- Appello del 6 Febbraio 2013 -

Esercizio 1)
Si è lanciata una moneta 14 volte e si sono ricavati i seguenti esiti dove T = Testa e C = Croce.

T C T T C C C T T C T C T T

a) Determini la tipologia del carattere.
b) Fornisca una rappresentazione grafica dei dati.
c) Si indichino e si calcolino tutti gli indici di posizioni adeguati ai dati.
d) Si indichino gli indici di variabilità adeguati ai dati e, se possibile, se ne calcoli uno.
e) Indicare e calcolare gli indici di posizione validi se, invece di usare le lettere T e C, si fossero indicati con il numero 0 l'esito di Testa e con il numero 1 l'esito di Croce.

Esercizio 2)
Un laboratorio di analisi si serve di acido solforico da un fornitore che garantisce che la concentrazione del prodotto ha una valore pari al 4%. Per verificare la veridicità della dichiarazione del fornitore si sono misurate le concentrazioni presenti in 21 confezioni ottenendo le seguenti misurazioni:

4.0% 3.8% 4.1% 4.0% 4.2% 4.0% 3.9% 3.9% 4.0% 4.0% 4.0%
3.9% 4.1% 3.8% 4.0% 3.9% 4.2% 4.0% 4.0% 4.0% 4.2%

Il candidato verifichi con un test di ipotesi ad un opportuno livello di significatività la validità dell'asserzione del fornitore. Il candidato, prima di procedere al test, indichi le necessarie ipotesi e proceda al calcolo anche qualora queste non siano verificate

Esercizio 3)
Il candidato, utilizzando i dati riportati nell'Esercizio 2, stimi puntualmente e per intervallo la varianza della v. c. che indica la concentrazione di acido solforico di una boccetta prodotta dal fornitore descritto nell'Esercizio 2. Prima di procedere alla stima il candidato indichi le necessarie ipotesi e proceda al calcolo anche qualora queste non siano verificate.

Esercizio 4)
Si considerino i seguenti eventi dichiarati incompatibili.

E_1 : si ottenga $x<0.5$ dove x è estratto da una v. c. distribuita come una normale standardizzata.
E_2: si ottenga $y=0$ dove y è estratto da una v. c. distribuita come $Ber(0.8)$.

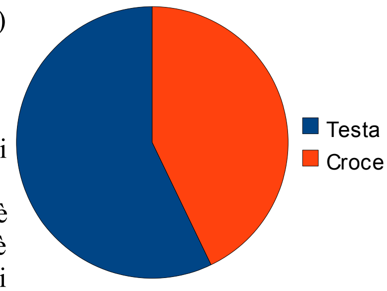
a) Il candidato calcoli le seguenti Probabilità $P(E_1)$; $P(E_2)$; $P(E_1 \cup E_2)$ $P(E_1 | E_2)$.
b) Il candidato fornisca la definizione di eventi statisticamente indipendenti ed indichi se i due eventi E_1 ed E_2 siano indipendenti.

- Prova in itinere del 6 Febbraio 2013 -
Svolgimento

a) Determini la tipologia del carattere.

Il carattere è di tipo qualitativo (in quanto espresso da esiti e non numeri) non ordinabile (gli esiti non possono essere ordinati).

b) Fornisca una rappresentazione grafica opportuna.



Per caratteri qualitativi non ordinabili sono possibili due tipologie di rappresentazioni grafiche: il diagramma a barre ed il diagramma a torta.
In questa soluzione si è scelto di rappresentare il secondo. Un diagramma a torta è costituito da una cerchio diviso in settori, uno per modalità, la cui ampiezza è proporzionale alla frequenza relativa della modalità ad essa legata. A lato si mostra il diagramma relativo ai dati in oggetto.

c*) Si indichino e si calcolino tutti gli indici di posizioni adeguati ai dati*.

L'indice di posizione di una serie di osservazioni indica il valore centrale che viene assunto dalla serie. Gli indici di posizione visti a lezione sono tre: moda, mediana e media. Per i caratteri in esame sono calcolabili sono i primi due. La moda (ovvero la modalità cui corrisponde la modalità maggiore) della serie è Testa.

d) Si indichino gli indici di variabilità adeguati ai dati e, se possibile, se ne calcoli uno.

