

**CORSO DI STATISTICA DI BASE (Prof. GIORGIO ALLEVA)**

**Anno Accademico 2019-2020, appello straordinario**

**Prova scritta del 5 maggio 2020**

**Esercizio 1.** Su un campione di 100 unità si sia rilevata la spesa mensile delle famiglie per la salute e i servizi sanitari secondo il titolo di studio della persona di riferimento della famiglia. Sulla base delle seguenti informazioni:

Titolo di studio della persona di riferimento della famiglia	n. di osservazioni	Spesa media (euro)	Varianza della spesa
Fino alla licenza media	55	108	484
Diploma di scuola secondaria superiore	31	126	1.024
Laurea o post laurea	14	150	1.225
Totale	100		

a) Indicare l'unità di misura della varianza;

*L'unità di misura della varianza è il quadrato di quella originaria (se euro, euro al quadrato)*

b) Stimare la spesa media nella popolazione;

*Media ponderata delle medie:  $108 \times 55 + 126 \times 31 + 150 \times 14 / 100 = 119,5$*

c) Indicare quale titolo di studio presenta comparativamente una maggiore variabilità della spesa;

*E' il titolo di studio con il minore coefficiente di variazione (rapporto tra deviazione standard e media).*

*I tre Cv sono: 0,20; 0,254; 0,233; quindi il titolo con maggiore variabilità è il diploma di scuola secondaria superiore*

d) Misurare la dipendenza in media della spesa dal titolo di studio;

*Occorre calcolare il rapporto di correlazione  $r^2$  (Il rapporto tra la devianza esterna e la devianza totale).*

*La devianza esterna (devianza delle medie condizionate) =*

*$(108-119,5)^2 \times 55 + (126-119,5)^2 \times 31 + (150-119,5)^2 \times 14 = 21.606,8$*

*La devianza totale è la somma della devianza esterna ed interna.*

*La devianza interna (somma delle devianze parziali) =  $55 \times 484 + 31 \times 1.024 + 14 \times 1.225 = 75.514$*

*Pertanto  $r^2 = 21.606,8 / (21.606,8 + 75.514) = 0,22$ .*

e) Determinare l'intervallo di confidenza della spesa media delle famiglie con persona di riferimento in possesso di licenza elementare o diploma di scuola media inferiore (con  $\alpha = 0,05$ );

media campionaria = 108	t (0,025)	s/rad n	t x s/rad(n)	e-inf	e-sup
s corretto = $22 \times 55 / 54$	2,010	3,021	6,073	101,927	114,073
n = 55					

f) Stimare la proporzione di spesa delle famiglie con persona di riferimento in possesso del diploma di scuola secondaria superiore e verificare con un test statistico se nella popolazione di riferimento possa essere considerata superiore a 0,30 (con  $\alpha = 0,05$ ).

Spesa Dipl = $126 \times 31 = 3.906$	Rn=0,327	test su p
Spesa Tot = $119,5 \times 100 = 11.950$	n=31	$H_1: p > 0,30$
Proporzione = 0,327		$H_0: p_0 = 0,30$

z (0,05)	p(1-p)	p(1-p)/n	<- rad	z x rad	val cr
1,645	0,210	0,00677	0,082	0,135393	0,44

Non rifiuto, accetto  $H_0$

**Esercizio 2.** Sulla base di 48 osservazioni, le medie delle variabili X e Y siano risultate rispettivamente pari a 12 e -18.

Si disponga anche della seguente matrice di varianze e covarianze tra le due variabili.

	Y	X
Y	800	
X	-400	360

a) Stimare l'equazione della retta di regressione di Y su X;

$$B_1 = -400/360 = -1,11 \quad B_0 = -18 + 12(-1,11) = -4,67 \quad \text{e dunque } \hat{Y} = -4,67 - 1,11 X$$

b) Calcolare il coefficiente di correlazione tra le due variabili;

$$r = -400 / (800^{0,5} \times 360^{0,5}) = -0,75$$

c) Calcolare la devianza spiegata di Y.

$$Dev(Y) = 48 \times 800 = 22.627,4 \quad \text{e} \quad DevRes(Y) = r^2 Dev(Y) = 0,75 \times 22627,4 = 10.056,3$$

d) Aggiungendo una seconda variabile esplicativa, indicare in quale intervallo di valori sarà compresa la misura della bontà di adattamento.

$$0,75 \leq R^2 \leq 1$$

**Esercizio 3.** La probabilità di essere contagiati dal Coronavirus è stimata essere pari a 0,06.

Si conosca inoltre che il tasso di letalità tra i contagiati è pari a 0,138 nella popolazione con 80 o più anni di età e che nell'anno precedente all'epidemia la probabilità di morte per un individuo di 80 anni sia risultata pari a 0,038. Si determini la probabilità che un deceduto di 80 anni di età nel corso degli ultimi mesi di quest'anno sia dovuto al Covid.

Indicando con  $p(I)$ : probabilità di essere infetto e con  $p(M)$ : probabilità di morte, allora le informazioni disponibili sono:

$$p(I) = 0,06 \quad \text{e quindi } p(\text{non}I) = 0,94$$

$$p(M|I) = 0,138$$

$$p(M|\text{non}I) = 0,038.$$

$$\text{La probabilità richiesta è } p(I/M) = 0,06 \times 0,138 / (0,06 \times 0,138 + 0,94 \times 0,038) = 0,188$$

## QUESITI

**(PUNTI 2 per risposta corretta, PUNTI -1 per risposta sbagliata, PUNTI 0 per assenza di risposta)**

**Q1.** L'indice di Cramer è:

- A. Una misura di variabilità
- B. Una misura della dipendenza assoluta
- C. Una misura di dipendenza monotona

**Q2.** Nell'impostazione classica, l'errore di seconda specie in un test statistico è:

- A. rifiutare  $H_0$  quando è vera
- B. accettare  $H_0$  quando è falsa
- C. nessuna delle precedenti

**Q3.** Il prezzo di un bene sia raddoppiato nel corso di un decennio. Questo comporta che:

- A. il numero indice dei prezzi è 2
- B. l'incremento nel decennio è stato del 200%
- C. sono vere entrambe

**Q4.** Considerando il modello di regressione  $\log Y = B_0 + B_1 X$ , l'espressione del parametro  $B_1$  è:

- A.  $Cov(\log X, \log Y) / Var(\log X)$
- B.  $Cov(X, \log Y) / Var(\log X)$
- C.  $Cov(X, \log Y) / Var(X)$