Lecture 2 Sept 27

Pop 1. $N(\mu_1, \sigma^2)$ Let $Y_1^1, \ldots, Y_{n_1}^1$ iid $N(\mu_1, \sigma^2)$

Pop 2. $N(\mu_2, \sigma^2)$ Let $Y_1^2 \ldots Y_{n_2}^2$ iid $N(\mu_2, \sigma^2)$

$$\delta = \mu_2 - \mu_1$$

$$\hat{\mu}_2 = \bar{Y}^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i^2$$

$$\hat{\mu}_1 = \bar{Y}^1 = \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i^1$$

Using MGF technique: $\bar{Y}^2 \sim N(\mu_2, \sigma^2/n_2)$

$\bar{Y}^1 \sim N(\mu_1, \sigma^2/n_1)$

$$\hat{\delta} = \bar{Y}^2 - \bar{Y}^1 \sim N\left(\overset{\delta}{\mu_2 - \mu_1}, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

Recall: If $Y_1^1, \ldots Y_{n_1}^1$ iid $N(\mu_1, \sigma^2)$

$\Rightarrow$ (1) $\bar{Y}^1 \sim N(\mu_1, \sigma^2/n_1)$

(2) $$\frac{(n_1-1) S_1^2}{\sigma^2} = \frac{\Sigma (Y_i^1 - \bar{Y}^1)^2}{\sigma^2} \sim \chi^2(n_1-1)$$

(3) S_1^2 and $\bar{Y}^1$ are statistically independent

$P(S_1^2 \in I$ and $\bar{Y}^1 \in A)$

$= P(S_1^2 \in I) \cdot P(\bar{Y}^1 \in A)$

To estimate the common pop'n variance σ^2,

we use the sample variances S_1^2, S_2^2

Let S_p^2 be the pooled variance estimator

$$S_p^2 = w_1 S_1^2 + w_2 S_2^2$$

s.t. $0 < w_1, w_2 < 1$ and $w_1 + w_2 = 1$

$$S_p^2 = \frac{(n_1-1)}{(n_1-1)+(n_2-1)} S_1^2 + \frac{(n_2-1)}{(n_1-1)+(n_2-1)} S_2^2$$

$$\Rightarrow \frac{(n_1+n_2-2)\, S_p^2}{\sigma^2} = \underbrace{\frac{(n_1-1) S_1^2}{\sigma^2}}_{\chi^2(n_1-1)} + \underbrace{\frac{(n_2-1) S_2^2}{\sigma^2}}_{\chi^2(n_2-1)}$$

indep

$$\Rightarrow \frac{(n_1+n_2-2)\, S_p^2}{\sigma^2} \sim \chi^2(df = n_1+n_2-2)$$

<u>So far:</u>

(A) $$\bar{Y}^1 - \bar{Y}^2 \sim N\Big(\overset{\delta}{\mu_1-\mu_2},\ \sigma^2\Big(\frac{1}{n_1}+\frac{1}{n_2}\Big)\Big)$$

$$\Rightarrow \frac{(\bar{Y}^1-\bar{Y}^2)-\delta}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0,1)$$

When $H_0: \delta = 0 \quad (\mu_1 = \mu_2)$

$$\frac{(\bar{Y}^1 - \bar{Y}^2)}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$$

(B) $$\frac{(n_1 + n_2 - 2)\, S_p^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$$

(C) (A) & (B) are independent

$(\bar{Y}^1, \bar{Y}^2) \qquad (S_p^2 : S_1^2, S_2^2)$

(D) Recall: If $Z \sim N(0, 1)$, $Q \sim \chi^2(g)$

Z and Q are independent

$$\Rightarrow \quad t = \frac{Z}{\sqrt{Q/g}} \sim \text{Student's } t \quad df = g$$

