Chapter 5: Comparison of Several Mean vectors (Part 1)

Zhixiang ZHANG

University of Macau

zhixzhang@um.edu.mo

February 24, 2025

Introduction

- The ideas developed in Chapter 4 can be extended to handle problems involving the comparison of several mean vectors.
- The theories will rest on assumptions of multivariate normal distributions or large sample sizes. Large S. S. C.L.T. MVN
    - Small sample case:

      MVN assumptions: usually F distribution
      Not MVN: not considered
    - Large sample case: usually chi-squared distribution

Introduction

Lecture 4: $\mu = \mu_1 - \mu_2 = 0$

Lecture 5: $\mu_1 = \mu_2$ part 1

- We begin by considering pairs of mean vectors. Measurements are often recorded under different sets of experimental conditions to see whether the responses differ significantly over these sets. For example, the efficacy of a new drug or of an advertising campaign may be determined by comparing measurements before the "treatment" (drug or advertising) with those after the treatment.
  medicine new strategy

  $H_0: \vec{\mu_1} = \vec{\mu_2} = \cdots = \vec{\mu_m}$ part 2
- Later we discuss comparisons among several mean vectors arranged according to treatment levels. The corresponding test statistics depend upon a partitioning of the total variation into pieces of variation attributable to the treatment sources and error, known as MANOVA.

Table of Contents

Paired Comparison

Let $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ be two p-variate responses for observation $i(i = 1, \ldots, n)$.
Example: SAT pre-class test grades and post-class grades
Pre-class grades:

$$\mathbf{x}_{1i} = (\text{ Quant } = 640, \text{ Analyt } = 610, \text{ Verbal } = 490)$$

Post-class grades:

$$\mathbf{x}_{2i} = (\text{ Quant } = 680, \text{ Analyt } = 620, \text{ Verbal } = 560)$$

Does the class has an effect on SAT score?

Paired Comparison

1. Calculate $\mathbf{d}_i = \mathbf{x}_{1i} - \mathbf{x}_{2i}$, $i = 1, 2, \cdots, n$.

care about difference

Denote $\delta = E[d_i]$, testing $H_0: \delta = 0$

criteria

Assume $d_i \overset{i.i.d}{\sim} N(\mu, \Sigma)$

$$\overline{\mathbf{d}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{d}_i,$$

Testing: $T^2 = n\bar{d}^\top S_d^{-1} \bar{d} \sim \frac{(n-1)p}{n-p} F_{p,n-p}$ of H_0 is true

$$\mathbf{S}_d = \frac{1}{n-1} \sum_{i=1}^{n} \left(\mathbf{d}_i - \overline{\mathbf{d}}\right) \left(\mathbf{d}_i - \overline{\mathbf{d}}\right)^\top$$

2. Under the assumptions that $\mathbf{d}_1, \cdots \mathbf{d}_n$ are independent samples from $N(\boldsymbol{\delta}, \boldsymbol{\Sigma})$, then $T^2 = n(\overline{\mathbf{d}} - \boldsymbol{\delta})^\top \mathbf{S}_d^{-1} (\overline{\mathbf{d}} - \boldsymbol{\delta}) \sim T^2_{p,n-1}$[1] $\sim \frac{(n-1)p}{n-p} F_{p,n-p}$
Same follow-up analyses as in one-sample test/intervals/regions apply here. For large sample inference, we can use χ^2_p.

[1] Here and below, we use the notation $T^2_{p,\nu}$ to represent a variable with distribution the same as $N_p(\mathbf{0}, \boldsymbol{\Sigma})^\top \left[\frac{1}{\nu} \mathbf{W}_{p,\nu}(\boldsymbol{\Sigma})\right]^{-1} N_p(\mathbf{0}, \boldsymbol{\Sigma})$

$T^2_{p,n-1}$ not Hotelling's T^2

$$H_0: \delta = 0 \begin{cases} T^2 > \boxed{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)} \text{ reject, Gaussian Case} \\ T^2 > \chi_p^2(\alpha) \text{ reject if } n \text{ is large, large sample case} \end{cases}$$

$d_i = X_{1i} - X_{2i}$ $\quad i = 1, 2, \cdots, n$

currently $\delta = E(d_i) = 0$

or $H_0: \mu_1 = \mu_2$

$$H_0: a^T\delta = 0 \begin{cases} \left| \dfrac{a^T\bar{d}}{\sqrt{\frac{a^TSa}{n}}} \right| > t_{n-1}\left(\frac{\alpha}{2}\right) \text{ Gaussian Case} \\ \left| \dfrac{a^T\bar{d}}{\sqrt{\frac{a^TSa}{n}}} \right| > Z\left(\frac{\alpha}{2}\right) \text{ large sample case} \end{cases}$$

quantile of Standard normal

$H_0: a_1^T\delta = 0, \cdots, a_n^T\delta = 0$

$$\frac{a_1^T\bar{d}}{\sqrt{\frac{a_1^TSa_1}{n}}} > t_{n-1}\left(\frac{\alpha_1}{2}\right), \cdots, \frac{a_n^T\bar{d}}{\sqrt{\frac{a_n^TSa_n}{n}}} > t_{n-1}\left(\frac{\alpha_n}{2}\right) \text{ when } \sum_{i=1}^n \alpha_i = \alpha$$

