Tecniche statistiche di analisi del cambiamento

02-Ripasso inferenziale: Inferenza
(v. 1.2a, 9 ottobre 2018)
versione per stampa

Germano Rossi[1]
germano.rossi@unimib.it

[1]Dipartimento di Psicologia, Università di Milano-Bicocca

2018-19

Introduzione alla statistica inferenziale

- L'inferenza statistica serve per risolvere due problemi della ricerca

1) Essere sicuri dei risultati ottenuti dalla ricerca
2) Poter fare la ricerca su un campione anziché sulla popolazione

- Il punto 1 si risolve tramite il procedimento della "verifica d'ipotesi"
- il punto 2 tramite le procedure di campionamento

- Non sempre però questo è sufficiente

Logica del campionamento

- Chi si occupa di comportamento necessita di studiare il comportamento delle persone (la popolazione) e di trarre delle conclusioni
- Gli psicologi, di solito, possono misurare però solo una piccola parte di queste persone
- Per questo motivo, la maggior parte della ricerca in psicologia si basa su un piccolo campione di dati da cui derivano affermazioni generali
- La statistica descrittiva si applica a dati di qualsiasi ampiezza (in termini di casi statistici)
- Per cui le statistiche descrittive valgono sia per un campione sia per una popolazione

Campione e popolazione

- Il campionamento è l'estrazione di una parte della popolazione (secondo determinati criteri) per poterla studiare più agevolmente
- Ricordiamoci che una popolazione è l'insieme di tutti i casi statistici possibili con le caratteristiche che intendiamo studiare
- In altre parole: il campione è costituito da tutte le **misurazioni che ho fatto** in "questa" raccolta di dati
- La popolazione è costituita da tutte le misurazioni **che avrei potuto fare** in questa raccolta dati
- Il campione è sempre finito
- La popolazione può essere finita o infinita

Campione e popolazione

- Il termine “finito” indica che **esiste un numero** che rappresenta il massimo dei casi statistici considerabili (Ad es. *tutti gli studenti immatricolati a Psicologia nell’a.a. 2014/15*)
- Il termine “infinito” che **non esiste un numero** massimo di casi statistici (Ad es. *tutti tempi di reazione a un certo stimolo*)
- Se conosciamo le caratteristiche della popolazione, sarà facile
  - estrarre un campione che ben la rappresenti
  - riconoscere se il campione in esame rappresenta bene la popolazioni oppure no

Campione e popolazione

- Molto spesso nelle scienze sociali non si conoscono le caratteristiche della popolazione
- Se non la conosciamo dovremo cercare di estrarre un campione che sia una buona stima della popolazione (rappresentatività)
- Dal momento che non sempre conosciamo le caratteristiche della popolazione, le statistiche descrittive dei campioni sono usate come stima delle analoghe statistiche della popolazione
- Non abbiamo la certezza che queste stime siano “vere” ma sono le stime “migliori” dal momento che non conosciamo nulla!

Rappresentatività

- Generalizziamo il concetto di “buona stima” dicendo che il campione dev’essere rappresentativo
- Il campione selezionato “dovrebbe” rappresentare “in piccolo” la popolazione che si vuol studiare... quindi il campione dovrebbe avere le stesse caratteristiche della popolazione (e nella stessa proporzione)
- Non sempre si può fare (soprattutto in ambito clinico)
- Perciò, spesso si fa il contrario: posso generalizzare i risultati di “questo” campione ad una popolazione che abbia le stesse caratteristiche del campione
- Sulla base del campione rappresentativo, estendiamo i dati ottenuti all’intera popolazione (con le dovute cautele), tramite l’**inferenza statistica**

Rappresentatività

- Una volta selezionate le variabili che ci interessa studiare (*variabili dipendenti*),
- si individuano anche delle variabili che si ritengono importanti o che possono essere/produrre influenza sulle *variabili indipendenti*.
- Il campione dovrebbe distribuirsi (in queste variabili) proporzionalmente alla popolazione
- Un modo generalmente usato per avere la rappresentatività è quella della selezione casuale dei casi statistici dalla popolazione
- Questi campioni sono chiamati *campioni casuali* della popolazione
- In italiano, "casuale" ha più un significato di *arbitrario*, *informale*, *quello che capita*... Ma ha anche un significato diverso

