Multivariate Analysis
Chapter 5: One Sample Inference

Shaobo Jin

Department of Mathematics

Intended Learning Outcome

Through this chapter, you should be able to

1. test multivariate normal mean,
2. construct confidence region and simultaneous confidence intervals for normal means.

Test Univariate Normal Mean: t Test

We want to test

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0$$

If X_1, X_2, ..., X_n denote a random sample from a normal population, then the test statistic is

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

where

$$S^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2.$$

The statistic t follows a t distribution with $n-1$ degrees of freedom if H_0 is true. We reject H_0 if $|t|$ is too large.

Confidence Interval of t Test

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0$$

If the significance level is α, we reject H_0 if the observed t statistic satisfies

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

The $1 - \alpha$ confidence interval for μ is

$$\bar{X} - t_{n-1}\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1}\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{n}}.$$

The realized confidence interval

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

collects all values μ that would not be rejected by the level α test.

Univariate to Multivariate Test

- The t test is equivalent to rejecting H_0 if

$$t^2 = n\left(\bar{X} - \mu_0\right) S^{-2} \left(\bar{X} - \mu_0\right)$$

  is too large.

- For testing

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0,$$

  a generalization is to reject H_0 if

$$T^2 = n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)^T \boldsymbol{S}^{-1} \left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right),$$

  is too large, where

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

Structure of T^2

The univariate t test is equivalent to

$$\begin{aligned} t^2 &= n\left(\bar{X}-\mu_0\right) S^{-2}\left(\bar{X}-\mu_0\right) \\ &= \underbrace{\sqrt{n}\left(\bar{X}-\mu_0\right)}_{\text{Normal}} \left[\frac{1}{n-1} \underbrace{(n-1) S^2}_{\text{Scaled chi-square}}\right]^{-1} \underbrace{\sqrt{n}\left(\bar{X}-\mu_0\right)}_{\text{Normal}}. \end{aligned}$$

The T^2 is equivalent to

$$\begin{aligned} T^2 &= n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)^T \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right) \\ &= \underbrace{\sqrt{n}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)^T}_{\text{Normal}} \left[\frac{1}{n-1} \underbrace{(n-1)\boldsymbol{S}}_{\text{Wishart}}\right]^{-1} \underbrace{\sqrt{n}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)}_{\text{Normal}} \end{aligned}$$

Hotelling's T^2 Distribution

Definition: Hotelling's T^2 Distribution

Suppose that $\boldsymbol{X} \sim N_p\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$ and $\boldsymbol{M} \sim W_p\left(\boldsymbol{\Sigma}, n\right)$. Then

$$T^2 = n\boldsymbol{X}^T\boldsymbol{M}^{-1}\boldsymbol{X} \quad \sim \quad T^2\left(p, n\right),$$

a Hotelling's T^2 distribution.

Proposition: Hotelling's T^2 Distribution And F Distribution

If $T^2 \sim T^2\left(p, n-1\right)$, then

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

where $F_{p,n-p}$ denotes a random variable with an F distribution with p and $n-p$ degrees of freedom.

Hotelling's T^2

The statistic

$$T^2 = n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)^T \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)$$

is called Hotelling's T^2.

Hotelling's T^2

Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be a random sample from an $N_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ population. Then

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

under H_0.

Invariance Property

Hotelling's T^2 is invariant under affine transformation

$$\boldsymbol{Y}_{p\times 1} = \boldsymbol{C}_{p\times p}\boldsymbol{X}_{p\times 1} + \boldsymbol{d},$$

where $\boldsymbol{C}$ is nonsingular. Hence, we can change the scale of data.

