n AMPIEZZA CAMPIONE

x_i OSSERVAZIONE i ESIMA $\{x_i\}$ CAMPIONE

X_i V.A. $\{X_i\}$ FAMIGLIA DI V.A.

$X_i = x_i$ DETERMINAZIONE REALIZZAZIONE

$Y = f(X_i)$ UNA V.A. FUNZIONE DI V.A. PUÒ DIPENDERE DA PARAMETRI

$T = f(X_i)$ STATISTICA (È UNA V.A.) NON DIPENDE DA ϑ

$t = f(x_i)$ REALIZZAZIONE (È UN NUMERO)

ϑ PARAMETRO DA STIMARE (È UN NUMERO)

T STIMATORE (È UNA V.A.)

t STIMA (È UN NUMERO)

$D_T = \mathbb{E}T - \vartheta$ DISTORSIONE

$\mathbb{E}T - \vartheta = 0$ NON DISTORTO (CORRETTO)

$\lim_{n \to \infty} \mathbb{E}T_n - \vartheta = 0$ ASINTOTICAMENTE CORRETTO

$\lim EQM(T_n) = 0$ CONSISTENTE

$$EQM(T_n) = \mathbb{E}\left[(T_n - \theta)^2\right] = \begin{cases} Var\, T_n + D_{T_n}^2 & T_n \text{ GENERICO} \\ Var\, T_n & T_n \text{ NON DISTORTO} \end{cases}$$

EFFICIENZA RELATIVA DI DUE STIMATORI T_n E Θ_n

SE $\dfrac{EQM(T_n)}{EQM(\Theta_n)} < 1 \Rightarrow T_n$ È IL PIÙ EFFICIENTE

Definizione 2.1.2. La **frequenza assoluta** $f_a(k)$ relativa alla k-esima classe è il numero di osservazioni che ricadono in quella classe.

$$f_a(k) = \#C_k \quad (k = 1, \dots, N_c)$$

Definizione 2.1.3. La **frequenza relativa** $f_r(k)$ della k-esima classe è il rapporto $f_a(k)/n$
Proprietà: $\sum_{k=1}^{N_c} f_r(k) = 1$.

Definizione 2.1.4. La **frequenza percentuale** $f_p(k)$ è la quantità $f_p(k) = f_r(k) \cdot 100$.
Proprietà: $\sum_{k=1}^{N_c} f_p(k) = 100$.

Definizione 2.1.5. La **frequenza assoluta cumulativa** $F_a(k)$ della k-esima classe è il numero totale delle osservazioni che ricadono nelle classi fino a k-esima compresa:

$$F_a(k) = \sum_{j=1}^{k} f_a(j).$$

Definizione 2.1.6. La **frequenza relativa cumulativa** è il rapporto $F_r(k) = F_a(k)/n \equiv \sum_{j=1}^{k} f_r(j)$, ed è sempre compresa fra 0 ed 1.
Proprietà: F_r è una funzione non decrescente e $F_r(N_c) = 1$.

Definizione 2.1.7. La **frequenza percentuale cumulativa** $F_p(k)$ è la quantità $F_p(k) = F_r(k) \cdot 100$. Proprietà: F_p è una funzione non decrescente e $F_p(N_c) = 100$.

- **media** o **media campionaria** di n dati numerici $\{x_i\}_{i=1}^{n}$ $\quad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \sum_{k=1}^{N_c} f_r(k)\overline{x}_k$

- **Varianza campionaria**

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\overline{x}^2 = \frac{n}{n-1}\left[\overline{x^2} - (\overline{x})^2\right] = \frac{1}{n-1}\left(\sum_i x_i^2 - \frac{(\sum x_i)^2}{n}\right)$$

Varianza calcolata in base ai dati raggruppati:

$$\frac{n}{n-1}\left(\sum_{k=1}^{N_c} f_r(k)\overline{x}_k^2 - \overline{x}^2\right) = \frac{n}{n-1}\sum_{k=1}^{N_c} f_r(k)\left(\overline{x}_k - \overline{x}\right)^2$$

