ $\begin{pmatrix} 6 \\ 1 \end{pmatrix}$

$0+1$ $\begin{pmatrix} 0 \\ -1 \end{pmatrix}(0,-1) + \begin{pmatrix} 1 \\ 1 \end{pmatrix}(1,1)$

$0 + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\frac{1}{2}$

$$S_{11} = \frac{(6-6)^2 + (7-6)^2 + (5-6)^2 + (6-6)^2}{4} = \frac{1}{2}$$

$$S_{22} = \frac{(0-1)^2 + (2-1)^2 + (1-1)^2 + (1-1)^2}{4} = \frac{1}{2}$$

$$S_{33} = \frac{1}{4}(1+4+4+1) = \frac{5}{2} \qquad S_{23} = \frac{1}{4}(1+2+0+0) = \frac{3}{4}$$

$$S_{12} = \frac{1}{4}(0+1+0+0) = \frac{1}{4}$$

$$S_{13} = \frac{1}{4}(0+2+2+0) = 1$$

$$\hat{\mu} = \begin{pmatrix} 6 \\ 1 \\ 4 \end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & 1 \\ 1/4 & \frac{1}{2} & 3/4 \\ 1 & 3/4 & 5/2 \end{pmatrix}$$

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

Step2 $Y = \begin{pmatrix} X \\ \hline 0 \\ 3 \end{pmatrix} \sim N\left(\begin{pmatrix} 6 \\ \hline 1 \\ 4 \end{pmatrix}, \left(\begin{array}{c|cc} 1/2 & 1/4 & 1 \\ \hline 1/4 & 1/2 & 3/4 \\ 1 & 3/4 & 5/2 \end{array}\right) \right)$

Use conditional dist of Y_1 given $\begin{pmatrix} Y_2 \\ Y_3 \end{pmatrix}$

$$\underset{\downarrow}{\underline{\text{mean}}} = \underset{\hat{\mu}_1}{6} + \left(\frac{1}{4}, 1\right) \begin{pmatrix} 1/2 & 3/4 \\ 3/4 & 5/2 \end{pmatrix}^{-1} \begin{pmatrix} \overset{y_2}{0} - \overset{\hat{\mu}_2}{1} \\ \underset{y_3}{3} - \underset{\hat{\mu}_3}{4} \end{pmatrix} = \tilde{\mu}^{(1)} + \hat{S}_{12} \hat{S}_{22}^{-1} \left(\tilde{x}_j^{(2)} - \tilde{\hat{\mu}}^{(2)} \right)$$

$$\mathbb{E}\left(Y_1 \middle| \begin{pmatrix} Y_2 \\ Y_3 \end{pmatrix} \right)$$

$$\mathbb{E}\left(Y_1^2 \middle| \begin{pmatrix} Y_2 \\ Y_3 \end{pmatrix} \right) = Var\left(Y_1 \middle| \begin{pmatrix} Y_2 \\ Y_3 \end{pmatrix} \right) + \left(\mathbb{E}\left(Y_1 \middle| \begin{pmatrix} Y_2 \\ Y_3 \end{pmatrix} \right) \right)^2$$

$$= \frac{1}{2} - \left(\frac{1}{4} \ 1 \right) \begin{pmatrix} 1/2 & 3/4 \\ 3/4 & 5/2 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{4} \\ 1 \end{pmatrix} + \left(\text{mean} \right)^2 .$$

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \\ 5 \end{pmatrix} \sim N_3\left(\begin{pmatrix} 6 \\ 1 \\ 4 \end{pmatrix}, \left(\begin{array}{cc|c} 1/2 & 1/4 & 1 \\ 1/4 & 1/2 & 3/4 \\ \hline 1 & 3/4 & 5/2 \end{array}\right) \right)$$

Use conditional dist of $\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$ given Z_3.

$$\underset{\uparrow}{\underline{\text{mean}}} = \begin{pmatrix} 6 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 3/4 \end{pmatrix} \left(\frac{2}{5} \right) (5-4).$$

$$\mathbb{E}\left(\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \middle| \underset{5}{\overset{}{Z_3}} \right)$$

$$\mathbb{E}\left(\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} (Z_1, Z_2) \middle| Z_3 \right) = \underline{\mathbb{E}\left(\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \middle| Z_3 \right) \mathbb{E}\left(\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \middle| Z_3 \right)'}$$

$$+ \underbrace{\begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{pmatrix} - \begin{pmatrix} 1 \\ 3/4 \end{pmatrix} \left(\frac{2}{5} \right) \left(1, \frac{3}{4} \right)}_{S_{11}^{-1}} .$$

$$= \left(\mathbb{E}(Y_1 | Z_2) \right)^2$$

$$cov\left(\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \middle| Z_3 \right)$$

$$cov(Z_1, Z_2 \mid Z_3)$$

<u>Conditional Distribution</u>

Suppose $\underset{p\times 1}{\tilde{X}} = \begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{pmatrix} \begin{array}{l} q\times 1 \\ (p-q)\times 1 \end{array} \sim N_p(\tilde{\mu}, \Sigma)$, where

$$\tilde{\mu} = \begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix} \begin{array}{l} q\times 1 \\ (p-q)\times 1 \end{array}, \quad \Sigma = \left(\begin{array}{c|c} \underset{q\times q}{\Sigma_{11}} & \underset{q\times(p-q)}{\Sigma_{12}} \\ \hline \underset{(p-q)\times q}{\Sigma_{21}} & \underset{(p-q)\times(p-q)}{\Sigma_{22}} \end{array} \right)$$

Given $\tilde{X}_2 = \tilde{x}_2$, the conditional distribution of $\tilde{X}_1$ is

$$N_q\left(\tilde{\mu}_1 + \Sigma_{12} \Sigma_{22}^{-1} (\tilde{x}_2 - \tilde{\mu}_2), \ \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right)$$

Given $\tilde{X}_1 = \tilde{x}_1$, the conditional distribution of $\tilde{X}_2$ is

$$N_{p-q}\left(\tilde{\mu}_2 + \Sigma_{21} \Sigma_{11}^{-1} (\tilde{x}_1 - \tilde{\mu}_1), \ \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right).$$

Ex $\underset{\underset{\sim}{X}}{\begin{pmatrix}X_1\\X_2\\X_3\end{pmatrix}} \sim N_3\left(\underset{\underset{\sim}{\mu}}{\begin{pmatrix}-1\\0\\2\end{pmatrix}}, \underset{\Sigma}{\begin{pmatrix}3&-1&1\\-2&2&0\\1&0&1\end{pmatrix}}\right)$

Find conditional distribution of

(a) $\begin{pmatrix}X_1\\X_3\end{pmatrix} \mid X_2 = 5$.

(b) $X_1 \mid (X_2=5, X_3=2)$

(a) $\begin{pmatrix}X_1\\X_3\\X_2\end{pmatrix} \sim N_3\left(\begin{pmatrix}-1\\2\\0\end{pmatrix}, \begin{pmatrix}3&1&-1\\1&1&0\\-1&0&2\end{pmatrix}\right)$ with Σ_{11}, Σ_{12}, Σ_{21}, Σ_{22}

$$\begin{pmatrix}X_1\\X_3\end{pmatrix} \sim N_2\left(\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2),\ \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

$$= N_2\left(\begin{pmatrix}-1\\2\end{pmatrix}+\begin{pmatrix}-1\\0\end{pmatrix}\left(\frac{1}{2}\right)(5-0),\ \begin{pmatrix}3&1\\1&1\end{pmatrix}-\begin{pmatrix}-1\\0\end{pmatrix}\left(\frac{1}{2}\right)(-1,0)\right)$$

$$= N_2\left(\begin{pmatrix}-7/2\\2\end{pmatrix}, \begin{pmatrix}5/2&1\\1&1\end{pmatrix}\right)$$

(b) $\begin{pmatrix}X_1\\X_2\\X_3\end{pmatrix} \sim N_3\left(\begin{pmatrix}-1\\0\\2\end{pmatrix}, \begin{pmatrix}3&-1&1\\-1&2&0\\1&0&1\end{pmatrix}\right)$ with Σ_{11}, Σ_{12}, Σ_{21}, Σ_{22}

$$X_1 \sim N_1\left(\underset{\sim}{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}\underbrace{(\underset{\sim}{x}_2-\underset{\sim}{\mu}_2)}_{\begin{pmatrix}x_2-\mu_2\\x_3-\mu_3\end{pmatrix}},\ \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

$$= N_1\left(-1+(-1,1)\begin{pmatrix}2&0\\0&1\end{pmatrix}^{-1}\begin{pmatrix}5-0\\-2-2\end{pmatrix},\ 3-(-1,1)\begin{pmatrix}2&0\\0&1\end{pmatrix}^{-1}\begin{pmatrix}-1\\1\end{pmatrix}\right)$$

$$= N_1\left(\frac{-15}{2}, \frac{3}{2}\right).$$

Math A7800 — 11/14/17, Tue

Tuesday is Friday schedule (11/21)

(12/7) Thurs

Missing Data

Ex $X = \begin{pmatrix}?&0&3\\7&2&6\\5&1&2\\?&?&5\end{pmatrix}$ $\quad X_1,\dots,X_4 \sim N_3(\mu,\Sigma)$.

want MLE for μ and Σ.

Step 1 Get an initial $\tilde{\mu}, \tilde{\Sigma}$.

$$\tilde{\mu} = \begin{pmatrix}6\\1\\4\end{pmatrix},\quad \tilde{\Sigma} = \begin{pmatrix}1/2&1/4&1\\1/4&1/2&3/4\\1&3/4&5/2\end{pmatrix}.$$

Usual MLE of μ and Σ are $\bar{X}$ and $\boxed{\frac{1}{n}\sum_{i=1}^{n} \underset{\sim}{X}_i \underset{\sim}{X}_i' - \bar{\underset{\sim}{X}}\bar{\underset{\sim}{X}}'}$.

$$\tilde{\mu}^{new} = \mathbb{E}(\bar{X} \mid \text{known } x_{ij}\text{'s})$$

$$\bar{X} = \begin{pmatrix}\dfrac{x_{11}+7+5+x_{41}}{4}\\[2ex] \dfrac{0+2+1+x_{42}}{4}\\[2ex] \dfrac{3+6+2+5}{4}\end{pmatrix}$$

replace x_{11}, x_{41}, and x_{42} with their conditional expectations to obtain $\tilde{\mu}^{new}$

$$X_{11} \mid (X_{12}, X_{13}) \sim N_1\Bigg(\underbrace{\underset{\mu_1}{6} + \left(\tfrac{1}{4}, 1\right) \overset{S_{22}}{\begin{pmatrix} 1/2 & 3/4 \\ 3/4 & 5/2 \end{pmatrix}}^{-1} \begin{pmatrix} 0-1 \\ 3-4 \end{pmatrix}}_{\text{"mean"}} ,\ \underbrace{\underset{S_{11}}{\tfrac{1}{2}} - \underset{S_{12}}{\left(\tfrac{1}{4}, 1\right)} \left(S_{22} \right)^{-1} \overset{S_{21}}{\begin{pmatrix} 1/4 \\ 1 \end{pmatrix}}}_{\text{"variance"}} \Bigg)$$

$$\begin{pmatrix} X_{12} - \mu_2 \\ X_{13} - \mu_3 \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 1/2 & 1/4 & 1 \\ 1/4 & 1/2 & 3/4 \\ 1 & 3/4 & 5/2 \end{pmatrix}, \quad \begin{array}{cc} S_{11} & S_{12} \\ S_{21} & S_{22} \end{array} \qquad \hat{\mu} = \begin{pmatrix} 6 \\ 1 \\ 4 \end{pmatrix}$$

$$X_{11} \mid (X_{12}, X_{13}) \sim N_1(5.727, 0.09)$$

$$\begin{pmatrix} X_{41} \\ X_{42} \end{pmatrix} \Bigg| X_{43} \sim N_2\left(\underset{\mu_1}{\begin{pmatrix} 6 \\ 1 \end{pmatrix}} + \underset{\Sigma_{12}}{\begin{pmatrix} 1 \\ 3/4 \end{pmatrix}} \underset{\Sigma_{22}^{-1}}{\left(\tfrac{5}{2}\right)^{-1}} \underset{(x_2 - \mu_2)}{(5-4)},\ \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix} - \begin{pmatrix} 1 \\ 3/4 \end{pmatrix} \left(\tfrac{2}{5}\right) \left(1, \tfrac{3}{4}\right) \right)$$

