UNIVERSITÀ
DI SIENA
1240

Precorso di Statistica per le Lauree Magistrali

Gianni Betti

Lezione 4 Ottobre 2023 - Ore 10-12

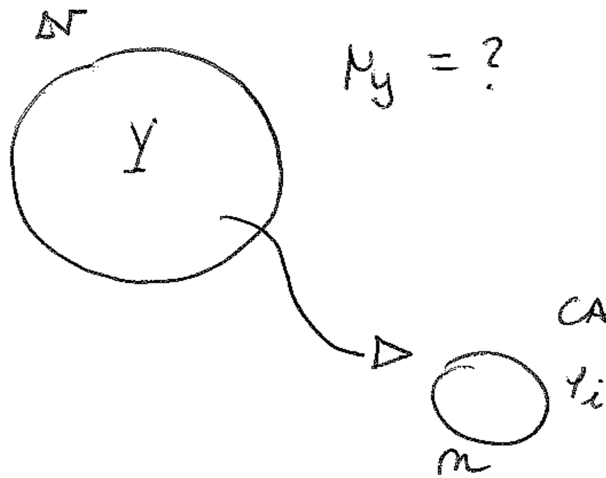
Intervalli di confidenza per la media della popolazione (3.3 Stock & Watson)

- Nelle lezioni precedenti è stata definita la stima della media di una popolazione (**paragrafo 3.1**) e la verifica di ipotesi circa la media della popolazione (**paragrafo 3.2**).
- Il concetto fondamentale che ne è emerso consiste nel fatto che – a causa degli errori campionari – è impossibile inferire l'esatto valore medio di Y nella popolazione usando solo l'informazione nel campione.

Ricapitolando:

