Chapter 4: Inferences about a mean vector

Zhixiang ZHANG

University of Macau

zhixzhang@um.edu.mo

February 10, 2025

Motivating example

In 1985, the USDA commissioned a study of women's nutrition. Nutrient intake was measured for a random sample of 737 women aged 25 – 50 years. The table below shows the recommended daily intake of five nutritional components and the sample means for all variables:

Variable	Recommended Intake (μ_o)	Mean
Calcium	1000mg	624.0mg
Iron	15mg	11.1mg
Protein	60 g	65.8 g
Vitamin A	800μg	839.6μg
Vitamin C	75mg	78.9mg

One of the questions of interest is whether women meet the federal nutritional intake guidelines.

Motivating example

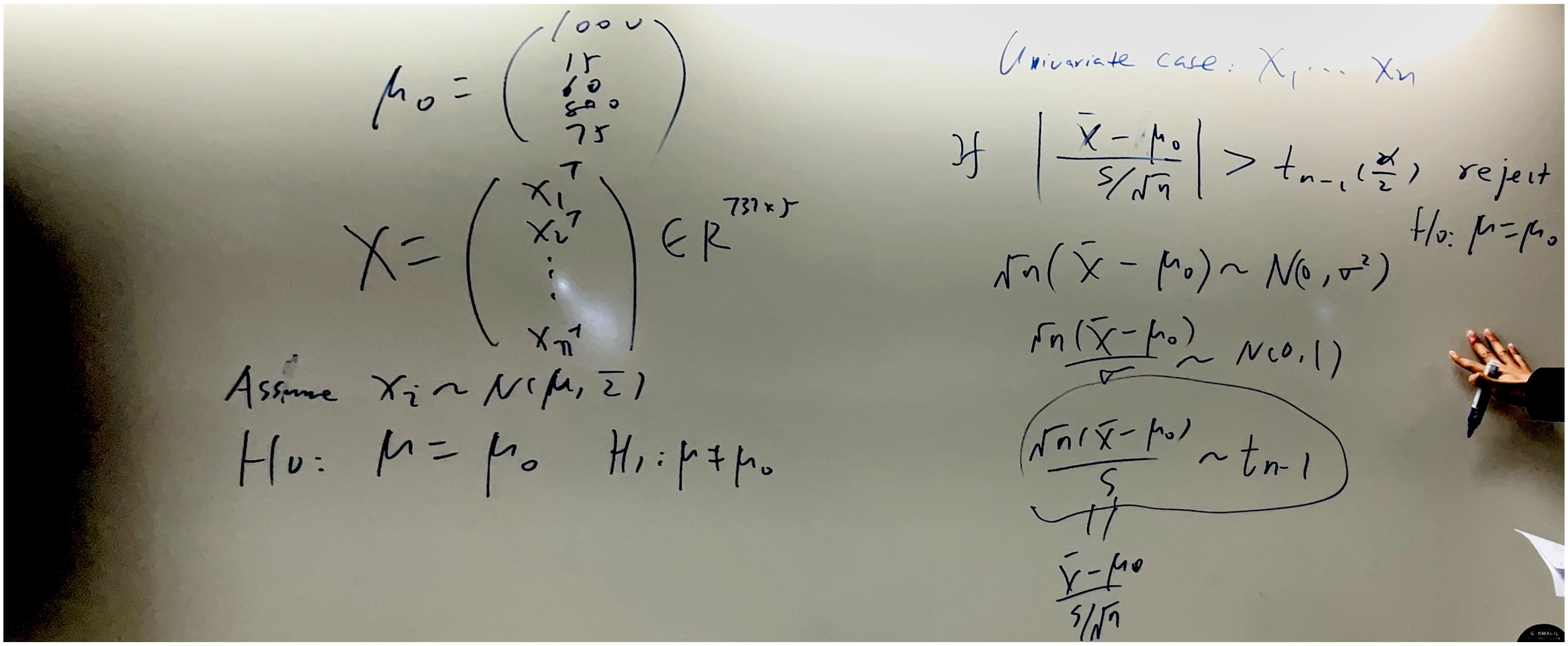
The hypothesis of interest is that women meet nutritional standards for all nutritional components. This null hypothesis would be rejected if women fail to meet nutritional standards on any one or more of these nutritional variables.
In mathematical notation, the null hypothesis is the population mean vector $\boldsymbol{\mu}$ equals the hypothesized mean vector $\boldsymbol{\mu}_0$ as shown below:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$$

Recap: Inference on univariate mean

First recalling the univariate theory for determining whether a specific value μ_0 is a plausible value for the population mean μ, i.e., consider the hypothesis test: $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$.
If $X_1, X_2, \ldots, X_n$ denote a random sample from a normal population (with known variance σ^2), the appropriate test statistic is

$$z = \frac{(\bar{X} - \mu_0)}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \text{ under } H_0.$$



$\bar{x} = \begin{bmatrix} \bar{x}_1^T \\ \bar{x}_2^T \\ \vdots \\ \bar{x}_n^T \end{bmatrix}$ and $\mu_0 = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0n} \end{bmatrix}$ are not easy to compare when it is in multidimensional.

Recap

Usually in practice, σ is not known, use

$$t = \frac{(\bar{X} - \mu_0)}{s/\sqrt{n}} \sim t_{n-1}, \text{ where } s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2$$

We do not reject H_0 at level α if

$$\left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right| \leq t_{n-1}(\alpha/2).$$

The $100(1-\alpha)\%$ confidence interval is given as

$$\left[\bar{x} - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}\ ,\ \bar{x} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}\right]$$

Multivariate case

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \qquad \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right|^2 \sim \chi_1^2$$

In the univariate case, by taking the square, $\Leftrightarrow n\frac{(\bar{x}-\mu_0)^2}{\sigma^2} \sim \chi_1^2$

$$n\frac{(\bar{X}-\mu_0)^2}{\sigma^2} \sim \chi_1^2 \text{ under } H_0.$$

$Q = \sum_{i=1}^{d} Z_i^2$ where $Z_i \overset{iid}{\sim} N(0,1)$ then $Q \sim \chi_d^2$

Multivariate generalization ($\boldsymbol{\Sigma}$ is known):

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix} \quad \text{vs.} \quad \underbrace{H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0}_{\substack{\text{At least one } \mu_i \\ \text{is not equal to } \mu_{0i}}}$$

$$n\left(\bar{\mathbf{x}}-\boldsymbol{\mu}_0\right)^\top \boldsymbol{\Sigma}^{-1}\left(\bar{\mathbf{x}}-\boldsymbol{\mu}_0\right) \sim \chi_p^2 \quad \text{under } H_0$$

Maha... distance of $\bar{X}$ to D with mean μ_0 and covariance $\frac{\Sigma}{n}$

Why $n(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}_0) \sim \chi^2_p$

$\operatorname{dist}_M(X,\mu) = \sqrt{(X-\mu)^T\Sigma^{-1}(X-\mu)}$ X to Q with mean μ and $\mathrm{Cov}=\Sigma$

