TEST DIFFERENZA DI PROPORZIONI

Supponiamo di avere due campioni $X_1, \dots, X_{n_1} \sim Be(p_X)$ e $Y_1, \dots, Y_{n_2} \sim Be(p_Y)$ ind.; abbiamo le stime

$$\hat{p}_X = \frac{1}{n_1} \sum_1^{n_1} {}_i X_i = \frac{m_1}{n_1}, \quad \text{con } m_1 := \sum_1^{n_1} {}_i X_i$$

$$\hat{p}_Y = \frac{m_2}{n_2}, \quad \text{con } m_2 = \sum_1^{n_2} {}_i Y_i$$

Vogliamo fare un test di ipotesi $H_0: \mu_X = \mu_Y$ e $H_1: \mu_X \neq \mu_Y$.
Se n_1 e n_2 sono "grandi", allora $\hat{p}_X \approx N\left(p_X, \frac{p_X(1-p_X)}{n_1}\right)$ e $\hat{p}_Y \approx N\left(p_Y, \frac{p_Y(1-p_Y)}{n_2}\right)$, e $\hat{p}_X$ e $\hat{p}_Y$ sono ind., quindi

$$\hat{p}_Y - \hat{p}_X \approx N\left(p_Y - p_X, \frac{p_X(1-p_X)}{n_1} + \frac{p_Y(1-p_Y)}{n_2}\right)$$

Sotto H_0, $p_X = p_Y = p$, quindi

$$\hat{p}_Y - \hat{p}_X \approx N\left(0, p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

La dev. st. di $\hat{p}_Y - \hat{p}_X$ può essere stimata da

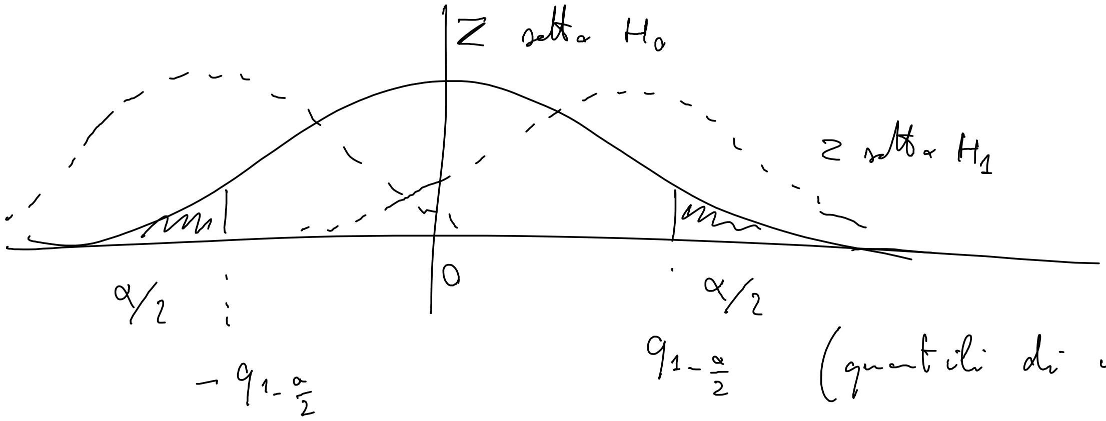
$$S_{\hat{p}_Y - \hat{p}_X} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

<u>errore standard della differenza di proporzioni</u>, con

errore standard della differenza di proporzioni, con

$$\hat{p} = \frac{m_1 + m_2}{n_1 + n_2} = \frac{\sum_1^{n_1} X_i + \sum_1^{n_2} Y_i}{n_1 + n_2}$$

$$Z = \frac{\hat{p}_Y - \hat{p}_X}{s_{\hat{p}_Y - \hat{p}_X}} \approx N(0,1) \text{ sotto } H_0$$



(quantili di una $N(0,1)$)

La logica è identica al test di Student. Regola:

- se $|Z| < q_{1-\frac{\alpha}{2}}$, accetto H_0
- se $|Z| > q_{1-\frac{\alpha}{2}}$, rifiuto H_0 e accetto H_1

esempio: alotano contro morfina

alotano: $8 = m_1$ morti su $61 = n_1$

morfina: $10 = m_2$ morti su $67 = n_2$

$$\hat{p}_A = \frac{8}{61} = 0{,}131 \qquad \hat{p}_M = \frac{m_2}{n_2} = \frac{10}{67} = 0{,}149$$

Vogliamo fare un test di ipotesi $H_0: p_A = p_M$ e $H_1: p_A \neq p_M$

$$\hat{p} = \frac{8+10}{61+67} = \frac{18}{128} = 0,141$$

$$S_{\hat{p}_A - \hat{p}_M} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{0,141 \times (1-0,141)\left(\frac{1}{61}+\frac{1}{67}\right)}$$

$$= 0,062$$

$$n_1 \hat{p}_A = n_1 \cdot \frac{m_1}{n_1} = \cancel{61} \times \frac{8}{\cancel{61}} = 8 > 5$$

$$n_1(1-\hat{p}_A) = 61\left(1 - \frac{8}{61}\right) = \cancel{61} \times \frac{61-8}{\cancel{61}} = 53 > 5$$

oppure: $\hat{p}_A < 0,5$, quindi $1 - \hat{p}_A > 0,5 > \hat{p}_A$

$$n_1(1-\hat{p}_A) > n_1 \hat{p}_A > 5$$

$n_2 \hat{p}_M = \cancel{67} \times \frac{10}{\cancel{67}} = 10 > 5$; siccome $\hat{p}_M < 0,5$,

$n_2(1-\hat{p}_M) > n_2 \hat{p}_M > 5 \Rightarrow$ appr. normale

$$Z = \frac{\hat{p}_A - \hat{p}_M}{S_{\hat{p}_A - \hat{p}_M}} = \frac{0,131 - 0,149}{0,062} = -0,30$$

se ad esempio $\alpha = 0,05$, allora il val. critico è $q_{1-\frac{\alpha}{2}} =$

$= q_{0,975} = 1,96 > |Z| \Rightarrow$ accetto H_0

cerchiamo il valore P, tale che $q_{1-\frac{P}{2}} = |Z| = 0,30$

$$1 - \frac{P}{2} = F_Z(0,30) = 0,61791$$

$$\frac{p}{2} = 1 - 0,61791 = 0,38209$$

$$p = 2 \times 0,38209 \simeq 0,764$$

Si possono anche calcolare intervalli di confidenza per la differenza di proporzioni, ed eventualmente usarli per test di ipotesi.

Cerchiamo $I_{1-\alpha}$ tale che $\mathbb{P}\{p_X - p_Y \in I_{1-\alpha}\} = 1-\alpha$

Con gli stessi passaggi del caso gaussiano, arriviamo a dire che $I_{1-\alpha}$ ha estremi $\hat{p}_X - \hat{p}_Y \pm q_{1-\frac{\alpha}{2}} S_{\hat{p}_X - \hat{p}_Y}$.

Se poi vogliamo effettuare un test con $H_0: p_X = p_Y$ e $H_1: p_X \neq p_Y$ la regola è:

- se $0 \in I_{1-\alpha}$, accetto H_0
- se $0 \notin I_{1-\alpha}$, rifiuto H_0 e accetto H_1