

Università di Siena

Corso di STATISTICA

Parte terza: Test di verifica delle ipotesi

Andrea Garulli, Antonello Giannitrapani, Simone Paoletti

Master E^2C

Centro per lo Studio dei Sistemi Complessi

Università di Siena

email: statistica@csc.unisi.it

3.1 Verifica di ipotesi statistiche

- ✓ Problemi di *decision making*
- ✓ Ipotesi Statistiche
- ✓ Verifica di ipotesi
  - ◊ obiettivo
  - ◊ esito
  - ◊ tipi di errori
- ✓ Livello di significatività e potenza di un test
- ✓ Statistica di test e regione di rifiuto
- ✓ Procedura *"operativa"*
- ✓ Conclusioni

Esempio 1

"Un'agenzia pubblicitaria ha la sensazione che, a causa dell'eccessiva violenza in alcuni programmi televisivi, una certa classe di popolazione adulta abbia iniziato a guardare i programmi dei ragazzi. Un cliente sarebbe interessato a delle inserzioni pubblicitarie durante questi programmi, a patto che almeno il 20% della popolazione adulta li guardi. Per questo motivo, l'agenzia conduce uno studio su un campione di 200 adulti, dal quale risulta che il 24% degli intervistati guarda i programmi dei ragazzi."

$\Rightarrow$ Conviene al cliente investire?

$\Rightarrow$ Effettivamente almeno il 20% della popolazione adulta guarda quei programmi?

Esempio 2

"Un azienda produttrice di detergenti deve decidere se migliorare la propria linea di produzione di saponette. Recentemente, infatti, un'associazione di consumatori ha vivacemente protestato, poiché molti esemplari pesano significativamente meno dei 500 grammi dichiarati dall'azienda. A tal fine, un campione casuale di 25 saponette viene estratto da un lotto pronto per la consegna. Risulta che il peso medio è pari a 498 grammi, con una deviazione standard di 4 grammi. Le specifiche di prodotto che l'azienda deve garantire sono 500 ± 2 *grammi."*

$\Rightarrow$ La linea di produzione attuale va migliorata?

$\Rightarrow$ Il prodotto rispetta effettivamente le specifiche?

Esempio 3

"I sistemi di irrigazione sono progettati per distribuire acqua uniformemente su una zona agricola. Un'azienda produttrice di tali sistemi è interessata a verificare la validità di una nuova tecnologia, che dovrebbe garantire maggiore uniformità rispetto ai sistemi attuali. Una misura dell'uniformità di irrigazione è data dalla variazione di quantità d'acqua (deviazione standard) distribuita in diverse locazioni. I sistemi normalmente utilizzati garantiscono variazioni di 0.1 cm/hr. Effettuando un'esperimento, con 25 rilevazioni casuali in punti differenti, si è rilevata una variazione di 0.078 cm/hr, col nuovo metodo"

$\Rightarrow$ Conviene utilizzare la nuova tecnologia?

$\Rightarrow$ Le prestazioni offerte dal nuovo sistema sono realmente diverse da quello in uso?

Problema

> *"Prendere delle decisioni avendo a disposizione un certo numero di informazioni incerte."*

Ad esempio

- ✓ check up dentistico;
- ✓ validità di un prodotto farmaceutico;
- ✓ bontà di un processo produttivo;
- ✓ metal-detector aeroportuali;
- ✓ sistemi di allarmi per auto;
- ✓ motore di ricerca per il reperimento di informazioni sul web;
- ✓ ...

Verifica di ipotesi statistiche

La verifica di ipotesi statistiche (*hypothesis testing*) è uno strumento di supporto alle decisioni.

⇒ Procedure formali che consentono di *prendere delle decisioni* a partire da

- un insieme di *dati incerti*;
- *proprietà statistiche* dei dati disponibili.

⇒ Tecniche statistiche progettate per estrapolare informazioni da un campione per effettuare inferenze su una popolazione allo scopo di prendere delle decisioni.

⇒ Regole di decisione che consentono di utilizzare i dati incerti disponibili per discriminare fra *due ipotesi antagoniste* relative al mondo esterno.

Ipotesi statistiche

Un'ipotesi è un'asserzione relativa ad alcune proprietà statistiche di una o più variabili aleatorie.

Ipotesi parametriche

- Affermazioni riguardanti il valore di uno o più parametri incogniti della distribuzione di una o più variabili aleatorie.
  Es.:
  - $E[\mathbf{x}] = \mu_0$
  - $Var(\mathbf{x}) = \sigma_0^2$
  - $E[\mathbf{x} - \mathbf{y}] = \mu_x - \mu_y = 0$

Ipotesi non parametriche

- Affermazioni riguardanti il tipo di distribuzione di una o più variabili aleatorie.
  Es.:
  - $\mathbf{x} \sim \mathcal{N}(0, 1)$

Ipotesi nulla ed ipotesi alternativa

Null hypothesis $\mathbf{H_0}$

- In genere è formulata come un'uguaglianza, ed esprime una condizione di indifferenza, in funzione della quale non occorre intraprendere alcuna azione.

Alternative hypothesis $\mathbf{H_1}$

- Ipotesi rispetto alla quale testare l'ipotesi nulla $\mathbf{H_0}$, esprime l'esistenza di una qualche differenza.

Le due ipotesi devono esprimere condizioni *mutuamente esclusive*.

Esito di una procedura di test

Metodo scientifico

- confutare ipotesi insoddisfacenti (*status quo*) proponendone di nuove, migliori e verificabili.

Analogamente, l'obiettivo di una procedura di test è *rifiutare l'ipotesi nulla* $\mathbf{H_0}$ in favore dell'ipotesi alternativa $\mathbf{H_1}$, sulla base dei dati sperimentali disponibili.

Esito di una procedura di test

- ✓ *Rifiutare* $\mathbf{H_0}$ in favore di $\mathbf{H_1}$
- ✓ *Non rifiutare* $\mathbf{H_0}$

Terminologia
Quando i dati non consentono di rifiutare l'ipotesi $\mathbf{H_0}$, si usa dire, *impropriamente*, che il test suggerisce di accettare l'ipotesi $\mathbf{H_0}$.

Analogia col sistema processuale

✓**Ipotesi nulla**

Presunzione di innocenza $\Longleftrightarrow$ Ipotesi $\mathbf{H_0}$

✓**Obiettivo**

Provare la colpevolezza $\Longleftrightarrow$ Rifiutare l'ipotesi $\mathbf{H_0}$

✓**Regola di decisione**

Raccolta prove $\Longleftrightarrow$ Evidenza sperimentale

Vaglio della giuria $\Longleftrightarrow$ Test statistico

✓**Possibili esiti**

Colpevole $\Longleftrightarrow$ Rifiuto ipotesi $\mathbf{H_0}$

Non colpevole $\Longleftrightarrow$ Non rifiuto ipotesi $\mathbf{H_0}$

N.B. *La giuria non emette un verdetto di* innocenza, *bensì di* non colpevolezza.

