

Homework # 4 (da consegnare entro il 4.11.2010)

Esercizio A

La seguente tabella riporta il numero di pagine trascritte (X) e il numero di errori di battitura (Y) relativi a 10 documenti:

<i>Documento</i>	1	2	3	4	5	6	7	8	9	10
<i>n. pagine (X)</i>	5	5	10	12	10	10	12	15	15	15
<i>n. errori (Y)</i>	2	3	3	3	3	3	4	4	4	4

- A.1 Quali sono le unità statistiche? Qual è la natura delle variabili?
- A.2 Costruire la tabella di contingenza della variabile doppia (X, Y)
- A.3 Derivare la distribuzione di frequenza relativa per la v.s. condizionata $Y|X=10$ e farne una rappresentazione grafica
- A.4 Calcolare media, mediana e moda della distribuzione del punto precedente
- A.5 Costruire la tabella di contingenza teorica di indipendenza statistica e dire se X e Y sono indipendenti

Esercizio B

I 100 impiegati di un'azienda sono stati classificati in base al loro peso corporeo (Y : sottopeso, peso normale, sovrappeso) e alla loro propensione al fumo (X : fumatori, non fumatori). I risultati ottenuti sono stati i seguenti:

- 1/5 degli impiegati ha dichiarato di essere fumatore abituale; di questi metà è sottopeso, 1/4 ha peso normale e 1/4 è sovrappeso;
- fra i non fumatori, i 3/4 hanno un peso normale, mentre 1/4 è sovrappeso.

Si costruisca la tabella a doppia entrata contenente le frequenze relative congiunte e marginali per la variabile bivariata (X, Y); in base alla tabella, si fornisca la quota di impiegati che non sono fumatori e sono sovrappeso.

Esercizio C

Un'industria di cioccolata ha confezionato i cioccolatini prodotti in 6 diversi tipi di scatole, che si differenziano per colore e forma. Nella tabella seguente è riportata la distribuzione doppia di frequenza delle scatole di cioccolatini vendute in un mese.

<i>Forma</i>	<i>Colore</i>		
	<i>rosso</i>	<i>verde</i>	<i>Totale</i>
<i>quadrata</i>	800	1600	2400
<i>rettangolare</i>	300	250	550
<i>esagonale</i>	1950	200	2150
<i>Totale</i>	3050	2050	5100

- C.1 Costruire la distribuzione di frequenza relativa condizionata delle scatole rosse per forma.
- C.2 Costruire la distribuzione di frequenza relativa marginale delle scatole in base al colore.
- C.3 Misurare l'associazione tra i due caratteri mediante un opportuno indice.
- C.4 Costruire la distribuzione bivariata teorica delle frequenze assolute in caso di indipendenza.

Esercizio D

La Tabella 2 riporta la distribuzione per numero di *esami sostenuti* alla fine del primo anno e *scuola di provenienza* degli iscritti ad un certo corso di laurea in una dato anno accademico.

Tabella 2 –(dati fittizi)

esami	scuola di provenienza	
	liceo	altro
0	40	60
1	50	50
2	60	40
3	70	30
4	80	20
totale	300	200

- D.1 costruire la distribuzione marginale per *esame*;
- D.2 costruire la distribuzione condizionata di *esame* per gli studenti che provengono da un liceo;
- D.3 senza calcolare indici, dire se le due variabili considerate sono indipendenti in media (motivare la risposta fornita);
- D.4 calcolare il rapporto di correlazione e commentare i risultati ottenuti.

Esercizio E

Ad un campione di 40 persone con reddito annuo netto tra 10000 e 20000 euro è stato chiesto di indicare il reddito annuo netto (X) e la spesa per viaggi turistici (Y). Le statistiche utili per la regressione lineare di Y su X sono: $\bar{x} = 14300.000$; $\bar{y} = 1720.034$; $\sum_i (x_i - \bar{x})(y_i - \bar{y}) = 61530724.367$; $\sum_i (x_i - \bar{x})^2 = 588400000.000$; $\sum_i (y_i - \bar{y})^2 = 15662170.482$; $\sum_i (\hat{y}_i - \bar{y})^2 = 6434449.424$;

- E.1 Determinare l'inclinazione della retta di minimi quadrati e interpretare il valore ottenuto (spiegare in termini reddito e spesa per viaggi)
- E.2 Determinare l'intercetta della retta di minimi quadrati e interpretare il valore ottenuto
- E.3 Calcolare l'indice di determinazione e commentare il suo valore
- E.4 L'individuo numero 30 del campione ha dichiarato un reddito di 18000 euro e una spesa di 2683.96 euro: calcolare la spesa prevista dal modello di regressione e il corrispondente residuo