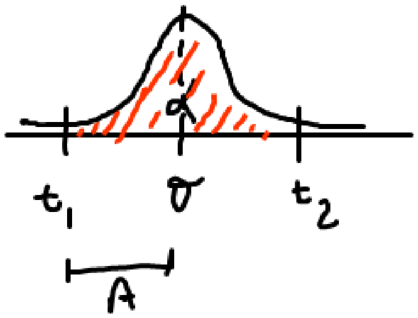
IdC e TEST PARAMETRICI

per MEDIA	con VARIANZA	NOTA
		INCOGNITA
per VARIANZA	con MEDIA	NOTA
		INCOGNITA
per FREQUENZA		
cfr 2 MEDIE	con VARIANZE	NOTE
cfr 2 MEDIE	con VARIANZE	INCOGNITE UGUALI
cfr 2 MEDIE	DATI ACCOPPIATI	
cfr 2 FREQUENZE		

TEST NON PARAMETRICI

ADATTAMENTO

INDIPENDENZA



$$\mathbb{P}\left[T_1 < \vartheta < T_2\right] = \alpha$$

α LIVELLO DI CONFIDENZA

$(t_1 ; t_2)$ INTERVALLO DI CONFIDENZA PER ϑ

IdC PER LA MEDIA μ DI UNA POPOLAZIONE NORMALE CON VARIANZA σ^2 NOTA / INCOGNITA

DA USARE SOLO SE LA POPOLAZIONE È NORMALE (PERCHÉ LO DICE IL TESTO O PERCHÉ IL CAMPIONE HA SUPERATO UN TEST DI NORMALITÀ)

- CON VARIANZA NOTA

BILATERO: $[\mu_- ; \mu_+]$ CON $\mu_\pm = \overline{X}_n \pm z_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

MAGGIORAZIONE: $(-\infty ; \mu_+]$ CON $\mu_+ = \overline{X}_n + z_\alpha \frac{\sigma}{\sqrt{n}}$

MINORAZIONE: $[\mu_- ; +\infty)$ CON $\mu_- = \overline{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}}$

- CON VARIANZA INCOGNITA

BILATERO: $[\mu_- ; \mu_+]$ CON $\mu_\pm = \overline{X}_n \pm t_{\frac{1+\alpha}{2}}(n-1) \frac{S}{\sqrt{n}}$

MAGGIORAZIONE: $(-\infty ; \mu_+]$ CON $\mu_+ = \overline{X}_n + t_\alpha(n-1) \frac{S}{\sqrt{n}}$

MINORAZIONE: $[\mu_- ; +\infty)$ CON $\mu_- = \overline{X}_n - t_\alpha(n-1) \frac{S}{\sqrt{n}}$

DOVE $t_\beta(\tau)$ È IL QUANTILE DI LIVELLO β DI UNA t-STUDENT CON τ GRADI DI LIBERTÀ

E S È LA DEV. STANDARD CAMPIONARIA

IdC PER LA MEDIA μ DI UNA POPOLAZIONE ASINTOTICAMENTE NORMALE CON VARIANZA σ^2 NOTA / INCOGNITA

DA USARE SOLO SE LA POPOLAZIONE NON È NORMALE A PATTO CHE SIA SODDISFATTA LA CONDIZIONE DI APPLICABILITÀ $n \geq 30$ OVVERO CHE IL CAMPIONE SIA SUFFICIENTEMENTE NUMEROSO. IN TAL CASO IL CAMPIONE, PUR NON ESSENDO NORMALE, PUÒ ESSERE CONSIDERATO ASINTOTICAMENTE NORMALE. L'INTERVALLO PUÒ ESSERE CALCOLATO E LE FORMULE SONO LE STESSE.

CASO PARTICOLARE: CAMPIONE BERNOULLIANO

CAMBIANO LE CONDIZIONI E LE FORMULE.

