Esercizi di Statistica per Biotecnologie

Francesca Pizzorni Ferrarese

Esercitazione V – Inferenza II

Es 1

Una ditta produttrice di lampadine sostiene che il valore atteso della durata delle lampadine prodotte è di 1600 ore, con uno scarto quadratico medio pari a 120 ore.
Estraendo un campione di 100 lampadine si è calcolata una durata media di 1570 ore.
Stabilire se l'affermazione del produttore è corretta, usando come ipotesi alternativa che la durata media sia
a) inferiore a quella dichiarata;
b) diversa da quella dichiarata.
Usare in entrambi i casi il livello di significatività $\alpha = 0.05$ e il livello di significatività $\alpha = 0.01$.

Es 2

Il contenuto dichiarato delle bottiglie di sciroppo è 330ml.
Scegliendo un campione di 30 bottiglie, si riscontra un contenuto medio di 328 ml, con uno scarto quadratico medio pari a 3.2 ml.
In base a questi dati si può ritenere che la ditta produttrice inganni il consumatore? Si scelga il livello di significatività dell'1%.

Es 3

Cinque monete sono state lanciate 1000 volte, e a ciascun lancio è stato osservato il numero di teste; nella tabella è riportato il numero di lanci nei quali sono state ottenute 0, 1, …, 5 teste.

N° teste	0	1	2	3	4	5
Frequenza osservata	38	144	342	287	164	25

Stabilire se almeno una moneta si può ritenere truccata.

Es 4

Per stabilire l'efficacia di un vaccino anti-influenzale è stata condotta una ricerca, somministrando il vaccino a 500 persone e controllando il loro stato di salute in un anno; lo stesso controllo è stato fatto per un gruppo di altre 500 persone non vaccinate; in base ai risultati dell'esperimento si è ottenuta la seguente tabella

Frequenze osservate				
	nessuna influenza	*una influenza*	*più di una influenza*	*Totale*
vaccinati	252	145	103	500
non vaccinati	224	136	140	500
Totale	476	281	243	1000

Si può ritenere che il vaccino sia efficace, ossia sottoponendosi alla vaccinazione si ha un minor rischio di contrarre la malattia, oppure il vaccino non è efficace?

Es 5

Dall'esame del colore dei capelli dei bambini di una certa regione, si sono ricavati i seguenti dati

Frequenze osservate						
	biondo	*rosso*	*castano*	*bruno*	*nero*	*Totale*
maschi	592	119	849	504	36	2100
femmine	544	97	677	451	14	1783
Totale	1136	216	1526	955	50	3883

Il colore dei capelli è indipendente dal sesso?

Es 6

Si vuole determinare se una slot machine è ancora in tolleranza.
Determinare quante prove debbono essere effettuate sapendo che la slot machine dovrebbe reagire con le seguenti probabilità:

perdere 70%
vincere una volta la posta 15 %
vincere 3 volte la posta 10 %
vincere 10 volte la posta 3 %
vincere 20 volte la posta 1.5 %
vincere 50 volte la posta 0.5 %

ESERCITAZIONE 8

① valore atteso 1600 ore

s.q.m. 120 ore

n = 100 campione $\bar{x}$ = 1570 ore

Faccio un'ipotesi su una proprietà della distribuzione teorica della popolazione e vedo se le osservazioni consentono di accettarla.

Un test di ipotesi cerca di verificare se un'affermazione è vera o falsa
↓
ipotesi nulla.

Se il test dà esito "negativo" si usa dire che l'ipotesi nulla viene rifiutata e si accetta l'ipotesi alternativa.

- si suppone H_0 vera
- si calcola la distribuzione di uno stimatore
  - corretto per il parametro θ descritto in H_0
  - calcolato da un campione di dimensione N
- si fissa un livello di significatività α
- si trova una regione di accettazione (A)
- si stima puntualmente il parametro θ dal campione c.
  - se il valore è interno ad A accetto l'ipotesi H_0
  - se il valore è esterno ad A rifiuto l'ipotesi H_0.

b) $H_0: E[P] = 1600$ $H_1: E[P] \neq 1600$

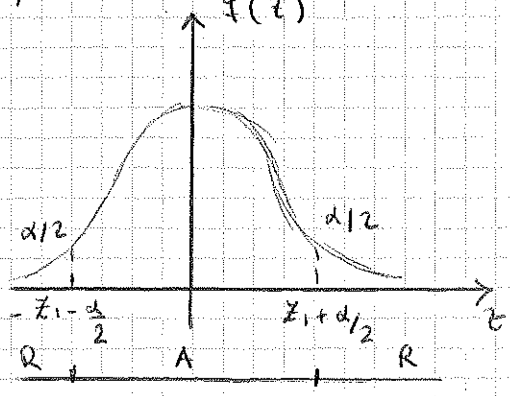
H_0 + test: $E[P] = 1600$ $Var[P] = 14400$

$n = 100 \rightarrow \bar{X} \sim N\left(E[P]; \frac{Var[P]}{n}\right) = N(1600; 144)$

$H_1 \rightarrow$ test a 2 code $\alpha = 0{,}05 \rightarrow$ valori critici $-1{,}96; 1{,}96$

$\Rightarrow A = [-1{,}96; 1{,}96]$

Standardizzo $\bar{x}$ $z_{\bar{x}} = \dfrac{\bar{x} - 1600}{\frac{120}{10}} = \dfrac{1570 - 1600}{12} = -2{,}5$



Il valore $-2{,}5$ cade al di fuori dell'intervallo avente come estremi i valori critici, cioè appartiene alla regione di rifiuto.

Perciò si rifiuta l'ipotesi nulla al livello di significatività $\alpha = 0,05$

$\alpha = 0,01$ I valori critici sono $-2,576$; $2,576 \to A = [-2,576; 2,576]$

Il valore $-2,5$ ricade fra questi estremi $\Rightarrow$ non si rifiuta l'ipotesi per $\alpha = 0,01$.

a) $H_0 : E[P] = 1600$ $H_1 : E[P] < 1600$

H_0 + Testo $\to$ $E[P] = 1600$ $Var[P] = 16400$

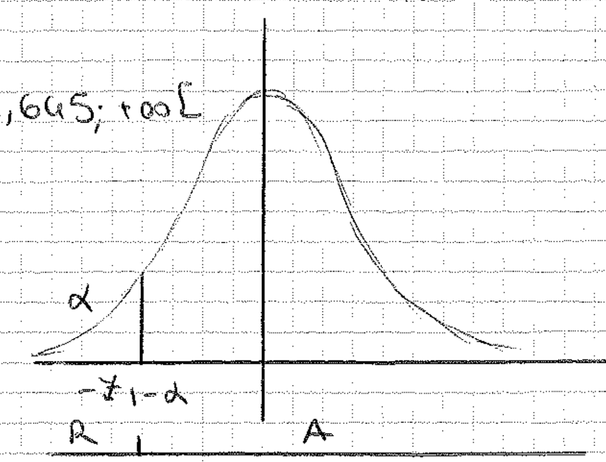
$n = 100$ $\bar{X} \sim N\left(E[P]; \frac{Var[P]}{n}\right) = N\left(1600; \frac{16400}{100}\right)$

H_1 Test unilaterale ad una coda

$\to A = [-1,645; +\infty[$

$\alpha = 0,05 \to$ valore critico $-1,645$

La regola di decisione consiste nel scartare l'ipotesi se il valore della (rifiutare) statistica z ottenuto dai dati del campione è $< -1,645$



il valore era $-2,5 < -1,645 \to$ rifiuto l'ipotesi nulla con livello di significatività $\alpha = 0,05$

$\to A = [-2,326; +\infty[$

con $\alpha = 0,01 \to$ valore critico $-2,326$ anche in questo caso rifiuto.

(2) In questo caso la varianza non è nota $\Rightarrow$ si stima con la varianza campionaria s^2

$H_0 : E[P] = 330$ $H_1 : E[P] < 330$

$n = 30 \to \bar{X} \sim N\left(E[P]; \frac{s^2}{n}\right) = N\left(330; \frac{3,2^2}{30}\right) =$

$= N(330; 0,341)$

$H_1 \to$ test ad 1 coda, unilaterale sx.

standardizzo $\bar{x}$ $z_{\bar{x}} = \frac{\bar{x} - 330}{\frac{3,2}{\sqrt{30}}} = \frac{328-330}{0,584} = -3,42$

$\alpha = 0,01 \to$ valore critico $-z_{0,01} = -2,33 \to A = [-2,33; +\infty[$

$z_{\bar{x}} \notin A \to$ rifiuto H_0 : c'è una significativa evoluzione, che risulta una flessione.