AVETE VISTO CHE

$\mu_y = ?$



CAMPIONE

$$\bar{Y} = \frac{\sum y_i}{n}$$

$$E(\bar{Y}) = \mu_y$$

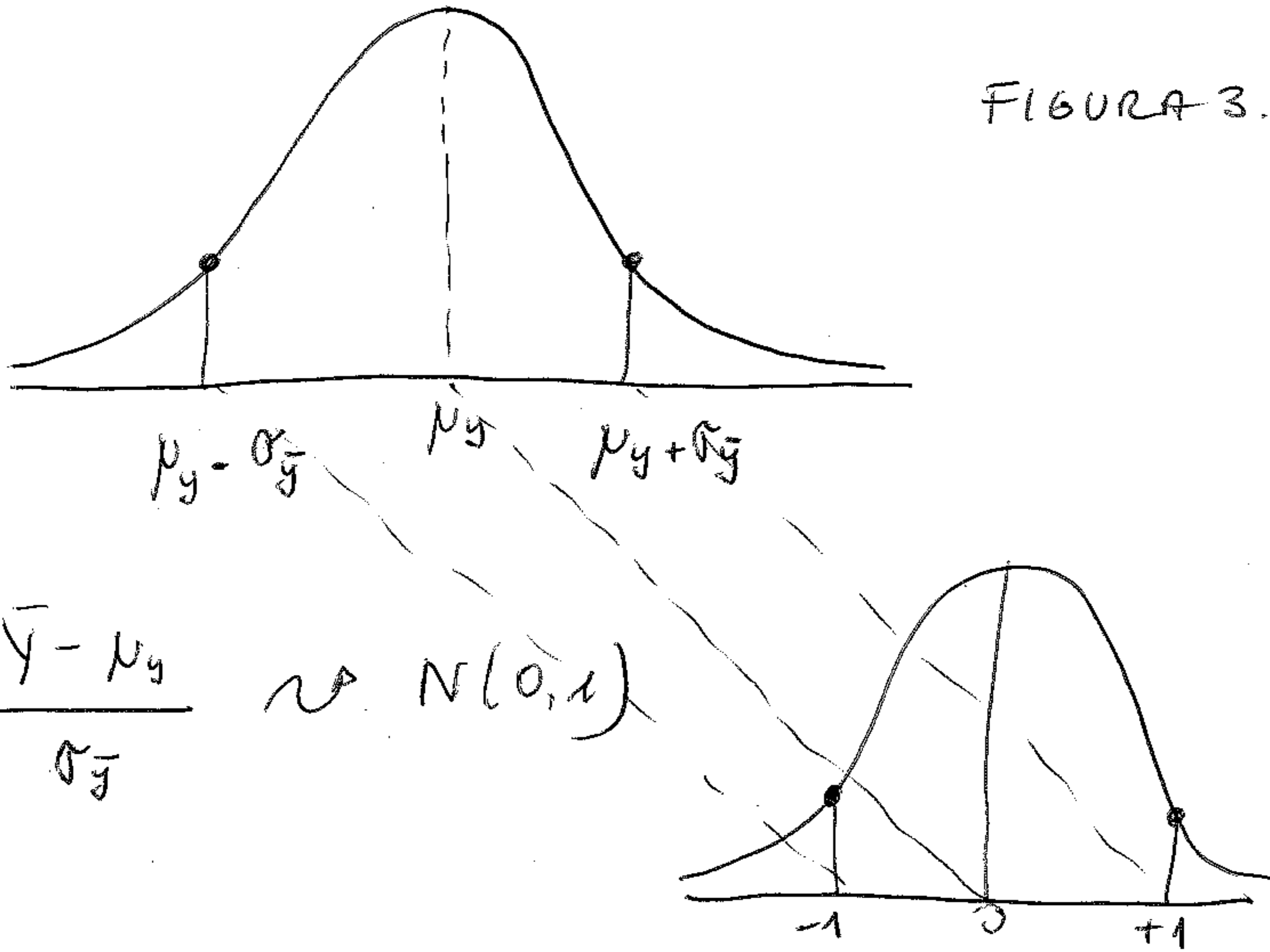
$$V(\bar{Y}) = \sigma_y^2 / n$$

SE n è grande la distribuzione di $\bar{Y}$ è Normale

$$\bar{Y} \sim N\left(\mu_y , \sigma^2_{\bar{y}} = \sigma^2_y / n\right)$$

… se standardizziamo la normale:

FIGURA 3.1



$$\frac{\bar{Y} - \mu_y}{\sigma_{\bar{y}}} \sim N(0,1)$$

CASO 1: varianza popolazione nota (p.57)

quindi se σ_y^2 è nota test su μ_y:

$H_0: \mu_y = \mu_{y,0}$

$$\frac{\bar{y} - \mu_{y,0}}{\sigma_{\bar{y}}^2} \sim N(0,1)$$

TUTTO è NOTO

CASO 2: varianza popolazione ignota (p.59)

se σ^2_y NON è noto ALLORA VIENE STIMATO

$$s^2_y = \frac{1}{n-1} \sum (y_i - \bar{y})^2 \qquad (3.7)$$

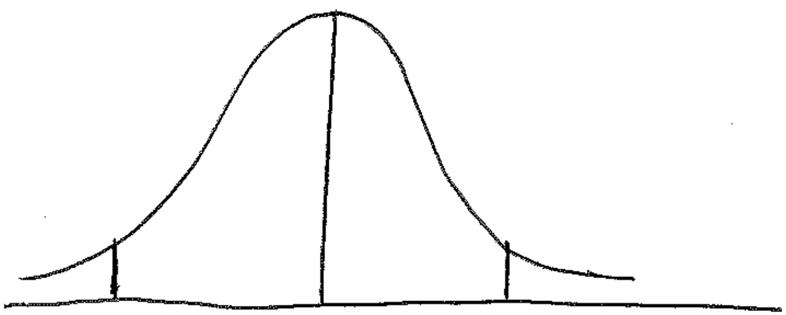
$$\frac{\bar{y} - \mu_{y,0}}{SE(\bar{y}) = s_y / \sqrt{n}}$$

$$\sigma_{\bar{y}} = \sigma_y / \sqrt{n}$$

p. 59

CASO 2: varianza popolazione ignota (p.59)

questo non si distribuisce più come $N(0,1)$ ma come t di Student



μ_y

$$Pr\left\{ \left| \frac{\bar{Y} - \mu_y}{SE(\bar{Y})} \right| \le 1.96 \right\} = 95\%$$

$$\Downarrow$$

$$\left| \bar{Y} - \mu_y \right| = 1.96\ SE(\bar{Y})$$

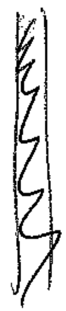
$$\mu_{\bar{y}} = \left\{ \bar{Y} \pm 1{,}96\ SE(\bar{Y}) \right\}$$

PIÙ STRETTO 1,64

PIÙ LARGO 2,58

CONCETTO CHIAVE 3.7

P. 63



95%

90%

99%

Intervalli di confidenza per la media della popolazione (3.3 Stock & Watson)

