

STATISTICA CORSO BASE. Prova scritta del 10-1-2018. Tempo: 2 ore

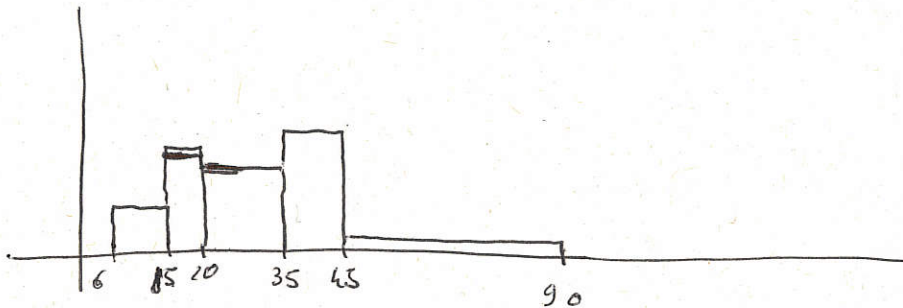
Cognome e Nome: Matricola:

Ex 1 La seguente tabella contiene la distribuzione di frequenze relative dell'età (in anni) degli utilizzatori di Internet in un determinato collettivo statistico:

classi di età	frequenze relative	\bar{x}_i	d_i	f_i/d_i	F_i
6 - 15	0.11	10,5	9	0,012	0.11
15 - 20	0.10	17,5	5	0,02	0.21
20 - 35	0.29	27,5	15	0,019	0.5
35 - 45	0.23	40	10	0,023	0.73
45 - 90	0.27	67,5	45	0,006	1
totale	1				

2 pt

1. Rappresentare graficamente, in modo opportuno, la distribuzione.



1 pt

2. Qual è la classe modale?

35-45

2 pt

3. Calcolare la media e la varianza della distribuzione dell'età degli utilizzatori di Internet?

$$\mu = \sum \bar{x}_i \cdot f_i = 38.305 \quad \sigma^2 = \sum \bar{x}_i^2 \cdot f_i - \mu^2 = 1860,253 - 1467,273 = 393$$

3 pt

4. Indicare la classe dove cade il primo quartile, riportare un valore puntuale per il terzo quartile e calcolare il valore della funzione di ripartizione $F(40)$

q_1 ~~cade nella classe 15-20~~ cade nella classe ~~15-20~~ 20-35

q_3 cade nella classe 45-90. Inoltre $q_3 = 45 + \frac{\frac{3}{4} - 0.73}{1 - 0.73} \cdot 45 = 48.33$

$F(40) = 0.5 + 0.023 \cdot (40 - 35) = 0.615$

Ex 2 Ad un campione di 200 ragazzi sotto i 30 anni è stato chiesto che tipo di sistema operativo avessero sul loro computer ottenendo la seguente distribuzione doppia relativamente ai caratteri $Y = \text{sistema operativo}$ e $X = \text{età}$

Y \ X	14-18	18-24	24-30	
Mac	25	30	42	97
Windows	50	32	21	103
Linux	2	5	8	15
	77	67	71	215

pt 1
pt 2

1. Completare la tabella precedente riportando le distribuzioni marginali dei caratteri X e Y
2. Riportare le distribuzioni relative del carattere X condizionatamente alle tre diverse modalità di Y

	MAC	WINDOWS	LINUX
14-18	0.258	0.425	0.133
18-24	0.309	0.311	0.333
24-30	0.433	0.204	0.533

pt 2

3. Stabilire sull'analisi delle sole distribuzioni condizionate se l'indice χ^2 per la misura della dipendenza debba essere positivo o nullo motivando la risposta

L'indice χ^2 risulterà positivo in quanto le distribuzioni condizionate sono diverse tra loro.

pt 3

4. Calcolare l'indice χ^2

tabella delle contingenze

	34,74	30,23	32,03
	36,89	32,10	34,01
	5,37	4,67	4,95

$$\chi^2 = \sum_{ij} \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} = 19,49$$