COND. DI APPLICABILITÀ: $n\bar{x}_n \geq 5 \qquad n(1-\bar{x}_n) \geq 5$

BILATERO: $[p_- ; p_+]$ CON $p_\pm = \bar{X}_n \pm z_{\frac{1+\alpha}{2}} \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}$

MAGGIORAZIONE: $(-\infty ; p_+]$ CON $p_+ = \bar{X}_n + z_\alpha \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}$

MINORAZIONE: $[p_- ; +\infty)$ CON $p_- = \bar{X}_n - z_\alpha \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}$

	H_0 è vera	H_0 è falsa
Rifiutiamo H_0	*Errore di I specie*	Decisione corretta
Accettiamo H_0	Decisione corretta	*Errore di II specie*

α $\quad \pi = 1-\beta$
$1-\alpha$ $\quad \beta$

$\alpha = P[\text{RIFIUTO } H_0 \mid H_0 \text{ VERA}]$

$\beta = P[\text{ACCETTO } H_0 \mid H_0 \text{ FALSA}]$

Le tappe di un test

1. Sulla base del tipo di esperimento si sceglie un modello statistico per gli esiti, con uno o più parametri θ incogniti.
2. Si scelgono H_0 e H_1 (affermazioni sul valore di θ) in modo che sia più grave commettere l'errore di I specie che quello di II (equivalentemente, H_1 sarà l'ipotesi che riterremo vera **solo** in presenza di **forte** evidenza).
3. Si sceglie una statistica $T = t(X_1, \ldots, X_n)$ (in genere stimatore di θ).
4. Si sceglie il livello di significatività $\alpha \in [0, 1]$, cioè la **massima** probabilità che tolleriamo di commettere un errore di I specie.
5. In dipendenza da T e α si sceglie la regione di rifiuto $\mathcal{R} = ((x_1, \ldots, x_n) \,:\, t(x_1, \ldots, x_n) \in I)$.
6. Si esegue il campionamento (= si fanno gli esperimenti e si osservano dei dati, realizzazioni del campione casuale).
7. Si calcola la statistica T sulla base del campione osservato.
8. Si decide: se il valore calcolato della statistica sta in I allora **rifiutiamo** H_0 (e accettiamo H_1) $\Longrightarrow$ conclusione forte.
   Altrimenti **accettiamo** H_0 (e rifiutiamo H_1), dicendo che *non c'è evidenza statistica che H_0 sia falsa.* $\Longrightarrow$ conclusione debole.

Il p-value è il più basso livello di significatività per cui i dati osservati portano a rifiutare H_0.

Ricordiamo che più è basso il livello di significatività e più risulta difficile rifiutare H_0 (cioè ci vogliono prove più evidenti).

Perciò un **p-value basso** (0.01 o 0.001...) indicherà che è **molto probabile che H_0 sia falsa** (anche se *potrebbe* essere vera! pensate alla moneta equilibrata che dia 700 teste in 1000 lanci...).

Un **p-value alto** (0.10 o 0.30...) indicherà che è **poco (o per nulla) evidente che H_0 sia falsa**.

Test di livello α per μ

Popolazione $\mathcal{N}(\mu, \sigma^2)$; varianza σ^2 nota.
μ_0 numero reale fissato.

$$\text{Statistica: } Z_n = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}.$$

H_0	H_1	Rifiutiamo H_0 se	p-value
$\mu \leq \mu_0$	$\mu > \mu_0$	$z_n > z_{1-\alpha}$	$\Phi(-z_n)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$z_n < -z_{1-\alpha}$	$\Phi(z_n)$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\|z_n\| > z_{1-\frac{\alpha}{2}}$	$2 - 2\Phi(\|z_n\|)$

dove z_n= valore di Z_n calcolato dal campione; z_β= quantile β della $\mathcal{N}(0,1)$.

