Statistica e Laboratorio

9. Inferenza statistica: verifica delle ipotesi

Paolo Vidoni

Dipartimento di Scienze Economiche e Statistiche
Università di Udine
via Tomadini 30/a - Udine
paolo.vidoni@uniud.it

https://elearning.uniud.it/

Sommario

Sommario

- Introduzione
- Test statistici per la verifica delle ipotesi
- Test sulla media e sulla varianza
- Test per il confronto tra medie e tra varianze
- Test su proporzioni

Introduzione

Con le procedure di verifica di ipotesi si vuole verificare, sulla base dell'osservazione campionaria e del modello statistico, se una certa congettura o ipotesi sulla popolazione (fenomeno) oggetto di indagine sia accettabile o meno.

Si è propensi ad accettare l'ipotesi se essa è in accordo con i dati campionari osservati x.

Nell'ambito dell'*inferenza statistica parametrica*, l'**ipotesi** (**statistica**) è una affermazione o una congettura sul parametro ignoto θ.

L'ipotesi è **semplice** se specifica in modo completo la popolazione (il processo generatore dei dati); ad esempio, $\theta = 3$.

L'ipotesi è **composta** se non specifica in modo completo la popolazione (il processo generatore dei dati); ad esempio, $\theta > 3$, $\theta < 3$, $\theta \neq 3$.

Un **test statistico** è un procedimento che consente di rifiutare o non rifiutare (e quindi accettare) un'ipotesi statistica.

La decisione dipende dalla discrepanza, più o meno accentuata, che si rileva tra quanto ci si attende sulla base dell'ipotesi formulata e quanto si osserva nel campione.

L'ipotesi sottoposta a verifica viene chiamata **ipotesi nulla**; nel seguito si considereranno *ipotesi nulle semplici* del tipo

$$H_0 \, : \, \theta = \theta_0$$

dove θ_0 un valore fissato dello spazio parametrico.

L'ipotesi nulla è, in genere, un'assunzione semplificatrice sul modello statistico o che descrive le conoscenze attuali sulla popolazione (fenomeno) oggetto di indagine.

In alcuni contesti, l'ipotesi nulla viene formulata sulla base di specifiche richieste, come ad esempio, definizioni contrattuali, obiettivi di qualità, ecc.

Congetture non contemplate da H_0 costituiscono l'**ipotesi alternativa** H_1. Nel seguito si considerano i seguenti casi:

- $H_1 : \theta > \theta_0$, detta **alternativa unilaterale destra**;
- $H_1 : \theta < \theta_0$, detta **alternativa unilaterale sinistra**;
- $H_1 : \theta \neq \theta_0$, detta **alternativa bilaterale**.

Tipicamente interessa provare l'inaccettabilità di H_0 a favore di H_1, ammettendo una certa quota di errore; in questo caso si dice che il **test** è **significativo** (contro H_0).

La conclusione di un test statistico, che porta ad non rifiutare l'ipotesi nulla o a rifiutarla a favore dell'alternativa, contempla la possibilità di errore.

Non si dice nulla circa la verità o falsità di H_0, si afferma unicamente che l'ipotesi risulta più o meno ragionevole, alla luce dei dati e del modello statistico.

Esempio. *Nascite.* Dalle pubblicazioni dell'ISTAT si ricava che nella regione Friuli Venezia-Giulia nel 1991 si sono registrati 9127 nati vivi, di cui 4608 maschi e 4519 femmine.

Si assume che il numero di maschi osservati nelle 9127 nascite sia descritto una variabile casuale $Bi(9127, p)$, con p la probabilità che il nascituro sia maschio.

Sulla base dei dati e del modello, si vuole verificare l'ipotesi che ci sia equiprobabilità tra i sessi alla nascita: $H_0 : p = 1/2$. $\Diamond$

Esempio. *Antipiretico.* Si vuole indagare se un nuovo antipiretico sia più efficace del migliore attualmente in commercio: H_0 traduce l'ipotesi che l'efficacia dei due prodotti sia equivalente, mentre H_1 attribuisce maggiore efficacia al nuovo prodotto.

Si è interessati a verificare se, alla luce del campione osservato, può essere ragionevole abbandonare l'ipotesi H_0 a favore di H_1. $\Diamond$

Esempio. *Batterie*. Una industria produce batterie elettriche con durata media dichiarata di $\mu = 36$ mesi. L'acquirente vuole accertarsi che la durata media non sia più bassa di quella dichiarata. Si osserva la durata di $n = 40$ batterie.

Si considera, di fatto, un campione casuale semplice $X_1, \ldots, X_{40}$ da una popolazione $N(\mu, \sigma^2)$, con media μ ignota e varianza $\sigma^2 = 9$ supposta nota.

L'acquirente vuole verificare l'ipotesi $H_0 : \mu = 36$ contro l'alternativa, per lui sfavorevole, $H_1 : \mu < 36$.

Dal momento che si vuole fare inferenza su μ, si considera, come statistica campionaria, la media campionaria $\bar{X}_{40}$: se si osservano *valori piccoli* (rispetto a $\mu = 36$) per $\bar{X}_{40}$ si è portati a rifiutare H_0.

Anche se questa procedura di decisione sembra condivisibile, non è chiaro il significato concreto dell'espressione *valori piccoli*.

Si può pensare di definire una **soglia critica** $c \leq 0$, in modo tale da **rifiutare** H_0 **se si verifica l'evento** $(\bar{X}_{40} - 36)/(3/\sqrt{40}) \leq c$, cioè se il valore osservato di $\bar{X}_{40}$ è significativamente più piccolo di 36.

Si determina c di modo che risulti fissata ad un valore sufficientemente piccolo, ad esempio $\alpha = 0.05$, la probabilità di rifiutare H_0, se è vera, cioè

$$P_0\left(\frac{\bar{X}_{40} - 36}{3/\sqrt{40}} \leq c\right) = 0.05,$$

dove P_0 indica la probabilità sotto H_0, cioè nell'ipotesi che $\mu = 36$.

Poiché, sotto H_0, $(\bar{X}_{40} - 36)/(3\sqrt{40}) \sim N(0, 1)$, si conclude che $c = -z_{0.05} = -1.645$, con $z_{0.05}$ il **valore critico** di livello 0.05.

Quindi, se si osserva $\bar{x}_{40} = 35$, si ha $(\bar{x}_{40} - 36)/(3/\sqrt{40}) = -2.11$ e si rifiuta H_0 a favore di H_1. Si tollera una probabilità di rifiutare erroneamente H_0 pari a 0.05.

La differenza tra la media osservata $\bar{x}_{40} = 35$ e il valore medio definito dall'ipotesi nulla $\mu = 36$ è stata valutata rispetto allo standard error della media campionaria $3/\sqrt{40}$.

In alternativa, avendo osservato $\bar{x}_{40} = 35$, si può essere interessati a valutare se questo **risultato campionario** risulta **in accordo con l'ipotesi nulla**.

A tal fine si potrebbe calcolare la probabilità che si verifichi un evento uguale a quello osservato, o anche più a favore di H_1, quando l'ipotesi H_0 è vera; più precisamente,

$$P_0(\bar{X}_{40} \leq 35) = P_0\left(\frac{\bar{X}_{40} - 36}{3/\sqrt{40}} \leq \frac{35 - 36}{3/\sqrt{40}}\right)$$

$$= P_0(Z \leq -2.11) = 0.017,$$

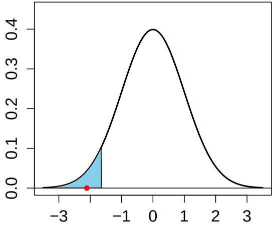
essendo $Z \sim N(0, 1)$.

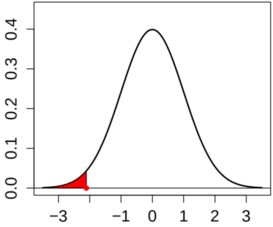
Quindi, nonostante l'apparenza, il valore $\bar{x}_{40} = 35$ è un valore anomalmente piccolo per $\bar{X}_{40}$, se fosse vera H_0.

C'è una *bassa conformità tra il campione osservato e l'ipotesi nulla*, nella direzione dell'ipotesi alternativa.

Nel grafico di sinistra viene considerata la soglia critica, individuata da un'area sulla coda destra pari a 0.05 (in azzurro), e il valore osservato della media campionaria standardizzata supponendo vera H_0 (punto rosso).

Nel grafico di destra viene individuata in rosso la probabilità che la media campionaria standardizzata, supponendo vera H_0, assuma un valore minore o uguale a quello osservato (punto rosso).





◊

Sommario

Statistica test

Nel seguito, non si svilupperà la teoria della verifica di ipotesi. Si forniranno delle indicazioni pratiche, utili per la costruzione di test di ipotesi per particolari problemi statistici.

Dato un campione casuale (semplice) $X_1, \ldots, X_n$ e le ipotesi H_0 e H_1, riferite al parametro ignoto θ, si chiama **statistica test** una statistica campionaria $T = t(X_1, \ldots, X_n)$ che permette di evidenziare se sia più ragionevole non rifiutare H_0 o rifiutare H_0 in favore di H_1.

In genere, si sceglie come statistica test uno *stimatore per* θ (o *una sua trasformata*) per il quale risulta nota, in forma esatta o approssimata, la distribuzione di probabilità (per lo meno sotto H_0).

Come detto in precedenza, un test statistico non una procedura libera da errori.

Si possono individuare due tipi di errore:

- rifiutare H_0 quando è vera: **errore di I tipo**;
- non rifiutare H_0 quando è falsa: **errore di II tipo**.

