# Esame 17/07/23

DOMANDA 1 (2 punti se giusta, -0.67 se errata)

Sono stati raccolti dieci dati relativi ad una variabile “x”. Quattro di questi dati hanno valore 5, tre hanno valore 2 e tre hanno valore 0. La media campionaria $\bar{x}$ vale:

a) 13/5 [✓]
b) 5
c) 26/5
d) 6
e) Non rispondo

$$\frac{4\cdot 5+3\cdot 2+3\cdot 0}{10}=\frac{20+6}{10}=\frac{26}{10}=\frac{13}{5}$$

DOMANDA 2 (2 punti se giusta, -0.67 se errata)

Un'urna contiene due palline gialle, due rosse e due verdi. Se si estraggono a caso 3 palline distinte, qual è la probabilità che le tre palline estratte abbiano colore diverso?

$P(A|B)=\frac{P(A\cap B)}{PB}$

$\underset{\text{(se indip.)}}{=}\frac{P(A)\cdot P(B)}{P(B)}$

$\binom{n}{k}=\frac{n!}{k!(n-k)!}$

a) 1/2
b) 1/3
c) 2/5 [✓]
d) 1/27
e) Non rispondo

2G 2R 2V = 6P

$$P\binom{\text{3 palline}}{\text{diverse}}=\frac{3!}{\binom{6}{2}}=\frac{3!}{\frac{6!}{2!\,4!}}=3!\cdot\frac{2!\cdot 4!}{6!}=\frac{3!\cdot 2!}{6\cdot 5}=\frac{12}{30}=\frac{6}{15}=\frac{2}{5}$$

$\frac{3!}{\binom{6}{2}}$ → Probabilità di trovare 1 colore distinto 3 volte

DOMANDA 3 (2 punti se giusta, -0.67 se errata)

Un evento A, che ha probabilità 1/3, è indipendente da un evento B. Inoltre $\mathbb{P}(A\cup B)=1/2$. Allora $\mathbb{P}(B)$ vale:

a) 1/6
b) 1/4 [✓]
c) 1/12
d) 1/3
e) Non rispondo

$P(A)=\frac{1}{3}$

$P(A\cup B)=\frac{1}{2}$

$P(A\cup B)=P(A)+P(B)-P(A\cap B)$

$P(A\cup B)=P(A)+P(B)-P(A)\cdot P(B)$

$$P(A \cup B) - P(A) = P(B)(1 - P(A))$$

$$P(B) = \frac{P(A \cup B) - P(A)}{1 - P(A)} = \frac{\frac{1}{2} - \frac{1}{3}}{1 - \frac{1}{3}} = \frac{\frac{1}{6}}{\frac{2}{3}} = \frac{1}{\cancel{6}_2} \cdot \frac{\cancel{3}^1}{2} = \frac{1}{4}$$

<u>DOMANDA 4</u> (2 punti se giusta, -0.67 se errata)

Il valor medio di una variabile aleatoria X discreta con densità "p" è data dalla formula:

a) $\sum_X x\, p(x)$ [✓]
b) $\sum_1^n p(X_i)\, /n$
c) $\sum_x p(x)$
d) $\sum_1^n X_i\, /n$
e) Non rispondo

$$\int_{-\infty}^{+\infty} x\, p(x)\, dx = \sum_x x\, p(x)$$

<u>DOMANDA 5</u> (2 punti se giusta, -0.67 se errata)

In una certa popolazione la circonferenza del polso ha una distribuzione normale di media 23cm e varianza di 9cm$^2$. Tobia afferma che esattamente il 5% di tale popolazione ha la lunghezza del polso più lunga della propria. Qual è la lunghezza del polso di Tobia? (si tenga conto che $z_{0.05}$=1.645 e si ricordi di esprimere il risultato con 2 cifre decimali).

