SIMULAZIONE
DI
ESAME
ESERCIZI

ESERCIZIO 1.

Alcuni autori hanno studiato se la depressione possa essere associata a indici sierologici di processi autoimmunitari o infezioni attive virali. A tal fine è stato misurato il livello di interluchina-2 (IL-2) in quattro gruppi di pazienti (i dati sono presentati nella pagina successiva).

1.1 Verificare se i livelli di interluchina 2 differiscono significativamente nei quattro gruppi sia con il metodo parametrico (1.1.a) che non parametrico (1.1.b)

1.2. Effettuare i confronti multipli con i metodi a voi noti se il confronto tra gruppi di tipo parametrico è risultato statisticamente significativo

1.3 Verificare se esiste una differenza significativa nei livelli di IL-2 tra i soggetti sani e i soggetti con depressione maggiore senza melanconia, sia con il metodo parametrico (1.3.a) che non parametrico (1.3.b)

Sani	Depressione minore	Depressione maggiore senza melanconia	Depressione maggiore con melanconia
92	634	253	242
259	305	271	283
157	324	254	354
220	250	316	517
240	306	303	292
203	369	225	439
190	428	363	444
244	324	288	348
182	655	349	230
192	395	237	255

1.1.a

92	634	253	242
259	305	271	283
157	324	254	354
220	250	316	517
240	306	303	292
203	369	225	439
190	428	363	444
244	324	288	348
182	655	349	230
192	395	237	255

n_i	10	10	10	10	N=40
$T_i = \sum_j y_{ij}$	1979	3990	2859	3404	$T^2/N=3740545.6$
$S_i = \sum_j y_{ij}^2$	413027	1765464	836939	1245408	S=4260838
T_i^2	3916441	15920100	8173881	11587216	
T_i^2 / n_i	391644.1	1592010.0	817388.1	1158721.7	$\sum \frac{T_i^2}{n_i} = 3959763{,}8$

1.1.a Calcolo delle devianze

DEVIANZA TOTALE =

$$\sum\left(y_{ij}-\bar{y}\right)^2 = S - T^2/N = 4260838 - 3740545.6 = 520292.4$$

DEVIANZA ENTRO GRUPPI =

$$\sum\left(y_{ij}-\bar{y}_i\right)^2 = S - \sum\left(T_i^2/n_i\right) = 301074.2$$

DEVIANZA TRA GRUPPI =

$$\sum\left(\bar{y}_i-\bar{y}\right)^2 = \sum\left(T_i^2/n_i\right) - T^2/N = 3959763.8 - 3740545.6$$

$$= 219218.2$$

1.1.a

Statistica test

A N O V A

Sorgenti di variazione	DEVIANZE	G.L.	VARIANZE	F_{CAL}
TRA GRUPPI	219218.2	3	73072.73	8.74
ENTRO GRUPPI	301074.2	36	8363.17	
TOTALE	520292.4	39		

Dove F=Varianza tra gruppi/varianza entro gruppi = 8.74

Poiché F cal =8.74 F tab = 2.9, rifiuto l'ipotesi nulla e concludo che almeno due gruppi differiscono tra loro

1.2.a

Confronti multipli

$$LSD = t_{\alpha/2} \sqrt{(2\ Var\ Residua/n)}$$

$t_{\alpha/2}$ Valore della t-Student con gradi di libertà N-k

$t\ \alpha/2 = 2.032$ $\quad$ $LSD = 2.032 \sqrt{2 \times 8363.17/10} = 83.10$

Medie	197.9	285.9	340.4	399.0

Diff14 = 399.0 – 197.9 = 201.1 > 83.10 ⇒ medie differenti

Diff24 = 399.0 – 285.9 = 113.1 > 83.10 ⇒ medie differenti

Diff34 = 399.0 – 340.4 = 58.6 > 83.10 ⇒ medie uguali

Diff13 = 340.4 – 197.9 = 142.5 > 83.10 ⇒ medie differenti

Diff23 = 340.4 – 285.9 = 54.5 > 83.10 ⇒ medie uguali

Diff12 = 285.9 – 197.9 = 88.0 > 83.10 ⇒ medie differenti

1.2.b

Confronti multipli

$$\left[(\bar{y}_r-\bar{y}_s)-t_{1-\alpha/2m,g.l.}\sqrt{\frac{2S^2_{residua}}{n}}\le\mu_r-\mu_s\le(\bar{y}_r-\bar{y}_s)+t_{1-\alpha/2m,g.l.}\sqrt{\frac{2S^2_{residua}}{n}}\right]$$