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$\hat{\mu} = \begin{pmatrix} 6 \\ 1 \\ 4 \end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix} 1/2 & 1/4 & 1 \\ 1/4 & 1/2 & 3/4 \\ 1 & 3/4 & 5/2 \end{pmatrix}$$

$$\begin{pmatrix} X_{41} \\ X_{42} \end{pmatrix} \Bigg| X_{43} \sim N_2\left(\begin{pmatrix} 6.4 \\ 1.3 \end{pmatrix}, \begin{pmatrix} 1/10 & -1/10 \\ -1/10 & 11/40 \end{pmatrix} \right)$$

$$\tilde{\mu}^{new} = \begin{pmatrix} \dfrac{\overset{\hat{X}_{11}}{5.272} + 7 + 5 + \overset{\hat{X}_{41}}{6.4}}{4} \\[2ex] \dfrac{0 + 2 + 1 + \overset{\hat{X}_{42}}{1.3}}{4} \\[2ex] \dfrac{3 + 6 + 2 + 5}{4} \end{pmatrix}$$

$$\tilde{\Sigma}^{new} = \frac{1}{4} \sum_{i=1}^{4} \underset{\sim}{x}_i \underset{\sim}{x}_i' - \underbrace{\tilde{\mu}^{new} \tilde{\mu}^{new\prime}}_{\text{"known."}}$$

$$X = \begin{pmatrix} X_{11} & 0 & 3 \\ 7 & 2 & 6 \\ 5 & 1 & 2 \\ X_{41} & X_{42} & 5 \end{pmatrix} \begin{array}{l} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \\ \underset{\sim}{x}_3 \\ \underset{\sim}{x}_4 \end{array}$$

$$\sum_{i=1}^{4} x_i x_i' = \underset{3\times1}{\begin{pmatrix} X_{11} \\ 0 \\ 3 \end{pmatrix}} \underset{1\times3}{(X_{11}, 0, 3)} + \begin{pmatrix} 7 \\ 2 \\ 6 \end{pmatrix} (7, 2, 6) + \begin{pmatrix} 5 \\ 1 \\ 2 \end{pmatrix} (5, 1, 2) + \begin{pmatrix} X_{41} \\ X_{42} \\ 5 \end{pmatrix} (X_{41}, X_{42}, 5).$$

Replace the unknown quantities by their conditional expectations.

$$\tilde{\Sigma}^{new} = \frac{1}{4} \left(\begin{pmatrix} E(X_{11}^2 \mid \cdots) & 0 & 3E(X_{11} \mid \cdots) \\ 0 & 0 & 0 \\ 3E(X_{11} \mid \cdots) & 0 & 9 \end{pmatrix} + \begin{pmatrix} 49 & 14 & 42 \\ 14 & 4 & 12 \\ 42 & 12 & 36 \end{pmatrix} + \begin{pmatrix} 25 & 5 & 10 \\ 5 & 1 & 2 \\ 10 & 2 & 4 \end{pmatrix} + \begin{pmatrix} E\begin{pmatrix} X_{41} \\ X_{42} \end{pmatrix}(X_{41}, X_{42}) & 5E\begin{pmatrix} X_{41} \\ X_{42} \end{pmatrix} \\ 5E(X_{41}, X_{42}) & 25 \end{pmatrix} \right).$$

$$- \tilde{\mu}^{new} \tilde{\mu}^{new\prime}$$

$$E(X_{11}^2 \mid \cdots) = (5.727)^2 + 0.09 \quad \Leftarrow E(X^2) = (E(X))^2 + var(X).$$

$$E\left[\begin{pmatrix} X_{41} \\ X_{42} \end{pmatrix} (X_{41}, X_{42}) \right] = \begin{pmatrix} 6.4 \\ 1.3 \end{pmatrix} (6.4, 1.3) + \begin{pmatrix} 1/10 & -1/10 \\ -1/10 & 11/40 \end{pmatrix}$$

$X \quad X'$

$$E(XY) = cov(X, Y) + E(X)E(Y)$$

$$☆\ E(\underset{\sim}{X}\underset{\sim}{X}') = E(\underset{\sim}{X})E(\underset{\sim}{X}') + var(\underset{\sim}{X})$$

$$Var(X) = E(X^2) - (E(X))^2$$

Chapter 6. Multiple Comparison.

Two sample t-test

setup $X_1, \dots, X_{n_1} \overset{iid}{\sim} N(\mu_1, \sigma^2)$

$Y_1, \dots, Y_{n_2} \overset{iid}{\sim} N(\mu_2, \sigma^2)$

Two sample from normal with possibly different sample mean and sample size.

$H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$

$$T = \frac{\bar{X}-\mu_0}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$$

In this case, the standard test is

$$\boxed{T = \frac{\bar{X}-\bar{Y}}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{n_1+n_2-1}} \text{ d.f. under } H_0$$

$$\frac{n(\bar{X}-\mu_0)^2}{S^2}$$

$$S^2 = \frac{1}{n_1+n_2-1}\left(\sum_{i=1}^{n_1}(X_i-\hat{\mu})^2 + \sum_{i=1}^{n_2}(Y_i-\hat{\mu})^2\right), \text{ where}$$

$$\hat{\mu} = \frac{1}{n_1+n_2}\left(\sum_{i=1}^{n_1} X_i + \sum_{i=1}^{n_2} Y_i\right)$$

So, reject H_0 if $\boxed{|T| > t_{n_1+n_2-1}\left(\frac{\alpha}{2}\right)}$

$$\left.\begin{array}{l}\bar{X} \sim N\left(\mu_1, \frac{\sigma^2}{n_1}\right) \\ \bar{Y} \sim N\left(\mu_2, \frac{\sigma^2}{n_2}\right)\end{array}\right\} \text{ independent}$$

$\boxed{var(X-Y) = var(X) + var(Y)}$

$$\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)$$

under H_0, $\frac{\bar{X}-\bar{Y}}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0,1)$.

$$T = \frac{\frac{\bar{X}-\bar{Y}}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}}{\sqrt{\frac{S^2}{\sigma^2}}}$$

$\frac{\bar{X}-\bar{Y}}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0,1)$, $(n_1+n_2-1)S^2 \sim \chi^2_{n_1+n_2-1}$ — Independent

So, $T \sim t_{n_1+n_2-1}$ under H_0.

Paired t-test

$(X_1, Y_1), \dots, (X_n, Y_n)$ are iid $N_2\left(\begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix}, \Sigma\right)$

before "treatment" (X_n), after "treatment" (Y_n)

$H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$.

Look at $\boxed{D_i = Y_i - X_i}$.

Then $\boxed{D_i \sim N(\underbrace{\mu_2-\mu_1}_{\mu}, \sigma^2)}$

$H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$

So, $D_1, D_2, \dots, D_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$.

Reject H_0 if $\boxed{\left|\frac{\bar{D}}{S_D/\sqrt{n}}\right| > t_{n-1}\left(\frac{\alpha}{2}\right)}$, where $S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i-\bar{D})^2$.

treatment 효과가 ~~없다~~ consist 하다.

Multivariate case (Paired Comparisons)

Suppose $(\boldsymbol{X}_1, \boldsymbol{Y}_1), \dots, (\boldsymbol{X}_n, \boldsymbol{Y}_n)$ are pairs of p-variate vectors corresponding to treatment 1 and 2 applied to n-subjects.

$$\boxed{\boldsymbol{\delta} = \mathbb{E}(\boldsymbol{Y}_1) - \mathbb{E}(\boldsymbol{X}_1)}$$

(1) test $H_0 : \boldsymbol{\delta} = \boldsymbol{0}$ vs $H_1 : \boldsymbol{\delta} \neq \boldsymbol{0}$.

← 효과가 없다.

(2) Confidence region for $\boldsymbol{\delta}$.

(3) Simultaneous CI for δ_i's

$(\boldsymbol{X}_i, \boldsymbol{Y}_i)$'s are i.i.d. But $\boldsymbol{X}_i, \boldsymbol{Y}_i$ are dependent.

Define $\boldsymbol{D}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i$. Then $\boldsymbol{D}_i$'s are iid and

$\boldsymbol{D}_i \sim N_p(\boldsymbol{\delta}, \Sigma)$ for some Σ.

sol

(1) $T^2 = n\bar{\boldsymbol{D}}' S_D^{-1} \bar{\boldsymbol{D}}$ * There are two cases: Normal and Not.

ⓐ reject H_0 if $\boxed{T^2 > \frac{p(n-1)}{n-p} F_{p,n-p}(\alpha)}$, where

$$S_D = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{D}_i - \bar{\boldsymbol{D}})(\boldsymbol{D}_i - \bar{\boldsymbol{D}})'.$$

assume that $\boxed{(\boldsymbol{X}_i, \boldsymbol{Y}_i) \sim N_{2p}(\ ,\ )}$.

large sample test

by

ⓑ The distribution of $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ is not known (may not be normal)

When $n-p$ is large, then T^2 (approximately) follows χ^2_p.

$$\underbrace{\frac{p(n-1)}{(n-p)} F_{p,n-p}}_{\text{Normal case}} \Rightarrow \underbrace{\frac{p(n-1)}{(n-p)} \frac{\chi^2_p/p}{\chi^2_{n-p}/n-p}}_{\text{Non-normal case}} \approx \frac{\chi^2_p}{\underbrace{\chi^2_{n-p}/n-p}_{\to 1}} \approx \chi^2_p.$$

Then, we reject H_0 if $T^2 > \chi^2_p(\alpha)$.

(2) CR for $\boldsymbol{\delta}$.

(a) $\left\{ \boldsymbol{\delta} : (\bar{\boldsymbol{D}} - \boldsymbol{\delta})' S_D^{-1} (\bar{\boldsymbol{D}} - \boldsymbol{\delta}) \le \frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha) \right\}$

(b) $\left\{ \boldsymbol{\delta} : (\bar{\boldsymbol{D}} - \boldsymbol{\delta})' S_D^{-1} (\bar{\boldsymbol{D}} - \boldsymbol{\delta}) \le \frac{1}{n} \chi^2_p(\alpha) \right\}$.

(3) Scheffe CI.

(a) Simultaneous CI for δ_i is

$$\bar{D}_i \pm \sqrt{\frac{(S_D)_{ii}}{n}} \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p}(\alpha)}.$$

(b) $\bar{D}_i \pm \sqrt{\frac{(S_D)_{ii}}{n}} \sqrt{\chi^2_p(\alpha)}$.