Rappresentatività: Casuale o randomizzato

- Quando metto la mano nel sacchetto con i numeri della tombola, *non guardo dentro al sacchetto* proprio per poter estrarre un numero casuale…
- **Casuale** (in statistica) significa appunto che non uso strategie per selezionare un caso statistico a scapito di un altro…
- I libri italiani di statistica usano sia *casuale* sia *randomizzato*. Ma intendono lo stesso concetto
- In inglese, “random” enfatizza il fatto che tutti gli *eventi possibili* hanno la stessa possibilità di essere selezionati
- Un campione casuale (o randomizzato) è quindi uno dei *possibili campioni estraibili* da quella popolazione
- Inoltre, *tutti i casi selezionati per quel campione hanno avuto la stessa probabilità di essere selezionati*

Estrazione casuale

- Esistono numerosi modi per selezionare un campione casuale
- In molti casi si tratta di identificare ogni caso in qualche modo (ad es. con un numero)
- Poi è possibile (ad es.)
  - mettere tutti gli identificatori in un “contenitore”, da cui si selezionano “alla cieca” fino a raggiungere il numero di casi stabilito per il campione
  - usare un numero casuale generato in qualche modo (computer, calcolatrice, tavole dei numeri casuali) per selezionare
  - si ordinano gli identificatori e si selezionano quelli che sono in una certa posizione (ad es. 1 ogni 20)

Estrazione casuale

- In teoria, dopo aver selezionato un caso, dovremmo ri-immetterlo nel mucchio; altrimenti gli altri non avranno la stessa probabilità dei precedenti (1/n, 1/(n-1), 1/(n-2)...)
- non si fa, perché (con popolazioni grandi o infinite) la differenza delle probabilità è piccolissima
  - 1 su 1.000.000 = 0.0000010000 (1 milione)
  - 1 su 999.900 = 0.0000010001 (1 milione-100)

Verifica d'ipotesi

- Per verificare un'ipotesi, serve di averne una
- L'ipotesi che noi facciamo, implica che qualcosa influenzi, modifichi, cambi qualcos'altro: è un'ipotesi *rivoluzionaria*
- Ma proprio per questo, di solito, è generica
- Allora si lavora su un'ipotesi contraria, di solito, chiamata "ipotesi nulla" e indicata con H_0
- che afferma che non ci sarà nessun cambiamento, nessuna diversità, che tutto resterà uguale a prima; è l'*ipotesi conservativa*

Verifica d'ipotesi

- Ad es. noi ipotizziamo che bere bevande contenenti *steroli vegetali*, migliori la nostra salute perché diminuiscono il colesterolo
- Per verificarla decidiamo di usare una misura del colesterolo nel sangue
- Cosa ci aspettiamo? che diminuisca
- Di quanto? Non lo sappiamo, ma potremmo dire "abbastanza per escludere una diminuzione casuale"
- Perciò potremmo considerare il valore medio di colestero
- e ipotizzare che, dopo l'assunzione di steroli vegetali, il nostro colesterolo diminuisca in modo *non casuale*

Verifica d'ipotesi

- Un'ipotesi che ci permette di "lavorare" è: la media del colesterolo prima e dopo l'assunzione di steroli non cambia (ovvero $M_{prima} = M_{dopo}$
- Se la differenza fra prima e dopo è sufficientemente grande per escludere che sia casuale. . . .
- Chiamiamo quest'**ipotesi** come **nulla**, mentre l'ipotesi da cui è scaturita la nostra ricerca sarà l'**ipotesi alternativa** (o H_1
- Noi vorremmo che fosse vera l'alternativa, ma dobbiamo prima dimostrare che quella nulla è falsa!