Popolazione $\mathcal{N}(\mu, \sigma^2)$; varianza σ^2 ignota.
μ_0 numero reale fissato.

$$\text{Statistica: } T_n = \frac{\overline{X}_n - \mu_0}{\sqrt{S_n^2/n}}.$$

H_0	H_1	Rifiutiamo H_0 se	p-value=$\overline{\alpha}$ tale che
$\mu \leq \mu_0$	$\mu > \mu_0$	$t_n > t_{1-\alpha}(n-1)$	$t_{1-\overline{\alpha}}(n-1) = t_n$
$\mu \geq \mu_0$	$\mu < \mu_0$	$t_n < -t_{1-\alpha}(n-1)$	$t_{1-\overline{\alpha}}(n-1) = -t_n$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\|t_n\| > t_{1-\frac{\alpha}{2}}(n-1)$	$t_{1-\overline{\alpha}/2}(n-1) = \|t_n\|$

dove t_n= valore di T_n calcolato dal campione; $t_\beta(n-1)$= quantile β della $t(n-1)$.

SE LA POPOLAZIONE NON È NORMALE, PER POTER PARLARE DI POPOLAZIONE ASINTOTICAMENTE NORMALE E CONDURRE IL TEST, DEVE ESSERE VERIFICATA LA CONDIZIONE $n \geq 30$.

CASO PARTICOLARE: Test su una frequenza (Bernoulli)

CONDIZIONE DI APPLICABILITÀ: $n\overline{x}_n \geq 5$ e $n(1-\overline{x}_n) \geq 5$.

Test di livello α per p

$$\text{Statistica: } Z_n = \frac{\overline{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}}.$$

H_0	H_1	Rifiutiamo H_0 se	p-value
$p = p_0$	$p \neq p_0$	$\|z_n\| > z_{1-\alpha/2}$	$2 - 2\Phi(\|z_n\|)$
$p \leq p_0$	$p > p_0$	$z_n > z_{1-\alpha}$	$\Phi(-z_n)$
$p \geq p_0$	$p < p_0$	$z_n < -z_{1-\alpha}$	$\Phi(z_n)$

dove z_n= valore di Z_n calcolato dal campione; z_β= quantile β della $\mathcal{N}(0,1)$.

Test per due medie - campioni accoppiati

H_0 per 2 medie	H_0 per 1 media
$\mu_X = \mu_Y + \delta$	$\mu_Z = \delta$
$\mu_X \leq \mu_Y + \delta$	$\mu_Z \leq \delta$
$\mu_X \geq \mu_Y + \delta$	$\mu_Z \geq \delta$

Test per due medie - campioni indipendenti

Test per due medie - campioni indipendenti, varianze note

H_0	H_1	Rifiuto H_0 se	p-value
$\mu_X - \mu_Y = \delta$	$\mu_X - \mu_Y \neq \delta$	$\lvert z_{n,m}\rvert > z_{1-\alpha/2}$	$2 - 2\Phi(\lvert z_{n,m}\rvert)$
$\mu_X - \mu_Y \leq \delta$	$\mu_X - \mu_Y > \delta$	$z_{n,m} > z_{1-\alpha}$	$\Phi(-z_{n,m})$
$\mu_X - \mu_Y \geq \delta$	$\mu_X - \mu_Y < \delta$	$z_{n,m} < -z_{1-\alpha}$	$\Phi(z_{n,m})$

$$\text{dove } z_{n,m} = \frac{\overline{x}_n - \overline{y}_m - \delta}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

OMOSCHEDASTICITÀ
MA UGUALI

Test per due medie - campioni indipendenti, varianze ignote

H_0	H_1	Rifiuto H_0 se	p-value$=\overline{\alpha}$ con
$\mu_X - \mu_Y = \delta$	$\mu_X - \mu_Y \neq \delta$	$\lvert\hat{t}\rvert > t_{1-\alpha/2}(k)$	$t_{1-\overline{\alpha}/2}(k) = \lvert\hat{t}\rvert$
$\mu_X - \mu_Y \leq \delta$	$\mu_X - \mu_Y > \delta$	$\hat{t} > t_{1-\alpha}(k)$	$t_{1-\overline{\alpha}}(k) = \hat{t}$
$\mu_X - \mu_Y \geq \delta$	$\mu_X - \mu_Y < \delta$	$t < -t_{1-\alpha}(k)$	$t_{1-\overline{\alpha}}(k) = -t$

dove $k = n + m - 2$,

$$\hat{t} = \frac{\overline{x}_n - \overline{y}_m - \delta}{s\sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

$$S = \sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}.$$

↳ DEV. STD COMBINATA
(POOLED)

