DOMANDA N° 3 (12 punti)

In media, i lavoratori con livello di istruzione più elevato hanno retribuzioni più elevate rispetto ai lavoratori con livello di istruzione inferiore. Usando dati provenienti dal Current Population Survey del marzo 2013, si stima la relazione tra la retribuzione oraria e il numero di anni di istruzione per un campione di 2829 lavoratori a tempo pieno negli Stati Uniti di età compresa tra 29 e 30 anni, con anni di istruzione compresi tra 6 e 18:

$$\widehat{Earnings} = \underset{(0,72)}{10,23} + \underset{(0,03)}{0,63} * YearsEducation, \qquad R^2 = 0,180 \quad SER = 10,88$$

a. Si spieghi che cosa indicano i valori 10,23 e 0,63 dei coefficienti. (*2 punti*)

b. È possibile dare un'interpretazione economica alla costante? (*1 punto*)

Il coefficiente 0,63 indica l'effetto marginale di years of education su Earning, signica che ci si aspetta un aumento di earnigs di 0,63 ogni anno in più di istruzione. 10,23 è l'intercetta della retta di regressione stimata, che indica in livello della retta di regressione, è indicherebbe il livello di retribuzione per un componente del campione con anni di istruzione pari a 0, nel nostro caso non è utile dare interpretazione economica in quanto il nostro campione è composta da individui che hanno anni di istruzione compresi tra 6 e 18.

c. Un lavoratore di 30 anni selezionato casualmente ha un livello di istruzione pari a 16 anni (ovvero è un laureato). Qual è la retribuzione oraria attesa del lavoratore? (*1 punto*)

*Earning= 10,23+0,63*16= 20,31*

d. Si costruisca un intervallo di confidenza di livello 95% per β_1, il coefficiente angolare della regressione. (*2 punti*)

[B1-1,96*SE(B1);B1+1,96*SE(B1)]

[0,63-1,96*(O,O3); 0,63+1,96*(0,03)]

0,5712<=b1<=0,6888

e. Si conduca un test d'ipotesi con alternativa bilaterale per l'ipotesi nulla $H_0: \beta_1 = 0$. Si rifiuta l'ipotesi nulla al livello di significatività del 5%? E all'1%? (*4 punti*)

H0=B1=0

Livello di significatività 5%

Tact= 0.63-0/0,03= 21>1,96 quindi si rifiuta l'ipotesi nulla di B0=0 a livello di significatività del 5%

Livello di significatività 1%

|tact|>2,58 si rifiuta l'ipotesi nulla anche a livello di significatività dell'1%

Classification : Internal

f. Cosa indicano R^2 e SER? (*2 punti*)

Rquadro e Ser sono le misure di bontà della retta di regressione stimata che mi permette di verificare se ho stimato bene la retta. Nello specifico Rquadro è la frazione di varianza di Y spiegata in X ed è data dal rapporto della somma delle deviazioni quadratiche dei valori di Y dalla loro media e la somma degli scarti quadratici medi totali degli Y dalla loro media, tale misura di bontà è indipendente dall'unità di misura a differenza di SER che è espressa nella misura della variabile dipendente e misura la dispersione delle osservazioni attorno alla retta di regressione stimata.

DOMANDA N° 2 (12 punti)

Si supponga che un ricercatore, utilizzando i dati sulla dimensione delle classi (CS) e il punteggio medio in un test per 1150 classi relative al terzo livello d'istruzione, stimi la regressione OLS,

(TestScore)ˆ = 340,63-6,52 × CS R^2= 0,48,SER = 11,85

(16,21) (2,67)

a. L'anno scorso una certa classe era composta da 20 studenti e quest'anno da 25. Qual è la predizione della regressione per quanto riguarda una variazione nel punteggio medio nei test per la classe? (3 punti)

-6.52*(25-20)= -24.65

b. Si costruisca un intervallo di confidenza di livello 99% per β_0, l'intercetta della regressione. (2 punti)

[(340.63-2.58*16.21) ; (340.63+2.58*16.21)]