- It follows from Result 2.6 that

$$\begin{aligned} \bar{\boldsymbol{Y}} &= \boldsymbol{C}\bar{\boldsymbol{X}} + \boldsymbol{d}, \\ \boldsymbol{S}_y &= \boldsymbol{C}\boldsymbol{S}\boldsymbol{C}^T. \end{aligned}$$

- Since

$$\boldsymbol{\mu}_Y = \boldsymbol{C}\boldsymbol{\mu} + \boldsymbol{d},$$

  then

$$n\left(\bar{\boldsymbol{Y}} - \boldsymbol{\mu}_{Y,0}\right)^T \boldsymbol{S}_y^{-1}\left(\bar{\boldsymbol{Y}} - \boldsymbol{\mu}_{Y,0}\right) = n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)^T \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right).$$

General Likelihood Ratio Test

Suppose that $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ form a random sample from $\boldsymbol{X} \sim f(\boldsymbol{x}; \boldsymbol{\theta})$ with parameter vector $\boldsymbol{\theta}$. We want to test

$$H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0.$$

- The likelihood is

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{n} f(\boldsymbol{x}_j; \boldsymbol{\theta}).$$

- A likelihood ratio test (LRT) rejects H_0 if

$$\Lambda = \frac{\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} L(\boldsymbol{\theta})}{\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta})} < c$$

  for some suitably chosen constant c. The choice of c often depends on the distribution of Λ.

Distribution of LRT Statistic

The exact distribution of Λ is typically unknown.

Result 5.2

Under some regularity assumptions, $-2\log\Lambda$ converges in distribution to $\chi^2_{v-v_0}$ under H_0. Here v is dimension of the parameter space $\boldsymbol{\Theta}$, and v_0 is the dimension of $\boldsymbol{\Theta}_0$.
The asymptotic size α LRT rejects H_0 if $-2\log\lambda\left(\boldsymbol{x}\right) \geq \chi^2_{v-v_0}\left(\alpha\right)$.

Likelihood for Testing Normal Mean

Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population.

1. The likelihood under H_0 is

$$\begin{aligned} L(\boldsymbol{\Sigma}) &= \exp\left\{-\frac{np}{2}\log(2\pi) - \frac{n}{2}\log\det(\boldsymbol{\Sigma}) \right. \\ &\qquad \left. -\frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{X}_j - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_j - \boldsymbol{\mu}_0)\right\}. \end{aligned}$$

2. The likelihood under $H_0 \cup H_1$ is

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \exp\left\{-\frac{np}{2}\log(2\pi) - \frac{n}{2}\log\det(\boldsymbol{\Sigma}) \right. \\ &\qquad \left. -\frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{X}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_j - \boldsymbol{\mu})\right\}. \end{aligned}$$

MLE

1. The log-likelihood under H_0 is

$$\text{const} - \frac{n}{2}\log\det(\boldsymbol{\Sigma}) - \frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{X}_j - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_j - \boldsymbol{\mu}_0).$$

The MLE is

$$\frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{X}_j - \boldsymbol{\mu}_0)(\boldsymbol{X}_j - \boldsymbol{\mu}_0)^T = \arg\max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}).$$

2. The log-likelihood under $H_0 \cup H_1$ is

$$\text{const} - \frac{n}{2}\log\det(\boldsymbol{\Sigma}) - \frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{X}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_j - \boldsymbol{\mu}).$$

The MLE is

$$\left(\bar{\boldsymbol{X}},\ \boldsymbol{S}_n\right) = \arg\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}).$$

LRT for Normal Mean

The likelihood ratio is

$$\begin{aligned}\Lambda &= \frac{\max\limits_{\boldsymbol{\Sigma}} L\left(\boldsymbol{\Sigma}\right)}{\max\limits_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)} \\ &= \left[\frac{\det\left(\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^T\right)}{\det\left(\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\boldsymbol{\mu}_0\right)\left(\boldsymbol{X}_j-\boldsymbol{\mu}_0\right)^T\right)}\right]^{n/2}.\end{aligned}$$

We reject $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ if Λ is too small.

The statistic $\Lambda^{2/n}$ is called Wilks' lambda.