Errori di una procedura di test

Una procedura di verifica di ipotesi statistiche è soggetta a due tipi di errori:

- ◊ **Tipo I** Rifiutare l'ipotesi $\mathbf{H_0}$ quando $\mathbf{H_0}$ è in realtà vera
- ◊ **Tipo II** Non rifiutare l'ipotesi $\mathbf{H_0}$ quando $\mathbf{H_0}$ è in realtà falsa

Ossrvazioni

- Non è possibile eliminare questi due tipi di errore. È possibile solamente ridurre la loro frequenza al minimo e conoscere la probabilità con la quale avvengono.
- La statistica è la scienza che permette di scegliere e prendere decisioni non perché immune da errori, ma perché fornisce la probabilità di errare, associata ad ogni scelta; quindi di conoscere il rischio che si corre, se la scelta si dimostrasse errata.

Livello di significatività e potenza di un test

Si consideri un generico test, per il quale

- Pr{errore Tipo I} $= \alpha$
- Pr{errore Tipo II} $= \beta$

Allora, tale test ha:

✓ *livello di significatività* α

✓ *potenza* pari ad $1 - \beta$.

Obiettivo

$\Rightarrow$ Test che minimizzi le probabilità α e β di entrambi i tipi di errore

Livello di significatività e potenza di un test

- Un test con un livello di significatività α ha una probabilità di rifiutare *erroneamente* l'ipotesi nulla pari ad α.
- Un test con potenza pari ad $1 - \beta$ ha una probabilità di non rifiutare *erroneamente* l'ipotesi nulla pari a β.

Problema

$\Diamond$ Se si abbassa il livello di significatività, cioè la probabilità di commettere errori di Tipo I (α), si accresce quella dell'errore di Tipo II (β), e viceversa. L'unico modo per ridurle entrambi è aumentare il numero di dati.

Soluzione

$\Rightarrow$ Fissato un livello di significatività desiderato, si cerca di minimizzare β (cioè massimizzare la potenza).

Valori tipici per α sono 0.05, 0.01, 0.0001.

Statistica di test

Una *statistica di test* è una funzione dei dati raccolti (campione), di cui è nota (esattamente o in maniera approssimata) la distribuzione di probabilità, supponendo sia vera l'ipotesi nulla.

La statistica di test da utilizzare dipende da

- ◊ l'ipotesi $\mathbf{H_0}$ che si intende verificare
- ◊ la distribuzione di probabilità della popolazione da cui si estrae il campione delle osservazioni (conoscenza a priori)

Esempio. (Test sulla media μ, con varianza σ^2 nota)

Ipotesi	Statistica di test
$\mathbf{H_0} : \mu \leq \mu_0$ $\mathbf{H_1} : \mu > \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Regione di rifiuto

La *regione di rifiuto* è l'insieme $\mathcal{C} \subseteq \mathcal{R}$ costitutito da tutti e soli i valori della statistica di test che consentono di rifiutare l'ipotesi nulla.

La regione di rifiuto comprende tutti quei valori della statistica di test che sarebbero "poco probabili" se l'ipotesi nulla fosse vera.
Essa dipende da:

- livello di significatività α;
- ipotesi alternativa $\mathbf{H_1}$;
- distribuzione della statistica di test.

Esempio. (Test sulla media μ, con varianza σ^2 nota)

Ipotesi	Statistica di test	Regione di rifiuto
$\mathbf{H_0}: \mu \leq \mu_0$ $\mathbf{H_1}: \mu > \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$Z > Z_\alpha$, Z_α : se $X \sim \mathcal{N}(0,1)$ $Pr\{X > Z_\alpha\} = \alpha$

Verifica di ipotesi: procedura

1. A partire dal problema, formulare opportunamente le ipotesi
2. Scegliere il test statistico appropriato
3. Raccogliere i dati
4. Calcolare il valore della statistica di test
5. Calcolare la regione di rifiuto
6. Scgegliere l'ipotesi approppriata

Esempio 1

"Un'agenzia pubblicitaria ha la sensazione che, a causa dell'eccessiva violenza in alcuni programmi televisivi, una certa classe di popolazione adulta abbia iniziato a guardare i programmi dei ragazzi. Un cliente sarebbe interessato a delle inserzioni pubblicitarie durante questi programmi, a patto che almeno il 20% della popolazione adulta li guardi. Per questo motivo, l'agenzia conduce uno studio su un campione di 200 adulti, dal quale risulta che il 24% degli intervistati guarda i programmi dei ragazzi."

$\Rightarrow$ Conviene al cliente investire?

$\Rightarrow$ Effettivamente almeno il 20% della popolazione adulta guarda quei programmi?

Esempio 1: formulazione delle ipotesi

Il cliente è interessato a verificare che almeno il 20% della popolazione adulta guardi i programmi dei ragazzi

Il problema può essere formulato in termini di verifica di *ipotesi parametriche* relative alle *proporzioni* di una certa popolazione. Sia π la frazione di popolazione avente una certa caratteristica (es. *"guardare i programmi dei ragazzi"*).

- Ipotesi
  - $\mathbf{H_0}$: $\pi \leq 0.20$
  - $\mathbf{H_1}$: $\pi > 0.20$

Esempio 1: scelta del test

Nel caso in cui le ipotesi riguardino le proporzioni di una certa popolazione, ed il campione disponibile sia sufficientemente numeroso, si utilizza un cosiddetto $Z - test$.

La statistica di test da calcolare è

$$Z = \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}}$$

dove

- p è la frazione del campione avente la caratteristica desiderata
- n è la numerosità del campione
- π_0 è il valore della proporzione ipotizzato

Esempio 1: calcolo della statistica di test

Dopo aver intervistato un campione casuale di 200 persone adulte, è possibile calcolare il particolare valore della statistica in corrispondenza dei dati osservati:

- $p = 0.24$
- $n = 200$
- $\pi_0 = 0.20$

La statistica di test vale

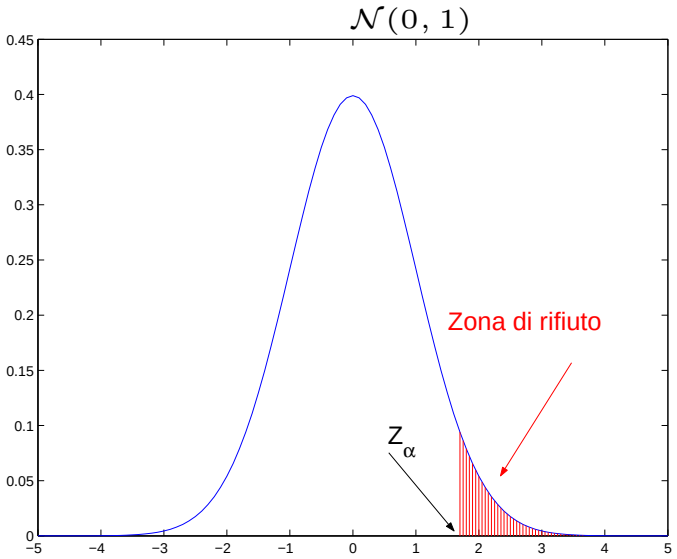
$$Z = \frac{p - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} = \frac{0.24 - 0.20}{\sqrt{0.20(1-0.20)/200}} = 1.41$$

Esempio 1: calcolo della regione di rifiuto

La statistica di test $Z = \frac{p-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$ è una v.a. normale a media nulla e varianza unitaria, sotto l'ipotesi $\pi = \pi_0$.
Sia $Z \sim \mathcal{N}(0,1)$ e si definisca Z_α:

$$\Pr\{Z > Z_\alpha\} = \alpha$$

Fissato $\alpha = 0.05$, si ha che $Z_{0.05} = 1.65$



Regione di rifiuto

$\Rightarrow$ Rifiutare l'ipotesi $\mathbf{H_0}$ se $Z > Z_\alpha$

Osservazione

✓ La regione di rifiuto dipende dal livello di significatività, dalla numerosità del campione e dal tipo di ipotesi alternativa.

