STATS 206
Applied Multivariate Analysis
Lecture 3: Inferences About a Mean Vector

Prathapasinghe Dharmawansa

Department of Statistics
Stanford University
Autumn 2013

Agenda

- Hotelling's T^2 test
- Likelihood ratio test and Hotelling's T^2
- Confidence regions
- Large sample inferences
- Dealing with missing values
- Effect of correlation among observations

Plausibility (Hypothesis Test) of a Mean Vector
Problem Formulation

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: random sample from a normal population $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- Our focus: the population mean $\boldsymbol{\mu}$
  - Is $\boldsymbol{\mu}_0$ (a known vector) plausible for $\boldsymbol{\mu}$?
- Formulation as a test of the two competing hypotheses:

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

H_0: the null hypothesis; H_1: the alternative hypothesis

- How to proceed? Start with the univariate case first (next page)

Plausibility (Hypothesis Test) of a Mean
The Univariate Case ($p = 1$)

$X_1, \ldots, X_n$: normal sample with mean μ

- Test of the two hypotheses: $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$
- Let $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$, $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2$

Test statistic : $t = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}} \sim$ Student's t with $(n-1)$ deg. of freedom

Squared distance : $t^2 = n(\overline{X} - \mu_0)(s^2)^{-1}(\overline{X} - \mu_0)$

– Reject H_0 at significance level α if and only if the observed

$$n(\overline{x} - \mu_0)(s^2)^{-1}(\overline{x} - \mu_0) > t^2_{n-1}(\alpha/2)$$

$t_{n-1}(\alpha/2)$: the upper $100(\alpha/2)$th percentile of the t dist. (above)

Plausibility (Hypothesis Test) of a Mean Vector
The multivariate Case ($p \geq 2$) (i)

$$\mathbf{X}_1, \ldots, \mathbf{X}_n\text{: normal sample} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- Testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$
- Generalization

$$p = 1 : t^2 = n(\overline{X} - \mu_0)(s^2)^{-1}(\overline{X} - \mu_0)$$

$p \geq 2$: Test statistic : Hotelling's T^2

$$T^2 = n(\overline{\mathbf{X}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}_0) \sim \frac{(n-1)p}{(n-p)} F_{p,n-p}$$

  - $\overline{\mathbf{X}}_{p\times 1} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{X}_j, \quad \mathbf{S}_{p\times p} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$
  - $\boldsymbol{\mu}_0 = [\mu_{10}, \ldots, \mu_{p0}]'$
  - F_{c_1,c_2}: F-dist. with c_1 and c_2 deg. of freedom

Plausibility (Hypothesis Test) of a Mean Vector

The multivariate Case ($p \geq 2$) (ii)

$$\mathbf{X}_1, \ldots, \mathbf{X}_n: \text{ normal sample } \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- $F_{c_1,c_2}(\alpha)$: the upper (100α)th percentile of F_{c_1,c_2}

  For all values of $\boldsymbol{\mu}, \boldsymbol{\Sigma}$

$$T^2 = n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \sim \frac{(n-1)p}{(n-p)} F_{p,n-p}$$

$$\Longrightarrow \; P\left[T^2 > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right]$$

$$= P\left[n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right] = \alpha$$

$$\overline{\mathbf{X}} = \tfrac{1}{n}\textstyle\sum_{j=1}^{n} \mathbf{X}_j \qquad \mathbf{S} = \tfrac{1}{n-1}\textstyle\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$$

Plausibility (Hypothesis Test) of a Mean Vector
The multivariate Case ($p \geq 2$) (iii)

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: normal sample $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ and $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$

- $F_{c_1,c_2}(\alpha)$: the upper (100α)th percentile of F_{c_1,c_2}
  Reject H_0 at significance level α if and only if the observed

$$n(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0) > \frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)$$

Plausibility (Hypothesis Test) of a Mean Vector

The multivariate Case ($p \geq 2$) (iv)

Remark: recall $(n-1)\mathbf{S} = \sum_{j=1}^{n} (\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})' \sim \mathbf{W}_{p,n-1}(\mathbf{\Sigma})$

$$p = 1 : t^2 = \underbrace{\sqrt{n}(\overline{X} - \mu_0)}_{\text{univariate normal}} \left(\underbrace{s^2}_{\substack{\text{scaled } \chi^2 \text{ random variable} \\ \text{divided by deg. of freedom}}} \right)^{-1} \underbrace{\sqrt{n}(\overline{X} - \mu_0)}_{\text{univariate normal}}$$

$$p \geq 2 : T^2 = \left[\underbrace{\sqrt{n}(\overline{\mathbf{X}} - \boldsymbol{\mu}_0)}_{N_p(\mathbf{0},\mathbf{\Sigma})} \right]' \left(\underbrace{\mathbf{S}}_{\substack{\text{Wishart random matrix} \\ (\mathbf{W}_{p,n-1}(\mathbf{\Sigma})) \text{ divided by} \\ \text{deg. of freedom}}} \right)^{-1} \left[\underbrace{\sqrt{n}(\overline{\mathbf{X}} - \boldsymbol{\mu}_0)}_{N_p(\mathbf{0},\mathbf{\Sigma})} \right]$$