Wilks' Lambda and Hotelling's T^2

Result 5.1

Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Then the test that rejects $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ if

$$T^2 \quad > \quad \frac{(n-1)\,p}{n-p} F_{p,n-p}(\alpha)$$

is equivalent to the likelihood ratio test of testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_1$ because of

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

(monotonic decreasing).

Confidence Region

Let $\boldsymbol{X}$ be a data matrix and $\boldsymbol{\theta}$ be a vector of population parameters. The region $R(\boldsymbol{X})$ is said to be a $100(1-\alpha)\%$ confidence region if

$$P\{R(\boldsymbol{X}) \text{ covers the true } \boldsymbol{\theta}\} = 1-\alpha.$$

The probability is evaluated under the true value of $\boldsymbol{\theta}$.

- Confidence interval is a special case of confidence region.
- Consider testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. If $\boldsymbol{\theta}_0 \in R(\boldsymbol{X})$, then we cannot reject H_0.

Confidence Region for Normal Mean

Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Then

$$n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right)^T \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right) \sim \frac{(n-1)\,p}{n-p} F_{p,n-p}$$

and

$$P\left\{n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right)^T \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right) \leq \frac{(n-1)\,p}{n-p} F_{p,n-p}(\alpha)\right\} = 1 - \alpha.$$

A confidence region can be

$$R(\boldsymbol{X}) = \left\{\boldsymbol{\mu};\; n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right)^T \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right) \leq \frac{(n-1)\,p}{n-p} F_{p,n-p}(\alpha)\right\}.$$

Confidence Region As Random Quantity

The confidence region is a random quantity. The probability

$$P\left\{R\left(\boldsymbol{X}\right) \text{ covers the true } \boldsymbol{\theta}\right\} = 1-\alpha$$

is evaluated for the random matrix $\boldsymbol{X}$. If we plug in our data, the realized confidence region has no uncertainty.

- For normal mean, the random confidence region is

$$\left\{\boldsymbol{\mu};\ n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^T \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right) \leq \frac{\left(n-1\right)p}{n-p} F_{p,n-p}\left(\alpha\right)\right\}.$$

- Once we plug in our data, the realized confidence region

$$\left\{\boldsymbol{\mu};\ n\left(\bar{\boldsymbol{x}}-\boldsymbol{\mu}\right)^T \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{x}}-\boldsymbol{\mu}\right) \leq \frac{\left(n-1\right)p}{n-p} F_{p,n-p}\left(\alpha\right)\right\}$$

  defines an ellipsoid with the center $\bar{\boldsymbol{x}}$.

Confidence for Components or Functions

The confidence region

$$R(\boldsymbol{X}) = \left\{\boldsymbol{\mu};\ n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^T \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right) \leq \frac{(n-1)\,p}{n-p} F_{p,n-p}(\alpha)\right\}$$

describes the whole vector $\boldsymbol{\mu}$ (joint knowledge). But we often need to make statements about each component (μ_1 or μ_2) or functions of components (e.g., $\mu_1 - \mu_2$).

Confidence Intervals for Linear Combinations of Normal Mean

Let $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. By Result 4.2, for a fixed vector $\boldsymbol{a}$,

$$Z = \boldsymbol{a}^T \boldsymbol{X} \sim N\left(\boldsymbol{a}^T \boldsymbol{\mu}, \boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a}\right).$$

If we have a random sample $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$, then the sample mean and sample variance of $Z_j = \boldsymbol{a}^T \boldsymbol{X}_j$ are

$$\bar{Z} = \boldsymbol{a}^T \bar{\boldsymbol{X}} \quad \text{and} \quad S_Z^2 = \boldsymbol{a}^T \boldsymbol{S} \boldsymbol{a},$$

where $\boldsymbol{S}$ is the sample covariance matrix of $\boldsymbol{X}_j$'s.

