What to do today ?

Part I. Introduction and Preparation

Part II. Inference under Multivariate Normal Distribution (Chp 4-7)

II.1 Multivariate Normal Distribution (Chp 4)

II.1.1 Multivariate Normal Distribution $MN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (Chp 4.1-2)

II.1.2 Estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (Chp 4.3)

II.1.3 Properties of $\bar{\mathbf{X}}$ and $\mathbf{S}$ (Chp 4.4-5)

II.1.4 More on Normality (Chp 4.6-8)

II.2 Inferences on Mean Vector (Chp 5)

II.2.1 Introduction (Chp5.1-2a)

II.2.2 Hotelling's T^2 and Likelihood Ratio Test (Chp5.2b-3)

II.3 Comparisons of Several Mean Vector (Chp 6)

II.4 Multivariate Linear Regression (Chp 7)

II.1.2 Estimation of μ and Σ - Review

Consider $\mathbf{X} \sim MN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: what are $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$?
Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are iid observations on $\mathbf{X}$, how to use the data to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$?

II.1.2A. By Method of Moments

- $E(\mathbf{X}) = \boldsymbol{\mu}$ and thus the MME $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} = \sum_{i=1}^{n} \mathbf{X}_i \big/ n$, the sample mean.
- $V(\mathbf{X}) = \boldsymbol{\Sigma} = E(\mathbf{X}\mathbf{X}^{'}) - \big(E\mathbf{X}\big)\big(E\mathbf{X}\big)^{'}$ and thus the MME

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^{'} - \bar{\mathbf{X}}\bar{\mathbf{X}}^{'}.$$

$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^{'} = \frac{n-1}{n}\mathbf{S}$.

II.1.2B. By Maximum Likelihood Estimation

The likelihood function is

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\Big[-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^{'}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})^{'}\Big].$$

The *MLE* $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}$ and $\hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\mathbf{S}$ are the same as the *MME*.

II.1.3 Properties of $\bar{\mathbf{X}}$ and S - Review

- *Unbiased estimators*: $E(\bar{\mathbf{X}}) = \boldsymbol{\mu}$ and $E(\mathbf{S}) = \boldsymbol{\Sigma}$, while $E(\hat{\boldsymbol{\Sigma}}) = \frac{n-1}{n}\boldsymbol{\Sigma}$.

- *Relationship*: $\bar{\mathbf{X}}$ and $\mathbf{S}$ are independent; so are $\bar{\mathbf{X}}$ and $\hat{\boldsymbol{\Sigma}}$.

- *Limits:* $\bar{\mathbf{X}} \to \boldsymbol{\mu}$ and $\mathbf{S} \to \boldsymbol{\Sigma}$ as $n \to \infty$.

- *Distributions:*
  - $\bar{\mathbf{X}} = \sum_{i=1}^{n} \mathbf{X}_i / n \sim MN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$ by noting

    $$\bar{\mathbf{X}} = \frac{1}{n}(\mathbf{I}, \ldots, \mathbf{I}) \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}.$$

  - $n\hat{\boldsymbol{\Sigma}} = (n-1)\mathbf{S} = \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$ follows a *Wishart* distribution with degree of freedom $n - 1$.

Definition. If $\mathbf{Z}_1, \ldots, \mathbf{Z}_m$ are indpt and follow $MN_p(\mathbf{0}, \mathbf{\Sigma})$, the distribution of $\sum_{j=1}^{m} \mathbf{Z}_j \mathbf{Z}_j'$ is the Wishart Distribution $W_p(\mathbf{\Sigma}, m)$.

- Special case of $p = 1$ and $\mathbf{\Sigma} = \sigma^2 = 1$: χ_m^2-distn
- Suppose $\mathbf{W}_1 \sim W_p(\mathbf{\Sigma}, m_1)$ and $\mathbf{W}_2 \sim W_p(\mathbf{\Sigma}, m_2)$. If $\mathbf{W}_1$ and $\mathbf{W}_2$ are indpt, $\mathbf{W}_1 + \mathbf{W}_2 \sim W_p(\mathbf{\Sigma}, m_1 + m_2)$
- If $\mathbf{W} \sim W_p(\mathbf{\Sigma}, m)$, $\mathbf{C}\mathbf{W}\mathbf{C}' \sim W_q(\mathbf{C}\mathbf{\Sigma}\mathbf{C}', m)$ when $\mathbf{C}$ is $q \times p$ matrix with rank of q.