Obiettivi della statistica inferenziale

Ci sono sostanzialmente tre obiettivi:

1. Stimare il valore puntuale dei parametri della popolazione
2. Calcolare la probabilità di ottenere un certo valore di una certa statistica in base alle caratteristiche (parametri) di una certa popolazione
3. Determinare la stima intervallare della statistica calcolata

Ci sono anche degli altri obiettivi:

- se c'è un effetto (ad es. un cambiamento), quanto è grande?
- se abbiamo accettato l'ipotesi alternativa, qual è la probabilità di aver fatto la scelta giusta?

Obiettivi della statistica inferenziale

- Ci sono diversi modi per rispondere a queste domande
- e ci servono dei ragionamenti ulteriori che implicano la *distribuzione campionaria*

- In pratica, finora ci siamo occupati di campioni, interessandoci ai singoli individui (distribuzioni di frequenza)
- Adesso ci interessiamo ai gruppi (popolazioni) che hanno le stesse caratteristiche del nostro campione

Distribuzione campionaria

- Se estraiamo un campione da una popolazione e il campione è rappresentativo di quella popolazione, il campione dovrebbe avere gli stessi indici statistici
- Ovviamente non è sempre vero
- Ma possiamo vedere/calcolare/studiare quanto potrebbero differire le statistiche calcolate su un campione rispetto ai parametri della popolazione da cui sono state tratte
- Per questo useremo campioni estratti da una popolazione come se fossero “casi statistici”
- E ci concentreremo sulla media (ma potremmo rifare lo stesso discorso su altre statistiche descrittive)

Distribuzione campionaria

- Ipotizziamo di estrarre un campione di 100 casi da una popolazione e di calcolare la media di una certa variabile
- Usiamo la variabile **Fondamentalismo** (calcolata su un campione di 659 persone) come popolazione. La sua media è $\mu = 90.3915$
- Estraiamo un campione casuale di 100 persone e calcoliamo la media di questo campione: M=91.46
- Estraiamo altri 20 campioni di ampiezza 100 dalla stessa popolazione e calcoliamo la media per ciascuno:

  87.83, 90.63, 91.90, 91.99, 90.10, 90.80, 93.84, 90.80, 89.80, 90.12, 90.71, 88.56, 89.67, 90.76, 87.77, 90.51, 89.78, 90.68, 90.40, 89.27

Distribuzione campionaria

Media	Scarto	
87.83	-2.56	
90.63	0.24	
91.90	1.51	
91.99	1.60	
90.10	-0.29	
90.80	0.41	
93.84	3.45	max
90.80	0.41	
89.80	-0.59	
90.12	-0.27	
90.71	0.32	
88.56	-1.83	
89.67	-0.72	
90.76	0.37	
87.77	-2.62	
90.51	0.12	
89.78	-0.61	
90.68	0.29	
90.40	0.01	min
89.27	-1.12	
91.46	1.07	

- Conoscendo la media della popolazione, possiamo sapere di quanto sono “buone” le medie dei singoli campioni
- Poiché vengono dalla stessa popolazione, la media di ogni campione estratto tenderà ad oscillare attorno alla media della popolazione
- La più vicina a μ ha uno scarto di 0.01; la più lontana uno scarto di 3.45
- Estraendo un solo campione non abbiamo nessuna sicurezza che sia il “migliore”

Distribuzione campionaria

Media	Scarto	
87.83	-2.56	
90.63	0.24	
91.90	1.51	
91.99	1.60	
90.10	-0.29	
90.80	0.41	
93.84	3.45	max
90.80	0.41	
89.80	-0.59	
90.12	-0.27	
90.71	0.32	
88.56	-1.83	
89.67	-0.72	
90.76	0.37	
87.77	-2.62	
90.51	0.12	
89.78	-0.61	
90.68	0.29	
90.40	0.01	min
89.27	-1.12	
91.46	1.07	

- Ma la media delle 21 medie, avrà un valore sicuramente più vicino alla media della popolazione: 90.3915