Gaussian Case

reject when at least one of them happen.

$$\frac{a_1^T\bar{d}}{\sqrt{\frac{a_1^TSa_1}{n}}} > Z\left(\frac{\alpha_1}{2}\right), \cdots, \frac{a_n^T\bar{d}}{\sqrt{\frac{a_n^TSa_n}{n}}} > Z\left(\frac{\alpha_n}{2}\right)$$

quantile of sd normal

reject when at least one of them happen.

Matched pair design

$d_i = X_{1i} - X_{2i}$ on assumpeion of i.i.d.

but in reality this is not reasonable because it is sometimes related to other facts.

In practice, an appropriate pairing of units and a randomized assignment of treatments can enhance the statistical analysis.

like HDX's D to flip the coin of average out accumulation

Experimental Design for Paired Comparisons

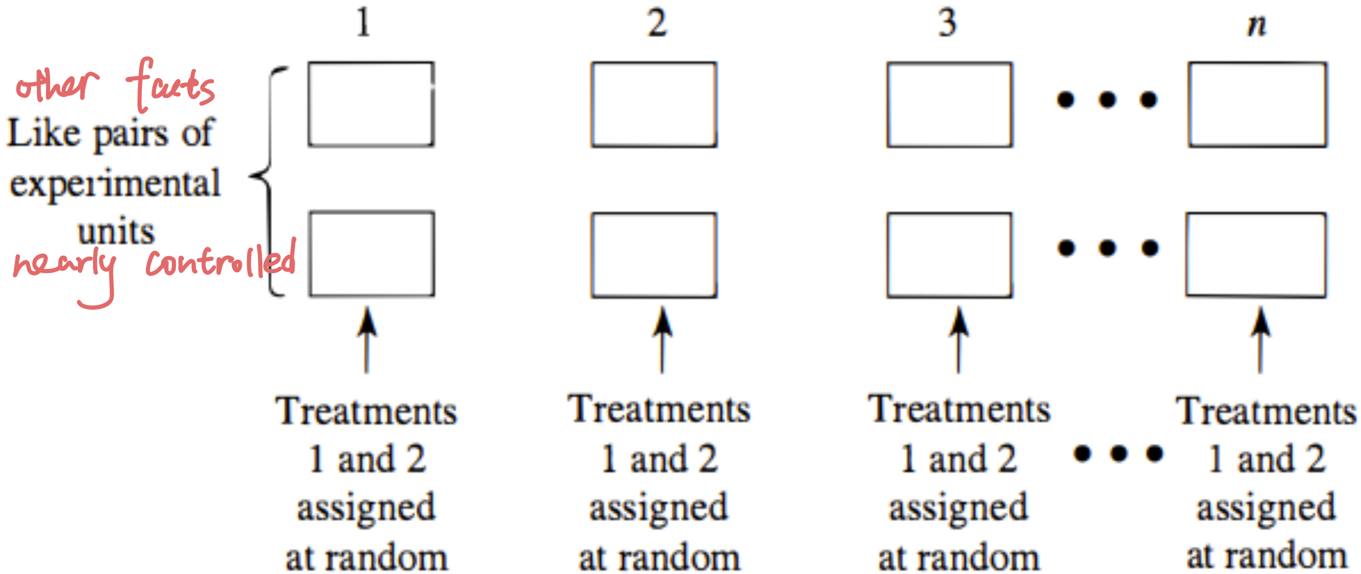


Figure: A random assignment of treatment 1 to one unit and treatment 2 to the other unit will help eliminate the systematic effects of uncontrolled sources of variation. Randomization can be implemented by flipping a coin. A separate independent randomization is conducted for each pair.

$p(\text{choosing Treatment}_i) = 0.5$ for $i = 1, 2$

Matched pair design

We can also think of each observation

$$\underset{2p\times 1}{\mathbf{x}_i} = \begin{bmatrix} \mathbf{x}_{1i} \\ \mathbf{x}_{2i} \end{bmatrix}$$

$$\underset{2p\times 1}{\overline{\mathbf{x}}} = \begin{bmatrix} \overline{\mathbf{x}}_1 \\ \overline{\mathbf{x}}_2 \end{bmatrix} \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\underset{2p\times 2p}{\mathbf{S}} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \qquad S = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{x})(X_i - \bar{x})^T$$

