

Test di Ipotesi

Corso di Statistica di base

Giancarlo Ferrari

Test di ipotesi per una media campionaria

- Nelle lezioni precedenti abbiamo studiato gli intervalli di confidenza per le medie campionarie ed abbiamo visto come lo studio di tali intervalli ci aiuta a comprendere qual è l'intervallo all'interno del quale è contenuta la vera media della popolazione
- Questo può essere un modo per studiare quanto un parametro stimato della popolazione possa essere più o meno vicino a valori attesi o già noti
- In ogni caso abbiamo tratto le conclusioni basandoci sulle osservazioni effettuate su un campione

Test di ipotesi per una media campionaria

- Un altro approccio per stabilire quanto vicino (o lontano) possa essere un valore campionario da un parametro è attraverso l'esecuzione di un *Test statistico di ipotesi*
- Il test di ipotesi è una regola per decidere se un particolare si può ritenere appartenente alla popolazione dal quale è stato estratto
- Più in particolare è possibile stabilire se il campione estratto appartiene alla classe dei 'molto probabili' o dei 'poco probabili' o in maniera più semplice è una tecnica attraverso la quale si può misurare la discrepanza tra quanto si osserva nel campione e quanto si prevede nell'ipotesi nulla
- Ad esempio l'ipotesi nulla potrebbe riguardare la verifica che la media di una popolazione sia pari a μ_0

Test di ipotesi per una media campionaria

- Tale verifica potrebbe basarsi sul valore della media stimata su un campione
- L'approccio metodologico per verificare se la media campionaria differisce da un determinato valore è di formulare una ipotesi (Ipotesi 0 o ipotesi nulla) che viene convenzionalmente indicata con H_0
- L'ipotesi nulla è formulata in modo tale da voler verificare se la media campionaria NON differisce dalla popolazione dalla quale il campione è estratto

Test di ipotesi per una media campionaria

- All'ipotesi nulla corrisponde una ipotesi alternativa H_a
- L'ipotesi alternativa è speculare alla ipotesi nulla e le due ipotesi coprono tutti i possibili valori di μ a significare che una delle due deve essere vera
- Il passo successivo per l'esecuzione del test di ipotesi è di selezionare un campione dalla popolazione che si vuole sottoporre al test e calcolarne il valore medio \bar{x}
- Dal confronto tra la media \bar{x} di questo campione con la media postulata μ_0 vogliamo sapere se la differenza tra la media del campione e la media ipotizzata è tale da poter concludere che non vi è evidenza di differenza (oppure che vi è evidenza)

Test di ipotesi per una media campionaria

- La regola di decisione per stabilire se un campione appartiene o no ad una popolazione è di stimare la probabilità di osservare una deviazione pari o minore o maggiore rispetto a quella attesa sotto l'ipotesi nulla H_0
- Per poter rifiutare l'ipotesi nulla H_0 (e di conseguenza accettare l'ipotesi alternativa H_a) la probabilità del risultato osservato deve essere bassa QUANTO BASSA?

Test di ipotesi per una media campionaria

- Nella maggior parte dei casi si sceglie lo 0,05 come **valore critico** di probabilità e così rifiutiamo H_0 se la differenza osservata ha probabilità di verificarsi $\leq 0,05$
- In questo caso diciamo che i dati non sono compatibili con l'ipotesi nulla e il risultato del test è definito come *statisticamente significativo a livello dello 0,05*
- Il livello di significatività è indicato con α a significare che la probabilità di rifiutare erroneamente H_0 (quando essa è vera) è pari a 0,05

Test di ipotesi per una media campionaria

- Ad esempio si vuole verificare se il livello medio di colesterolo sierico della popolazione maschile di ipertesi fumatori μ sia uguale alla media della popolazione generale maschile di età compresa tra 20 e 74 anni ($\mu_0 = 211 \text{ mg}/100\text{ml}$; $\sigma_0 = 46 \text{ mg}/100\text{ml}$) che si assume essere distribuita normalmente

L'ipotesi 0 (ipotesi nulla) sarà: $H_0: \mu = \mu_0 = 211 \text{ mg}/100\text{ml}$

- L'ipotesi alternativa è una seconda ipotesi che contraddice H_0

Ipotesi alternativa: $H_a: \mu \neq 211 \text{ mg}/100\text{ml}$

Test di ipotesi per una media campionaria

- Una cosa importante da osservare prima di eseguire il test è che siamo interessati a deviazioni sia maggiori che inferiori rispetto al livello medio della popolazione e che pertanto eseguiamo un test bilaterale
- Il test di ipotesi si basa sulle nostre conoscenze sulla distribuzione campionaria della media e sulla possibilità di poter standardizzare la distribuzione campionaria in base ai valori della deviato standardizzata:

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

Test di ipotesi per una media campionaria

- Possiamo procedere in due modi
- MODO 1: Abbiamo già visto in precedenza che una media campionaria \bar{x} si distribuisce normalmente (per valori di n sufficientemente grande, in genere $n > 30$) con media $\bar{x} = \mu_0$ ed errore standard

$$SE = \frac{\sigma_0}{\sqrt{n}}$$

Test di ipotesi per una media campionaria

- Supponiamo di aver estratto un campione di 40 individui dalla popolazione degli ipertesi fumatori ed aver ottenuto un valore medio pari a $\bar{x} = 217 \text{ mg}/100\text{ml}$
- Sappiamo inoltre che (se estraessimo gli infiniti possibili campioni di dimensioni $n = 40$) il 95% delle medie campionarie saranno comprese nell'intervallo $\mu_0 \pm 1,96 \cdot SE$
- Nel nostro esempio diventa $211 \pm 1,96 \cdot \frac{46}{\sqrt{40}} = 211 \pm 14,25$

Test di ipotesi per una media campionaria

- Ci aspettiamo quindi che il 95% dei campioni di dimensione 20 oscillino tra valori pari a:

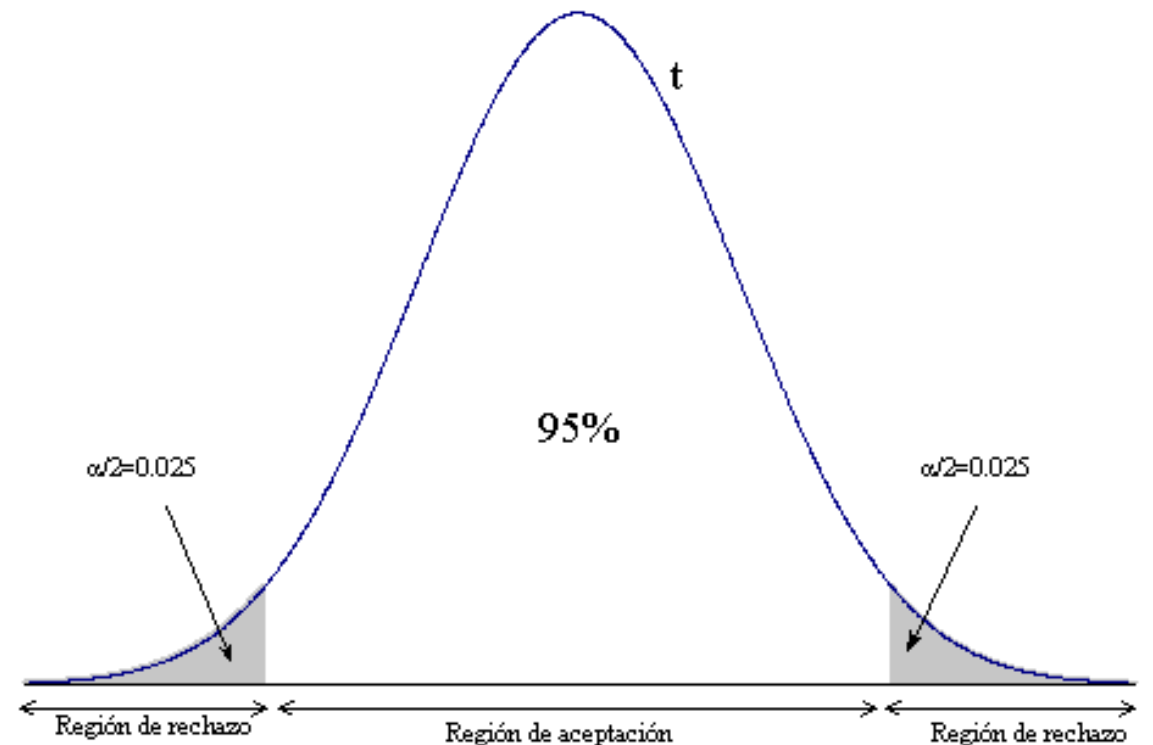
$$\textit{Limite inferiore} = 211 - 14,25 = 196,74$$

$$\textit{Limite superiore} = 211 + 14,25 = 225,25$$

- Se avessimo voluto ottenere un intervallo contenente il 99% dei valori delle media campionarie avremmo dovuto moltiplicare per 2,58 invece che per 1,96

Test di ipotesi per una media campionaria

- Il risultato ottenuto attraverso il nostro campione di 40 individui ci dice che la stima del livello di colesterolo è $\bar{x} = 217$
- Il valore pari a 217 è uno dei valori che ci saremmo aspettati di osservare il 95% delle volte poiché è compreso nell'intervallo $\pm 1,96$ SE



Test di ipotesi per una media campionaria

- MODO 2: sappiamo che per ciascun valore della media campionaria possiamo costruire una Deviata standardizzata

