Verifica di ipotesi sulla varianza

Singola varianza ignota della popolazione

Luccio 9.1.1

Consideriamo un campione casuale di n osservazioni indipendenti estratto da $N(\mu, \sigma)$:

$$W = \frac{(n-1)s^2}{\sigma^2}$$

Segue la distribuzione
Chi-quadrato con $n-1$ gradi di libertà.

Una ipotesi su σ nella popolazione può essere sottoposta a verifica calcolando W sulla base del valore s^2(priva di errore sistematico) fornita dal campione e dal valore σ_0^2 specificato dall'ipotesi nulla.

$$H_0: \sigma^2 = \sigma_0^2$$

$$W = \frac{(n-1)s^2}{\sigma_0^2}$$

$$H_1: \sigma^2 > \sigma_0^2$$

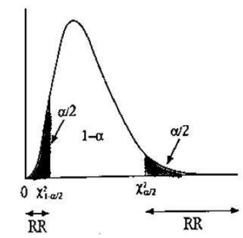
Tanto maggiore è s^2, tanto più forte sarà l'evidenza a sostengo di $H_1: \sigma^2 > \sigma_0^2$.

Se s^2è grande relativamente a σ^2 ne segue che anche il valore di $W = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2$sarà grande.

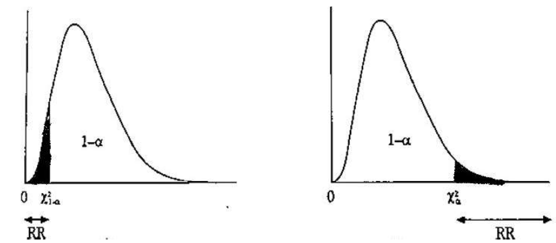
$H_1\colon \sigma^2 < \sigma_0^2$

Tanto minore è s^2, tanto più forte
sarà l'evidenza a sostengo di $H_1\colon \sigma^2 < \sigma_0^2$.

Se s^2 è piccolo relativamente a σ^2 ne segue che
anche il valore di $W = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2$ sarà piccolo.



(a) $H_1 : \sigma^2 \neq \sigma_0^2$,



(b) $H_1 : \sigma^2 < \sigma_0^2$ e (c) $H_1 : \sigma^2 > \sigma_0^2$.

Esempio 9.1.
Test per bambini di 8-9 anni

$N(100, \sqrt{16})$

Ipotesi sperimentale.
La varianza *diminuirebbe* se il test venisse somministrato a bambini tra i 9 e 10 anni

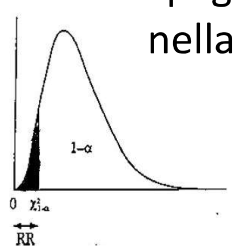
1- La formualzione delle ipotesi

$$H_0\colon \sigma^2 \geq 16$$
$$H_1\colon \sigma^2 < 16$$

2- La scelta del livello di signficatività

$$\alpha = 0{,}05$$

Tavola 4 pag. 294. Valore critico che racchiude nella coda destra $1 - \alpha = .950$



$$\chi^2_{1-\alpha} = 17.7084$$

3- La scelta della statistica appropriata per sottoporre a verifica l'ipotesi nulla

$$W = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(30-1)9.4}{16}$$

$$= 17.037 < 17.7084$$

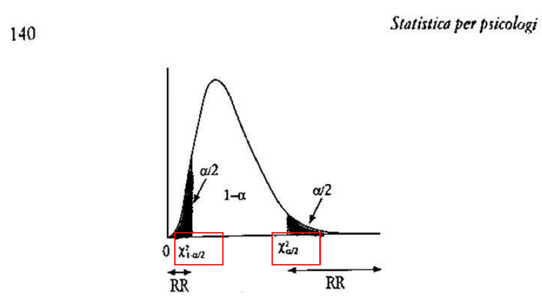
4- Interpretazione

Dato che il valore osservato è minore del valore critico, l'ipotesi nulla può essere rigettata.

Con ipotesi bidirezionale:

$$H_0: \sigma^2 = 16$$
$$H_1: \sigma^2 \neq 16$$

Vanno identificati i due valori critici delle **code** (*non simmetriche*, se non per gradi di libertà molto grandi) sinistra, $\chi^2_{1-\alpha/2} = 16.0471$ e destra, $\chi^2_{\alpha/2} = 45.7223$.



$$16.0471 < 17.037 < 45.7223$$

L'adozione di un'ipotesi sostantiva bidirezionale non avrebbe consentito nel caso presente di rifiutare l'ipotesi nulla.

