Lezione 3

La verifica delle ipotesi:
principi generali

Argomenti della lezione:

- Ipotesi statistiche
- Ipotesi sulla media

Indicatore campionario: $\overline{X}$

Il campione è stato estratto da una popolazione con parametro μ?

Procedura di controllo

Distanza tra valore ottenuto e valore atteso

- **Può essere attribuita al caso?**
- **L'indicatore ottenuto proviene da una popolazione la cui media è uguale a quella di riferimento? Es: $\mu = \mu_0$**

Ipotesi sulla media:

Ipotesi nulla H0: $\mu=\mu_0$

Ipotesi alternativa H1: $\mu\neq\mu_0$

Media della popolazione (μ) rispetto a una media di riferimento (μ_0)

- **L'ipotesi nulla è vera ⇒ distanza attribuibile a variazioni casuali**
- **L'ipotesi nulla è falsa ⇒ distanza attribuibile a una causa e non al caso. Potrebbe essere vera una ipotesi alternativa (H1)**

Decisione

- si assume come vera l'ipotesi nulla
- si fa riferimento a una distribuzione costruita sulla base di H_0
- si definisce un valore critico che, con una certa probabilità, consente di accettare o respingere H_0

Valore critico

Dato un livello di probabilità α il valore critico separa

i risultati poco probabili (respingere H_0)

dai risultati più probabili (accettare H_0)

Decisione di tipo probabilistico

- si può accettare H0 quando è falsa o non accettarla quando è vera

Decisione che comporta un rischio

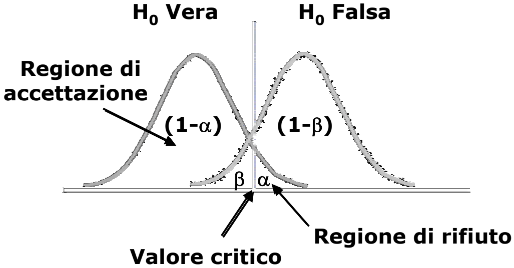
Il rischio di rifiutare H_0 quando è vera è pari ad α

$\alpha = .05 \Rightarrow$
si decide di correre il rischio di sbagliare nel rifiutare H_0 con una probabilità di 0.05

- rifiutare H0 quando è vera $\Rightarrow$ errore di I tipo
- accettare H0 quando è falsa $\Rightarrow$ errore di II tipo

Inversamente proporzionali

	H_0 vera	H_0 falsa
accetto H_0	Decisione corretta ($p=1-\alpha$)	Errore di II tipo ($p=\beta$)
rifiuto H_0	Errore di I tipo ($p=\alpha$)	Decisione corretta ($p=1-\beta$)



Modi di formulare le ipotesi

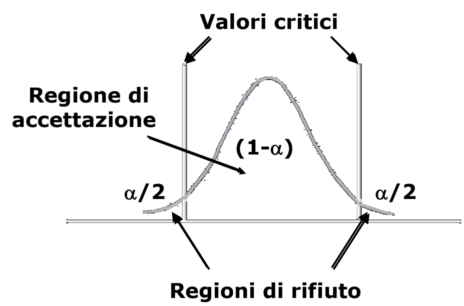
H_0 (ipotesi nulla): è sempre di uguaglianza

$\mu = \mu_0$

H_1 (ipotesi alternativa) è sempre di differenza:

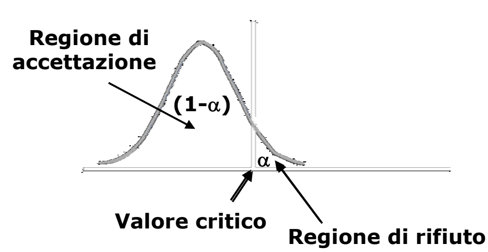
bidirezionale (≠) se è differenza senza direzione

$\mu \neq \mu_0$



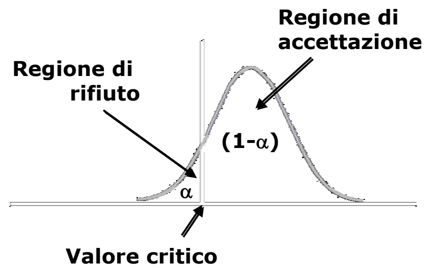
Monodirezionale destra (>) si ipotizza una differenza positiva

$\mu > \mu_0$



Monodirezionale sinistra (<) si ipotizza una differenza negativa

$\mu < \mu_0$



Ipotesi sulla media caso di un campione

Confronto tra media della popolazione da cui è estratto il campione (μ) e media di una popolazione di confronto (μ_0)

Esempio:

Variabile: test di abilità manuale
Popolazione: giovani di leva
Media della popolazione: 100

Campione di soggetti che hanno seguito un corso di formazione:

$n = 80$; $\bar{x} = 105.62$; $s = 20$

Ipotesi di ricerca:
i soggetti del campione sono più abili di quelli della popolazione