298.21<=B0<=382.45

c. Si calcoli il valore-p di un test bilaterale avente come ipotesi nulla H_0:β_1=0 e come ipotesi alternativa H_1:β_1≠0. Si rifiuta l'ipotesi nulla (i) al livello di significatività del 5%? (ii) Al livello dell'1%? (3 punti)

t= -6.52-0/2.67=-2.44

(se non chiede valore p posso fare |t|>valore critico, nel caso in esame posso rifiutare l'ipotesi nulla a livello di significatività del 5% ma non dell'1% (2.44<2.58)

Valore p= 2*ϕ|-(tact)|= 2*0.0073=0.0146

Si rifiuta l'ipotesi nulla se valore p<0.05 o 0.1

Nel nostro caso si può rifiutare l'ipotesi nulla al 5% ma non all'1%

d. Si costruisca ora un intervallo di confidenza di livello 95% per β_1, la pendenza della regressione. Il risultato è in linea con le conclusioni tratte nel punto precedente? Si motivi la risposta. (2 punti)

[(-6.52-1.96*2.67);(-6.52+1.96*2.67)]

-11.75<=B1<=-1.2868

La risposta è in linea con le conclusioni che si rifiuta l'ipotesi nulla a livello di significatività del 5% in quanto nel nostro intervallo non è compreso lo 0

e. Quali sarebbero le conclusioni di un test unilaterale avente come ipotesi nulla H_0: β_1≤1,28 e come ipotesi alternativa H_1:β_1>1,28? Senza calcolare il valore-p, si motivi la risposta. (2 punti)

è corretto che non ci sia il livello di significatività?

Si rifiuta l'ipotesi nulla a favore dell'alternativa per valori grandi e negativi della statistica t.

DOMANDA N° 3 (6 punti)

Siano X e Y variabili casuali discrete con la seguente distribuzione congiunta:

		Valore di Y			
		2	**6**	**10**	**a**
Valore di X	**3**	0,08	0,16	0,05	0.29
	6	0,15	0,07	0,14	0.36
	9	0,19	0,11	0,05	0.35
a		0.42	0.34	0.24	1

Ovvero, $Pr(X = 3, Y = 6) = 0{,}16$ e così via.

a. Calcolare la distribuzione di probabilità di Y e la distribuzione di probabilità di X. (*2 punti*)
(vedi tabella)

b. Calcolare la distribuzione di probabilità di Y data $X = 9$. (*2 punti*)
0.19/0.35= 0.54
0.11/0.35=0.31
0.05/0.35=0.14

c. Si calcoli $E(Y)$ e $E(X)$. (*2 punti*)

Classification : Internal

<u>E(y)=2*0.42+6*0.34+10*0.24=5.28</u>

<u>E(x)= 3*0.29+6*0.36+9*0.35=6.18</u>

DOMANDA 1.

Il dottore di Anna è preoccupato che lei possa soffrire di diabete gestazionale (alto livello di glucosio nel sangue durante la gravidanza). È presente una certa variabilità sia nel reale livello di glucosio nel sangue, sia nei risultati del test che lo misura.

Una paziente è affetta da diabete gestazionale se il livello di glucosio, un'ora dopo aver ingerito una bevanda zuccherata, è superiore ai 140 milligrammi per decilitro (mg/dl). Il livello di glucosio di Anna varia secondo una distribuzione Normale con media $\mu = 125\, mg/dl$ e $\sigma = 10\, mg/dl$.

a) Se si fa una singola misurazione di glucosio, qual è la probabilità che ad Anna sia diagnosticato il diabete gestazionale?

Y>140 mg/dl

N(125mg, 10mg/dl)

