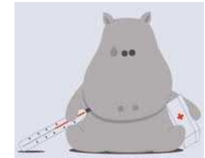
Test d’Ipotesi - Esempi



ESEMPIO 1 - Si vuole verificare se le lattine di caffè confezionate automaticamente da una ditta, contengono in media il peso dichiarato di $\mu = 250\,g$.

A tale scopo si prende un campione di 50 lattine, se ne pesa il conenuto e si calcola il peso medio.

Naturalmente questo peso non è in generale uguale al peso medio di tutte le lattine

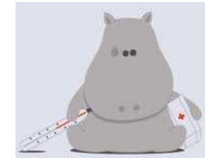
Ma quando questo risultato si allontana abbastanza da $\mu = 250\,g$ da permetterci di rifiutare il peso dichiarato dalla ditta con ragionevole probabilità di non sbagliare?

ESEMPIO 2 - Si vuole sottoporre a test l'affermazione di un produttore di vernici secondo cui il tempo medio di asciugatura di una nuova vernice è non superiore a 30 minuti.

A questo scopo si prende un campione di 40 lattine di vernice, si effettuano 40 prove di verniciatura con la vernice delle diverse confezioni e si calcola il tempo medio di asciugatura.

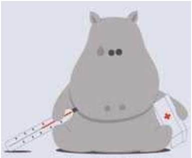
Quando il tempo medio si allontana abbastanza dal tempo dichiarato da permetterci di rifiutare l'affermazione del produttore con ragionevole probabilità di non sbagliare?

Test d'Ipotesi - Definizioni



- Un'ipotesi formulata in termini di parametri di una popolazione, come media o varianza, è detta **ipotesi statitistica**.
- Il procedimento che consente di rigettare o accettare un'ipotesi statistica, utilizzando i dati di un campione abbastanza numeroso, viene chiamato **test d'ipotesi**.
- Le possibili conclusioni di un test d'ipotesi sono:
  - l'ipotesi statistica è rigettata
  - l'ipotesi statistica non è rigettata

Test d'Ipotesi - Ipotesi Zero



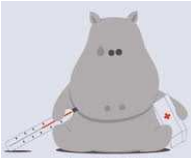
Noi ci occuperemo della localizzazione della media perchè è spesso utile e inoltre è quella in cui possiamo usare il teorema centrale del limite.

La distribuzione gaussiana delle medie consente di sottoporre ad esame critico ipotesi effettuate su una popolazione.

1. Si suppone che venga fatta una affermazione che localizzi la media μ della popolazione (**ipotesi zero**).
2. Per verificare l'attendibilità dell'ipotesi, si seleziona un campione casuale sufficientemente grande ($n > 30$) di cui si calcola media campionaria $\overline{x}$ e deviazione standard campionaria s .

3. Si misura la distanza, in termini di deviazioni standard, di μ dalla media osservata “sul campo” $\overline{x}$.
4. Quanto più $\overline{x}$ si allontana da μ, tanto più diventiamo sospettosi circa la validità dell’ipotesi riguardante la media e siamo condotti a **rigettare** l’ipotesi.

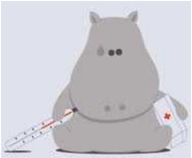
Test d'Ipotesi - Errori



- Se l'ipotesi zero è vera, ma erroneamente viene rigettata, si commette un errore di primo tipo
- Se l'ipotesi zero è falsa , ma erroneamente non viene rigettata, si commette un errore di secondo tipo

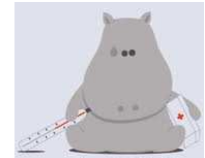
NOI CI OCCUPEREMO DI "RENDERE PICCOLO" l' ERRORE DI PRIMO TIPO

Test d'Ipotesi - Livello Significatività



- Così facendo ci assumiamo un rischio, quello che il campione scelto avesse media $\overline{x}$ realmente molto lontana da μ e che la media μ fosse accettabile.
- Il livello di rischio di prendere una decisione sbagliata, che siamo disposti a correre, dipende dalle circostanze.
- Solitamente si accetta un rischio dell' 1% o del 5%. Il rischio di prendere la decisione sbagliata sulla scorta dei dati del campione è detto **livello di significatività** del test.

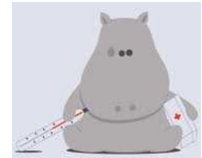
Test d'Ipotesi - Esempio 1



Esempio 1 - È stato affermato che il peso medio degli individui adulti di una certa nazione è $\mu = 68.5\,kg$. Volendo sottoporre questa ipotesi a verifica, si considera un campione casuale di 625 individui che vengono pesati. Si ottiene un valor medio campionario $\overline{x} = 69.1\,kg$ con una deviazione standard campionaria $s = 7\,kg$.