Example 1: Testing a Multivariate Mean Vector

Sweat Data

X_1: Sweat rate; X_2: Sodium; X_3: Potassium (from $n = 20$ individuals)							
No	X_1	X_2	X_3	No	X_1	X_2	X_3
1	3.7	48.5	9.3	11	3.9	36.9	12.7
2	5.7	65.1	8	12	4.5	58.8	12.3
3	3.8	47.2	10.9	13	3.5	27.8	9.8
4	3.2	53.2	12	14	4.5	40.2	8.4
5	3.1	55.5	9.7	15	1.5	13.5	10.1
6	4.6	36.1	7.9	16	8.5	56.4	7.1
7	2.4	24.8	14	17	4.5	71.6	8.2
8	7.2	33.1	7.6	18	6.5	52.8	10.9
9	6.7	47.4	8.5	19	4.1	44.1	11.2
10	5.4	54.1	11.3	20	5.5	40.9	9.4

Table 5.1 in the textbook

Example 1: Testing a Multivariate Mean Vector
Calculating T^2 of the Sweat Data

- Assuming the data are multivariate normal (need to be checked first!)
- $\boldsymbol{\mu}_0 = [4, 50, 10]'$
- Test hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$
- $\underline{\alpha = 0.10}$
- It can be shown that $(n = 20, p = 3)$

$$T^2 = 9.7388 > \underbrace{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\Big|_{\substack{\alpha=0.1\\ n=20\\ p=3}} = 8.1726}_{\text{critical value here}}$$

$\Longrightarrow$ Reject H_0 at $\alpha = 0.1$ (level of significance)

Example 1: Testing a Multivariate Mean Vector

Calculating T^2 of the Sweat Data (MATLAB)

```
%%%%%% Calculating Hotelling's T^2 (T2)%%%%%%
load T5_1.dat
mu_0 = [4; 50; 10];
X = T5_1;
[n, p] = size(X);
X_Mean = (1/n) * X' * ones(n,1);
S = (1/(n-1)) * X' * (eye(n) - (1/n) * ones(n,1) * ones(1,n)) * X;
T2 = n * (X_Mean - mu_0)' * inv(S) * (X_Mean - mu_0);

%%%%%% Evaluating the critical value (CV)%%%%%%
alpha = 0.10;
CV = (n-1) * p * finv(1-alpha, p, n-p)/(n-p);
```

Invariance Property of Hotelling's T^2

Transformation: $\mathbf{Y}_{p\times 1} = \mathbf{C}_{p\times p}\mathbf{X}_{p\times 1} + \mathbf{d}_{p\times 1}$ ($\mathbf{C}$: non-singular)

- Previous techniques

$$\overline{\mathbf{y}} = \mathbf{C}\overline{\mathbf{x}} + \mathbf{d},\ \boldsymbol{\mu}_{\mathbf{Y}} = \mathbf{C}\boldsymbol{\mu} + \mathbf{d}$$

$$\mathbf{S}_{\mathbf{y}} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{y}_j - \overline{\mathbf{y}})(\mathbf{y}_j - \overline{\mathbf{y}})' = \mathbf{CSC}'$$

- Let $\boldsymbol{\mu}_{\mathbf{Y},0} = \mathbf{C}\boldsymbol{\mu}_0 + \mathbf{d}$

$$\begin{aligned} T^2 &= n(\overline{\mathbf{y}} - \boldsymbol{\mu}_{\mathbf{y},0})'\mathbf{S}_{\mathbf{y}}^{-1}(\overline{\mathbf{y}} - \boldsymbol{\mu}_{\mathbf{y},0}) \quad \text{(for } \mathbf{y}_1, \ldots, \mathbf{y}_n\text{)} \\ &= n[\mathbf{C}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)]'(\mathbf{CSC}')^{-1}[\mathbf{C}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)] \\ &= n(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{C}'(\mathbf{C}')^{-1}\mathbf{S}^{-1}\mathbf{C}^{-1}\mathbf{C}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0) \\ &= n(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0) \quad \text{(for } \mathbf{x}_1, \ldots, \mathbf{x}_n\text{)} \end{aligned}$$

- Testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ of $\mathbf{x} \Longleftrightarrow$ testing $H_0 : \boldsymbol{\mu}_{\mathbf{Y}} = \boldsymbol{\mu}_{\mathbf{Y},0}$ of $\mathbf{y}$

Likelihood Ratio Tests of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ (i)

- Likelihood ratio test of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$
- Recall: Maximum of the multivariate normal likelihood [unknown $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$]

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

where the maximum is achieved using the following estimates

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})'$$

$$\widehat{\boldsymbol{\mu}} = \overline{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j$$

Likelihood Ratio Tests of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ (ii)

- Under the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ (unknown $\boldsymbol{\Sigma}$ only) (multivariate normal distribution)

$$L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left(-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{x}_j - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j - \boldsymbol{\mu}_0)\right)$$

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

where the maximum here is achieved using the following estimate

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

Likelihood Ratio Tests of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ (iii)

- (Normal population) Is $\boldsymbol{\mu}_0$ plausible for $\boldsymbol{\mu}$? Check

$$\underbrace{\text{Likelihood Ratio} = \Lambda = \frac{\max_{\boldsymbol{\Sigma}} \; L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} \; L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} = \left(\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widehat{\boldsymbol{\Sigma}}_0|} \right)^{n/2}}_{\text{Likelihood ratio statistic}}$$

- Equivalently, check Wilks' lambda $= \Lambda^{2/n} = |\widehat{\boldsymbol{\Sigma}}|/|\widehat{\boldsymbol{\Sigma}}_0|$
- c_α: the lower (100α)th percentile of the dist. of Λ

  Likelihood ratio test rejects H_0 if

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

Likelihood Ratio Tests of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ (iv)

Hotelling's T^2 and Wilks' lambda $\Lambda^{2/n}$

$$\boxed{\Lambda^{2/n} = |\widehat{\mathbf{\Sigma}}|/|\widehat{\mathbf{\Sigma}}_0| = \left(1 + \tfrac{T^2}{n-1}\right)^{-1}}$$

- Proof:

$$\widehat{\mathbf{\Sigma}}_0 = \widehat{\mathbf{\Sigma}} + (\overline{\mathbf{x}} - \boldsymbol{\mu}_0)(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)' \quad \text{(Need some algebra here!)}$$

$$|\widehat{\mathbf{\Sigma}}_0| = |\widehat{\mathbf{\Sigma}} + (\overline{\mathbf{x}} - \boldsymbol{\mu}_0)(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)'| \overset{(*)}{=} |\widehat{\mathbf{\Sigma}}||1 + (\overline{\mathbf{x}} - \boldsymbol{\mu}_0)'\widehat{\mathbf{\Sigma}}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)|$$

$$= |\widehat{\mathbf{\Sigma}}| \left| 1 + \frac{n}{n-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0) \right| = |\widehat{\mathbf{\Sigma}}| \left(1 + \frac{T^2}{n-1}\right)$$

$$(*) : |\mathbf{A}_{m\times m} + \mathbf{a}_{m\times 1}(\mathbf{b}_{m\times 1})'| = |\mathbf{A}|(1 + \mathbf{b}'\mathbf{A}^{-1}\mathbf{a}) \quad (\mathbf{A} \text{ non-singular})$$

- For normal samples, test based on $T^2 \Longleftrightarrow$ test based on Λ ($\Lambda^{2/n}$)
- $T^2 = (n-1)(|\widehat{\mathbf{\Sigma}}_0|/|\widehat{\mathbf{\Sigma}}| - 1)$: no need to calculate $\mathbf{S}^{-1}$

General Likelihood Ratio Method

- General observations: $\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n$
- $\boldsymbol{\theta}$: vector of unknown population parameters ($\boldsymbol{\theta} \in \boldsymbol{\Theta}$)
- $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ vs. $H_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0$ ($\boldsymbol{\Theta}_0 \subset \boldsymbol{\Theta}$; $\boldsymbol{\Theta}$: whole parameter space)
- General likelihood ratio test statistic: Λ

  Reject H_0 if

$$\Lambda = \frac{\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \; L(\boldsymbol{\theta})}{\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \; L(\boldsymbol{\theta})} < \; c \text{ (a suitably chosen constant)}$$

- Remark:

$$-2 \ln \Lambda \;\xrightarrow{\text{approximately}}\; \chi^2_{\nu - \nu_0} \quad (n \to \infty)$$