Procedura di verifica delle ipotesi si utilizzano le proprietà della Distribuzione Campionaria della Media (DCM)

ricordando che $\mu_{\bar{x}}=\mu$ e $\sigma_{\bar{x}}=\sigma/\sqrt{n}$

Formulazione delle ipotesi

$H0: \mu_{\bar{x}} = \mu_0$

La media della popolazione da cui è estratto il campione ($\mu_{\bar{x}}$), è uguale a quella della popolazione di riferimento (μ_0)

Formulazione delle ipotesi

H1: $\mu_{\bar{x}} > \mu_0$

La media della popolazione da cui è estratto il campione è maggiore di quella della popolazione di riferimento (ipotesi alternativa monodirezionale destra)

Scelta del "livello critico": definizione di un rischio accettabile di commettere l'errore del I tipo

$\alpha = 0.05$

Definizione dei parametri della DCM

Se è vera l'ipotesi nulla, il campione appartiene alla distribuzione campionaria delle medie che ha parametri:

$\mu_{\bar{x}} = \mu_0 = 100$

$\hat{\sigma}_{\bar{x}} = s / \sqrt{n-1} = 20/\sqrt{80-1} = 2.25$

Definizione di z_c per $\alpha = 0.05$, e ipotesi monodirezionale destra

$z_c = +1.64$

Regola decisionale

Se $z \geq$ di z_c rifiuto H0

Se $z < z_c$ accetto H0

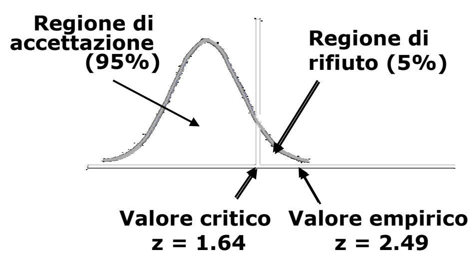
Calcolo di z

$$z = \frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{105.62-100}{2.25} = \frac{5.62}{2.25} = 2.49$$

Decisione

Poiché $z > z_c$ rifiuto H0.
Il corso di addestramento ha un effetto positivo sull'abilità manuale

$\alpha = .05$



Verifica delle ipotesi sulla media della popolazione quando σ non è noto e n≤ 30

La distribuzione campionaria della media non ha forma normale. Si fa riferimento alla distribuzione "t di Student" con (n-1) gdl

Invece di calcolare z, si calcola

$$t = \frac{\bar{X}-\mu_{\bar{x}}}{\hat{\sigma}_{\bar{x}}} \quad \text{dove} \quad \hat{\sigma}_{\bar{x}} = s/\sqrt{(n-1)}$$

Tavola della t di Student

Ipotesi alternativa monodirezionale o bidirezionale

Variabile: livello di apprendimento
Popolazione: studenti di I media
Media della popolazione: 20

Campione di soggetti che partecipa ad un corso sperimentale:

$n = 17;\ \bar{x} = 21.52;\ s = 5.2$

Formulazione delle ipotesi

H0: $\mu_{\bar{x}} = \mu_0 = 20$

La media della popolazione da cui è estratto il campione ($\mu_{\bar{x}}$), è uguale a quella della popolazione di riferimento (μ_0)

Formulazione delle ipotesi

H1: $\mu_{\bar{x}} > \mu_0$

La media della popolazione da cui è estratto il campione è maggiore di quella della popolazione di riferimento (ipotesi alternativa monodirezionale destra)

Scelta del "livello critico", cioè definizione del rischio che si accetta di correre di commettere errore di I tipo

$\alpha = 0.05$

Definizione dei parametri della DCM

Se è vera l'ipotesi nulla, il campione appartiene alla distribuzione campionarie delle medie che ha parametri:

$$\mu_{\bar{x}} = \mu = 20$$

$$\hat{\sigma}_{\bar{x}} = s / \sqrt{n-1} = 5.2/\sqrt{17-1} = 1.3$$

Definizione di t_c per $\alpha = 0.05$, 16 gdl e ipotesi monodirezionale destra

$$t_c = +1.746$$

Regola decisionale

Se $t \geq$ di t_c rifiuto H0
Se $t < t_c$ accetto H0

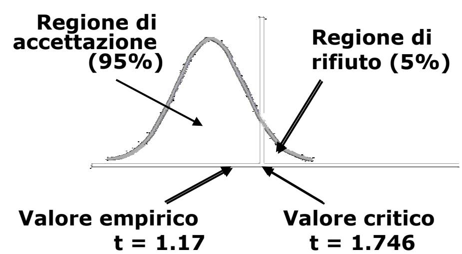
Calcolo di t

$$t = \frac{\bar{X}-\mu_{\bar{x}}}{\hat{\sigma}_{\bar{x}}} = \frac{21.52-20}{1.3} = \frac{1.52}{1.3} = 1.17$$

Decisione

Poiché $t < t_c$ accetto H0
il corso sperimentale non sembra influire in modo significativo sul livello di apprendimento

$\alpha = .05$



CONCLUSIONE

- ➔ **Ipotesi statistiche sulla media**
- ➔ **Processo di decisione**
- ➔ **Distribuzione normale e t di Student**