In our situation: Under H_0: $\bar{X} \sim N(\mu_0, \frac{\Sigma}{n})$

So is Maha distance of $\bar{X}$ to Q with mean μ_0 and $\mathrm{Cov}=\frac{\Sigma}{n}$

Square

If $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

prove: Denote: $Z = \Sigma^{\frac{1}{2}}(X-\mu) \in R^p$

$Z = \begin{pmatrix} z_1 \\ \vdots \\ z_p \end{pmatrix} \sim N(0, I_p)$ standard normal z_i and independent

$\Sigma^{-\frac{1}{2}}(X-\mu) \sim N(0, I_p)$

$$(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \underbrace{\left[\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x}-\boldsymbol{\mu})\right]^\top}_{\mathbf{z}^\top} \underbrace{\left[\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x}-\boldsymbol{\mu})\right]}_{\sim N_p(\mathbf{0},\mathbf{I})}$$

$X-\mu \sim N(0,\Sigma)$

$\Sigma^{-\frac{1}{2}}(X-\mu) \sim N(0, \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}})$

$\Leftrightarrow \Sigma^{-\frac{1}{2}}(X-\mu) \sim N(0, I_p)$

then $Z^TZ = (X-\mu)^T(\Sigma^{-\frac{1}{2}})^T\Sigma^{-\frac{1}{2}}(X-\mu)$

$$= \sum_{i=1}^p z_i^2 \quad \text{(sum of } p \text{ indep. squared } \mathcal{N}(0,1))$$

$= (X-\mu)^T\Sigma^{-1}(X-\mu)$

$$\stackrel{d}{=} \chi^2_p$$

$\Leftrightarrow \sum_{i=1}^p z_i^2 = (X-\mu)^T\Sigma^{-1}(X-\mu)$

We know that under H_0, $n^{1/2}(\bar{\mathbf{x}}-\boldsymbol{\mu}_0) \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, then use above.

$\chi_p^2 \stackrel{d}{=} (n^{1/2}(\bar{x}-\mu_0)-0)^T\Sigma^{-1}(n^{1/2}(\bar{x}-\mu_0)-0) \stackrel{d}{=} \chi_p^2$

$n(\bar{x}-\mu_0)^T\Sigma^{-1}(\bar{x}-\mu_0) \stackrel{d}{=} \chi_p^2$

shown

$\bar{x} \sim N(\mu_0, \frac{\Sigma}{n})$

then $(\bar{x}-\mu_0)^T(\frac{\Sigma}{n})^{-1}(X-\mu_0) \sim \chi_p^2$

$n(\bar{x}-\mu_0)^T\Sigma^{-1}(X-\mu_0) \sim \chi_p^2$ shown.

Hotelling's T^2

$n(\bar{x}-\mu_0)^T S^{-1}(\bar{x}-\mu_0) \sim F \ldots\ldots$

$\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \sim N(0,1)$ $\quad \frac{\sqrt{n}(\bar{x}-\mu_0)}{S} \sim t_{n-1} \stackrel{d}{=} \frac{N(0,1)}{\sqrt{\frac{\sum z_i^2}{n-1}}}$

$n(\bar{X}-\mu_0)^T \Sigma^{-1}(x_0-\mu) \sim N(0,1)^2$ $\quad n(\bar{x}-\mu_0)^T S^{-1}(x_0-\mu)$

More frequently in practice, $\boldsymbol{\Sigma}$ is unknown.
A natural generalization of the squared distance is its multivariate analog

$$\underline{T^2 = (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \left(\frac{1}{n}\mathbf{S}\right)^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) = n\,(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)} \sim F \ldots\ldots$$

distribution might have some change like t^2

where

$$\underset{(p\times 1)}{\bar{\mathbf{x}}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j, \qquad \underset{(p\times p)}{\mathbf{S}} = \frac{1}{n-1}\sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^\top$$

The statistic T^2 is called Hotelling's T^2 in honor of Harold Hotelling, a pioneer in multivariate statistics

Some properties of Hotelling's T^2

X_i^T is either far left or right of μ_0

- If the observed statistical distance T^2 is too large-that is, if $\bar{\mathbf{x}}$ is "too far" from $\boldsymbol{\mu}_0$-the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is rejected.
- Alternative hypothesis is two-sided (no such thing as "$H_1 : \boldsymbol{\mu} > \boldsymbol{\mu}_0$ ")
- We must have $n > p$ otherwise $\mathbf{S}^{-1}$ does not exist.

$p = 20$, $n = 40$ ×

S^{-1} does not exist if $n \le p$

$p = 20$, $n = 200$ ✓
1000

$S = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T$

normal $\frac{p}{n} < 0.1$

fact: $\forall A$, $\text{rank}(AA^T) = \text{rank}(A^TA) = \text{rank}(A)$

Distribution of Hotelling's T^2

rank(S) = rank $\left(\begin{bmatrix} x_1-\bar{x} \\ x_2-\bar{x} \\ \vdots \\ x_n-\bar{x} \end{bmatrix} \right)$

$S \in R^{p \times p}$

$= \text{rank}(B(1)) = \text{rank}((x_1-\bar{x}, x_2-\bar{x}, x_3-\bar{x}, \cdots, x_n-\bar{x}))$

$\leq \min\{n-1, p\}$

if $X \in R^{10 \times 20}$

$(X^T X)^{-1}$ not exist

If $n-1<p$ then rank(S) $= n-1 < p$ $\therefore$ not full rank because $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

$\therefore S^{-1}$ not exist

Under H_0,

$X^T X \in R^{p \times p}$ if $n<p$

$$T^2 \text{ is distributed as } \frac{(n-1)p}{(n-p)} F_{p,n-p}$$

$\text{rank}(X^T X) = \text{rank}(X) \leq \min(n, p) = n < p$

where $F_{p,n-p}$ denotes a random variable with an F-distribution with p and $n-p$ d.f. At the α level of significance, we reject H_0 if the observed

$\therefore X^T X$ not full rank as $X^T X \in R^{p \times p}$ in inverteible

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

$F_{1,n-1} = (t_{n-1})^2$

Nutrition example

The sample mean is given in the table on the previous page, the sample variance-covariance matrix:

$$S = \begin{pmatrix} 157829.4 & 940.1 & 6075.8 & 102411.1 & 6701.6 \\ 940.1 & 35.8 & 114.1 & 2383.2 & 137.7 \\ 6075.8 & 114.1 & 934.9 & 7330.1 & 477.2 \\ 102411.1 & 2383.2 & 7330.1 & 2668452.4 & 22063.3 \\ 6701.6 & 137.7 & 477.2 & 22063.3 & 5416.3 \end{pmatrix}$$

Hotelling's T-square comes out to be: $T^2 = 1758.54$. Compute $\frac{p(n-1)}{n-p}F_{p,n-p}(\alpha) = \frac{5(737-1)}{737-5}F_{5,732}(0.01) = 15.29$. Since $T^2 > 15.29$, we can reject the null hypothesis that the average dietary intake meets recommendations at the significance level of 0.01.