Con livello di significatività del 5%, quale sarà la conclusione ?

Test d'Ipotesi - Esempio 1



Misuriamo la distanza $\overline{x} - \mu = 0.6\,kg$ della media campionaria dalla stima di μ contenuta nell'ipotesi zero in termini di deviazione standard.

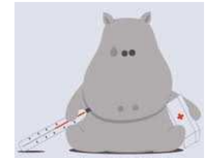
Poiché $$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} \simeq \frac{s}{\sqrt{n}} = \frac{7}{25},$$

il numero u di deviazioni standard di cui $\overline{x}$ si allontana da μ soddisfa la relazione $0.6 = \frac{7}{25}\,u$ da cui segue $u = 2.14\,.$

Nell'intervallo $[\,\mu - 1.96\,\frac{s}{\sqrt{n}}\,,\,\mu + 1.96\,\frac{s}{\sqrt{n}}]$ cade il 95% delle medie campionarie.

Poiché la media campionaria analizzata cade fuori da questo intervallo, siamo autorizzati a rigettare l'ipotesi su μ, assumendoci un rischio del 5%.

Test d'Ipotesi - Esempio 1



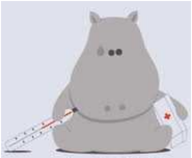
Con livello di significatività del 1%, quale sarà la conclusione ?

Se siamo siamo disposti ad assumerci solo l'1% di rischio di prendere la decisione sbagliata, l'esito del test è diverso.

Nell'intervallo $[\mu - 2.58\frac{s}{\sqrt{n}}, \mu + 2.58\frac{s}{\sqrt{n}}]$ cade il 99% delle medie campionarie.

Poiché la media campionaria $\overline{x}$ appartiene a questo intervallo, l'ipotesi su μ è compatibile con il risultato ottenuto dal campione casuale, con un livello di significatività è dell'1%

Test d'Ipotesi - Esempio 2

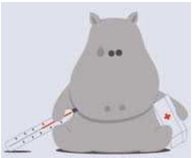


Esempio 2 - Una compagnia aerea afferma che il peso medio del bagaglio dei passeggeri dei suoi voli di linea è $\mu = 19.8\,kg$.

Per sottoporre a verifica tale ipotesi si considera un campione casuale di 324 passeggeri da cui emerge il peso medio campionario $\overline{x} = 20.3\,kg$, con scarto quadratico medio campionario $s = 3.6\,kg$.

Con livello di significatività del 1%, quale sarà la conclusione ?

Test d'Ipotesi - Esempio 2



Si ha $\quad \dfrac{s}{\sqrt{n}} = \dfrac{3.6}{18} = 0.2$

scrivendo $\quad \overline{x} - \mu = 0.5 = (0.2)\, u \quad$ segue che $u = 2.5\,.$

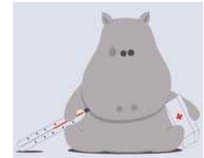
La media campionaria dista, da quella contenuta nell'ipotesi zero, 2.5 scarti quadratici medi e dunque è ancora interna all'intervallo $(\,2.50\,<\,2.58\,)$

$$[\,\mu - 2.58\frac{s}{\sqrt{n}}\,,\,\mu + 2.58\frac{s}{\sqrt{n}}\,]$$

entro cui cade il 99% dei dati.

L'esito del test è: <u>non esistono elementi sufficienti per rigettare l'ipotesi</u> su μ, al livello di significatività richiesto.

Test d'Ipotesi - Esempio 2



Se i valori di $\overline{x}$ ed s sono ottenuti da un campione di 400 passeggeri, come cambiano le conclusioni del test ?

Se il campione è formato da 400 passeggeri: $$\frac{s}{\sqrt{n}} = \frac{3.6}{20} = 0.18$$

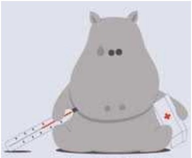
che corrisponde ad una distanza tra $\overline{x}$ e μ pari a 2.78 scarti quadratici medi. Dunque $\overline{x}$ cade fuori dall'intervallo ($2.78 > 2.58$)

$$[\,\mu - 2.58\frac{s}{\sqrt{n}}\,,\,\mu + 2.58\frac{s}{\sqrt{n}}\,]$$

La conclusione del test è: rigetto l'ipotesi formulata su μ, con un margine di rischio dell'1%.

