Тестирање на основу 2 независна узорка

1. - $\mathcal{X}=(X_1, X_2, \ldots, X_{n_1})$ ПСУ из $\mathcal{N}(m_1, \sigma_1^2)$
   - $\mathcal{Y}=(Y_1, Y_2, \ldots, Y_{n_2})$ ПСУ из $\mathcal{N}(m_2, \sigma_2^2)$
   - $\mathcal{X}$ и $\mathcal{Y}$ независни

Извршићемо тестирање:

$$H_0: m_1 - m_2 = m_0 \quad vs \quad H_1: m_1 - m_2 \overset{>}{\underset{<}{\neq}} m_0$$

Покушајмо да нађемо неку тест статистику за овакво тестирање. Нека је нпр. $H_1: m_1 - m_2 > m_0$

1. корак: - да видимо кад ћемо на основу ова два узорка поверовати да је $m_1 - m_2 > m_0$
   - Знамо ли неку оцену за $m_1 - m_2$? Да!
   - $\bar{X}_{n_1}$ је непристрасна оцена за m_1 и $\bar{Y}_{n_2}$ је непристрасна оцена за m_2, па је $E(\bar{X}_{n_1} - \bar{Y}_{n_2}) = E\bar{X}_{n_1} - E\bar{Y}_{n_2} = m_1 - m_2$.

   $\Rightarrow \bar{X}_{n_1} - \bar{Y}_{n_2}$ је непристрасна оцена за $m_1 - m_2$

Што је $\bar{x}_{n_1} - \bar{y}_{n_2}$ веће од m_0, расте наше уверење да је $m_1 - m_2 > m_0$, па би критична област могла бити: $W = \{\bar{X}_{n_1} - \bar{Y}_{n_2} - m_0 > c_1\}$

2. корак: - запишимо W у облику: $\{T > c_1\}$, где је T статистика коју ћемо користити за тестирање. T треба да буде т.д. знамо да јој одредимо расподелу под H_0.

Сада разликујемо случајеве када су нам σ_1^2 и σ_2^2 познати и када нису.

1.1. σ_1^2 и σ_2^2 познати

Показали смо раније да: $$\frac{\bar{X}_{n_1}-\bar{Y}_{n_2}-(m_1-m_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0,1) \quad (*)$$

Запишимо обл. W овако:

$$W=\left\{\bar{X}_{n_1}-\bar{Y}_{n_2}-m_0>c_1\right\}=\left\{\frac{\bar{X}_{n_1}-\bar{Y}_{n_2}-m_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}>\frac{c_1}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right\}$$

Да не бисмо вукли $\frac{c_1}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$, означићемо га са c,

а видимо и да ћемо за T бирати $\frac{\bar{X}_{n_1}-\bar{Y}_{n_2}-m_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$,

јер је под H_0 m_0 права вредност разлике m_1 и m_2, па ће по $(*)$ T тада имати познату расподелу ($\mathcal{N}(0,1)$)

Дакле, $$\boxed{T=\frac{\bar{X}_{n_1}-\bar{Y}_{n_2}-m_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}}, \quad \boxed{W=\{T>c\}}$$

За $H_1: m_1-m_2<m_0 \rightarrow W=\{T<c\}$

За $H_2: m_1-m_2\neq m_0 \rightarrow W=\{|T|>c\}$

1.2. σ_1^2 и σ_2^2 непознати

Сада ћемо извршити тестирање о једнакости σ_1^2 и σ_2^2. У зависности од тога да ли су оне једнаке или не разликујемо тест статистику.

Дакле, тестираћемо:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad vs \quad H_1: \sigma_1^2 \neq \sigma_2^2$$

Запажање: H_0 и H_1 су комплементарне хипотезе, тј. оне покривају све могуће скупове вредности σ_1^2 и σ_2^2, па ако не прихватимо H_1, можемо рећи да прихватамо H_0.

- Како ово тестирати? $\tilde{S}_{n_1}^2$ и $\tilde{S}_{n_2}^2$ су непристрасне оцене за σ_1^2 и σ_2^2, па бисмо очекивали да, у случају да је H_0 тачна, $\frac{\tilde{S}_{n_1}^2}{\tilde{S}_{n_2}^2}$ буде блиско 1.