③ Spesso è utile fare ipotesi sulla distribuzione di frequenza di una v.c.
Nel nostro caso l'esperimento può essere descritto mediante n realizzazioni i.i.d di una v.c.

Ho: tutte le monete sono oneste da cui la una distribuzione di riferimento $\rightarrow Bin(5, 1/2)$ $\quad p = \frac{1}{2} \quad n = 5$

H_1: almeno una moneta è truccata.

Definisco le frequenze teoriche: frequenza attesa se Ho è vera $\hat{n}_i = n p_i$
contingenza: scarto fra frequenza rilevata e teorica
$c_i = n_i - \hat{n}$

Se Ho è vera è verosimile che tutte le contingenze (in valore assoluto) siano piccole.
Se H_1 è vera è verosimile che almeno una contingenza (in valore assoluto) sia elevata.
La contingenza è la base di un indicatore per quantificare l'aderenza di una distribuzione teorica ad una reale (Ho)
Indicatore di Pizzetti - Pearson: $\sum_{i=1}^{M} \frac{c_i^2}{\hat{n}_i} = \sum_{i=1}^{M} \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}$

Le quadrato elimina il − e pesa molto i valori alti
Il rapporto serve per scalare contemporaneamente i contributi

i ← n° di teste	n_i	$\hat{n}_i$	$(n_i - \hat{n}_i)$	$(n_i - \hat{n}_i)^2$	$(n_i - \hat{n}_i)^2 / \hat{n}_i$
0	38	31,25	6,75	45,5625	1,46
1	144	156,25	−12,25	150,0625	0,96
2	342	312,5	29,5	870,25	2,78
3	287	312,5	−25,5	650,25	2,08
4	164	156,25	7,75	60,0625	0,38
5	25	31,25	−6,25	39,0625	1,25

→ Binomiale $P(X=0) = \binom{n}{k} p^k q^{n-k} = \binom{5}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = 0,03125$

$P(X=1) = \binom{5}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^4 = \frac{5!}{1!\,4!} \cdot \frac{1}{2} \cdot \frac{1}{16} = 0,15625$

$P(X=2) = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{5!}{2!\,3!} \cdot \frac{1}{4} \cdot \frac{1}{8} = 0,3125$

$P(X=3) = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = \frac{5!}{3!\,2!} \cdot \frac{1}{8} \cdot \frac{1}{4} = 0{,}3125$

$P(X=4) = \binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right) = \frac{5!}{4!\,1!} \cdot \frac{1}{16} \cdot \frac{1}{2} = 0{,}15625$

$P(X=5) = \binom{5}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 = \frac{5!}{5!\,0!} \cdot \frac{1}{32} \cdot 1 = 0{,}03125$

Moltiplichiamo per 1000 questi risultati.

$\longrightarrow \sum_{i=0}^{5} \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i} = 8{,}91$

La strategia di test riguarda la d.d.p. dello stimatore

Teorema: si dimostra che al ↑ della dimensione del campione (n) allora si ha che

$$\sum_{i=1}^{M} \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i} \sim \chi^2(\nu) \quad \leftarrow \text{T.L.C.}$$

↑ parametri liberi (g.d.l.) di P $(M-1)$

Il valore della statistica chi-quadro calcolato dal campione è

$\chi^2 = 8{,}91 \leftarrow$ stimatore

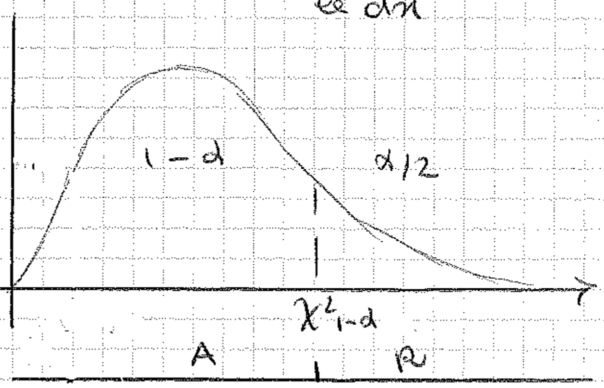
$A = [0; \chi^2_{1-\alpha}(M-1)]$

gradi di libertà $\rightarrow M-1 = 5$

Test ad una coda, unilatera a dx

Il valore critico a livello di significatività $\alpha = 0{,}05$

$\chi^2_{\underbrace{0,95}_{1-\alpha}}(5) = 11{,}1 \ldots$



$\Rightarrow A = [\,0\;;\;11{,}1\,]$

$8{,}91 \in A \Rightarrow$ accetto $H_0 \rightarrow$ c'è buon adattamento dei dati alla distribuzione binomiale.