Medie 197.9 285.9 340.4 399.0

$\alpha/12 = 0.004$ $t = 2.7$

Int 1-4 201.1 ± 2.7 x 40.9 = 201.1 ± 110.43 ⇨ 90.67 --- 311.53

Int 2-4 113.1 ± 2.7 x 40.9 = 113.1 ± 110.43 ⇨ 2.67 --- 223.53

Int 3-4 58.6 ± 2.7 x 40.9 = 58.6 ± 110.43 ⇨ -51.83 --- 169.03

Int 1-3 142.5 ± 2.7 x 40.9 = 142.5 ± 110.43 ⇨ 32.07 --- 252.93

Int 2-3 54.5 ± 2.7 x 40.9 = 54.5 ± 110.43 ⇨ -55.93 --- 164.93

Int 1-2 88.0 ± 2.7 x 40.9 = 88.0 ± 110.43 ⇨ -22.43 --- 198.43

1.1.b

METODO NON PARAMETRICO - Kruskall Wallis

Sani	R1	Dep.Min	R2	D.mags mel	R3	D.Mag con mel	R4
92	1	634	39	253	15	242	12
259	18	305	24	271	19	283	20
157	2	324	27.5	254	16	354	31
220	7	250	14	316	26	517	38
240	11	306	25	303	23	292	22
203	6	369	33	225	8	439	36
190	4	428	35	363	32	444	37
244	13	324	27.5	288	21	348	29
182	3	655	40	349	30	230	9
192	5	395	34	237	10	255	17
	70		299		200		251

1.1.b

Statistica test e svolgimento

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

Dove:
K= numero dei campioni
nj = numero di osservazioni nel j-esimo campioni
n = numero totale delle osservazioni
Rj =somma dei ranghi nel j-esimo campioni

Per il nostro insieme di dati:

$$\boldsymbol{H} = \frac{12}{40(40+1)} \left[\frac{70^2}{10} + \frac{299^2}{10} + \frac{200^2}{10} + \frac{251^2}{10} \right] - 3(40+1) = 21.03$$

1.1.b

Decisione

Per α=0,05 e g.l= k-1=3 $\chi 2$ = 7.81

Rifiuto l'ipotesi nulla

i livelli di interluchina 2 differiscono significativamente nei quattro gruppi

1.1.b Correzione della statistica H

Poichè ci sono osservazioni con il medesimo valore (ties) bisogna correggere la statistica H

$$1-\sum T \Big/ n^3-n$$

$$T=t^3-t$$

$$H_{corr}=\frac{H}{1-\left(\frac{\sum T}{n^3-n}\right)}$$

1.1.b

Correzione della Statistica H

$T=t^3-t=2^3-2=6$

Correzione = 1 - (6/ 40^3-40) = 0,999

$$H_{corr} = \frac{H}{1-\frac{\sum T}{n^3-n}} = \frac{21.03}{0.999} = 21.05$$

Rifiuto l'ipotesi nulla

i livelli di interluchina 2 differiscono significativamente

1.3.a Confronto tra campioni indipendenti

Verificare ad un livello di significatività α=0.05 se le medie delle due popolazioni sono uguali oppure diverse:

$$\begin{cases} H_0: & \mu_1 = \mu_2 \\ H_1: & \mu_1 \neq \mu_2 \end{cases}$$

$$\bar{x} = \left(\sum x_i\right) / n$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - \left(\sum x_i\right)^2 / n}{n-1}$$

Gruppo Sani: **n=10** **$\Sigma x_i = 1979$** **$\Sigma x_i^2 = 413027$** **$\bar{x} = 197.9$** **$S^2 = 2375.88$**

Gruppo depressione Maggiore senza melanconia :

n=10 **$\Sigma x_i = 2859$** **$\Sigma x_i^2 = 836939$** **$\bar{x} = 285.9$** **$S^2 = 2172.33$**

1.3.a Confronto tra campioni indipendenti omogenità delle varianze

$$\begin{cases} H_0: & \sigma^2_1 = \sigma^2_2 \\ H_1: & \sigma^2_1 \neq \sigma^2_2 \end{cases}$$

$$F = S_1^2/S_2^2 = 2375.88/2172.33 = 1.09$$

Poiché $F_{tab} = F_{9,9,0.05} = 3.18 > F_{cal} = 1.09$ ⇒

Varianze sono omogenee

E' possibile calcolare la varianza comune

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2} = \frac{21382.9 + 19550.9}{18} = 2274.1$$