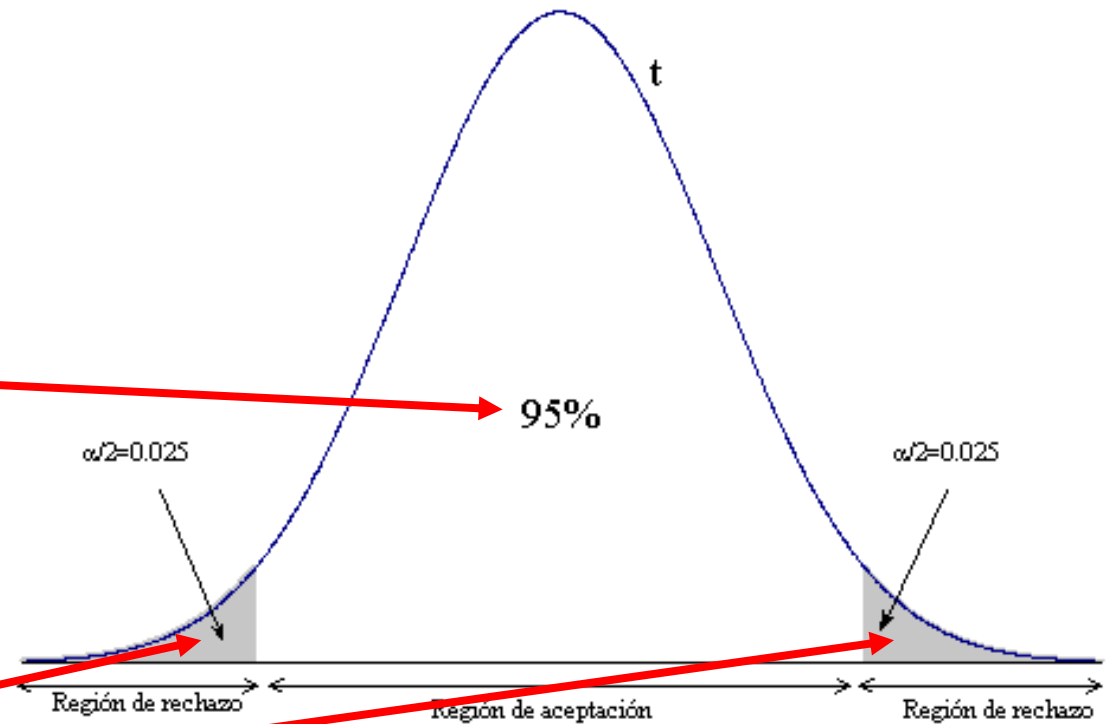
$$Z = \frac{\bar{x} - \mu_0}{\sigma_0 / \sqrt{n}}$$

- La devziata standardizzata è distribuita normalmente con media 0 e deviazione standard 1

Test di ipotesi per una media campionaria

- Abbiamo così formulato una Regola attraverso la quale accettiamo o rifiutiamo l'ipotesi nulla.

- Regione di accettazione di H_0



- Regioni di rifiuto di H_0

Test di ipotesi per una media campionaria

- Un test di ipotesi può essere paragonato ad un processo giudiziario
- L'individuo processato può essere innocente o colpevole ma per legge fino alla deposizione della sentenza è considerato innocente (ipotesi nulla)
- Dopo la raccolta delle prove la giuria considera l'imputato colpevole o innocente
- Se l'imputato è innocente ed il verdetto della giuria lo dichiara innocente allora è stato espresso il verdetto giusto.
- Il verdetto è giusto anche nel caso in cui l'imputato è colpevole e viene dichiarato tale

Test di ipotesi per una media campionaria

| Verdetto della giuria | Imputato | |
|-----------------------|--------------|--------------|
| | Innocente | Colpevole |
| Non colpevole | Corretto | Non corretto |
| Colpevole | Non corretto | Corretto |

Allo stesso modo la media reale della popolazione può essere o non essere μ_0

| Risultato del test | Popolazione | |
|-----------------------------|-------------------|----------------------|
| | $\bar{x} = \mu_0$ | $\bar{x} \neq \mu_0$ |
| Ipotesi nulla non rifiutata | Corretto | Non corretto |
| Ipotesi nulla rifiutata | Non corretto | Corretto |

Test di ipotesi per una media campionaria

- Nel test di ipotesi è quindi possibile commettere due tipi di errori
- Il primo tipo di errore (Errore di I tipo) α lo si può commettere quando l'ipotesi nulla è vera ma la si rifiuta
- Il secondo tipo di errore (Errore di II tipo) β lo si può commettere quando l'ipotesi nulla è falsa ma non la si rifiuta

Test di ipotesi per una media campionaria

- Gli errori sono possibili perché il test di ipotesi si basa sulla distribuzione dei valori delle medie campionarie
- L'argomento sui due tipi di errore verrà ripreso più avanti e per ora consideriamo solo l'errore α

Test di ipotesi per una media campionaria

- La domanda alla quale si tenta di rispondere quando si effettua un test di ipotesi è:

I dati campionari che indicazioni danno rispetto alla popolazione con cui lo si sta confrontando? E' uno dei possibili campioni che avrei potuto estrarre dalla popolazione con la quale lo sto confrontando? Qual è la probabilità di osservare un valore almeno uguale o superiore/inferiore a quello osservato?