Applying (A) to (D):

$$\frac{\dfrac{\bar{y}^1 - \bar{y}^2}{\sqrt{\cancel{\sigma^2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}}{\sqrt{\dfrac{\cancel{(n_1+n_2-2)}\, S_p^2}{\cancel{\sigma^2}} \Big/ \cancel{(n_1+n_2-2)}}} \sim t(df = n_1+n_2-2)$$

$$\left\{ \frac{\bar{y}^1 - \bar{y}^2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t(df = n_1+n_2-2) \right.$$

similar form $\qquad \dfrac{\bar{y}^1 - \bar{y}^2}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

FORMAL HYPOTHESIS TESTING

$H_0: \ \delta = 0 \qquad (\mu_1 = \mu_2)$

$H_1: \ \delta \neq 0 \qquad (\mu_1 > \mu_2 \text{ OR } \mu_1 < \mu_2)$

Test Statistic

$$T = \frac{\bar{Y}^1 - \bar{Y}^2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Decision Rule

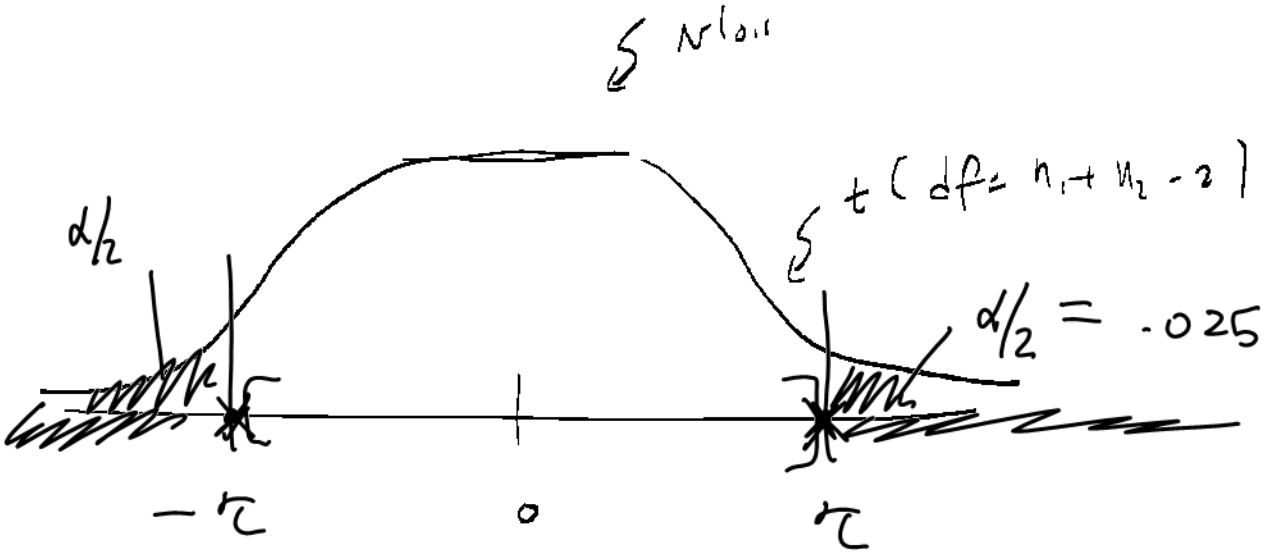
		Decision H_0	Decision H_1
TRUTH	H_0	No Error	Type I
	H_1	Type II	No Error

$$P(\text{Type I error}) = \alpha, \quad \alpha \text{ small number} \in (0,1)$$

usually 0.05.
0.01

When H_0 is true:

$$T \sim t(df = n_1 + n_2 - 2)$$



$\tau =$ 97.5th percentile of a $t(df = n_1 + n_2 - 2)$

$= t\left((1-\alpha/2)\times 100\%,\ n_1 + n_2 - 2\right)$

If n_1 & n_2 large (say ≥ 50)

then $\tau \approx 1.96$

Given data $y^1_1, \ldots, y^1_{n_1} \Rightarrow \bar{y}^1, s_1^2$

$y^2_1 \ldots y^2_{n_2} \Rightarrow \bar{y}^1, s_2^2$

$$t_{obs} = \frac{\bar{y}^1 - \bar{y}^2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

Reject H_0 if $|t_{obs}| > c$.