A $1-\alpha$ confidence interval for $\mu_Z = \boldsymbol{a}^T \boldsymbol{\mu}$ is

$$\boldsymbol{a}^T \bar{\boldsymbol{X}} - t_{n-1}\left(\frac{\alpha}{2}\right) \frac{\sqrt{\boldsymbol{a}^T \boldsymbol{S} \boldsymbol{a}}}{\sqrt{n}} \leq \mu_Z \leq \boldsymbol{a}^T \bar{\boldsymbol{X}} + t_{n-1}\left(\frac{\alpha}{2}\right) \frac{\sqrt{\boldsymbol{a}^T \boldsymbol{S} \boldsymbol{a}}}{\sqrt{n}}.$$

Example

Confidence Interval

Consider

$$\boldsymbol{a}^T\bar{\boldsymbol{X}} - t_{n-1}\left(\frac{\alpha}{2}\right)\frac{\sqrt{\boldsymbol{a}^T\boldsymbol{S}\boldsymbol{a}}}{\sqrt{n}} \leq \mu_Z \leq \boldsymbol{a}^T\bar{\boldsymbol{X}} + t_{n-1}\left(\frac{\alpha}{2}\right)\frac{\sqrt{\boldsymbol{a}^T\boldsymbol{S}\boldsymbol{a}}}{\sqrt{n}}.$$

- Find a $1-\alpha$ confidence interval for μ_1.
- Find a $1-\alpha$ confidence interval for $\mu_1-\mu_2$.

Individual Confidence Statements

- For different statements about $\boldsymbol{\mu}$, we typically choose different $\boldsymbol{a}$ based on the

$$\boldsymbol{a}^T \bar{\boldsymbol{X}} - t_{n-1}\left(\frac{\alpha}{2}\right)\frac{\sqrt{\boldsymbol{a}^T \boldsymbol{S}\boldsymbol{a}}}{\sqrt{n}} \leq \mu_Z \leq \boldsymbol{a}^T \bar{\boldsymbol{X}} + t_{n-1}\left(\frac{\alpha}{2}\right)\frac{\sqrt{\boldsymbol{a}^T \boldsymbol{S}\boldsymbol{a}}}{\sqrt{n}}.$$

- Each statement (with its own $\boldsymbol{a}$) is associated with confidence coefficient $1-\alpha$.
- However, when we put several statements together, the associated confidence coefficient is typically less than $1-\alpha$.

Simultaneous Confidence Statements

It is better to have a common $\boldsymbol{a}$ such that there is a probability of $1-\alpha$ that all confidence intervals for $\boldsymbol{a}^T\boldsymbol{\mu}$ obtained by varying $\boldsymbol{a}$ will be true. The statements hold jointly for all $\boldsymbol{a}$.

- The t confidence interval is equivalent to

$$\frac{n\left(\boldsymbol{a}^T\boldsymbol{\mu}-\boldsymbol{a}^T\bar{\boldsymbol{X}}\right)^2}{\boldsymbol{a}^T\boldsymbol{S}\boldsymbol{a}} \leq t_{n-1}^2\left(\frac{\alpha}{2}\right).$$

- We want to find a value c^2 such that

$$\frac{n\left(\boldsymbol{a}^T\boldsymbol{\mu}-\boldsymbol{a}^T\bar{\boldsymbol{X}}\right)^2}{\boldsymbol{a}^T\boldsymbol{S}\boldsymbol{a}} \leq c^2$$

for all choices of $\boldsymbol{a}$, such as

$$\max_{\boldsymbol{a}}\frac{n\left(\boldsymbol{a}^T\boldsymbol{\mu}-\boldsymbol{a}^T\bar{\boldsymbol{X}}\right)^2}{\boldsymbol{a}^T\boldsymbol{S}\boldsymbol{a}} \leq c^2.$$

Simultaneous Confidence Statements

Lemma (External)

Let $\boldsymbol{A}$ and $\boldsymbol{B} > 0$ be two symmetric matrices. The maximum value of

$$\frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{B} \boldsymbol{x}}, \quad \boldsymbol{x} \neq \boldsymbol{0},$$

is attained when $\boldsymbol{x}$ is the eigenvector of $\boldsymbol{B}^{-1}\boldsymbol{A}$ corresponding to the largest eigenvalue of $\boldsymbol{B}^{-1}\boldsymbol{A}$. Its maximum value is the largest eigenvalue of $\boldsymbol{B}^{-1}\boldsymbol{A}$.