Bonferroni CI

(a) Find simultaneous CI for $\underset{(Y_1-X_1)}{\delta_1}, \underset{(Y_3-X_3)}{\delta_3}, \underset{(Y_4-X_4)}{\delta_4}$ (using Bonferroni)

$$\bar{D}_i \pm \sqrt{\frac{(S_D)_{ii}}{n}}\, t_{n-1}\left(\frac{\alpha}{2 \cdot 3}\right)$$

(b) $\bar{D}_i \pm \sqrt{\frac{(S_D)_{ii}}{n}}\, z\left(\frac{\alpha}{2 \cdot 3}\right)$ ← standard Normal.

Suppose for each of n-subjects, we apply q-treatments (one after the other) with treatment effect $\mu_1, \mu_2, \ldots, \mu_q$.

Subjects treatment 1 ...... treatment q

A treatment contrast is a linear combination of $\mu_1, \ldots, \mu_q$, where sum is zero.

For example, $\mu_1 - \mu_2$, $\mu_1 + \mu_3 - \mu_2 - \mu_1$.

first case

$$H_0: \begin{pmatrix} \mu_1-\mu_2 \\ \mu_1-\mu_3 \\ \vdots \\ \mu_1-\mu_q \end{pmatrix} = \underset{\sim}{0} \quad \text{vs} \quad H_1: H_0 \text{ is false}.$$

second case

$$H_0: \begin{pmatrix} \mu_2-\mu_1 \\ \mu_3-\mu_2 \\ \vdots \\ \mu_q-\mu_{q-1} \end{pmatrix} = \underset{\sim}{0} \quad \text{vs} \quad H_1: H_0 \text{ is false}.$$

Math A7800 11/16/17, Thur

Midterm 3 ~ 12/7, Thur.

Repeated measure treatment effect. (a repeated measures Design for Comparing Treatments)

Each of n-subjects is given q treatments

X_{ji} = measurement for jth subject and ith treatment,

$j = 1, 2, \ldots, n$; $i = 1, 2, \ldots, q$

$$\underset{\sim}{X}_j = \begin{pmatrix} X_{j1} \\ \vdots \\ X_{jq} \end{pmatrix}, \ \mathbb{E}(\underset{\sim}{X}_j) = \underset{\sim}{\mu}, \ \operatorname{cov}(\underset{\sim}{X}_j) = \Sigma$$

So, $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n$ are iid with mean $\underset{\sim}{\mu}$ and covariance Σ.

$H_0: \mu_1 = \cdots = \mu_q$ vs H_1: H_0 is false.

Rewrite $H_0: C\underset{\sim}{\mu} = 0$, where

$$C = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix}_{(q-1)\times q}$$

rank$(C) = q-1$.

Note: There are other choices of C

e.g. $\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$ having size $(q-1)\times q$ and rank $= q-1$.

$$\begin{pmatrix} \mu_1-\mu_2 \\ \mu_1-\mu_3 \\ \vdots \\ \mu_1-\mu_q \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & 0 \\ 1 & 0 & \cdots & 0 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{pmatrix} = C\underset{\sim}{\mu}$$

comparing treatments characterized by C

Such C is called contrast matrix.

We need to test $H_0 : C\underset{\sim}{\mu} = \underset{\sim}{0}$ vs $H_1 : C\underset{\sim}{\mu} \neq \underset{\sim}{0}$.

Use T^2 test.

$$T^2 = n(C\underset{\sim}{\bar{x}})'(CSC')^{-1}C\underset{\sim}{\bar{x}}$$

Under H_0, $T^2 \sim \frac{(q-1)(n-1)}{n-q+1} F_{q-1,\, n-q+1}$

Reject H_0 at α level of significance, if

$$T^2 > \frac{(q-1)(n-1)}{n-q+1} F_{q-1,\, n-q+1}(\alpha).$$

Note:

The value of T^2 does not depend on the particular choice of contrast matrix.

Simultaneous (Sheffe or T^2) $100(1-\alpha)\%$ CI for contrasts $\underset{\sim}{c}'\underset{\sim}{\mu}$.

Ans $\quad \underset{\sim}{c}'\underset{\sim}{\bar{x}} \pm \sqrt{\frac{\underset{\sim}{c}'S\underset{\sim}{c}}{n}} \sqrt{\frac{(q-1)(n-1)}{n-q+1} F_{q-1,\, n-q+1}(\alpha)}$

Ex (Example 6.2)

Four treatments applied to each of 19 dogs.

Halothane	High	Low
Present	•	•
absent	•	•

CO_2 pressure

T_1 : H absent, CO_2 high | μ_1

T_2 : H absent, CO_2 low | μ_2

T_3 : H present, CO_2 high | μ_3

T_4 : H present, CO_2 low | μ_4 — measurements are heartbeats.

treatment effects

The contrasts of interest:

$\mu_3 + \mu_4 - \mu_1 - \mu_2$ (H contrast)

$\mu_1 + \mu_3 - \mu_2 - \mu_4$ (CO_2 contrast)

$\mu_1 + \mu_4 - \mu_2 - \mu_3$ (influence of H on CO_2 effect)

$$C = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

$\underset{\sim}{\bar{x}}$ and S are given from data

We compute $\boxed{T^2 = 19(C\bar{\underset{\sim}{X}})'(CSC')^{-1}(C\bar{\underset{\sim}{X}})}$

compare with $\boxed{\frac{3\times 18}{16} F_{3,16}(\alpha)}$.

Read the example carefully. There H_0 is rejected at 5% level. Also, the CI for H contrast does not contain 0, meaning Halothane has a positive effect in both CO_2 conditions.

Comparison of two population means

(Comparing Mean Vectors from Two Populations.)

So, $\boxed{\underset{\sim}{X}_{1,1}, \underset{\sim}{X}_{1,2}, \cdots, \underset{\sim}{X}_{1,n_1}}$ are iid from $N_p(\underset{\sim}{\mu}_1, \Sigma)$

$\boxed{\underset{\sim}{X}_{2,1}, \underset{\sim}{X}_{2,2}, \cdots, \underset{\sim}{X}_{2,n_2}} \overset{iid}{\sim} N_p(\underset{\sim}{\mu}_2, \Sigma)$. covariance 不變

$H_0: \underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2 = \underset{\sim}{\delta}_0$ vs $H_1: \underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2 \neq \underset{\sim}{\delta}_0$. "$\Sigma = \Sigma$"

Suppose the sample mean and sample covariance matrix.

population 1: $\bar{\underset{\sim}{X}}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} \underset{\sim}{X}_{1j}$, $\boxed{S_1 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(\underset{\sim}{X}_{1j} - \bar{\underset{\sim}{X}}_1)(\underset{\sim}{X}_{1j} - \bar{\underset{\sim}{X}}_1)'}$

population 2: $\bar{\underset{\sim}{X}}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} \underset{\sim}{X}_{2j}$, $\boxed{S_1 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(\underset{\sim}{X}_{2j} - \bar{\underset{\sim}{X}}_2)(\underset{\sim}{X}_{2j} - \bar{\underset{\sim}{X}}_2)'}$

combined sample covariance matrix:

$$S_{pooled} = \frac{\sum_{j=1}^{n_1}(\underset{\sim}{X}_{1j} - \bar{\underset{\sim}{X}}_1)(\underset{\sim}{X}_{1j} - \bar{\underset{\sim}{X}}_1)' + \sum_{j=1}^{n_2}(\underset{\sim}{X}_{2j} - \bar{\underset{\sim}{X}}_2)(\underset{\sim}{X}_{2j} - \bar{\underset{\sim}{X}}_2)'}{n_1 + n_2 - 2}$$

$$= \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

$$= \frac{n_1 - 1}{n_1 + n_2 - 2} S_1 + \frac{n_2 - 1}{n_1 + n_2 - 2} S_2$$

$\bar{X}_1 \sim N_p(\mu_1, \frac{1}{n_1}\Sigma)$, $\bar{X}_2 \sim N_p(\mu_2, \frac{1}{n_2}\Sigma)$

$\bar{X}_1$ and $\bar{X}_2$ are independent.

$$\bar{X}_1 - \bar{X}_2 \sim N_p\left(\mu_1 - \mu_2, \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\Sigma\right)$$

$$(\bar{X}_1 - \bar{X}_2 - \underbrace{(\mu_1 - \mu_2)}_{\delta_0})' \left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\Sigma\right]^{-1} (\bar{X}_1 - \bar{X}_2 - \underbrace{(\mu_1 - \mu_2)}_{\delta_0}) \sim \chi^2_p.$$

e.g. $Y \sim N(\mu, \Sigma)$

$(Y-\mu)'\Sigma^{-1}(Y-\mu) \sim \chi^2_p$.

Since Σ is unknown, we replace Σ by pooled.

Define $T^2 = (\bar{X}_1 - \bar{X}_2 - \delta_0)' \left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right) S_{pooled}\right]^{-1} (\bar{X}_1 - \bar{X}_2 - \delta_0)$

under H_0, $T^2 \sim \underbrace{\frac{p(n_1+n_2-2)}{n_1+n_2-p-1} F_{p,\, n_1+n_2-p-1}(\alpha)}_{C}$. $\quad$ ($n-1$) = (n_1+n_2-1)

So, reject H_0 if $T^2 > C$.

$(n_1-1)S_1 \sim W_{n_1-1}(\Sigma)$

$(n_2-1)S_2 \sim W_{n_2-1}(\Sigma)$ $\Big\}$ independent.

$(n_1-1)S_1 + (n_2-1)S_2 = (n_1+n_2-2)S_{pooled} \sim W_{n_1+n_2-2}(\Sigma)$

$\mu_{1i} - \mu_{2i}$: $(\bar{X}_{1i} - \bar{X}_{2i}) \pm C\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right) S^{(ii)}_{pooled}}$ for $i = 1, 2, \dots, p$.

$$T^2 = \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1/2} (\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2))' [S_{pooled}]^{-1} \underbrace{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1/2} (\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2))}_{\sim N_p(0,\Sigma)}$$

$\frac{1}{\sqrt{n}}(\bar{X}-\mu) \sim N(0, \Sigma)$

So, $T^2 = (N_p(0,\Sigma))' \left(\frac{W_{n_1+n_2-2}(\Sigma)}{n_1+n_2-2}\right)^{-1} (N_p(0,\Sigma)) =$

$(N_p(0,\Sigma))' \left(\frac{W_k(\Sigma)}{k}\right)^{-1} (N_p(0,\Sigma)) \sim \frac{p \cdot k}{k+1-p} F_{p,\, k+1-p}(\alpha)$.

Hence, $T^2 \sim \frac{p(n_1+n_2-2)}{n_1+n_2-p-1} F_{p,\, n_1+n_2-p-1}(\alpha)$

Next case "$\Sigma \neq \Sigma$" (also, general case)

(a) unequal variance : $X_{1,1}, \dots, X_{1,n_1}$ are iid with mean μ_1 and covariance Σ_1. (covariance가 틀림)

$X_{2,1}, \dots, X_{2,n_2}$ are iid with mean μ_2 and covariance Σ_2.

$H_0 : \mu_1 - \mu_2 = \delta_0$ vs $H_1 : \mu_1 - \mu_2 \neq \delta_0$

$\frac{1}{\sqrt{n}}(\bar{X}-\mu) \sim N(0, \Sigma)$.