Medie	Scarto	
90.39		Media popolazione
90.35	-0.04	Media dei campioni

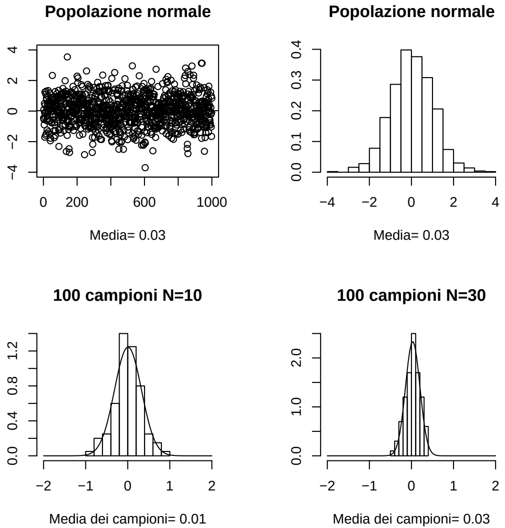
Distribuzione campionaria

- Anziché 20 campioni ne potremmo estrarre 10 mila, avremmo 10 mila medie e potremmo costruire una distribuzione di frequenza di quelle medie
- L'importante è che ogni campione sia casuale, ovvero
  - ogni caso di un singolo campione ha la stessa probabilità di essere estratto degli altri
  - ogni possibile campione estraibile dalla popolazione ha la stessa probabilità degli altri
- La distribuzione di frequenza che costruiremmo con le medie dei campioni si chiama **distribuzione campionaria delle medie**
- Se il numero di campioni estratto è sufficientemente elevato, le medie dei campioni tenderanno a distribuirsi secondo la curva della normale

Distribuzione campionaria

- Se effettivamente estraessimo un numero elevatissimo di campioni da una popolazione (metodo Monte Carlo), avremmo una *distribuzione sperimentale*, mentre quella su cui noi lavoreremo è una *distribuzione teorica*
- La distribuzione campionaria delle medie si basa sul **teorema del limite centrale** che afferma che, all'aumentare dell'*ampiezza dei campioni*, la distribuzione campionaria della media si avvicinerà sempre più ad una distribuzione normale, indipendentemente dalla forma delle misurazioni individuali
- Se una variabile si distribuisce normalmente, anche piccoli campioni produrranno una distribuzione campionaria normale
- Con variabili non normali, la distribuzione campionaria deve avere N uguale a 30 o 40

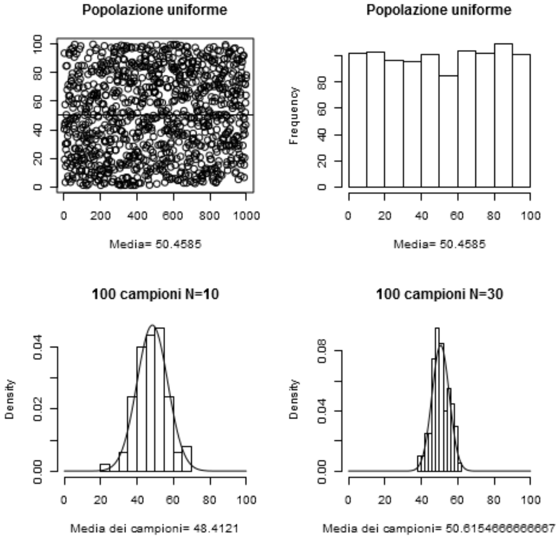
Distribuzione campionaria delle medie



A partire da una popolazione **distribuita normalmente** (1000 casi, valori -4; 4)

- abbiamo il grafico dei valori [1]
- l'istogramma delle frequenze [2]
- l'istogramma con normale di 100 campioni di ampiezza 10 [3]
- l'istogramma con normale di 100 campioni di ampiezza 30 [4]

Distribuzione campionaria delle medie



A partire da una popolazione **uniformemente distribuita** (1000 casi, valori 1-100)

- abbiamo il grafico dei valori [1]
- l'istogramma delle frequenze [2]
- l'istogramma con normale di 100 campioni di ampiezza 10 [3]
- l'istogramma con normale di 100 campioni di ampiezza 30 [4]

Distribuzione campionaria

- La distribuzione campionaria è una distribuzione di probabilità e per una numerosità (N) del campione superiore o uguale a 30, tende verso una curva stabile (e "normale") con

$$M_{\overline{X}} = \mu \quad \text{e} \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}$$

