Statistica Matematica A

Docente: Dott. F. Zucca

II prova in itinere

Lecco 31/1/2005

Tempo a disposizione: 2h30

<u>Candidato</u>

Cognome: .... **Nome:** ....

Corso di Laurea: ....

Matricola: .... **Anno di corso:** ....

1 Domande a risposta multipla

Ogni domanda ha una sola risposta esatta. Le risposte sbagliate sono contate negativamente per un terzo del valore della risposta esatta.

1. **Che cos'è il p-value?**

☐ Il più grande valore di significatività al quale si rifiuta l'ipotesi nulla.
☒ L'estremo inferiore dei livelli di significatività ai quali si rifiuta l'ipotesi nulla.
☐ Il più basso valore del livello di significatività che rende il test applicabile.
☐ La probabilità di commettere un errore di prima specie.

2. **Quale dei seguenti stimatori è corretto per il parametro λ di una legge di Poisson?**

☐ $n^{-1}\sum_{i=1}^n X_i^2$
☐ $(n+1)^{-1}\sum_{i=1}^n X_i$
☐ $\sum_{i=1}^n X_i$
☒ $(n-1)^{-1}\sum_{i=1}^n (X_i - n^{-1}\sum_{j=1}^n X_j)^2$

3. **Stiamo cercando un'intervallo di confidenza bilatero per la media a varianza nota; dovendo ridurre l'ampiezza dell'intervallo cosa possiamo fare?**

☐ Aumentare il livello di confidenza.
☒ Aumentare l'ampiezza del campione.
☐ Diminuire l'ampiezza del campione.
☐ Aumentare la varianza

4. **Siano $X_1, \cdots, X_n$ variabili indipendenti di legge esponenziale di parametro $1/3$. Tenendo conto del Teorema del limite centrale come possiamo meglio approssimare $\sum_{i=1}^{n} X_i$?**

☐ Con una legge di Poisson di parametro $3n$.
☐ Con una legge normale $\mathcal{N}(0, 1)$.
☒ Con una legge normale $\mathcal{N}(3n, 9n)$.
☐ Con una legge Binomiale $B(3n, 1/3n)$.

2 Esercizi

Esercizio 1 Robinson Crusoe è stato estromesso dalla propria capanna da Venerdì. Per ultimare la sua nuova capanna a che gli ci vorrebbero ancora 50 giorni di lavoro al ritmo di 8 ore di lavoro al giorno. Purtroppo mancano 55 giorni all'inizio della stagione delle piogge e negli ultimi 20 giorni egli ha lavorato complessivamente solo 135 ore. Si supponga che le ore di lavoro giornaliere siano casuali con media μ e varianza σ^2 ignote e siano indipendenti in giorni differenti.

1. Supponendo che la varianza campionaria delle ore giornaliere dei lavoro (calcolata dal campione dei 20 giorni appena trascorsi) sia $s_{20}^2 = 2$ calcolare un'intervallo di confidenza al 95% per il valore atteso delle ore lavorative giornaliere.

2. Robinson si chiede se sia plausibile l'ipotesi $\mu \geq 8$ al 5%.

3. Se il valore atteso μ e la varianza σ^2 coincidessero, rispettivamente con la media campionaria e la varianza campionaria, qual è la probabilità che Robinson termini la capanna prima dell'arrivo della stagione delle piogge? Qual è il numero minimo di giorni necessari affinché la capanna sia ultimata con una probabilità non inferiore al 99%?

Soluzione.

1. Sia X_i la variabile che descrive il numero di ore lavorative il giorno i-esimo; supponiamo che ammetta media finita e varianza finita (quasi certamente non nulla visto che $s_{20}^2 \neq 0$). Sappiamo che $n = 20$, $s_{20}^2 = 2$ e $\bar{x}_{20} = 135/20 = 6.75$. L'intervallo di confidenza si calcola come

$$IC(\alpha) := \left[\bar{x}_n - t_{(1+\alpha)/2}(n-1)\frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{(1+\alpha)/2}(n-1)\frac{s_n}{\sqrt{n}}\right]$$

essendo $t_{0.975}(19) \approx 2.093$, $s_n = \sqrt{2}$ si ha

$$IC(0.95) = [6.75 - 0.6619, 6.75 + 0.6619] = [6.0881, 7.4119].$$

2. L'ipotesi nulla $H_0 : \mu \geq 8$ (contro $H_1 : \mu < 8$) è accettata al 5% se e solo se 8 appartiene all'intervallo di confidenza $I^s(0.95)$ infinito a sinistra a livello 95%; ovviamente $I^s(0.95) \subset (-\infty, 7.4119]$ pertanto l'ipotesi nulla è rifiutata.

