Esercitazione di riepilogo

Esercizio 1

Si consideri la seguente distribuzione doppia del PUNTEGGIO ottenuto ad un test psico-attitudinale da 200 soggetti classificati per SESSO:

Voto \ Sesso	uomini	donne	Totale
6	18	8	**26**
7	28	14	**42**
8	20	36	**56**
9	18	28	**46**
10	12	18	**30**
Totale	**96**	**104**	**200**

a) Si controlli se la variabilità nella distribuzione maschile è diversa dalla variabilità in quella femminile
b) Si verifichi se il punteggio conseguito è indipendente dal sesso
c) Si confronti la media conseguita dagli uomini con quella ottenuta dalle donne

Soluzione

a) media uomini $= \mu_x = \dfrac{\sum(x)*n_i}{N} = \dfrac{(6*18)+(7*28)+...+(10*12)}{96} = 7{,}77$

media donne $= \mu_y = \dfrac{\sum(x)*n_i}{N} = \dfrac{(6*8)+(7*14)+...+(10*18)}{104} = 8{,}33$

VAR(uomini) $= \dfrac{\sum(x - \bar{x}_M)*n_i}{N-1} = \dfrac{(6-7{,}77)+(7-7{,}77)+...+(10-7{,}77)}{5} = 2{,}05$

VAR(donne) $= \dfrac{\sum(x - \bar{x}_F)*n_i}{N-1} = \dfrac{(6-8{,}33)+(7-8{,}33)+...+(10-8{,}33)}{5} = 2{,}09$

Test sulla differenza tra varianze

IPOTESI $H_0: \ \sigma_M^2 = \sigma_F^2$

$H_1: \ \sigma_M^2 \neq \sigma_F^2$

STAT TEST $\dfrac{s_M^2}{s_F^2} = \dfrac{2{,}05}{2{,}09} = 0{,}98$

VALORE CRITICO $F_\alpha, n_{M-1}; n_{F-1} = F_{0,05,95,103} = 1{,}37$

DECISIONE 1,37>0,98 Accetto H_0, le due varianze non sono significativamente diverse b)

Test d'indipendenza

tabella osservata

Voto \ Sesso	uomini	donne	Totale
6	18	8	**26**
7	28	14	**42**
8	20	36	**56**
9	18	28	**46**
10	12	18	**30**
Totale	**96**	**104**	**200**

Tabella teorica

Voto \ Sesso	uomini	donne	Totale
6	96*26/200	104*26/200	**26**
7	...	...	**42**
8	...	...	**56**
9	...	...	**46**
10	...	104*30/200	**30**
Totale	**96**	**104**	**200**

Ipotesi

$$H_0 = \sum_i \sum_j \frac{(nij - \hat{n}ij)^2}{\hat{n}ij} = 0$$

$$H_1 = \sum_i \sum_j \frac{(nij - \hat{n}ij)^2}{\hat{n}ij} > 0$$

Statistica Test = $\sum_i \sum_j \frac{(nij - \hat{n}ij)^2}{\hat{n}ij}$

Valore test (Chi quadro) = 16,2 (valore calcolato tramite la tabella di calcolo)

Valore critico = $\chi^2_{0,05,4}$ df= (r-1)*(c-1) = 4*1=4 → 9,488

Decisione Rifiuto $H_{0,}$ ossia l'ipotesi di indipendenza tra i due caratteri

c) test sulla differenza tra medie (varianza incognita)

IPOTESI $H_0 : \mu_M - \mu_F = 0$

... $H_1 : \mu_M - \mu_F \neq 0$

Statistica test: $\dfrac{\bar{x}_M - \bar{x}_F}{s\sqrt{\dfrac{1}{n_M} + \dfrac{1}{n_F}}}$ in cui $S^2 = \dfrac{(n_M - 1)s_M^2 + (n_F - 1)s_F^2}{n_M + n_F - 2}$

Valore critico $\pm t_{\alpha/2; n_M + n_F - 2} = \pm t_{0,025;198} = \pm 1{,}96$

Valore test:

$$s = \sqrt{\frac{(96-1)*1{,}68 + (104-1)*1{,}413}{198}} = \sqrt{\frac{159{,}6 + 145{,}23}{198}} = 1{,}24$$

Da cui $\dfrac{7{,}77 - 8{,}33}{1{,}24\sqrt{\dfrac{1}{96} + \dfrac{1}{104}}} = -3{,}2$

DECISIONE : Rifiuto H_0 e concludo che le medie dei due gruppi sono significativamente diverse

Esercizio n 2

Misurare la relazione esistente tra il Tempo di percorrenza (in secondi) ed il Peso (in Kg) rilevati su un campione di 6 giovani atleti impegnati in una gara di 100 metri.

PESO	TEMPO
45	15
52	18
38	16
41	13
51	17
50	19

$$\mu_x = \frac{277}{6} = 46{,}16 \qquad \mu_y = \frac{98}{6} = 16{,}33$$

$$\sigma^2{}_x = 27{,}8 \qquad \sigma^2{}_y = 3{,}9$$

$$\mu(x \cdot y) = (45 \cdot 15 + 52 \cdot 18 + ... + 50 \cdot 19 / 6) = 4569 / 6 = 761{,}5$$

$Cov_{x,y} =$ $\mu(x \cdot y) - (\mu_x \cdot \mu_y) = 761{,}5 - (46{,}16)*(16{,}33) = 7{,}7$

$Corr_{x,y} =$ $\rho_{x,y} = \frac{Cov_{x,y}}{\sigma_x \cdot \sigma_y} = \frac{7{,}7}{5{,}27 * 1{,}97} = 0{,}74 \quad \rho^2 = 0{,}74^2 = 0{,}55$

STIMA DELLA RETTA DI REGRESSIONE

$$\hat{y} = a + b_1 x_i$$

$$b = \frac{Cov_{x,y}}{\sigma^2{}_x} = \frac{7{,}7}{27{,}8} = 0{,}277$$

$$a = \overline{y} - b\overline{x} = 16{,}33 - (0{,}277 \cdot 46{,}16) = 3{,}54$$

$$\hat{y} = a + bx_i = 3{,}54 + 0{,}277 \cdot x_i$$

Verificare la validità dei risultati:

Di regola, dopo aver stimato il modello di regressione si sottopone ad ipotesi nulla:

$$H_0 = \beta = 0$$

In tal caso (β=0), il valore atteso della Y è costante e pari a per qualsiasi valore di X. Ciò implica l'assenza di un legame in media tra Y ed X e di conseguenza il modello di regressione è inutile.

Supponiamo che la mia ipotesi alternativa sia:

$$H_1 = \beta \neq 0$$

y_i	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$	$e_i^2 = (y_i - \hat{y}_i)^2$
15	16	-1	1
18	17,94	0,06	0,0036
16	14	2	4
13	14,9	-1,9	3,61
17	17,67	-0,67	0,4489
19	17,39	1,61	2,5921

$$S_b^2 = \frac{\sum e_i^2}{n-2} = \frac{11,6546}{4} = 2,9$$

La statistica test è data da:

Xtest= $\dfrac{b-\beta}{\sqrt{\dfrac{s_b}{\sum (x_i - \bar{x})^2}}}$ = $= \dfrac{0,277}{\sqrt{\dfrac{2,9}{166,83}}} = 2,1$

Essendo $t_{n-2,\alpha/2} = t_{4,0,025} = 2,776$, Accettiamo H_0 e concludiamo il peso degli atleti non incide sul tempo di percorrenza.

Per un'ulteriore conferma costruiamo un test sull' R^2 (indice di determinazione lineare).

$H_0 : R^2 = 0$ vs $H_1 : R^2 > 0$

Sotto l'ipotesi nulla la statistica test

$$Xtest = \frac{R^2(n-2)}{(1-R^2)}$$

Ha una distribuzione *F* con n-2 gradi di libertà e il valore osservato della statistica è dato da:

$$Vtest == \frac{0{,}55(4)}{1-0{,}55} = 4{,}88$$

Per n=6 la statistica test si distribuisce come una v.c $F_{\alpha,1,\ n-2}$ Al livello di significatività α=0,05, il valore critico risulta $F_{0,05,1,4} = 7{,}71$

Accetto H_0 e concludo che non c'è una relazione lineare tra le due variabili

Esercizio 3

In un esame scritto composto da 4 domande a risposta multipla, uno studente che non ha per nulla studiato intende valutare la probabilità di superare la prova. Sapendo che ciascuna domanda ha quattro opzioni possibili

a) costruire la vc "numero di risposte esatte"
b) calcolare la media e la varianza di questa distribuzione
c) calcolare la probabilità che venga bocciato, sapendo che per per superare l prova occorrono almeno tre domande giuste.

Svolgimento

n=4

p=0,25

1-p=0,75

X~Bin(4,0,25)

$$P(X = x) = \binom{n}{x} p_x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

(X=0) = $\frac{4!}{0!4!} * 0{,}25^0 (1-0{,}25)^4 = 0{,}316$

(X=1) = $\frac{4!}{1!3!} * 0{,}25^1 (1-0{,}25)^3 = 0{,}421$

(X=2) = $\frac{4!}{2!2!} * 0{,}25^2 (1-0{,}25)^2 = 0{,}21$

(X=3) = $\frac{4!}{3!1!} * 0{,}25^3 (1-0{,}25)^1 = 0{,}0468$

(X=4) = $\frac{4!}{4!0!} * 0{,}25^4 (1-0{,}25)^0 = 0{,}0039$

b) $E(X) = n \cdot p = 4 \cdot 0{,}25 = 1$

$Var(X) = n \cdot p \cdot (1 - p) = 4 \cdot 0{,}25 \cdot (1 - 0{,}25) = 0{,}75$