07.07.2015 -appello di STATISTICA per studenti ENE - docente: E. Piazza

obbligatorio - n. iscrizione sulla lista ☐ **se non ve lo ricordate siete fritti; o no?**

il presente elaborato si compone di 4 (quattro) pagine

Cognome_____________________________ Nome__________________________ matr.n.________

Avvertimento: nello svolgimento degli esercizi se una quantità è indicata con il simbolo x continuare a chiamarla x, e se si chiama W non chiamarla X, e se si chiama t_i non chiamarla x_i, e se si chiama $\bar{T}_n$ non indicarla con $\bar{X}_n$. Se volete scrivere dipendenti non scrivete indipendenti, se dovete sottrarre non sommate. Si fanno troppi errori di distrazione. Mi raccomando: concentrazione.

Si consiglia di lavorare con 3 decimali, arrotondando opportunamente.

Problema 1. *Il Dottor H. vuole conoscere l'angolo θ compreso fra l'ipotenusa e il cateto orizzontale di un fregio sull'isola di Rodi. Ha a disposizione uno strumento per la misura delle lunghezze, in cm, affetto da errore (additivo) casuale $\epsilon \sim N(0\,,\,0.01)$. Ordina quindi che vengano effettuate 13 misure indipendenti $X_1, \dots, X_{13}$ della lunghezza x del cateto orizzontale e 15 misure indipendenti $Y_1, \dots, Y_{15}$ della lunghezza y del cateto verticale. Chiaramente il Dottor H. sa che*

$$\theta = \arctan \frac{y}{x}.$$

1 **1.1** *Esprimere il risultato X_k della k-esima misura in funzione di x e del corrispondente errore ϵ_k. Determinare quindi la distribuzione di X_k.*

$$X_k = x + \epsilon_k \sim N(x\,,\,0.01) \text{ per ogni } k.$$

1 **1.2** *Calcolare la probabilità di ottenere $|X_1 - x| > 0.1$ cm.*

$$\mathbb{P}(|X_1 - x| > 0.1) = 2\left(1 - \Phi\left(\frac{0.1}{0.1}\right)\right) = 0.3173.$$

2 **1.3** *Introdurre stimatori opportuni $\widehat{X}$ e $\widehat{Y}$ di x e y sulla base delle misure ordinate. Ricordando che lo scarto quadratico medio MSE di uno stimatore Q di un parametro θ è dato da $E[(Q-\vartheta)^2]$, specificare distribuzione, distorsione, errore quadratico medio.*

$$\widehat{X} = \overline{X}_{13} \sim N\left(x\,,\,\frac{0.01}{13}\right), \qquad \text{bias}(\widehat{X}) = 0, \qquad \text{MSE}(\widehat{X}) = \frac{0.01}{13},$$

$$\widehat{Y} = \overline{Y}_{15} \sim N\left(y\,,\,\frac{0.01}{15}\right), \qquad \text{bias}(\widehat{Y}) = 0, \qquad \text{MSE}(\widehat{Y}) = \frac{0.01}{15}.$$

1 **1.4** *Calcolare la probabilità di ottenere $|\widehat{X} - x| > 0.1$ cm.*

$$\mathbb{P}(|\widehat{X} - x| > 0.1) = 2\left(1 - \Phi\left(\frac{0.1}{0.1/\sqrt{13}}\right)\right) = 0.0003115.$$

1 **1.5** *Proporre uno stimatore $\widehat{\Theta}$ per θ sulla base delle misure ordinate.*

$$\widehat{\Theta} = \arctan \frac{\overline{Y}_{15}}{\overline{X}_{13}}.$$

2 **1.6** *Determinare, in funzione di x e y, eventualmente in modo approssimato: media, varianza, distorsione ed errore quadratico medio di $\widehat{\Theta}$.*
Applicando il metodo delta, ovvero la formula di propagazione degli errori:

$$\mathbb{E}[\widehat{\Theta}] \simeq \theta, \qquad \text{Var}(\widehat{\Theta}) \simeq 0.01\left(\frac{1}{13}\frac{x^2}{(x^2+y^2)^2} + \frac{1}{15}\frac{y^2}{(x^2+y^2)^2}\right),$$

$$\text{bias}(\widehat{\Theta}) \simeq 0, \qquad \text{MSE}(\widehat{\Theta}) \simeq 0.01\left(\frac{1}{13}\frac{x^2}{(x^2+y^2)^2} + \frac{1}{15}\frac{y^2}{(x^2+y^2)^2}\right).$$