1.3.a Confronto tra campioni indipendenti – Statistica test

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

$$\boldsymbol{t_{calc}} = \frac{285.9 - 197.9}{\sqrt{\frac{2274.1}{10} + \frac{2274.1}{10}}} = 4.13$$

$t_{tab} = t_{18} = 2.10 < t_{cal} = 4.13$ ⇨ rifiuto H_0 ⇨

le medie dei due gruppi sono diverse

1.3.b Confronto tra campioni indipendenti test della somma dei ranghi

Se la distribuzione non è gaussiana utilizziamo il metodo non parametrico della Somma dei Ranghi

Gruppo Sani	92	259	157	220	240	203	190	244	182	192
Gr. Depr.Mag. senza melan.	242	283	354	517	292	439	444	348	230	255

	92	157	182	190	192	203	220	225	237	240
R	1	2	3	4	5	6	7	8	9	10

	244	253	254	295	271	288	303	316	349	363
R	11	12	13	14	15	16	17	18	19	20

Σ R sani = 63 Σ R patologici = 147

Per $\alpha = 0.05$ Intervallo dei ranghi 78 - 132

Rifiuto l'uguaglianza dei due gruppi

ESERCIZIO 2.

Sono stati utilizzati due metodi per determinare in 15 soggetti l'efficacia di un antibiotico per il trattamento della tubercolosi. I logaritmi dei titoli ottenuti con i due metodi sono i seguenti:

Metodo A	3.3	2.4	2.7	2.4	2.1	2.1	3.0	2.2	2.4	2.1	2.4	2.0	3.0	2.0	2.1
Metodo B	4.1	3.8	3.6	3.2	2.9	3.2	3.9	2.8	3.4	3.3	3.3	2.9	3.5	3.1	2.7

2.1 Determinare la retta di regressione ipotizzando che il metodo A è affetto da errore trascurabile

2.2 Verificare l'ipotesi nulla b=0 con tutti i metodi conosciuti

2.3 Calcolare il coefficiente di determinazione

2.4 Studiare la relazione esistente tra le due metodiche con il metodo parametrico (2.4.a) e non parametrico (2.4.b), ipotizzando che entrambi i metodi non siano affetti da errore trascurabile

2.5 Verificare se esiste una differenza significativa tra le due metodiche sia con il metodo parametrico (2.5.a) che non parametrico (2.5.b)

2.1 Calcoli necessari per lo svolgimento

$$\sum x_i = 36.2 \qquad \sum x_i^2 = 89.7$$

$$\sum y_i = 49.7 \qquad \sum y_i^2 = 167.05$$

$$\sum x_i y_i = 121.90$$

$$\overline{x} = 2.42$$

$$\overline{y} = 3.32$$

2.1 Determinazione dei parametri

effettuando i conti si ha:

$$\bar{x} = 36.2/15 = 2.42$$
$$\bar{y} = 49.7/15 = 3.32$$

$$\hat{b} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i y_i - \dfrac{\sum x_i \sum y_i}{n}}{\left[\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}\right]} =$$

$$= \frac{121.90 - \left(36.2 \times 49.7\right)/15}{89.7 - (36.2)^2/15} = 0.84$$

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x} = 3.32 - 0.84 \times 2.42 = 1.29$$

2.2.a Calcolo delle devianze

Per il nostro esempio

DEV. TOTALE =

$$\sum (y_i - \bar{y})^2 = \sum y_i^2 - (\sum y_i)^2 / n =$$

$$= 167.05 - 49.7^2 / 15 = 2.38$$

DEV. REGRESSIONE=

$$\sum (\hat{y}_i - \bar{y})^2 = b^2 \cdot \sum (x_i - \bar{x})^2 =$$

$$= \frac{[\sum (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum (x_i - \bar{x})^2} = 1.64$$

DEV. RESIDUA =

$$\text{TOTALE} - \text{REGRES.} = 2.38 - 1.64 = 0.74$$

2.2.a Nell'esempio

A N O V A

Sorgenti di variazione	DEVIANZE	G.L.	VARIANZE	F_{CAL}
REGRESSIONE	1.64	1	1.64	28.81
RESIDUA	0.74	13	0.06	
TOTALE	2.38	14		