	H_0 vera	H_1 vera
Rifiuto H_0	Errore I tipo	OK
Accetto H_0	OK	Errore II tipo

Le probabilità di commettere i due tipi di errore sono indicate, rispettivamente, con α e β:

$$\begin{aligned} \alpha &= P(\text{errore di I tipo}) = P(\text{rifiutare } H_0 \text{ quando è vera}) \\ \beta &= P(\text{errore di II tipo}) = P(\text{non rifiutare } H_0 \text{ quando è falsa}). \end{aligned}$$

La probabilità $1-\alpha$ di non rifiutare H_0 quando è vera è detta **livello di protezione** del test, mentre la probabilità $1-\beta$ di rifiutare H_0 quando è falsa è detta **potenza** del test.

Livello di significatività e regione di rifiuto

Le probabilità α e β sono interdipendenti e non è possibile minimizzarle contemporaneamente.

In genere, si fissa la probabilità d'errore di I tipo α, detta **livello di significatività** (ad esempio, $\alpha = 0.1, 0.05, 0.01, 0.001$), e si cerca il test che, a parità di α, presenta potenza $1 - \beta$ più elevata. A parità di altre condizioni, la potenza cresce con n.

La *procedura* che verrà usualmente adottata per definire un test statistico per l'ipotesi nulla $H_0 : \theta = \theta_0$ è la seguente:

- si individua una opportuna statistica test T, per la quale risulta nota, in forma esatta o approssimata, la distribuzione di probabilità sotto H_0;
- si fissa il livello di significatività α;
- tenendo conto di tale vincolo su α, si individuano le **regioni di rifiuto** (**regione critica**) R_α e **di accettazione** A_α per H_0;
- se il valore osservato t^{oss} di T è in R_α (A_α) si rifiuta (non si rifiuta) H_0. In caso di rifiuto il test è detto significativo al $100\alpha\%$.

Come in precedenza, si indica con P_0 la probabilità calcolata supponendo vera H_0 (sotto H_0).

Si considerano i seguenti casi:

1) $H_1 : \theta > \theta_0$ (*alternativa unilaterale destra*)

Se valori grandi per T non sono conformi all'ipotesi nulla H_0, nella direzione di H_1, la regione di rifiuto sarà

$$R_\alpha = \{t \in S_T \,:\, t \geq c_\alpha\},$$

con c_α tale che $P_0(T \geq c_\alpha) = \alpha$.

2) $H_1 : \theta < \theta_0$ (*alternativa unilaterale sinistra*)

Se valori piccoli per T non sono conformi all'ipotesi nulla H_0, nella direzione di H_1, la regione di rifiuto sarà

$$R_\alpha = \{t \in S_T \,:\, t \leq d_\alpha\},$$

con d_α tale che $P_0(T \leq d_\alpha) = \alpha$.

3) $H_1 : \theta \neq \theta_0$ (*alternativa bilaterale*)

Se sia valori piccoli che valori grandi per T non sono conformi all'ipotesi nulla H_0, nella direzione di H_1, la regione di rifiuto sarà

$$R_\alpha = \{t \in S_T \, : \, t \leq d_{\alpha/2} \text{ oppure } t \geq c_{\alpha/2}\},$$

con $d_{\alpha/2}$ tale che $P_0(T \leq d_{\alpha/2}) = \alpha/2$ e $c_{\alpha/2}$ tale che $P_0(T \geq c_{\alpha/2}) = \alpha/2$.

Le regioni di accettazione A_α si individuano in modo speculare.

R_α e A_α determinano una partizione del supporto S_T della statistica test T e individuano la soglia o le soglie critiche, per i valori osservati di T, oltre le quali si rifiuta H_0.

Nel caso 1) e nel caso 2) il test è detto, rispettivamente, **test unilaterale destro e sinistro di livello** α, mentre nel caso 3) si parla di **test bilaterale di livello** α.

P-value

Non è sempre soddisfacente concludere la procedura di verifica di ipotesi con un semplice rifiuto o non rifiuto dell'ipotesi nulla. Per motivare tale decisione può essere utile specificare la forza del risultato ottenuto.

Si può quindi cercare di *quantificare il grado di conformità tra i dati campionari e l'ipotesi nulla*.

Si definisce **livello di significatività osservato** (**P-value**) il più piccolo livello di significatività, indicato con α^{oss}, che conduce al rifiuto di H_0.

Considerando i tre casi analizzati precedenza, indicato con t^{oss} il valore osservato di T, si ha che:

se 1) $H_1 : \theta > \theta_0$ allora $\alpha^{oss} = P_0(T \geq t^{oss})$;

se 2) $H_1 : \theta < \theta_0$ allora $\alpha^{oss} = P_0(T \leq t^{oss})$;

se 3) $H_1 : \theta \neq \theta_0$ allora $\alpha^{oss} = 2 \min\{P_0(T \leq t^{oss}), P_0(T \geq t^{oss})\}$.

Quindi, α^{oss} corrisponde alla probabilità che si verifichi un evento uguale a quello osservato, o anche più a favore di H_1, quando l'ipotesi H_0 è vera.

Ovviamente, $\alpha^{oss} \in [0, 1]$ e, se si osserva un valore per α^{oss} vicino a zero, allora l'osservazione t^{oss} risulta anomala per T se fosse vera H_0 (ciò indica una scarsa conformità tra dati e H_0).

Usualmente, avendo a disposizione il livello di significatività osservato, si fanno le seguenti considerazioni:

- se $\alpha^{oss} \leq 0.01$, si rifiuta H_0 con sicurezza;
- se $0.01 < \alpha^{oss} \leq 0.05$, si rifiuta H_0 con riserva;
- se $0.05 < \alpha^{oss} \leq 0.10$, c'è incertezza sulla decisione;
- se $\alpha^{oss} > 0.10$, non si rifiuta H_0.

In ogni caso, la decisione finale sarà sempre basata anche su considerazioni legate al contesto applicativo.

Per campioni di dimensione elevata la decisione di rifiutare H_0 si considera solo se α^{oss} è decisamente piccolo, ad esempio se $\alpha^{oss} \leq 0.01$.

Il calcolo del livello di significatività osservato, in molti casi, non è agevole, ma viene fornito automaticamente dalle procedure di verifica di ipotesi implementate nei pacchetti statistici.

Considerando il livello di significatività osservato, è possibile specificare la seguente **regola di decisione** tra H_0 e H_1, che risulta equivalente alla precedente, basata sulla regione di rifiuto.

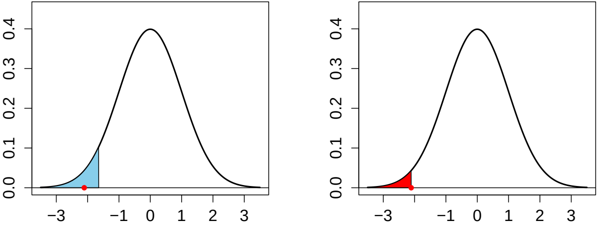
Avendo fissato il livello di significatività α e calcolato il livello di significatività osservato α^{oss},

- se $\alpha^{oss} \leq \alpha$, allora $t^{oss} \in R_\alpha$ e **si rifiuta** H_0;
- se $\alpha^{oss} > \alpha$, allora $t^{oss} \in A_\alpha$ e **non si rifiuta** H_0.

Esempio. *Batterie* (continua). Avendo osservato $\bar{x}_n = 35$, ci si chiede se la differenza con $\mu = 36$ (ipotesi nulla) sia abbastanza elevata da portare al rifiuto di H_0. Si considera, come statistica test, la media campionaria standardizzata sotto H_0; si osserva $z^{oss} = -2.11$.

Poiché l'alternativa H_1 è unilaterale sinistra, il P-value è pari a $\alpha^{oss} = P_0(\bar{X}_n \leq 35) = P_0(Z \leq -2.11) = 0.017$ e corrisponde all'area in rosso nel grafico di destra. Se $\alpha = 0.05$, che è l'area in azzurro nel grafico di sinistra, si individua come soglia di rifiuto (inferiore) il valore -1.645.

Entrambe le regole di decisione portano al rifiuto di H_0, se $\alpha = 0.05$.



◊

Sommario

Verifica di ipotesi per la media di una popolazione normale

Sia $X_1, \ldots, X_n$ un campione casuale semplice da una popolazione $N(\mu, \sigma^2)$, con *varianza* σ^2 *nota*. Dato un valore μ_0 per la media, si vuole verificare l'ipotesi nulla

$$H_0 : \mu = \mu_0$$

ad un livello di significatività α fissato. La **statistica test** è la **media campionaria standardizzata**

$$Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$$

che, sotto H_0, ha distribuzione $N(0,1)$. Il test è chiamato **test z**.

La **regione di rifiuto** R_α si determina considerando il vincolo α sulla probabilità dell'errore di I tipo e corrisponde a valori per Z che sono sintomo di allontanamento da H_0 nella direzione di H_1.

Si considerano i seguenti casi:

1) $H_1 : \mu > \mu_0$ (*alternativa unilaterale destra*)

Poiché valori grandi per Z non sono conformi all'ipotesi nulla H_0, nella direzione di H_1, si ha che

$$R_\alpha = \{z \in \mathbf{R} \, : \, z \geq z_\alpha\},$$

con z_α valore critico tale che $P_0(Z \geq z_\alpha) = \alpha$.

2) $H_1 : \mu < \mu_0$ (*alternativa unilaterale sinistra*)

Poiché valori piccoli per Z non sono conformi all'ipotesi nulla H_0, nella direzione di H_1, si ha che

$$R_\alpha = \{z \in \mathbf{R} \, : \, z \leq -z_\alpha\},$$

con z_α valore critico tale che $P_0(Z \geq z_\alpha) = \alpha$, per la simmetria della funzione di densità normale.

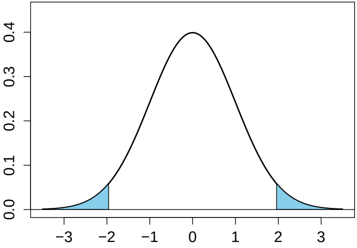
3) $H_1 : \mu \neq \mu_0$ (*alternativa bilaterale*)

Poiché sia valori grandi che valori piccoli per Z non sono conformi all'ipotesi nulla H_0, nella direzione di H_1, si ha che

$$R_\alpha = \{z \in \mathbf{R} \,:\, z \leq -z_{\alpha/2} \text{ o } z \geq z_{\alpha/2}\} = \{z \in \mathbf{R} \,:\, |\, z \,| \geq z_{\alpha/2}\},$$

con $z_{\alpha/2}$ valore critico tale che $P_0(Z \geq z_{\alpha/2}) = \alpha/2$.

In tutti e tre i casi si ha che $P(\text{rifiutare } H_0 \text{ se è vera}) = \alpha$ e R_α è data da valori riferiti alla/e coda/e della distribuzione di probabilità di Z sotto H_0 (poco probabili se H_0 è vera). Per il caso 3), con $\alpha = 0.05$, la regione di rifiuto è individuata dalle due code simmetriche con area 0.025



È possibile determinare il **livello di significatività osservato** (**P-value**), con riferimento alle tre tipologie di ipotesi alternativa.

Essendo z^{oss} il valore osservato della statistica test Z, si ha che:

se 1) $H_1 : \mu > \mu_0$, $\alpha^{oss} = P_0(Z \geq z^{oss}) = 1 - \Phi(z^{oss})$;

se 2) $H_1 : \mu < \mu_0$, $\alpha^{oss} = P_0(Z \leq z^{oss}) = \Phi(z^{oss})$;

se 3) $H_1 : \mu \neq \mu_0$, $\alpha^{oss} = 2\min\{\Phi(z^{oss}), 1 - \Phi(z^{oss})\}$.

Esempio. *Contenitori*. Si desidera valutare la resistenza alla pressione interna di una certa tipologia di contenitori di vetro. In particolare si vuole verificare se la resistenza media supera 175 psi.

Per tale fenomeno è ragionevole assumere il modello $N(\mu, \sigma^2)$, dove, sulla base di esperienze precedenti, si pone $\sigma^2 = 100$. Il sistema di ipotesi che si considera è

$$H_0 : \mu = 175 \quad \textit{vs} \quad H_1 : \mu > 175$$

Viene selezionato un campione casuale semplice di $n = 25$ contenitori che vengono sottoposti a pressione idrostatica fino alla rottura, al fine di misurarne la resistenza.

Si ottiene una pressione media di $\bar{x}_{25} = 181.28$, e pertanto

$$z = \frac{\bar{x}_{25} - 175}{10/\sqrt{25}} = 3.14\,.$$

Avendo fissato $\alpha = 0.05$, si ottiene $z_{0.05} = 1.645$, quindi si rifiuta H_0 essendo $3.14 > 1.645$. Si dice quindi che il test è risultato significativo contro H_0 ad un livello di significatività $\alpha = 0.05$.

Il P-value corrisponde a

$$\alpha^{oss} = 1 - \Phi(3.14) = 0.00085;$$

quindi si rifiuta H_0 con sicurezza, poiché i dati evidenziano un accordo molto scarso con l'ipotesi nulla. $\Diamond$

Esempio. *Bilancia*. Si vuole verificare se una bilancia è stata calibrata di modo tale che, in media, fornisca il vero valore del peso dell'oggetto che si considera.

Si considera un peso di 25 gr e si assume che una generica pesata sia descritta da una variabile casuale $N(\mu, \sigma^2)$; è nota la precisione della bilancia, che corrisponde ad una varianza $\sigma^2 = 45$

Si effettuano $n = 10$ pesate e si ottiene $\bar{x}_{10} = 21.98$. Il sistema di ipotesi che si considera è

$$H_0 : \mu = 25 \quad \textit{vs} \quad H_1 : \mu \neq 25$$

e si assume un livello di significatività $\alpha = 0.05$. In questo caso, $\alpha/2 = 0.025$, $z_{0.025} = 1.96$ e $R_{0.05} = \{z \in \mathbf{R} : | z | \geq 1.96\}$, e il valore della statistica test è $z = (21.98 - 25)/(6.71/\sqrt{10}) = -1.42$.

Quindi l'ipotesi H_0 non viene rifiutata. Tuttavia il grado di conformità tra H_0 e i dati non è molto alto dal momento che $\alpha^{oss} = 2 \min\{\Phi(-1.42), 1 - \Phi(-1.42)\} = 0.155$. $\Diamond$

Se la *varianza* σ^2 è *ignota*, si utilizza al suo posto la varianza campionaria corretta S_c^2 e si considera, come **statistica test**, la **media campionaria studentizzata**

$$T = \frac{\bar{X}_n - \mu_0}{S_c/\sqrt{n}},$$

che, sotto H_0, ha distribuzione $t(n-1)$. Il test è detto **test t**.

Fissato il livello di significatività α e indicato con t un generico valore per T, per la regione di rifiuto si evidenziano i seguenti casi:

se 1) $H_1 : \mu > \mu_0$, $R_\alpha = \{t \in \mathbf{R} \, : \, t \geq t_\alpha\}$;

se 2) $H_1 : \mu < \mu_0$, $R_\alpha = \{t \in \mathbf{R} \, : \, t \leq -t_\alpha\}$;

se 3) $H_1 : \mu \neq \mu_0$, $R_\alpha = \{t \in \mathbf{R} \, : | \, t \, | \geq t_{\alpha/2}\}$,

con t_α e $t_{\alpha/2}$ valori critici associati alla distribuzione $t(n-1)$.

I **P-value** si ottengono come nel caso di varianza nota, sostituendo $\Phi(\cdot)$ con la funzione di ripartizione di una $t(n-1)$.

Esempio. *Spese mediche.* Sulla base dell'esperienza passata si assume che l'importo in euro delle spese mediche per famiglia in un mese segue una distribuzione $N(66, \sigma^2)$, con σ^2 ignota.

Si considerano le spese mensili di un campione casuale semplice di $n = 56$ famiglie e si osserva una spesa media di $\bar{x}_{56} = 60.57$ euro, con $s_c^2 = 15.23$.

Per valutare se la spesa media sia effettivamente diminuita si definisce il sistema di ipotesi è

$$H_0 \, : \, \mu = 66 \quad \textit{vs} \quad H_1 \, : \, \mu < 66.$$

Essendo ignota la varianza, si considera un test t con un livello di significatività $\alpha = 0.05$. In questo caso, considerando la distribuzione $t(55)$, $t_{0.05} = 1.67$ e $R_{0.05} = \{t \in \mathbf{R} \, : \, t \leq -1.67\}$; il valore della statistica test è $t = (60.57 - 66)/(3.9/\sqrt{56}) = -10.41$.

Quindi l'ipotesi H_0 viene rifiutata. Il grado di conformità tra H_0 e i dati è molto scarso, infatti $\alpha^{oss} = F_T(-10.2) = 6.51 \cdot 10^{-15}$. $\Diamond$

Esempio. *Bilancia* (continua). Si considera un problema analogo al precedente, con l'unica differenza che la varianza σ^2 è ignota.

Si effettuano le $n = 10$ misurazioni e si osserva $\bar{x}_{10} = 21.98$ e $\sum_{i=1}^{10} x_i^2/10 = 519.9$. Si ha

$$H_0 \,:\, \mu = 25 \quad \textit{vs} \quad H_1 \,:\, \mu \neq 25$$

e si assume un livello di significatività $\alpha = 0.05$. Poiché

$$s^2 = \frac{1}{10}\sum_{i=1}^{10} x_i - \bar{x}_{10}^2 = 36.78, \quad s_c^2 = \frac{10}{9}\, s^2 = 40.86,$$

il valore della statistica test è $t = (21.98 - 25)/(6.39/\sqrt{10}) = -1.49$. In questo caso, considerando una distribuzione $t(9)$, $t_{0.025} = 2.262$ e $R_{0.05} = \{t \in \mathbf{R} : \mid t \mid \geq 2.262\}$; l'ipotesi H_0 non viene rifiutata.

Tuttavia il grado di conformità tra H_0 e i dati non è molto alto dal momento che $\alpha^{oss} = 2\min\{F_T(-1.49), 1 - F_T(-1.49)\} = 0.17$. $\Diamond$

Verifica di ipotesi per la media di una popolazione qualsiasi

Sia $X_1, \ldots, X_n$ un campione casuale semplice da una popolazione non normale con media μ. Si vuole verificare l'ipotesi nulla

$$H_0 \,:\, \mu = \mu_0,$$

ad un livello di significatività α fissato.

Per il teorema limite centrale, se la numerosità campionaria n è sufficientemente elevata, si determina, con relativa facilità, un test per μ con *livello di significatività α approssimato*.

Come **statistica test** si considera la **media campionaria standardizzata**

$$Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$$

che, sotto H_0, ha distribuzione approssimata $N(0,1)$. Se il valore di σ non è noto sotto H_0, lo si può stimare con $\hat{\sigma}$.

La **regione di rifiuto** e il **livello di significatività osservato (approssimato)** si definiscono come nel caso del test sulla media di una popolazione normale con varianza nota.