Дакле, што је овај количник даљи од 1, то смо више сигурни у H_1, односно:

- што је $\frac{\tilde{S}_{n_1}^2}{\tilde{S}_{n_2}^2}$ ближе 0, све више верујемо да је σ_2^2 веће,
а што је $\frac{\tilde{S}_{n_1}^2}{\tilde{S}_{n_2}^2}$ веће од 1, верујемо да је σ_1^2 веће

Зато ћемо W потражити у облику:

$$W = \left\{0 < \frac{\tilde{S}_{n_1}^2}{\tilde{S}_{n_2}^2} < c_1\right\} \cup \left\{\frac{\tilde{S}_{n_1}^2}{\tilde{S}_{n_2}^2} > c_2\right\}$$

ово не морамо писати јер је $\frac{\tilde{S}_{n_1}^2}{\tilde{S}_{n_2}^2}$ сигурно позитивна

За T можемо узети $\frac{\tilde{S}_{n_1}^2}{\tilde{S}_{n_2}^2}$, јер знамо њену расподелу под H_0.

Ево и зашто: при H_0 је $\sigma_1^2 = \sigma_2^2$, односно $\frac{\sigma_1^2}{\sigma_2^2} = 1$.

Зато је тада

$$T = \frac{\tilde{S}_{n_1}^2}{\tilde{S}_{n_2}^2} = \frac{\tilde{S}_{n_1}^2}{\tilde{S}_{n_2}^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} = \frac{\frac{\tilde{S}_{n_1}^2}{\sigma_1^2}}{\frac{\tilde{S}_{n_2}^2}{\sigma_2^2}} = \frac{\frac{X}{n_1-1}}{\frac{Y}{n_2-1}},$$

где је $X = (n_1-1)\frac{\tilde{S}_{n_1}^2}{\sigma_1^2}$, а $Y = (n_2-1)\frac{\tilde{S}_{n_2}^2}{\sigma_2^2}$.

Како су X и Y независне, а показали смо раније и да $X \sim \chi^2_{n_1-1}$ и $Y \sim \chi^2_{n_2-1}$, следи: $\boxed{T \sim \mathcal{F}_{n_1-1, n_2-1}}$

Дакле, за тестирање $H_0: \sigma_1^2 = \sigma_2^2$ vs $H_1: \sigma_1^2 \neq \sigma_2^2$ користимо:

$$T = \frac{\tilde{S}_{n_1}^2}{\tilde{S}_{n_2}^2}, \quad W = \{T < c_1\} \cup \{T > c_2\}$$

c_1 и c_2 ћемо као и раније бирати т.д.

$$P_{H_0}\{T < c_1\} = \frac{\alpha}{2} = P_{H_0}\{T > c_2\},$$ где ће α бити неки нови ниво значајности који ћемо сами бирати

- Ако прихватимо да је $\sigma_1^2 = \sigma_2^2$, тест статистика за тестирање $H_0: m_1 - m_2 = m_0$ vs $H_1: m_1 - m_2 > m_0$ је:

$$\boxed{T = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - m_0}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}$$ и при H_0 $\quad T \sim t_{n_1+n_2-2}$

S^2 је оцена за σ_1^2, односно σ_2^2 и тада:

$$S^2 = \frac{(n_1-1)\tilde{S}_{n_1}^2 + (n_2-1)\tilde{S}_{n_2}^2}{n_1+n_2-2}$$

Критична област је $W=\{T>c\}$, а за друге случајеве:

$H_1: \; m_1-m_2 < m_0 \to W=\{T<c\}$

$H_2: \; m_1-m_2 \neq m_0 \to W=\{|T|>c\}$

Ако одбацимо једнакост дисперзија, тј. прихватимо: $\sigma_1^2 \neq \sigma_2^2$, тест статистика за тестирање

$H_0: m_1 - m_2 = m_0$ vs $H_1: m_1-m_2 > m_0$ је

$$\boxed{T = \frac{\bar{X}_{n} - \bar{Y}_{n} - m_0}{\sqrt{\frac{\tilde{S}_{n_1}^2}{n_1} + \frac{\tilde{S}_{n_2}^2}{n_2}}}}$$ и она при H_0 има t_ν расподелу

ово ће бити дато

Критична област: $W=\{T>c\}$, а у другим случајевима:

$H_1: \; m_1-m_2 \neq m_0 \to W=\{|T|>c\}$

$H_1: \; m_1-m_2 < m_0 \to W=\{T<c\}$

Видети 40. задатак у збирци. (у задатку је $m_0=0$)

---

2. $X = (X_1, X_2, \dots, X_{n_1}) \sim Ber(p_1)$

$Y = (Y_1, Y_2, \dots, Y_{n_2}) \sim Ber(p_2)$

- X и Y независни, n_1 и n_2 велики

Тестирајмо сада:

$$H_0: p_1 - p_2 = p_0 \quad vs \quad H_1: p_1 - p_2 \gtrless p_0$$

Нека је $H_1: p_1 - p_2 > p_0$, остале случајеве ћемо само прокоментарисати.