Ex 3 Dai dati in possesso di una compagnia telefonica risulta che i minuti di traffico in uscita effettuati dai ragazzi tra i 18 e i 24 anni in una settimana possano considerarsi come una variabile casuale di tipo Normale con media 60 e varianza 625. La stessa compagnia telefonica prevede un cambio di tariffa e quindi dei costi aggiuntivi a carico del cliente quando si superano i 100 minuti in una settimana

pt

2

1. Calcolare la probabilità che un cliente della compagnia superi i 100 minuti in una settimana

$$P(N(60, 625) > 100) = P(N(0,1) > 1,6) = 0,0548$$

2

2. Se la compagnia volesse applicare dei costi aggiuntivi solo al 3% dei suoi clienti, a quale soglia dovrebbe fissare il cambio di tariffa? Si tratta di trovare il valore $x_{0,97}$ tale che

$$0,97 = P(N(60, 625) < x_{0,97}) \text{ , Poiché } 0,97 = P(N(0,1) < \frac{x_{0,97} - 60}{25}) = \Phi(\frac{x_{0,97} - 60}{25})$$

$$= P(N(0,1) < 1,88) \quad x_{0,97} = 60 + 25 \cdot 1,88 = 107$$

2

3. Qual è la probabilità che un cliente superi per quattro settimane la soglia dei 100 minuti, assumendo che ogni settimana il suo traffico in uscita sia indipendente da quello delle altre settimane?

$$(0,0548)^4 = 0,000009$$

2

4. Dieci amici sono clienti della compagnia. Qual è la probabilità che almeno uno di essi superi la soglia dei 100 minuti?

$$P(\text{BINOMIALE}(p=0,0548, n=10) \geq 1) = 1 - \binom{10}{0} 0,0548^0 (1-0,0548)^{10}$$

$$= 1 - (1-0,0548)^{10} \approx 1 - 0,569 = 0,431$$

Ex 4 Gli amministratori comunali della cittadina di *Informatica* negli ultimi anni hanno concesso incentivi economici ai loro cittadini per l'acquisto di computers portatili ed hanno creato una rete WIFI con copertura totale sul territorio comunale. Per vedere quali siano stati gli effetti degli investimenti effettuati, decidono di selezionare un campione di utilizzatori di Internet tra i residenti a Informatica al fine di confrontare le due ipotesi:

$$H_0 : p = 0.39$$

$$H_a : p > 0.39$$

riguardanti la proporzione p di utilizzatori di Portatili con WIFI ad Informatica. Scelgono a caso $n = 100$ utilizzatori di Internet tra i residenti nel comune: 50 di loro dichiarano di collegarsi ad Internet prevalentemente attraverso un Portatile con collegamento WIFI.

pt 2 1. Quale risulta il valore della statistica test?

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.5 - 0.39}{\sqrt{\frac{0.39 \cdot 0.61}{100}}} = 2,255$$

pt 1 2. Con riferimento al punto precedente, quanto vale il corrispondente P -value?
 0.0130 0.0228 0.9861

pt 1 3. Sulla base delle informazioni raccolte, cosa concludono gli amministratori comunali al livello di significatività del 5%?
 non c'è evidenza sufficiente per rifiutare l'ipotesi nulla
 i dati forniscono evidenza sufficiente per il rifiuto dell'ipotesi nulla
 le informazioni a disposizione non sono sufficienti a prendere una decisione

pt 2 4. Quali valori del livello di significatività porterebbero ad una conclusione diversa?
 $\alpha \leq 0.01$

pt 2 5. Sulla base dei risultati del campione calcolare un intervallo di confidenza al 95% per la proporzione p di utilizzatori di portatili.

$$\left[\hat{p} \pm z_{0.975} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] = 0.5 \pm \sqrt{\frac{0.52}{100}} \times 1.96 = [0,402, 0,598]$$