- Risposta aperta: 27.935

$\mu = 23\,cm$ $\sigma^2 = 9\,cm^2$ $z_{0,05} = 1,645$

$z = \frac{x - \mu}{\sigma}$ $1,645 = \frac{x - 23}{3}$ $3 \cdot 1,645 = x - 23$

$x = 3 \cdot 1,645 + 23 = 27,93$

<u>DOMANDA 6</u> (2 punti se giusta, -0.67 se errata)

Una variabile aleatoria di Poisson X di parametro λ è tale che $\mathbb{P}(X = 0) = 1\, /\, e^2$. Allora il valore di λ è:

a) -2
b) 2 [✓]
c) 4
d) 1
e) Non rispondo

$P(X=0) = \frac{1}{e^2}$ $X \sim P_0(\lambda)$ $P(X=x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$

$\frac{1}{e^2} = \frac{1 \cdot e^{-\lambda}}{1}$ $e^{-\lambda} = e^{-2}$ $-\lambda = -2$ $\lambda = 2$

DOMANDA 7 (2 punti se giusta, -0.67 se errata)

Sia $X_1$, $X_2$, $X_3$ un campione casuale di ampiezza 3, estratto da una popolazione di media μ incognita. Per quale valore di "a" lo stimatore per μ definito come $T = {}^{1}\!/_{4}(X_1 + X_2) + aX_3$ risulta corretto?

a) 1/3
b) 1/2 [✓]
c) 3/5
d) 2/5
e) Non rispondo

$$T = \frac{1}{4}(x_1 + x_2) + a x_3$$

$$\text{bias}(T) = E(T) - \theta = E\left(\frac{1}{4}(x_1 + x_2) + a x_3\right) - \mu = 0$$

Quando il bias é 0 lo stimatore é corretto

$$\frac{1}{4}(\mu + \mu) + a\mu - \mu = 0$$

$$\frac{1}{2}\mu + a\mu - \mu = 0$$

$$a\mu = +\frac{1}{2}\mu + \mu$$

$$a = \frac{+\frac{1}{2}\cancel{\mu} + \cancel{\mu}}{\cancel{\mu}}$$

$$a = +\frac{1}{2}$$

DOMANDA 8 (2 punti se giusta, -0.67 se errata)

Si vuole determinare, all'interno di una grande popolazione, la percentuale "p" di individui con una certa caratteristica genetica. I dati su 1000 individui mostrano che 81 di essi possiede tale caratteristica. Calcolare l'estremo superiore di un intervallo di confidenza bilaterale per "p" di livello 0.99 (ricordare che $z_{0.005}$=2.575; esprimere il risultato con 3 cifre decimali).

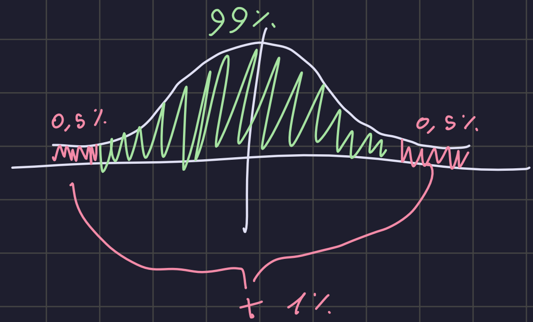
- Risposta aperta: 0.103

Livello 0.99 = Confidenza 0,01 → bilaterale = $\frac{0,01}{2} = 0,005$

$Z_{0,005} = 2,575$

$\sigma^2 = \hat{\theta}(1 - \hat{\theta})$ ↓ Varianza di Bernoulli

$\hat{\theta} = \frac{81}{1000} = 0,081$




$$L_{sup} = \hat{\theta} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$L_{sup} = \hat{\theta} + z_{0,005} \cdot \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} = 0,081 + 2,575 \cdot \sqrt{\frac{0,081(1-0,081)}{1000}} = 0.103$$

DOMANDA 9 (2 punti se giusta, -0.67 se errata)

Viene effettuato un test per verificare che una certa ipotesi nulla $H_0$. Se il valore-p sui dati vale 0.003, si può concludere che:

a) I dati sono in forte disaccordo con $H_0$ [✓]
b) $H_0$ è falsa
c) I dati sono in forte accordo con $H_0$
d) $H_0$ è vera
e) Non rispondo

pvalue=0,003

Se p-value é molto piccolo l'ipotesi nulla é rifiutabile

## DOMANDA 10 (7 punti, parte di laboratorio)

Spiegare passo passo il seguente script R, commentandone il risultato:

```
1  set.seed(123)
2  dado1 <- sample(x = 1:6, size = 10000, replace = TRUE,
3    prob = c(rep((1 - 3/6)/5,5), 3/6))
4  dado2 <- sample(x = 1:6, size = 10000, replace = TRUE)
5  dadi_diff <- dado1 - dado2
6  barplot(table(dadi_diff), horiz = FALSE, main = "Dado 1 - Dado 2",
7    xlab = "Differenza", ylab = "Frequenza assoluta")
8  p_diff_5 <- cumsum(dadi_diff == 5)/1:10000
9  plot(x = 1:10000, y = p_diff_5, type = "l", log = "x",
10   main = "P(dado1 - dado2 = 5)",
11   xlab = "Numero di lanci",
12   ylab = "Probabilità")
13 abline(h = 3/6*1/6, col = "red", lty = "dashed")
14 abline(h = 1/6*1/6, col = "blue", lty = "dotted")
15 legend("topright", col = c("red", "blue"),
16   lty = c("dashed", "dotted"),
17   legend = c("dado1: non equo, dado2: equo", "dadi equi"))
```

1-4 Setta un seed e genera dei dati casuali da mettere nelle variabili dado1 e dado2, precisamente genera 10000 campioni con valori da 1 a 6 con rimpiazzo. Il dado 1 ha il 10% di probabilità di ottenere da 1 a 5 e il 50% di probabilità di ottenere 6.

5 Calcola la differenza tra i 2 dadi.

6-7 Crea un grafico a barre della differenza tra i dadi con le seguenti opzioni: barre verticali, un titolo e delle etichette sugli assi.

8 Probabilità che la differenza tra i 2 dadi dia come risultato 5

9-12 Crea un grafico a linea che mappa la probabilità che la differenza dei dadi sia 5 al numero di lanci. Le opzioni del grafico sono:
tipo di grafico a linea, log "x" comprime i dati sull'asse x man mano che si va verso destra, un titolo e delle etichette sugli assi.

13-14 Aggiunge due linee orizzontali che indicano l'andamento asintotico dei dati ottenuti.
La prima (a riga 13) indica la probabilità che la differenza tra il dado 1 non equo e il dado 2 equo sia 5, che è rossa e tratteggiata.
La seconda (a riga 14) indica la probabilità che la differenzat tra 2 dadi equi sia 5, che è blu e puntinata.

15-17 Aggiunge al grafico una legenda in alto a destra che spiega a cosa si riferiscono le rette a riga 13 e 14.

DOMANDA 11 (7 punti, parte di laboratorio)

Un magazzino di stoffe possiede diversi rotoli di stoffa. Il proprietario si è accorto che alcuni rotoli sono bucati e si chiede se esista una relazione di tipo lineare tra la lunghezza dei rotoli e il numero di buchi. Procede quindi a registrare, per ciascun rotolo, la lunghezza e il numero di buchi. Il seguente oggetto contiene i dati raccolti:

```
stoffe <- data.frame(
    rotolo = c("Rosso", "Blu", "Verde", "Damascato", "Fiori", "Pois",
        "Optical", "Geometrico", "Bimbo", "Leopardo", "Zebra"),
    lunghezza = c(7, 14.7, 8.7, 11.9, 7.9, 5.1, 8.3, 6.5, 9.2, 5.7, 12.7),
    buchi = c(3, 7, 4, 4, 1, 1, 3, 1, 4, 0, 3))
```

Rispondi ai seguenti punti riportando i comandi R utilizzati:

1) Calcolare il coefficiente di correlazione lineare tra il numero di buchi e la lunghezza della stoffa;
2) Calcolare i coefficienti di un modello di regressione lineare semplice per descrivere la relazione tra la variabile dipendente "buchi" e la variabile indipendente "lunghezza" dei rotoli. Quanto vale l'intercetta? E il coefficiente angolare?
3) Utilizzando il comando plot() genera il grafico a dispersione per i valori osservati di lunghezza e numero di buchi. Poi con il comando abline() sovrapponi al grafico la retta di regressione stimata colorandola di rosso. Ricorda di dare il giusto nome agli assi e al titolo del grafico.
4) Il proprietario si accorge di avere un altro rotolo che ha dimenticato di inserire nella struttura dati iniziale. Non ha contato i buchi ma sa che il rotolo è lungo 9.2 metri. Utilizzando il modello di regressione che hai stimato quanti buchi sono previsti per una tale lunghezza (arrotonda all’intero più vicino)?