$$H_0 : E(X_{1i} - X_{2i}) = 0$$

Interest is in $\mathbf{C}\mathbf{x}_i$, where

name: contrast matrix

$$\underset{(p\times 2p)}{\mathbf{C}} = \begin{array}{c} \begin{array}{cccccccc} (1) & & & (p) & (p+1) & & & (2p) \end{array} \\ \begin{bmatrix} 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \end{array} \underset{(2p\times 1)}{\begin{bmatrix} X_{1i1} \\ \vdots \\ X_{1ip} \\ X_{2i1} \\ \vdots \\ X_{2ip} \end{bmatrix}} = \begin{bmatrix} X_{1i1} - X_{2i1} \\ \vdots \\ X_{1ip-1} - X_{2ip-1} \\ X_{1ip} - X_{2ip} \end{bmatrix}$$

$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$

Let $C = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & 1 & -1 & \\ & & & \ddots & \end{pmatrix}_{(k-1 \times k)}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}$ So $\mu_1 = \cdots = \mu_k \Leftrightarrow C\mu = 0$

Matched pair design comparison

Denote: $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$

$\underset{(p\times 2p)(2p\times 1)}{C\mu} = \mu_1 - \mu_2 \in R^{p\times 1}$

because $C\mu = \begin{pmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \\ \vdots \\ \mu_{k-1} - \mu_k \end{pmatrix} = 0 \Leftrightarrow \begin{cases} \mu_1 = \mu_2 \\ \mu_2 = \mu_3 \\ \vdots \\ \mu_{k-1} = \mu_k \end{cases} \Leftrightarrow \mu_1 = \cdots = \mu_k$

Note that

$$\mathbf{d}_i = \mathbf{C}\mathbf{x}_i$$

$$\bar{\mathbf{d}} = \mathbf{C}\bar{\mathbf{x}}$$

$$\mathbf{S}_d = \mathbf{C}\mathbf{S}\mathbf{C}^\top$$

because $C\bar{x}$ is approximation $C\mu$ and $E[\bar{x}] = \mu \Rightarrow E(C\bar{x}) = CE(\bar{x}) = C\mu$

$N_m(\mu, C\Sigma C^T)$

If $X_i \overset{i.i.d.}{\sim} N_p(\mu, \Sigma)$, for any $C \in R^{m\times p}$, $m<p$, the Hotelling's T^2 based on $\{Cx_1, Cx_2, \cdots, Cx_n\}$ which is $n[C(\bar{x}-\mu)]^T (CSC^T)^{-1} [C(\bar{x}-\mu)]$

$\sim \frac{(n-1)m}{n-m} F_{m,n-m}$ dist. doesn't dependent on C

$\sim T^2_{m,n-1}$

here $m = k-1$

because $C \in R^{k-1\times k}$

Under the assumption that $\mathbf{d}_i$ are independent $N(0, \mathbf{\Sigma})$

$$T^2 = n\bar{\mathbf{x}}^\top \mathbf{C}^\top \left(\mathbf{C}\mathbf{S}\mathbf{C}^\top\right)^{-1} \mathbf{C}\bar{\mathbf{x}} \sim T^2_{p,n-1}$$

$$\sim \frac{(n-1)p}{n-p} F_{p,n-p}$$

if C invertible

$n(\bar{x}-\mu)^T C^T (C^T)^{-1} S^{-1} C^{-1} C(\bar{x}-\mu)$

$= n(\bar{x}-\mu)^T S^{-1} (\bar{x}-\mu)$

no dependence on C

$$N(\mu, \Sigma)^T \left(\frac{1}{n-1} W(n-1, \Sigma)\right)^{-1} N(\mu, \Sigma)$$

$$\sim \frac{(n-1)p}{n-p} F_{p,n-p}$$

Table of Contents

Setting : Group 1: $X_{11}, X_{12}, \cdots, X_{1n_1}$ μ_1, Σ_1

n_1 observations, each has p features

Group 2: $X_{21}, X_{22}, \cdots, X_{2n_2}$ μ_2, Σ_2

n_2 observation, each has p features.

H_0 : $\mu_1 = \mu_2$

Comparing means from two populations

Interest in $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ (difference in two population means). Generalizing Hotelling's T^2 to two populations.

Assumptions:

- $\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1n_1}$ is a random sample of size n_1 from a p-variate population with mean vector $\boldsymbol{\mu}_1$ and covariance matrix $\boldsymbol{\Sigma}_1$.
- $\mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2n_2}$ is a random sample of size n_1 from a p-variate population with mean vector $\boldsymbol{\mu}_2$ and covariance matrix $\boldsymbol{\Sigma}_2$. [2]
- The two samples are independent

Let

$$\overline{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}, \quad i = 1, 2$$

$$\mathbf{S}_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \left(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i\right) \left(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i\right)^\top, \quad i = 1, 2$$

[2] Note that two groups do not need have the same number of samples

In addition: $\Sigma_1 = \Sigma_2$, X_{1i} X_{2j} MVN

If in addition assume $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ and two samples are MVN, then

$E[\bar{x}_1 - \bar{x}_2] = E[\bar{x}_1] - E[\bar{x}_2] = \mu_1 - \mu_2.$

$$(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) \sim N_p\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{\Sigma}\right).$$

$(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2) \sim N_p(0, Cov(\bar{x}_1 - \bar{x}_2)) = N_p(0, E[[(\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)][(\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)]^T])$

$N_p(0, Cov(\bar{x}_1) + Cov(\bar{x}_2) - Cov(\bar{x}_1, \bar{x}_2) - Cov(\bar{x}_2, \bar{x}_1))$

$= N_p(0, \frac{\Sigma}{n_1} + \frac{\Sigma}{n_2} - 0 - 0)$

$= N_p(0, (\frac{1}{n_1} + \frac{1}{n_2})\Sigma)$

Note that

$$(n_1 - 1)\,\mathbf{S}_1 \sim W_p(n_1 - 1, \boldsymbol{\Sigma})$$
$$(n_2 - 1)\,\mathbf{S}_2 \sim W_p(n_2 - 1, \boldsymbol{\Sigma})$$

and

$$\underbrace{(n_1 - 1)\,\mathbf{S}_1 + (n_2 - 1)\,\mathbf{S}_2}_{=(n_1+n_2-2)\mathbf{S}_{pooled}} \sim W_p(n_1 + n_2 - 2, \boldsymbol{\Sigma}).$$

Define

$$\mathbf{S}_{pooled} = \frac{\sum_{j=1}^{n_1}(\mathbf{x}_{1j} - \overline{\mathbf{x}}_1)(\mathbf{x}_{1j} - \overline{\mathbf{x}}_1)^\top + \sum_{j=1}^{n_2}(\mathbf{x}_{2j} - \overline{\mathbf{x}}_2)(\mathbf{x}_{2j} - \overline{\mathbf{x}}_2)^\top}{n_1 + n_2 - 2}$$
$$= \frac{n_1 - 1}{n_1 + n_2 - 2}\mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\mathbf{S}_2$$

then

$$E[\mathbf{S}_{pooled}] = \boldsymbol{\Sigma}$$

Test $H_0: \mu_1=\mu_2 \in R^p$

Population 1 $\quad x_{11}, x_{12} \cdots x_{1n_1}$ $\quad X_{1i} \sim N_p(\mu_1, \Sigma_1)$

Population 2 $\quad x_{21}, x_{22}, \cdots x_{2n_2}$ $\quad X_{2i} \sim N_p(\mu_2, \Sigma_2)$

If $\Sigma_1 = \Sigma_2 \ (=\Sigma)$, then

$$\bar{x}_1 - \bar{x}_2 \sim N_p\left(\mu_1-\mu_2, \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\Sigma\right)$$

How to estimate Σ?

$$S_1 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_{1i}-\bar{x}_1)(X_{1i}-\bar{x}_1)^T \quad E[S_1]=\Sigma$$

$$S_2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(X_{2i}-\bar{x}_2)(X_{2i}-\bar{x}_2)^T \quad E[S_2]=\Sigma$$

a better estimate:

$$S_{pooled} := \frac{n_1-1}{n_1+n_2-2}S_1 + \frac{n_2-1}{n_1+n_2-2}S_2, \quad E[S_{pooled}]=\Sigma$$

Use S_{pooled} to estimate Σ

$$[(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)]^T\Big[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\underbrace{\Sigma}_{\downarrow\, S_{pooled}}\Big]^{-1}[(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)] \sim \chi_p^2$$

$\sim T^2_{p,n_1+n_2-2}$

by: if $z \sim N_p(0,\Sigma)$ independent of $V \sim W_p(n-1,\Sigma)$

then $z^T\left(\frac{V}{n-1}\right)^{-1}z \sim T_{p,n-1}$

$\sim \frac{(n-1)p}{n-p}F_{p,n-p}$

$(n_1+n_2-2)S_{pooled}$

$= (n_1-1)S_1 + (n_2-1)S_2$

$= W_p(n_1-1,\Sigma) + W_p(n_2-1,\Sigma) \sim W_p(n_1+n_2-2,\Sigma)$

independent

Do Test:

So reject H_0 if

$$(\bar{x}_1-\bar{x}_2)^T\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_{pooled}\right]^{-1}(\bar{x}_1-\bar{x}_2) > \frac{(n_1+n_2-2)p}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1}(\alpha)$$

$$Cov(\bar{x}_1-\bar{x}_2) = \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\Sigma$$

$$\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,\Sigma)$$

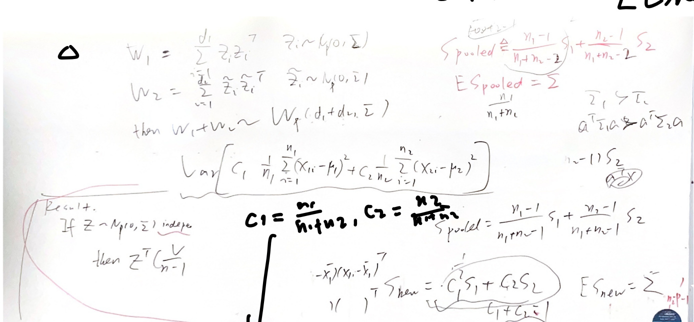
$W_1 = \sum_{i=1}^{d_1} z_i z_i^T$, $z_i \in N_p(0,\Sigma)$

$W_2 = \sum_{i=1}^{d_2} \tilde{z}_i \tilde{z}_i^T$, $\tilde{z}_i \in N_p(0,\Sigma)$

then $w_1 + w_2 \sim W_p(d_1+d_2, \Sigma)$

We may could $S_{new} := c_1S_1 + c_2S_2$

$c_1 + c_2 = 1$ $\quad E[S_{new}] = \Sigma$



$c_1 = \frac{n_1}{n_1+n_2}, \ c_2 = \frac{n_2}{n_1+n_2}$

Sample mean: $\bar{x}_2$: $\frac{n_2-1}{n_1+n_2-2}$

When $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$ and two samples are MVN

Recall that

$$(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) \sim N_p\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{\Sigma}\right)$$

then

$$\begin{aligned} T^2 &= [\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]^\top \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{pooled}\right]^{-1} [\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \\ &\sim T^2_{p,n_1+n_2-2} \\ &\sim \frac{(n_1 + n_2 - 2)\,p}{(n_1 + n_2 - 2) - p + 1} F_{p,(n_1+n_2-2)-p+1} \end{aligned}$$

Follow-up analyses

- "t-interval":

$$\mathbf{a}^\top \overline{\mathbf{x}}_1 - \mathbf{a}^\top \overline{\mathbf{x}}_2 \pm t_{n_1+n_2-2}(\frac{\alpha}{2})\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{a}^\top \mathbf{S}_{pooled}\mathbf{a}}$$

- "Bonferroni interval":

$$\mathbf{a}_i^\top \overline{\mathbf{x}}_1 - \mathbf{a}_i^\top \overline{\mathbf{x}}_2 \pm t_{n_1+n_2-2}\left(\frac{\alpha}{2k}\right)\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{a}_i^\top \mathbf{S}_{pooled}\mathbf{a}_i}$$

for $i = 1, \cdots, k$;

- "T^2-interval"

$$\mathbf{a}^\top \overline{\mathbf{x}}_1 - \mathbf{a}^\top \overline{\mathbf{x}}_2 \pm \sqrt{T^2_{\alpha,p,n_1+n_2-2}}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{a}^\top \mathbf{S}_{pooled}\mathbf{a}}$$

where $T^2_{\alpha,p,n_1+n_2-2} \equiv \frac{(n_1+n_2-2)p}{(n_1+n_2-2)-p+1}F_{p,(n_1+n_2-2)-p+1}(\alpha)$

When $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$

- When $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, only assuming MVN is not enough, so we consider the large sample case.
- As $n_1 \to \infty$, $n_2 \to \infty$, by the central limit theorem, $\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2$ is approximately $N_p\left[\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, n_1^{-1}\boldsymbol{\Sigma}_1 + n_2^{-1}\boldsymbol{\Sigma}_2\right]$, thus

$$[\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]^\top \left(\frac{1}{n_1}\boldsymbol{\Sigma}_1 + \frac{1}{n_2}\boldsymbol{\Sigma}_2\right)^{-1} [\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \to \chi^2_p$$

  When n_1 and n_2 are large, with high probability, $\mathbf{S}_1 \to \boldsymbol{\Sigma}_1$, $\mathbf{S}_2 \to \boldsymbol{\Sigma}_2$. Consequently, the approximation holds with $\mathbf{S}_1$ and $\mathbf{S}_2$ in place of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, respectively; and we have

$$[\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]^\top \left[\frac{\mathbf{S}_1}{n_1} + \frac{\mathbf{S}_2}{n_2}\right]^{-1} [\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \to \chi^2_p$$

- Follow-up analysis in test/confidence intervals/regions apply.