Si considerano i seguenti casi interessanti:

- se $X_i \sim Ber(p)$, $i = 1, \ldots, n$, allora $H_0 : p = p_0$ e la statistica test per la media p si specifica con $p_0 = \mu_0$, $\hat{p} = \bar{X}_n =$ e $\sigma = \sqrt{p_0(1 - p_0)}$ (*test su una proporzione*);
- se $X_i \sim P(\lambda)$, $i = 1, \ldots, n$, allora $H_0 : \lambda = \lambda_0$ e la statistica test per la media λ si specifica con $\lambda_0 = \mu_0$, $\hat{\lambda} = \bar{X}_n$ e $\sigma = \sqrt{\lambda_0}$.

Con campioni bernoulliani, se H_1 è bilaterale, si può anche utilizzare la statistica test Z^2 che, sotto H_0, ha distribuzione approssimata $\chi^2(1)$.

Per distribuzioni non normali esistono anche metodi più accurati, generalmente implementati nei software statistici, che risultano utili per il calcolo di test sulla media nel caso di *piccoli campioni*.

Esempio. *Carta di credito.* È noto che il 20% dei possessori di carta di credito la utilizza abitualmente per gli acquisti.

Dopo una campagna mirata a pubblicizzare l'uso della carta di credito, in un campione di 120 individui, 36 dichiarano di utilizzarla per gli acquisti. Ci si chiede se la campagna pubblicitaria ha modificato la proporzione iniziale di utilizzatori.

Il campione osservato, di dimensione $n = 120$, proviene da una popolazione $Ber(p)$, con p ignoto. Si considerano le ipotesi

$$H_0 : p = 0.2 \quad \textit{vs} \quad H_1 : p \neq 0.2$$

con un livello di significatività (approssimato) $\alpha = 0.05$. In questo caso, $\hat{p} = 36/120 = 0.3$, $\sqrt{p_0(1 - p_0)} = 0.4$, $z_{0.025} = 1.96$ e $R_{0.05} = \{z \in \mathbf{R} : | z | \geq 1.96\}$, mentre il valore della statistica test è $z = (0.3 - 0.2)/(0.4/\sqrt{120}) = 2.74$.

Quindi l'ipotesi H_0 viene rifiutata e il test risulta significativo al livello (approssimato) $\alpha = 0.05$. Il livello di significatività osservato (approssimato) è $\alpha^{oss} \doteq 2 \min\{\Phi(2.74), 1 - \Phi(2.74)\} = 0.006$. $\Diamond$

Esempio. *Chip.* La direzione di un reparto produttivo afferma che il numero medio di circuiti integrati difettosi prodotti ogni giorno è pari a 25, ma l'osservazione di un campione casuale semplice, relativo a $n = 75$ giorni, ha registrato un totale di 2110 chip difettosi.

Si utilizza il modello $P(\lambda)$ per descrivere il numero giornaliero di chip difettosi e si considerano le ipotesi

$$H_0 : \lambda = 25 \quad \textit{vs} \quad H_1 : \lambda > 25$$

con un livello di significatività (approssimato) $\alpha = 0.01$. In questo caso, $\hat{\lambda} = 2110/75 = 28.13$, $\sqrt{\lambda_0} = 5$, $z_{0.01} = 2.33$ e la regione di rifiuto è $R_{0.01} = \{z \in \mathbf{R} \, : \, z \geq 2.33\}$, mentre il valore della statistica test è $z = (28.13 - 25)/(5/\sqrt{75}) = 5.43$.

Quindi l'ipotesi H_0 viene rifiutata e il test risulta significativo al livello (approssimato) $\alpha = 0.01$. Il livello di significatività osservato (approssimato) è $\alpha^{oss} \doteq 1 - \Phi(5.43) = 2.86 \cdot 10^{-8}$. $\Diamond$

Esempio. *Lampadina.* La durata in ore di un certo tipo di lampadine può essere descritta dal modello $Esp(\lambda)$. L'esperienza passata indica che $\lambda = 0.005$, cioè che la durata media è $\mu = 200$.

Considerando un campione casuale semplice di dimensione $n = 70$, si ottiene una durata media osservata di $\bar{x}_{70} = 249.5$. Si desidera verificare le ipotesi

$$H_0 : \mu = 200 \quad \textit{vs} \quad H_1 : \mu > 200$$

con un livello di significatività (approssimato) $\alpha = 0.01$. In questo caso, sotto H_0, $\sigma = \mu_0 = 200$, $z_{0.01} = 2.33$ e la regione di rifiuto è $R_{0.01} = \{z \in \mathbf{R} \,:\, z \geq 2.33\}$, mentre il valore della statistica test è $z = (249.5 - 200)/(200/\sqrt{70}) = 2.07$.

Quindi l'ipotesi H_0 non viene rifiutata. Tuttavia il grado di conformità tra H_0 e i dati è basso, infatti il livello di significatività osservato (approssimato) è $\alpha^{oss} \doteq 1 - \Phi(2.07) = 0.02$, valore leggermente superiore a $\alpha = 0.01$. $\diamondsuit$

Verifica di ipotesi per la varianza di una popolazione normale

Sia $X_1, \ldots, X_n$ un campione casuale semplice da una popolazione $N(\mu, \sigma^2)$. Dato un valore σ_0^2 per la varianza, si vuole verificare l'ipotesi nulla

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

ad un livello di significatività α fissato. La **statistica test** è la **varianza campionaria** S^2, opportunamente modificata,

$$K = \frac{nS^2}{\sigma_0^2}$$

che, sotto H_0, ha distribuzione $\chi^2(n-1)$.

La **regione di rifiuto** R_α si determina considerando il vincolo α sulla probabilità dell'errore di I tipo e corrisponde a valori per K che sono sintomo di allontanamento da H_0 nella direzione di H_1.

Si considerano i seguenti casi:

1) $H_1 : \sigma^2 > \sigma_0^2$ (*alternativa unilaterale destra*)

Poiché valori grandi per K non sono conformi all'ipotesi nulla H_0, nella direzione di H_1, si ha che

$$R_\alpha = \{k \in \mathbf{R}^+ \,:\, k \geq \chi^2_\alpha\},$$

con χ^2_α valore critico tale che $P_0(K \geq \chi^2_\alpha) = \alpha$.

2) $H_1 : \sigma^2 < \sigma_0^2$ (*alternativa unilaterale sinistra*)

Poiché valori piccoli per K non sono conformi all'ipotesi nulla H_0, nella direzione di H_1, si ha che

$$R_\alpha = \{k \in \mathbf{R}^+ \,:\, k \leq \chi^2_{1-\alpha}\},$$

con $\chi^2_{1-\alpha}$ valore critico tale che $P_0(K \geq \chi^2_{1-\alpha}) = 1 - \alpha$.

3) $H_1 : \sigma^2 \neq \sigma_0^2$ (*alternativa bilaterale*)

Poiché sia valori grandi che valori piccoli per K non sono conformi all'ipotesi nulla H_0, nella direzione di H_1, si ha che

$$R_\alpha = \{k \in \mathbf{R}^+ \, : \, k \leq \chi^2_{1-\alpha/2} \text{ o } k \geq \chi^2_{\alpha/2}\},$$

con $\chi^2_{\alpha/2}$ e $\chi^2_{1-\alpha/2}$ valori critici tali che $P_0(K \geq \chi^2_{\alpha/2}) = \alpha/2$ e $P_0(K \geq \chi^2_{1-\alpha/2}) = 1 - \alpha/2$.

Si noti che, a differenza dei test sulla media, nel caso di alternativa bilaterale non è possibile determinare i valori critici invocando la simmetria della distribuzione di probabilità della statistica test.

La statistica test si può ottenere partendo anche dalla varianza campionaria corretta S_c^2, infatti

$$K = \frac{(n-1)S_c^2}{\sigma_0^2}.$$

È possibile determinare il **livello di significatività osservato** (**P-value**), con riferimento alle tre tipologie di ipotesi alternativa.

Essendo k^{oss} il valore osservato della statistica test K, si ha che:

se 1) $H_1 : \sigma^2 > \sigma_0^2$, $\alpha^{oss} = P_0(K \geq k^{oss}) = 1 - F_K(k^{oss})$;

se 2) $H_1 : \sigma^2 < \sigma_0^2$, $\alpha^{oss} = P_0(K \leq k^{oss}) = F_K(k^{oss})$;

se 3) $H_1 : \sigma^2 \neq \sigma_0^2$, poiché K non ha densità simmetrica,

$$\alpha^{oss} = 2 \min \{F_K(k^{oss}), 1 - F_K(k^{oss})\},$$

con $F_K(\cdot)$ la funzione di ripartizione di una $\chi^2(n-1)$.

Esempio. *Catalizzatore*. Nell'ambito di un processo produttivo, si vuole studiare la quantità, in Kg, di sottoprodotti di una certa reazione chimica in presenza di un catalizzatore. Se il *catalizzatore è attivo*, la quantità di sottoprodotti è descritta da una $N(\mu, 3.83)$, con μ ignoto.

Si osservano le quantità prodotte in $n = 13$ reazioni chimiche, considerando un *catalizzatore esaurito*. Si vuole verificare

$$H_0 : \sigma^2 = 3.83 \quad \mathit{vs} \quad H_1 : \sigma^2 \neq 3.83$$

con un livello di significatività $\alpha = 0.05$.