Test per due frequenze

Abbiamo due popolazioni, la $X_1, \ldots, X_n$ sia $\mathcal{B}(p_1)$ e la $Y_1, \ldots, Y_m$ sia $\mathcal{B}(p_2)$, con

$$n\overline{x}_n \geq 5, \; n(1-\overline{x}_n) \geq 5, \qquad m\overline{y}_m \geq 5, \; m(1-\overline{y}_m) \geq 5.$$

Test per due frequenze

H_0	H_1	Rifiuto H_0 se	p-value
$p_1 = p_2$	$p_1 \neq p_2$	$\lvert z_{n,m}\rvert > z_{1-\alpha/2}$	$2 - 2\Phi(\lvert z_{n,m}\rvert)$
$p_1 \leq p_2$	$p_1 > p_2$	$z_{n,m} > z_{1-\alpha}$	$\Phi(-z_{n,m})$
$p_1 \geq p_2$	$p_1 < p_2$	$z_{n,m} < -z_{1-\alpha}$	$\Phi(z_{n,m})$

$$\text{dove } z_{n,m} = \frac{\overline{x}_n - \overline{y}_m - \delta}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}}$$

$$\hat{p} = (\overline{x}_n n + \overline{y}_m m)/(n+m).$$

SONO AMMESSE SIA
LA t-STUDENT
CHE LA NORMALE (PIÙ $k = n+m-2$ È GRANDE MEGLIO È)

INTERVALLI DI CONFIDENZA PER LA DIFFERENZA DI DUE MEDIE

1) POP. NORMALI, INDIPENDENTI, A VARIANZA NOTA

$$\left[\overline{x}_{n_X} - \overline{y}_{n_Y} + q_{(1-\alpha)/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}},\ \overline{x}_{n_X} - \overline{y}_{n_Y} + q_{(1+\alpha)/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\right]$$

2) POP. NORMALI, INDIPENDENTI, OMOSCHEDASTICITÀ

$$\left[\overline{x}_{n_X} - \overline{y}_{n_Y} + t_{\frac{1-\alpha}{2}}(n_X+n_Y-2)\sqrt{\frac{s_p^2}{n_X} + \frac{s_p^2}{n_Y}},\ \overline{x}_{n_X} - \overline{y}_{n_Y} + t_{\frac{1+\alpha}{2}}(n_X+n_Y-2)\sqrt{\frac{s_p^2}{n_X} + \frac{s_p^2}{n_Y}}\right]$$

dove

$$s_p^2 := \frac{(n_X-1)s_X^2 + (n_Y-1)s_Y^2}{n_X+n_Y-2}.$$

3) POP. BERNOULLIANE (DUE FREQUENZE), CONDIZIONI TLC SODDISFATTE

$$\left[\overline{p}_{n_X} - \overline{p}_{n_Y} + q_{\frac{1-\alpha}{2}}\sqrt{\frac{\overline{p}_{n_X}(1-\overline{p}_{n_X})}{n_X} + \frac{\overline{p}_{n_Y}(1-\overline{p}_{n_Y})}{n_Y}},\ \overline{p}_{n_X} - \overline{p}_{n_Y} + q_{\frac{1+\alpha}{2}}\sqrt{\frac{\overline{p}_{n_X}(1-\overline{p}_{n_X})}{n_X} + \frac{\overline{p}_{n_Y}(1-\overline{p}_{n_Y})}{n_Y}}\right]$$

Intervallo per σ^2 di popolazione $\mathcal{N}(\mu, \sigma^2)$, con μ nota.