If $x \sim N_p(\mu, \Sigma)$

$(x-\mu)^T \Sigma^{-1} (x-\mu) \sim \chi_p^2$

When $\Sigma_1 \neq \Sigma_2$

As $n \to \infty$

$$[\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)]^T \left(\frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2\right)^{-1} [\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)] \to \chi_p^2$$

$$\frac{1}{n_1}S_1 + \frac{1}{n_2}S_2$$

$\to \chi_p^2$ ✓

$S \to \Sigma$

no nail biting

Table of Contents

Comparing of several means

A medical study that aims to compare the effectiveness of several different treatment approaches (Treatment A, Treatment B, ...) for managing a particular chronic disease, such as diabetes. The study could assess various health markers, including blood sugar levels, cholesterol levels, and blood pressure.
The collected data can be organized as

		Treatment			
		1	2	$\dots$	g
	1	$\mathbf{x}_{11}$	$\mathbf{x}_{21}$	$\dots$	$\mathbf{x}_{g_1}$
	2	$\mathbf{x}_{12}$	$\mathbf{x}_{22}$	$\dots$	$\mathbf{x}_{g_2}$
Subjects	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	n_i	$\mathbf{x}_{1n_1}$	$\mathbf{x}_{2n_2}$	$\dots$	$\mathbf{x}_{gn_g}$

where each $\mathbf{x}_{\ell j}$ is p-variate vector.

Basic Assumptions

Assumptions needed for statistical inference.

- $\mathbf{x}_{l1}, \mathbf{x}_{l2}, \ldots, \mathbf{x}_{ln_l}$ is a random sample of size n_l from a population with means μ_l for $l = 1, \ldots, g$ (i.e., observations within populations are independent and representative of their populations).
- Random samples from different populations are independent.
- All populations have the same covariance matrix, $\mathbf{\Sigma}$.
- $\mathbf{x}_{lj} \sim \mathcal{N}(\boldsymbol{\mu}_l, \mathbf{\Sigma})$; that is, each population is multivariate normal.

If a population is not multivariate normal, then for large n_l central limit theorem may "kick-in".

Aim: $H_0: \mu_1 = \mu_2 = \cdots = \mu_g$

$X_{lj} \sim N(\mu_l, \Sigma)$

no nail biting

MANOVA

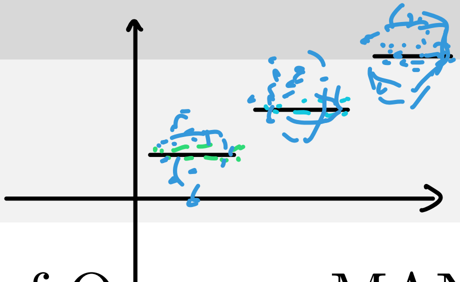
multivariate analysis of variance.

p>1 : each X_{lj} is a vector.

One-way ANOVA

$\mu_b = \mu + \mu_b - \mu$ b-treatment effect.

common mean for all groups
ground mean



We start by considering the case where $p = 1$ of One-way MANOVA, also known as One-way ANOVA

- Assumptions: $X_{lj} \sim \mathcal{N}\left(\mu_l, \sigma^2\right)$ i.i.d for $j = 1, \ldots, n_l$ and $l = 1, \ldots, g$.
- Hypotheses:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_g \quad \text{versus} \quad H_1 : \text{ not } H_0$$

- We usually express μ_l as the sum of a grand mean and deviations from the grand mean

$$\underbrace{\mu_l}_{\ell^{\text{th}} \text{ pop. mean}} = \underbrace{\mu}_{\text{grand mean}} + \underbrace{\mu_l - \mu}_{\ell^{\text{th}} \text{ pop. treatment effect}}$$
$$= \mu + \tau_l$$

- If $\mu_1 = \mu_2 = \cdots = \mu_g$, then an equivalent way to write the null hypothesis is

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0$$

The model for an observation

$$X_{lj} = \mu + \tau_l + \epsilon_{lj}$$

where $\epsilon_{lj} \sim \mathcal{N}\left(0, \sigma^2\right)$ and independent.

- ϵ_{lj} is "random error".
- We typically impose the condition $\sum_{l=1}^{g} \tau_l = 0$ as an identification constraint.
- The decomposition of an observation is

$$\underbrace{X_{lj}}_{\text{observation}} = \underbrace{\bar{X}}_{\substack{\text{overall}\\ \text{sample}\\ \text{mean}}} + \underbrace{\left(\bar{X}_l - \bar{X}\right)}_{\substack{\text{estimated}\\ \text{treatment}\\ \text{effect}}} + \underbrace{\left(X_{lj} - \bar{X}_l\right)}_{\substack{\text{residual}\\ \text{"error"}}}$$

- $\bar{X}$ is the estimator of μ
- $\hat{\tau}_l = \left(\bar{X}_l - \bar{X}\right)$ is the estimator of τ_l
- $\left(X_{lj} - \bar{X}_l\right)$ is the estimator of ϵ_{lj}.

Sum of squares

The total sum of squares:

$$SS_{\text{total}} = \sum_{l=1}^{g} \sum_{j=1}^{n_l} (X_{lj} - \bar{X})(X_{lj} - \bar{X})^\top$$

Remark: Here we use the "transpose" to indicate that the decomposition discussed in the next two slides also applies to multivariate cases, i.e., you can directly replace X_{lj} with $\mathbf{x}_{lj}$, $\bar{X}_l$ with $\bar{\mathbf{x}}_l$...