Pr(y>140) = 140-125/10= 1.5 Y= a – 0.9332= 0.0668

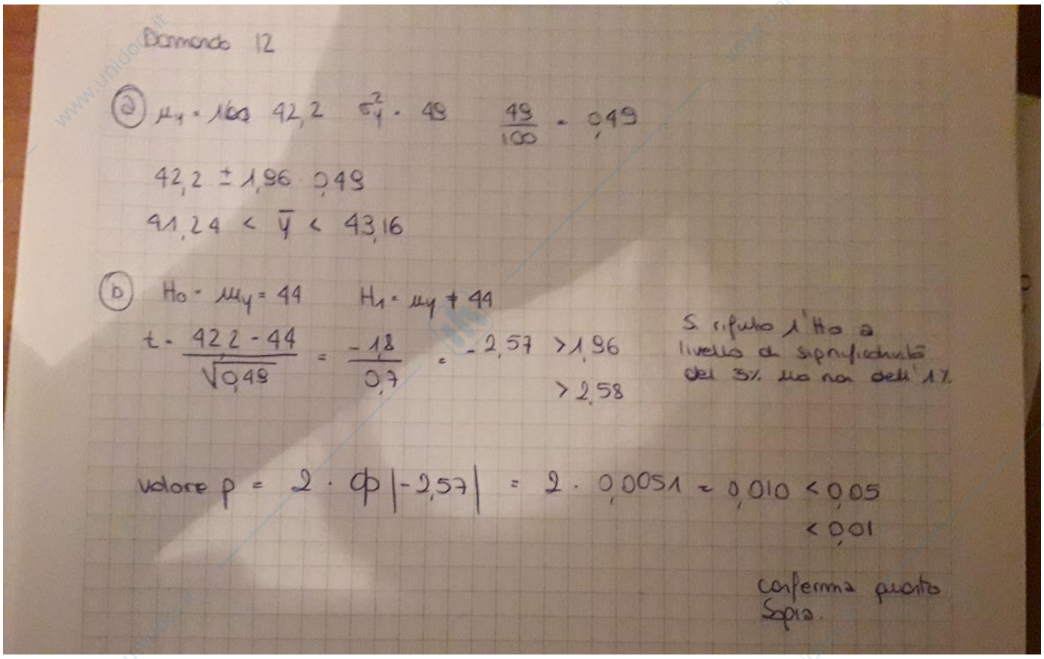
DOMANDA 12.

Dalla popolazione dei docenti universitari è stato estratto un campione casuale di 100 docenti di sesso femminile rilevandone l'età.

a) Determinare l'intervallo di confidenza a livello 95% per l'età media, sapendo che l'età media del campione delle 100 donne osservate è pari a 42.2 e che nella popolazione dei docenti di sesso femminile la variabile età presenta distribuzione Normale con varianza pari a 49;

b) Si vuole verificare l'ipotesi che l'età media sia pari a 44 anni contro l'ipotesi alternativa bilaterale. Cosa possiamo concludere a livello di significatività 0.05? E se il livello di significatività fosse 0.1?

Classification : Internal


Domanda 12

(a) $\mu_1 = 160$ 42,2 $\sigma_1^2 = 49$ $\frac{49}{100} = 0{,}49$

$42{,}2 \pm 1{,}96 \cdot 0{,}49$

$41{,}24 < \bar{y} < 43{,}16$

(b) $H_0: \mu_y = 44$ $H_1: \mu_y \neq 44$

$t = \frac{42{,}2 - 44}{\sqrt{0{,}49}} = \frac{-1{,}8}{0{,}7} = -2{,}57 > 1{,}96$

$> 2{,}58$

Si rifiuta l'Ho a livello di significatività del 5% ma non dell'1%

valore $p = 2 \cdot \Phi|-2{,}57| = 2 \cdot 0{,}0051 = 0{,}010 < 0{,}05$

$< 0{,}01$

conferma quanto sopra.

DOMANDA 9.

Si stima che il 30% degli adulti negli Stati Uniti siano obesi, che il 3% siano diabetici e che il 2% siano sia obesi che diabetici. Determina la probabilità che un individuo scelto casualmente

a) sia diabetico se è obeso;
b) sia obeso se è diabetico.

Pr(x) è la probabibilità che un individuo scelto casualmente sia OBESO

Pr(y) la probabilità che un individuo scelto casualmente sia DIABETICO

Abbiamo che Pr(x)=0,30 Pr(y) = 0,03 Pr(X,Y)= 0,02

a) Pr(y|x)= 0,02/0,03= 0,066
b) Pr(x|y)= 0,02/0,03= 0,66

DOMANDA 10.