$$\left(\frac{nT_n^2}{\chi^2_{\frac{1+\alpha}{2}}(n)}, \frac{nT_n^2}{\chi^2_{\frac{1-\alpha}{2}}(n)} \right)$$

$$T_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

$$\sigma^2 \in \left[0, \frac{nT_n^2}{\chi^2_{1-\alpha}(n)}\right] \qquad \sigma^2 \in \left[\frac{nT_n^2}{\chi^2_{\alpha}(n)}, +\infty\right)$$

Intervallo bilatero per varianza di pop. $\mathcal{N}(\mu, \sigma^2)$, con μ ignota.

$$\left(\frac{(n-1)S_n^2}{\chi^2_{\frac{1+\alpha}{2}}(n-1)}, \frac{(n-1)S_n^2}{\chi^2_{\frac{1-\alpha}{2}}(n-1)} \right)$$

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

$$\sigma^2 \in \left[0, \frac{(n-1)S_n^2}{\chi^2_{1-\alpha}(n-1)}\right] \qquad \sigma^2 \in \left[\frac{(n-1)S_n^2}{\chi^2_{\alpha}(n-1)}, +\infty\right)$$

Test per la varianza di una popolazione $\mathcal{N}(\mu, \sigma^2)$, conosciamo μ e con σ_0^2 numero fissato,

Test di livello α per σ^2

Statistica: $W_n = \dfrac{nT_n^2}{\sigma_0^2}$.

H_0	H_1	Rifiutiamo H_0 se	p-value=$\overline{\alpha}$ tale che
$\sigma^2 \le \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$w_n > \chi^2_{1-\alpha}(n)$	$\chi^2_{1-\overline{\alpha}}(n) = w_n$
$\sigma^2 \ge \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$w_n < \chi^2_{\alpha}(n)$	$\chi^2_{\overline{\alpha}}(n) = w_n$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \ne \sigma_0^2$	$w_n > \chi^2_{1-\frac{\alpha}{2}}(n)$ oppure $w_n < \chi^2_{\frac{\alpha}{2}}(n)$	$\overline{\alpha}=2\min(\alpha_0, 1-\alpha_0)$ dove $w_n^2=\chi^2_{\alpha_0}(n)$

dove w_n= valore di W_n calcolato dal campione;
$\chi^2_\beta(n)$= quantile β della $\chi^2(n)$.

Test per la varianza di una popolazione $\mathcal{N}(\mu, \sigma^2)$, NON conosciamo μ e con σ_0^2 numero fissato,

Test di livello α per σ^2

Statistica: $W_n = \dfrac{(n-1)S_n^2}{\sigma^2}$.

H_0	H_1	Rifiutiamo H_0 se	p-value=$\overline{\alpha}$ tale che
$\sigma^2 \le \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$w_n > \chi^2_{1-\alpha}(n-1)$	$\chi^2_{1-\overline{\alpha}}(n-1) = w_n$
$\sigma^2 \ge \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$w_n < \chi^2_{\alpha}(n-1)$	$\chi^2_{\overline{\alpha}}(n-1) = w_n$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \ne \sigma_0^2$	$w_n > \chi^2_{1-\frac{\alpha}{2}}(n-1)$ opp. $w_n < \chi^2_{\frac{\alpha}{2}}(n-1)$	$\overline{\alpha}=2\min(\alpha_0, 1-\alpha_0)$ dove $w_n^2=\chi^2_{\alpha_0}(n-1)$

dove w_n= valore di W_n calcolato dal campione;
$\chi^2_\beta(n-1)$= quantile β della $\chi^2(n-1)$.

$$\left[\frac{s_X^2}{s_Y^2} F_{(1-\alpha)/2}(n_Y-1, n_X-1),\ \frac{s_X^2}{s_Y^2} F_{(1+\alpha)/2}(n_Y-1, n_X-1)\right]$$