II.1.4 More on Normality (Chp 4.6-8)

II.1.4A Assessing normality assumption

- Check on the univariate marginal distributions by Q-Q plot (textbook p177-180)
- Use the fact that $(\mathbf{X}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}_i - \boldsymbol{\mu}) \sim \chi_p^2$ if $\mathbf{X}_i \sim MV_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $i = 1, \ldots, n$
  - The proportion of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ that fall in $\left\{\mathbf{x} : (\mathbf{x} - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}) \le \chi_p^2(.95)\right\}$ should be about .95 if the data are from a normal distn.
  - By *Chi-square plot*, check on whether $d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$ for $i = 1, \ldots, n$ form a sample from χ_p^2.

Example. (Textbook Example 4.14, p186) Four measurements of stiffness.

II.1.4 More on Normality (Chp 4.6-8)

II.1.4B Detecting outliers and data cleaning

- Make a dot plot for each variable.
- Make a scatter plot for each pair of variable.
- Calculate the standardized values $z_{ik} = (x_{ik} - \bar{x}_k) \big/ \sqrt{s_{kk}}$ for $i = 1, \ldots, n$ and $k = 1, \ldots, p$, and examine those for unusually large/small values.
- Calculate $d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$ for $i = 1, \ldots, n$, and examine those for unusually large/small values. (In a Chi-square plot, check the points farthest from the origin.)

Example. (cont'd) (Textbook Example 4.15, p190)

II.1.4 More on Normality (Chp 4.6-8)

II.1.4C Examples of transformation to near normality

- If the orginal scale is *counts*, consider $\sqrt{y}$.

- If the orginal scale is *proportions*, consider $logit(y) = \frac{1}{2}\log\left(\frac{y}{1-y}\right)$.

- If the orginal scale is *correlations*, consider Fisher's transformation $z(y) = \frac{1}{2}\log\left(\frac{1+y}{1-y}\right)$.

- In general, consider the Box-Cox transformation:

$$x^{(\lambda)} = \begin{cases} \frac{y^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \ln y & \lambda = 0 \end{cases}$$

To use λ^* that maximizes

$$l(\lambda) = -\frac{n}{2}\ln\left[\frac{1}{n}\sum_{i=1}^{n}(x_i^{(\lambda)} - \bar{x}^{(\lambda)})^2\right] + (\lambda - 1)\sum_{i=1}^{n}\ln x_i$$

II.2 Inferences on Mean Vector (Chp 5)

II.2.1 Introduction (Chp5.1-2a)

Consider $\mathbf{X} \sim MN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: what are $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, provided $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are iid observations on $\mathbf{X}$?

- How to use the data to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$?

  e.g. the MME and MLE: $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}$ and $\hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\mathbf{S}$

  sufficiently to address all the related issues?

- How to use the data to test on hypotheses about $\boldsymbol{\mu}$ or/and $\boldsymbol{\Sigma}$?

What did we do with a univariate normal r.v.?

Recall ... ... Consider r.v. $X \sim N(\mu, \sigma^2)$ with iid observations $X_1, \dots, X_n$.

- *Point Estimation.*
  - the MME and MLE of μ: $\hat{\mu} = \bar{X}$
  - the MME and MLE of σ^2: $\hat{\sigma}^2 = \frac{n-1}{n} s^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$

- *Interval Estimation.* e.g. 95% *Confidence Interval of* μ
  - 95% CI of μ with known σ^2:

$$\Big(\underbrace{\bar{X} - 1.96\sqrt{\tfrac{\sigma^2}{n}}}_{\text{lower limit}}, \quad \underbrace{\bar{X} + 1.96\sqrt{\tfrac{\sigma^2}{n}}}_{\text{upper limit}}\Big)$$

$1.96 = z_{0.025}$ with $Z \sim N(0, 1)$ and $P(Z > z_{0.025}) = 2.5\%$.

Interpretation:

To say that we are 95% confident is shorthand for "95% of all possible samples of a given size from this population will result in an interval that captures the unknown parameter."

Recall ... ... Consider r.v. $X \sim N(\mu, \sigma^2)$ with iid observations $X_1, \dots, X_n$.