By the lemma,

$$\max_a \frac{n\left[\boldsymbol{a}^T\left(\bar{\boldsymbol{x}}-\boldsymbol{\mu}\right)\right]^2}{\boldsymbol{a}^T \boldsymbol{S} \boldsymbol{a}} = n\left(\bar{\boldsymbol{x}}-\boldsymbol{\mu}\right) \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{x}}-\boldsymbol{\mu}\right) = T^2,$$

where $\boldsymbol{a}$ is proportional to $\boldsymbol{S}^{-1}\left(\bar{\boldsymbol{x}}-\boldsymbol{\mu}\right)$.

T^2 Intervals

Result 5.3

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

$$\boldsymbol{a}^T\bar{\boldsymbol{X}} \pm \sqrt{\frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)}\frac{\sqrt{\boldsymbol{a}^T\boldsymbol{S}\boldsymbol{a}}}{\sqrt{n}}$$

will cover $\boldsymbol{a}^T\boldsymbol{\mu}$ with probability $1-\alpha$. The interval is called T^2 interval.

Comparing to

$$\boldsymbol{a}^T\bar{\boldsymbol{X}} - t_{n-1}\left(\frac{\alpha}{2}\right)\frac{\sqrt{\boldsymbol{a}^T\boldsymbol{S}\boldsymbol{a}}}{\sqrt{n}} \leq \mu_Z \leq \boldsymbol{a}^T\bar{\boldsymbol{X}} + t_{n-1}\left(\frac{\alpha}{2}\right)\frac{\sqrt{\boldsymbol{a}^T\boldsymbol{S}\boldsymbol{a}}}{\sqrt{n}},$$

the major difference is the critical value.

Connection

$$t_{n-1}\left(\frac{\alpha}{2}\right) \quad \text{versus} \quad \sqrt{\frac{p\left(n-1\right)}{n-p}F_{p,n-p}\left(\alpha\right)}$$

- The T^2 interval is generally longer.
- A connection between the t distribution and the F distribution is that, if $X \sim t\left(m\right)$, then $X^2 \sim F\left(1, m\right)$. If $p = 1$, then

$$\sqrt{\frac{p\left(n-1\right)}{n-p}F_{p,n-p}\left(\alpha\right)} = t_{n-1}\left(\frac{\alpha}{2}\right).$$

Joint Statements for Marginal Mean

Using Result 5.3, we can say that

$$\bar{X}_1 - \sqrt{\frac{p\left(n-1\right)}{n-p}F_{p,n-p}\left(\alpha\right)}\frac{\sqrt{S_{11}}}{\sqrt{n}} \leq \mu_1 \leq \bar{X}_1 + \sqrt{\frac{p\left(n-1\right)}{n-p}F_{p,n-p}\left(\alpha\right)}\frac{\sqrt{S_{11}}}{\sqrt{n}}$$

$$\bar{X}_2 - \sqrt{\frac{p\left(n-1\right)}{n-p}F_{p,n-p}\left(\alpha\right)}\frac{\sqrt{S_{22}}}{\sqrt{n}} \leq \mu_2 \leq \bar{X}_2 + \sqrt{\frac{p\left(n-1\right)}{n-p}F_{p,n-p}\left(\alpha\right)}\frac{\sqrt{S_{22}}}{\sqrt{n}}$$

$$\vdots$$

$$\bar{X}_p - \sqrt{\frac{p\left(n-1\right)}{n-p}F_{p,n-p}\left(\alpha\right)}\frac{\sqrt{S_{pp}}}{\sqrt{n}} \leq \mu_p \leq \bar{X}_p + \sqrt{\frac{p\left(n-1\right)}{n-p}F_{p,n-p}\left(\alpha\right)}\frac{\sqrt{S_{pp}}}{\sqrt{n}}$$

hold simultaneously with confidence coefficient $1-\alpha$. These intervals are the shadows of the confidence ellipsoid on each component axis.