Test di livello α per due varianze

$$\text{Stima: } f = \frac{s_X^2}{s_Y^2}$$

H_0	H_1	Rifiutare H_0 se	P-value
$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 \neq \sigma_Y^2$	$f < F_{\alpha/2}(n_X-1, n_Y-1)$ oppure $f > F_{1-\alpha/2}(n_X-1, n_Y-1)$	$\overline{\alpha} = 2\min(\alpha_0, 1-\alpha_0)$
$\sigma_X^2 \neq \sigma_Y^2$	$\sigma_X^2 = \sigma_Y^2$	$f > F_{(1-\alpha)/2}(n_X-1, n_Y-1)$ e $f < F_{(1+\alpha)/2}(n_X-1, n_Y-1)$	$\overline{\alpha} = \|2\alpha_0 - 1\|$
$\sigma_X^2 \leq \sigma_Y^2$	$\sigma_X^2 > \sigma_Y^2$	$f > F_{1-\alpha}(n_X-1, n_Y-1)$	$f = F_{1-\overline{\alpha}}(n_X-1, n_Y-1)$
$\sigma_X^2 \geq \sigma_Y^2$	$\sigma_X^2 < \sigma_Y^2$	$f < F_{\alpha}(n_X-1, n_Y-1)$	$f = F_{\overline{\alpha}}(n_X-1, n_Y-1)$

dove $F_{\alpha_0}(n_X-1, n_Y-1) = f$.

1. Dividiamo l'insieme dei possibili valori che le singole osservazioni possono assumere in k classi: $C_1, C_2, \ldots, C_k$.
2. Chiamiamo p_i la probabilità che una osservazione appartenga alla classe C_i.
3. Decidiamo di fare n osservazioni.
4. Ognuna delle classi ha una frequenza assoluta teorica np_i e una frequenza assoluta osservata $f_a(i)$.
5. La statistica di riferimento è

$$Q = \sum_{i=1}^{k} \frac{(np_i - f_a(i))^2}{np_i}$$

La statistica, a seconda dell'occorrenza, ha altre espressioni equivalenti:

$$Q := \sum_{i=1}^{N_c} \frac{(np_i - f_a(i))^2}{np_i} = n \sum_{i=1}^{N_c} \frac{(p_i - f_r(i))^2}{p_i}$$
$$= \sum_{i=1}^{N_c} \frac{f_a(i)^2}{np_i} - n = n \left(\sum_{i=1}^{N_c} \frac{f_r(i)^2}{p_i} - 1 \right)$$

Regola pratica per approssimare

Le approssimazioni $Q \approx \chi^2(k-1)$ e $Q \approx \chi^2(k-1-r)$ valgono se le probabilità p_i soddisfano $np_i \geq 5$ per ogni i.

Dato un campione casuale $X_1, \ldots, X_n$, se le probabilità p_i sono determinate *senza stimare parametri* e se $np_i \geq 5$ per ogni $i \in \{1, \ldots, k\}$, allora il test di adattamento è

H_0	$\mathbb{P}(X_1 \in C_i) = p_i$ per ogni $i \in \{1, \ldots, k\}$
H_1	$\exists i$ tale che $\mathbb{P}(X_1 \in C_i) \neq p_i$
Rifiutiamo H_0 se	$q > \chi^2_{1-\alpha}(k-1)$
p-value: $\bar{\alpha}$ tale che	$q = \chi^2_{1-\bar{\alpha}}(k-1)$

dove $\chi^2_{1-\alpha}(k-1)$ è il quantile $1-\alpha$ della legge $\chi^2(k-1)$.

Dato un campione casuale $X_1, \ldots, X_n$, se le probabilità p_i sono determinate *stimando r parametri* e se $np_i \geq 5$ per ogni $i \in \{1, \ldots, k\}$, allora il test di adattamento è

H_0	$\mathbb{P}(X_1 \in C_i) = p_i$ per ogni $i \in \{1, \ldots, k\}$
H_1	$\exists i$ tale che $\mathbb{P}(X_1 \in C_i) \neq p_i$
Rifiutiamo H_0 se	$q > \chi^2_{1-\alpha}(k-r-1)$
p-value: $\bar{\alpha}$ tale che	$q = \chi^2_{1-\bar{\alpha}}(k-r-1)$

dove $\chi^2_{1-\alpha}(k-r-1)$ è il quantile $1-\alpha$ della legge $\chi^2(k-r-1)$.