Il test bidirezionale è meno potente.

Verifica di ipotesi sulla varianza

Singola varianza ignota della popolazione

Approssimazione normale alla distribuzione χ^2

Luccio 9.1.3

La distribuzione χ^2 si approssima alla distribuzione normale con il crescere dei gradi di libertà.

Per grandi campioni avremo che

$$z = \frac{\chi^2 - \nu}{\sqrt{2\nu}}$$

Da cui

$$\chi^2 = z\sqrt{2\nu} + \nu$$

Esempio 9.3.

$\chi^2_{0.05}$ con **5** gradi di libertà

Tavola 4 → 11.0705

$$\chi^2 = z_{0.05}\sqrt{2\nu} + \nu =$$

$$1.645\sqrt{10} + 5 = 10{,}20195$$ **−8%**

Esempio 9.3.

$\chi^2_{0.05}$ con 10 gradi di libertà

Tavola 4 → 18.3070

$$\chi^2 = z_{0.05}\sqrt{2\nu} + \nu =$$

$$1.645\sqrt{20} + 10 = 17{,}3566$$ **−5%**

Esempio 9.3.

$\chi^2_{0.05}$ con 30 gradi di libertà

Tavola 4 → 43.7730

$$\chi^2 = z_{0.05}\sqrt{2\nu} + \nu =$$

$$1.645\sqrt{60} + 30 = 42{,}7421$$ **−2%**

Esempio 9.3.

$\chi^2_{0.05}$ con 100 gradi di libertà

PC (Excel =CHISQ.INV.RT(0,05;100))
→ 124,3421

$$\chi^2 = z_{0.05}\sqrt{2\nu} + \nu =$$

$$1.645\sqrt{200} + 100 = 123{,}26 \quad -1\%$$

Esempio 9.1.

$$H_0: \sigma^2 \geq 16$$
$$H_1: \sigma^2 < 16$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(30-1)9.4}{16}$$

$$= 17.037$$

Excel: =1-CHISQ.DIST.RT(17,037;29)
P-valore = 0,038396

$$z = \frac{\chi^2 - \nu}{\sqrt{2\nu}}$$

$$z = \frac{17{,}037 - 29}{\sqrt{2 \times 29}} = -1.57$$

Tabella 1: per 1.57 la coda di destra è
P-valore = 0,058

Esempio 9.2.

$$n = 25$$

Popolazione normale.
La varianza del campione è $s^2 = 5$.

Calcolare i limiti di un intervallo di fiducia per σ^2, con $\alpha = 0.05$.

Esempio 9.2.

$\chi^2_{0.025}$ con 24 gradi di libertà

Tavola 4 → 39.3641

$\chi^2 = z_{0.025}\sqrt{2\nu} + \nu =$

$1.96\sqrt{48} + 24 = 37.57903$

Esempio 9.2.

$\chi^2_{0.975}$ con 24 gradi di libertà

Tavola 4 → 12.4011

$\chi^2 = z_{0.975}\sqrt{2\nu} + \nu =$

$-1.96\sqrt{48} + 24 = 10{,}421$

Esempio 9.2.

$$P\left(\frac{\nu s^2}{\chi^2_{0.025}} \leq \sigma^2 \leq \frac{\nu s^2}{\chi^2_{0.975}}\right) = .95$$

$$P\left(\frac{\nu s^2}{z_{0.025}\sqrt{2\nu} + \nu} \leq \sigma^2 \leq \frac{\nu s^2}{z_{0.975}\sqrt{2\nu} + \nu}\right) = .95$$

Utilizzando l'approssimazione normale

Esempio 9.2.

$$P\left(\frac{24 \cdot 5}{39.3641} \leq \sigma^2 \leq \frac{24 \cdot 5}{12.4011}\right) = .95$$

$$P\left(\frac{24 \cdot 5}{37.5790} \leq \sigma^2 \leq \frac{24 \cdot 5}{10{,}421}\right) = .95$$

Utilizzando l'approssimazione normale

Esempio 9.2.

$$P(3{,}048 \leq \sigma^2 \leq 9{,}677) = .95$$

$$P(3{,}193 \leq \sigma^2 \leq 11{,}515) = .95$$

Utilizzando l'approssimazione normale

Verifica di ipotesi sulla varianza

Il confronto tra le varianze di due campioni estratti da due popolazioni normali

La distribuzione χ^2 fornisce la base per lo sviluppo di procedure inferenziali riguardanti il confronto tra le varianze di due campioni estratti da due popolazioni normali.