Se il campione osservato è uno di quelli che si sarebbe osservato raramente (meno del 5% o dell'1% delle volte) allora questo autorizza a sollevare dubbi se effettivamente il campione appartiene alla popolazione con la quale lo si sta confrontando

Test di ipotesi per una media campionaria

- ESEMPIO N. 2: Supponiamo di conoscere che la popolazione dei Marziani ha un'altezza media pari a 135 cm e deviazione standard uguale a 12
- Definiamo quindi $\mu_0 = 135$ e $\sigma_0 = 12$
- Facciamo un viaggio su Venere e raccogliamo i dati sulle altezze di 100 individui
- L'ipotesi che vogliamo sottoporre a verifica è: l'altezza media dei venusiani è uguale o diversa a quella dei marziani?

Test di ipotesi per una media campionaria

- L'ipotesi 0 (che chiamiamo H_0) è: $\bar{x} = \mu_0$
- L'ipotesi alternativa (che chiamiamo H_a) è $\bar{x} \neq \mu_0$

Dove \bar{x} è la media campionaria calcolata sul campione di 100 individui abitanti su Venere

Poniamo il caso che $\bar{x}_{ven} = 136,5$

Test di ipotesi per una media campionaria

- Prima di eseguire un test di ipotesi si deve sempre decidere se si è interessati alle deviazioni da μ_0 che possono verificarsi in entrambe le direzioni (test bilaterale) oppure in una sola direzione (test unilaterale)
- La decisione dovrebbe essere presa prima di selezionare il campione casuale e non essere influenzata dal risultato ottenuto nel campione
- Se si sa a priori che il valore di \bar{x} non può essere minore di μ_0 i valori di \bar{x} che forniranno evidenza contraria all'ipotesi nulla sono solo quelli maggiori di μ_0

Test di ipotesi per una media campionaria

- In questo caso l'ipotesi nulla può essere più correttamente definita come:

$$H_0: \bar{x} \leq \mu_0$$

- L'ipotesi alternativa può quindi essere formulata come:

$$H_a: \bar{x} > \mu_0$$

Test di ipotesi per una media campionaria

Esempio n. 3

- Si è stabilito su un gran numero di cani affetti da tumore in una sede particolare e in un particolare stadio clinico che il tempo medio di sopravvivenza dalla diagnosi è pari 38,3 mesi con una deviazione standard di 43,3 mesi
- In uno studio clinico 100 cani furono trattati con una nuova tecnica terapeutica e per essi il tempo medio di sopravvivenza fu 46,9 mesi
- Tale apparente aumento si può spiegare solo per una fluttuazione casuale o può essere attribuibile alla nuova terapia?

Test di ipotesi per una media campionaria

- In questo caso siamo interessati a verificare che la nuova terapia aumenti i tempi di sopravvivenza e quindi siamo interessati ad effettuare un test unilaterale
- Possiamo formulare l'ipotesi nulla come segue:

$$H_0: \bar{x} \leq \mu_0$$

- L'ipotesi alternativa sarà:

$$H_a: \bar{x} > \mu_0$$

Test di ipotesi per una media campionaria

- Come livello di significatività scegliamo il 5% ($\alpha \leq 0,05$) che corrisponde ad un valore di Z pari a 1,64
- Ponendo quindi $n = 100$ e $\bar{x} = 46,9$ si ottiene il seguente valore della Deviata Standardizzata:

$$Z = \frac{46,9 - 38,3}{\frac{43,3}{\sqrt{100}}} = \frac{8,6}{4,33} = 2,0$$

Test di ipotesi con varianza ignota

Test di ipotesi per una media campionaria (con varianza ignota)

- Un caso un po' diverso dal precedente è quando ci si trovi a voler fare una verifica su una media campionaria ma non si dispone di informazioni sulla varianza (e quindi sulla deviazione standard) della popolazione
- In questo caso l'unica informazione sulla varianza della popolazione potrebbe essere quella contenuta nel campione e sembra ragionevole poter sostituire σ con la stima della deviazione standard s

Test di ipotesi per una media campionaria (con varianza ignota)

- In questo caso è opportuno utilizzare la distribuzione t di Student:

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

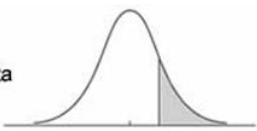
- La t di Student segue una distribuzione non molto dissimile da quella di Z (con media 0 e deviazione standard 1) a condizione che n sia sufficientemente grande poiché se n è piccolo s può differire considerevolmente da σ

Test di ipotesi per una media campionaria (con varianza ignota)

- Abbiamo già visto che t segue una distribuzione nota come *distribuzione t con $n-1$ gradi di libertà*
- Al crescere dei gradi di libertà la distribuzione t tende alla distribuzione normale standard Z

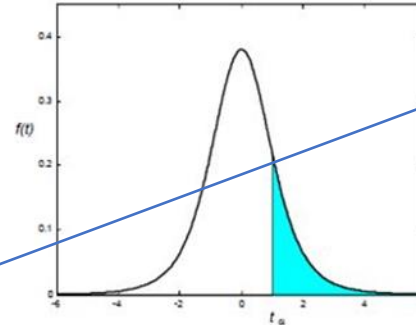
Test di ipotesi per una media campionaria (con varianza ignota)

Tavola 1B: Aree della distribuzione normale standardizzata



| Z | 0 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0 | 0.5000 | 0.4960 | 0.4920 | 0.4880 | 0.4840 | 0.4801 | 0.4761 | 0.4721 | 0.4681 | 0.4641 |
| 0.1 | 0.4602 | 0.4562 | 0.4522 | 0.4483 | 0.4443 | 0.4404 | 0.4364 | 0.4325 | 0.4286 | 0.4247 |
| 0.2 | 0.4207 | 0.4168 | 0.4129 | 0.4090 | 0.4052 | 0.4013 | 0.3974 | 0.3936 | 0.3897 | 0.3859 |
| 0.3 | 0.3821 | 0.3783 | 0.3745 | 0.3707 | 0.3669 | 0.3632 | 0.3594 | 0.3557 | 0.3520 | 0.3483 |
| 0.4 | 0.3446 | 0.3409 | 0.3372 | 0.3336 | 0.3300 | 0.3264 | 0.3228 | 0.3192 | 0.3156 | 0.3121 |
| 0.5 | 0.3085 | 0.3050 | 0.3015 | 0.2981 | 0.2946 | 0.2912 | 0.2877 | 0.2843 | 0.2810 | 0.2776 |
| 0.6 | 0.2743 | 0.2709 | 0.2676 | 0.2643 | 0.2611 | 0.2578 | 0.2546 | 0.2514 | 0.2483 | 0.2451 |
| 0.7 | 0.2420 | 0.2389 | 0.2358 | 0.2327 | 0.2296 | 0.2266 | 0.2236 | 0.2206 | 0.2177 | 0.2148 |
| 0.8 | 0.2119 | 0.2090 | 0.2061 | 0.2033 | 0.2005 | 0.1977 | 0.1949 | 0.1922 | 0.1894 | 0.1867 |
| 0.9 | 0.1841 | 0.1814 | 0.1788 | 0.1762 | 0.1736 | 0.1711 | 0.1685 | 0.1660 | 0.1635 | 0.1611 |
| 1 | 0.1587 | 0.1562 | 0.1539 | 0.1515 | 0.1492 | 0.1469 | 0.1446 | 0.1423 | 0.1401 | 0.1379 |
| 1.1 | 0.1357 | 0.1335 | 0.1314 | 0.1292 | 0.1271 | 0.1251 | 0.1230 | 0.1210 | 0.1190 | 0.1170 |
| 1.2 | 0.1151 | 0.1131 | 0.1112 | 0.1093 | 0.1075 | 0.1056 | 0.1038 | 0.1020 | 0.1003 | 0.0985 |
| 1.3 | 0.0968 | 0.0951 | 0.0934 | 0.0918 | 0.0901 | 0.0885 | 0.0869 | 0.0853 | 0.0838 | 0.0823 |
| 1.4 | 0.0808 | 0.0793 | 0.0778 | 0.0764 | 0.0749 | 0.0735 | 0.0721 | 0.0708 | 0.0694 | 0.0681 |
| 1.5 | 0.0668 | 0.0655 | 0.0643 | 0.0630 | 0.0618 | 0.0606 | 0.0594 | 0.0582 | 0.0571 | 0.0559 |
| 1.6 | 0.0548 | 0.0537 | 0.0526 | 0.0516 | 0.0505 | 0.0495 | 0.0485 | 0.0475 | 0.0465 | 0.0455 |
| 1.7 | 0.0446 | 0.0436 | 0.0427 | 0.0418 | 0.0409 | 0.0401 | 0.0392 | 0.0384 | 0.0375 | 0.0367 |
| 1.8 | 0.0359 | 0.0351 | 0.0344 | 0.0336 | 0.0329 | 0.0322 | 0.0314 | 0.0307 | 0.0301 | 0.0294 |
| 1.9 | 0.0287 | 0.0281 | 0.0274 | 0.0268 | 0.0262 | 0.0256 | 0.0250 | 0.0244 | 0.0239 | 0.0233 |
| 2 | 0.0228 | 0.0222 | 0.0217 | 0.0212 | 0.0207 | 0.0202 | 0.0197 | 0.0192 | 0.0188 | 0.0183 |
| 2.1 | 0.0179 | 0.0174 | 0.0170 | 0.0166 | 0.0162 | 0.0158 | 0.0154 | 0.0150 | 0.0146 | 0.0143 |
| 2.2 | 0.0139 | 0.0136 | 0.0132 | 0.0129 | 0.0125 | 0.0122 | 0.0119 | 0.0116 | 0.0113 | 0.0110 |
| 2.3 | 0.0107 | 0.0104 | 0.0102 | 0.0099 | 0.0096 | 0.0094 | 0.0091 | 0.0089 | 0.0087 | 0.0084 |
| 2.4 | 0.0082 | 0.0080 | 0.0078 | 0.0075 | 0.0073 | 0.0071 | 0.0069 | 0.0068 | 0.0066 | 0.0064 |
| 2.5 | 0.0062 | 0.0060 | 0.0059 | 0.0057 | 0.0055 | 0.0054 | 0.0052 | 0.0051 | 0.0049 | 0.0048 |
| 2.6 | 0.0047 | 0.0045 | 0.0044 | 0.0043 | 0.0041 | 0.0040 | 0.0039 | 0.0038 | 0.0037 | 0.0036 |
| 2.7 | 0.0035 | 0.0034 | 0.0033 | 0.0032 | 0.0031 | 0.0030 | 0.0029 | 0.0028 | 0.0027 | 0.0026 |
| 2.8 | 0.0026 | 0.0025 | 0.0024 | 0.0023 | 0.0023 | 0.0022 | 0.0021 | 0.0021 | 0.0020 | 0.0019 |
| 2.9 | 0.0019 | 0.0018 | 0.0018 | 0.0017 | 0.0016 | 0.0016 | 0.0015 | 0.0015 | 0.0014 | 0.0014 |
| 3 | 0.0013 | 0.0013 | 0.0013 | 0.0012 | 0.0012 | 0.0011 | 0.0011 | 0.0011 | 0.0010 | 0.0010 |
| 3.1 | 0.0010 | 0.0009 | 0.0009 | 0.0009 | 0.0008 | 0.0008 | 0.0008 | 0.0008 | 0.0007 | 0.0007 |
| 3.2 | 0.0007 | 0.0007 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0005 | 0.0005 | 0.0005 |
| 3.3 | 0.0005 | 0.0005 | 0.0005 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0003 |
| 3.4 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0002 |
| 3.5 | 0.0002 | 0.0002 | 0.0002 | 0.0002 | 0.0002 | 0.0002 | 0.0002 | 0.0002 | 0.0002 | 0.0002 |
| 3.6 | 0.0002 | 0.0002 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| 3.7 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| 3.8 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| 3.9 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 4 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