Esempio 1: scelta dell'ipotesi

Regola di decisione

◊ Rifiutare l'ipotesi $\mathbf{H_0}$ se $Z > Z_\alpha$

Nell'esempio in questione:

- $\mathbf{H_0}$: *"Non più del* 20% *della popolazione adulta guarda i programmi dei ragazzi"*
- $Z = 1.41$
- $Z_{0.05} = 1.65$

Conclusione

⇒ L'ipotesi nulla *non può essere rifiutata* con un *livello di significatività* $\alpha = 0.05$, sulla base dell'evidenza sperimentale.

Interpretazione

✓ Le interviste raccolte sono compatibili con l'ipotesi secondo la quale *"non più del* 20% *degli adulti guarda la tv dei ragazzi"*. Il risultato campionario del 24% può essere dovuto al caso.

⇒ Non esistono dati sufficienti per indurre il cliente ad investire.

Esempio 2

"Un azienda produttrice di detergenti deve decidere se migliorare la propria linea di produzione di saponette. Recentemente, infatti, un'associazione di consumatori ha vivacemente protestato, poiché molti esemplari pesano significativamente meno dei 500 grammi dichiarati dall'azienda. A tal fine, un campione casuale di 25 saponette viene estratto da un lotto pronto per la consegna. Risulta che il peso medio è pari a 498 grammi, con una deviazione standard di 4 grammi. Le specifiche di prodotto che l'azienda deve garantire sono 500 ± 2 *grammi."*

$\Rightarrow$ La linea di produzione attuale va migliorata?

$\Rightarrow$ Il prodotto rispetta effettivamente le specifiche?

Esempio 2: formulazione delle ipotesi

Il produttore è interessato a verificare che il peso medio di una saponetta sia almeno pari a 500 grammi, come da specifiche.

Supponendo che il peso di un esemplare del prodotto sia assimilabile ad una v.a. X avente distribuzione normale, è possibile formulare il problema in termini di verifica di *ipotesi parametriche* relative al *valor medio* di una v.a. normale.
Sia $X \sim \mathcal{N}(\mu, \sigma^2)$, con σ^2 ignota .

- Ipotesi parametriche
  - $\mathbf{H_0}$: $\mu \geq 500$
  - $\mathbf{H_1}$: $\mu < 500$

Esempio 2: scelta del test

Nel caso in cui le ipotesi riguardino il valor medio di una v.a. normale, con varianza ignota, si utilizza un cosiddetto *t-test*.

La statistica di test da calcolare è

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

dove

- $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ è la media campionaria
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2$ è la varianza campionaria
- n è la numerosità del campione
- μ_0 è il valore della media ipotizzato

Esempio 2: calcolo della statistica di test

Dopo aver estratto un campione casuale di 25 saponette, è possibile calcolare il particolare valore della statistica in corrispondenza dei dati osservati:

- $\bar{X} = 498$
- $S = 4$
- $n = 25$
- $\mu_0 = 500$

La statistica di test vale

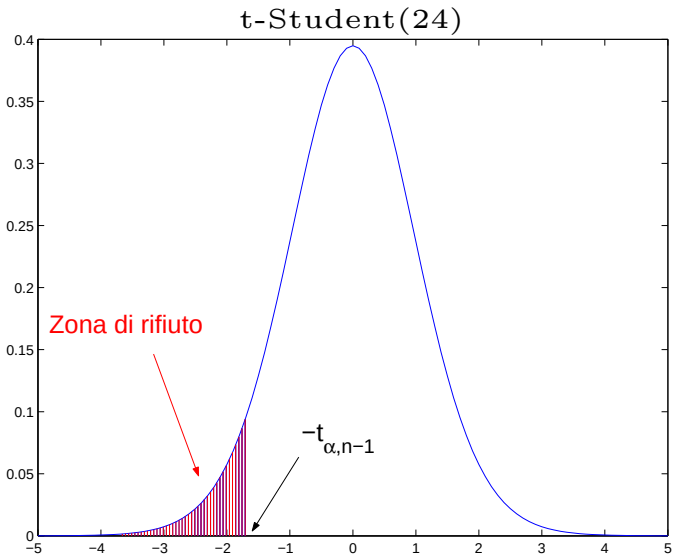
$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{498 - 500}{4/\sqrt{25}} = -2.5$$

Esempio 2: calcolo della regione di rifiuto

La statistica di test $t = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ è distribuita secondo una t-Student con $n-1$ gradi di libertà, sotto l'ipotesi $\mu = \mu_0$.
Sia $T \sim t(n-1)$ e si definisca $t_{\alpha,n-1}$:

$$\Pr\{T > t_{\alpha,n-1}\} = \alpha$$

Fissato $\alpha = 0.05$, si ha che $t_{0.05,24} = 1.71$



Regione di rifiuto

$\Rightarrow$ Rifiutare l'ipotesi $\mathbf{H_0}$ se $t < -t_{\alpha,n-1}$

Osservazione

✓ La regione di rifiuto dipende dal livello di significatività, dalla numerosità del campione e dal tipo di ipotesi alternativa.

Esempio 2: scelta dell'ipotesi

Regola di decisione

◊ Rifiutare l'ipotesi $\mathbf{H_0}$ se $t < -t_{\alpha,n-1}$

Nell'esempio in questione:

- $\mathbf{H_0}$: *"Il peso medio di una saponetta è almeno 500 grammi"*
- $t = -2.5$
- $-t_{0.05,24} = -1.71$

Conclusione

⇒ L'ipotesi nulla *può essere rifiutata* con un *livello di significatività* $\alpha = 0.05$, sulla base dell'evidenza sperimentale.

Interpretazione

✓ L'evidenza sperimentale consente di rigettare la tesi secondo la quale le saponette, in media, pesano 500 grammi. Il peso medio di 498 grammi, osservato dal produttore, non è dovuto al caso e il prodotto non rientra nelle specifiche di peso pari a 500 ± 2 grammi.

⇒ Occorre migliorare la linea di produzione.