- $\sigma_{\overline{X}}$ è la *deviazione standard delle medie* anche conosciuta come **errore standard della media**
- indica quanto affidabile è ciascuna media campionaria
- valori piccoli indicano che, estraendo più campioni, le medie sarebbero abbastanza vicine fra loro
- valori grandi, che le medie campionarie sarebbero abbastanza disperse attorno a μ

Errore standard della media

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}$$

- È più piccolo della deviazione standard della popolazione, perché
- Singoli “punteggi estremi” (anomali) sono più probabili di “medie estreme”, quindi la distribuzione delle medie sarà meno variabile rispetto alla popolazione
- Al crescere di N, le medie campionarie sono più raggruppate e l’errore standard diventa sempre più piccolo

Verifica d'ipotesi

La moneta è truccata?

Evento	p	p_{cum}
10T	0,00098	0,00098
9T	0,00977	0,01075
8T	0,04395	0,05470
7T	0,11719	0,17189
6T	0,20508	0,37697
5T	0,24609	0,62306
4T	0,20508	0,82814
3T	0,11719	0,94533
2T	0,04395	0,98928
1T	0,00977	0,99905
0T	0,00098	1,00003

- Se lancio 10 volte una moneta e cade 10 volte sulla stessa faccia, è truccata?
- Se non fosse truccata, quante probabilità avrei di ottenere 10 volte una stessa faccia?
- 10 volte -> P(10)=0.00098*2=0.00196
- la probabilità è così bassa che la moneta è quasi sicuramente truccata!

La moneta è truccata?

Evento	p	p_{cum}
10T	0,00098	0,00098
9T	0,00977	0,01075
8T	0,04395	0,05470
7T	0,11719	0,17189
6T	0,20508	0,37697
5T	0,24609	0,62306
4T	0,20508	0,82814
3T	0,11719	0,94533
2T	0,04395	0,98928
1T	0,00977	0,99905
0T	0,00098	1,00003

- E se ottenessi 9 volte la stessa faccia?
- 9 volte -> P(9) = 0.00977
- ma se escono 9 facce avrebbero potuto essere anche 10, quindi sommiamo
- almeno 9 volte -> p(10)+p(9)=0.01075*2=0.0215
- quindi una probabilità di 2 su 100
- È sufficientemente piccola?

La moneta è truccata?

- Per rispondere, devo stabilire un limite di probabilità, sotto il quale decido che la moneta è truccata e sopra che non lo è!
- Se la probabilità che accada quell’evento è alta, allora “non è truccata”
- Se la probabilità è molto bassa (e l’evento accade) allora la moneta è truccata

Truccata	Non truccata	Truccata

Verifica di ipotesi

- Possibilità 1 (ipotesi nulla): la moneta NON È truccata

$$P(t) = P(c) = 0.5$$

- Possibilità 2 (ipotesi alternativa): la moneta È truccata

$$P(t) \neq P(c) \neq 0.5$$

- L'ipotesi nulla (indicata anche come H_0) è tale, perché si basa su informazioni che abbiamo già e di cui siamo sicuri (una moneta non truccata ha probabilità 1/2)
- L'ipotesi alternativa (indicata come H_1) è l'ipotesi che contrapponiamo a quella nulla

Ipotesi nulla e alternativa

- L'ipotesi nulla è l'unica su cui si possono effettivamente fare calcoli
- L'ipotesi alternativa apre, invece, ad un insieme di possibilità ($P(t) = 0.4$; $P(t) = 0.3$; $P(t) = .2 \ldots$) che non è possibile verificare tutte contemporaneamente
- Se l'ipotesi nulla si dimostra **probabile**, la accetteremo per vera.
- Se l'ipotesi nulla si dimostra **improbabile**, opteremo per quella alternativa
- L'ipotesi alternativa la verifichiamo "per assurdo", ovvero dimostrando **probabilmente falsa** l'ipotesi nulla