$\bar{X}_1 \overset{CLT}{\approx} N_p(\mu_1, \frac{1}{n_1}\Sigma_1)$

$\bar{X}_2 \overset{CLT}{\approx} N_p(\mu_2, \frac{1}{n_2}\Sigma_2)$ $\Big\}$ independent.

$$\bar{X}_1 - \bar{X}_2 \approx N_p\left(\mu_1 - \mu_2, \frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2\right)$$

$$(\underset{\sim}{\bar{X}}_1-\underset{\sim}{\bar{X}}_2-(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2))'\left(\frac{1}{n_1}\Sigma_1+\frac{1}{n_2}\Sigma_2\right)^{-1}(\underset{\sim}{\bar{X}}_1-\underset{\sim}{\bar{X}}_2-(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2))$$

$$\approx \chi^2_p.$$

If n_1-p, n_2-p are both large,

then replace Σ_i by S_i.

$$\boxed{T^2=(\underset{\sim}{\bar{X}}_1-\underset{\sim}{\bar{X}}_2-\cancel{\underset{\sim}{\mu}_1+\underset{\sim}{\mu}_2})'\left(\frac{1}{n_1}S_1+\frac{1}{n_2}S_2\right)^{-1}(\underset{\sim}{\bar{X}}_1-\underset{\sim}{\bar{X}}_2-\cancel{\underset{\sim}{\mu}_1+\underset{\sim}{\mu}_2}) \underset{\text{approxim.}}{\sim} \chi^2_p}$$

$\underset{\sim}{\delta}_0=\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2$ $\quad\quad \underset{\sim}{\delta}_0$

Reject H_0 if $T^2>\chi^2_p(\alpha)$.

Q1: Find CI for $\underset{\sim}{a}'(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)$

$100(1-\alpha)\%$

Q2: Find simultaneous CI for $\underset{\sim}{a}'(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)$, $\forall\ \underset{\sim}{a}$ with confidence $100(1-\alpha)\%$.

sol Q1 $\quad \underset{\sim}{a}'(\underset{\sim}{\bar{X}}_1-\underset{\sim}{\bar{X}}_2)\sim N\left(\underset{\sim}{a}'(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2),\ \underset{\sim}{a}'\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\Sigma\,\underset{\sim}{a}\right).$

$$\boxed{\frac{\underset{\sim}{a}'(\underset{\sim}{\bar{X}}_1-\underset{\sim}{\bar{X}}_2)-\underset{\sim}{a}'(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\underset{\sim}{a}'S_{pooled}\,\underset{\sim}{a}}}\sim t_{n_1+n_2-2}(\alpha).}$$

For $\Sigma_1=\Sigma_2$,

CI for $\underset{\sim}{a}'(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)$ is $\underset{\sim}{a}'(\underset{\sim}{\bar{X}}_1-\underset{\sim}{\bar{X}}_2)\pm\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\underset{\sim}{a}'S_{pooled}\,\underset{\sim}{a}}\ *\ \left(t_{n_1+n_2-2}(\alpha)\right)$

$$S_{pooled}=\frac{n_1-1}{n_1+n_2-2}S_1+\frac{n_2-1}{n_1+n_2-2}S_2$$

For unequal variance, $\Sigma_1\neq\Sigma_2$, the CI is

$$\underset{\sim}{a}'(\underset{\sim}{\bar{X}}_1-\underset{\sim}{\bar{X}}_2)\pm\sqrt{\frac{1}{n_1}\underset{\sim}{a}'S_1\underset{\sim}{a}+\frac{1}{n_1}\underset{\sim}{a}'S_2\underset{\sim}{a}}\left(Z\left(\frac{\alpha}{2}\right)\right)$$

Normal percentile.

sol Q2 $\quad$ $\Sigma_1=\Sigma_2$

$$\underset{\sim}{a}'(\underset{\sim}{\bar{X}}_1-\underset{\sim}{\bar{X}}_2)\pm\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\underset{\sim}{a}'S_{pooled}\,\underset{\sim}{a}}\ \sqrt{\frac{p(n_1+n_2-2)}{n_1+n_2-p-1}F_{p,\,n_1+n_2-p-1}(\alpha)}$$

When $\Sigma_1\neq\Sigma_2$, then simultaneous CI is

$$\underset{\sim}{a}'(\underset{\sim}{\bar{X}}_1-\underset{\sim}{\bar{X}}_2)\pm\sqrt{\frac{1}{n_1}\underset{\sim}{a}'S_1\underset{\sim}{a}+\frac{1}{n_2}\underset{\sim}{a}'S_2\underset{\sim}{a}}\ \sqrt{\chi^2_p(\alpha)}$$

ANOVA (Analysis of Variance)

There are g treatments, $T_1, T_2, \dots, T_g$.

Treatment T_i is applied to n_i objects.

X_{Li} = measurement corresponding to ith subject of treatment T_i, $L = 1, 2, \dots, g$, $i = 1, 2, \dots, n$.

$$X = \begin{pmatrix} X_{11} \\ \vdots \\ X_{1n_1} \\ X_{21} \\ \vdots \\ X_{2n_2} \\ \vdots \\ X_{g1} \\ \vdots \\ X_{gn_g} \end{pmatrix}_{n\times 1} \qquad n = \sum_{i=1}^{g} n_i . \qquad \begin{pmatrix} x_{11} & \cdots & x_{1n_1} \\ x_{21} & \cdots & x_{2n_2} \\ & \vdots & \\ x_{g1} & \cdots & x_{gn_g} \end{pmatrix}$$

$$X_{Li} = \underbrace{\mu}_{\text{overall average}} + \underbrace{\tau_L}_{\text{treatment effect}} + \underbrace{\varepsilon_{L,i}}_{\text{error randomness.}}$$

We will test

$H_0 : \tau_1 = \tau_2 = \cdots = \tau_g$ VS $H_1 :$ H_0 is false.

$$\bar{X} = \text{combined sample mean} = \frac{\sum_{L=1}^{g}\sum_{i=1}^{n_L} X_{Li}}{n}$$

Within block mean

$$\bar{X}_L = \frac{1}{n_L}\sum_{i=1}^{n_L} X_{Li}$$

Total sum of square $= \sum_{L=1}^{g}\sum_{i=1}^{n_L}(X_{Li}-\bar{X})^2$.

$$= \sum_{L=1}^{g}\sum_{i=1}^{n_L}(X_{Li}-\bar{X}_L+\bar{X}_L-\bar{X})^2$$

$$= \sum_{L=1}^{g}\sum_{i=1}^{n_L}(X_{Li}-\bar{X}_L)^2 + \sum_{L=1}^{g} n_L(\bar{X}_L-\bar{X})^2 + 2\sum_{L=1}^{g}\sum_{i=1}^{n_L}(X_{Li}-\bar{X}_L)(\bar{X}_L-\bar{X}) \to 0$$

// Residual sum of square (within Block) + Treatment sum of square (within Block).

= Residual sum of square (within Block) + Treatment sum of square (within Block).

ANOVA Table

Source	statistic	d.f.	Standardized statistic
treatment	$\sum_{L=1}^{g} n_L(\bar{X}_L-\bar{X})^2$	$g-1$	$\frac{\text{Treat SS}}{g-1}$
residual	$\sum_{L=1}^{g}\sum_{i=1}^{n_L}(\bar{X}_{Li}-\bar{X}_L)^2$	$n-g$	$\frac{\text{Residual SS}}{n-g}$
Total	$\sum_{L=1}^{g}\sum_{i=1}^{n_L}(\bar{X}_{Li}-\bar{X})^2$	$n-1$	Sample variance

$$F = \frac{\frac{\text{treat SS}}{g-1}}{\frac{\text{Residual SS}}{n-g}} \sim F_{g-1,\, n-g} \quad \text{under } H_0.$$

Brooke schasteen-smith

PhD_Applications_Minwoobae.pdf

Math A7800 11/28/17

ANOVA (last time)

Analysis of variance

We are discussing one-way ANOVA.

Recall Here the model was X_{li} is the measurement.

For the i^{th} observation in the l^{th} treatment group,

$l = 1, 2, \ldots, g$ For fixed l, $i = 1, 2, \ldots, n_l$

Total number of measurement $n = \sum_{l=1}^{g} n_l$.

Model for one-way ANOVA

$\tau_1, \ldots, \tau_g$ satisfies $\sum_{l=1}^{g} n_l \tau_l = 0$.

$$\boxed{X_{li} = \underbrace{\mu}_{\text{overall effect}} + \underbrace{\tau_l}_{\text{treatment effect for } l^{th} \text{ treatment group}} + \underbrace{\varepsilon_{li}}_{\text{error (measurement or something else)}}}$$

$E[\varepsilon_{li}] = 0$ and ε_{li}'s are independent.

Additionally, sometimes one assumes $\varepsilon_{li} \sim N(0, \sigma^2)$ iid

Given $\{X_{li}\}$, $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{l=1}^{g}\sum_{i=1}^{n_l} X_{li}$.

$\hat{\tau}_l = \bar{X}_l - \bar{X}$, where $\bar{X}_l = \frac{1}{n_l}\sum_{i=1}^{n_l} X_{li}$

$\hat{\varepsilon}_{li} = X_{li} - \bar{X}_l$.

$X_{Li} = \hat{\mu} + \hat{\tau}_L + \hat{\varepsilon}_{Li}$

Check $\sum_{L=1}^{g} \hat{\tau}_L n_L = 0$.

Express $\bar{X}$ in terms of $\bar{X}_1, \dots, \bar{X}_g$

$$\bar{X} = \frac{n_1 \bar{X}_1 + \cdots + n_g \bar{X}_g}{n_1 + \cdots + n_g}$$

so, $\sum_{L=1}^{g} n_L \hat{\tau}_L = \sum_{L=1}^{g} n_L \bar{X}_L - n\bar{X} = 0$

$\sum_L \sum_i X_{Li}$

ANOVA Table

$\sum_{L=1}^{g} n_L(\bar{X}_L - \bar{X})^2$

Source	Statistic	d.f.	Mean S.S
treatment	$\sum_{L=1}^{g}\sum_{i=1}^{n_L}(\bar{X}_L - \bar{X})^2$	$g-1$	$\frac{\text{treat SS}}{g-1}$
residual	$\sum_{L=1}^{g}\sum_{i=1}^{n_L}(\bar{X}_{Li} - \bar{X}_L)^2$	$n-g$ $\sum_{L=1}^{g}(n_L-1)$	$\frac{\text{Residual SS}}{n-g}$
Total (corrected) Sum of Square	$\sum_{L=1}^{g}\sum_{i=1}^{n_L}(\bar{X}_{Li} - \bar{X})^2$	$n-1$	sample variance

$F = \frac{\text{treat SS}}{g-1}$

$\sum_{L=1}^{g} n_L - 1$

$\sum_{L=1}^{g} n_L - g = n-g$.

TSS (total sum of square) "uncorrected"

$= \sum_{L=1}^{g}\sum_{i=1}^{n_L} X_{Li}^2$

$=$ total S.S corrected $+ n\bar{X}^2$.

$$F = \frac{\text{treat SS}}{\text{res. SS}} \cdot \frac{\sum_{L=1}^{g} n_L - g}{g-1}.$$

For testing $H_0: \tau_1 = \cdots = \tau_g = 0$ VS $H_1: H_0$ is false

$F \sim F_{g-1, \sum_{L=1}^{g} n_L - g}$ under H_0.

Reject H_0 if $F > F_{g-1, \sum_{L=1}^{g} n_L - g}(\alpha)$ at level α.

MANOVA (Multivariate ANOVA)

$\underset{\sim}{X}_{Li}$ is the p-variate vector observation for the i^{th} subject under L^{th} treatment.

Model $\underset{\sim}{X}_{Li} = \underset{\sim}{\mu} + \underset{\sim}{\tau}_L + \underset{\sim}{\varepsilon}_{Li}$ (all three are 3×1 vectors).

$\underset{\sim}{\mu}$ = overall effect

$\underset{\sim}{\tau}_L$ = treatment effect from L^{th} treatment level.

$\underset{\sim}{\varepsilon}_{Li}$ = error.

$\underset{\sim}{\tau}_L$'s satisfy $\sum_{L=1}^{g} n_L \underset{\sim}{\tau}_L = \underset{\sim}{0}$

$\underset{\sim}{\varepsilon}_{Li} \overset{iid}{\sim} N_p(\underset{\sim}{0}, \Sigma)$, Σ is unknown.

$$\underset{\sim}{X}_{Li} = \underbrace{\bar{\underset{\sim}{X}}}_{\hat{\mu}} + \underbrace{(\bar{\underset{\sim}{X}}_L - \bar{\underset{\sim}{X}})}_{\hat{\tau}_L} + \underbrace{(\underset{\sim}{X}_{Li} - \bar{\underset{\sim}{X}}_L)}_{\hat{\varepsilon}_{Li}},$$ where

$\bar{\underset{\sim}{X}}_L = \frac{1}{n_L}\sum_{i=1}^{n_L} \underset{\sim}{X}_{Li}$

$\bar{\underset{\sim}{X}} = \sum_{L=1}^{g} \frac{n_L}{n} \bar{\underset{\sim}{X}}_L$

$$\text{Total SS (sum of squares and cross product)} = \sum_{l=1}^{g}\sum_{i=1}^{n_l} \underset{\sim}{x}_{li}\, \underset{\sim}{x}_{li}' = \underbrace{\text{Total Corrected SS}} + n\bar{\underset{\sim}{x}}\bar{\underset{\sim}{x}}'$$

$$= \sum_{l=1}^{g}\sum_{i=1}^{n_l}(\underset{\sim}{x}_{li}-\bar{\underset{\sim}{x}})(\underset{\sim}{x}_{li}-\bar{\underset{\sim}{x}})' = B+W$$ Total (corrected for the mean)

(Treatment) Treat SS (sum of squares and cross product)

$$= B = \sum_{l=1}^{g}\sum_{i=1}^{n_l}(\bar{\underset{\sim}{x}}_l-\bar{\underset{\sim}{x}})(\bar{\underset{\sim}{x}}_l-\bar{\underset{\sim}{x}})' = \sum_{l=1}^{g} n_l(\bar{\underset{\sim}{x}}_l-\bar{\underset{\sim}{x}})(\bar{\underset{\sim}{x}}_l-\bar{\underset{\sim}{x}})', \quad d.f. = g-1.$$

(Error) Residual Sum of Squares and cross product

$$= W = \sum_{l=1}^{g}\sum_{i=1}^{n_l}(\underset{\sim}{x}_{li}-\bar{\underset{\sim}{x}}_l)(\underset{\sim}{x}_{li}-\bar{\underset{\sim}{x}}_l)', \quad d.f. = \sum_{l=1}^{g} n_l - g.$$

Test $H_0: \underset{\sim}{\tau}_1 = \cdots = \underset{\sim}{\tau}_L = \underset{\sim}{0}$ ← treatment effect

Wilk's Λ — Bartlett's large sample test

$$\Lambda^* = \frac{|W|}{|B+W|}.$$ We will reject H_0 for small values of Λ^*.

For large n, $-(n-1-\frac{p+g}{2})\log \Lambda^* \sim \chi^2_{p(g-1)}(\alpha)$

Table 6.3 Distribution of Wilks' Λ^*

p	g = # of groups	Sampling distribution for multivariate normal data
p=1	g ≥ 2	$\left(\frac{\sum n_l - g}{g-1}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{g-1,\, \sum_{l=1}^{g} n_l - g}$
p=2	g ≥ 2	$\left(\frac{\sum_{l=1}^{g} n_l - g-1}{g-1}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1),\, 2(\sum_{l=1}^{g} n_l - g-1)}$
p ≥ 1	g = 2	$\left(\frac{\sum_{l=1}^{g} n_l - p-1}{p}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{p,\, \sum_{l=1}^{g} n_l - p-1}$
p ≥ 1	g = 3	$\left(\frac{\sum_{l=1}^{g} n_l - p-2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p,\, 2(\sum_{i=1}^{g} n_l - p-2)}$

Ex Suppose there are 3 treatment levels with # of subjects 3, 2, 3.

$$\bar{\underset{\sim}{x}}_1 = \begin{pmatrix}-1\\0\end{pmatrix}, \quad S_1 = \begin{pmatrix}5 & 4\\4 & 5\end{pmatrix}$$

$$\bar{\underset{\sim}{x}}_2 = \begin{pmatrix}1\\0\end{pmatrix}, \quad S_2 = \begin{pmatrix}5 & -4\\-4 & 5\end{pmatrix}$$

$$\bar{\underset{\sim}{x}}_3 = \begin{pmatrix}1\\1\end{pmatrix}, \quad S_3 = \begin{pmatrix}3 & 1\\1 & 1\end{pmatrix}$$

(1) Test $H_0: \underset{\sim}{\tau}_1 = \underset{\sim}{\tau}_2 = \underset{\sim}{\tau}_3 = \underset{\sim}{0}$ using (a) Wilk and (b) Bartlett at 5% level.

(2) Using only the first component, obtain MANOVA table.

$g=3,\ n_1=3,\ n_2=2,\ n_3=3,\ n=8$.

(1) $B = \sum_{l=1}^{g} n_l(\bar{\underset{\sim}{x}}_l-\bar{\underset{\sim}{x}})(\bar{\underset{\sim}{x}}_l-\bar{\underset{\sim}{x}})'$, ~~$\bar{x} = \frac{3}{8}$~~

$$\bar{\underset{\sim}{x}} = \frac{3}{8}\begin{pmatrix}-1\\0\end{pmatrix} + \frac{2}{8}\begin{pmatrix}1\\0\end{pmatrix} + \frac{3}{8}\begin{pmatrix}1\\1\end{pmatrix}$$

$$= 3(\bar{\underset{\sim}{x}}_1-\bar{\underset{\sim}{x}})(\bar{\underset{\sim}{x}}_1-\bar{\underset{\sim}{x}})' + 2(\bar{\underset{\sim}{x}}_2-\bar{\underset{\sim}{x}})(\bar{\underset{\sim}{x}}_2-\bar{\underset{\sim}{x}})' + 3(\bar{\underset{\sim}{x}}_3-\bar{\underset{\sim}{x}})(\bar{\underset{\sim}{x}}_3-\bar{\underset{\sim}{x}})'$$

$$W = (n_1-1)S_1 + (n_2-1)S_2 + (n_3-1)S_3 = 2S_1 + S_2 + 2S_3.$$

Compute $\Lambda^* = \frac{|W|}{|B+W|}$ ← determinant

Using the third dist,

$$U = \frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \cdot \frac{8-3-1}{3-1} \text{. Compare with } F_{4,8}(0.05)$$

Reject H_0 if $U > F_{4,8}(0.05)$.

Bartlett's

$$-\left((8-1)-\frac{2+3}{2}\right)\log(\Lambda^*) = V$$

$$\log_{10}(x) = \frac{\log x}{\log 10}$$

Compare with $\chi^2_4(0.05)$.

Reject H_0 if $V > \chi^2_4(0.05)$.

Simultaneous CI for $\tau_{Li} - \tau_{ki}$, $i = 1, 2, \dots, p$
$k, L = 1, 2, \dots, g$.

k = # intervals looking for
$= p\,\frac{g(g-1)}{2}$

$\hat{\tau}_{Li} = (\bar{X}_L - \bar{X})_i$, $\hat{\tau}_{ki} = (\bar{X}_k - \bar{X})_i$.

$\hat{\tau}_{Li} - \hat{\tau}_{ki} = (\bar{X}_L - \bar{X}_k)_i$

Recall $\underset{\sim}{X}_{Li} \sim N_p(\underset{\sim}{\mu} + \underset{\sim}{\tau}_L, \Sigma)$

$\bar{\underset{\sim}{X}}_L \sim N_p(\underset{\sim}{\mu} + \underset{\sim}{\tau}_L, \frac{1}{n_L}\Sigma)$

$\bar{\underset{\sim}{X}}_k \sim N_p(\underset{\sim}{\mu} + \underset{\sim}{\tau}_k, \frac{1}{n_k}\Sigma)$

$\bar{\underset{\sim}{X}}_L - \bar{\underset{\sim}{X}}_k \sim N_p\left(\underset{\sim}{\tau}_L - \underset{\sim}{\tau}_k, \left(\frac{1}{n_L}+\frac{1}{n_k}\right)\Sigma\right)$

$\hat{\tau}_{Li} - \hat{\tau}_{ki} \sim N\left(\tau_{Li} - \tau_{ki}, \left(\frac{1}{n_L}+\frac{1}{n_k}\right)\sigma_{ii}\right)$.

σ_{ii} is estimated by w_{ii}, where W is the within group covariance (sample) matrix.

$$\frac{\hat{\tau}_{Li} - \hat{\tau}_{ki} - (\tau_{Li} - \tau_{ki})}{\sqrt{\left(\frac{1}{n_L}+\frac{1}{n_k}\right)\frac{w_{ii}}{n-g}}} \sim t_{n-g}$$

σ_{ii} is estimated by $\frac{w_{ii}}{n-g}$.

For one difference,

$$\tau_{Li} - \tau_{ki} : \quad \hat{\tau}_{Li} - \hat{\tau}_{ki} \pm \sqrt{\left(\frac{1}{n_L}+\frac{1}{n_k}\right)\frac{w_{ii}}{n-g}}\; t_{n-g}\left(\frac{\alpha}{2}\right).$$

If we want m-simultaneous CI, then only change $t(\frac{\alpha}{2})$ to $t(\frac{\alpha}{2m})$.

$m = \frac{pg(g-1)}{2}$

upto ~~chapter 6~~ (today) for Exam 3
allow 2 pages sheets.

Practice for Exam3

1. 6.5
2. 6.10
3. Consider the data of Exercise 6.14. Suppose Factor 1 corresponds to Treatment and Factor 2 corresponds to **replications**. So, there are 3 treatment levels and for each level there are 4 bivariate measurement vectors. Describe how will you obtain the MANOVA table. How would you test for no treatment effect? How would you find 95% simultaneous CI for all possible treatment effect differences?

4. Suppose $x_1, \dots, x_n$ are iid from a bivariate **normal distribution**. How would you test the equality of the two components of the mean vector at 10% level of significance?

5. Suppose we have an incomplete data matrix from a bivariate normal ~~distrib~~ population.

$$X = \begin{bmatrix} ? & 3 \\ 3 & 1 \\ 1 & -1 \end{bmatrix}$$

Describe the iterative procedure (initial estimate and one update) to obtain the MLE of the parameters.

Exercise 6.14

		Factor 2: level 1	level 2	level 3	level 4
Factor 1	Level 1	$\binom{14}{8}$	$\binom{6}{2}$	$\binom{8}{2}$	$\binom{16}{-4}$
	Level 2	$\binom{1}{6}$	$\binom{5}{12}$	$\binom{0}{15}$	$\binom{2}{7}$
	Level 3	$\binom{3}{-2}$	$\binom{-2}{7}$	$\binom{-11}{1}$	$\binom{-6}{6}$

$\binom{-10}{-6}$ $\quad$ $\binom{-16}{4}$

Math A7800 $\qquad$ 11/30/17, Thur

Mid-term3 up to 11/28/17. (2pages sheets)

Test of equality of Covariance (Testing for equality of Covariance Matrices)

Set up $\underset{\sim}{X}_{li} \sim N_p(\underset{\sim}{\mu}_l, \Sigma_l)$, $i = 1, 2, \dots, n_l$ and $l = 1, 2, \dots, g$.