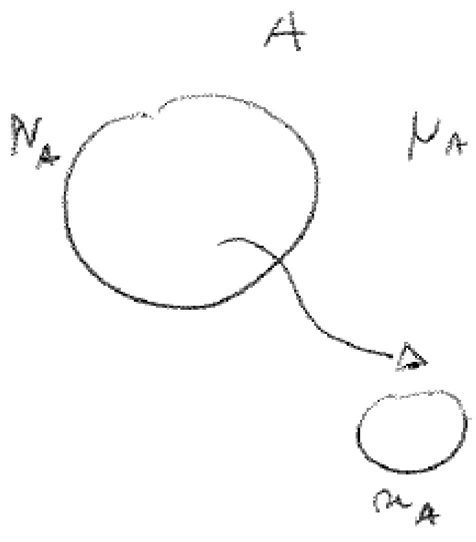
- Come detto – a causa degli errori campionari – è impossibile inferire l'esatto valore medio di Y nella popolazione usando solo l'informazione nel campione.
- Tuttavia è possibile utilizzare i dati relativi a un campione casuale per costruire un insieme di valori che contiene la vera media della popolazione – con una probabilità prefissata. Tale insieme prende il nome di **intervallo di confidenza**, e la probabilità prefissata **livello di confidenza**.

Confronto tra medie di popolazioni diverse (3.4 Stock & Watson) - 1

- Il libro di testo introduce questo argomento con una domanda “I neolaureati e le neolaureate guadagnano mediamente lo stesso ammontare?”
- Questa domanda richiede di confrontare le medie delle distribuzioni di due diverse popolazioni.
- Vediamo come verificare ipotesi e costruire intervalli di confidenza.

Confronto tra medie di popolazioni diverse (3.4 Stock & Watson) - 2

- Ipotizziamo in generale di avere due popolazioni, A e B, e da queste di estrarre 2 campioni indipendenti di numerosità n_A e n_B

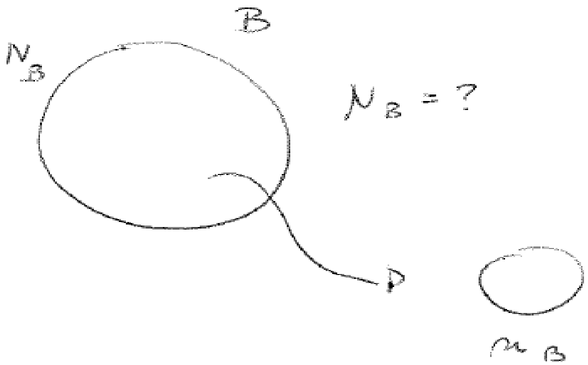


$\mu_A = ?$

CAMPIONE

$$\bar{Y}_A = \frac{\sum Y_{Ai}}{n_A}$$

Confronto tra medie di popolazioni diverse (3.4 Stock & Watson) - 3



$\mu_B = ?$

CAMPIONE

$$\bar{Y}_B = \frac{\sum Y_{Bi}}{n_B}$$

- Riprendendo l'esempio del confronto del reddito medio tra maschi (m) e femmine (w), si ha:

ESEMPIO P. 63 DIFFERENZA REDDITO MEDIO TRA UOMINI E DONNE

$$H_0 : \mu_m - \mu_w = d_0 = \mu_A - \mu_B \quad (3.12)$$

$$H_1 : \mu_m - \mu_w \neq d_0$$

Definizione dello stimatore differenza tra medie

$\bar{Y}_m - \bar{Y}_w$ è uno stimatore di $\mu_m - \mu_w$

DOBBIAMO CONOSCERE LA DISTRIBUZIONE DI

$$\bar{Y}_m - \bar{Y}_w$$

$$\bar{Y}_m \overset{A}{\sim} N\left(\mu_m, \frac{\sigma_m^2}{n_m}\right)$$

$$\bar{Y}_w \overset{A}{\sim} N\left(\mu_w, \frac{\sigma_w^2}{n_w}\right)$$

Proprietà dello stimatore differenza tra medie

ESSENDO LE 2 POPOLAZIONI <u>INDIPENDENTI</u>:

$$E(\bar{Y}_m - \bar{Y}_w) = E(\bar{Y}_m) - E(\bar{Y}_w) = \mu_m - \mu_w$$

$$VAR(\bar{Y}_m - \bar{Y}_w) = VAR(\bar{Y}_m) + VAR(\bar{Y}_w) - \underbrace{2\,COV(\bar{Y}_m, \bar{Y}_w)}_{=0}$$

(SI VEDA ANCHE 3.7)

$$= \frac{\sigma_m^2}{n_m} + \frac{\sigma_w^2}{n_w}$$