Esempio 3

"I sistemi di irrigazione sono progettati per distribuire acqua uniformemente su una zona agricola. Un'azienda produttrice di tali sistemi è interessata a verificare la validità di una nuova tecnologia, che dovrebbe garantire maggiore uniformità rispetto ai sistemi attuali. Una misura dell'uniformità di irrigazione è data dalla variazione di quantità d'acqua (deviazione standard) distribuita in diverse locazioni. I sistemi normalmente utilizzati garantiscono variazioni di 0.1 cm/hr. Effettuando un'esperimento, con 25 rilevazioni casuali in punti differenti, si è rilevata una deviazione standard di 0.078 cm/hr, col nuovo metodo"

$\Rightarrow$ Conviene utilizzare la nuova tecnologia?

$\Rightarrow$ Le prestazioni offerte dal nuovo sistema sono realmente diverse da quello in uso?

Esempio 3: formulazione delle ipotesi

Il produttore è interessato a verificare che la variazione di acqua distribuita in diverse zone sia diversa da 0.1 cm/hr, prestazione garantita dagli attuali sistemi di irrigazione.

Supponendo che i cm/hr di acqua distribuiti in una certa locazione siano assimilabili ad una v.a. X avente distribuzione normale, è possibile formulare il problema in termini di verifica di *ipotesi parametriche* relative alla *varianza* di una v.a. normale.
Sia $X \sim \mathcal{N}(\mu, \sigma^2)$, con μ ignoto.

- Ipotesi parametriche
  - $\mathbf{H_0}: \ \sigma^2 = 0.1^2$
  - $\mathbf{H_1}: \ \sigma^2 \neq 0.1^2$

Esempio 3: scelta del test

Nel caso in cui le ipotesi riguardino la varianza di una v.a. normale, con valor medio ignoto, si utilizza un cosiddetto *test* χ^2.

La statistica di test da calcolare è

$$\chi^2_c = \frac{(n-1)S^2}{\sigma_0^2}$$

dove

- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2$ è la varianza campionaria
- n è la numerosità del campione
- σ_0^2 è il valore della varianza ipotizzato

Esempio 3: calcolo della statistica di test

Dopo aver misurato la quantità di acqua distribuita in 25 locazioni diverse, è possibile calcolare il particolare valore della statistica in corrispondenza dei dati osservati:

- $S^2 = 0.078^2$
- $n = 25$
- $\sigma_0^2 = 0.1^2$

La statistica di test vale

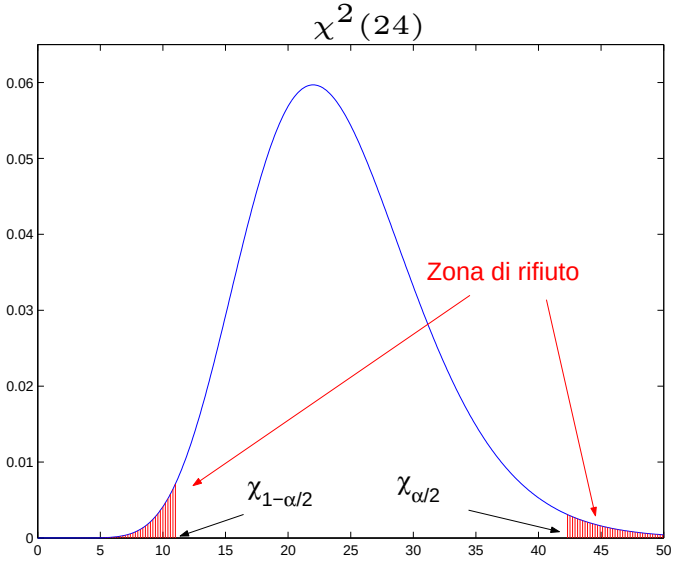
$$\chi_c^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(25-1)0.078^2}{0.1^2} = 14.60$$

Esempio 3: calcolo della regione di rifiuto

La statistica di test $\chi^2_c = \frac{(n-1)S^2}{\sigma_0^2}$ è distribuita secondo una χ^2 con $n-1$ gradi di libertà, sotto l'ipotesi $\sigma^2 = \sigma_0$. Sia $\chi^2_c \sim \chi^2(n-1)$.
Si definiscano $\chi^2_{1-\alpha/2,n-1}$ e $\chi^2_{\alpha/2,n-1}$:

$$\Pr\{\chi^2_c > \chi^2_{1-\alpha/2,n-1}\} = 1-\alpha/2$$
$$\Pr\{\chi^2_c > \chi^2_{\alpha/2,n-1}\} = \alpha/2$$

Fissato $\alpha = 0.025$, si ha che
$\chi^2_{1-0.0125,24} = 11.20$ e $\chi^2_{0.0125,24} = 42.12$



Regione di rifiuto

⇒ Rifiutare l'ipotesi $\mathbf{H_0}$ se $\chi^2_c < \chi^2_{1-\alpha/2,n-1}$ oppure $\chi^2_c > \chi^2_{\alpha/2,n-1}$

Osservazione

✓ La regione di rifiuto dipende dal livello di significatività, dalla numerosità del campione e dal tipo di ipotesi alternativa.

Esempio 3: scelta dell'ipotesi

Regola di decisione

◊ Rifiutare l'ipotesi $\mathbf{H_0}$ se $\chi^2_c > \chi^2_{1-\alpha/2,n-1}$ oppure $\chi^2_c < \chi^2_{\alpha/2,n-1}$.

Nell'esempio in questione:

- $\mathbf{H_0}$: *"La variazione di acqua distribuita da zona a zona è* 0.1^2 *[cm/hr]*^{2}*"*
- $\chi^2_c = 14.60$
- $\chi^2_{1-0.0125,24} = 11.20$ e $\chi^2_{0.0125,24} = 42.12$

Conclusione

⇒ L'ipotesi nulla *non può essere rifiutata* con un *livello di significatività* $\alpha = 0.025$, sulla base dell'evidenza sperimentale.

Interpretazione

✓ Con i dati raccolti, non è possibile affermare che la nuova tecnologia fornisca prestazioni diverse dagli attuali sistemi di irrigazione.

⇒ Effettuare un'indagine sperimentale più accurata, prima di adottare il nuovo metodo di irrigazione.

Conclusioni

"... la verifica di ipotesi statistiche non è una tecnica per dilettare gli statistici. Essa svolge un ruolo fondamentale in problemi di decision making. *La stima campionaria di alcuni parametri (media, varianza, ...) non dovrebbe spingere il manager a trarre conclusioni affrettate. La validazione statistica è uno strumento cruciale per prendere la decisione giusta. Perciò, l'affermazione:*

> *Ciò che è significativo per un manager può non essere statisticamente significativo.*
> *Ciò che non è significativo per un manager può essere statisticamente significativo.*

è profondamente vera."

P.K Viswanathan - Adjunct Professor and Management Consultant - Chennai, India.

3.2 Teoria statistica dei test

Ipotesi parametriche

- ✓ Definizioni
- ✓ Test uniformemente più potenti
- ✓ Lemma di Neyman-Pearson
  - ◊ Esempio
- ✓ Rapporto di verosimiglianza genralizzato (GLR)
- ✓ Test basati sul GLR

Goodness of fit

- ✓ Test Chi-Quadro
- ✓ Test di Kolmogorov-Smirnov

Ipotesi parametriche

Sia $\mathbf{X} = (X_1, X_2, ..., X_n)$ un campione estratto da una popolazione distribuita secondo una CDF $F(\mathbf{x}|\theta)$, avente *forma funzionale nota*, ma dipendente da un *vettore incognito di parametri* $\theta \in \Theta \subseteq \mathcal{R}^p$.