Esempio. Confrontare la precisione di due strumenti di misura

Consideriamo due campioni *indipendenti* costituiti da n_1 e n_2 osservazioni casuali, con varianze campionarie (corrette) s^2_1 e s^2_2, e tratti da due popolazioni normali con varianze σ^2_1 e σ^2_2.

$$W_1 = \frac{(n_1 - 1)s_1^2}{\sigma_1^2} \sim \chi_1^2$$

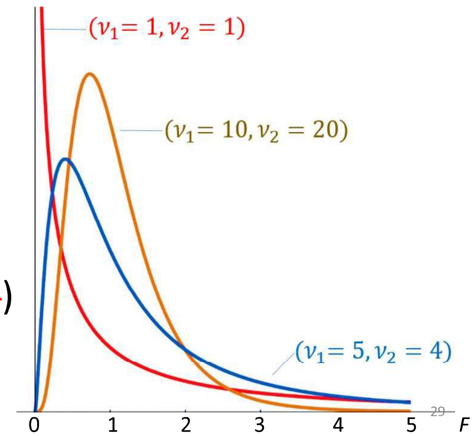
$$W_2 = \frac{(n_2 - 1)s_2^2}{\sigma_2^2} \sim \chi_2^2$$

con $\nu_1 = (n_1 - 1)$ e $\nu_2 = (n_2 - 1)$ g.d.libertà.

$$F = \frac{\frac{\chi_1^2}{\nu_1}}{\frac{\chi_2^2}{\nu_2}}$$

Segue una distribuzione *F* con ν_1 gradi di libertà al numeratore e ν_2 gradi di libertà al denominatore.

Tabelle
3a (α=.05) ,
3b (α=.025) , e
3c (α=.01 ~~non .001~~)



$$F = \frac{W_1/\nu_1}{W_2/\nu_2} = \frac{[(n_1-1)s_1^2/\sigma_1^2]/(n_1-1)}{[(n_2-1)s_2^2/\sigma_2^2]/(n_2-1)} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

Segue una distribuzione *F* con $\nu_1 = (n_1 - 1)$ gradi di libertà al numeratore e $\nu_2 = (n_2 - 1)$ gradi di libertà al denominatore.

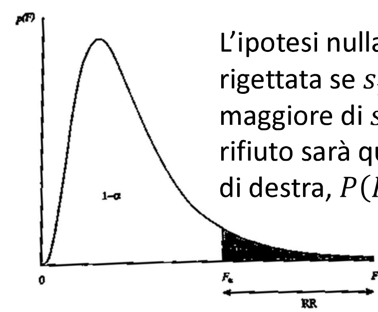
$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma^2$$

$$F = \frac{s_1^2/\sigma^2}{s_2^2/\sigma^2} = \frac{s_1^2}{s_2^2}$$

Segue una distribuzione *F* con $\nu_1 = (n_1 - 1)$ gradi di libertà al numeratore e $\nu_2 = (n_2 - 1)$ gradi di libertà al denominatore.

$$H_1: \sigma_1^2 > \sigma_2^2$$

$$F = \frac{s_1^2}{s_2^2}, \text{con } s_1^2 > s_2^2$$



L'ipotesi nulla potrà essere rigettata se s_1^2 è molto maggiore di s_2^2. La regione di rifiuto sarà quindi la coda di destra, $P(F > F_\alpha) = \alpha$

Figura 9.4. Regione di rifiuto per $H_0: \sigma_1^2 = \sigma_2^2$ e $H_1: \sigma_1^2 > \sigma_2^2$.

$H_1 : \sigma_1^2 \neq \sigma_2^2$

$$F = \frac{s_1^2}{s_2^2}, \text{con } s_1^2 > s_2^2$$

L'ipotesi nulla potrà essere rigettata se il valore calcolato di *F* cade nella regione di rifiuto $\boldsymbol{\alpha/2}$ della coda di destra della distribuzione. Il nuovo valore critico sarà quindi $F_{\alpha/2}$.

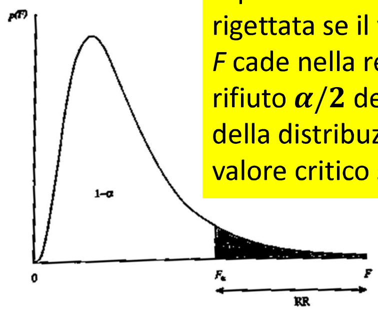


Figura 9.4. Regione di rifiuto per $H_0 : \sigma_1^2 = \sigma_2^2$ *e* $H_1 : \sigma_1^2 > \sigma_2^2$