Distribuzione t di Student



Ad esempio:
gdl=9 $\alpha=0.05$:
 $t_{0.05}=1.833$

Se gdl=9, trovare il valore di t_α tale che la somma dell'area a destra di t_α e dell'area a sinistra di $-t_\alpha$ vale $\alpha = 0.05$

Area totale delle due code = $\alpha = 0.05$
area a destra di t_α (una coda) = $\alpha/2 = 0.025$.

$t_\alpha = t_{0.025} = 2.262$

| v | $\alpha = 0.10$ | $\alpha = 0.05$ | $\alpha = 0.025$ | $\alpha = 0.01$ | $\alpha = 0.005$ | v |
|----------|-----------------|-----------------|------------------|-----------------|------------------|----------|
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 1 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 2 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 3 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 4 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 6 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 7 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 8 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 9 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 10 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 11 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 12 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 13 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 14 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 15 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 16 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 17 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 18 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 19 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 20 |
| 21 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 21 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 22 |
| 23 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 23 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 24 |
| 25 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 25 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 26 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 27 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 28 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 29 |
| ∞ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | ∞ |

Test di ipotesi per una media campionaria (con varianza ignota)

- Riprendiamo l'esempio precedente dove si voleva verificare se il tempo di sopravvivenza di cani sottoposti ad una particolare terapia (tempo di sopravvivenza pari a 46,3) era da ritenersi appartenente ad una popolazione di cani dove il valore di sopravvivenza era noto e pari a 38,3 mesi
- In questo caso però (a differenza del precedente esempio) si assume di non sapere il valore della deviazione standard della popolazione ed inoltre che questa verrà stimata da un campione di cani pari a 30
- Supponiamo che tale valore di deviazione standard sia stato stimato dai dati campionari essere pari a 43,3 (uguale al precedente)