Sulla base delle osservazioni $\mathbf{x} = (x_1, x_2, ..., x_n)$, determinare la "validità" di ipotesi riguardanti il vettore di parametri θ.

Osservazione

- Tipicamente θ coincide con il valor medio μ e/o la varianza σ^2.

Caso notevole

Consideriamo il *campionamento* di una v.a. scalare $X \sim f_X(x|\theta)$.

- n ripetizioni indipendenti dello stesso esperimento
- le n osservazioni della v.a. X possono essere considerate come un'unica realizzazione della v.a. vettoriale $\mathbf{X} = (X_1, \ldots, X_n)$, dove le v.a. X_i sono indipendenti e identicamente distribuite $(i.i.d.)$

La densità di probabilità congiunta di $\mathbf{X}$ è

$$f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

perché le v.a. X_i sono indipendenti.

Definizioni

- *Null Hypothesis:* $\mathbf{H_0} : \theta \in \Theta_0 \subset \Theta$
- *Alternative Hypothesis:* $\mathbf{H_1} : \theta \in \Theta_1 \subseteq \Theta - \Theta_0$

Le ipotesi $\mathbf{H_i}$ possono essere ***semplici*** o ***composte*** a seconda che l'insieme Θ_i contenga uno o più elementi.

✓ *Errore di Tipo I*: rifiutare l'ipotesi H_0 quando è vera;

✓ *Errore di Tipo II*: non rifiutare l'ipotesi H_0 quando H_0 è falsa.

Un **test** è una funzione $\varphi(\mathbf{x}) : \mathcal{R}^n \to [0, 1]$, indicante la probabilità di rifiutare H_0 avendo osservato il campione $\mathbf{x}$.

Caso notevole

✓ Se $\varphi(\mathbf{x}) : \mathcal{R}^n \to \{0, 1\}$ il test è detto *deterministico*

Definizioni

- Un test φ ha *livello di significatività* α se:

$$E_\theta[\varphi(\mathbf{x})] \leq \alpha, \qquad \forall \theta \in \Theta_0$$

- Un test φ ha *potenza* pari a $1-\beta$ rispetto all'alternativa θ_1 se:

$$E_{\theta_1}[\varphi(\mathbf{x})] = 1-\beta, \qquad \theta_1 \in \Theta_1$$

Osservazione

$\Rightarrow$ α=Pr{errore Tipo I}

$\Rightarrow$ β=Pr{errore Tipo II}

Obiettivo

✓ Minimizzare contemporaneamente α e β.

Test uniformemente più potenti

Problema

◊ Riducendo α aumenta β, e viceversa

Strategia

⇒ Fissato un livello di significatività $\bar{\alpha}$ desiderato, cercare il test che massimizzi la potenza $1 - \beta$.

$$\begin{aligned} \max_{\varphi} \quad & E_\theta[\varphi(\mathbf{x})], && \theta \in \Theta_1 \\ \text{s.t.} \quad & E_\theta[\varphi(\mathbf{x})] \leq \bar{\alpha}, && \forall \theta \in \Theta_0 \end{aligned}$$

Un test φ^*:

$$\begin{aligned} E_\theta[\varphi^*(\mathbf{x})] &\leq \bar{\alpha}, && \forall \theta \in \Theta_0 \\ E_\theta[\varphi^*(\mathbf{x})] &\geq E_\theta[\varphi(\mathbf{x})] && \forall \theta \in \Theta_1 \quad \forall \varphi \text{ con significatività } \bar{\alpha} \end{aligned}$$

è detto test ***uniformemente più potente (UMP)*** fra tutti i test di pari significatività $\bar{\alpha}$.

Lemma di Neyman-Pearson

Siano: $H_0: \theta = \theta_0$ e $H_1: \theta = \theta_1$ *ipotesi semplici.*
Si definisca il *rapporto delle verosimiglianze (LR)* $\lambda(\mathbf{x}) = \frac{f(\mathbf{x}|\theta_1)}{f(\mathbf{x}|\theta_0)}$.
Allora, un test del tipo

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{se } \lambda(\mathbf{x}) \geq k_\alpha \\ 0 & \text{se } \lambda(\mathbf{x}) < k_\alpha \end{cases}$$

è il più potente fra tutti i test di pari significatività.

Ossservazioni

- ✓ Il lemma vale solo nel caso di *ipotesi semplici.*
- ✓ Sotto opportune ipotesi, è possibile estendere il lemma precedente al caso di ipotesi alternativa composta, ma unilaterale (*one sided*).
- ✓ In generale, quando l'ipotesi alternativa è composta, non esistono test uniformemente più potenti.

Esempio (1/6)

Si consideri la v.a. $X \sim \mathcal{N}(\mu, 1)$, con μ ignoto.
Ipotesi sulla media

- $\mathbf{H_0}$: $\mu = \mu_0 = 1$
- $\mathbf{H_1}$: $\mu = \mu_1 = 2$

Sia $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ il campione osservato.
In virtù del lemma di Neyman-Pearson, occorre considerare il rapporto delle verosimiglianze:

$$\lambda(\mathbf{x}) = \frac{f(\mathbf{x}|\mu = 2)}{f(\mathbf{x}|\mu = 1)}$$

dove

$$f(\mathbf{x}|\mu = 2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-2)^2}{2}}$$

$$f(\mathbf{x}|\mu = 1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-1)^2}{2}}$$

Esempio (2/6)

Fatto

- $\lambda(\mathbf{x}) = e^{\frac{n}{\sigma^2}\left((\mu_1-\mu_0)\bar{X}-\frac{1}{2}(\mu_1^2-\mu_0^2)\right)}$

Dal lemma di Neyman-Pearson, un test del tipo

$$\varphi(\mathbf{x}) = \begin{cases} 1 \text{ (rifiutare } H_0) & \text{se } \lambda(\mathbf{x}) \geq k_\alpha \\ 0 \text{ (non rifiutare } H_0) & \text{se } \lambda(\mathbf{x}) < k_\alpha \end{cases}$$

è uniformemente più potente.

Osservazione

- ✓ $\lambda(\mathbf{x}) \geq k_\alpha \Leftrightarrow \bar{X} \geq k_\alpha^{'}$

con $k_\alpha^{'} = \dfrac{(\mu_1+\mu_0)}{2} + \dfrac{\sigma^2 \log k_\alpha}{n(\mu_1-\mu_0)}$

Esempio (3/6)

Il valore critico k'_α si determina a partire dal livello di significatività α desiderato:

- $\alpha = \Pr\{$ rifiutare H_0 quando è vera $\}=\Pr\{ \bar{X} > k'_\alpha \mid \mu = 1\}$

Da cui si ricava

- $k'_\alpha = 1 + \frac{Z_\alpha}{\sqrt{n}}$

dove Z_α : $\Pr\{Z > Z_\alpha\} = \alpha$, se $Z \sim \mathcal{N}(0,1)$.