Essendo $F_{tab} = F_{1,13} = 4.67 < F_{cal} = 28.81$ rifiuto H_0

Dove F_{cal}= Varianza Regressione/Varianza residua = 1.64/0.06 = 28.81

2.2.b Verifica di ipotesi su b – test t

H_0: b = 0

H_1: b ≠ 0

$$ES(b) = \sqrt{\frac{\text{var}\,res}{devx}} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2 \big/ N-2}{\sum (x_i - \bar{x})^2}} = 0.16$$

$$T = \frac{b - b_0}{ES(b)} = \frac{0.84}{0.16} = 5.25$$

Poiché T cal=5.25 > T tab=2.16 si rifiuta l'ipotesi nulla e quindi il coefficiente di regressione è significativamente diverso da zero

2.3 Coefficiente di determinazione

$$R^2 = \frac{devregr}{devtot} = \frac{\sum\left(\hat{y}_i - \bar{y}\right)^2}{\sum\left(y_i - \bar{y}\right)^2} = \frac{1.64}{2.38} = 0.69$$

La regressione spiega circa il 69% dei dati osservati

2.4.a

COEFFICIENTE DI CORRELAZIONE DI PEARSON

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} =$$

$$= \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left[\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right]\left[\sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}\right]}} =$$

$$= \frac{1.96}{2.36} = 0.83$$

2.4.a Verifica di ipotesi per l'indipendenza

IPOTESI

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

STATISTICA TEST

$$\boldsymbol{T} = \boldsymbol{r}\sqrt{\frac{\boldsymbol{n}-2}{1-\boldsymbol{r}^2}} = 0.83\sqrt{\frac{13}{1-0.83^2}} = 5.36$$

Essendo $t = 5.36 > t_{tab} = 2.160$ si rifiuta H_0

2.4.b

Metodo non parametrico-Coefficiente di correlazione di Spearman

A	B	RA	RB	d_i	di^2
3.3	4.1	15	15	0	0
2.4	3.8	9.5	13	-3.5	12.5
2.7	3.6	12	12	0	0
2.4	3.2	9.5	6.5	3	9
2.1	2.9	4.5	3.5	1	1
2.1	3.2	4.5	6.5	-2	4
3.0	3.9	13.5	14	-0.5	0.25
2.2	2.8	7	2	5	25
2.4	3.4	9.5	10	-0.5	0.25
2.1	3.3	4.5	8.5	-4	16
2.4	3.3	9.5	8.5	1	1
2.0	2.9	1.5	3.5	-2	4
3.0	3.5	13.5	11	2.5	6.25
2.0	3.1	1.5	5	-3.5	12.5
2.1	2.7	4.5	1	3.5	12.5
					104.25

2.4.b Coefficiente di correlazione di Spearman

$$r_s = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)} =$$

$$= 1 - \frac{6 \times 104.25}{15(15^2 - 1)} = 1 - \frac{625.5}{3360} = 0{,}813$$

Essendo r_s cal=0.813 > r_s tab=0.525 si rifiuta l'ipotesi nulla, i due test sono correlati

2.5.a

Si valuti l'esistenza di una differenza significativa tra i due metodi - Metodo parametrico

A	B	d_i	d_i^2
3.3	4.1	-0.8	0.64
2.4	3.8	-1.4	1.96
2.7	3.6	-0.9	0.81
2.4	3.2	-0.8	0.64
2.1	2.9	-0.8	0.64
2.1	3.2	-1.1	1.21
3.0	3.9	-0.9	0.81
2.2	2.8	-0.6	0.36
2.4	3.4	-1.0	1.0
2.1	3.3	-1.2	1.44
2.4	3.3	-0.9	0.81
2.0	2.9	-0.9	0.81
3.0	3.5	-0.5	0.25
2.0	3.1	-1.1	1.21
2.1	2.7	-0.6	0.36
Σ		-13.5	12.95

$$\begin{cases} H_0: & \mu_d = 0 \\ H_1: & \mu_d > 0 \end{cases}$$

2.5.a Confronto tra campioni appaiati test t- student

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-13.5}{15} = -0.9$$

$$S_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1} = \frac{\sum d_i^2 - \left(\sum d_i\right)^2 / n}{n-1} = \frac{12.95 - \frac{(-13.5)^2}{15}}{15-1} = 0.0571$$

$$t = \frac{\bar{d} - \mu_d}{S_d / \sqrt{n}} \qquad t_{calc} = \frac{-0.9}{\sqrt{\frac{0.0571}{15}}} = -14.59$$