Another perspective on the T^2 statistic

We introduced the T^2-statistic by analogy with the univariate squared distance t^2.

There is a general principle for constructing test procedures called the likelihood ratio method, and the T^2-statistic can be derived as the likelihood ratio test of H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$.

Likelihood ratio tests have several optimal properties for reasonably large samples. (beyond the scope of this course)

$X_i \in R^p$, $i = 1, \dots, n$ $\quad X_i \overset{iid}{\sim} N(\mu, \Sigma)$

$H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$

$$\Lambda = \frac{\max_{\mu = \mu_0} L(\mu)}{\max_{\mu \in R^p} L(\mu)}$$

$H_0: \theta = \theta_0$

Generally: $\Lambda = \dfrac{\max_{\theta = \theta_0} L(\theta)}{\max_{\theta \in R^p} L(\theta)}$ (close to 1 if H_0 is true)

$(0 \leq \Lambda \leq 1)$

Intuitive understanding

$H_0: \mu = \mu_0$ (suppose Σ is known)

$$L(\mu = \mu_0) = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{\frac{\sum_{i=1}^{n} (x_i - \mu_0)^T \Sigma^{-1} (x_i - \mu_0)}{2}}$$

If H_0 is true, then we expect $\bar{x} \approx \mu_0$

$$L(\mu = \mu_0) \approx \max_{\mu \in R^p} L(\mu) = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^T \Sigma^{-1} (x_i - \bar{x})}{2}}$$

Basic idea of Likelihood Ratio Tests

Basic idea of Likelihood Ratio Tests:

- Θ_0 = a set of unknown parameters under H_0 (e.g., $\boldsymbol{\Sigma}$).
- Θ = the set of unknown parameters under the alternative hypothesis (model), which is more general (e.g., $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$).
- $\mathcal{L}(\cdot)$ is the likelihood function. It is a function of parameters that indicates "how likely Θ (or Θ_0) is given the data".

The general form of Likelihood Ratio Statistic is

$$\Lambda = \frac{\max \mathcal{L}(\Theta_0)}{\max \mathcal{L}(\Theta)}$$ general approach

- $\Lambda \in [0, 1]$.
- If Λ is relatively small, then the data are not likely to have occurred under $H_0 \longrightarrow$ Reject H_0.

Likelihood ratio test for mean vector

- We know that the maximum of the multivariate normal likelihood as $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are varied over their possible values is given by

$$\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\hat{\boldsymbol{\Sigma}}|^{n/2}} e^{-np/2}$$

  with MLE:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{j=1}^{n} (\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})^\top \quad \text{and } \hat{\boldsymbol{\mu}} = \overline{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$$

- Under $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, the likelihood specializes to

$$\max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}\left|\hat{\boldsymbol{\Sigma}}_0\right|^{n/2}} e^{-np/2}$$

  with MLE

$$\hat{\boldsymbol{\Sigma}}_0 = \frac{1}{n}\sum_{j=1}^{n} (\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)^\top$$

Likelihood ratio test for mean vector

Therefore, the likelihood ratio test of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ rejects H_0 if

$$\Lambda = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{\left|\hat{\boldsymbol{\Sigma}}_0\right|} \right)^{n/2} = \left(\frac{\left| \sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^\top \right|}{\left| \sum_{j=1}^{n} (\mathbf{x}_j - \boldsymbol{\mu}_0)(\mathbf{x}_j - \boldsymbol{\mu}_0)^\top \right|} \right)^{n/2} < c_\alpha$$

where c_α is the lower (100α)-th percentile of the distribution of Λ.

Result. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Then the Hotelling's T^2 test is equivalent to the likelihood ratio test of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ because

Specisific case

$x_1, \ldots, x_n, \ldots \quad n \to \infty$

$-2\log \dfrac{\max_{\theta=\theta_0} \prod_{i=1}^{n} f_i(\theta)}{\max_{\theta \in R^p} \prod f(\theta)} \sim \chi^2_d$

because this is

because of linear algebra.

$$\Lambda^{2/n} = \left(1 + \frac{T^2}{(n-1)}\right)^{-1}.$$

$\Lambda \to 1 \Rightarrow T^2 \to 0$

which means we accept H_0

$X_i \in R^p \quad i = 1 \ldots n \quad X_i \overset{i.i.d.}{\sim} N(\mu, \Sigma)$

$H_0: \mu = \mu_0, \ H_1: \mu \neq \mu_0$

$\Lambda = \dfrac{\max_{\mu=\mu_0} L(\mu)}{\max_{\mu \in R^p} L(\mu)}$

$\dfrac{\max_{\mu=\mu_0, \Sigma} L(\mu, \Sigma)}{\max_{\mu, \Sigma} L(\mu, \Sigma)}$

$H_0: \theta = \theta_0$

Generally: $\Lambda = \dfrac{\max_{\theta=\theta_0} L(\theta_0)}{\max_{\theta \in R^p} L(\theta)}$ (close to 1 if H_0 is true)

$(0 \le \Lambda \le 1)$

Compare the goodness of fit of data to model under H_0 and full parameter space

If H_0 is true $L(\mu = \mu$

the general definition : if T^2 is inference tend to reject H_0.

Confidence region

Confidence "interval" Region.

- Our goal is to make inferences about populations from samples. In univariate statistics, we form confidence intervals; we'll generalize this to multivariate confidence region.
- A *confidence region* is a region of likely true parameter $\boldsymbol{\theta}$ (in this Chapter, $\boldsymbol{\theta}$ is $\boldsymbol{\mu}$). This region is determined by the data, and for the moment, we shall denote it by $R(\mathbf{X})$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^\top$ is the data matrix.
- The region $R(\mathbf{X})$ is said to be a $100(1-\alpha)\%$ confidence region if, before the sample is selected,

if give specific sample points, the resulted P should be a fixed value.

random quantity.

$$P[R(\mathbf{X}) \text{ will cover the true } \boldsymbol{\theta}] = 1 - \alpha$$

Here is a P that is determined without sampling, depending on R(X) and θ

This probability is calculated under the true, but unknown, value of $\boldsymbol{\theta}$.

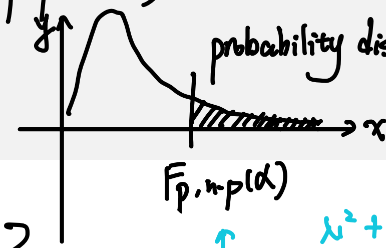
R(X) is r.v. and θ is unknown.