$H_0 : \Sigma_1 = \Sigma_2 = \cdots \Sigma_g$ VS $H_1 : H_0$ is false.

We will do LRT (Likelihood Ratio Test) and approx its null dist. by $\chi^2_{(?)}$

Obtain the sample covariance matrix $S_1, \dots, S_g$, where

$$S_l = \frac{1}{n_l - 1} \sum_{i=1}^{n_l} (\underset{\sim}{x}_{li} - \bar{\underset{\sim}{x}}_l)(\underset{\sim}{x}_{li} - \bar{\underset{\sim}{x}}_l)', \text{ where}$$

$$\bar{\underset{\sim}{x}}_l = \frac{1}{n_l} \sum_{i=1}^{n_l} \underset{\sim}{x}_{li}.$$

Then obtain

$$S_{pooled} = \frac{(n_1 - 1)S_1 + \cdots + (n_g - 1)S_g}{\sum_{l=1}^{g} (n_l - 1)}$$

The LR $\Lambda = \prod_{l=1}^{g} \left(\frac{|S_l|}{|S_{pooled}|} \right)^{\frac{(n_l - 1)}{2}}$

So, $-2 \log \Lambda = \sum_{l=1}^{g} (n_l - 1) \log |S_{pooled}| - \sum_{l=1}^{g} (n_l - 1) \log |S_l|$

$= M$ (call) "Box's M statistic"

We expect M to be small when H_0 is true, and M to be large when H_0 is false.

So, reject H_0 when M is large.

Box's test for Equality of Covariance Matrices

$$u = \left[\sum_{l=1}^{g} \frac{1}{n_l-1} - \frac{1}{\sum_{l=1}^{g}(n_l-1)}\right]\frac{2p^2+3p-1}{6(p+1)(g-1)}.$$

Then $(1-u)M \sim \chi^2_\nu$ under H_0,

where $\nu = g\left(p+\frac{p(p+1)}{2}\right) - gp - \frac{p(p+1)}{2}$

$$= (g-1)\frac{p(p+1)}{2} = \frac{1}{2}p(p+1)(g-1)$$

Reject H_0 at level α if $(1-\alpha)M > \chi^2_\nu(\alpha)$.

Two-way MANOVA (with Fixed effect and interactions)

In an experiment, the outcome measurement may be influenced by one of the two 'factors' present in the experimental condition or by the interaction of the two factors.

Factor 1 has g levels

Factor 2 has b levels

For each of gb factor pairs, the experiment is repeated n times

X_{lki} = measurement for the i th experiment with Fac. 1 level l, Fac. 2 level k, $i=1,2,\dots,n$, $l=1,2,\dots,g$, $k=1,2,\dots,b$.

Total number of measurement $= gbn$

$$X_{lki} = \underbrace{\mu}_{\text{overall effect}} + \underbrace{\tau_l}_{\text{Factor 1 effect}} + \underbrace{\beta_k}_{\text{Factor 2 effect}} + \underbrace{\gamma_{lk}}_{\text{interaction effect}} + \underbrace{\varepsilon_{lki}}_{\text{error}}$$

1. $\varepsilon_{lki} \overset{iid}{\sim} N_p(0,\Sigma)$
2. $\sum_{l=1}^{g}\tau_l = \sum_{k=1}^{b}\beta_k = \sum_{l=1}^{g}\gamma_{lk} = \sum_{k=1}^{b}\gamma_{lk} = 0$.

Define

$\bar{X}$ = combined mean

$\bar{X}_{l\cdot} = \frac{1}{bn}\sum_{k=1}^{b}\sum_{i=1}^{n} X_{lki}$, $l=1,\dots,g$.

l level of factor 1 fix 'g'

$\bar{X}_{\cdot k} = \frac{1}{gn}\sum_{l=1}^{g}\sum_{i=1}^{n} X_{lki}$, $k=1,\dots,b$

k level of factor 2; fix 'b'.

$\bar{X}_{lk} = \frac{1}{n}\sum_{i=1}^{n} X_{lki}$, $l=1,2,\dots,g$, $k=1,2,\dots,b$.

$$X_{lki} = \bar{X} + (\bar{X}_{l\cdot} - \bar{X}) + (\bar{X}_{\cdot k} - \bar{X}) + (\bar{X}_{lk} - \bar{X}_{l\cdot} - \bar{X}_{\cdot k} + \bar{X}) + (X_{lki} - \bar{X}_{lk})$$

$$= \hat{\mu} + \hat{\tau}_l + \hat{\beta}_k + \hat{\gamma}_{lk} + \hat{\varepsilon}_{lki}.$$

m

Two-way MANOVA

Source	SSP (sum of squares and cross products)	d.f
Factor 1	$SSP_{fac1} = \sum_{l=1}^{g} bn(\underset{\sim}{\bar{x}}_{l\cdot} - \underset{\sim}{\bar{x}})(\underset{\sim}{\bar{x}}_{l\cdot} - \underset{\sim}{\bar{x}})'$	$g-1$
Factor 2	$SSP_{fac2} = \sum_{k=1}^{b} gn(\underset{\sim}{\bar{x}}_{\cdot k} - \underset{\sim}{\bar{x}})(\underset{\sim}{\bar{x}}_{\cdot k} - \underset{\sim}{\bar{x}})'$	$b-1$
Interaction	$SSP_{int} = \sum_{l=1}^{g}\sum_{k=1}^{b} n(\underset{\sim}{\bar{x}}_{lk} - \underset{\sim}{\bar{x}}_{l\cdot} - \underset{\sim}{\bar{x}}_{\cdot k} + \underset{\sim}{\bar{x}})(\underset{\sim}{\bar{x}}_{lk} - \underset{\sim}{\bar{x}}_{l\cdot} - \underset{\sim}{\bar{x}}_{\cdot k} + \underset{\sim}{\bar{x}})'$	$(g-1)(b-1)$
Residual	$SSP_{res} = \sum_{l=1}^{g}\sum_{k=1}^{b}\sum_{i=1}^{n} (\underset{\sim}{x}_{lki} - \underset{\sim}{\bar{x}}_{lk})(\underset{\sim}{x}_{lki} - \underset{\sim}{\bar{x}}_{lk})'$	$gb(n-1)$
Total (corrected)	$SSP_{cor} = \sum_{l=1}^{g}\sum_{k=1}^{b}\sum_{i=1}^{n} (\underset{\sim}{x}_{lki} - \underset{\sim}{\bar{x}})(\underset{\sim}{x}_{lki} - \underset{\sim}{\bar{x}})'$	$gbn-1$

$$\underset{(\text{uncorrected})}{TSS} = \sum_{l=1}^{g}\sum_{k=1}^{b}\sum_{i=1}^{n} \underset{\sim}{x}_{lki}\underset{\sim}{x}_{lki}' = SSP_{tot\,(\text{uncorrected})}$$

$$= \sum_{l}\sum_{k}\sum_{i} (\underset{\sim}{x}_{lki} - \underset{\sim}{\bar{x}})(\underset{\sim}{x}_{lki} - \underset{\sim}{\bar{x}})' + gbn\,\underset{\sim}{\bar{x}}\,\underset{\sim}{\bar{x}}'$$

$$= SSP_{tot(cor)} + gbn\,\underset{\sim}{\bar{x}}\,\underset{\sim}{\bar{x}}'$$

$$\begin{cases} H_0 : \text{No interaction effect, } \underset{\sim}{\gamma}_{lk} = \underset{\sim}{0}, \forall\, l, k \\ H_1 : H_0 \text{ is false.} \end{cases}$$

reject H_0 at level α if $\Lambda = \dfrac{|SSP_{res}|}{|SSP_{int} + SSP_{res}|}$ is small.

Bartlett's test

Reject H_0 if at level α $-\left[gb(n-1) - \dfrac{p+1-(g-1)(b-1)}{2}\right]\log\Lambda > \chi^2_{(g-1)(b-1)p}(\alpha)$

- Test 1 $\begin{cases} H_0 : \text{Fact 1 has no effect, i.e., } \underset{\sim}{\tau}_1 = \cdots = \underset{\sim}{\tau}_g = \underset{\sim}{0} \\ H_1 : H_0 \text{ is false.} \end{cases}$

$$\Lambda = \frac{|SSP_{res}|}{|SSP_{fac1} + SSP_{res}|}$$

The Bartlett's test rejects H_0 at level α if

$$-\left[gb(n-1) - \frac{p+1-(g-1)}{2}\right]\log\Lambda > \chi^2_{p(g-1)}(\alpha)$$

- Test 2 $\begin{cases} H_0 : \text{Fact 2 has no effect} \\ H_1 : H_0 \text{ is false} \end{cases}$

$$\Lambda = \frac{|SSP_{res}|}{|SSP_{fact2} + SSP_{res}|}$$

The Bartlett's test rejects H_0 at level α if

$$-\left[gb(n-1) - \frac{p+1-(b-1)}{2}\right]\log\Lambda > \chi^2_{p(b-1)}(\alpha)$$

$$\hat{\Sigma} = \frac{SSP_{res}}{gp(g-1)} = E.$$

$\hat{\Sigma} = SSP_{res} = E$

Find $100(1-\alpha)\%$ CI for

$\tau_{Li} - \tau_{mi}$, $i = 1, 2, \dots, p$, $m = 1, 2, \dots, g$.

$\hat{\underset{\sim}{\tau}}_L = \underset{\sim}{\bar{X}}_{L\cdot} - \underset{\sim}{\bar{X}}$, $\quad \hat{\underset{\sim}{\tau}}_L - \hat{\underset{\sim}{\tau}}_m = \underset{\sim}{\bar{X}}_{L\cdot} - \underset{\sim}{\bar{X}}_m$

$\widehat{\tau_{Li} - \tau_{mi}} = \bar{X}_{L\cdot i} - \bar{X}_{m\cdot i}$

CI for $\tau_{Li} - \tau_{mi}$ is

$$\boxed{\bar{X}_{L\cdot i} - \bar{X}_{m\cdot i} \pm t_\nu\left(\frac{\alpha}{2}\right)\sqrt{\frac{E_{ii}}{\nu}\frac{2}{bn}}, \ \nu = gb(n-1).}$$

Q: $100(1-\alpha)\%$ simultaneous CI for $\tau_{Li} - \tau_{mi}$ for all i and L, m.

Total # of intervals $= p \cdot \binom{g}{2} = \frac{pg(g-1)}{2}$

Ans $\bar{X}_{L\cdot i} - \bar{X}_{m\cdot i} \pm \sqrt{\frac{E_{ii}}{\nu}\frac{2}{bn}} \; t_{\underbrace{gb(n-1)}_{\nu}}\left(\frac{\alpha}{pg(g-1)}\right)$

Find $100(1-\alpha)\%$ CI for $\beta_{ki} - \beta_{qi}$.

Ans $\bar{X}_{\cdot ki} - \bar{X}_{\cdot qi} \pm t_{\underbrace{gb(n-1)}_{\nu}}\left(\frac{\alpha}{2}\right)\sqrt{\frac{E_{ii}}{\nu}\frac{2}{gn}}$.

Q: $100(1-\alpha)\%$ simultaneous CI for $\beta_{ki} - \beta_{qi}$ for all i, k, q.

Ans $\bar{X}_{\cdot ki} - \bar{X}_{\cdot qi} \pm t_{gb(n-1)}\left(\frac{\alpha}{pb(b-1)}\right)\sqrt{\frac{E_{ii}}{\nu}\frac{2}{gn}}$.

Chapter 7. Multivariate Linear Regression Models.