(4) Ci vogliamo chiedere se le 2 variabili (qualitative) "vaccinazione" e "minor n° di influenze" sono indipendenti oppure no.

Date le osservazioni di una bivariata, la v.c. P ottiene:

- $M = M_x M_y$ modalità
- d.d.p. $\hat{p}_{i,j} = \frac{n_{i,+}}{n} \cdot \frac{n_{+,j}}{n} \quad i = 1,2 \ldots M_x,\ j = 1,2 \ldots M_y$

che sarebbe la bivariata solo se vi è indipendenza

↑ $P(X = x_i \cap Y = y_j) = P(X = x_i)P(Y = y_j)$

L'indipendenza viene testata con un test di aderenza teorica
alla distribuzione

⇒ ① calcolare le frequenze teoriche

$$\hat{n}_{i,j} = n\,\hat{p}_{i,j} \quad \forall i,j \quad = \frac{n_{i,+}\, n_{+,j}}{n}$$

↓

		nessuna infrazione	una infrazione	più di una infrazione
X	bocciati	238	140,5	121,5
	non bocciati	238	140,5	121,5

Convergenza verificata $\hat{n}_{i,j} > 5 \ \forall i,j$

$$\Rightarrow \sum_{i=1}^{M_x}\sum_{j=1}^{M_y} \frac{(n_{i,j}-\hat{n}_{i,j})^2}{\hat{n}_{i,j}} \sim \chi^2((M_x-1)(M_y-1))$$

$$\sim \chi^2((2-1)(3-1)) = \chi^2(2)$$

$\chi^2_{0,99}(2)$

$\alpha = 0,01 \longrightarrow A = [0 \ ; 9,21] \longleftarrow A = [0 ; \chi^2_{1-\alpha}((M_x-1)(M_y-1))]$

$$\sum_{i=1}^{M_x}\sum_{j=1}^{M_y} \frac{(n_{i,j}-\hat{n}_{i,j})^2}{\hat{n}_{i,j}} = 7,57$$ → lo stimatore è interno ad A

(*)

⇒ accetto H_0: le variabili sono non legate.

⑥

i	p_i	$\hat{n}_i$ ← deve essere > 5 $\forall i$
1	0,7	
2	0,15	
3	0,10	
4	0,03	
5	0,015	
6	0,005 ← 1000 dadi fare uno ×1000 prove.	

↑ Molti autori ritengono che si abbia una buona convergenza della legge quando tutte le frequenze teoriche sono > 5.

(*)
$$= \frac{(252-238)^2}{238} + \frac{(145-140,5)^2}{140,5} + \frac{(103-121,5)^2}{121,5} +$$
$$+ \frac{(224-238)^2}{238} + \frac{(136-140,5)^2}{140,5} + \frac{(140-121,5)^2}{121,5} = 7,57$$

⑤ Costruiamo la tabella delle frequenze teoriche.

		Y			
	bi.	r	c	br.	m
Maschi	614,37	116,82	825,29	516,48	27,04
X Femmine	521,63	99,18	700,71	438,52	22,96

$$\sum_{i=1}^{H_x} \sum_{j=1}^{H_y} \frac{(n_{i,j} - \hat{n}_{i,j})^2}{\hat{n}_{i,j}} = 10,47$$

$$\chi^2_{1-\alpha}((H_x-1)(H_y-1)) = \qquad \alpha = 0,01$$

$$= \chi^2_{0,99}(1)(4) = \chi^2_{0,99}(4) = 13,1$$

$$\Rightarrow A = [0 ; 13,1] \qquad 10,47 \in A \rightarrow H_0$$

$$A = [0 ; 9,49] \qquad \alpha = 0,05$$

$$10,47 \notin A \rightarrow H_1$$

I risultati relativi ai 2 livelli di significatività non sono in accordo → necessità di indagini più approfondite.