Nel censimento del 2000 ogni persona residente negli USA doveva scegliere da un lungo elenco la propria razza. La categoria "Ispanico/latino" è un caso a parte poiché in essa vi possono essere tante razze diverse. Se scegliamo un residente negli USA in modo casuale, in base ai dati del censimento del 2000 abbiamo le seguenti probabilità:

	Ispanici	Non ispanici
Asiatici	0.000	0.036
Neri	0.003	0.121
Bianchi	0.060	0.691
Altro	0.062	0.027

a) Verifica che questa tabella di probabilità sia corretta.

La tabella risulta corretta in quanto le probabilità marginali sommano a 1

b) Quanto vale la probabilità che un americano scelto in modo casuale sia ispanico?

y=1 ispanico y=0 non ispanico

Pr(y=1)= 0,000 + 0,003 + 0,060 + 0,062= 0,68

c) I bianchi di origine non ispanica rappresentano da sempre la maggioranza di residenti negli USA. Quale è la probabilità che un americano scelto in modo casuale non sia membro di questo gruppo?

X=0 Asiatici x=1 Neri x=2 Bianchi X=3 Altro

Pr(x ≠2; y ≠ 0) = 1 – Pr(x=2 y=0) = 1-0,691= 0,309

Esercizi

3.1 In una popolazione $\mu_Y = 75$ e $\sigma_Y^2 = 45$. Si usi il teorema limite centrale per trovare:

a. $\Pr(\overline{Y} < 73)$ in un campione casuale di numerosità $n = 50$.

b. $\Pr(76 < \overline{Y} < 78)$ in un campione casuale di numerosità $n = 90$.

c. $\Pr(\overline{Y} > 68)$ in un campione casuale di numerosità $n = 120$.

3.1

$\mu_y = 75$ $\sigma_y^2 = 45$ $\sigma_{\bar{y}}^2 = \frac{45}{50} = 0,9$

nelle soluzioni viene riportato $1-\Phi(-2,10)$ che si utilizzerebbe nel caso di $Pr(\bar{y} > 73)$

ⓐ $Pr(\bar{y} < 73)$ $n = 50$

$\frac{y-75}{\sqrt{0,9}} < \frac{73-75}{\sqrt{0,9}}$ $\hat{y} < \frac{-2}{0,95}$ $y = \Phi\, -2,10$ $y = 0,0179$

ⓑ $Pr(76 < \bar{y} < 78)$ $n = 90$ $\sigma_y^2 = \frac{45}{90} = 0,5$

$Pr\left(\frac{76-75}{\sqrt{0,5}} < \frac{\bar{y}-75}{\sqrt{0,5}} < \frac{78-75}{\sqrt{0,5}}\right)$ $1,41 < \bar{y} < 4,23$

$\bar{y} = \Phi 4,23 - \Phi 1,41 = 1 - 0,9207 = 0,078$

ⓒ $Pr(\bar{y} > 68)$ $n = 120$ $45/120 = 0,38$

$\bar{y} > \frac{68-75}{\sqrt{0,38}} = \frac{-7}{0,62}$ $\bar{y} = 1 - \Phi\, -11,28$ $y = 1$

pari a 0

DOMANDA 13.

Supponiamo che la vita (in ore) di una lampadina da 75 watt sia approssimativamente normalmente distribuita ed abbia una deviazione standard pari a σ = 25 ore. Un campione di 20 lampadine ha una media (campionaria) di vita $\bar{x}$ = 1014 ore.

A) È ragionevole supporre che la media di vita delle lampadine superi 1000 ore? Usare un livello di significatività α = 5%. E a un livello di significatività dell'1%?

Esercizio 3

X è distr. con $N(\mu, 625)$ $n = 20$ $\bar{x} = 1014$

(1) $H_0 = X \leq 1000$ vs $H_1 > 1000$

$$t = \frac{1014 - 1000}{25/\sqrt{20}} = \frac{14}{5{,}59} = 2{,}50 > 1{,}64 \quad > 2{,}33$$

$$p = 1 - \Phi\, 2{,}50 = 1 - \overset{0{,}9938}{\cancel{0{,}9929}} = 0{,}0062 < 0{,}01 \quad < 0{,}05$$

Si è ragionevole pensare che $\bar{X}$ sia > 1000 in quanto ad entrambi i livelli di significatività si rifiuta l'H_0 di $X \leq 1000$.

Classification : Internal

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari