

Es. 7.3 PAG. 215

$\hat{\pi}_1$ = stima % di adulti fumatori in America nel 1965 = 0.419 (41.9%)
 $\hat{\pi}_2$ = " " " " 2003 = 0.215 (21.5%)

a) $\hat{\pi}_1 - \hat{\pi}_2 = 0.419 - 0.215 = 0.204$

b) $se_1 = 0.02$
 $se_2 = 0.02$
 $(se) = \sqrt{(se_1)^2 + (se_2)^2} = \sqrt{(0.02)^2 + (0.02)^2} = \sqrt{0.0008} = 0.028$

quanto precisamente descrive con quale precisione il corrispondente parametro (differenza
 0.204 stima $\pi_1 - \pi_2$ tra 2 proporzioni, $\pi_1 - \pi_2$, viene stimato).
 $\pi_1 - \pi_2$ $\hookrightarrow$ variabilit  della stima nei possibili differenziali campionari con
 la stessa numerosita'.

Es. 7.4 PAG. 215

$\hat{\pi}_D = 0.52$ $n = 1008$
 $\hat{\pi}_L = 0.14$ $n = 1002$

a) $0.52 - 0.14 = 0.38$

b) $se_D = 0.0157$
 $se_L = 0.110$
 $se_{D-L} = \sqrt{(se_D)^2 + (se_L)^2} = \sqrt{(0.0157)^2 + (0.110)^2} = \sqrt{0.01466} = 0.121$

NB se della differenza e' sempre piu' grande di se di entrambe le singole
 stime campionarie
 $\bar{y}_2 - \bar{y}_1$ e' piu' lontano da $\mu_2 - \mu_1$ di quanto $\bar{y}_1$ (o $\bar{y}_2$) sia da μ_1 (o μ_2).

Es. 7.5 PAG. 215

$\bar{y}_1 = 140$ $se_1 = 2$
 $\bar{y}_2 = 164$ $se_2 = 2$

a) $\bar{y}_2 - \bar{y}_1 = 164 - 140 = 24$

$se = \sqrt{(se_1)^2 + (se_2)^2} = \sqrt{2^2 + 2^2} = 2.828$

b) $\bar{y}_2 = 1.17 \bar{y}_1 \rightarrow \frac{\bar{y}_2}{\bar{y}_1} = 1.17 \rightarrow \frac{164}{140} = 1.17$

Il peso del 2002 e' del 17% piu' alto che nel 1962.

c) se $\bar{y}_1 = 166$ $\bar{y}_2 = 191$ differenza $\Rightarrow \bar{y}_2 - \bar{y}_1 = 191 - 166 = 25$ $\hookrightarrow$ stima della differenza delle medie di due popolaz. tramite la differenza delle medie campionarie.

RAPPORTO $\Rightarrow \frac{\bar{y}_2}{\bar{y}_1} = \frac{191}{166} = 1.15$ $\Rightarrow$ la $\bar{y}_2$ ^{media campionaria} è il 15% più grande di $\bar{y}_1$ (1.15 volte)
 NB Rapporto pari a 1 se i 2 parametri sono uguali.

• Es. 7.7 PAG. 215

$$\hat{\pi}_1 = \frac{832}{100.000} = 0.00832 \quad (\text{tasso di incarcerazione in proporzione})$$

$$\hat{\pi}_2 = \frac{58}{100.000} = 0.00058$$

a) Rischio RELATIVO $\rightarrow$ rapporto tra 2 proporzioni.

$$\frac{\hat{\pi}_1}{\hat{\pi}_2} = \frac{0.00832}{0.00058} = \frac{832}{58} = 14.345$$

Rischio relativo: ^{tasso} di incarcerazione per gli uomini è 14 volte quello delle donne.

b) $\hat{\pi}_1 - \hat{\pi}_2 = 0.00832 - 0.00058 = 0.00774$ Difficile interpretazione

c) La misura preferibile per sintetizzare i dati è il rapporto (rischio relativo); è molto più informativo della differenza quando entrambe le proporzioni sono vicine allo 0.

• Es. 7.10 PAG. 215

$$\hat{\pi}_1 = 0.60$$

$$\hat{\pi}_2 = 0.57$$

99% IC per $\hat{\pi}_2 - \hat{\pi}_1$

$$(\hat{\pi}_2 - \hat{\pi}_1) \pm z \cdot se$$

$\downarrow$
2.57

$$\Rightarrow (-0.07, 0.01) \Rightarrow -0.03 \pm 0.04$$

NB IC del 99%!

In generale: si può dedurre che $\hat{\pi}_2$ può essere o di 0.07 più piccolo di $\hat{\pi}_1$ o di 0.01 più grande di $\hat{\pi}_1$, con un livello di fiducia del 99%.