Procedimento di verifica

- L'ipotesi alternativa può essere di due tipi: semplice/composta e mono-/bi-direzionale
  - semplice (H_1: $\mu = 20$)
  - composta (H_1: $100 \leq \mu \leq 120$)

  - mono-direzionale (H_1: $\mu > 100$) o (H_1: $\mu < 100$)
  - bi-direzionale (H_1: $\mu \neq 100$)
- Calcoleremo la probabilità che la statistica calcolata sul nostro campione possa corrispondere a quella stimata della popolazione
- Non avremo mai una risposta sicura ma solo la probabilità di un errore!
- Ovvero: qualunque decisione prenderemo (H_0 o H_1), ci sarà sempre la possibilità che la nostra scelta sia sbagliata.

Verifica d'ipotesi: esempio 1

- Supponiamo di voler sapere se i bambini che crescono in famiglie che hanno animali domestici hanno QI diversi da quelli dei bambini senza animali domestici.
- Nella popolazione generale, il QI è distribuito normalmente con $\mu = 100$ e $\sigma = 15$
- Raccogliamo un campione casuale di 25 soggetti (N = 25) che vivono con animali domestici e misuriamo il loro QI. La media è $\overline{X} = 103.48$
- Le ipotesi nulle e alternative sono:

$$H_0 : \mu = 100$$

$$H_1 : \mu > 100$$

Verifica d’ipotesi: esempio 1

Abbiamo

- Una media di riferimento della popolazione: 100
- Una dev.st. di riferimento: 15
- La media di un campione: 103.45

- Sono 3 informazioni che ritroviamo anche nei punti z
- Possiamo usare i punti z per trovare la posizione di un gruppo rispetto a tutti gli altri gruppi della stessa ampiezza?
- Sì

I punti z per le medie campionarie

- Usiamo la distribuzione campionaria delle medie
- Calcoliamo il punto z e poi troviamo l'area corrispondente (ovvero la porobabilità)

$$z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{N}}}$$

In questo caso il punteggio grezzo è la media del campione, la media di riferimento è quella della popolazione e la deviazione standard per cui dividiamo è l'errore standard della media campionaria

Verifica d'ipotesi

- Calcoliamo il punto z della nostra media:

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} =$$

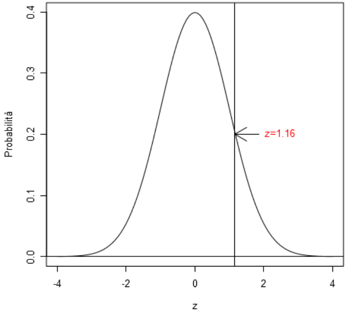
$$\frac{103.48 - 100}{\frac{15}{\sqrt{25}}} = 1.16$$

- Cerchiamo il punto z nella tavola A e troviamo l'area corrispondente

$$z(1.16) = .3730 = 37.30\%$$

$$50 + 37.30 = 87.30\%$$

$$50 - 37.30 = 12.70\%$$

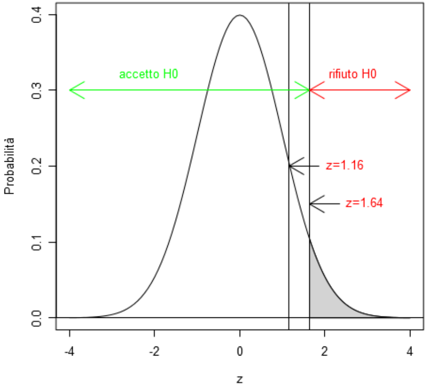


- Valore *p*: la probabilità associata al risultato, è il .1270

Criterio di significatività

Ci sono due possibili percorsi:

- Valore p: la probabilità associata al risultato
- Valore critico: il punto z che separa l'area di accettazione da quella di rifiuto di H_0
- se decidiamo di correre un rischio massimo del 5%,
- con p, la confronteremo direttamente con il 5%, quindi .127 vs. .05



- con il valore critico, cercheremo il punto z corrispondente ad un'area di 50%-5%=45% (punto z=1.64) e lo confronteremo con 1.16.