ν: dimension of $\boldsymbol{\Theta}$; ν_0: dimension of $\boldsymbol{\Theta}_0$

Confidence Regions (i)

- For a general random sample $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]'$
  Unknown parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (parameter space)

  $$R(\mathbf{X}) : 100(1-\alpha)\% \text{ confidence region} \iff \underbrace{P[\boldsymbol{\theta} \in R(\mathbf{X})]}_{\substack{\text{calculated under the} \\ \text{true but unkown } \boldsymbol{\theta}}} = 1-\alpha$$

- Recall: for a p-dimensional normal population, $\forall \boldsymbol{\mu}, \boldsymbol{\Sigma}$

  $$P\left[T^2 = n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \le \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right] = 1-\alpha$$

  $$\mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$$

  $F_{c_1,c_2}(\alpha)$: the upper (100α)th percentile of F_{c_1,c_2}

Confidence Regions (ii)

- Given $\mathbf{x}_1, \ldots, \mathbf{x}_n$: observations from a p-dim normal population

  $100(1-\alpha)\%$ confidence region for the mean $\boldsymbol{\mu}$ of a p-dim. normal dist.

$$\left\{ \boldsymbol{\mu} : n(\overline{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}) \leq \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha) \right\}$$

$$\left[\overline{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j, \quad \mathbf{S} = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})' \right]$$

- Relation to testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$

  The above confidence region consists of all $\boldsymbol{\mu}_0$ with which H_0 would not be rejected by the T^2 test at significance level α.

Confidence Regions (iii)

- $100(1-\alpha)\%$ confidence region for the mean $\boldsymbol{\mu}$ of a p-dim. normal dist.

$$\underbrace{\left\{\boldsymbol{\mu} : n(\overline{\mathbf{x}}-\boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu}) \le \frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha) = c^2\right\}}_{100(1-\alpha)\%\ \text{confidence ellipsoid}}$$

- Given that: $\mathbf{S}\mathbf{e}_i = \lambda_i \mathbf{e}_i, i = 1, \ldots, p$

Center of the confidence ellipsoid $= \overline{\mathbf{x}}$

$$\text{Axes of the confidence ellipsoid} = \pm\, c\sqrt{\lambda_i}\,\mathbf{e}_i$$
$$= \pm\,\sqrt{\lambda_i}\sqrt{\frac{(n-1)p}{n(n-p)}F_{p,n-p}(\alpha)}\,\mathbf{e}_i$$

(similar to constant prob. density contour analysis of $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$)

Simultaneous Confidence Intervals (i)

$$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Longrightarrow Z = \mathbf{a}'\mathbf{X} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}); \quad \mathbf{a} \neq \mathbf{0}$$

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: random sample $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; $\quad Z_j = \mathbf{a}'\mathbf{X}_j, \forall j$

$\overline{\mathbf{x}}$ ($\mathbf{S}$): sample mean (cov. matrix) from $\mathbf{x}_1, \ldots, \mathbf{x}_n$

- Sample mean and var. of $z_1, \ldots, z_n$: $\quad \overline{z} = \mathbf{a}'\overline{\mathbf{x}}$, $s_z^2 = \mathbf{a}'\mathbf{S}\mathbf{a}$
- $100(1-\alpha)\%$ confidence interval for $\mu_Z = \mathbf{a}'\boldsymbol{\mu}$ (given $\mathbf{a}$, unknown σ_Z^2)

$$\mathbf{a}'\overline{\mathbf{x}} - t_{n-1}(\alpha/2)\frac{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}}}{\sqrt{n}} \leq \mathbf{a}'\boldsymbol{\mu} \leq \mathbf{a}'\overline{\mathbf{x}} + t_{n-1}(\alpha/2)\frac{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}}}{\sqrt{n}}$$

using Student's t : $|t| = \left|\frac{\overline{z} - \mu_Z}{s_z/\sqrt{n}}\right| = \left|\frac{\sqrt{n}(\mathbf{a}'\overline{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu})}{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}}}\right| \leq t_{n-1}(\alpha/2)$

Simultaneous Confidence Intervals (ii)

- A simultaneous region satisfies: $\forall \mathbf{a} : t^2 = \left(\frac{\sqrt{n}(\mathbf{a}'\overline{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu})}{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}}}\right)^2 \leq c^2$
- If $\max_{\mathbf{a}} t^2 \leq c^2$, then the above holds.