- Intervallo include lo zero quindi è plausibile che $\hat{\pi}_2 - \hat{\pi}_1 = 0$, cioè che $\hat{\pi}_2 = \hat{\pi}_1$ (no cambiamento).
- La stima della diff. è negativa, quindi è plausibile un calo nei sostenitori (ma non rilevante, perché IC include 0).
- Aumento perché in intervallo anche val. positivi, ma trascurabili (più sbilanciato verso val. neg.)

• Es. 7.11 PAG. 215

39.9% di 12708
 48.2% di 8783

$$\hat{\pi}_1 = 0.399$$

$$\hat{\pi}_2 = 0.482$$

a) se per $\hat{\pi}_2 - \hat{\pi}_1$

$$se = \sqrt{(se_1)^2 + (se_2)^2}$$

$$se = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} = \sqrt{\frac{0.399(0.601)}{12708} + \frac{0.482(0.518)}{8783}} = 0.0069$$

(78)

b) 95% IC per $\hat{\pi}_2 - \hat{\pi}_1$

$$(\hat{\pi}_2 - \hat{\pi}_1) \pm z \cdot se$$

$$(0.482 - 0.399) \pm 1.96 \cdot 0.0069 = 0.083 \pm 0.012 \quad (0.071, 0.095)$$

stima
marginale
d'errore

$\hat{\pi}_2$ è tra 0.071 e 0.095
più grande di $\hat{\pi}_1$

• Es 7.14 PAG 216

$$\hat{\pi}_1 = \frac{989}{1503} = 0.658$$

$$\hat{\pi}_2 = \frac{704}{1968} = 0.358$$

La tabella è una TABELLA DI CONTINGENZA

Riga: categorie variabile esplicativa che definisce i gruppi da confrontare

Colonna: categoria della variabile risposta

Celle: frequenze

π_1 e π_2 (parametri) $\rightarrow$ PROBAB. CONDIZIONATE

$$a) \hat{\pi}_1 - \hat{\pi}_2 = 0.658 - 0.358 = 0.30$$

$$se = \sqrt{\frac{(0.658)(0.342)}{1503} + \frac{(0.358)(0.642)}{1968}} = 0.0163$$

b) 95% IC per $\hat{\pi}_1 - \hat{\pi}_2$

$$0.30 \pm 1.96 \cdot 0.0163 = 0.30 \pm 0.03 \quad (0.27, 0.33) \rightarrow \text{tutti valori positivi}$$

$$\hat{\pi}_1 > \hat{\pi}_2$$

$$c) \hat{\pi}_1 = \frac{514}{1503} = 0.342$$

$\rightarrow$ nella differenza cambia il segno

$$\hat{\pi}_1 - \hat{\pi}_2 = 0.342 - 0.642 = -0.30$$

$$\hat{\pi}_2 = \frac{1264}{1968} = 0.642$$

$\rightarrow$ se rimane uguale!

$\rightarrow$ 95% IC : stima differenza con segno $\neq$
ma il margine d'errore sarà uguale

$$-0.30 \pm 0.03$$

$(-0.27, -0.33) \rightarrow$ estremi opposti
rispetto a 95% IC ottenuta prima.
(tutti valori negativi).

• Es. 7.15 PAG. 216

(campioni indipendenti, n grande)

$$\hat{\pi}_1 = \frac{153}{411} = 0.372$$

$$\hat{\pi}_2 = \frac{166}{472} = 0.352$$

a) $H_0: \pi_1 - \pi_2 = 0$

$H_a: \pi_1 - \pi_2 \neq 0$

b) Stima della proporz. di pop. presunta H_0 vera

↳ sotto assunzione di H_0 che $\pi_1 = \pi_2$, stimo il valore comune di π_1 e π_2 attraverso la proporz. campionaria per l'intero campione

$$\hat{\pi} = \frac{(153 + 166)}{(411 + 472)} = \frac{319}{883} = 0.36$$

↳ stima congiuntata $\Rightarrow$ raggruppo le osservazioni dei due campioni.

$$\begin{aligned} se_0 &= \sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{(0.36)(0.64)\left(\frac{1}{411} + \frac{1}{472}\right)} \\ &= \sqrt{0.2304 \cdot \left(\frac{1}{411} + \frac{1}{472}\right)} \\ &= \sqrt{0.0010487195} \\ &= 0.0324 \end{aligned}$$

$$z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{se_0} = \frac{0.372 - 0.352}{0.0324} = 0.62$$

c) P-valore? TAVOLA A $\rightarrow$ (0.2672) $\cdot 2 = 0.5344$

IP. BIDIREZ.