- *Interval Estimation.* e.g. 95% *Confidence Interval of* μ
  - 95% CI of μ with unknown σ^2:

$$(\bar{X} - t_{0.025}(n-1)\sqrt{\frac{s^2}{n}}, \quad \bar{X} + t_{0.025}(n-1)\sqrt{\frac{s^2}{n}})$$

s^2: the sample variance; $t_{\alpha/2}(n-1)$ – the $\alpha/2$-right tail of the Student t-distribution with df $= n-1$.

What if $\mathbf{X} \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *with iid observations* $\mathbf{X}_1, \dots, \mathbf{X}_n$*?* $\Longrightarrow$ **Chp5.4: Part II.2.3**

Recall ... ... Consider r.v. $X \sim N(\mu, \sigma^2)$ with iid observations $X_1, \ldots, X_n$.

- *Hypothesis Tests about* μ. e.g. $H_0 : \mu = \mu_0$ by *Wald-type* Testing
  - When σ^2 is known:
    - *test statistic.* $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1)$ under H_0
    - *rejection region.* with pre-determined α,
      (i) when $H_1 : \mu \neq \mu_0$: $\mathcal{R} = \left\{ z : |z| > z_{\alpha/2} \right\}$
      (ii) when $H_1 : \mu < \mu_0$: $\mathcal{R} = \left\{ z : z < -z_{\alpha} \right\}$
      (iii) when $H_1 : \mu > \mu_0$: $\mathcal{R} = \left\{ z : z > z_{\alpha} \right\}$
    - *making decision.* Reject H_0 if $Z_{obs} \in \mathcal{R}$; otherwise, not reject H_0.

    Z-test procedure.
  - When σ^2 is unknown:
    - *test statistic.* $T = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \sim t(n-1)$ under H_0 with s^2 is the sample variance.
    - *rejection region*, with pre-determined α,
      (i) when $H_1 : \mu \neq \mu_0$: $\mathcal{R} = \left\{ t : |t| > t_{\alpha/2}(n-1) \right\}$
      (ii) when $H_1 : \mu < \mu_0$: $\mathcal{R} = \left\{ t : t < -t_{\alpha}(n-1) \right\}$
      (iii) when $H_1 : \mu > \mu_0$: $\mathcal{R} = \left\{ t : t > t_{\alpha}(n-1) \right\}$
  - *making decision.* Reject H_0 if $T_{obs} \in \mathcal{R}$; otherwise, not reject H_0.

    T-test procedure.

What if $\mathbf{X} \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *with iid observations* $\mathbf{X}_1, \ldots, \mathbf{X}_n$*?* $\Longrightarrow$ **Chp**

Recall ... ... Consider r.v. $X \sim N(\mu, \sigma^2)$ with iid observations $X_1, \dots, X_n$.

- *Hypothesis Tests about* μ. e.g. $H_0 : \mu = \mu_0$ by *Likelihood Ratio Test*
  - When σ^2 is known:
    - *test statistic.* $R(\mu_0) = \frac{L(\mu_0, \sigma^2 | X_1, \dots, X_n)}{L(\hat{\mu}, \sigma^2 | X_1, \dots, X_n)}$ with $W = -2 \log R(\mu_0) \sim \chi^2(1)$ under H_0
    - *rejection region.* with pre-determined α, when $H_1 : \mu \neq \mu_0$: $\mathcal{R} = \left\{ w : w > \chi^2_\alpha \right\}$
    - *making decision.* Reject H_0 if $W_{obs} \in \mathcal{R}$; otherwise, not reject H_0.
  - When σ^2 is unknown:
    - *test statistic.* $R(\mu_0) = \frac{L(\mu_0, \hat{\sigma}_0^2 | X_1, \dots, X_n)}{L(\hat{\mu}, \hat{\sigma}^2 | X_1, \dots, X_n)}$ with $W = -2 \log R(\mu_0) \sim \chi^2(1)$ under H_0 approximately.
    - *rejection region*, with pre-determined α, when $H_1 : \mu \neq \mu_0$: $\mathcal{R} = \left\{ w : w > \chi^2_\alpha \right\}$
    - *making decision.* Reject H_0 if $W_{obs} \in \mathcal{R}$; otherwise, not reject H_0.