History

one response variable : y

A bunch of explanatory variables $Z_1, \dots, Z_p$.

Linear

$\mathbb{E}(Y)$ is a linear function of $Z_1, \dots, Z_p$.

In particular, $\mathbb{E}(Y) = \beta_0 + \beta_0 Z_1 + \dots + \beta_r Z_r$ for some (unknown) coefficients β_i's.

$Y = \mathbb{E}(Y) + \text{error}$.

The linear model: $Y = \underbrace{\beta_0 + \beta_1 Z_1 + \dots + \beta_r Z_r}_{\text{model}} + \underbrace{\varepsilon}_{\text{random}}$

Observed data

We observe n-samples. For these observed samples, we know the values of the response and explanatory variable.

$$\underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \quad \begin{bmatrix} Z_{11} & \cdots & Z_{1r} \\ \vdots & \ddots & \\ Z_{n1} & \cdots & Z_{nr} \end{bmatrix}$$

The i^{th} row corresponds to the i^{th} sample.

Z_{ij} = value of i^{th} sample, j^{th} variable.

Define

$$Z = \begin{bmatrix} 1 & Z_{11} & \cdots & Z_{1r} \\ 1 & \vdots & \ddots & \\ \vdots & & & \\ 1 & Z_{n1} & \cdots & Z_{nr} \end{bmatrix}, \quad \underset{\sim}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix}, \quad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

β_0

Linear Model

$$\underset{\sim}{Y} = \underset{n\times(r+1)}{Z}\ \underset{(r+1)\times 1}{\underset{\sim}{\beta}} + \underset{n\times 1}{\underset{\sim}{\varepsilon}}$$

$\underset{\sim}{\beta}$ is unknown.

Assumptions about error

$\mathbb{E}(\varepsilon_i) = 0$, $var(\varepsilon_i) = \sigma^2$.

$cov(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$.

so, $cov(\underset{\sim}{\varepsilon}) = \sigma^2 \underset{\sim}{I}$.

First goal: Estimate $\underset{\sim}{\beta}$

We obtain $\hat{\beta}$ by minimizing the sum of squared error.

$$\sum_{i=1}^{n} (y_i - Z_{i*}\underset{\sim}{\beta})^2$$

Z_{i*} = i^{th} row of Z.

Linear Regression

In the last class, we defined the Linear Regression Model.

$$\underset{\sim}{Y} = Z\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

$$\underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$Z = \begin{pmatrix} 1 & Z_{11} & \cdots & Z_{1r} \\ 1 & \vdots & & \\ \vdots & & & \\ 1 & Z_{n1} & \cdots & Z_{nr} \end{pmatrix}_{n \times (r+1)}$$

Y is called the response (or dependent) variable.

$Z = (Z_1, \ldots, Z_r)$ are called independent (or predictor or feature)

$\underset{\sim}{\varepsilon}$ is additive error.

one of the goals is to be able to predict the value of y based on corresponding values of Z.

There are other goals. e.g. variable selection.

For predicting the y value, we need to estimate the unknown parameter $\underset{\sim}{\beta}$.

For this, we rely on a 'training data', say a sample of size n for which both y and Z values are known.

We represent the y values in $\underset{\sim}{y}$ and z values in $Z_{n\times(r+1)}$

Then, we come up with an estimator ('good')

$$\hat{\underset{\sim}{\beta}} = \hat{\beta}(\underset{\sim}{y}, Z)$$

For future observations, or for those cases where y value is not known, but z value is known. (say z_0 r×1 vectors).

Then the predicted value of y is

$$\hat\beta_0+\hat\beta_1 z_{01}+\cdots+\hat\beta_r z_{0r} = [1, z_{01},\ldots,z_{0r}]\hat{\underset{\sim}{\beta}}.$$

For estimating $\hat\beta$, we will use least squares method.

We want to β that minimizes

$$S(\underset{\sim}{b})=\sum_{i=1}^{n} (y_i-b_0-b_1 z_{i1}-\cdots b_r z_{ir})^2$$

$$=(\underset{\sim}{y}-Z\underset{\sim}{b})'(\underset{\sim}{y}-Z\underset{\sim}{b})$$

$$\begin{bmatrix} \frac{\partial S}{\partial b_0} \\ \vdots \\ \frac{\partial S}{\partial b_r} \end{bmatrix} = -2Z'(\underset{\sim}{y}-Z\underset{\sim}{b}) \quad \text{set} = 0.$$

$$Z'Z\hat{\underset{\sim}{\beta}} = Z'\underset{\sim}{y}$$

Assume that Z has rank r+1 "full rank."

Then (Z'Z) is non-singular.

so, $\hat{\underset{\sim}{\beta}} = (Z'Z)^{-1}Z\underset{\sim}{y}$

$\underset{\sim}{y} = Z\underset{\sim}{\beta}$

$$\hat{\underset{\sim}{y}} = \text{Fitted values of } \underset{\sim}{y} = Z\hat{\underset{\sim}{\beta}} = \underbrace{Z(Z'Z)^{-1}Z'}_{P_Z \text{ (projection matrix)}}\underset{\sim}{y} = P_Z\underset{\sim}{y}.$$

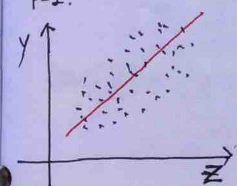
Where P_Z is symmetric $P_Z^2 = P_Z$,

$\hat{\underset{\sim}{\beta}} = (Z'Z)^{-1}Z'\underset{\sim}{y}$.

$(I-P_Z)^2 = I-P_Z$, $I-P_Z$ is symmetric.

Ex height vs weight
Z Y

p=1.

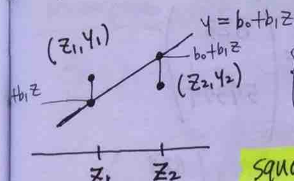


Plot of training sample data

Goal : Fit a straight line

$$Y = \beta_0+\beta_1 Z.$$

"$\hat{y}$" $= Z\hat{\underset{\sim}{\beta}}$



error for $z_1 = y_1-(b_0+b_1z_1)$

error for $z_2 = y_2-(b_0+b_1z_2)$

$\hat{y} = Z\hat{\underset{\sim}{\beta}}$

squared total error

$$=(y_1-b_0-b_1z_1)^2+(y_2-b_0-b_1z_2)^2$$

Ex Suppose height and weight of 5 persons are given below

$$\begin{array}{cc} H & W \end{array}$$
$$\begin{bmatrix} 73 & 230 \\ 70 & 140 \\ 65 & 141 \\ 71 & 180 \\ 67 & 132 \end{bmatrix}$$

Q: Predict the weight ($\underset{\sim}{y}$) of a person having height 60.

$$\underset{\sim}{Y} = \begin{bmatrix} 230 \\ 140 \\ 141 \\ 180 \\ 132 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 73 \\ 1 & 70 \\ 1 & 65 \\ 1 & 71 \\ 1 & 67 \end{bmatrix}$$

$$\underset{\sim}{\hat\beta} = \begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \end{bmatrix} = (Z'Z)^{-1} Z' \underset{\sim}{Y}$$

$$Z'Z = \begin{pmatrix} 5 & 346 \\ 346 & 23984 \end{pmatrix}$$

$$(Z'Z)^{-1} = \begin{pmatrix} 117.5 & -1.696 \\ -1.696 & 0.0245 \end{pmatrix}, \quad Z'\underset{\sim}{y} = \begin{pmatrix} 823 \\ 57379 \end{pmatrix}$$

$$\underset{\sim}{\hat\beta} = \begin{pmatrix} \hat\beta_0 \\ \hat\beta_1 \end{pmatrix} = (Z'Z)^{-1} Z'\underset{\sim}{y} = \begin{pmatrix} 117.5 & -1.696 \\ -1.696 & 0.0245 \end{pmatrix} \begin{pmatrix} 823 \\ 57379 \end{pmatrix}$$

Ans: $\hat\beta_0 + \hat\beta_1 \cdot \underset{\uparrow}{60} = W_{60}$

Properties of $\underset{\sim}{\hat\beta}$

Recall the assumption on the error $\underset{\sim}{\varepsilon}$ was

$$\mathbb{E}[\underset{\sim}{\varepsilon}] = \underset{\sim}{0}, \quad \text{cov}(\underset{\sim}{\varepsilon}) = \sigma^2 \underset{\sim}{I}.$$

$$\mathbb{E}(\underset{\sim}{y}) = Z\underset{\sim}{\beta}, \quad \text{cov}(\underset{\sim}{y}) = \sigma^2 \underset{\sim}{I}$$

$$\mathbb{E}(\underset{\sim}{\hat\beta}) = \mathbb{E}\big[\overbrace{(Z'Z)^{-1}Z'}^{\text{"constant"}}\underset{\sim}{y}\big]$$
$$= (Z'Z)^{-1}Z'Z\underset{\sim}{\beta} = \underset{\sim}{\beta}.$$

$$\text{cov}(\underset{\sim}{\hat\beta}) = (Z'Z)^{-1}Z'(\sigma^2 \underset{\sim}{I})Z(Z'Z)^{-1}$$
$$= \sigma^2 (Z'Z)^{-1}$$

Decompose sum of squares

$$\sum_{i=1}^n (y_i - \bar y)^2 = \text{Total SS (corrected)}$$
$$= \underbrace{\sum_{i=1}^n (y_i - \hat y_i)^2}_{\text{Residual S.S.}} + \underbrace{\sum_{i=1}^n (\hat y_i - \bar y)^2}_{\text{Regression SS.}}$$

$$\hat{\sigma^2} = \frac{\text{Residual SS}}{n - \underset{\text{rank of } Z}{(r+1)}} = s^2$$

Multiple correlation coefficient

$$\boxed{R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

R^2 lies between $(0,1)$

$R^2 \approx 0 \Rightarrow$ bad prediction.

R^2 close to $1 \Rightarrow$ good prediction.

$\hat{y} = P_Z y$

$y - \hat{y} = \hat{\varepsilon} = (I - P_Z) y$.

Residual SS $= (y - \hat{y})'(y - \hat{y}) = \varepsilon'\varepsilon$

$= y'(I - P_Z) y$

$\hat{y} \cdot \hat{\varepsilon} = 0 = \hat{y}'\hat{\varepsilon}$.

Total SS (corrected)

$= y'(I - P_Z) y$, $P_1 = 1(1'1)^{-1}1' = \frac{1}{n} 1 1'$

so, Regression SS $= y'(P_Z - P_1) y$

Now, assume ε (error vector) $\sim N_n(0, \sigma^2 I)$

Then $\boxed{y \sim N(Z\beta, \sigma^2 I)}$

Q: $\hat{\beta}_{MLE} = \hat{\beta}_{LS}$ (same $\hat{\beta}$)

$= (Z'Z)^{-1} Z' y$

$\hat{\sigma}^2_{MLE} = \frac{1}{n}$ (Residual SS). $r+1 = \text{rank}(Z)$.

$= \frac{n-(r+1)}{n} S^2$

$\hat{\beta} \sim N_{r+1}\left(\beta, \underbrace{\sigma^2 (Z'Z)^{-1}}_{\text{cov}(\hat{\beta})}\right)$

CI for β_i ? — ①

CR for β ? — ②

$\frac{\hat{\beta}_i - \beta_i}{\sqrt{S^2((Z'Z)^{-1})_{ii}}} \sim t_{n-r-1}$

① $\hat{\beta}_i \sim N(\beta_i, \underset{\uparrow\, S^2}{\sigma^2}((Z'Z)^{-1})_{ii})$

$\hat{\beta}_i \pm \sqrt{S((Z'Z)^{-1})_{ii}}\, t_{n-r-1}\left(\frac{\alpha}{2}\right)$.