Nota

✓ Il valore Z_α si ricava dalle tabelle.

✓ In MATLAB®

```
>> alpha=0.05;
>> Z_alpha=norminv(1-alpha);
```

Esempio (4/6)

La potenza del test è pari a $1-\beta$, dove

- $\beta = \Pr\{$ non rifiutare H_0 quando è falsa $\}=\Pr\{\ \bar{X} < k'_\alpha \mid \mu = 2\}$

Da cui si ricava

- $1-\beta = 1-\Pr\{\ \bar{X} < 1+\frac{Z_\alpha}{\sqrt{n}} \mid \mu = 2\} = 1-\Pr\{Z < Z_\alpha - \sqrt{n}\}$

con $Z \sim \mathcal{N}(0,1)$.

Nota

- ✓ Il valore della probabilità si ricava dalle tabelle.
- ✓ In MATLAB®

```
>> n=10;
>> alpha=0.05;
>> Z_alpha=norminv(1-alpha);
>> beta=normcdf(Z_alpha-sqrt(n));
```

Esempio (5/6)

Siano:

- $n = 3$
- $x_1 = 1.23,\ x_2 = 1.57,\ x_3 = 1.49$
- livello di significatività desiderato $\alpha = 0.1$

Test

$$\varphi(\mathbf{x}) = \begin{cases} 1 \text{ (rifiutare } H_0) & \text{se } \bar{X} \geq 1 + \frac{Z_\alpha}{\sqrt{n}} \\ 0 \text{ (non rifiutare } H_0) & \text{se } \bar{X} < 1 + \frac{Z_\alpha}{\sqrt{n}} \end{cases}$$

Con tale scelta di α, e col numero n di dati a disposizione, si ottiene che il valore per rifiutare l'ipotesi $\mathbf{H_0}$ è:

✓ $1 + \frac{Z_{0.1}}{\sqrt{3}} = 1.74$

Poichè la media campionaria $\bar{X} = 1.43 < 1.74$, non rifiutare l'ipotesi $\mathbf{H_0} : \mu = 1$ in favore dell'ipotesi alternativa $\mathbf{H_1} : \mu = 2$, col livello di significatività fissato.

Esempio (6/6)

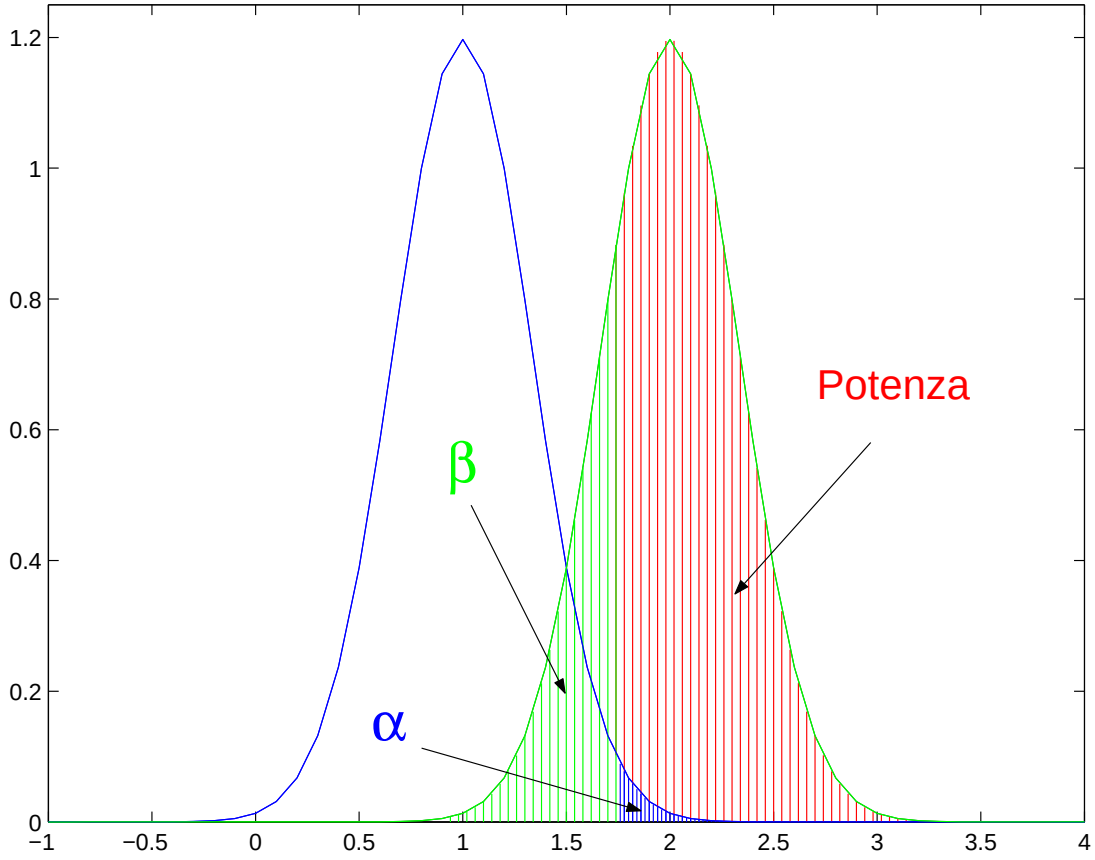


Figure 1: Andamento di $f_{\bar{X}}(x|\mu = 1)$ (blu) e $f_{\bar{X}}(x|\mu = 2)$ (verde)

Rapporto delle verosimiglianze

Problema

- ◊ Test uniformemente più potenti non esistono per un'ampia classe di problemi

Idea

$$\Rightarrow \text{ Considerare il rapporto delle verosimiglianze } r(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_1} f(\mathbf{x}|\theta)}{\sup_{\theta \in \Theta_0} f(\mathbf{x}|\theta)}$$

Nota

- ✓ Il sup è necessario in quanto, in generale, le ipotesi sono composte
  $\Rightarrow$ Θ_i contengono più di un elemento.

Interpretazione

- $\frac{\texttt{Miglior spiegazione dei dati secondo } H_1}{\texttt{Miglior spiegazione dei dati secondo } H_0}$

Rapporto di verosimiglianza generalizzato

Il rapporto delle verosimiglianze $r(\mathbf{x})$ è difficile da calcolare esattamente, per cui si preferisce usare il

$\Rightarrow$ *Rapporto di verosimiglianza generalizzato (GLR)*

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} f(\mathbf{x}|\theta)}{\sup_{\theta \in \Theta} f(\mathbf{x}|\theta)}$$

Osservazioni

- ✓ $0 \leq \lambda(\mathbf{x}) \leq 1$
- ✓ Se $\lambda(\mathbf{x}) \ll 1$ l'ipotesi H_0 è poco plausibile.
- ✓ Se $\lambda(\mathbf{x}) \simeq 1$ l'ipotesi H_0 è molto plausibile.