$P(\mu \in R(x)) = 1-\alpha$

$n(\bar{x}-\mu)^T S^{-1}(\bar{x}-\mu) \sim \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$

Confidence region of $\boldsymbol{\mu}$

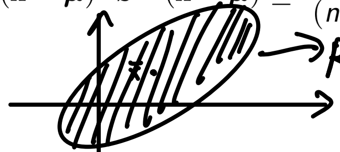
$P(\mu \in R(x)) = 1-\alpha$



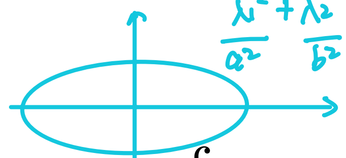
right quantile

$$\left\{\boldsymbol{\mu} : n(\bar{\mathbf{x}}-\boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \le \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right\}$$

If $p=2$:

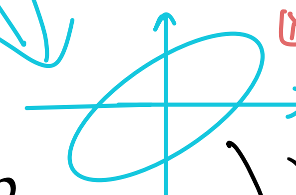


why ellipsoid?



$\frac{\lambda_1^2}{a^2}+\frac{\lambda_2^2}{b^2}=1$

$(\lambda_1,\lambda_2)\begin{pmatrix}\frac{1}{a^2} & 0\\ 0 & \frac{1}{b^2}\end{pmatrix}\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}=1$



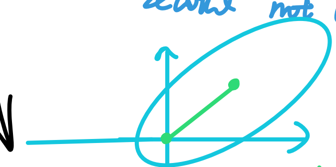
$(\lambda_1,\lambda_2)\begin{pmatrix}\sin\theta & \cos\theta\\ -\cos\theta & \sin\theta\end{pmatrix}\begin{pmatrix}\frac{1}{a^2} & 0\\ 0 & \frac{1}{b^2}\end{pmatrix}\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}$

become not "0"

The confidence region for the mean $\boldsymbol{\mu}$ of a p-dimensional normal population is

$$\left\{\boldsymbol{\mu} : n(\bar{\mathbf{x}}-\boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \le \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right\}$$

whatever the values of the unknown $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

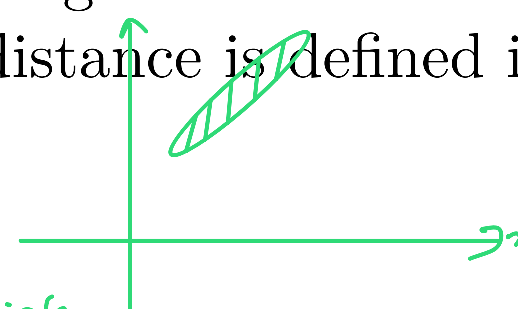


only centre changed

$(\lambda_1-\mu_1, \lambda_2-\mu_2)$

A geometric interpretation is that the plausible $\boldsymbol{\mu}$ (at $1-\alpha$ level) are those points within the ellipsoid centered at $\bar{\mathbf{x}}$, having a distance $\{(n-1)pF_{p,n-p}(\alpha)/[n(n-p)]\}^{1/2}$ to $\bar{\mathbf{x}}$, where the distance is defined in terms of $\mathbf{S}^{-1}$.

Strong positive corr ⇒



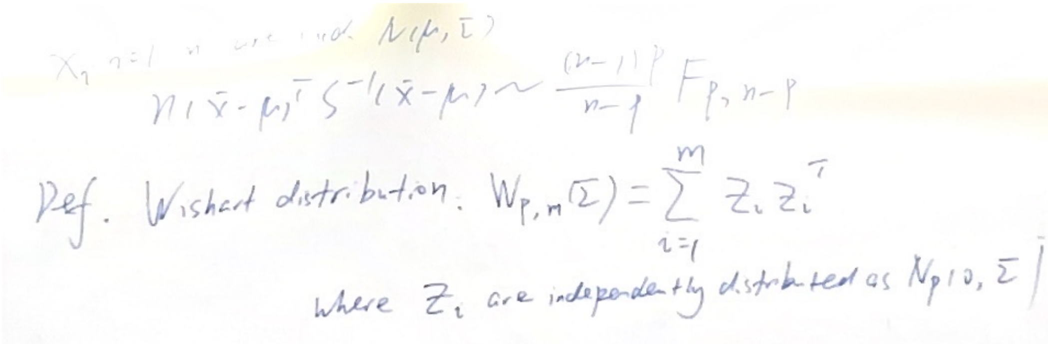
if $S = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$, then it is circle, which means $\frac{1}{a^2}=\frac{1}{b^2}$

$p=3$: ellipsoid: $\sum_{i=1}^{p} \frac{x_i^2}{a_i^2} = 1$

$$\Downarrow$$

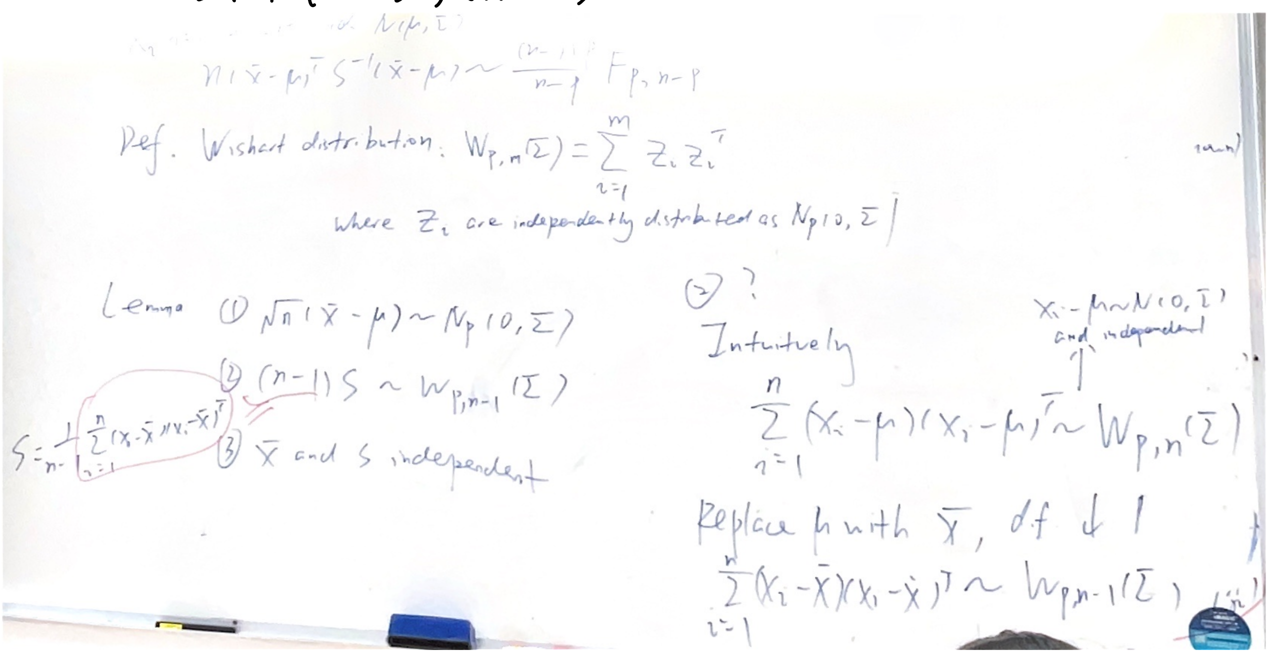
$$(x-\mu)^T \Sigma^{-1} (x-\mu) = x^T \overbrace{\begin{pmatrix} \frac{1}{a_1^2} & & 0 \\ & \frac{1}{a_2^2} & \\ 0 & & \frac{1}{a_3^2} \end{pmatrix}}^{\Sigma^{-1}} x$$