Bonferroni Correction

Suppose that we want to consider m components of $\boldsymbol{\mu}$. Let C_i be the event that

$$\bar{X}_i - t_{n-1}\left(\frac{\alpha_i}{2}\right)\frac{\sqrt{S_{ii}}}{\sqrt{n}} \leq \mu_i \leq \bar{x}_1 + t_{n-1}\left(\frac{\alpha_i}{2}\right)\frac{\sqrt{S_{ii}}}{\sqrt{n}}.$$

Then,

$$\begin{aligned} P\left(C_i \text{ holds for all } i\right) &= 1 - P\left(C_i \text{ does not hold for some } i\right) \\ &\geq 1 - \sum_{i=1}^{m} P\left(C_i \text{ does not hold}\right) \\ &= 1 - \sum_{i=1}^{m} \alpha_i. \end{aligned}$$

In order to let $1-\alpha$ be a lower bound, we can choose new significance levels α_i such that

$$\alpha = 1 - \sum_{i=1}^{m} \alpha_i.$$

Bonferroni Confidence Interval

By the Bonferroni correction,

$$
\begin{gathered}
\bar{X}_1 - t_{n-1}\left(\frac{\alpha}{2m}\right)\frac{\sqrt{S_{11}}}{\sqrt{n}} \leq \mu_1 \leq \bar{X}_1 + t_{n-1}\left(\frac{\alpha}{2m}\right)\frac{\sqrt{S_{11}}}{\sqrt{n}} \\
\bar{X}_2 - t_{n-1}\left(\frac{\alpha}{2m}\right)\frac{\sqrt{S_{22}}}{\sqrt{n}} \leq \mu_2 \leq \bar{X}_2 + t_{n-1}\left(\frac{\alpha}{2m}\right)\frac{\sqrt{S_{22}}}{\sqrt{n}} \\
\vdots \\
\bar{X}_m - t_{n-1}\left(\frac{\alpha}{2m}\right)\frac{\sqrt{S_{mm}}}{\sqrt{n}} \leq \mu_m \leq \bar{X}_m + t_{n-1}\left(\frac{\alpha}{2m}\right)\frac{\sqrt{S_{mm}}}{\sqrt{n}}
\end{gathered}
$$

hold simultaneously with confidence coefficient $1-\alpha$.

Normality Assumption

- The normal assumption is needed for the above statistical tests, confidence regions, and confidence intervals.
- When the sample size is large, inference can often be made without the assumption of a normal population.
  - For a sufficiently large n,

$$\sqrt{n}\boldsymbol{S}^{-1/2}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right) \text{ is approximately } N_p\left(\mathbf{0},\boldsymbol{I}\right),$$
$$n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^T\boldsymbol{S}^{-1}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right) \text{ is approximately } \chi^2_p,$$

    even if $\boldsymbol{X}$ is nonnormal.

Large n

Result 5.4, Revised

Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be a random sample from a population with mean $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. Consider testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. The test statistic $T^2 = n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)^T \boldsymbol{S}^{-1}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)$ converges in distribution to χ_p^2 under H_0. The significance level is approximately α if we reject $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ when

$$n\left(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0\right)\boldsymbol{S}^{-1}\left(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0\right) \quad > \quad \chi_p^2\left(\alpha\right).$$