② $100(1-\alpha)\%$ CR of β is given by

$(\beta - \hat{\beta})'(Z'Z)(\beta - \hat{\beta}) \leq (r+1) S^2 F_{r+1,\, n-r-1}(\alpha)$

Q3: $H_0 : \underset{\sim}{\beta} = \underset{\sim}{0}$ vs $H_1 : \underset{\sim}{\beta} \neq \underset{\sim}{0}$

Reject H_0 at level α if $\hat{\underset{\sim}{\beta}}' Z' Z \hat{\underset{\sim}{\beta}} > (r+1) s^2 * F_{r+1, n-r-1}(\alpha)$

For $H_0 : \beta_i = 0$ vs $H_1 : \beta_i \neq 0$,

Reject H_0 at level α,

if $0 \notin \hat{\beta}_i \pm \sqrt{s^2 \left((Z'Z)^{-1} \right)_{ii}}$

$x_2 = \beta_0 + \beta_1 x_1$

$x_3 = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$(1 \ x_1 \ x_2) \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$

$x_4 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

$(1 \ x_1 \ x_2 \ x_3) \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$

6.14 (b) Two-way MANOVA.

$\underbrace{\underset{\sim}{X}_{lki}}_{2\times 1}$ for $l = 1, 2, 3$, $k = 1, 2, 3, 4$, $i = 1, 2$

Table 1 (in Q13) Factor 2 (k)

$i=1$ case

Factor 1 (l)	1	2	3	4
1	$X_{111} = \begin{pmatrix} 6 \\ 8 \end{pmatrix}$	$X_{121} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$	$X_{131} = \begin{pmatrix} 8 \\ 12 \end{pmatrix}$	$X_{141} = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$
2	$X_{211} = \begin{pmatrix} 3 \\ 8 \end{pmatrix}$	$X_{221} = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$	$X_{231} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$	$X_{241} = \begin{pmatrix} -4 \\ 3 \end{pmatrix}$
3	$X_{311} = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$	$X_{321} = \begin{pmatrix} -4 \\ -5 \end{pmatrix}$	$X_{331} = \begin{pmatrix} 3 \\ -3 \end{pmatrix}$	$X_{341} = \begin{pmatrix} -4 \\ -6 \end{pmatrix}$

Table 2 (in Q14)

$i=2$ case

Factor 1	1	2	3	4
1	$X_{112} = \begin{pmatrix} 14 \\ 8 \end{pmatrix}$	$X_{122} = \begin{pmatrix} 6 \\ 2 \end{pmatrix}$,	$X_{132} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$,	$X_{142} = \begin{pmatrix} 16 \\ -4 \end{pmatrix}$
2	$X_{212} =$	$X_{222} =$	$X_{232} =$	$X_{242} =$
3	$X_{312} =$	$X_{322} =$	$X_{332} =$	$X_{342} =$

$\bar{X}_{11\cdot} = ($

<u>Need to compute</u> $\bar{X}_{L\cdot}$, $\bar{X}_{\cdot k}$, $\bar{X}_{Lk\cdot}$, $\bar{X}$.

$\bar{X}_{1\cdot}$ = average of ~~first~~ all the vectors present in the first row of Table 1 and 2.

$\bar{X}_{2\cdot}$ = ---- second row, $\bar{X}_{3\cdot}$ = --- third row.

$\bar{X}_{\cdot 1}$ = average of all the vectors in the first column of Table 1 and 2.

$\bar{X}_{\cdot 2}$ = ---- second column, --

$\bar{X}_{\cdot 3}$ = --- third column ---

$\bar{X}_{\cdot 4}$ = --- forth column ---

$\bar{X}_{Lk\cdot} = \frac{1}{2}(X_{Lk1} + X_{Lk2})$ = average of the $(L,k)^{th}$ vectors of Table 1 and 2.

$\bar{X}$ = average over vectors.

= average of $\bar{X}_{1\cdot}, \bar{X}_{2\cdot}, \bar{X}_{3\cdot}$

= average of $\bar{X}_{\cdot 1}, \bar{X}_{\cdot 2}, \bar{X}_{\cdot 3}, \bar{X}_{\cdot 4}$

= average of $\bar{X}_{11\cdot}, \cdots, \bar{X}_{34\cdot}$.

$$SSP_{Fac1} = \sum_{L=1}^{3}\sum_{k=1}^{4}\sum_{i=1}^{2} (\bar{X}_{L\cdot} - \bar{X})(\bar{X}_{L\cdot} - \bar{X})'$$

$$= 8\Big[(\bar{X}_{1\cdot} - \bar{X})(\bar{X}_{1\cdot} - \bar{X})' + (\bar{X}_{2\cdot} - \bar{X})(\bar{X}_{2\cdot} - \bar{X})' + (\bar{X}_{3\cdot} - \bar{X})(\bar{X}_{3\cdot} - \bar{X})'\Big]$$

$$SSP_{Fac2} = \sum_{L=1}^{3}\sum_{k=1}^{4}\sum_{i=1}^{2} (\bar{X}_{\cdot k} - \bar{X})(\bar{X}_{\cdot k} - \bar{X})'$$

$$= 6\Big[(\bar{X}_{\cdot 1} - \bar{X})(\bar{X}_{\cdot 1} - \bar{X})' + (\bar{X}_{\cdot 2} - \bar{X})(\bar{X}_{\cdot 2} - \bar{X})' + (\bar{X}_{\cdot 3} - \bar{X})(\bar{X}_{\cdot 3} - \bar{X})' + (\bar{X}_{\cdot 4} - \bar{X})(\bar{X}_{\cdot 4} - \bar{X})'\Big].$$

$$SSP_{int} = \sum_{L}\sum_{k}\sum_{i} (\bar{X}_{Lk\cdot} - \bar{X}_{L\cdot} - \bar{X}_{\cdot k} + \bar{X})(\bar{X}_{Lk\cdot} - \bar{X}_{L\cdot} - \bar{X}_{\cdot k} + \bar{X})'$$

$$= 2\Big[(\overbrace{\cancel{\bar{X}_{11\cdot} - \bar{X}}}^{\bar{X}_{11\cdot} - \bar{X}_{1\cdot} - \bar{X}_{\cdot 1} + \bar{X}})(\quad)' + (\overbrace{\cancel{\bar{X}_{12\cdot} - \bar{X}}}^{\bar{X}_{12\cdot} - \bar{X}_{1\cdot} - \bar{X}_{\cdot 2} + \bar{X}})(\quad)' + \cdots + (\underbrace{\cancel{\bar{X}_{34} - \bar{X}}}_{\bar{X}_{34\cdot} - \bar{X}_{3\cdot} - \bar{X}_{\cdot 4} + \bar{X}})(\uparrow)'\Big]$$

$$SSP_{err} = \sum_{L=1}^{3}\sum_{k=1}^{4}\sum_{i=1}^{2} (X_{Lki} - \bar{X}_{Lk\cdot})(X_{Lki} - \bar{X}_{Lk\cdot})'$$

<u>test "interaction"</u>

6.18 $\bar{X}_1 = \begin{pmatrix} 4.9 \\ 4.6 \\ 3.9 \end{pmatrix}$, $\bar{X}_2 = \begin{pmatrix} 4.7 \\ 4.4 \\ 3.7 \end{pmatrix}$

$S_{pooled} = (\quad)$ $n_1 = 24$, $n_2 = 24$

$H_0: \underset{\sim}{\mu}_1 = \underset{\sim}{\mu}_2$ vs $H_1: \underset{\sim}{\mu}_1 \neq \underset{\sim}{\mu}_2$.

$$\bar{X}_1 - \bar{X}_2 \sim N_3\left(\underset{\sim}{0}, \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\Sigma\right)$$

$$(\bar{X}_1 - \bar{X}_2)'\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_{pooled}\right]^{-1}(\bar{X}_1 - \bar{X}_2)$$

(a)

$$\sim \frac{(n_1+n_2-2)P}{n_1+n_2-P-1} F_{P,n_1+n_2-P-1}(\alpha)$$

Compute $T^2 = (\bar{X}_1 - \bar{X}_2)'(\quad)^{-1}(\bar{X}_1 - \bar{X}_2)$.

Reject H_0 at level α if $T^2 > \frac{(n_1+n_2-2)P}{n_1+n_1-P-1} F_{P,n_1+n_2-P-1}(\alpha)$.

(c) $\mu_{1i} - \mu_{2i}$ has estimator

$$\underbrace{\bar{X}_{1i} - \bar{X}_{2i}}_{\text{"midpoint"}} \pm \sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_{pooled}^{(ii)}}\; t_{n_1+n_2-2}\left(\frac{\alpha}{2\cdot K}\right)$$

Bonferroni

K

7.2 $Y = \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon$

$\underset{\sim}{Y}$, $\underset{\sim}{Z} = \begin{pmatrix} Z_{11} & Z_{12} \\ \vdots & \\ Z_{n1} & Z_{n2} \end{pmatrix}$

$$\underset{2\times 1}{\hat{\underset{\sim}{\beta}}} = (Z'Z)^{-1}Z'\underset{\sim}{Y}.$$

7.9 $y_{j1} = \beta_{01} + \beta_{11} z_{j1} + \varepsilon_{j1}$

$y_{j2} = \beta_{02} + \beta_{12} z_{j1} + \varepsilon_{j2}$, $j = 1,2,3,4,5$.

$Z = \begin{pmatrix} 1 & z_1 \\ 1 & \vdots \\ \vdots & \vdots \\ 1 & z_n \end{pmatrix}$ $\underset{\sim}{Y}_1 = \begin{pmatrix} y_{11} \\ \vdots \\ y_{n1} \end{pmatrix}$, $\underset{\sim}{Y}_2 = \begin{pmatrix} y_{12} \\ \vdots \\ y_{n2} \end{pmatrix}$

$$\begin{pmatrix} \hat{\beta}_{01} \\ \hat{\beta}_{11} \end{pmatrix} = (Z'Z)^{-1}Z'\underset{\sim}{Y}_1$$

$$\begin{pmatrix} \hat{\beta}_{02} \\ \hat{\beta}_{12} \end{pmatrix} = (Z'Z)^{-1}Z'\underset{\sim}{Y}_2$$

$\hat{\beta} \sim N_2\left(\beta, \sigma^2 (Z'Z)^{-1}\right)$ $\qquad s^2 = \frac{SS_{err}}{n-2}$

Find CI for $\beta_0 + \beta_1 Z_{new}$

Look at $\hat{\beta}_0 + \hat{\beta}_1 Z_{new} \pm \sqrt{s^2 f}\; t_{n-2}\left(\frac{\alpha}{2}\right)$.

$$[1 \;\; Z_{new}]\,\hat{\beta} \sim N\left(\beta_0 + \beta_1 Z_{new},\; [1,\, Z_{new}]\,\sigma^2 (Z'Z)^{-1} \begin{pmatrix} 1 \\ Z_{new} \end{pmatrix}\right)$$

$$= \sigma^2 f$$

Final — 6 – 8:15 pm

3 { Ch. 2 – 1
Ch 3 – 1

7 problems

proof

MLE Linear Algebra

Ch. 4, 5, 6 – 5 problems
(2, 2, 1)

How to get Scheffe Interval?
proof

Bonferroni why we divide by k?

"calculator & 4 pages (note)"

How formula comes from for CR


ellipse

$S = \frac{1}{n-1} \Sigma(\quad)$ why unbiased of Σ.