Per dati come qualitativi non è possibile determinare alcun indice di variabilità.

e) Indicare e calcolare gli indici di posizione validi se, invece di usare le lettere T e C, si fossero indicati con il numero 0 l'esito di testa e con il numero 1 l'esito di croce.

Nel caso di indici numerici (quantitativi) è possibile calcolare tutti gli indici visti a lezione.

- **Moda:** questo indice risulta calcolato in maniera analoga al punto precedente; moda = 0
- **Mediana:** questo indice identifica l'osservazione che bipartisce le osservazioni ordinate. Nel caso in esame l'osservazione deve essere preceduta da (N-1)/2 = 13/2 = 6.5 osservazioni. Poiché il numero non è tondo si esegue la media fra la settima e l'ottava osservazione. Essendo entrambe pari a 0 si ha che la mediana è 0.
- **Media:** la media di osservazioni con frequenze non unitarie è data dalla seguente formula:

$$\frac{1}{N}\sum_{i=1}^{M} n_i x_i = \frac{1}{14}(8*0+6*1) = \frac{6}{14} = 0.4286$$

Esercizio 2)

Nel testo si effettuano diverse misure di una grandezza ignota da stimare. Possiamo modellare il processo di generazione dei dati come eseguire diverse estrazioni dalla variabile casuale

P :*concentrazione di acido solforico in una confezione*

avente distribuzione ignota. Si sono effettuate $n = 21$ osservazioni i. i. d. in cui si son rilevate $M = 5$ modalità.
Continuando con il modello precedentemente definito l'esercizio richiede di verificare l'ipotesi che che $E[P]=4$. L'ipotesi alternativa è quella che nel nostro caso è dannosa ovvero che la concentrazione sia diversa da quanto dichiarato. Riassumendo si hanno le seguenti ipotesi:

$$H_0\colon E[P]=4 \qquad\qquad H_1\colon E[P] \neq 4$$

La scelta fra queste ipotesi può essere realizzata, utilizzando gli strumenti forniti nel corso, solo se la le estrazioni sono indipendenti ed identicamente distribuite e se la dimensione del campione è pari ad almeno 30 unità. Avendo a disposizione solo 11 valori quest'ultima condizione non è verificata. Ciononostante il testo richiede di procedere in ogni caso al test.

Il primo passo in un test di ipotesi consiste nel determinare la distribuzione limite dello stimatore. In questo caso lo stimatore di riferimento è la media campionaria $\overline{X}$ la cui distribuzione limite è $\overline{X} \sim N\left(E[P]; \frac{Var[P]}{n}\right)$. Per valutare la distribuzione limite si deve stimare puntualmente la varianza tramite la varianza campionaria s^2.

$$\hat{Var}[P] = s^2 = \frac{1}{n-1}\sum_{i=1}^{M} n_i (x_i - \bar{x})^2 = \frac{1}{20}\sum_{i=1}^{5} n_i (x_i - 4)^2 = 0.013$$

Si ottiene che la distribuzione limite dello stimatore è la seguente:

$$\overline{X} \sim N(E[P]; \frac{Var[P]}{n}) = N(4; \frac{0.013}{21}) = N(4; 0.0006)$$

Fissata la convergenza dello stimatore si procede alla determinazione della regione di accettazione posta per comodità sulla normale standardizzata. Essendo l'ipotesi alternativa un'ipotesi di disuguaglianza il tipo di test è detto a due code (o simmetrico). Ponendo un livello di significatività pari al *90%* si ottengono i seguenti valori critici

$$z_{inf} = -1.65 \qquad\qquad z_{sup} = 1.65$$

Determinata la regione di accettazione *A=[-1.65; 1.65]* possiamo ora standardizzare il valore dello stimatore calcolato sui dati reali e vedere se esso è interno alla regione. Sfruttando i conti fatti in tabella 1 ricaviamo il valore dello stimatore

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{M} n_i x_i = \frac{1}{21} 84 = 4$$

Che standardizzato diviene.

$$z_{\bar{x}} = \frac{\bar{x} - E[P]}{\sqrt{Var[P]}} = \frac{4-4}{\sqrt{0.0006}} = 0$$

Poiché il valore appartiene all'intervallo di accettazione, si accetta l'ipotesi nulla e si rifiuta l'ipotesi alternativa. Si ritiene dunque veritiera la dichiarazione del fornitore ad un livello di significatività del 90%.