With the normality assumption, we reject H_0 when

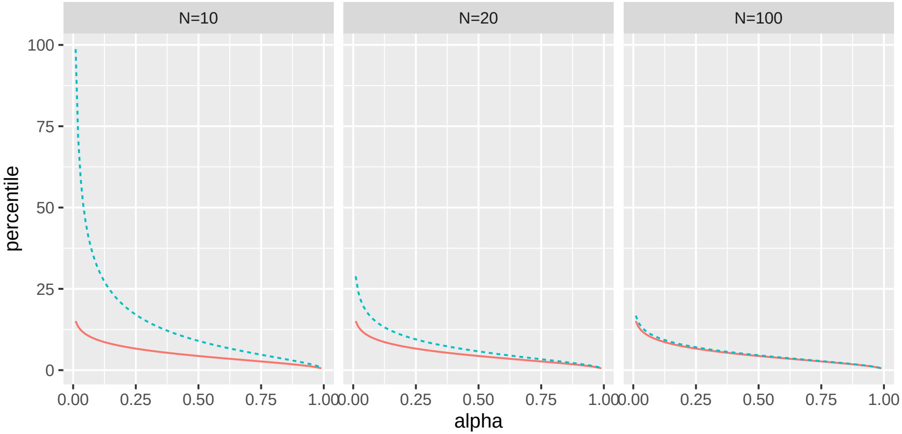
$$n\left(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0\right)\boldsymbol{S}^{-1}\left(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0\right) \quad > \quad \frac{\left(n-1\right)p}{n-p}F_{p,n-p}\left(\alpha\right).$$

Same test statistic, but different critical values (F versus χ^2)!

F versus χ^2

As $n \to \infty$,

$$\frac{(n-1)\,p}{n-p} F_{p,n-p}(\alpha) \quad \to \quad \chi^2_p(\alpha)$$



Large Sample Inference

Result 5.5

Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., $\boldsymbol{X}_n$ be a random sample from a population with mean $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. When $n-p$ is large,

$$\boldsymbol{a}^T\bar{\boldsymbol{X}} - \sqrt{\chi_p^2(\alpha)\frac{\boldsymbol{a}^T\boldsymbol{S}\boldsymbol{a}}{n}} \leq \mu_Z \leq \boldsymbol{a}^T\bar{\boldsymbol{X}} + \sqrt{\chi_p^2(\alpha)\frac{\boldsymbol{a}^T\boldsymbol{S}\boldsymbol{a}}{n}}$$

will contain $\boldsymbol{a}^T\boldsymbol{\mu}$ for every $\boldsymbol{a}$ with probability approximately $1-\alpha$. Consequently, we can make the simultaneous confidence statements

$$\bar{X}_i \pm \sqrt{\chi_p^2(\alpha)\frac{S_{ii}}{n}} \quad \text{contains} \quad \mu_i, \text{ for all } i,$$

and, in addition, for all pairs (μ_i, μ_k), the sample mean-centered ellipses

$$n\begin{bmatrix}\bar{X}_i-\mu_i & \bar{X}_k-\mu_k\end{bmatrix}\begin{bmatrix}S_{ii} & S_{ik}\\ S_{ik} & S_{kk}\end{bmatrix}^{-1}\begin{bmatrix}\bar{X}_i-\mu_i\\ \bar{X}_k-\mu_k\end{bmatrix} \leq \chi_p^2(\alpha) \text{ contain } (\mu_i, \mu_k).$$

Comparison

For a sufficiently large sample size,

1. the one-at-a-time confidence interval for individual means are

$$\bar{X}_i - z\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_{ii}}{n}} \leq \mu_i \leq \bar{X}_i + z\left(\frac{\alpha}{2}\right)\sqrt{\frac{S_{ii}}{n}}$$

2. the simultaneous T^2 intervals from the ellipsoid are

$$\bar{X}_i - \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{S_{ii}}{n}} \leq \mu_i \leq \bar{X}_i + \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{S_{ii}}{n}}$$

3. the Bonferroni simultaneous confidence intervals are

$$\bar{X}_i - z\left(\frac{\alpha}{2p}\right)\sqrt{\frac{S_{ii}}{n}} \leq \mu_i \leq \bar{X}_i + z\left(\frac{\alpha}{2p}\right)\sqrt{\frac{S_{ii}}{n}}$$