Test di ipotesi per una media campionaria (con varianza ignota)

- Questa è una situazione in cui sarebbe inappropriato usare la distribuzione Z perché la σ non è nota
- Il valore di t è pertanto:

$$t = \frac{46,9 - 38,3}{43,3 / \sqrt{30}} = 1,08$$

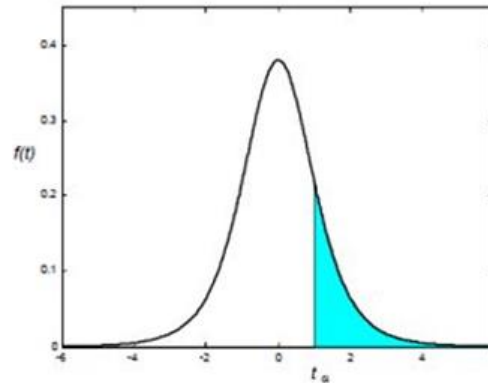
Test di ipotesi per una media campionaria (con varianza ignota)

$$t = 1,08$$

Il valore di t è inferiore al valore 1,699 in corrispondenza del valore critico di $\alpha = 0,05$ (per g.l. = 29)

Non vi sono elementi per rifiutare l'ipotesi nulla che la nuova terapia abbia un valore medio diverso da quello noto della popolazione

Distribuzione
t di Student



Ad esempio:
gdl=9 $\alpha=0.05$:
 $t_{0.05}=1.833$

Se gdl=9, trovare il valore di t_{α} tale che la somma dell'area a destra di t_{α} e dell'area a sinistra di $-t_{\alpha}$ vale $\alpha = 0.05$

Area totale delle due code = $\alpha = 0.05$
area a destra di t_{α} (una coda) = $\alpha/2 = 0.025$.

$$t_{\alpha} = t_{0.025} = 2.262$$

| v | $\alpha = 0.10$ | $\alpha = 0.05$ | $\alpha = 0.025$ | $\alpha = 0.01$ | $\alpha = 0.005$ | v |
|----------|-----------------|-----------------|------------------|-----------------|------------------|----------|
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 1 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 2 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 3 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 4 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 6 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 7 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 8 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 9 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 10 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 11 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 12 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 13 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 14 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 15 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 16 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 17 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 18 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 19 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 20 |
| 21 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 21 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 22 |
| 23 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 23 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 24 |
| 25 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 25 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 26 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 27 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 28 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 29 |
| ∞ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | ∞ |

Test di ipotesi per una media campionaria (con varianza ignota)

- E' importante osservare che il test di significatività ha prodotto un risultato diverso (opposto) da quello prodotto con numerosità campionaria pari a 100
- E' importante anche osservare la diversa modalità di lettura della tavola della distribuzione t di Student dove gli incroci tra le righe corrispondenti ai diversi gradi di libertà (la numerosità campionaria) e le colonne corrispondenti ai diversi livelli di errore α contengono i valori di t (e non le aree, corrispondenti ai valori delle probabilità, come nelle tavole della distribuzione Z)

Test di ipotesi per una proporzione campionaria

Test di ipotesi per una proporzione campionaria

- Quanto visto fino ad ora si applica a variabili di tipo quantitativo e può essere applicato indifferentemente per variabili di tipo continuo o discreto
- Lo stesso approccio lo si può seguire nella verifica di ipotesi per variabili di tipo qualitativo e dove quindi le inferenze vengono effettuate sulle frequenze
- Nello studio delle frequenze siamo interessati in genere alla proporzione (più che al numero di volte) con cui si verifica un evento

Test di ipotesi per una proporzione campionaria

- C'è un esempio in letteratura medica che è estremamente importante
- Nella metà del 19mo secolo l'ospedale universitario di Vienna aveva due reparti di ostetricia.
- Ogni anno nascevano in ciascun reparto circa 3.500 bambini.
- Vi erano però due differenze importanti tra i due reparti

Test di ipotesi per una proporzione campionaria

- Nel primo reparto tutte le nascite erano supervisionate da ostetrici e studenti in medicina mentre nel secondo reparto la supervisione era effettuata da levatrici
- Nel primo reparto la proporzione di donne morte per febbre puerperale era compresa tra 0,17 e 0,23 mentre nel secondo reparto era circa 0,017
- Ignac Semmelweiss, assistente del professore di ostetricia era convinto che questa differenza non fosse soltanto dovuta al caso

Test di ipotesi per una proporzione campionaria

- La sua ricerca lo portò a concludere che questa differenza esisteva perché gli ostetrici e gli studenti eseguivano molte necroscopie al giorno e poiché la teoria microbica delle malattie non era stata ancora proposta non si praticava alcuna norma igienica e si passava senza alcuna precauzione dalla sala di necroscopia alla sala parto
- Ritenendo che ciò fosse la base del problema Semmelweis modificò la procedura e si prodigò affinché gli ostetrici si lavassero le mani in una soluzione clorata prima di assistere ad un parto