$$\begin{aligned} \max_{\mathbf{a}} \; t^2 &= \max_{\mathbf{a}} \; \frac{n\left(\mathbf{a}'(\overline{\mathbf{x}} - \boldsymbol{\mu})\right)^2}{\mathbf{a}'\mathbf{S}\mathbf{a}} = n\left[\max_{\mathbf{a}} \; \frac{\left(\mathbf{a}'(\overline{\mathbf{x}} - \boldsymbol{\mu})\right)^2}{\mathbf{a}'\mathbf{S}\mathbf{a}}\right] \\ &= n(\overline{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}) \\ &= T^2 \end{aligned}$$

Maximum achieved with $\mathbf{a} = b \cdot \mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu})$ ($b \neq 0$: constant)

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

Simultaneous Confidence Intervals (iii)

If: $\mathbf{X}_1, \ldots, \mathbf{X}_n$: random sample $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ($\boldsymbol{\Sigma}$ positive definite)

Then: simultaneously for all $\mathbf{a}$,

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

will contain $\mathbf{a}'\boldsymbol{\mu}$ with prob. $(1-\alpha)$.

- The simultaneous intervals are referred to as T^2-intervals.

Simultaneous Confidence Intervals (iv)

- Choices of $\mathbf{a} = [1, 0, \ldots, 0]', \ [0, 1, \ldots, 0]', \ \ldots, \ [0, \ldots, 0, 1]'$

$$\Rightarrow \overline{x}_j - \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{jj}}{n}} \leq \mu_j \leq \overline{x}_j + \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{jj}}{n}}, \quad j = 1, \ldots, p$$

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

Simultaneous Confidence Intervals (v)

- Let $\mathbf{a} = [0, \ldots, 0, a_i, 0, \ldots, 0, a_k, 0, \ldots, 0]'$, $\quad a_i = 1, a_k = -1$

$$\Rightarrow \overline{x}_i - \overline{x}_k - \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{ii} - 2s_{ik} + s_{kk}}{n}}$$

$$\leq \mu_i - \mu_k$$

$$\leq \overline{x}_i - \overline{x}_k + \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{ii} - 2s_{ik} + s_{kk}}{n}}$$

One-at-a-Time Intervals

- Consider each component μ_j one at a time, for $j = 1, \ldots, p$:

$$\overline{x}_j - t_{n-1}(\alpha/2)\sqrt{\frac{s_{jj}}{n}} \leq \mu_j \leq \overline{x}_j + t_{n-1}(\alpha/2)\sqrt{\frac{s_{jj}}{n}}$$

  - No clue of P[all t-intervals above contain the μ_j's] in general
  - When X_i's are independent

$$P[\text{all } t\text{-intervals above contain the } \mu_j\text{'s}] = (1-\alpha)^p$$

  - The T^2-interval is wider than the individual interval for each μ_j. (Comparison next page)

Comparison of T^2 and One-at-a-Time Intervals

Critical distance multipliers for both intervals $(1 - \alpha = 0.95)$

sample size		$\sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p}(0.05)}$	
n	$t_{n-1}(0.025)$	$p = 4$	$p = 10$
15	2.1448	4.1338	11.5144
25	2.0639	3.6032	6.3796
50	2.0096	3.3117	5.0444
100	1.9842	3.1897	4.6166
∞	1.9600	3.0750	4.2769

Bonferroni Method for Multiple Comparisons (i)

- Motivation: in practice
  - T^2 intervals can be too wide to use
  - Only m (finite) linear combinations needed: $\mathbf{a}_1, \ldots, \mathbf{a}_m$
    $\overset{\text{Bonferroni}}{\Longrightarrow}$ Constructing shorter intervals than T^2
- C_i: confidence statement about $\mathbf{a}_i'\boldsymbol{\mu}$, $P(C_i \text{ true}) = 1 - \alpha_i$, $i = 1, \ldots, m$

$$
\begin{aligned}
P[\text{all } C_i\text{'s true}] &= 1 - P[\text{at least one } C_i \text{ false}] \\
&\overset{\text{Bonferroni}}{\geq} 1 - \sum_{i=1}^{m} P[C_i \text{ false}] = 1 - \sum_{i=1}^{m}(1 - P[C_i \text{ true}]) \\
&= 1 - (\alpha_1 + \ldots + \alpha_m)
\end{aligned}
$$

control of overall type-I error rate $\sum_i \alpha_i$; flexibility in choosing each α_i

Bonferroni Method for Multiple Comparisons (ii)