GLR Test

"Rifiutare l'ipotesi $\mathbf{H_0}$ *se e solo se* $\lambda(\mathbf{x}) < c$*."*
La costante c è determinata a partire dal livello di significatività α desiderato per il test:

$$\sup_{\theta \in \Theta_0} \Pr_\theta\{\lambda(\mathbf{x}) < c\} = \alpha$$

Osservazioni

- ✓ La maggior parte dei test statistici è basata sul GLR.
- ✓ La conoscenza della distribuzione di $\lambda(\mathbf{x})$ consente di calcolare la significatività del test.
- ✓ È possibile considerare ipotesi in cui vi siano più parametri incogniti.

Procedura operativa

1. Calcolare il GLR $\lambda(\mathbf{x})$.
2. Esprimere $\lambda(\mathbf{x})$ in funzione di una statistica $T(\mathbf{x})$ con distribuzione nota.
3. Riformulare il test $\lambda(\mathbf{x}) < c_\alpha$ come $T(\mathbf{x}) \lessgtr t^\alpha$, con t^α ricavato dalla distribuzione di $T(\mathbf{x})$, fissato il livello di significatività α.

Esempio

H_0	Statistica	Distribuzione
$\sigma^2 = \sigma_0^2$	$\frac{(n-1)S^2}{\sigma_0^2}$	$\chi^2(n-1)$
$\mu = \mu_0$	$\frac{\sqrt{n}(\bar{x}-\mu_0)}{S}$	$t(n-1)$
$\sigma_1^2/\sigma_2^2 = 1$	S_1^2/S_2^2	$F(n_1-1, n_2-1)$

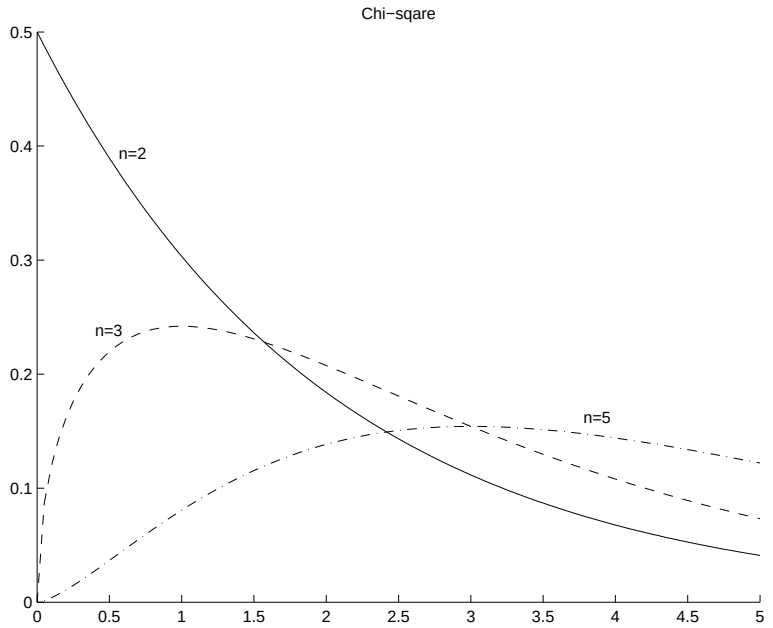
Test χ^2

- Date n v.a. $X_i \sim \mathcal{N}(\mu, \sigma^2)$, allora

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

- Date n osservazioni indipendenti di una v.a. $X \sim \mathcal{N}(\mu, \sigma^2)$, la varianza campionaria S^2 è tale che

$$\chi^2_c = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$



$\Rightarrow$ La statistica χ^2_c si utilizza per verificare ipotesi riguardanti il valore della *varianza di una popolazione*, anche con media ignota.

```
>> n=25;
>> alpha=0.05;
>> x_alpha=chi2inv(1-alpha,n-1);
```

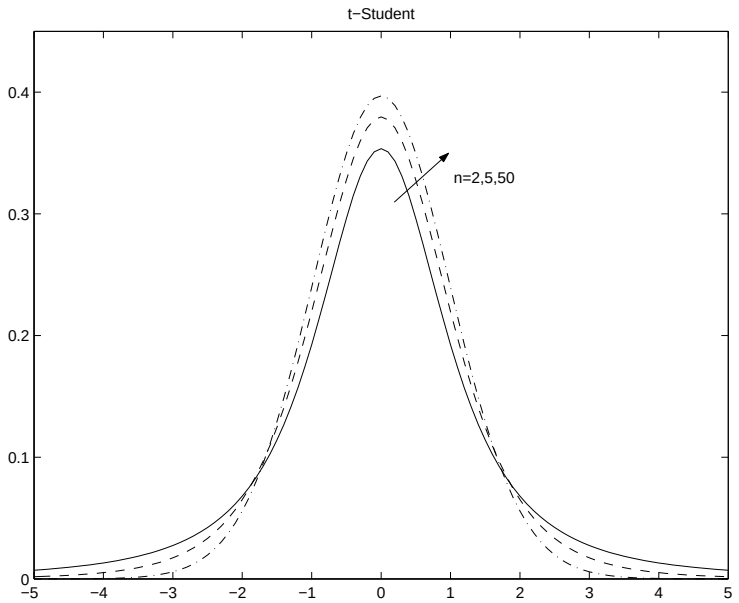
t-Test

- Date 2 v.a. $X \sim \mathcal{N}(0,1)$, $Y \sim \chi^2(n)$, allora

$$\frac{X}{\sqrt{Y/n}} \sim t(n)$$

- Date n osservazioni indipendenti di una v.a. $X \sim \mathcal{N}(\mu, \sigma^2)$, la media campionaria $\bar{X}$ e la deviazione standard campionaria S sono tali che

$$t_c = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$



$\Rightarrow$ La statistica t_c si utilizza per verificare ipotesi riguardanti il valore della *media di una o due popolazioni*, con *varianza ignota.*

```
>> ttest(X,m,alpha,tail)
>> ttest2(X,Y,alpha,tail)
```

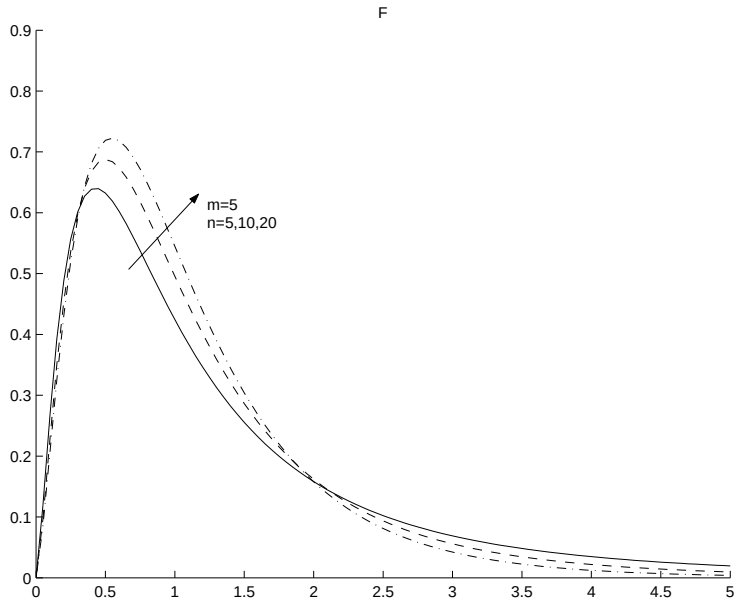
F-Test

- Date 2 v.a. $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$, allora

$$\frac{X/n_1}{Y/n_2} \sim F(n1, n2)$$

- Date (n_1, n_2) osservazioni indipendenti di due v.a. $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, la varianze campionarie S_1^2 e S_2^2 sono tali che

$$F_c = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$$



$\Rightarrow$ La statistica F_c si utilizza per verificare ipotesi riguardanti il valore delle *varianze di due popolazioni*, anche con medie ignote.

```
>> n1=25; n2=30;
>> alpha=0.05;
>> F_alpha=finv(1-alpha,n1-1,n2-1);
```