2 **1.7** *Eseguite le misure, si trova*

$$\overline{x}_{13} = 115.00 \text{ cm}, \qquad s_x^2 = 0.0103 \text{ cm}^2, \qquad \overline{y}_{15} = 186.64 \text{ cm}, \qquad s_y^2 = 0.00989 \text{ cm}^2.$$

Fornire una stima puntuale di θ e del corrispondente errore quadratico medio.

$$\widehat{\theta} = \arctan\frac{\overline{y}_{15}}{\overline{x}_{13}} = 1.0185795331 = 58.36°,$$

$$\text{MSE}(\widehat{\Theta}) \simeq 0.01\left(\frac{1}{13}\frac{115^2}{(115^2+186.64^2)^2} + \frac{1}{15}\frac{186.64^2}{(115^2+186.64^2)^2}\right) = 1,4\cdot 10^{-8}.$$

Problema 2. *Supponendo di non conoscere alcuna cifra decimale di π, il matematico Mario Lazzarini vuole inferire statisticamente sul suo valore lanciando n aghi di lunghezza ℓ su un pavimento a strisce distanti $\frac{6\ell}{5}$, cosicché la probabilità che un singolo ago intersechi una qualche linea del pavimento valga*

$$p = \frac{5}{3\pi}.$$

Sia X_k la variabile che indica se il k-esimo ago interseca una linea del pavimento.

2 **2.1** *Che variabili sono le variabili X_k e come sono distribuite?*

$$X_k \sim B(p); \forall k = 1, \dots, n.$$

va di Bernoulli di parametro p.

In particolare il Lazzarini vuole confutare statisticamente la tesi, formulata da un progetto di legge del Parlamento dell'Indiana, secondo cui

$$\pi = 3.2.$$

2 **2.2** *Introdurre un opportuno test statistico utile allo scopo del Lazzarini, specificando: ipotesi nulla, ipotesi alternativa, regione critica di livello α, condizioni di applicabilità.*

Posto $\pi_0 = 3.2$ la sua corrispondente probabilità $p_0 = \frac{5}{3\cdot 3.2} = 0.520833$, e detta $\overline{X}_n$ la media campionaria delle osservazioni, poniamo:

$$H_0 : p = p_0 \qquad \text{vs} \qquad H_1 : p \neq p_0$$

$$R_C : |\overline{X}_n - p_0| > \sqrt{\frac{p_0(1-p_0)}{n}}\, z_{\alpha/2} = \sqrt{\frac{p_0(1-p_0)}{n}}\,\Phi^{-1}(1-\frac{\alpha}{2})$$

purché $\quad n \cdot p_0 \geq 5, \quad n\,(1-p_0) \geq 5.$

Il Lazzarini lancia quindi i suoi 3˙408 aghi, contando 1˙808 intersezioni.

2 **2.4** *Stimare puntualmente p e π.*

$$\widehat{p} = \frac{1˙808}{3˙408} = 0.530516 \implies \widehat{\pi} = \frac{5}{3p_0} = \frac{5}{3\cdot 0.530516} = 3.141592920.$$

2 **2.5** *Calcolare il p-value dei dati per il test introdotto.*

$$\frac{|\overline{x}_n - p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} = 1.13$$

$$\text{p-value} = 2(1-\Phi(1.13)) = 2(1-0.8708) = 0.2584$$

2 **2.6** *Trarre le debite conclusioni al livello scelto $\alpha = 0.01$.*

I dati raccolti non consentono di rifiutare all'1% l'ipotesi nulla: potrebbe essere $\pi = 3.2$.

2.7 *Se la conclusione tratta fosse errata, l'errore commesso sarebbe del primo o del secondo tipo?*

Se la conclusione tratta fosse errata, l'errore commesso sarebbe del secondo tipo.