- Let $z_i = \mathbf{a}_i'\mathbf{x}_i$; $s_{z,ii} = \mathbf{a}_i'\mathbf{S}\mathbf{a}_i$; without prior info., $\alpha_i = \frac{\alpha}{m}, \forall i$

$$P\left[\overline{z}_i \pm t_{n-1}(\alpha/(2m))\sqrt{s_{z,ii}/n} \text{ contains } \mathbf{a}_i'\boldsymbol{\mu}\right] = 1 - \alpha/m, \forall i$$

$$P\left[\overline{z}_i \pm t_{n-1}(\alpha/(2m))\sqrt{\frac{s_{z,ii}}{n}} \text{ contains } \mathbf{a}_i'\boldsymbol{\mu}, \underline{\forall i}\right] \geq 1 - \sum_{i=1}^{m} \frac{\alpha}{m} = 1 - \alpha$$

- Set $m = p$, $\mathbf{a}_i = [0, \ldots, 0, 1, 0, \ldots, 0]'$ (1 at the i-th entry only), $\forall i$:
  Then the following intervals (the Bonferroni intervals)

$$\overline{x}_i - t_{n-1}(\alpha/(2p))\sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \overline{x}_i + t_{n-1}(\alpha/(2p))\sqrt{\frac{s_{ii}}{n}}, \quad i = 1, \ldots, p$$

hold (simultaneously) with probability $\geq 1 - \alpha$

Comparison of T^2 and Bonferroni Intervals

- Comparing Bonferroni and T^2 intervals (same $\mathbf{a}_i$'s, $\alpha_i = \alpha/p$):

$$t_{n-1}(\alpha/(2p)) \leftarrow \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p}(\alpha)} \text{ (difference in intervals)}$$

$$\frac{\text{Length of Bonferroni interval}}{\text{Length of } T^2 \text{ interval}} = \frac{t_{n-1}(\alpha/(2p))}{\sqrt{\frac{p(n-1)}{n-p} F_{p,n-p}(\alpha)}}$$

- Table of the above ratio ($\alpha_i = \alpha/p, 1 - \alpha = 0.95$)

	p		
n	2	4	10
15	0.8766	0.6928	0.2888
25	0.8947	0.7494	0.4844
50	0.9060	0.7831	0.5828
100	0.9110	0.7976	0.6219
∞	0.9199	0.8122	0.6563

Large Sample Inferences about a Mean Vector (i)

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: random sample

(mean: $\boldsymbol{\mu}_{p\times 1}$, finite positive definite covariance matrix: $\boldsymbol{\Sigma}_{p\times p}$)

(no normality assumption!!!)

- Recall: for large $n-p$,

$$n(\overline{\mathbf{X}}-\boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}}-\boldsymbol{\mu}) \longrightarrow \chi_p^2 \text{ distributed}$$

$$\Longrightarrow \; P[n(\overline{\mathbf{X}}-\boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}}-\boldsymbol{\mu}) \le \chi_p^2(\alpha)] \approx 1-\alpha$$

$\chi_p^2(\alpha)$: the upper (100α)th percentile of χ_p^2

- Consequences: next page

Large Sample Inferences about a Mean Vector (ii)

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: random sample (mean: $\boldsymbol{\mu}$, finite positive def. cov.: $\boldsymbol{\Sigma}$)

(no normality assumption)

- (Asymptotic testing)
  For large $n-p$: at significance level approximately α:
  reject the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ in favor of $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ if
  $$n(\overline{\mathbf{X}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}_0) > \chi_p^2(\alpha).$$

- (Asymptotic simultaneous confidence intervals)
  For large $n-p$ and for every $\mathbf{a}$
  $$\mathbf{a}'\overline{\mathbf{X}} \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}$$
  will contain $\mathbf{a}'\boldsymbol{\mu}$ with probability approximately $1-\alpha$.

Note: For large n and large $n-p$, $[p(n-1)/(n-p)]F_{p,n-p} \approx \chi_p^2(\alpha)$

Dealing with Missing Observations in Normal Samples

- Important assumption: data are missing at random!
- One of the methods: the EM algorithm (see footnote for a ref.)
- Recall: (multivariate normal distribution)

1. (Cond. dist. of a subset) $\mathbf{Y} = [\mathbf{Y}_1' \ \mathbf{Y}_2']' \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $|\boldsymbol{\Sigma}_{22}| > 0$
   $\mathbf{Y}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$, $\mathbf{Y}_2 \sim N_{p-q}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$, $\mathrm{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) = \boldsymbol{\Sigma}_{12}$, $q \leq p$

$$\Longrightarrow \mathbf{Y}_1|_{\mathbf{Y}_2=\mathbf{y}_2} \sim N_q\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)$$