here $\mu = 0$



②: $S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$

$\Rightarrow (n-1)S = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$



$$\Downarrow$$

Distribution of $\bar{\mathbf{X}}$ and $\mathbf{S}$

dimension
$p=1$ $(\Sigma=1)$

Definition [Wishart distribution]

$$\begin{aligned} W_{p,m}(\boldsymbol{\Sigma}) &= \text{ Wishart distribution with } m \text{ d.f.} \\ &= \text{ distribution of } \sum_{j=1}^{m} \mathbf{Z}_j \mathbf{Z}_j^\top \end{aligned}$$

where $\mathbf{Z}_j$ are independently distributed as $N_p(\mathbf{0}, \boldsymbol{\Sigma})$.

Lemma [Distribution of sample mean and sample covariance]

Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are i.i.d. $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then ③ $\bar{\mathbf{x}}$ and $\mathbf{S}$ are independent, with

$$① \sqrt{n}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim N_p(0, \boldsymbol{\Sigma})$$ proved before

$$② \quad (n-1)\mathbf{S} \sim W_{p,n-1}(\boldsymbol{\Sigma}).$$

Distribution of Hotelling's T^2

An informal proof: We can write

$$T^2 = \underline{\sqrt{n}\left(\overline{\mathbf{x}} - \boldsymbol{\mu}_0\right)^\top} \left(\frac{\sum_{j=1}^{n} \left(\mathbf{x}_j - \overline{\mathbf{x}}\right)\left(\mathbf{x}_j - \overline{\mathbf{x}}\right)^\top}{n-1} \right)^{-1} \underline{\sqrt{n}\left(\overline{\mathbf{x}} - \boldsymbol{\mu}_0\right)}$$

$$\stackrel{d}{=} \underline{N_p(\mathbf{0}, \boldsymbol{\Sigma})^\top} \left[\frac{1}{n-1} W_{p,n-1}(\boldsymbol{\Sigma}) \right]^{-1} \underline{N_p(\mathbf{0}, \boldsymbol{\Sigma})} \qquad \sim \frac{(n-1)p}{n-p} F_{p,n-p}$$

We know the exact distribution of "Wishart term" and "Normal terms", and their independence. Through calculations, we can obtain

$$T^2 \sim \frac{(n-1)p}{(n-p)} F_{p,n-p}.$$

$$p = 1 \Rightarrow T^2 \stackrel{d}{=} t_{n-1}^2 = F_{1,n-1}$$

Alternatives to Confidence Regions

The confidence regions consider all the components of $\boldsymbol{\mu}$ jointly. We often desire a confidence statement (i.e, confidence interval) about individual components of $\boldsymbol{\mu}$ or several linear combinations of $\boldsymbol{\mu}$.

We want all such statements to hold simultaneously with some specified large probability; that is, want to make sure that the probability that any one of the confidence statements is incorrect is small.

Three ways of forming simultaneous confidence intervals considered:

- "one-at-a-time" intervals
- T^2 intervals
- Bonferroni

$\mu_1, \mu_2.$

$H_{01}: \mu_1 = 0, \; H_{02}: \mu_2 - \mu_3 = 0$

$H_{01}, H_{02}, \cdots, H_{0m}$

```
    SXy<-apply(sample(c(1,-1),n1,replace=TRUE,prob=c(0.5,0.5))*cbind(X1,y1),2,fhm)[which(rbinom(n1,1,gamma)!=0),]/sqrt(m)
    if(partial==0){g<-qr.solve(SXy[,1:p],SXy[,p+1])}
    else{SX<-SXy[,1:p];M<-solve(t(SX)%*%SX);g<-M%*%(t(X)%*%y)}
    return(list(r1=g,r2=sum(c*g),r3=SXy))
  }
```

“One-at-a-time” confidence interval

Interest in C.I.'s for individual components of $\boldsymbol{\mu}$ or linear combination $\mathbf{a}^\top \boldsymbol{\mu}$. Define $z = \mathbf{a}^\top \mathbf{x}$

$$z \sim N_1\left(\mathbf{a}^\top \boldsymbol{\mu}, \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}\right)$$

Note: $\mathbf{a}_1 = [0, 1, 0, \ldots, 0]$ will yield $\mathbf{a}_1^\top \boldsymbol{\mu} = \mu_2 = E[\bar{x}_2]$ and $\mathbf{a}_2 = [1, -1, 0, \ldots, 0]$ implies that $\mathbf{a}_2^\top \boldsymbol{\mu} = \mu_1 - \mu_2 = E[\mathbf{a}_2^\top \bar{\mathbf{x}}]$, etc.

$E[a^T \bar{x}] = E[a^T \mu]$

because $E[\bar{x}] = \mu$

“One-at-a-time” interval (t-interval)

A $100(1-\alpha)\%$ confidence interval for $\mathbf{a}^\top \boldsymbol{\mu}$ is

$a^T\bar{x} - a^T\mu \sim ?$

we know $\sqrt{n}(\bar{x} - \mu) \sim N(0, \Sigma)$, so

$\sqrt{n}(a^T\bar{x} - a^T\mu) \sim N(0, a^T\Sigma a)$

$\therefore a^T\bar{x} - a^T\mu \sim N(0, \frac{a^T\Sigma a}{n})$

$\Rightarrow \frac{\sqrt{n}(a^T\bar{x} - a^T\mu)}{\sqrt{a^T\Sigma a}} \sim N(0, 1)$, replace Σ with S

then $\frac{\sqrt{n}(a^T\bar{x} - a^T\mu)}{\sqrt{a^T S a}} \sim t_{n-1}$

$\therefore$ conclusion $\left[a^T\bar{x} - t_{n-1}(\frac{\alpha}{2})\sqrt{\frac{a^T S a}{n}}, a^T\bar{x} + t_{n-1}(\frac{\alpha}{2})\sqrt{\frac{a^T S a}{n}}\right]$

$$\mathbf{a}^\top \bar{\mathbf{x}} \pm t_{n-1}(\alpha/2)\sqrt{\frac{\mathbf{a}^\top \mathbf{S} \mathbf{a}}{n}}$$

limitation: consider coverage for a fixed $\vec{a}$ vector.

where $t_{n-1}(\alpha/2))$ is the upper $100(\alpha/2)$ percentile of Student's t-distribution with $df = n-1$. This can also be referred to as t-interval.

T^2 interval

expected longer than above

for $\forall \vec{a}$

$P(\text{for all } \vec{a}, a^T\mu \in [a^T\bar{x} - c\sqrt{\frac{a^TSa}{n}}, a^T\bar{x} + c\sqrt{\frac{a^TSa}{n}}]) = 1-\alpha$?

If we want to consider simultaneously all possible choices for the vector $\mathbf{a}$ such that the coverage rate over all of them is $100(1-\alpha)\%$;
Rewrite t-interval as

$$\left\{ n\frac{\left(\mathbf{a}^\top(\overline{\mathbf{x}}-\boldsymbol{\mu})\right)^2}{\mathbf{a}^\top\mathbf{S}\mathbf{a}} \le t^2_{n-1}(\alpha/2) \right\}$$

$n \times n$ $\quad n \times d$

Is there a bound which can replace $t^2_{n-1}(\alpha/2)$ and defines a C.R. that simultaneously contains $\mathbf{a}^\top\boldsymbol{\mu}$ for all $\mathbf{a}$? to
Preliminary result: For $\underset{p\times p}{\mathbf{B}}$ p.d. and $\mathbf{x} \neq \mathbf{0}$ $\quad d \in \mathbb{R}^p$

$$\max_{\mathbf{x}\neq 0} \frac{(\mathbf{x}^\top\mathbf{d})^2}{\mathbf{x}^\top\mathbf{B}\mathbf{x}} = \mathbf{d}^\top\mathbf{B}^{-1}\mathbf{d}$$

$\Uparrow$ substitute

with maximum attained when $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{d}$, $c \neq 0$

For any $u, v \in \mathbb{R}^p$

$(u^Tv)^2 \le (u^Tu)(v^Tv)$

$u^Tv = \langle u, v \rangle = \|u\|\cdot\|v\| \cos\theta$

$\cos\theta = 1$

when $\theta = 0$ or $\theta = 180$

$\therefore u = cv$

T^2 interval

$(x^\top d)^2 = (x^\top B^{\frac{1}{2}} B^{-\frac{1}{2}} d)^2$

$\le x^\top B x \, d^\top B^{-1} d$ "=" hold when

$\therefore \frac{(x^\top d)^2}{x^\top B x} \le d^\top B^{-1} d$ $B^{\frac{1}{2}} x = c B^{-\frac{1}{2}} d$

So, $\max_{\mathbf{a}\neq 0} \frac{\left(\mathbf{a}^\top(\overline{\mathbf{x}}-\boldsymbol{\mu})\right)^2}{\mathbf{a}^\top \mathbf{S}\mathbf{a}} = n(\overline{\mathbf{x}}-\boldsymbol{\mu})^\top \mathbf{S}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) = T^2$ with maximum at

$\therefore x = cB^{-1} d$

$$\mathbf{a} = c\mathbf{S}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}), \quad c \neq 0$$

$a^\top \bar{x} \pm c\sqrt{\frac{a^\top S a}{n}}$

T^2 interval

So we choose $c^2 = \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population with Σ positive definite. Then, simultaneously for all $\mathbf{a}$, the interval

$c = \sqrt{\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)}$

$$\left(\mathbf{a}^\top\overline{\mathbf{x}} - \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)\mathbf{a}^\top\mathbf{S}\mathbf{a}}, \quad \mathbf{a}^\top\overline{\mathbf{x}} + \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)\mathbf{a}^\top\mathbf{S}\mathbf{a}}\right)$$

Smallest length interval to achieve $1-\alpha$

will contain $\mathbf{a}^\top\boldsymbol{\mu}$ with probability $1-\alpha$.

as for any fixed a

$P(a^\top\mu \in [\ ]) \ge 1-\alpha$

These are called T^2 intervals because the coverage probability is determined by the distribution of T^2.

Simultaneously C.I. for several components

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ -x_{21} & -x_{22} \end{bmatrix}$$

Often, we focus on just a few confidence statements, like creating confidence intervals for different components of $\boldsymbol{\mu}$, while keeping an overall confidence level $1 - \alpha$.

It is possible to do better than the T^2 intervals as T^2 intervals can be too wide (less precise)

FWER

$\mu_1, \dots, \mu_k$

$I_1, \dots, I_k$

C.I.s

overall confidence is $1-\alpha$

FWER: Suppose $I_1, \dots, I_k$,

for each CI, confidence interval is $1-\alpha$.

Consider finding C.I.s for k components of $\boldsymbol{\mu}$. What is the overall error if we use confidence level $1-\alpha$ to control individual component?

- Familywise error rate (FWER)

I_j cover μ_j

$\Downarrow$

$$\begin{aligned} &\Pr\{ \text{ at least one C.I. is "wrong" } \} \\ &= 1 - \Pr\{ \text{ no C.I.'s are wrong } \} \quad p(\mu_i \in I_i) = 1-\alpha \\ &= 1 - (1-\alpha)^k \quad \text{assuming independence of C.I.'s} \\ &\geq \alpha \end{aligned}$$

Example: if $\alpha = .05$, FWER for $k = 10$ is $1 - (.95)^{10} \cong .40$; as k increase, this tends to 1.

- If the coverage rate is $100(1-\alpha)\%$ for one interval, then the overall familywise coverage rate could be much less than $100(1-\alpha)\%$.

If want overall control, each p_i of I_i should be large for success rate

If α very small, $(1-\alpha)^k \approx 1-k\alpha$

$1-(1-\alpha)^k \approx k\alpha \geq \alpha$

Bonferroni Method

逆向思维

Could be seen as successful rate

$[\quad]_{1-\alpha} \; [\quad]_{1-\alpha}$

$\Downarrow$

$P(\mu_i \in I_i, i=1,2) < 1-\alpha$

prefer this

$P(\mu_i \in \tilde{I}_i, i=1,2) \geq 1-\alpha$

$= 1 - P(\{\mu_1 \notin \tilde{I}_1\} \cup \{\mu_2 \notin \tilde{I}_2\})$

Suppose the j th confidence interval has coverage probability $1-\alpha_j$, i.e. $P\left(\mathbf{a}_j^\top \boldsymbol{\beta} \in I_j\right) = 1-\alpha_j$. Then

$P(A \cup B) \leq P(A) + P(B)$

$\geq 1 - \sum_{i=1}^{2} P(\mu_i \notin \tilde{I}_i)$

$$P\left(\mathbf{a}_j^\top \boldsymbol{\beta} \in I_j, \forall j \in \{1, \ldots, k\}\right) = 1 - P\left(\mathbf{a}_j^\top \boldsymbol{\beta} \notin I_j \text{ for some } j\right)$$

$= 1 - \underbrace{(\alpha_1 + \alpha_2)}_{\text{not success rate}} = 1-\alpha$

$$\geq 1 - \sum_{j=1}^{k} P\left(\mathbf{a}_j^\top \boldsymbol{\beta} \notin I_j\right) = 1 - \sum_{j=1}^{k} \alpha_j.$$

The Bonferroni method proceeds by setting the α_j so that $\sum_{j=1}^{k} \alpha_j = \alpha$, (usually $\alpha_j = \alpha/k$). Then

$$P(\text{all confidence intervals contain their true values})$$

$$= P\left(\mathbf{a}_j^\top \boldsymbol{\beta} \in I_j, \forall j \in \{1, \ldots, k\}\right) \geq 1 - \alpha.$$

Bonferroni method

The main features of the Bonferroni method are that it is simple to use, and that it is conservative (i.e. the inequality actual coverage probability is greater than claimed and the confidence intervals are wider than they have to be). But it is less conservative than T^2 interval.

Bonferroni intervals for components of $\boldsymbol{\mu}$

$I_1 \quad 1-\alpha_1$
$I_2 \quad 1-\alpha_2$
$\vdots \quad \vdots$
$I_m \quad 1-\alpha_m$
$\sum \alpha_i = \alpha$

Let us develop Bonferroni intervals for the components μ_i of $\boldsymbol{\mu}$. Lacking information on the relative importance of these components, we consider the individual t-intervals

$[\bar{X}_i - t_{n-1}(\frac{\alpha_i}{2})\sqrt{\frac{s_{11}}{n}}, \bar{X}_i + t_{n-1}(\frac{\alpha_i}{2})\sqrt{\frac{s_{ii}}{n}}]$

usually, take $\alpha_i = \frac{\alpha}{m}$

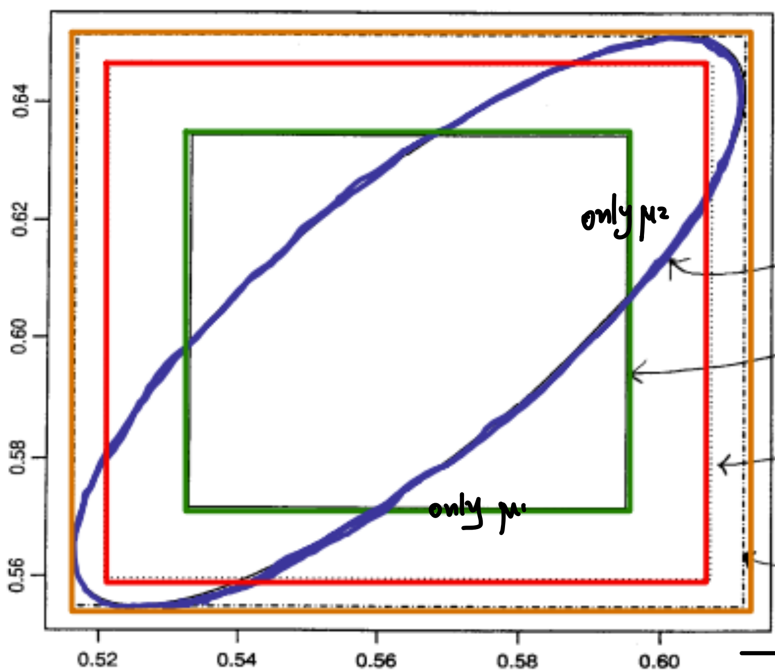
$$\bar{x}_i \pm t_{n-1}\left(\frac{\alpha_i}{2}\right)\sqrt{\frac{s_{ii}}{n}} \quad i = 1, 2, \ldots, m$$

with $\alpha_i = \alpha/m$. Since $P\left[\bar{X}_i \pm t_{n-1}(\alpha/2m)\sqrt{s_{ii}/n}\right.$ contains $\left.\mu_i\right] = 1 - \alpha/m$, $i = 1, 2, \ldots, m$, we have, from,

$$P\left[\bar{X}_i \pm t_{n-1}\left(\frac{\alpha}{2m}\right)\sqrt{\frac{s_{ii}}{n}} \text{ contains } \mu_\mathrm{i}, \text{ all i}\right] \geq 1 - \underbrace{\left(\frac{\alpha}{m} + \frac{\alpha}{m} + \cdots + \frac{\alpha}{m}\right)}_{m \text{ terms}}$$

$$= 1 - \alpha$$

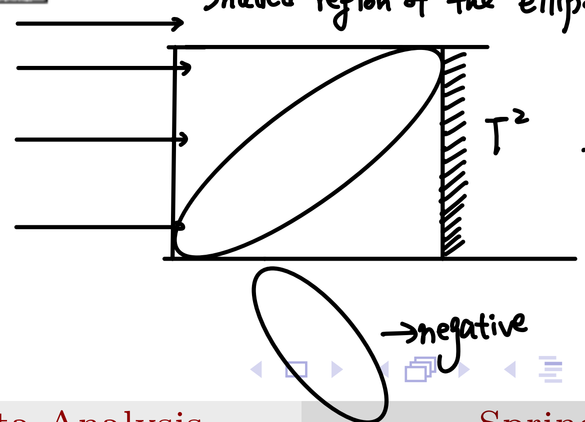
An illustration of several types of confidence intervals



$\boxed{\text{ex}}$ $p=2$

- Bivariate confidence region — H_0... T^2 ellipse construction
- t intervals
- Bonferroni intervals
- T^2 intervals

Shadow of ellipse
Shaded region of the ellipse



Multiple hypothesis testing

We next look at multiple hypothesis testing problems. In many studies several hypothesis will be tested, we may be testing a drug on several different cell lines, or measuring a response at several different time points after a application of treatment.

The challenge is controlling the Familtywise Type-I error in multiple hypothesis testing. This is actually the same to the problem arising in constructing several confidence intervals simultaneously.

Multiple hypothesis testing

H_0 H_1

Type I error α

P(reject H_0 | H_0 is true) $= \alpha$ control Type-I error

(0.05) usually α is small

H_0: no effect

H_1: effect

Multiple hypothesis testing arises when a statistical analysis involves multiple simultaneous statistical tests: $H_{01}, \cdots H_{0k}$.

If use α for each individual test, then Familywise Type I error rate (FWER)

$\{H_{01}, H_{02}, \ldots, H_{0k}\}$ as a whole

use $\frac{\alpha}{k}$

$$
\begin{aligned}
&\Pr\{\text{at least reject one } H_{0j}\} \\
&= 1 - \Pr\{\text{ fail to reject any } H_{0j}\} \\
&= 1 - (1-\alpha)^k \quad \text{assuming independence of } H_{0j}\text{'s}
\end{aligned}
$$

Example: if $\alpha = .05$, FWER for $k = 10$ is $1 - (.95)^{10} \cong .40$; as k increase, this tends to 1.