3 **2.3** *Ricordando che*

$$\text{potenza} = P[\text{rifiuto } H_0 | H_0 \text{ è falsa}]$$

calcolare la potenza del test introdotto nel caso di $n = 3'408$ aghi, di un livello $\alpha = 0.01$, sapendo che il valore reale di pi-greco è $\pi = 3.141592920\ldots$, cioè il p vero è $p_1 = 0.530516$.
Per rifiutare H_0 la statistica test deve essere nella regione critica, cioè

$$\frac{|\overline{X}_n - p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_{\alpha/2}$$

La probabilità che ciò accada deve essere calcolata sotto H_0 falsa cioè $p = p_1 = 0.530516$ e non $p = p_0$. In questo caso
$\overline{X}_n \approx N(0.530516; \dfrac{p_1 \cdot (1-p_1)}{3408} = N(0.530516; 7.308\,4 \times 10^{-5})$ e $z_{\alpha/2} = z_{0.005} = \Phi^{-1}(0.995) = 2.576$.
La potenza richiesta vale

$$P[|\overline{X}_n - p_0| > z_{\alpha/2}\sqrt{\tfrac{p_0(1-p_0)}{n}}|p=p_1] = P[\{\overline{X}_n - p_0 < -\sqrt{\tfrac{p_0(1-p_0)}{n}}z_{\alpha/2}\} \cup \{\overline{X}_n - p_0 > \sqrt{\tfrac{p_0(1-p_0)}{n}}z_{\alpha/2}\}|p=p_1] =$$
$$P[\overline{X}_n < p_0 - \sqrt{\tfrac{p_0(1-p_0)}{n}}z_{\alpha/2}|p=p_1] + P[\overline{X}_n > p_0 + \sqrt{\tfrac{p_0(1-p_0)}{n}}z_{\alpha/2}|p=p_1] = \text{(standardizzando)} =$$
$$P[\frac{\overline{X}_n - p_1}{\sqrt{\dfrac{p_1(1-p_1)}{n}}} < \frac{p_0 - \sqrt{\frac{p_0(1-p_0)}{n}}z_{\alpha/2} - p_1}{\sqrt{\dfrac{p_1(1-p_1)}{n}}}] + P[\frac{\overline{X}_n - p_1}{\sqrt{\dfrac{p_1(1-p_1)}{n}}} > \frac{p_0 + \sqrt{\frac{p_0(1-p_0)}{n}}z_{\alpha/2} - p_1}{\sqrt{\dfrac{p_1(1-p_1)}{n}}}] =$$
$$= \Phi\left(\frac{p_0 - \sqrt{\frac{p_0(1-p_0)}{n}}z_{\alpha/2} - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right) + 1 - \Phi\left(\frac{p_0 + \sqrt{\frac{p_0(1-p_0)}{n}}z_{\alpha/2} - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right) =$$
$p_0 = 0.520833,\ p_1 = 0.530516$
$$= \Phi\left(\frac{0.520833 - \sqrt{\frac{0.520833(1-0.520833)}{3408}}2.576 - 0.530516}{\sqrt{\frac{0.530516(1-0.530516)}{3408}}}\right) + 1 - \Phi\left(\frac{0.520833 + \sqrt{\frac{0.520833(1-0.520833)}{3408}}2.576 - 0.530516}{\sqrt{\frac{0.530516(1-0.530516)}{3408}}}\right) =$$
$$= \Phi\left(\frac{-3.1727 \times 10^{-2}}{8.5489 \times 10^{-3}}\right) + 1 - \Phi\left(\frac{1.2361 \times 10^{-2}}{8.5489 \times 10^{-3}}\right)$$
$$= \Phi(-3.7112) + 1 - \Phi(1\,4459) \simeq 0 + 1 - 0.9265 = 0.0735$$

Problema 3. *L'osservazione congiunta di tre va X, Y e Z ha prodotto il campione di dimensione $n = 50$ i cui dati di sintesi sono riportati nella tabella seguente*

$\sum x_i$	$\sum y_i$	$\sum z_i$	$\sum x_i^2$	$\sum y_i^2$	$\sum z_i^2$	$\sum x_i y_i$	$\sum x_i z_i$	$\sum y_i z_i$
0	0	0	10	4	5	5	-5	-2.5

(1)