2. $\mathbf{X}_1, \ldots, \mathbf{X}_n$: sample $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ($\boldsymbol{\mu}, \boldsymbol{\Sigma}$ unknown, to be estimated)

$\overline{\mathbf{X}}$ and $\mathbf{S}$: sufficient statistics $\Longrightarrow$ so are:

$$\mathbf{T}_1 = \sum_j \mathbf{X}_j = n\overline{\mathbf{X}} \quad \text{and} \quad \mathbf{T}_2 = \sum_j \mathbf{X}_j\mathbf{X}_j' = (n-1)\mathbf{S} + n\overline{\mathbf{X}}\overline{\mathbf{X}}'$$

Dempster, Laird and Rubin, *Maximum Likelihood from Incomplete Data via the EM Algorithm*, Journal of the Royal Statistical Society, Series B (Methodological), Vol. 39, No. 1, (1977), pp. 1-38.

Dealing with Missing Observations in Normal Samples

The EM Algorithm (Assuming Data Missing at Random) (i)

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: p-dim. normal sample with unknown $\boldsymbol{\mu}, \boldsymbol{\Sigma}$

- Initializing $\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\mathbf{T}}_1, \widehat{\mathbf{T}}_2$ (for $\mathbf{T}_1, \mathbf{T}_2$, see prev. page)
- E-step: for each vector with missing values: $\mathbf{x}_j = [\mathbf{x}_j^{(1)'}, \mathbf{x}_j^{(2)'}]'$:

  ($\mathbf{x}_j^{(1)}$: missing, $\mathbf{x}_j^{(2)}$: available)

$$(\star)\ \widehat{\mathbf{x}}_j^{(1)} = \mathsf{E}(\mathbf{X}_j^{(1)} | \mathbf{x}_j^{(2)}; \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = \widehat{\boldsymbol{\mu}}_1 + \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}(\mathbf{x}_j^{(2)} - \widehat{\boldsymbol{\mu}}_2)$$

$$(\diamond)\ \widehat{\mathbf{x}_j^{(1)}\mathbf{x}_j^{(1)'}} = \mathsf{E}(\mathbf{X}_j^{(1)}\mathbf{X}_j^{(1)'} | \mathbf{x}_j^{(2)}; \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$$

$$= \widehat{\boldsymbol{\Sigma}}_{11} - \widehat{\boldsymbol{\Sigma}}_{12}\widehat{\boldsymbol{\Sigma}}_{22}^{-1}\widehat{\boldsymbol{\Sigma}}_{21} + \widehat{\mathbf{x}}_j^{(1)}\widehat{\mathbf{x}}_j^{(1)'}$$

$$(\circ)\ \widehat{\mathbf{x}_j^{(1)}\mathbf{x}_j^{(2)'}} = \mathsf{E}(\mathbf{X}_j^{(1)}\mathbf{X}_j^{(2)'} | \mathbf{x}_j^{(2)}; \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = \widehat{\mathbf{x}}_j^{(1)}\widehat{\mathbf{x}}_j^{(2)'}$$

($\star$): estimate of missing data; use ($\star$), ($\diamond$), ($\circ$) to update $\widehat{\mathbf{T}}_1$, $\widehat{\mathbf{T}}_2$

Dealing with Missing Observations in Normal Samples

The EM Algorithm (Assuming Data Missing at Random) (ii)

$\mathbf{X}_1, \ldots, \mathbf{X}_n$: p-dim. normal sample with unknown $\boldsymbol{\mu}, \boldsymbol{\Sigma}$

- M-step: compute the updates of ML estimates

$$\widehat{\boldsymbol{\mu}} = \frac{\widehat{\mathbf{T}}_1}{n}, \quad \widehat{\boldsymbol{\Sigma}} = \frac{\widehat{\mathbf{T}}_2}{n} - \widehat{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}'$$

- Recursively continue the above E-step and M-step until the difference between updates and previous values are smaller than a given small value.

Effect of Correlation among Observations (i)

- Suppose observations are collected over time
- Time correlation: using the first-order autoregressive (AR (1)) model

$$\mathbf{X}_t - \boldsymbol{\mu} = \boldsymbol{\Phi}(\mathbf{X}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})(\boldsymbol{\varepsilon}_t : \text{ i.i.d. over } t)$$