Indicator I

$[-1 \quad 0 \quad +1]$

The Bonferroni correction sets the significance cut-off at α/k, or $\alpha_1, \cdots, \alpha_k$ such that $\sum_{i=1}^{k} \alpha_i = \alpha$.

Indicator i

$[-2 \quad 0 \quad 2]$

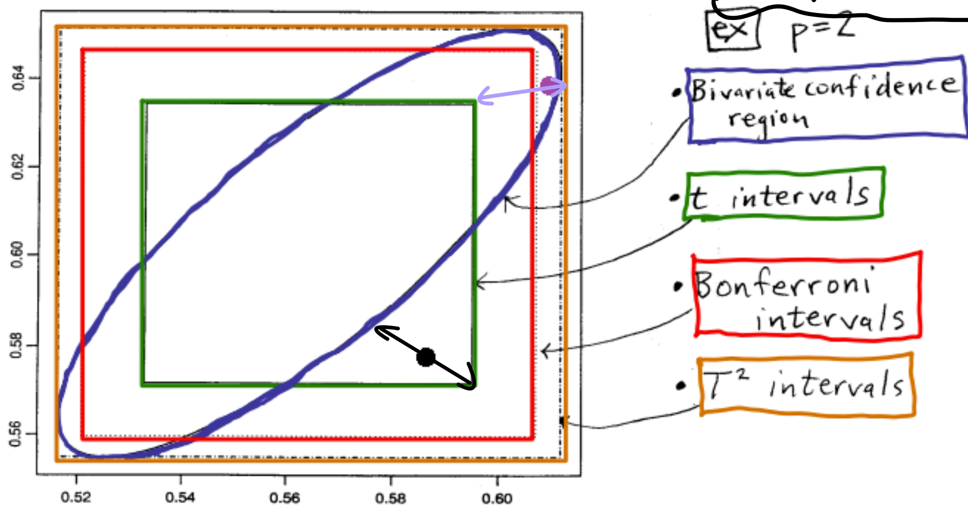
increase the [ ] for each i since when $i = 1, 2, \ldots, k$, there is sampling error.

Hypothesis testing of $H_0 : \mu = \mu_0$ may lead to some seemingly inconsistent results. For example,

- The multivariate tests may reject H_0, but the component means are within their respective confidence intervals for them (e.g., black dot).
- Separate t-tests for component means may be rejected, but you do not reject for multivariate (e.g., purple dot).

$$\begin{cases} H_0: \mu_1 = 0 & \text{rej} \\ H_0: \mu_2 = 0 & \text{rej} \\ H_0: \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \text{do not reject} \end{cases}$$

$\boxed{\text{ex}}$ $p=2$



Hotelling's T^2: $X_1, \cdots, X_m \sim$ i.i.d.

The confidence region is the only one that takes into consideration of the covariances.

1: $H_0: \mu_1 = \mu_2 = \cdots = \mu_m = 0$ $\quad \mu = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

2: or $H_{0i}: \mu_i = 0, i = 1, \cdots, m$

Bonferrnoi method is conservative

Assume i.i.d. and dimension not large, use I.

If no specific structure, $X_1, \dots, X_m$, not easy to use T^2, use Bonferrnoi method.

Use Bonferrnoi will be more flexible. (more daily scenario)

Similar as both don't consider covariance

important use $\boxed{S \to \Sigma}$

Assume $X_1, \dots, X_m \in \mathbb{R}^p$ some Σ_1, some Σ_2.

$X_1, \dots, X_{n_1} \in \mathbb{R}^p \to N(\mu_1, \Sigma_1)$

$X_{n_1+1}, \dots, X_m \to N(\mu_2, \Sigma_2)$

$S \overset{\times}{\to} \Sigma$ is not an estimate of Σ

The Bonferroni method tends to be a bit too conservative, in other words, it tends to lead to a high rate of false negatives.

This can be understood from the property that the Bonferroni intervals are wider than they have to be, thus it tends to fail to reject the null even if they are wrong. we only controlled Type-I error, but not control Type-II error.

Some advanced method: Benjamini–Hochberg procedure (beyond the scope of this course)

Large sample theory

$\sqrt{n}(\bar{x}-\mu) \Rightarrow N(0,\Sigma)$ as $n \to \infty$

$\boxed{\begin{array}{l}\text{Recall than } \sqrt{n}(\bar{x}-\mu) \sim N(0,\Sigma) \text{ if } X_i \sim MVN \\ n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu) \sim \chi_p^2\end{array}}$

$n(\bar{x}-\mu)^T\Sigma^{-1}(\bar{x}-\mu) \Rightarrow \chi_p^2$ as $n \to \infty$

$F_n(x) \to F(x)$ for any $x \in \mathbb{R}$

cdf of F_n cdf of F

$Z_n \to Z$

For any subset S : $\lim_{n\to\infty} P(Z_n \in S) \to P(Z \in S)$

Large sample theory

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ be independent observations from a population with mean $\boldsymbol{\mu}$ and finite (nonsingular) covariance $\boldsymbol{\Sigma}$. Then $\sqrt{n}(\bar{\mathbf{x}} - \boldsymbol{\mu})$ is approximately $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ and $n(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$ is approximately χ_p^2 as $n \to \infty$.

The first statement is the Central Limit Therorem (CLT) mentioned in Chapter 3.

$\lim_{n\to\infty} P(\theta \in R(x)) \approx 1-\alpha$

Large sample inference of mean vector

close to I if n is large

$n(\bar{x}-\mu)^T S^{-1}(\bar{x}-\mu) \sim \frac{(n-1)p}{n-p} F_{p,n-p} \sim \frac{(n-1)p}{n-p} \cdot \frac{\chi_p^2 / p}{\chi_{n-p}^2/(n-p)}$

law of large number

$\frac{\chi^2_{n-p}}{n-p} \to I$ as $n \to \infty$

$\frac{\sum_{i=1}^{n-p} z_i^2}{n-p} \to E[z_i^2] \to I$

When the sample size is large, tests of hypotheses and confidence regions for $\boldsymbol{\mu}$ can be constructed without the assumption of a normal population.

All large-sample inferences about $\boldsymbol{\mu}$ are based on a χ^2-distribution. We know that $(\bar{\mathbf{x}}-\boldsymbol{\mu})^\top \left(n^{-1}\mathbf{S}\right)^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) = n(\bar{\mathbf{x}}-\boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})$ is approximately χ^2 with p d.f., and thus,

C.I. using this way

$$P\left[n(\bar{\mathbf{x}}-\boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \leq \chi^2_p(\alpha)\right] \to 1-\alpha$$

as $n \to \infty$ where $\chi^2_p(\alpha)$ is the upper α-th percentile of the χ^2_p-distribution.

This immediately leads to large sample tests of hypotheses and confidence regions.

Summary

Summary

- Hotelling T^2
- Likelihood ratio test
- Confidence region
- One-at-a-time interval (t-interval)
- T^2 interval
- Bonferroni interval
- Comparison of several intervals
- Large sample inference

End