Nel prosieguo può velocizzare i conti sapere che $\begin{bmatrix} 50 & 0 & 0 \\ 0 & 10 & -5 \\ 0 & -5 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 0.02 & 0 & 0 \\ 0 & 0.2 & 0.2 \\ 0 & 0.2 & 0.4 \end{bmatrix}$

2 **3.1** *Si ipotizzi il legame lineare $Y_i = b_0 + b_1 x_i + \varepsilon_i$ e, sfruttando i valori appena ricavati e riportati in tabella (1), si effettui la stima puntuale dei coefficienti $\hat{b}_0$ e $\hat{b}_1$ della regressione lineare.*

$$\hat{b}_1 = \frac{s_{xy}}{s_{xx}} = \frac{\sum_{i=1}^{50}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{50}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{50} x_i y_i}{\sum_{i=1}^{50} x_i^2} = \frac{5}{10} = 0.5 \quad ; \quad \hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = 0$$

2 **3.2** *Ricordando l'espressione dell'intervallo di confidenza bilatero di livello γ per $\hat{b}_1$ cioè:*

$$\hat{b}_1 - t_{n-2}(\frac{1+\gamma}{2})\sqrt{\frac{SS_E}{(n-2)S_{XX}}} < b_1 < \hat{b}_1 + t_{n-2}(\frac{1+\gamma}{2})\sqrt{\frac{SS_E}{(n-2)S_{XX}}}$$

dove $SS_E = S_{YY} - \hat{b}_1^2 S_{XX}$ calcolare un intervallo di confidenza al livello 90% per il coefficiente b_1 della regressione (può essere utile ricordare che i quantili della t con 48 gradi di libertà possono essere ricavati dalla tavola della $N(0;1)$.

Per l'osservazione fatta abbiamo $t_{n-2}(\dfrac{1+\gamma}{2}) = \Phi^{-1}(\dfrac{1+\gamma}{2}) = \Phi^{-1}(0.95) = 1.645$. Inoltre:
$\bar{y} = 0$ quindi $S_{YY} = \sum_{i=1}^{50}(y_i - \bar{y})^2 = \sum_{i=1}^{50} y_i^2 = 4$ quindi $SS_E = 4 - (0.5)^2 \cdot 10 = 1.5$ e

$$\sqrt{\frac{SS_E}{(n-2)S_{XX}}} = \sqrt{\frac{1.5}{48 \cdot 10}} = 5.59 \times 10^{-2},$$

$$t_{n-2}\left(\frac{1+\gamma}{2}\right)\sqrt{\frac{SS_E}{(n-2)S_{XX}}} = 1.645 \cdot 5.59 \times 10^{-2} = 9.1956 \times 10^{-2},$$

perciò l'intervallo di confidenza cercato è l'intervallo della retta che ha per estremi 0.5 ± 0.092 cioè $(0.408, 0.592)$.

3 **3.3** *Ora si ipotizzi il legame multilineare $Y_i = b_0 + b_1 x_i + b_2 z_i + \varepsilon_i$ e, sempre sfruttando i valori riportati in tabella, si effettui la stima puntuale dei coefficienti $\hat{b}_0$, $\hat{b}_1$ e $\hat{b}_2$ della regressione multilineare.*

Si ponga al solito $X = \begin{bmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ \vdots & \vdots & \vdots \\ 1 & x_{50} & z_{50} \end{bmatrix}$, $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{50} \end{bmatrix}$ *e* $b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$ *e si usi la notazione matriciale.*

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} n & \sum_i x_i & \sum_i z_i \\ \sum_i x_i & \sum_i x_i^2 & \sum_i x_i z_i \\ \sum_i z_i & \sum_i x_i z_i & \sum_i z_i^2 \end{pmatrix} = \begin{pmatrix} 50 & 0 & 0 \\ 0 & 10 & -5 \\ 0 & -5 & 5 \end{pmatrix}$$

$$\mathbf{X}^T\mathbf{Y} = \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \\ \sum_i z_i y_i \end{pmatrix}$$

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y}) = \begin{pmatrix} 0.02 & 0 & 0 \\ 0 & 0.2 & 0.2 \\ 0 & 0.2 & 0.4 \end{pmatrix} \begin{pmatrix} 0 \\ 5 \\ -2.5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.5 \\ 0 \end{pmatrix}$$