  Also assuming $|\lambda_i(\boldsymbol{\Phi})| < 1, \forall i$
- Clearly,

$$\begin{aligned}\mathbf{X}_t - \boldsymbol{\mu} &= \boldsymbol{\varepsilon}_t + \boldsymbol{\Phi}(\mathbf{X}_{t-1} - \boldsymbol{\mu}) \\ &= \boldsymbol{\varepsilon}_t + \boldsymbol{\Phi}[\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Phi}(\mathbf{X}_{t-2} - \boldsymbol{\mu})] \\ &= \boldsymbol{\varepsilon}_t + \boldsymbol{\Phi}\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Phi}^2\boldsymbol{\varepsilon}_{t-2} + \dots\end{aligned}$$

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \mathsf{E}(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})' = \sum_{j=0}^{\infty} \boldsymbol{\Phi}^j \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \boldsymbol{\Phi}'^j$$

$$\mathsf{Cov}(\mathbf{X}_t, \mathbf{X}_{t-r}) = \boldsymbol{\Phi}^r \boldsymbol{\Sigma}_{\mathbf{x}}$$

Effect of Correlation among Observations (ii)

- Collect $\mathbf{X}_t$ for n consecutive times to get $\mathbf{X}_1, \ldots, \mathbf{X}_n$;
  It can be shown that

$$\mathbf{S} = \frac{1}{n-1}\sum_{t=1}^{n}(\mathbf{X}_t - \overline{\mathbf{X}})(\mathbf{X}_t - \overline{\mathbf{X}})' \to \boldsymbol{\Sigma}_{\mathbf{x}} \text{ (in prob.)} \quad (n \to \infty)$$

$$\overline{\mathbf{X}} \to \boldsymbol{\mu}, \quad \text{Cov}\left(n^{-1/2}\sum_{t=1}^{n}\mathbf{X}_t\right) \to \boldsymbol{\Psi} \text{ (in prob.)} \quad (n \to \infty)$$

where $\quad \boldsymbol{\Psi} = (\mathbf{I} - \boldsymbol{\Phi})^{-1}\boldsymbol{\Sigma}_{\mathbf{x}} + \boldsymbol{\Sigma}_{\mathbf{x}}(\mathbf{I} - \boldsymbol{\Phi}')^{-1} - \boldsymbol{\Sigma}_{\mathbf{x}}$

Furthermore, for large n

$$\sqrt{n}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \approx \text{ distributed as } N_p(\mathbf{0}, \boldsymbol{\Psi})$$

$$\Longrightarrow n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\boldsymbol{\Psi}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \approx \text{ distributed as } \chi_p^2$$

Effect of Correlation among Observations (iii)

- Large n, $\mathbf{X}_1, \ldots, \mathbf{X}_n$: i.i.d. ($\mathbf{S} \to \boldsymbol{\Sigma}_{\mathbf{x}}$ in prob.)

$$\Longrightarrow \left\{ \boldsymbol{\mu} : \underbrace{n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu})}_{\text{asymptotically } n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu}) \sim \chi_p^2} \leq \chi_p^2(0.05) \right\}$$

covers $\boldsymbol{\mu}$ with prob. 0.95 (asymptotically).

- Large n, $\mathbf{X}_1, \ldots, \mathbf{X}_n$: AR(1) $[\boldsymbol{\Phi} = \phi\mathbf{I}, (|\phi| < 1), \ \ \boldsymbol{\Psi} = \frac{1+\phi}{1-\phi}\boldsymbol{\Sigma}_{\mathbf{x}}]$,

$$\Longrightarrow \left\{ \boldsymbol{\mu} : \underbrace{n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu})}_{\text{asymptotically } n(\overline{\mathbf{X}} - \boldsymbol{\mu})'[\frac{1+\phi}{1-\phi}\boldsymbol{\Psi}^{-1}](\overline{\mathbf{X}} - \boldsymbol{\mu}) \sim \frac{1+\phi}{1-\phi}\chi_p^2} \leq \chi_p^2(0.05) \right\}$$

covers $\boldsymbol{\mu}$ with $\underbrace{P[\chi_p^2 \leq (1-\phi)(1+\phi)^{-1}\chi_p^2(0.05)]}_{\text{tabulated on next page}}$ (asymptotically).

Effect of Correlation among Observations (iv)

Table:

$$P\left[\chi_p^2 \leq (1-\phi)(1+\phi)^{-1}\chi_p^2(0.05)\right]$$

	ϕ			
p	-0.25	0	0.25	0.5
1	0.9886	0.95	0.8710	0.7422
2	0.9932	0.95	0.8343	0.6316
5	0.9976	0.95	0.7514	0.4052
10	0.9993	0.95	0.6413	0.1934
15	0.9997	0.95	0.5484	0.0902

Conclusion: correlation among observations can cause the coverage prob. to drop significantly (e.g., comparing some of the above with 0.95 in the case with i.i.d. observations).