Note that

$$\begin{aligned}
&(X_{lj} - \bar{X})(X_{lj} - \bar{X})^\top = \left[(X_{lj} - \bar{X}_l) + (\bar{X}_l - \bar{X})\right]\left[(X_{lj} - \bar{X}_l) + (\bar{X}_l - \bar{X})\right]^\top \\
&= \underbrace{(X_{lj} - \bar{X}_l)(X_{lj} - \bar{X}_l)^\top + (\bar{X}_l - \bar{X})(\bar{X}_l - \bar{X})^\top}_{\text{squares-products}} \\
&+ \underbrace{(X_{lj} - \bar{X}_l)(\bar{X}_l - \bar{X})^\top + (\bar{X}_l - \bar{X})(X_{lj} - \bar{X}_l)^\top}_{\text{cross-products}}
\end{aligned}$$

Sum of cross-products terms

$$\sum_{j=1}^{n_l} \left(X_{lj} - \overline{X}_l\right)\left(\overline{X}_l - \overline{X}\right)^\top = \left(\sum_{j=1}^{n_l} \left(X_{lj} - \overline{X}_l\right)\right)\left(\overline{X}_l - \overline{X}\right)^\top$$

$$= \left(\left(\sum_{j=1}^{n_l} X_{lj}\right) - n_l \overline{X}_l\right)\left(\overline{X}_l - \bar{X}\right)^\top$$

$$= n_l \underbrace{\left(\bar{X}_l - \bar{X}_l\right)}_{0}\left(\bar{X}_l - \bar{X}\right)^\top = 0$$

Sum of Squares decomposition

Now summing the rest over j and ℓ we get

$$\sum_{l=1}^{g}\sum_{j=1}^{n_l}\left(X_{lj}-\bar{X}\right)\left(X_{lj}-\bar{X}\right)^\top = \sum_{l=1}^{g} n_l\left(\overline{X}_l-\overline{X}\right)\left(\overline{X}_l-\overline{X}\right)^\top$$

$$+\sum_{l=1}^{g}\sum_{j=1}^{n_l}\left(X_{lj}-\overline{X}_l\right)\left(X_{lj}-\overline{X}_l\right)^\top$$

$$\begin{aligned}\text{Total SS} &= \text{Treatment} + \text{Residual}\\ &= \text{Between Groups} + \text{Within Groups}\\ &= \text{Hypothesis} + \text{Error}\end{aligned}$$

ANOVA table and test

Let $n_+ = \sum_{l=1}^{g} n_l$, the total sample size

Source of Variation	Sum of Squares	df
Treatment	$SS_{tr} = \sum_{l=1}^{g} n_l \left(\bar{X}_l - \bar{X}\right)^2$	$g-1$
Residual	$SS_{res} = \sum_{l=1}^{g} \sum_{j=1}^{n_l} \left(X_{lj} - \bar{X}_l\right)^2$	$n_+ - g$
Total	$= \sum_{l=1}^{g} \sum_{j=1}^{n_l} \left(X_{lj} - \bar{X}\right)^2$	$n_+ - 1$

Test statistic for $H_0 : \mu_1 = \cdots = \mu_g$ (or $H_0 : \tau_1 = \cdots = \tau_g$) and its sampling distribution are

$$F = \frac{SS_{tr}/(g-1)}{SS_{\text{res}} \,/\, (n_+ - g)} \sim \mathcal{F}_{(g-1),(n_+-g)}$$

ANOVA test

The ANOVA F-test rejects $H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0$ at level α if

$$F = \frac{SS_{\mathrm{tr}}/(g-1)}{SS_{\mathrm{res}}/\left(\sum_{\ell=1}^{g} n_\ell - g\right)} > F_{g-1,\Sigma n_\ell - g}(\alpha)$$

This is equivalent to rejecting H_0 for large values of $SS_{\mathrm{tr}}/SS_{\mathrm{res}}$ or for large values of $1 + SS_{\mathrm{tr}}/SS_{\mathrm{res}}$. The statistic appropriate for a multivariate generalization rejects H_0 for small values of the reciprocal

$$\frac{1}{1 + SS_{\mathrm{tr}}/SS_{\mathrm{res}}} = \frac{SS_{\mathrm{res}}}{SS_{res} + SS_{\mathrm{tr}}}$$