con livello $\alpha = 0.05$
Non rifiuto H_0 .

d) $\hat{\pi}_1 = 0.40$ $n = 652$
 $\hat{\pi}_2 = 0.256$ $n = 231$

DIFFERENZA $\rightarrow$ variabile sesso $\hat{\pi}_1 - \hat{\pi}_2 = 0.372 - 0.352 = 0.02$

variabile titolo di studio $= 0.40 - 0.256 = 0.144 \rightarrow$ maggiore influenza

• Es. 7.16 PAG. 216

a) Fanno più uso di maniguana i maschi o le femmine?

b) $H_0: \pi_1 - \pi_2 = 0$ con $\frac{\hat{\pi}_1}{\hat{\pi}_2} = \frac{\text{proporz. F}}{\text{proporz. M}}$

nel 95% IC per $\hat{\pi}_1 - \hat{\pi}_2$ ho solo valori negativi $\Rightarrow \hat{\pi}_1 < \hat{\pi}_2$ femmine fanno meno uso dei maschi.
PERO' non molto diff. rispetto allo 0.

c) P-valore = 0.020 $\rightarrow$ probab. sotto alle 2 code della normale standardizzata oltre al valore osservato della statistica test.
 Per un livello $\alpha = 0.05$ rifiuto H_0 .

• Es. 7.23 PAG. 217

$\bar{y}_1 = 1.8$ 95% IC per $\bar{y}_1 - \bar{y}_2 = (0.2, 0.6)$

$\bar{y}_2 = 1.4$ $t = 4.8$, $p = 0.000$

95% IC non comprende lo 0, ma solo valori positivi $\rightarrow \bar{y}_1 > \bar{y}_2$

$\hookrightarrow$ confermato da test T e P-valore altamente significativo

MA la differenza reale è molto piccola, come mostra IC

• Es. 7.24 PAG. 217

(medie, campioni indep., n grande)

$H_0: \mu_2 - \mu_1 = 0$
 $H_a: \mu_2 - \mu_1 \neq 0$

$t = \frac{\bar{y}_2 - \bar{y}_1}{se} = \frac{2.99 - 2.86}{0.103} = \frac{0.13}{0.103} = 1.26$ con $se = \sqrt{(se_1)^2 + (se_2)^2}$
 $= \sqrt{(0.07)^2 + (0.075)^2}$
 $= 0.103$

gdl = $\infty \rightarrow$ TAVOLA A $\rightarrow (0.1038) \cdot 2 = 0.208$ con $\alpha = 0.05$ non significativo

b) Sì.

95% IC per $\mu_2 - \mu_1 = (\bar{y}_2 - \bar{y}_1) \pm t \cdot se = 0.13 \pm 1.96 \cdot 0.103$
 0.13 ± 0.20
 $(-0.07, 0.33)$

c) No, per la regola empirica (deve essere asimm. dx/positiva)
 No, test robusto se n è molto grande $\rightarrow$ la distrib. campionaria delle diff. tra le medie campionarie ha forma approx. campanulone con n così grande

• Es. 7.28 PAG. 219

a) DIPENDENTI → i due campioni hanno gli stessi soggetti

b) plausibile che $\mu_1 = \mu_2$ perché IC comprende lo 0.

Sog	y_1	y_2	d
1	10	5	5
2	4	0	4
3	12	20	-8
4	2	6	-4
5	12	2	10
6	7	8	-1
7	65	12	33
8	1	25	-24
9	25	0	25
10	12	12	0

$$\bar{y}_d = \frac{(5+4+(-8)+\dots)}{10} = 4$$

Calcolo $s_d = \sqrt{\frac{\sum (d_i - \bar{y}_d)^2}{n-1}} = \sqrt{\frac{2352}{9}} = 16.166$

$$(d_i - \bar{y}_d)^2$$

$(5-4)^2 = 1$
$(4-4)^2 = 0$
$(-8-4)^2 = 144$
$(-4-4)^2 = 64$
$(10-4)^2 = 36$
$(-1-4)^2 = 25$
$(33-4)^2 = 841$
$(-24-4)^2 = 784$
$(25-4)^2 = 441$
$(0-4)^2 = 16$
$\Sigma 2352$

$$se = \frac{s_d}{\sqrt{n}} = \frac{16.166}{\sqrt{10}} = 5.113$$

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 - \mu_2 \neq 0$$

t test per campioni appaiati

$$t = \frac{\bar{y}_d}{se} = \frac{4}{5.113} = 0.782$$

$$gdl = n-1 = 9$$

$$P > (0.1) \cdot 2$$

$$P > 0.2$$

→ con $\alpha = 0.05$ non rifiuto H_0