Linear Models

$$Y_i^1 \sim N(\mu_1, \sigma^2)$$

$$Y_i^1 = \mu_1 + \varepsilon_i^1 ; \quad \varepsilon_i^1 \sim N(0, \sigma^2)$$

$$\begin{bmatrix} Y_1^1 \\ Y_2^1 \\ \vdots \\ Y_{n_1}^1 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \vdots \\ \mu_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1^1 \\ \vdots \\ \varepsilon_{n_1}^1 \end{bmatrix}$$

Similarly: $Y_i^2 \overset{iid}{\sim} N(\mu_2, \sigma^2)$

$$Y_i^2 = \mu_2 + \varepsilon_i^2, \quad \varepsilon_i^2 \sim N(0,\sigma^2)$$

$$\begin{bmatrix} Y_1^2 \\ \vdots \\ Y_{n_2}^2 \end{bmatrix} = \begin{bmatrix} \mu_2 \\ \vdots \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1^2 \\ \vdots \\ \varepsilon_{n_2}^2 \end{bmatrix}$$

Let $\underline{Y} = \begin{bmatrix} Y_1^1 \\ \vdots \\ Y_{n_1}^1 \\ \hline Y_1^2 \\ \vdots \\ Y_{n_2}^2 \end{bmatrix}_{N\times 1}$, $\underline{\varepsilon} = \begin{bmatrix} \varepsilon_1^1 \\ \vdots \\ \varepsilon_{n_1}^1 \\ \hline \varepsilon_1^2 \\ \vdots \\ \varepsilon_{n_2}^2 \end{bmatrix}_{N\times 1}$.

$N = n_1 + n_2$

$$\underline{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$\underset{N\times 1}{\underline{U}} = \underbrace{\begin{bmatrix} \mathbb{1}_{n_1} & \underline{0}_{n_1} \\ \underline{0}_{n_2} & \mathbb{1}_{n_2} \end{bmatrix}}_{X} \underset{\beta\ 2\times 1}{\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}} + \underset{N\times 1}{\underline{\varepsilon}},$$

X design matrix

$$\underline{\varepsilon} \sim N(\underline{0}\ \underset{N\times N}{I} \otimes \sigma^2)$$

$$\operatorname{cov}(\underline{\varepsilon}) = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}$$

The linear model:

$$\underline{U} = X \underline{\beta} + \underline{\varepsilon}, \qquad \underline{\varepsilon} \sim N(\underline{0}, I \otimes \sigma^2)$$

X: design matrix

$\underline{\beta}$: parameter vector

Goal: Estimate β

Inference on $\underline{\beta}$, $\quad \underline{c}\,\underline{\beta}$

$$(-1\ \ 1)\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$\mu_2 - \mu_1$$

Alternative Parameterization:

$$Y_i^1 = \mu_1 + \varepsilon_i^1 \qquad \delta = \mu_2 - \mu_1$$

$$Y_i^2 = \underbrace{\mu_2}_{\mu_1 + \delta} + \varepsilon_i^2$$

$$\underline{Y} = \begin{bmatrix} Y_1^1 \\ \vdots \\ Y_{n_1}^1 \\ \hline Y_1^2 \\ \vdots \\ Y_{n_2}^2 \end{bmatrix} = \underbrace{\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_1 \\ \hline \mu_1 + \delta \\ \vdots \\ \mu_1 + \delta \end{bmatrix}}_{\left[\begin{array}{c|c} \underline{1}_{n_1} & 0_{n_1} \\ \hline \underline{1}_{n_2} & \underline{1}_{n_2} \end{array}\right] \begin{pmatrix} \mu_1 \\ \delta \end{pmatrix}} + \underline{\varepsilon}$$