Слично као у претходном разматрању, како су $\bar{X}_{n_1}$ и $\bar{Y}_{n_2}$ непристрасне оцене за p_1 и p_2, редом,
$\bar{X}_{n_1} - \bar{Y}_{n_2}$ је непристрасна оцена за $p_1 - p_2$,
па у прилог H_1 иду вредности $\bar{x}_{n_1} - \bar{y}_{n_2}$ које су веће од p_0.

Зато ћемо W потражити у облику:

$$W = \{ \bar{X}_{n_1} - \bar{Y}_{n_2} - p_0 > c_1 \}$$

и покушати као и до сада да је сведемо на

$$W = \{ T > c \}$$

где ће T бити тест статистика.

- Шта може бити T? Подсетимо се следећег:

$\sum_{i=1}^{n_1} X_i$ се за велико n_1 може апроксимирати нормалном расподелом (чије параметре можемо наслутити), што су показали Муавр и Лаплас, док с друге стране тај резултат можемо видети и као посебан случај ЦГТ.

Дакле, за велико n_1: $\sum_{i=1}^{n_1} X_i \sim \mathcal{N}(n_1 p_1, n_1 p_1 (1-p_1))$

Одатле добијамо: $\bar{X}_{n_1} \sim \mathcal{N}(p_1, \frac{p_1}{n_1}(1-p_1))$

Слично: $\bar{Y}_{n_2} \sim \mathcal{N}(p_2, \frac{p_2}{n_2}(1-p_2))$

$\bar{X}_{n_1}$ и $\bar{Y}_{n_2}$ су независне, па важи:

$$\bar{X}_{n_1} - \bar{Y}_{n_2} \sim \mathcal{N}\left(p_1 - p_2, \frac{p_1}{n_1}(1-p_1) + \frac{p_2}{n_2}(1-p_2)\right)$$

$$\Rightarrow Z = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - (p_1 - p_2)}{\sqrt{\frac{p_1}{n_1}(1-p_1) + \frac{p_2}{n_2}(1-p_2)}} \sim \mathcal{N}(0,1)$$

стандардизација

Како за велике n_1 и n_2 по ЗВБ: $\bar{X}_{n_1} \approx p_1$ и $\bar{Y}_{n_2} \approx p_2$, следећа сл. вел. има приближно $\mathcal{N}(0,1)$ расподелу:

$$\tilde{Z} = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - (p_1 - p_2)}{\sqrt{\frac{\bar{X}_{n_1}}{n_1}(1-\bar{X}_{n_1}) + \frac{\bar{Y}_{n_2}}{n_2}(1-\bar{Y}_{n_2})}}$$

Сада видимо да за T можемо дирати:

$$\boxed{T = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - p_0}{\sqrt{\frac{\bar{X}_{n_1}}{n_1}(1-\bar{X}_{n_1}) + \frac{\bar{Y}_{n_2}}{n_2}(1-\bar{Y}_{n_2})}}},$$

јер је при H_0 p_0 права вредност за $p_1 - p_2$, па тада $T \sim \mathcal{N}(0,1)$, што смо управо показали.

Критична област: $W = \{T > c\}$.

За остале случајеве важи:

$H_1: p_1 - p_2 < p_0 \rightarrow W = \{T < c\}$

$H_1: p_1 - p_2 \neq p_0 \rightarrow W = \{|T| > c\}$

Видети 44. задатак. (У том задатку је $p_0=0$)

У специјалном случају, када је $p_0=0$, можемо користити и другу тест статистику, док ће критична област бити истог облика.

Како је Z ода када је $p_1=p_2=p$ једнака:

$$Z=\frac{\bar{X}_{n_1}-\bar{Y}_{n_2}}{\sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}},$$

а p се може оценити као из оба узорка као:

$$\hat{p}=\frac{n_1\bar{X}_{n_1}+n_2\bar{Y}_{n_2}}{n_1+n_2},$$

За тест статистику можемо узети и

$$\boxed{T=\frac{\bar{X}_{n_1}-\bar{Y}_{n_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}}$$