Dai dati si ricava che $\bar{x}_{13} = 11.7$ e $\sum_{i=1}^{13} x_i^2/13 = 142.78$, quindi

$$s^2 = \frac{1}{13}\sum_{i=1}^{13} x_i - \bar{x}_{13}^2 = 5.87.$$

Poiché $\alpha/2 = 0.025$ e, con riferimento a una distribuzione $\chi^2(12)$, $\chi^2_{1-0.025} = 4.404$ e $\chi^2_{0.025} = 23.337$, si ottiene la regione di rifiuto $R_\alpha = \{k \in \mathbf{R}^+ \,:\, k \leq 4.404 \text{ o } k \geq 23.337\}$.

Il valore della statistica test è $k = 13 \cdot 5.87/3.83 = 19.91$ e l'ipotesi H_0 non viene rifiutata. Il P-value è

$$\alpha^{oss} = 2\min\{F_K(19.91), 1 - F_K(19.91)\} = 0.137,$$

essendo $K \sim \chi^2(12)$. Non c'è sufficiente evidenza empirica per affermare che, se il catalizzatore è esaurito, si ha una modificazione di σ^2. $\Diamond$

Test per il confronto tra medie di popolazioni normali

Quando si hanno due o più campioni, ci si può chiedere se provengono dalla stessa popolazione o, in particolare, se provengono da popolazioni che hanno qualche caratteristica simile.

Si presenteranno alcuni test di ipotesi che permettono di confrontare valori di sintesi calcolati su due o più campioni.

L'obiettivo è fare inferenza sulla differenza tra parametri delle corrispondenti popolazioni, quali ad esempio, medie, varianze e proporzioni.

Esempio. *Inquinamento.* Per confrontare, in media, l'efficacia di due diverse metodiche, A e B, per contenere l'inquinamento atmosferico si sono analizzati i fumi prodotti da una certa industria.

Si è misurata la quantità di pulviscolo inquinante in g/min, con riferimento a due campioni indipendenti di fumi, ottenuti utilizzando, rispettivamente, il dispositivo anti-inquinante A e B. $\Diamond$

Siano date *due popolazioni normali indipendenti* con distribuzione di probabilità $N(\mu_1, \sigma_1^2)$ e $N(\mu_2, \sigma_2^2)$, con *varianze note*.

Si estraggono due campioni casuali semplici di dimensione n_1 e n_2 dalla prima e dalla seconda popolazione, rispettivamente.

Si vuole verificare l'ipotesi che le due popolazioni abbiano la stessa media e quindi si considera l'ipotesi nulla

$$H_0 \,:\, \mu_1 = \mu_2 \quad (\mu_1 - \mu_2 = 0)$$

ad un livello di significatività α fissato. La **statistica test** è la **differenza tra le medie campionarie** $\bar{X}_1$ e $\bar{X}_2$, calcolate sui due campioni, opportunamente standardizzata,

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

che, sotto H_0, ha distribuzione $N(0, 1)$.

La **regione di rifiuto** R_α e il **P-value**, in corrispondenza alle varie tipologie di ipotesi alternativa, coincidono con quelli individuati per la verifica di ipotesi sulla media di una popolazione normale.

Se le *varianze* sono *ignote e uguali* (*omoschedasticità*), la **statistica test** è la **differenza tra le medie campionarie studentizzata**,

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2(1/n_1 + 1/n_2)}}$$

che, sotto H_0, ha distribuzione $t(n_1 + n_2 - 2)$. Per la varianza comune $\sigma^2 = \sigma_1^2 = \sigma_2^2$ si è utilizzato lo **stimatore combinato**

$$S_p^2 = \frac{(n_1 - 1)S_{c1}^2 + (n_2 - 1)S_{c2}^2}{n_1 + n_2 - 2},$$

con S_{c1}^2 e S_{c2}^2 le varianze campionarie corrette basate sui due campioni.

Il test è detto **test t a due campioni**; la **regione di rifiuto** R_α e il **P-value**, in corrispondenza alle varie tipologie di ipotesi alternativa, si ottengono in modo analogo a quanto visto per il test t.

Se le *varianze* sono *ignote e diverse* (*eteroschedasticità*), si utilizza il **test di Welch**, con **statistica test**,

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_{c1}^2/n_1 + S_{c2}^2/n_2}}$$

con distribuzione, sotto H_0, complicata ma disponibile in molti software statistici, che forniscono automaticamente il **P-value**.

Può essere utile premettere al test di ipotesi una analisi grafica comparata dei due insiemi di dati campionari.

Esempio. *Inquinamento* (continua). Si considerano i due campioni casuali semplici, di dimensione $n_1 = n_2 = 180$, indipendenti, riferiti ai due dispositivi anti-inquinanti A e B. Si ottiene $\bar{x}_1 = 15.00$, $\bar{x}_2 = 15.02$, $s_{c1}^2 = 0.046$, $s_{c2}^2 = 0.099$.

Per verificare le ipotesi

$$H_0 : \mu_1 = \mu_2 \quad \textit{vs} \quad H_1 : \mu_1 \neq \mu_2,$$

con un livello di significatività $\alpha = 0.05$, si considera il test t a due campioni.

Visto che $\alpha/2 = 0.025$ e, per una distribuzione $t(358)$, il corrispondente valore critico è $t_{0.025} = 1.96$, si ottiene la regione di rifiuto $R_\alpha = \{t \in \mathbf{R} : \mid t \mid \geq 1.96\}$.

Poiché $s_p^2 = (0.046 \cdot 179 + 0.099 \cdot 179)/358 = 0.0727$ e la statistica test è $t = (15 - 15.02)/\sqrt{0.0727 \cdot (1/180 + 1/180)} = -0.662$, si conclude che l'ipotesi H_0 non va rifiutata.

Inoltre, il P-value è $\alpha^{oss} = 2 \min\{F_T(-0.662), 1 - F_T(-0.662)\} = 0.51$.

Si noti che le due varianze campionarie sembrano sensibilmente diverse. Se tale differenza risulta statisticamente significativa, si deve utilizzare il test di Welch. $\Diamond$

Esempio. *Vernici*. Si desidera confrontare il tempo di asciugatura in ore di due diverse vernici. Si considerano due campioni casuali semplici di $n_1 = 17$ e $n_2 = 18$ verniciature, rispettivamente.

Si ottengono i seguenti risultati, in ore: $\bar{x}_1 = 6.20$, $\bar{x}_2 = 8.55$, $s^2_{c1} = 4.25$, $s^2_{c2} = 4.28$. Per verificare le ipotesi

$$H_0 : \mu_1 = \mu_2 \quad \textsf{\textit{vs}} \quad H_1 : \mu_1 \neq \mu_2,$$

con un livello di significatività $\alpha = 0.01$, sotto ipotesi di normalità e omeschedasticità, si considera il test t a due campioni.

Visto che per una distribuzione $t(33)$, il corrispondente valore critico è $t_{0.005} = 2.73$, si ottiene $R_\alpha = \{t \in \mathbf{R} : \mid t \mid \geq 2.73\}$.

Poiché $s^2_p = (4.25 \cdot 16 + 4.28 \cdot 17)/33 = 4.27$ e la statistica test è $t = (6.20 - 8.55)/\sqrt{4.27 \cdot (1/17 + 1/18)} = -3.362$, l'ipotesi H_0 va rifiutata. I tempi medi di asciugatura sono significativamente diversi.

Inoltre, $\alpha^{oss} = 2 \min\{F_T(-3.362), 1 - F_T(-3.362)\} = 0.002$. $\Diamond$

L'utilizzazione dei test ora presentati richiede la *verifica preliminare dell'ipotesi di normalità* ed inoltre una valutazione in merito alla eventuale *omoschedasticità*.

Con tale obiettivo si possono utilizzare *metodi grafici*, come ad esempio, istogrammi, boxplot, q-q plot. Esistono anche metodi più formali basati su test di ipotesi; ad esempio, i **test di normailtà**, che rientrano nella famiglia dei *test di goodness of fit*.

Per la verifica dell'omoschedasticità, in seguito si presenterà un *opportuno test statistico*.

Nel caso del **confronto tra medie di popolazioni non normali**, se la *numerosità dei due campioni* è *abbastanza elevata*, si possono utilizzare le statistiche test viste in precedenza che, sotto H_0, hanno distribuzione approssimata $N(0, 1)$.

È possibile operare un **confronto tra medie di** due **popolazioni** anche quando **non** sono **indipendenti**. Ad esempio, quando si effettuano osservazioni ripetute sulle medesime unità statistiche (**dati appaiati**).

Esempio. *Pressione*. Si desidera valutare l'efficacia di un farmaco per abbassare la pressione sistolica.

A tale scopo, si misura le pressione ad un campione casuale semplice di $n = 15$ pazienti, *prima* e *dopo* il trattamento con il farmaco in esame. $\Diamond$

La verifica di ipotesi sulle medie di popolazioni, in ipotesi di dati appaiati, si effettua con metodi simili a quelli già studiati, senza trascurare la relazione tra le due popolazioni in esame.

Siano date *due popolazioni normali dipendenti*, con distribuzione di probabilità $N(\mu_1, \sigma_1^2)$ e $N(\mu_2, \sigma_2^2)$. Si estraggono due campioni casuali semplici (appaiati) di dimensione n dalla prima e dalla seconda popolazione, rispettivamente.

Si calcolano le differenze tra osservazioni corrispondenti; si ha un **campione casuale semplice di variabili casuali differenza** $D_i = X_{1i} - X_{2i}$, $i = 1, \ldots, n$, con media $E(D_i) = \mu_1 - \mu_2 = \mu_d$.

Si vuole verificare l'ipotesi che le due popolazioni abbiano la stessa media, cioè che le variabili casuali differenza abbiano media nulla,

$$H_0 : \mu_d = 0 \quad (\mu_1 = \mu_2)$$

ad un livello di significatività α fissato.

Limitando l'analisi alle differenze, ci si riconduce di fatto ad un problema di verifica di ipotesi sulla media di una popolazione normale.

Quindi si utilizzano le formule del **test t** applicate *al campione formato dalle n differenze* $d_i = x_{1i} - x_{2i}$, $i = 1, \ldots, n$.

Esempio. *Pressione* (continua). La misurazione della pressione sistolica ad un campione casuale semplice di $n = 15$ pazienti, *prima* e *dopo* il trattamento con il farmaco in esame, fornisce i seguenti risultati.

La media delle osservazioni prima del trattamento è $\bar{x}_1 = 144$, la media dopo il trattamento è $\bar{x}_2 = 122.53$, mentre la varianza campionaria corretta calcolata sulle differenze è $s^2_{cd} = 39.41$.

Per verificare le ipotesi

$$H_0 : \mu_d = 0 \quad \textit{vs} \quad H_1 : \mu_d > 0,$$

con un livello di significatività $\alpha = 0.01$, sotto ipotesi di normalità, si considera il test t a un campione applicato alle differenze.

Visto che per una distribuzione $t(14)$, il valore critico è $t_{0.01} = 2.62$, si ha $R_\alpha = \{t \in \mathbf{R} : t > 2.62\}$.

Poiché $t = (144 - 122.53)/\sqrt{39.41/15} = 13.24$ è il valore della statistica test, si rifiuta l'ipotesi H_0; il farmaco è efficace per abbassare la pressione.

Inoltre, il P-value è $\alpha^{oss} = 1 - F_T(13.24) = 1.31 \cdot 10^{-9}$. $\Diamond$

Test per il confronto tra varianze di popolazioni normali

Quando si hanno due o più campioni, ci si può chiedere se provengono da popolazioni con la stessa varianza (ignota), cioè se vale l'**ipotesi di omoschedasticità**.

Lo studio di tale ipotesi può anche essere utile come verifica preliminare all'applicazione di test per il confronto tra medie.

Siano date *due popolazioni normali indipendenti* con distribuzione di probabilità $N(\mu_1, \sigma_1^2)$ e $N(\mu_2, \sigma_2^2)$, con parametri ignoti. Si estraggono due campioni casuali semplici di dimensione n_1 e n_2 dalla prima e dalla seconda popolazione, rispettivamente.

Si vuole verificare l'ipotesi che le due popolazioni abbiano la stessa varianza e quindi si considera l'ipotesi nulla

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad (\sigma_1^2/\sigma_2^2 = 1)$$

ad un livello di significatività α fissato.

La **statistica test** è il **rapporto tra le varianze campionarie corrette** S^2_{c1} e S^2_{c2} calcolate nei due campioni,

$$F = \frac{S^2_{c1}}{S^2_{c2}}$$

che, sotto H_0, ha distribuzione *F* di Fisher $F(n_1 - 1, n_2 - 1)$. Il test ottenuto è chiamato **test F**.

La **regione di rifiuto** R_α si determina considerando il vincolo α e corrisponde a valori per F che sono sintomo di allontanamento da H_0 nella direzione di H_1. Si considerano i seguenti casi:

1) $H_1 : \sigma^2_1 > \sigma^2_2 \quad (\sigma^2_1/\sigma^2_2 > 1)$ (*alternativa unilaterale destra*)

Poiché valori per F molto maggiori di 1 non sono conformi ad H_0, nella direzione di H_1, si ha che

$$R_\alpha = \{f \in \mathbf{R}^+ \,:\, f \geq F_\alpha\},$$

con F_α valore critico tale che $P_0(F \geq F_\alpha) = \alpha$.

2) $H_1 : \sigma_1^2 < \sigma_2^2 \quad (\sigma_1^2/\sigma_2^2 < 1)$ (*alternativa unilaterale sinistra*)

Poiché valori per F molto minori di 1 non sono conformi ad H_0, nella direzione di H_1, si ha che

$$R_\alpha = \{f \in \mathbf{R}^+ \,:\, f \leq F_{1-\alpha}\},$$

con $F_{1-\alpha}$ valore critico tale che $P_0(F \geq F_{1-\alpha}) = 1 - \alpha$.

3) $H_1 : \sigma^2 < \sigma_0^2 \quad (\sigma_1^2/\sigma_2^2 \neq 1)$ (*alternativa bilaterale*)

Poiché valori per F molto minori o molto maggiori di 1 non sono conformi ad H_0, nella direzione di H_1, si ha che

$$R_\alpha = \{f \in \mathbf{R}^+ \,:\, f \leq F_{1-\alpha/2} \text{ o } f \geq F_{\alpha/2}\},$$

con $F_{\alpha/2}$ e $F_{1-\alpha/2}$ valori critici tali che $P_0(F \geq F_{\alpha/2}) = \alpha/2$ e $P_0(F \geq F_{1-\alpha/2}) = 1 - \alpha/2$.

Si noti che, a differenza dei test sulla media, nel caso di alternativa bilaterale non è possibile determinare i valori critici invocando la simmetria della distribuzione di probabilità della statistica test.

È possibile determinare il **livello di significatività osservato** (**P-value**), con riferimento alle tre tipologie di ipotesi alternativa.

Essendo f^{oss} il valore osservato della statistica test F, si ha che:

se 1) $H_1 : \sigma_1^2 > \sigma_2^2$, $\alpha^{oss} = P_0(F \geq f^{oss}) = 1 - F_F(f^{oss})$;

se 2) $H_1 : \sigma_1^2 < \sigma_2^2$, $\alpha^{oss} = P_0(F \leq f^{oss}) = F_F(f^{oss})$;

se 3) $H_1 : \sigma_1^2 \neq \sigma_2^2$, poiché F non ha densità simmetrica,

$$\alpha^{oss} = 2 \min \left\{ F_F(f^{oss}), 1 - F_F(f^{oss}) \right\},$$

con $F_F(\cdot)$ la funzione di ripartizione di una $F(n_1 - 1, n_2 - 1)$.

A differenza dei test sulle medie, i test sulle varianza sono molto sensibili all'assunzione di normalità.

Esempio. *Inquinamento* (continua). Si considerano i dati riferiti alle due metodiche anti-inquinamento. Dai dati campionari si ottiene $s_{c1}^2 = 0.046$, $s_{c2}^2 = 0.099$, con $n_1 = n_2 = 180$.

Per verificare le ipotesi

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \textsf{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2,$$

con un livello di significatività $\alpha = 0.01$, si considera il test F.

Visto che $\alpha/2 = 0.005$ e, per una distribuzione $F(179, 179)$, i valori critici sono $F_{0.005} = 1.47$ e $F_{0.995} = 0.68$, si definisce la regione di rifiuto $R_\alpha = \{f \in \mathbf{R}^+ \,:\, f \leq 0.68 \text{ o } f \geq 1.47\}$.

Poiché il valore della statistica test è $f = 0.046/0.099 = 0.46$, si conclude che l'ipotesi H_0 va rifiutata.

Inoltre, $\alpha^{oss} = 2 \min\{F_F(0.46), 1 - F_F(0.46)\} = 2.98 \cdot 10^{-7}$.

Quindi, essendo le varianze sensibilmente diverse, per confrontare le medie si sarebbe dovuto considerare il test di Welch. ♢

Sommario

Test su proporzioni

In precedenza si è già considerato il caso del *test su una proporzione*, nel caso di campioni di tipo bernoulliano, dove l'obiettivo era confrontare la proporzione campionaria con un valore ipotizzato di riferimento.

Un ulteriore caso che si presenta con molta frequenza nella pratica sperimentale riguarda il confronto tra due o più proporzioni campionarie.

In questo contesto non è disponibile un valore di riferimento teorico o ideale e il confronto coinvolge esclusivamente le proporzioni osservate riferite ai campioni in esame.

Le popolazioni da cui i campioni sono estratti vengono usualmente descritte dal modello bernoulliano o, più in generale, binomiale.

Oltre a ciò si presenteranno procedure per la verifica dell'ipotesi di dipendenza tra due variabili categoriali. Il test ottenuto può anche venir utilizzato per verificare se una variabile categoriale segue una specifica distribuzione di probabilità.

Test per il confronto tra proporzioni

Siano date *due popolazioni bernoulliane indipendenti* con distribuzione di probabilità $Ber(p_1)$ e $Ber(p_2)$, con p_1 e p_2 *proporzioni ignote*.

Si estraggono due campioni casuali semplici di dimensione n_1 e n_2 dalla prima e dalla seconda popolazione, rispettivamente.

Si vuole verificare l'ipotesi che i due campioni provengano dalla stessa popolazione e quindi si considera l'ipotesi nulla

$$H_0 \, : \, p_1 = p_2 \quad (p_1 - p_2 = 0)$$

ad un livello di significatività α fissato. La **statistica test** è la **differenza tra le proporzioni campionarie** $\hat{p}_1 = \bar{X}_1$ e $\hat{p}_2 = \bar{X}_2$, calcolate sui due campioni, opportunamente standardizzata,

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}.$$

Si definisce $\hat{p}$ lo **stimatore combinato** di $p_1 = p_2 = p$ sotto H_0

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2},$$

che corrisponde alla proporzione campionaria calcolata considerando insieme i due campioni.

Nel caso in cui n_1 e n_2 sono sufficientemente elevati, la statistica test Z ha distribuzione approssimata $N(0, 1)$.

La **regione di rifiuto** e il **livello di significatività osservato** (**approssimato**) si definiscono come nel caso del test sulla media di una popolazione non normale.

Se i campioni in esame non sono sufficientemente elevati, esistono soluzioni approssimate più accurate o soluzioni esatte, che sono di solito implementate nei software statistici.

Esempio. *Cerotti.* Ad un gruppo di $n_1 = 224$ fumatori che desiderano smettere di fumare viene prescritto il cerotto alla nicotina, mentre ad altri $n_2 = 245$ fumatori oltre al cerotto si prescrive anche un leggero antidepressivo.

Dopo sei mesi, 40 soggetti del primo gruppo e 87 del secondo avevano smesso di fumare. Ci si chiede se è utile associare cerotto ad antidepressivo.

Si considera il test per il confronto tra proporzioni ad un livello $\alpha = 0.05$, per verificare l'ipotesi $H_0 : p_1 = p_2$ *vs* $H_1 : p_1 \neq p_2$. La regione di rifiuto è $R_\alpha = \{z \in \mathbf{R} :\mid z \mid \geq 1.96\}$.

Poiché $\hat{p}_1 = 40/224 = 0.18$, $\hat{p}_2 = 87/245 = 0.36$ e $\hat{p} = 0.27$, il valore della statistica test è $z = -4.30$.

Si rifiuta H_0 e $\alpha^{oss} = 2\min\{\Phi(-4.30), 1 - \Phi(-4.30)\} = 1.73 \cdot 10^{-5}$.

L'efficacia della terapia *è significativamente diversa* (*migliora*) con l'utilizzazione dell'antidepressivo. $\Diamond$

Esempio. *Aspirina*. Alcuni anni or sono venne condotto uno studio per valutare l'effetto dell'aspirina per prevenire l'infarto.

Si costituì un gruppo di trattamento (somministrazione giornaliera di aspirina) con $n_1 = 11037$ individui e un gruppo di controllo (somministrazione giornaliera di placebo) con $n_2 = 11034$ individui.

Dopo un certo periodo si osservarono 139 infartuati nel primo gruppo e 239 nel secondo. Sulla base di questi dati preliminari si interruppe lo studio che doveva durare cinque anni.

Infatti, $\hat{p}_1 = 139/11037 = 0.0126$, $\hat{p}_2 = 239/11034 = 0.0217$ e $\hat{p} = 0.0171$, da cui si ottiene che il valore della statistica test è $z = -5.19$. Si vuole verificare $H_0 : p_1 = p_2$ *vs* $H_1 : p_1 < p_2$.

Poiché $\alpha^{oss} = \Phi(-5.19) = 1.05 \cdot 10^{-7}$, si rifiuta l'ipotesi che la proporzione di infartuati sia la stessa nel due casi.

La terapia con l'aspirina porta ad una riduzione significativa della proporzione di infartuati. $\diamondsuit$

È possibile operare un confronto tra proporzioni anche quando le due *popolazioni bernoulliane non* sono *indipendenti*.

Ad esempio, quando si vuole valutare la differenza tra le probabilità riferite ad un certo evento di interesse in due distinte condizioni sperimentali, con riferimento alle stesse unità statistiche (**dati appaiati**).

Con campioni di elevata numerosità, si può utilizzare il **test di McNemar**, che è basato sul confronto tra le discordanze osservate e le discordanza attese e che, sotto $H_0 : p_1 = p_2$, ha distribuzione approssimata $N(0, 1)$.

Il test per il confronto tra proporzioni si può estendere al caso in cui si hanno 2 o più *campioni multinomiali indipendenti*.

Il **modello multinomiale** estende il modello binomiale considerando il caso in cui si hanno $n \geq 1$ replicazioni indipendenti di un esperimento di base che può assumere due o più esiti.

Si hanno quindi $k \geq 2$ possibili categorie $x_1, \ldots, x_k$ e il modello binomiale si ottiene come caso particolare con $k = 2$ e $x_1 = 0, x_2 = 1$.

Si considerino $r \geq 2$ *campioni indipendenti* di numerosità fissata $n_1, \ldots, n_r$, riferiti a un *modello multinomiale* con $k \geq 2$ *categorie* $y_1, \ldots, y_k$. I dati possono venire riassunti nella seguente tabella

	y_1	y_2	$\ldots$	y_k	
campione 1	n_{11}	n_{12}	$\ldots$	n_{1k}	$n_{1+} = n_1$
campione 2	n_{21}	n_{22}	$\ldots$	n_{2k}	$n_{2+} = n_2$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
campione r	n_{r1}	n_{r2}	$\ldots$	n_{rk}	$n_{r+} = n_r$
	n_{+1}	n_{+2}	$\ldots$	n_{+k}	n

Si vuole verificare l'ipotesi che i campioni provengano dalla stessa popolazione multinomiale e quindi si considera l'ipotesi nulla

$$H_0 \,:\, p_{1j} = p_{2j} = \cdots = p_{rj}, \text{ per ogni } j = 1, \ldots, k,$$

dove p_{ij}, $i = 1, \ldots, r$ è la probabilità dell'esito y_j nella popolazione i-esima. L'ipotesi alternativa H_1 è la negazione di H_0.

Nel caso di confronto tra due proporzioni, i dati vengono organizzati in una tabella 2×2.

La **statistica test** è ottenuta generalizzando il quadrato della statistica test Z, riferita alla differenza tra le proporzioni campionarie, e corrisponde all'**indice chi-quadrato** definito in statistica descrittiva

$$K = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*},$$

dove $n_{ij}^* = n_{i+}n_{+j}/n$, $i = 1, \ldots, r$, $j = 1, \ldots, k$.

In genere, se $n_1, \ldots, n_r$ sono sufficientemente elevati, la statistica test ha, sotto H_0, distribuzione approssimata $\chi^2((r-1)(k-1))$ e valori (positivi) elevati indicano una bassa conformità con H_0.

Se i campioni in esame non sono sufficientemente elevati, esistono soluzioni esatte, che sono di solito implementate nei software statistici.

Esempio. *Farmaco*. Per sperimentare l'efficacia di un principio attivo si considerano tre farmaci: A (senza principio attivo), B (con una certa quantità di principio attivo) e C (con una quantità doppia di principio attivo).

Si effettua la sperimentazione su $r = 3$ campioni indipendenti con $n_1 = 126$, $n_2 = 85$ e $n_3 = 119$ pazienti ai quali si somministrano, rispettivamente, i farmaci A, B e C.

Con riferimento ad ogni paziente si osserva la variabile multinomiale "Esito finale", con $k = 3$ possibili categorie: $y_1 =$*peggiorato*, $y_2 =$*stazionario*, $y_3 =$*migliorato*.

Ci si chiede i tre campioni provengono dalla stessa popolazione multinomiale, cioè se le proporzioni delle tre tipologie di esiti sono ragionevolmente le stesse nei tre casi.

I dati osservati vengono riassunti nella seguente tabella 3×3

	peggiorato	*stazionario*	*migliorato*	
farmaco A	10	104	12	126
farmaco B	5	70	10	85
farmaco C	10	65	44	119
	25	239	66	330

Il valore osservato della statistica test χ^2 è $k = 35.537$. Posto $\alpha = 0.01$, la regione di rifiuto è $R_\alpha = \{k \in \mathbf{R}^+ \,:\, k > 13.28\}$, dove 13.28 è il valore critico di livello 0.01 di una variabile casuale $\chi^2(4)$.

Inoltre, si ha che $\alpha^{oss} = 1 - F_K(35.537) = 3.70 \cdot 10^{-7}$.

Quindi si rifiuta l'ipotesi che i tre campioni provengono dalla stessa popolazione multinomiale, cioè che le proporzioni delle tre tipologie di esiti siano le stesse nei tre casi. $\Diamond$

Esempio. *Fisioterapia*. Si vuole valutare il livello di soddisfazione dei clienti di un centro di fisioterapia, considerando separatamente coloro che svolgono attività ginnica e terapie.

Si considerano $r = 2$ campioni indipendenti con $n_1 = 227$ clienti che svolgono attività ginnica e $n_2 = 262$ che seguono terapie, rispettivamente. I dati sono riassunti nella seguente tabella

	soddisfatto	*non soddisfatto*	
ginnastica	163	64	227
terapie	154	108	262
	317	172	489

Ci si chiede se la proporzione di soddisfatti p_1 tra coloro che svolgono attività ginnica e la proporzione di soddisfatti p_2 tra coloro che seguono terapie è la stessa.

Si considerano le ipotesi $H_0 : p_1 = p_2$ *vs* $H_1 : p_1 \neq p_2$ e, per la verifica, si può seguire la procedura illustrata in precedenza basata sul test per il confronto tra due proporzioni, con $\alpha = 0.05$.

In questo caso, $\hat{p}_1 = 0.718$, $\hat{p}_2 = 0.588$, $\hat{p} = 0.648$ e, poiché il valore osservato della statistica test è $z = 3.009$ e $\alpha^{oss} = 0.0026$, si rifiuta H_0 ad un livello di significatività $\alpha = 0.05$.

Si giunge al medesimo risultato utilizzando il test chi-quadrato. In questo caso il valore osservato della statistica test è $k = 9.05$. Posto $\alpha = 0.05$, la regione di rifiuto è $R_\alpha = \{k \in \mathbf{R}^+ \,:\, k > 3.84\}$, dove 3.84 è il valore critico di livello 0.05 di una variabile casuale $\chi^2(1)$.

Quindi si rifiuta l'ipotesi nulla che i due campioni provengano dalla stessa popolazione bernoulliana e si conclude che il livello di soddisfazione dipende dal tipo di servizio offerto.

Inoltre, si ha che $\alpha^{oss} = 1 - F_K(9.05) = 0.0026$, che coincide con il valore trovato in precedenza.

Si noti che il quadrato della statistica test Z corrisponde a 9.05, valore osservato della statistica test chi-quadrato. $\diamondsuit$

Test chi-quadrato di indipendenza

Una **tabella di contingenza** descrive la frequenza con la quale le modalità (categorie) di due variabili qualitative X e Y vengono *congiuntamente* osservate in un campione di dimensione n.

Se X ha r categorie, $x_1, \ldots, x_r$, ed Y ha k categorie, $y_1, \ldots, y_k$, la tabella di contingenza contiene la frequenza n_{ij} con cui si sono osservate le $r \times k$ coppie (x_i, y_j), $i = 1, \ldots, r$, $j = 1, \ldots, k$,

	y_1	y_2	$\ldots$	y_k	
x_1	n_{11}	n_{12}	$\ldots$	n_{1k}	n_{1+}
x_2	n_{21}	n_{22}	$\ldots$	n_{2k}	n_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_r	n_{r1}	n_{r2}	$\ldots$	n_{rk}	n_{r+}
	n_{+1}	n_{+2}	$\ldots$	n_{+k}	n

dove n_{i+} e n_{+j} sono, rispettivamente, i totali di riga e di colonna e corrispondono alle *frequenze marginali* osservate associate alle variabili X e Y.

Se si dividono le frequenze assolute, congiunte e marginali, per la dimensione del campione n si ottengono le corrispondenti **frequenze relative** o **proporzioni osservate**.

In analogia con quanto affermato sia in statistica descrittiva che in probabilità, ci sarà **indipendenza** tra X e Y se le *frequenze congiunte* sono pari a

$$n_{ij}^* = \frac{n_{i+}n_{+j}}{n}, \, i = 1, \ldots, r, j = 1, \ldots, k.$$

Per verificare le ipotesi H_0 : indipendenza *vs* H_1 : dipendenza, si può considerare la seguente **statistica test**

$$K = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*},$$

che si basa sulla differenza tra le *frequenze osservate* n_{ij} e le *frequenze attese* in ipotesi di dipendenza n_{ij}^*. La forma della statistica test K è identica a quella considerata il precedenza ma il contesto di applicazione è diverso.

È evidente che valori elevati per la statistica test sono in disaccordo con l'ipotesi nulla.

Se il numero di osservazioni per ogni combinazione di modalità è sufficientemente elevato, la statistica test K ha, sotto H_0, distribuzione approssimata $\chi^2((r-1)(k-1))$.

Il test è detto **test chi-quadrato di indipendenza**; fissato il livello di significatività α, la regione di rifiuto corrisponde a

$$R_\alpha = \{k \in \mathbf{R}^+ \, : \, k > \chi^2_\alpha\},$$

mentre $\alpha^{oss} = 1 - F_K(k^{oss})$, con k^{oss} il valore osservato della statistica test, essendo χ^2_α e $F_K(\cdot)$ il valore critico di livello α e la funzione di ripartizione di una variabile casuale $\chi^2((r-1)(k-1))$.

L'approssimazione fornisce risultati soddisfacenti se tutte le frequenze attese n^*_{ij} sono almeno pari a 5. Per piccoli campioni, è preferibile utilizzare soluzioni esatte, come ad esempio il **test esatto di Fisher**, usualmente implementate nei software statistici.

Esempio. *Perni* (continua). In uno stabilimento industriale ci sono tre macchinari per la produzione di perni di acciaio, che devono rispettare le specifiche di diametro.

Per valutarne l'efficacia del procedimento produttivo si considera un campione di $n = 400$ perni per i quali si osservano i valori della variabile $X =$"Macchinario", con $r = 3$ categorie (*macchinario1*, *macchinario2*, *macchinario3*), e della variabile $Y =$"Diametro", con $k = 3$ categorie (*fine*, *ok*, *spesso*).

I dati vengono riassunti nella seguente tabella di contingenza

	fine	*ok*	*spesso*	
macchinario1	10	102	8	120
macchinario2	34	161	5	200
macchinario3	10	60	10	80
	54	323	23	400

Si vuole verificare se c'è dipendenza tra il tipo di macchinario utilizzato e la qualità dei perni: $H_0 : X$ e Y indipendenti *vs* $H_1 : X$ e Y dipendenti.

Per la verifica, si può considerare il test chi-quadrato di indipendenza. Dal momento che tabella con le frequenze attese nel caso di indipendenza corrisponde a

	fine	*ok*	*spesso*	
macchinario1	16.2	96.9	6.9 8	120
macchinario2	27.0	161.5	11.5	200
macchinario3	10.8	64.6	4.6	80
	54	323	23	400

la statistica test assume valore $k = 15.03$.

Posto $\alpha = 0.01$, la regione di rifiuto è $R_\alpha = \{k \in \mathbf{R}^+ \ : \ k > 13.28\}$, dove 13.28 è il valore critico di livello 0.01 di una variabile casuale $\chi^2(4)$.

Inoltre, si ha che $\alpha^{oss} = 1 - F_K(15.03) = 0.005$.

Quindi si rifiuta l'ipotesi che la qualità dei perni sia indipendente dal tipo di macchinario utilizzato. $\diamondsuit$

Test chi-quadrato di adattamento

Una variante del test chi-quadrato può essere utilizzata anche per verificare se una variabile categoriale ha una specifica distribuzione di probabilità. In questo caso il test è detto **test chi-quadrato di adattamento**.

Esempio. *Nascite*. Si vuole valutare se le nascite siamo distribuite in modo uniforme tra i giorni della settimana.

Indicata con X la variabile casuale categoriale "Giorno settimanale di nascita", con supporto $S_X = \{\text{lu}, \text{ma}, \text{me}, \text{gi}, \text{ve}, \text{sa}, \text{do}\}$, si vuole verificare se la sua distribuzione di probabilità è uniforme.

Posto $p_1 = P(X = \text{lu}), \ldots, p_7 = P(X = \text{do})$, si specificano le seguenti ipotesi

$$H_0 : p_i = 1/7, \text{ per ogni } i \quad \textit{vs} \quad H_1 : p_i \neq 1/7, \text{ per qualche } i.$$

Per decidere se rifiutare o non rifiutare H_0, si considera un campione di n nascite, distribuite per giorno della settimana, e si definisce un opportuno test di ipotesi. $\Diamond$

In generale, sia X una *variabile casuale categoriale* che può assumere $k > 1$ modalità distinte con probabilità p_i, $i = 1, \ldots, k$, rispettivamente, dove $p_i > 0$, $i = 1, \ldots, k$, e $\sum_{i=1}^{k} p_i = 1$.

Si osserva un campione costituito da n osservazioni indipendenti di X e si vuole verificare l'ipotesi nulla

$$H_0 \, : \, p_1 = p_1^0, \ldots, p_k = p_k^0,$$

cioè che X abbia una distribuzione di probabilità specificata da $p_1^0, \ldots, p_k^0$. L'ipotesi alternativa H_1 è la negazione di H_0.

Anche in questo caso si considera una **statistica test chi-quadrato**

$$K = \sum_{i=1}^{k} \frac{(n_i - n_i^*)^2}{n_i^*},$$

dove n_i è la *frequenza osservata* nel campione e $n_i^* = n p_i^0$ è la *frequenza attesa*, sotto H_0, della modalità i-esima.

È evidente che valori elevati per la statistica test sono in disaccordo con l'ipotesi nulla.

Se il numero di osservazioni nel campione, riferito alle k modalità, è sufficientemente elevato, la statistica test K ha, sotto H_0, distribuzione approssimata $\chi^2(k-1)$.

Fissato il livello di significatività α, la regione di rifiuto è

$$R_\alpha = \{k \in \mathbf{R}^+ \,:\, k > \chi^2_\alpha\},$$

mentre $\alpha^{oss} = 1 - F_K(k^{oss})$, con k^{oss} il valore osservato della statistica test, essendo χ^2_α e $F_K(\cdot)$ il valore critico di livello α e la funzione di ripartizione di una variabile casuale $\chi^2(k-1)$.

Il test χ^2 di adattamento si può applicare anche a variabili casuali discrete con un supporto finito e a variabili casuali continue con modalità raggruppate in un numero finito di classi.

Inoltre, può essere interpretato come una generalizzazione del *test su una proporzione*, che di fatto verifica se una variabile casuale binaria (con due possibili categorie) segue una specifica distribuzione bernoulliana.

Esempio. *Nascite* (continua). Con riferimento alla variabile casuale categoriale "Giorno settimanale di nascita", si dispone di $n = 140$ osservazioni campionarie con la seguente distribuzione

Giorno	lu	ma	me	gi	ve	sa	do
n_i	13	23	24	20	27	18	15

Sotto l'ipotesi nulla che tutte le probabilità siano $1/7$, le frequenze attese sono $n_i^* = np_i^0 = 140/7 = 20$, $i = 1, \ldots, 7$.

Il valore osservato della statistica test χ^2 è 7.6. Posto $\alpha = 0.05$, la regione di rifiuto è $R_\alpha = \{k \in \mathbf{R}^+ \ : \ k > 12.6\}$, dove 12.6 è il valore critico di livello 0.05 di una variabile casuale $\chi^2(6)$.

Quindi non si rifiuta H_0: non c'è una sufficiente evidenza empirica per affermare che la distribuzione delle nascite non è uniforme.

Inoltre, si ha che $\alpha^{oss} = 1 - F_{\chi^2}(12.6) = 0.27$. $\Diamond$