   Con calcoli espliciti, la regione critica è $\bar{x}_n < \mu_0 + t_\alpha(n-1)s_n/\sqrt{n}$. Essendo $t_{0.05}(19) \approx -1.7291$ allora la regione critica diviene $\bar{x}_n < 8 - 1.7291/\sqrt{10} \approx 8 - 0.5468 = 7.4532$; $\bar{x}_n = 6.75$ quindi rifiutiamo l'ipotesi nulla al 5%.

3. Dall'indipendenza delle variabili X_i che soddisfano $\mathbb{E}(X_i) = 6.75$ e $\mathrm{var}(X_i) = 2$, in virtù del Teorema Centrale del Limite si ha

$$\sum_{i=1}^n X_i \approx Y_n \sim \mathcal{N}(6.75n, 2n).$$

   Pertanto

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq 50 \cdot 8\right) = \mathbb{P}\left(\frac{\sum_{i=1}^n X_i - 6.75 \cdot n}{\sqrt{2n}} \geq \frac{400 - 6.75 \cdot n}{\sqrt{2n}}\right)$$
$$\approx 1 - \Phi\left(\frac{400 - 6.75 \cdot n}{\sqrt{2n}}\right).$$

   Nel caso $n = 55$ si ha

$$1 - \Phi\left(\frac{400 - 6.75n}{\sqrt{2n}}\right) \approx 1 - \Phi\left(\frac{400 - 6.75 \cdot 55}{\sqrt{110}}\right) \approx 1 - \Phi\left(2.7412\right) \approx 0.0031.$$

   Inoltre la disequazione in n

$$1 - \Phi\left(\frac{400 - 6.75n}{\sqrt{2n}}\right) \geq \alpha$$

   è equivalente a $\frac{400-6.75n}{\sqrt{2n}} \geq q_{1-\alpha}$ che, con l'usuale sostituzione $t = \sqrt{n}$ e tenendo conto che $q_{0.01} \approx -2.3263$, diviene equivalente a

$$\begin{cases} 6.75t^2 - 3.29t - 400 \geq 0 \\ t \geq 0 \end{cases}$$

   che ammette soluzione $t \geq 7.9456$ che implica $n \geq 7.9456^2 = 63.132$ cioè $n \geq 64$.

Esercizio 2 Il numero giornaliero X di passeggeri del treno “Old Rusty Trail” è riassunto nella seguente tabella che prende in considerazione un campione di 100 giorni.

Passeggeri	$X \leq 200$	$200 < X \leq 220$	$220 < X \leq 240$	$240 < X \leq 260$	$260 < X$
Numero giorni	5	33	50	12	0

Ci si chiede se i dati si accordino con una distribuzione di Poisson al 5%. (Sugg: una volta stimato il parametro della Poisson, calcolare le probabilità teoriche con un'opportuna approssimazione).

Soluzione. Essendo il parametro λ di una Poisson pari al valore atteso, stimiamo λ con $\bar{x}_{100}$ calcolato, in maniera approssimativa come

$$\bar{x}_{100} = \frac{1}{100}(5 \cdot 100 + 33 \cdot 210.5 + 50 \cdot 230.5 + 12 \cdot 250.5) = 219.775.$$

In virtù del Teorema Centrale del Limite $X \approx Y \sim \mathcal{N}(\lambda, \lambda)$, pertanto calcoliamo in maniera approssimata $\mathbb{P}(X < k) \approx \Phi((k + 0.5 - \lambda)/\sqrt{\lambda})$ (dove utilizziamo la correzione di continuità e k è un numero intero). Pertanto

k	200	220	240	260
$\mathbb{P}(X \leq k)$	0.0968	0.5196	0.919	0.9969

che ci permette di completare la tabella nel seguente modo

Passeggeri	$X \leq 200$	$200 < X \leq 220$	$220 < X \leq 240$	$240 < X \leq 260$	$260 < X$
$f_a(i)$	5	33	50	12	0
$f_a^{teor}(i)$	9.68	$51.96 - 9.68 = 42.28$	$91.9 - 51.96 = 39.94$	$99.69 - 91.9 = 7.79$	0.31

La regione di rifiuto è $Q > \chi^2_{1-\alpha}(N-r-1)$ dove $N = 4$ (bisogna accorpare le ultime due classi), $r = 1$, $\chi^2_{0.95}(2) = 5.991465$ e

$$Q := \sum_{i=1}^{4} \frac{f_a^2(i)}{f_a^{teor}(i)} - n = 8.7112$$

pertanto rifiutiamo l'ipotesi nulla. Alternativamente si potrebbe valutare il P-value

$$\bar{\alpha} = 1 - F_{\chi^2(N-r-1)}(Q) \in (0.01, 0.025).$$

Esercizio 3 Al fine di valutare la precisione di una bilancia, un oggetto di peso noto $p = 1$ grammo viene pesato $n = 10$ volte ottenendo le seguenti misure: 1.02, 1.01, 0.98, 0.99, 1.00, 0.99, 0.99, 1.02, 1.03, 1.00.
Si assuma che l'errore ε di misurazione della bilancia si distribuisca come una normale centrata di varianza incognita, ovvero $\varepsilon \sim N(0, \sigma^2)$.

1. Si stimi la varianza del peso $X = p + \varepsilon$ ad un livello di confidenza del 95%.

2. Sottoporre a test l'ipotesi nulla $\sigma^2 = 0.005$ al livello di significatività del 5%.

Soluzione.

1. Dobbiamo calcolare un intervallo di confidenza bilatero per la varianza di una popolazione normale con media nota.

   La statistica di riferimento segue dunque legge chi quadro con n gradi di libertà.

   Gli estremi dell'intervallo a livello α si otterranno al solito nel seguente modo:
$$I_{\sigma^2} = \left[\frac{nT^2}{\chi^2_{(1+\alpha)/2}(n)}, \frac{nT^2}{\chi^2_{(1-\alpha)/2}(n)}\right] \equiv \left[\frac{nT^2}{\chi^2_{0.975}(n)}, \frac{nT^2}{\chi^2_{0.025}(n)}\right]$$
   Poiché
$$nT^2 = \sum_{k=1}^{n}(x_i - \mu)^2 = 0.0025$$
   e
$$\chi^2_{0.975}(10) \simeq 20.4832 \quad , \quad \chi^2_{0.025}(10) \simeq 3.2470$$
   ricaviamo:
$$IC(0.95) \equiv [0.000122, 0.000770]\,.$$

2. Poiché il valore di riferimento $\sigma_0^2 = 0.005$ non é contenuto nell'intervallo bilatero al 95% calcolato nel punto precedente, rifiutiamo l'ipotesi nulla al 5%.
   Alternativamente la regione di rifiuto è
$$(-\infty, \chi^2_{\alpha/2}(n)) \cup (\chi^2_{1-\alpha/2}(n), +\infty).$$
   Essendo $\chi^2 = nT_n/\sigma_0^2 = 0.5 < 3.2470 = \chi^2_{0.025}(10)$ rifiutiamo l'ipotesi nulla al 5%.

Esercizio 4 Un test di ammissione alla facoltà di ingegneria è costituito da 500 domande a risposta multipla ognuna delle quali propone una risposta corretta e due sbagliate. L'esaminando potrà iscriversi solo se risponde esattamente ad almeno 170 domande. Poiché il test può essere svolto anche da casa tramite internet, un giovane burlone fa svolgere il test alla sua scimmia addomesticata la quale è in grado di rispondere a caso con ugual probabilità.

Qual è la probabilità che la scimmia superi il test e venga ammessa alla facoltà di ingegneria?

Soluzione. Sia X_i la v.a. che codifica con 1 la risposta esatta e con 0 quella sbagliata alla domanda i-esima e $X = X_1+...+X_n$ la v.a. che conta il numero di risposte corrette su n domande. Allora $X \sim Bin(n,p)$ dove $n = 500$ e $p = 1/3$.

La probabilità che la scimmia superi il test sarà data da:

$$P[X \geq 170] = 1 - P[X \leq 169] = 1 - 3^{-500} \sum_{k=0}^{169} \binom{500}{k} 2^{500-k}.$$

Il conto è impraticabile quindi procediamo per altra via sfruttando il TLC. Le condizioni di applicabilità ($np > 5$ e $n(1-p) > 5$) sono certamente soddisfatte quindi, ricordando la correzione di continuità:

$$1-P[X \leq 169] \simeq 1-\Phi\left(\frac{169.5 - np}{\sqrt{np(1-p)}}\right) \simeq 1-\Phi\left(0.2687\right) \simeq 1-0.6026 = 0.3974.$$