Criterio di significatività

- La regione critica si basa su un valore arbitrario, indicato con α, che è la probabilità di rifiutare H_0 quando, invece, è vera.
- Ci sono 2 tipi di errore:
  - Errore di primo tipo o α: l'errore di accettare per vera H_1 che, invece, è falsa ovvero di rifiutare H_0 che è invece vera
  - Errore di secondo tipo o β: l'errore di accettare per vera H_0 che, invece, è falsa ovvero rifiutare H_1 che invece è vera
  - Si chiama potenza di un test la sua capacità di accettare H_1 quando è vera [1-β]
  - Qualunque sia la decisione che prendiamo, corriamo un rischio calcolato
  - Il rischio viene calcolato tramite l'uso delle distribuzioni di probabilità

Relazioni fra errori e ipotesi

		Realtà	
		H_0 - Vera H_1 - Falsa	H_0 - Falsa H_1 - Vera
Ipotesi	Accetto H_0 Rifiuto H_1	Corretta $1-\alpha$	Errore II tipo β
	Rifiuto H_0 Accetto H_1	Errore I tipo α	Corretta $1-\beta$

- In psicologia si usano comunemente i seguenti valori di α:
  $\alpha = .05$ 5% *
  $\alpha = .01$ 1% **
  $\alpha = .001$ 0.1% ***

Verifica d'ipotesi

- La media del QI del campione di 25 bambini che vivono con animali domestici è di 103.48
- Questa media confrontata con i parametri della popolazione (tramite la distribuzione campionaria delle medie) sta a 1.16 dev. st. sopra $\mu_{\overline{X}}$
- E corrisponde all'87.30% (per H_0) o a 12.70% (per H_1)
- Ovvero, la probabilità di estrarre (da una popolazione con $\mu = 100$ e $\sigma = 15$) un campione di 25 bambini che abbiano un QI medio di 103.48, è di 12.70
- La probabilità di rifiutare H_0 (se vera) è del 12.70%
- Un evento abbastanza probabile, per cui possiamo concludere che vivere con animali domestici *non è connesso* ad un QI superiore alla media

Procedimento generale

La verifica d'ipotesi avviene sempre tramite

- Identificazione dell'ipotesi nulla (H_0) e ipotesi alternativa (H_1) (che è generalmente connessa con il test statistico scelto)
- Scelta di un test statistico e calcolo della relativa statistica (v_s)
- Scelto un determinato livello α, calcolo della probabilità associata (p) oppure identificazione del valore critico (v_c)
- Accettazione o rifiuto di H_0, in base alla scelta:

- **Con** p
- Se $p < \alpha$, rifiuto H_0
- Se $p > \alpha$, accetto H_0

- **Con** v_c (in genere)
- Se $|v_c| < |v_s|$, rifiuto H_0
- Se $|v_c| > |v_s|$, accetto H_0

Procedimento generale in pratica

Applicato al campione di 25 ragazzi con $M = 103.48$ e confrontato con una popolazione con $\mu = 100$ e $\sigma = 15$

- **Con** p
- Se $p < \alpha$, rifiuto H_0
- Se $p > \alpha$, accetto H_0

- $p = .1270$ (12.70%)
- $\alpha = .05$ (5%)
- Accetto H_0

- **Con** v_c (in genere)
- Se $|v_c| < |v_s|$, rifiuto H_0
- Se $|v_c| > |v_s|$, accetto H_0

- $v_s = 1.16$ (punto z)
- cerco il punto z corrispondente ad $\alpha = .05$ (5%)
- cerco nella tavola un'area pari a $.5000 - .0500 = .4500$ (45%) ed è (circa) 1.64 (o 1.65)
- $v_c = 1.64 > v_s = 1.03$: accetto H_0

Assunti richiesti

Il test statistico per la Media di un campione estratto da una determinata popolazione richiede alcuni assunti fondamentali:

- Gli individui nel campione sono stati selezionati in modo casuale e sono fra loro indipendenti rispetto alla popolazione
- La variabile misurata si distribuisce normalmente nella popolazione (ma considera anche il Teorema del Limite Centrale)