Test di ipotesi per una proporzione campionaria

- L'anno successivo la proporzione di donne morte era scesa a 0,012 nel primo reparto e a 0,013 nel secondo
- Come spesso accade con le innovazioni le sue scoperte non furono accettate e perse il posto di lavoro
- Il problema dello studio della differenza tra proporzioni non è dissimile dallo studio delle differenze tra valori medi di variabili di tipo quantitativo e nel caso dell'esempio di Semmelweis l'ipotesi da verificare era se le due proporzioni diverse potevano ritenersi provenienti dalla stessa popolazione oppure se vi erano differenze che ne potevano giustificare la differenza

Test di ipotesi per una proporzione campionaria

- Riprendiamo l'esempio fatto in precedenza sul trattamento di cani con una terapia dopo una diagnosi di un particolare tumore con la differenza che questa volta invece che i tempi di sopravvivenza vogliamo studiare la proporzione di cani che sopravvive ad un anno dalla diagnosi
- Supponiamo di conoscere che la proporzione di cani sopravvissuti ad un anno dalla diagnosi è pari a $\pi = 0,3$

Test di ipotesi per una proporzione campionaria

- L'ipotesi che si vuole verificare è se la nuova terapia non modifica o modifica la proporzione di sopravvivenza ad un anno. Quindi:

$$H_0: \hat{p} = \pi = 0,3$$

oppure

$$H_a: \hat{p} \neq \pi \neq 0,3$$

Test di ipotesi per una proporzione campionaria

- Si vuole eseguire un test bilaterale ad un livello di significatività di 0,05 (valore critico di $\alpha = 0,05$)
- A tale scopo viene esaminato un campione di 40 cani sottoposti alla nuova terapia sul quale si misura la proporzione che è ancora viva dopo 1 anno (poniamo che $\hat{p} = 0,5$)
- Verifichiamo che $n\hat{p} > 5$ e $n(1 - \hat{p}) > 5$

Test di ipotesi per una proporzione campionaria

- A questo punto possiamo utilizzare il test statistico Z

$$Z = \frac{\hat{p} - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}}$$

che per il campione pari a 40 diventa:

$$Z = \frac{0,5 - 0,3}{\sqrt{\frac{0,3(1 - 0,3)}{40}}} = 2,76$$

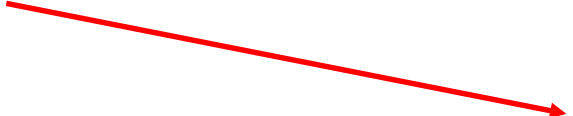
Test di ipotesi per una proporzione campionaria

- Un risultato abbastanza simile lo si sarebbe ottenuto utilizzando i valori delle frequenze
- Data l'ipotesi nulla con 20 individui su 40 che sopravvivono ad un anno e contro un valore atteso di $np = 12$ (su un numero di prove pari a 40) la quantità:

$$Z = \frac{|20 - 12| - \frac{1}{2}}{\sqrt{40 \cdot 0,3 \cdot (1 - 0,3)}} = 2,58$$

Test di ipotesi per una proporzione campionaria

- E' importante osservare che i due valori di Z stimati differiscono leggermente e il motivo di tale differenza è che con il secondo metodo è stata introdotta una correzione per la continuità che ha reso la stima più conservativa


$$Z = \frac{|20 - 12| - \frac{1}{2}}{\sqrt{40 \cdot 0,3 \cdot (1 - 0,3)}} = 2,58$$

Se la correzione per la continuità viene rimossa si ottiene lo stesso valore del metodo precedente

$$Z = \frac{20 - 12}{\sqrt{40 \cdot 0,3 \cdot (1 - 0,3)}} = 2,76$$

Conclusioni

- Negli esempi visti ci siamo occupati di utilizzare un test statistico di ipotesi per confrontare la media o la proporzione di una singola popolazione con un valore di μ_0 noto
- Abbiamo definito i criteri attraverso i quali poter concludere se vi sono sufficienti evidenze che il valore rilevato nel campione sia sufficientemente vicino da non fare rigettare l'ipotesi nulla oppure che il valore sia così 'lontano' dal valore atteso da mettere in dubbio la veridicità della ipotesi nulla
- Abbiamo trattato di variabili di tipo quantitativo e di tipo binomiale ed abbiamo tralasciato (per ora) le variabili di tipo nominale dove le modalità sono più di due

Conclusioni

- I concetti introdotti sono necessari per passare ai test di ipotesi effettuati tra le medie di due diverse popolazioni e dove i parametri di queste due popolazioni non sono noti
- Questa è la situazione più comune in molte applicazioni pratiche