Esercizio 3)

Continuando con il modello precedentemente fatto, l'esercizio richiede di stimare *E[X]* puntualmente e per intervallo. Le ipotesi richieste sono le medesime descritte al punto precedente per tanto restano valide le stesse considerazioni.

La stima puntuale può essere effettuata ricordando che la varianza viene stimata correttamente mediante la varianza campionaria. Il calcolo è già stato effettuato nello svolgimento del primo esercizio, ottenendo

$$\hat{Var}[P] = s^2 = \frac{1}{n-1} \sum_{i=1}^{M} n_i (x_i - \bar{x})^2 = \frac{1}{20} \sum_{i=1}^{5} n_i (x_i - 4)^2 = 0.013$$

Le stime per intervallo sono regolate dal livello di confidenza (*1-α*) che solitamente è fissato da chi svolge l'analisi dei dati. Nel caso in esame una scelta valida è porre *α = 10%*. Determinato il livello di confidenza la stima *I* è data dalla seguente formula:

$$I = \left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}; \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right] = \left[\frac{20 \cdot 0.013}{\chi^2_{0.95}(20)}; \frac{20 \cdot 0.013}{\chi^2_{0.05}(20)} \right] = \left[\frac{0.26}{31.410}; \frac{0.26}{10.851} \right] = [0.008; 0.024]$$

i	x_i	n_i	$n_i x_i$	$(x_i - \bar{x})^2$	$n_i(x_i - \bar{x})^2$
1	3.8	2	7.60	0.04	0.0800
2	3.9	4	15.60	0.01	0.0400
3	4.0	10	40.00	0.00	0.0000
4	4.1	2	8.20	0.01	0.0200
5	4.2	3	12.60	0.04	0.1200
Somma		21	84.00		0.2600

Tabella 1 - conti esercizio 2 e 3

Esercizio 4)

a) Il candidato calcoli le seguenti Probabilità: $P(E_1)$; $P(E_2)$; $P(E_1 \cup E_2)$ $P(E_1 | E_2)$.

Per il calcolo di $P(E_1)$, si deve calcolare l'area sottesa dalla curva della densità di probabilità della distribuzione normale da meno infinito a 0.5. Questa probabilità si può desumere dalle tavole. Si ottiene infatti che

$$P(E_1)=P(Z<0.50) = 69.15\%$$

La probabilità dell'evento E_2 è pari alla probabilità di avere un esito negativo (ovvero pari a zero) estraendo un valore da una distribuzione di Bernoulli. Ricordando che una generica distribuzioni di Bernoulli di parametro *p* (ovvero *Ber(p)*) sono possibili solo due esiti (0 ed 1) e che il parametro indica il la probabilità dell'esito positivo si ottiene immediatamente che:

$$P(E_2)=P(X=0) = 1 - P(X=1) = 1 - p = 1- 0.8 = 0.2$$

Propedeutico al calcolo delle altre due probabilità e il calcolo della probabilità dell'evento intersezione (ovvero che i due eventi si verifichino contemporaneamente). Per eventi incompatibili essa e nulla per definizione

$$P(E_1 \cap E_2)=0$$

Le restanti probabilità possono essere ricavate utilizzando la definizione assiomatica

$$P(E_1 \cup E_2)=P(E_1)+P(E_2)-P(E_1 \cap E_2)=0.6915+0.2-0=0.8915 \qquad P(E_1|E_2)=\frac{P(E_1 \cap E_2)}{P(E_2)}=\frac{0}{0.2}=0$$

b) Il candidato fornisca la definizione di eventi statisticamente indipendenti ed indichi se i due eventi E_1 ed E_2 siano indipendenti.

Due eventi si dicono statisticamente indipendenti se il verificarsi di un evento non altera la probabilità del verificarsi dell'altro. Pertanto se due eventi sono statisticamente indipendenti la probabilità di un evento condizionato è pari all'evento stesso. Nel punto precedente questa probabilità risulta nulla e differente da $P(E_1)$, quindi i due eventi non sono incompatibili.