L'esempio mostra il ruolo giocato dalla grandezza del campione nell'esito del test.

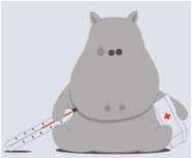
Test d'Ipotesi - Errori



- Se l'ipotesi zero è vera, ma erroneamente viene rigettata, si commette un errore di primo tipo
- Se l'ipotesi zero è falsa , ma erroneamente non viene rigettata, si commette un errore di secondo tipo

NOI CI OCCUPEREMO DI "RENDERE PICCOLO" l' ERRORE DI PRIMO TIPO

Test d'Ipotesi - Esempio 3



Esempio 3 - Supponiamo ora che l'ipotesi zero da vagliare sia formulata in questi termini:

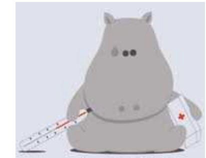
Il reddito medio μ di una famiglia che abita in una certa regione *non supera* i 12500 Euro ($\mu \leq 12500$).

Sottoporre a verifica tale ipotesi.

L'ipotesi sulla media è più sfumata rispetto ai casi precedenti in quanto si limita ad imporre un limite superiore alla media, non un valore preciso.

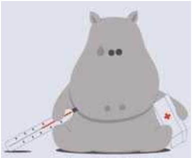
Si può adottare una procedura simile alla precedente.

Test d'Ipotesi - Esempio 3



- si considera la media ottenuta da un campione sufficientemente ampio e la si confronta con il limite superiore per μ di 12500 Euro.
- se $\overline{x}$ risulta inferiore a 12500 Euro non si ha alcun elemento per rigettare l'ipotesi
- i dubbi sulla bontà dell'ipotesi cominciano ad affiorare se $\overline{x}$ supera i 12500 Euro

Test d'Ipotesi - Esempio 3



Come si procede se $\overline{x}$ supera i 12500 Euro ?

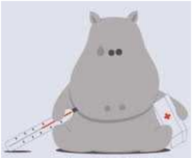
- Se si vuole un livello di significatività del 5%, si avranno elementi per rigettare l'ipotesi quando $\overline{x}$ si lascia a *sinistra* almeno il 95% dei dati. Questo accade non appena si esce dall'intervallo $(-\infty\,,\,\mu + 1.64\sigma_{\overline{x}}]$.
- Se il test sivuole un livello di significatività dell'1%, l'ipotesi diventa rigettabile (con un margine di rischio dell'1%) non appena $\overline{x}$ esce dall'intervallo $(-\infty, \mu + 2.33\sigma_{\overline{x}}]$ entro il quale cade il 99% dei dati.

Curva Gaussiana - 2

aree sottese dalla curva gaussiana sull'intervallo [μ , μ + z σ]

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,10	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,20	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,30	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,40	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,50	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,60	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,70	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,80	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,90	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,00	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,10	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,20	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,30	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,40	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,50	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,60	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,70	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,80	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,90	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	**0,4750**	0,4756	0,4761	0,4767
2,00	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,10	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,20	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,30	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,40	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,50	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	**0,4951**	0,4952
2,60	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,70	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,80	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,90	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,00	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Curva Gaussiana - 1



valori di u	Nell'intervallo $[\mu - u\sigma, \mu + u\sigma]$	Fuori dell'intervallo $[\mu - u\sigma, \mu + u\sigma]$	Nell'intervallo $[\mu + u\sigma, +\infty)$
0	0	1	$0,5$
$0,2$	$0,1586$	$0,8414$	$0,4207$
$0,4$	$0,3108$	$0,6892$	$0,3446$
$0,6$	$0,4514$	$0,5486$	$0,2743$
$0,8$	$0,5762$	$0,4238$	$0,2119$
1	$0,6826$	$0,3174$	$0,1587$
$1,2$	$0,7698$	$0,2302$	$0,1151$
$1,4$	$0,8384$	$0,1616$	$0,0808$
$1,6$	$0,8904$	$0,1096$	$0,0548$
$1,8$	$0,9282$	$0,0718$	$0,0359$
2	$0,9544$	$0,0456$	$0,0228$
$2,2$	$0,9722$	$0,0278$	$0,0139$
$2,4$	$0,9836$	$0,0164$	$0,0082$
$2,6$	$0,9906$	$0,0094$	$0,0047$
$2,8$	$0,9950$	$0,0050$	$0,0025$
3	$0,9974$	$0,0026$	$0,0013$
$3,2$	$0,9986$	$0,0014$	$0,0007$