Poiché t tab =2.16 < t cal = 14.59, rifiuto l'ipotesi nulla, le due metodiche sono significativamente differenti

2.5.b Si valuti l'esistenza di una differenza significativa tra i due metodi - Metodo non parametrico – Ranghi con segno

A	B	d_i	Rdif
3.3	4.1	-0.8	-5
2.4	3.8	-1.4	-15
2.7	3.6	-0.9	-8.5
2.4	3.2	-0.8	-5
2.1	2.9	-0.8	-5
2.1	3.2	-1.1	-12.5
3.0	3.9	-0.9	-8.5
2.2	2.8	-0.6	-2.5
2.4	3.4	-1.0	-11
2.1	3.3	-1.2	-14
2.4	3.3	-0.9	-8.5
2.0	2.9	-0.9	-8.5
3.0	3.5	-0.5	-1
2.0	3.1	-1.1	-12.5
2.1	2.7	-0.6	-2.5

Σ R- = 120

Intervallo tabulato
25 ---- 95

I due metodi sono significativamente differenti

ESERCIZIO 3.

Al fine di valutare la relazione tra radioesposizione e patologie tiroidee in una popolazione di lavoratori ospedalieri, sono stati reclutati 304 dipendenti radioesposti e 383 dipendenti non radioesposti. I risultati sono esposti nella tabella:

	Rx Esp.	Non Rx Esp.	Totale
Noduli	53	120	173
Tiroidite	14	23	37
Sani	237	240	477
Totale	304	383	687

3.1 Si valuti l'esistenza di una relazione tra esposizione e patologia tiroidea.

3.2 Limitando l'attenzione ai soggetti con patologia nodulare e sani si verifichi con tutti i metodi conosciuti la significatività della relazione e si determini l'odds ratio (3.3) e il suo intervallo di confidenza (3.3.a).

3.1

Tabella dei valori attesi

	Rx Esp.	Non Rx Esp.	Totale
Noduli	76,55	96,45	173,00
Tiroidite	16,37	20,63	37,00
Sani	211,07	265,93	477,00
Totale	304,00	383,00	687,00

$$E_{ij} = n_i * n_j / N = 304*173/687=76.55$$

$$G.l.=(r-1)*(c-1)=2$$

3.1

STATISTICA TEST

$$X^2 = \sum_{ij} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} =$$

$$= \frac{(53 - 76.55)^2}{76.55} + \ldots + \frac{(240 - 265.93)^2}{265.93} =$$

=7.25+0.34+3.18+5.75+0.27+2.53=19.33

Poiché χ^2 cal=19.33 < χ^2 tab=5.99
si conclude che c'è legame tra esposizione e patologia tiroidea

3.2.a

Confronto della proporzione di patologia nodulare tra esposti e non esposti – tabella 2 x 2.

	Rx Esp.	Non Rx Esp.	Totale
Noduli	53	120	173
Sani	237	240	477
Totale	290	360	650

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{650\cdot(53\cdot 240-120\cdot 237)^2}{173\cdot 477\cdot 290\cdot 360} = 18.64$$

Poiché χ^2 cal=18.64 < χ^2 tab=3.84
si conclude che c'è legame tra esposizione e patologia nodulare tiroidea

3.2.b Confronto della proporzione di patologia nodulare tra esposti e non esposti – test z.

n_1=290 n_2=360

x_1=53 x_2=120

p_1=0.18 p_2=0.33

$$p = \frac{53+120}{650} = 0.26$$

$$q = 1 - p = 0.74$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p \cdot q\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.18 - 0.33 - 0}{\sqrt{0.26 \cdot 0.74\left(\frac{1}{290} + \frac{1}{360}\right)}} = -4.32$$

Essendo z=4.32 >1.96 si rifiuta l'ipotesi nulla ⇨ le due proporzioni sono differenti

3.3.a Determinazione dell'Odds Ratio

OR = ad / bc = 53*240 / 120*237 = 12720 / 28440 = 0.45

La radioesposizione non è fattore di rischio per la patologia nodulare tiroidea

3.3.b Intervallo di confidenza dell'Odds Ratio

z = 1.96 χ^2 = preso dal test di significatività (vedi pag. 35)

$$OR^{\left[1 \pm \left(z / \sqrt{\chi^2} \right)\right]} = 0.313 – 0.646$$