Compito scritto dell'esame di Statistica e analisi dei dati (Crema) 01 settembre 2017	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e di 2 ore e 30 minuti. Durante la prova e possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu passaggi di calcolo, non sara preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara eventualmente rappresentato in quest'ultima forma).
- SOLO PER STUDENTI DEL CORSO ONLINE (VECCHIO ORDINAMENTO, Prof. Gianini): Indicare quale Modulo viene svolto al fine di completare una prova d'esame parziale precedentemente sostenuta con esito positivo
  - Modulo 1 ☐
  - Moduli 2 e 3 ☐

Modulo 1

Esercizio 1. Una moneta da 2 euro (il suo diametro é 25,75 mm) viene lanciata su un pavimento ricoperto con 100 mattonelle quadrate di lato $10cm$.

(a) Qual é la probabilitá che la moneta vada a finire internamente ad una mattonella? (cioé non tagli i lati dei quadrati)

Soluzione: Immaginiamo di lanciare la moneta su una mattonella quadrata di lato $10cm$. Fissiamo l'attenzione sul centro C della moneta e osserviamo che la probabilitá che esso cada in una regione R di area $a\ cm^2$, contenuta nel quadrato, é proporzionale all'area, ossia

$$P(C \in R) = \frac{a}{10^2}$$

La moneta cade all'interno della mattonella quando R é un quadrato concentrico alla mattonella e con lato $(10 - 2{,}575)cm$. Pertanto:

$$P(C \in R) = \frac{(10 - 2{,}575)^2}{10^2} \approx 0.551$$

Il risultato precedente continua a valere nel caso di un pavimento formato da n mattonelle quadrate, infatti

$$P(C \in R) = \frac{(10 - 2{,}575)^2 \times n}{10^2 \times n} \approx 0.551$$

Esercizio 2. Una scatola contiene: due palline bianche numerate "0"; otto palline bianche numerate "1"; quattro palline rosse numerate "0"; sei palline rosse numerate "1". Si estrae una pallina casualmente dalla scatola.

(a) Gli eventi "esce una pallina bianca" e "esce una pallina numerata con 0" sono indipendenti?

Soluzione: Il numero di palline é uguale a 20. Definiamo gli eventi seguenti:

$$A = \text{"esce una pallina bianca"};$$

$$B = \text{"esce una pallina numerata con 0"}.$$

Siccome ci sono 10 palline bianche e 6 palline numerate con 0,

$$P(A) = \frac{1}{2}, P(B) = \frac{6}{20}.$$

Si definisca l'evento $\{A \cap B\} =$ "esce una pallina bianca numerata con 0". Allora:

$$P(A \cap B) = \frac{2}{20} \neq \frac{3}{20} = P(A)P(B)$$

Dunque A e B non sono indipendenti.

(b) Spiegare brevemente il perché del risultato ottenuto in (a).

Soluzione: A e B non sono indipendenti in quanto la proporzione di palline numerate con 0 é diversa tra le palline bianche e le palline nere.

Esercizio 3. Consideriamo un test diagnostico per evidenziare la presenza di una patologia nel sangue. Il test ha probabilitá 0.8 di individuare correttamente la presenza della patologia, ma sappiamo che dá risultato positivo (ossia, indica la presenza della patologia) anche nel 20% degli esami compiuti su soggetti sani. Supponiamo che la presenza della patologia nella popolazione sia pari al 5%

(a) Determinare la probabilitá che un soggetto sia malato dopo che il test ha dato risultato positivo

Soluzione: (Teorema di Bayes)
Indichiamo con M l'evento "soggetto malato" e con T l'evento "test positivo".
I dati del problema ci dicono che $P(M) = 0.05, P(T \mid M) = 0.8, P(T \mid\sim M) = 0.2$
Usando Bayes:

$$P(M \mid T) = \frac{P(T \mid M)P(M)}{P(T \mid M)P(M) + P(T \mid\sim M)P(\sim M)} = \frac{4}{23} \approx 0.174. \quad (1)$$

(b) Determinare la probabilitá che un soggetto sia sano dopo che il test ha dato risultato negativo

Soluzione:
Usando sempre Bayes:

$$P(\sim M \mid\sim T) = \frac{P(\sim T \mid\sim M)P(\sim M)}{P(\sim T \mid\sim M)P(\sim M) + P(\sim T \mid M)P(M)} = \frac{76}{77} \approx 0.987. \quad (2)$$

Moduli 2 e 3

Esercizio 1. In media, si impiegano 10 minuti per guidare da casa all'universitá.

(a) Supponiamo per semplicitá che il tempo necessario, T , segua una legge esponenziale. Se parto da casa alle $8:45$, qual é la probabilitá di arrivare in universitá per le 9?

Soluzione: T segue una legge esponenziale di parametro

$$\lambda = \frac{1}{10};$$

Pertanto

$$P(T < 15) = 1 - e^{-\frac{15}{10}} \approx 77.7\%$$

(b) A che ora devo partire per avere una probabilitá del 99% di arrivare in tempo?

Soluzione: Per arrivare in tempo con una probabilitá del 99%, devo trovare qual é quel tempo t per cui $P(T \leq t) = 0.99$, ovvero

$$(1 - e^{-\lambda t}) = 0.99$$
$$-\lambda t = \log(0.01)$$
$$t = 10 \log(100) = 46.0517 min.$$

da cui partendo verso le $8:10$ ho (statisticamente)una ragionevole certezza di puntualitá.

Esercizio 2.

Al servizio di soccorso stradale di una certa cittá, aperto 24 ore su 24, arrivano in media 48 chiamate al giorno, due in media all'ora

(a) Calcolare la probabilitá che nella prima ora arrivino almeno due chiamate

Soluzione: (Distribuzione di Poisson) Possiamo assumere che le chiamate si distribuiscano con legge di Poisson. La variabile aleatoria X che descrive il numero di chiamate per ora ha media

$$\mu = 2,$$

pertanto la distribuzione di X si scrive

$$p_X(x) = \frac{2^x e^{-2}}{x!}.$$

Le probabilitá che non arrivino chiamate o che ne arrivi solo una sono date rispettivamente da

$$P(X = 0) = p_X(0) = \frac{2^0 e^{-2}}{0!} = e^{-2},$$
$$P(X = 1) = p_X(1) = \frac{2^1 e^{-2}}{1!} = 2e^{-2}.$$

Usando la probabilitá complementare si ha

$$P(X > 1) = 1 - P(0 \leq X \leq 1) = 1 - 3e^{-2}.$$

(b) Calcolare la probabilitá che il tempo di attesa fino alla prima chiamata di un certo giorno sia di almeno un'ora

Soluzione: Il quesito si risolve immediatamente notando che

$$P(T \geq 1) = P(X = 0) = e^{-2}.$$

Esercizio 3. Una macchina produce barre di acciaio a sezione circolare la cui lunghezza ottimale dovrebbe essere di 5 metri ed il diametro della sezione di 4 centimetri. Le barre effettivamente prodotte, che si suppongono tra loro indipendenti, hanno una lunghezza aleatoria con distribuzione normale di media $\mu_1 = 5\ m$ e scarto standard $\sigma_1 = 4\ cm$. Il diametro della sezione é una variabile aleatoria, indipendente dalla precedente, e con distribuzione normale di media $\mu_2 = 4\ cm$ e scarto standard $\sigma_2 = 0{,}8\ cm$. Una generica barra prodotta puó essere direttamente venduta senza modifiche se la sua lunghezza é compresa tra $4{,}95m$ e $5{,}05m$ e la sua sezione tra $2{,}8\ cm$ e $5{,}2\ cm$.

(a) Si verifichi che la probabilitá p di poter mettere in vendita senza modifiche una generica barra prodotta é 0.68

Soluzione: La variabile aleatoria l che descrive la lunghezza di una barra ha densitá di probabilitá normale. Dunque, standardizzando l con

$$z_l = \frac{l - \mu_1}{\sigma_1}$$

e facendo uso delle tavole della normale standard si ha che

$$P(495cm \leq l \leq 505cm) = \Phi(\frac{5}{4}) - \Phi(-\frac{5}{4}) \approx 0.788$$

Analogamente, per il diametro d della sezione:

$$P(2{,}8cm \leq d \leq 5{,}2cm) = \Phi(\frac{3}{2}) - \Phi(-\frac{3}{2}) \approx 0.866$$

Per indipendenza di l,d:

$$P(\{495cm \leq l \leq 505cm\} \cap \{2{,}8cm \leq d \leq 5{,}2cm\}) = P(495cm \leq l \leq 505cm) \times P(2{,}8cm \leq d \leq 5{,}2cm) \approx 0.788 \times 0.866 = 0.68$$

(b) Indicata con f_n la frequenza relativa delle barre direttamente vendibili su n barre prodotte, si esprima, in funzione di p, una stima della numerositá n necessaria affinché la probabilitá che f_n disti da p piú di 0.05 sia non superiore a 0.05

Soluzione: Sia X la variabile aleatoria che descrive il numero di barre vendibili, tale che $f_n = \frac{X}{n}$.
Tale variabile aleatoria ha distribuzione binomiale (numero barre vendibili / non vendibili) con media

$$\mu_X = np$$

e varianza

$$\sigma_X^2 = np(1-p)$$

dove $p \approx 0.68$ é la probabilitá calcolata in (a).
Il quesito richiede di determinare la probabilitá

$$P(|\frac{X}{n} - p| \geq 0.05) \leq 0.05 \tag{3}$$

Possiamo applicare il Teorema di Chebychev nella forma

$$P_X(|X - \mu_X| \geq r) \leq \frac{\sigma_X^2}{r^2},$$

ovvero

$$P_X(|X - np| \geq r) \leq \frac{np(1-p)}{r^2},$$

dove per confronto con eq. (3) si vede che $\frac{r}{n} = 0.05$. Il problema impone pertanto che il termine a destra della diseguaglianza di Chebychev sia ≤ 0.05. Dunque si ha che

$$\frac{np(1-p)}{n^2(0.05)^2} \leq 0.05$$

da cui, sostituendo i valori numerici noti si ottiene che

$$n \geq 1741$$