What if $\mathbf{X} \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *with iid observations* $\mathbf{X}_1, \dots, \mathbf{X}_n$*?* $\Longrightarrow$ **Chp 5.2b-3: Part II.2.2**

II.2.2 Hotelling's T^2 and Likelihood Ratio Test (Chp5.2b-3)

Suppose $\mathbf{X} \sim MN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with iid observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$

To test on $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs H_1 : otherwise when $\boldsymbol{\Sigma}$ is known

- *Test statistic.* Note that $\mathbf{Z} = \sqrt{n}\boldsymbol{\Sigma}^{-1/2}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0) \sim MN(\mathbf{0}, \mathbf{I})$ under H_0.
  Consider to use

$$W = \mathbf{Z}'\mathbf{Z} = n\left(\bar{\mathbf{X}} - \boldsymbol{\mu}_0\right)' \boldsymbol{\Sigma}^{-1}\left(\bar{\mathbf{X}} - \boldsymbol{\mu}_0\right) \sim \chi^2(p) \;\text{ under }\; H_0$$

- *Rejection region.* With pre-determined α,
  $\mathcal{R} = \left\{w : w > \chi^2_\alpha(p)\right\}$
- *Making decision.* Reject H_0 if $W_{obs} \in \mathcal{R}$; otherwise, not reject H_0.

Wald-test procedure.

Suppose $\mathbf{X} \sim MN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with iid observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$

To test on $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs H_1 : otherwise when $\boldsymbol{\Sigma}$ is unknown

- *Test statistic.* Is $\mathbf{T} = \sqrt{n}\mathbf{S}^{-1/2}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0) \sim MN(\mathbf{0}, \mathbf{I})$ under H_0?
  How about
  $W = \mathbf{T}'\mathbf{T} = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0) \sim \chi^2(p)$ under H_0?
  In fact, under H_0,
  $W = \mathbf{T}'\mathbf{T} = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0) \sim T^2(p, n-1)$

$T^2(p, n-1)$ is the *Hotelling's distribution*, the same as $\frac{(n-1)p}{n-p}F(p, n-p)$ with $F(p, n-p)$ the F-distribution with the two dfs p and $n-p$.

- *Rejection region.* With pre-determined α,
  $\mathcal{R} = \left\{w : w > T^2_\alpha(p, n-1)\right\}$
- *Making decision.* Reject H_0 if $W_{obs} \in \mathcal{R}$; otherwise, not reject H_0.

Hotelling's T^2-test procedure.

Example. (Textbook p214) Perspiration from 20 healthy females: three components X_1 = sweat rate, X_2 = sodium content, X_3 = potassium content. Test on $H_0 : \boldsymbol{\mu}^{'} = (4, 50, 10)$

II.2.2 Hotelling's T^2 and Likelihood Ratio Test (Chp5.2b-3)

Suppose $\mathbf{X} \sim MN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with iid observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$

To test on $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs H_1 : otherwise. When $\boldsymbol{\Sigma}$ is unknown by the *likelihood ratio test*

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\Big\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\Big\}$$

- *Test statistic.*

$$\Lambda = \frac{L(\boldsymbol{\mu}_0, \hat{\boldsymbol{\Sigma}}_0)}{L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})} = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{n/2}$$

  $\Lambda^{2/n}$ is called *Wilks's lambda* and

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

  under H_0.

- *Rejection region.* With pre-determined α, $\mathcal{R} = \{\lambda : \lambda < c\}$ with c determined by $P_{H_0}(\Lambda < c) = \alpha$.
- *Making decision.* Reject H_0 if $\Lambda_{obs} \in \mathcal{R}$; otherwise, not reject H_0.

What will we study next?

- *Part I. Introduction and Preparation*
- **Part II. Inference under Multivariate Normal Distribution (Chp 4-7)**
  - **II.1 Multivariate Normal Distribution (Chp 4)**
    - *II.1.1 Multivariate Normal Distribution $MN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (Chp 4.1-2)*
    - *II.1.2 Estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (Chp 4.3)*
    - *II.1.3 Properties of $\bar{\mathbf{X}}$ and $\mathbf{S}$ (Chp 4.4-5)*
    - *II.1.4 More on Normality (Chp 4.6-8)*
  - **II.2 Inferences on Mean Vector (Chp 5)**
  - **II.3 Comparisons of Several Mean Vector (Chp 6)**
  - *II.4 Multivariate Linear Regression (Chp 7)*