

# Tema d'esame di Statistica e analisi dei dati

Prova scritta del 7 febbraio 2018

## Esercizio 0

1. Siano  $A$  e  $B$  due eventi. Si esprima, in funzione di  $P(A)$ ,  $P(B)$  e  $P(B|A)$ , la probabilità  $P(A|B)$ .

## Esercizio 1

Sia  $X$  una variabile casuale bernoulliana di parametro  $p$ .

1. Quali sono i punti di massa di  $X$ ?
2. Si esprima, in funzione di  $p$ , la probabilità  $P(X \leq 0.5)$ .
3. Si controlli che  $Var(X) \leq 1/4$ .
4. Lo stimatore  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  è uno stimatore non distorto per il parametro  $p$ ? Si giustifichi la risposta.
5. Sia  $\epsilon > 0$  un numero reale e sia  $n > 1$  un intero. Dato un campione casuale  $X_1, \dots, X_n$  estratto dalla popolazione  $X$ , si controlli che

$$P(|\frac{1}{n} \sum_{i=1}^n X_i - p| \leq \epsilon) \geq 1 - \delta$$

è vera se:  $\delta \geq \frac{1}{4 \cdot \epsilon^2 n}$

6. Sia  $Y$  una variabile aleatoria binomiale di parametri  $n = 37$  e  $p = 0.35$ . La Figura 1 mostra il grafico della funzione di ripartizione di  $Y$ .  
Sia  $N$  una variabile aleatoria normale con stesso valore atteso e stessa varianza di  $Y$ . Sulla stessa figura tracciare a penna un grafico indicativo della funzione di ripartizione di  $N$ .
7. Calcolare le seguenti probabilità:
  - 7.1.  $P(|Y - E(Y)| \leq 1.5)$
  - 7.2.  $P(|N - E(N)| \leq 1.5)$

## Esercizio 2

Sul server del mercato del pesce ogni sera, alla chiusura, per ogni settore di pesca in cui sono divise le acque, viene creato un record con le seguenti informazioni:

- *giorno.settimana*: giorno della settimana,
- *settore.di.pesca*: settore di pesca,
- *settore.num*: settore di pesca ricodificato con valori numerici,
- *peso.pescato*: quantità di pesce complessiva pescata nel dato settore di pesca (espressa in quintali),
- *forza.del.mare*: condizioni del mare nel dato settore di pesca,
- *tempesta*: mare tempestoso,

Le colonne sono separate dal simbolo "," e i numeri reali sono stati registrati con il simbolo "." come separatore dei decimali. Per accedere al file collegarsi al sito [upload.di.unimi.it](http://upload.di.unimi.it), selezionare l'esame di *Statistica e analisi dei dati* e scaricare il file `pesce.csv`.

1. Quante sono le giornate lavorative in una settimana al mercato del pesce?
2. Per quanti giorni sono state raccolte le informazioni?
3. Quante volte si è pescato in condizioni di tempesta? Dare una risposta in termini di frequenza sia assoluta, sia relativa.
4. Tracciare un grafico rappresentativo della frequenza di *tempesta*.
5. Quanti sono i settori di pesca della parte di mare considerata?
6. Qual è la frequenza relativa del settore "A"?
7. Produrre la tabella delle frequenze congiunte tra *settore.di.pesca* e *tempesta*.
8. Utilizzare la tabella delle frequenze congiunte per stabilire, relativamente al settore di pesca "A", quanti sono stati i giorni di tempesta e quanti quelli senza tempesta.

Supponiamo, d'ora in poi, che sia verificata la condizione (realistica forse soltanto per alcune zone del pianeta) che le condizioni atmosferiche di un dato giorno siano indipendenti dalle condizioni atmosferiche di un altro giorno.

9. Sulla base dei dati a disposizione stimare la probabilità  $p_T$  che oggi ci sia tempesta.

10. Qual è la taglia del campione che avete utilizzato nella stima di  $p_T$ ?
11. Stimare la probabilità  $p_{T_A}$  che ci sia tempesta nel settore "A".
12. Un pescatore dice che ieri ha incontrato condizioni di tempesta. Qual è la probabilità che abbia lavorato nel settore di pesca "A"?
13. Qual è la taglia  $n_A$  del campione che avete utilizzato al punto 11. per la stima della probabilità che ci sia tempesta nel settore "A"?
14. Si determini un limite inferiore per la probabilità di aver compiuto, nella stima di  $p_{T_A}$ , un errore, in eccesso o in difetto, minore o uguale di 0.1.

## Esercizio 3

1. Il carattere *peso.pescato* è discreto o continuo?
2. Si tracci il grafico che meglio rappresenta la distribuzione del *peso.pescato*.
3. Calcolare la percentuale di giorni in cui al mercato è arrivata una quantità di pescato compresa tra i 10 e i 15 quintali.
4. Determinare i parametri della variabile casuale normale che potrebbe descrivere il *peso.pescato*.
5. Secondo il modello identificato al punto precedente, calcolare la probabilità che al mercato arrivi giornalmente una quantità di pescato compresa tra i 10 e i 15 quintali.
6. Si possono confrontare le risposte dei punti 3. e 5.?
7. In Tabella 1 trovate la matrice di covarianza dei caratteri del dataset. Tra quali caratteri sussiste una relazione lineare pronunciata?
8. Prendiamo in considerazione i caratteri *forza.del.mare* e *tempesta*. Tracciare un grafico opportuno per poter individuare l'esistenza di una relazione tra i due caratteri.
9. Il grafico ottenuto al punto precedente dovrebbe evidenziare l'esistenza di una precisa relazione tra i due caratteri. Caratterizzate tale relazione.

	giorno.settimana	peso.pescato	settore.num	forza.del.mare	tempesta
giorno.settimana	1.00	0.01	-0.04	-0.04	-0.06
peso.pescato	0.01	1.00	0.79	-0.68	-0.56
settore.num	-0.04	0.79	1.00	-0.15	-0.12
forza.del.mare	-0.04	-0.68	-0.15	1.00	0.84
tempesta	-0.06	-0.56	-0.12	0.84	1.00

Tabella 1: Matrice di correlazione

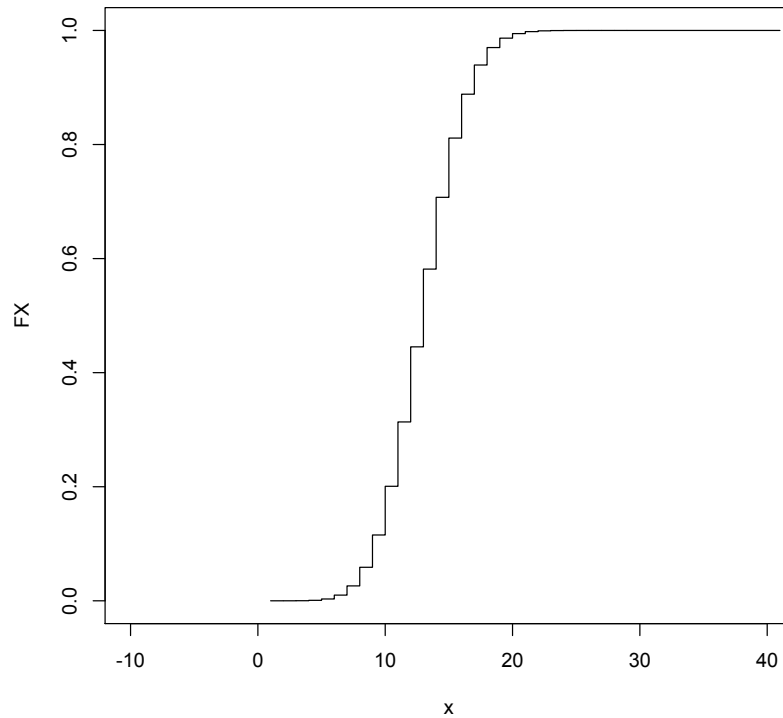


Figura 1: Il grafico della funzione di distribuzione di ripartizione di  $Y$

10. La relazione che avete individuato è compatibile con il coefficiente di correlazione presente nella matrice di correlazione? Giustificare la risposta.
11. In alcuni giorni non sono stati registrati valori per il carattere *peso.pescato*. Create una nuova variabile, chiamata **dati\_senza\_NA**, in cui memorizzare tutti i casi del dataset eccetto quelli che hanno il *peso.pescato* mancante e calcolate, a partire da questo nuovo dataset, il coefficiente di variazione del *peso.pescato*.
12. Dicono che il mare dall'altra parte della penisola garantisca una pescosità più stabile, cioè la variabilità del pesce pescato da quelle parti non è molto alta rispetto alla quantità media. Il bollettino dei pescatori riporta la tabella 2 per descrivere la quantità di pesce giornaliero in arrivo a un porto dell'altra parte della penisola. Siete d'accordo con questa diceria? Usate i dati della tabella per supportare la vostra risposta.

Tabella 2: Tabella descrittiva della quantità di pesce pescato altrove

MAX	40.1	DEV.ST	4
MEDIA	8.7	Primo quartile	3.5
MEDIANA	8.1	Secondo quartile	14.2