A closer look at Within group SS

$$
\begin{aligned}
\mathbf{W} =& \sum_{l=1}^{g} \sum_{j=1}^{n_l} \left(\mathbf{x}_{lj} - \overline{\mathbf{x}}_l\right) \left(\mathbf{x}_{lj} - \overline{\mathbf{x}}_l\right)^\top \\
=& \sum_{j=1}^{n_1} \left(\mathbf{x}_{1j} - \overline{\mathbf{x}}_1\right) \left(\mathbf{x}_{1j} - \overline{\mathbf{x}}_1\right)^\top + \sum_{j=1}^{n_2} \left(\mathbf{x}_{2j} - \overline{\mathbf{x}}_2\right) \left(\mathbf{x}_{2j} - \overline{\mathbf{x}}_2\right)^\top \\
& \cdots + \sum_{j=1}^{n_g} \left(\mathbf{x}_{gj} - \overline{\mathbf{x}}_g\right) \left(\mathbf{x}_{gj} - \overline{\mathbf{x}}_g\right)^\top \\
=& \mathbf{W}_1 + \mathbf{W}_2 + \ldots + \mathbf{W}_g \\
=& \left(n_1 - 1\right) \mathbf{S}_1 + \left(n_2 - 1\right) \mathbf{S}_2 + \cdots + \left(n_g - 1\right) \mathbf{S}_g
\end{aligned}
$$

$\mathbf{S}_i$ is the sample covariance matrix for the i-th group (treatment, condition, etc).

Test statistic

With respect to between groups SS,

$$\mathbf{B} = \sum_{l=1}^{g} n_l \left(\overline{\mathbf{x}}_l - \overline{\mathbf{x}}\right)\left(\overline{\mathbf{x}}_l - \overline{\mathbf{x}}\right)^\top = \sum_{l=1|}^{g} n_l \hat{\tau}_l \hat{\tau}_l^\top$$

- If $H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = \mathbf{0}$ is true, Then $\mathbf{B}$ should be "close" to $\mathbf{0}$.
- To test H_0, we consider the ratio of generalized SSs,

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|} = \frac{|\mathbf{W}|}{|\mathbf{T}|}$$

  where $\mathbf{T} = \mathbf{W} + \mathbf{B}$.
- Λ^* is known as "Wilk's Lambda".
- It's equivalent to likelihood ratio statistic.

Hypothesis testing with Λ^*

Λ^* is a ratio of generalized sampling variances

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{\prod_{i=1}^{p} \lambda_i}{\prod_{i=1}^{p} \lambda_i^*}$$

- where λ_i are eigenvalues of $\mathbf{W}$, and λ_i^* are eigenvalues of $\mathbf{T}$.
- If $H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0$ is true then B is close to 0

$$\Rightarrow \mathbf{T} \approx \mathbf{W} \Rightarrow \lambda_i \approx \lambda_i^* \implies \Lambda^* \text{ close to } 1.$$

- If $H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = \mathbf{0}$ is false then $\mathbf{B}$ is not close $\mathbf{0}$ $\implies \lambda_i < \lambda_i^* \implies \Lambda^*$ is "small".

The exact distribution of Λ^* can be derived for special cases of p and g.

Distribution of Wilks's Lambda Λ^*

No. of variables	No. of groups	Sampling distribution
$p = 1$	$g \geq 2$	$\left(\frac{\sum n_\ell - g}{g-1}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{g-1, \Sigma n_\ell - g}$
$p = 2$	$g \geq 2$	$\left(\frac{\sum n_\ell - g - 1}{g-1}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1), 2(\Sigma n_\ell - g - 1)}$
$p \geq 1$	$g = 2$	$\left(\frac{\sum n_\ell - p - 1}{p}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{p, \Sigma n_\ell - p - 1}$
$p \geq 1$	$g = 3$	$\left(\frac{\sum n_\ell - p - 2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p, 2(\sum n_\ell - p - 2)}$

If H_0 is true and $\sum n_\ell = n$ is large,

$$-\left(n - 1 - \frac{(p+g)}{2}\right) \ln \Lambda^*$$

has approximately a chi-square distribution with $p(g-1)$ d.f.

Summary of One-way MONAVA

MANOVA table:

Source of variation	Sum of Squares	df	Λ^*
Treatment (Between)	$\mathbf{B} = \sum_{l=1}^{g} n_l \left(\bar{\mathbf{x}}_l - \bar{\mathbf{x}}\right)\left(\bar{\mathbf{x}}_l - \bar{\mathbf{x}}\right)^\top$	$g-1$	$\lvert\mathbf{W}\rvert / \lvert\mathbf{T}\rvert$
Residual (Within)	$\mathbf{W} = \sum_{l=1}^{g} \sum_{j=1}^{n_l} \left(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l\right)\left(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l\right)^\top$	$n-g$	
Total (corrected for mean)	$\mathbf{T} = \mathbf{W} + \mathbf{B}$ $= \sum_{l=1}^{g} \sum_{j=1}^{n_l} \left(\mathbf{x}_{lj} - \bar{\mathbf{x}}\right)\left(\mathbf{x}_{lj} - \bar{\mathbf{x}}\right)^\top$	$n-1$	

Reject H_0 when the above-mentioned statistics involving Λ^* is greater than the associated critical values.