Goodness of fit

Sia $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ un campione aleatorio, corrispondente ad n realizzazioni indipendenti di una v.a. X.
Problema

- ◊ *"Il campione osservato proviene da una popolazione con distribuzione $F_0(x)$?"*

Osservazioni

- ✓ Non si fa alcuna ipotesi circa i parametri (media, varianza) della distribuzione $F_0(x)$.
- ✓ Si ipotizza *solo* la *forma funzionale* della distribuzione $F_0(x)$.

Ipotesi *non parametriche*

⇒ $\mathbf{H_0}$: $F(x) = F_0(x)$

⇒ $\mathbf{H_1}$: $F(x) \neq F_0(x)$

dove $F(x)$ denota la distribuzione *reale ignota* della v.a. X.

Test Chi-Quadro

Caso scalare, $X \in \mathcal{R}$.

- Si suddivida la retta reale in $k \ll n$ intervalli A_i.
- Si associ, ad ogni intervallo A_i, una v.a. *discreta* N_i indicante il numero di elementi del campione appartenente all'intervallo $i-$esimo.
- Supponendo vera l'ipotesi $\mathbf{H_0}$, è possibile calcolare la $P_{0i} = \Pr\{ X \in A_i \mid H_0\}$ del singolo esperimento.
- La densità congiunta degli n esperimenti è la multinomiale

$$f(N_1, \ldots, N_k | P_{01}, \ldots, P_{0k}) = \text{cost} \prod_{i=1}^{k} P_{0i}^{N_i}$$

Osservazione

✓ La $f(N_1, \ldots, N_k | P_{01}, \ldots, P_{0k})$ rappresenta la verosimiglianza, supponendo vera l'ipotesi $\mathbf{H_0}$.

Test Chi-Quadro

Idea

- Usare il rapporto di verosimiglianza

Problema

♢ La verosimiglianza $f(N_1, \ldots, N_k|H_1) = f(N_1, \ldots, N_k|P_{11}, \ldots, P_{1k})$, supponendo vera l'ipotesi H_1, non è nota.

Soluzione

✓ Stimo le probabilità P_{1i} sulla base delle frequenze raltive osservate $\hat{P}_{1i} = n_i/n$

Idea

✓ $\lambda(\mathbf{x}) = \frac{\texttt{densità ipotizzata}}{\texttt{densità osservata}}$

⇒ Si accetta l'ipotesi $\mathbf{H_0}$: $X \sim F_0(x)$ se $\lambda \simeq 1$, cioè se la distribuzione osservata è simile a quella ipotizzata

Test Chi-Quadro

Risultato

✓ $$\sum_{i=1}^{k} \frac{(n_i - nP_{0i})^2}{nP_{0i}} \sim \chi^2(k-1)$$

Regola di decisione

- Fissato il livello di significatività desiderato α, si accetta l'ipotesi $\mathbf{H_0}$ se

$$\sum_{i=1}^{k} \frac{(n_i - nP_{0i})^2}{nP_{0i}} < \chi_{\alpha,k-1}$$

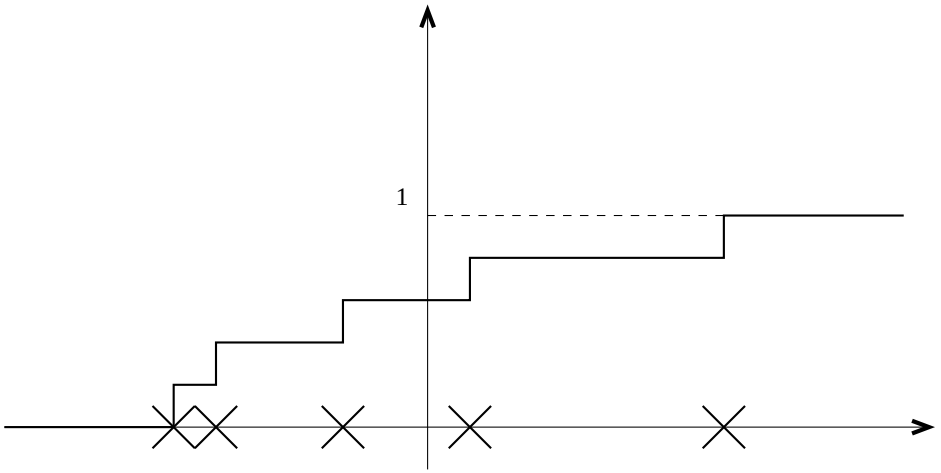
dove

- n_i= # elementi del campione appartenenti all'intervallo $i-$esimo A_i
- $P_{0i} = \Pr\{ X \in A_i \mid H_0\}$

Test di Kolmogorov-Smirnov

A partire dal campione $(x_1, x_2, ...x_n)$ si approssima la distribuzione di X con la *distribuzione empirica*

$$F^*(x) = \begin{cases} 0 & \text{se } x < x_{(1)} \\ \frac{r}{n} & \text{se } x_{(r)} \leq x < x_{(r+1)} \\ 1 & \text{se } x \geq x_{(n)} \end{cases}$$



Test di Kolmogorov-Smirnov

Si consideri la *statistica K-S*

- $D_n = \sup_x |F^*(x) - F_0(x)|$

Idea

✓ Accettare l'ipotesi H_0 se D_n è *"sufficientemente piccola"*.

Proprietà

- La distribuzione di D_n è indipendente dalla distribuzione ipotizzata $F_0(x)$.

Teorema.
Supponendo che l'ipotesi H_0 sia vera

$$\lim_{n\to\infty} \Pr\{D_n > \frac{z}{\sqrt{n}}\} = 2\sum_{r=1}^{\infty}(-1)^{r-1}e^{-2r^2z^2}$$

Test di Kolmogorov-Smirnov

- Fissato il livello di significatività desiderato α, dalle tavole è possiblile ricavare il valore $\frac{z}{\sqrt{n}}$ per cui

$$\Pr\{D_n > \frac{z}{\sqrt{n}}\} = \alpha$$

Regola di decisione

$\Rightarrow$ Accettare l'ipotesi H_0 se $D_n < \frac{z}{\sqrt{n}}$.

In MATLAB®

```
>> alpha=0.05

>> kstest(X,CDF,alpha,tail)
```