Indicazioni Ris. 26/01/2017

Es 1 (tutti). Si considerino i seguenti dati

-1.7 -1.9 14.1 2.9 1.9 15.8 -8.0 6.9

1) Si stimino un indice di dispersione e uno di posizione nel caso in cui i dati derivino da una distribuzione gaussiana.
2) Si stimino un indice di dispersione e uno di posizione nel caso in cui i dati non derivino da una distribuzione gaussiana
3) Si stimi un parametro che possa essere utilizzato per distinguere il dato da uno gaussiano

I primi due punti sono relativamente semplici. L'unico punto che vorrei segnalare è il 3. In questo caso può essere utile un parametro che per dati gaussiani è solitamente nullo. Sono adatti curtosi e skewness. Altrimenti qualsiasi parametro che descrive la non simmetria del dato

Es 2 (tutti)

Si confrontano due dispositivi utilizzati per il riempimento di contenitori. I responsabili del processo dubitano che le due macchine abbiano le stesse prestazioni. Vengono eseguite le seguenti misure, espresse in grammi, sui contenitori riempiti dai due dispositivi.

Dispositivo 1 16.03 16.01 16.04 15.96 16.05 15.98 16.05 16.02 16.02 15.99

Dispositivo 2 16.02 16.03 15.97 16.04 16.02 15.96 16.01 16.01 15.99 16.00

Eseguire un test per verificare i dubbi dei responsabili del processo $(\alpha = 0.05)$
Inoltre si esegua una stima dell'intervallo di confidenza al 95% della differenza tra le medie delle misure dei due dispositivi.

Abbiamo detto che i dati provengono da una distribuzione gaussiana. Questo ci permette di utilizzare un test per la media di tipo parametrico. In questo caso si utilizza un test t per dati non appaiati. È sbagliato ad esempio fare la differenza di ogni coppia di misure: tali misure non sono appaiate in nessun modo. In questo caso è importante definire l'ipotesi alternativa bilatera e quindi il t critico è pari a t 0.025 con 18 gdl pari quindi a 2.101. La regione di accettazione si estende quindi da -2.101 a 2.101. Va inoltre stimata la varianza combinando le due varianze (pooled). Il t che si ottiene è pari a 0.7989.

L'intervallo di confidenza si ottiene utilizzando un approccio standard e considerando la differenza tra le medie campionarie come base di partenza per la ricerca dell'intervallo (ovvero l'intervallo di confidenza sarà centrato attorno a tale valore campionario).

Es 3 (tutti)
In un test randomizzato, a doppio cieco, si misura il livello cognitivo in soggetti con MCI (Mild Cognitive Impairment). Si distingue un gruppo di controllo che assume placebo e un secondo gruppo che assume Ginkgo Biloba. Nella tabella sono mostrati il numero dei soggetti nei due gruppi raggruppati in funzione delle variazioni osservate. Alla fine dello studio possiamo dire che c'è stato un effetto del gingko-biloba sulle funzioni cognitive rispetto al placebo $(\alpha = 0.05)$?

	Variazione indice Cognitivo (variazione negativa==miglioramento)				
	-4	da -3 a -2	da -1 a 1	Da 2 a 3	Da 3 a 4
Placebo	10	11	19	11	24
Gingko Biloba	22	18	12	7	16

Possiamo dire che le funzioni cognitive sono migliorate con l'acquisizione del gingko biloba? Motivare.

Si utilizza il test del chi quadro. Vediamo la tabella delle frequenze attese nel caso in cui non ci sia una differenza nella distribuzione tra le due popolazioni. In grassetto le frequenze in questo caso.

	Variazione indice Cognitivo (variazione negativa==miglioramento)					
	-4	*da -3 a -2*	*da -1 a 1*	*Da 2 a 3*	*Da 3 a 4*	
Placebo	*10* ***16***	*11* ***14.5***	*19* ***15.5***	*11* ***9***	*24* ***20***	*75*
Gingko Biloba	*22* ***16***	*18* ***14.5***	*12* ***15.5***	*7* ***9***	*16* ***20***	*75*
	32	*29*	*31*	*18*	*40*	*150*

$$\chi^2 = \frac{(10-16)^2}{16} + \frac{(22-16)^2}{16} + \frac{(11-14.5)^2}{14.5} + \frac{(18-14.5)^2}{14.5} + \frac{(19-15.5)^2}{15.5} + \frac{(12-15.5)^2}{15.5} + \frac{(11-9)^2}{9} + \frac{(7-9)^2}{9} + \frac{(24-20)^2}{20} + \frac{(16-20)^2}{20} = 10.77$$

dalla tabella del χ^2 con 4 gdl si ottiene il valore critico 9.488. Quindi possiamo affermare che c'è un effetto del gingko biloba.

Si deve notare che tale test non ci dice se c'è stato un miglioramento o meno, ma solo che le distribuzioni delle frequenze sono differenti. Per vedere una direzionalità dovremmo creare una tabella 2x2.

Es 4.1
In uno studio relativo all'analisi del segnale vocale e stati emotivi è stato stimato un parametro (Ampl*) che misura l'ampiezza delle variazioni di intonazione. Sono state valutate 18 frasi, identiche, pronunciate ognuna da un attore differente che ha simulato una tra le seguenti emozioni: rabbia, noia, felicità e stato neutro. I valori del parametro in ciascuna frase, raggruppati per emozione simulata, sono presentati nella seguente tabella. È possibile dire che esiste una variazione significativa del parametro studiato al variare dell'emozione simulata ($\alpha = 0.05$)?

Rabbia	Stato Neutro	Noia	Felicità
1.2	-0.4	-0.3	1.4
0.3	0.5	0.4	1.6
0.9	0.4	-0.7	0.7
0.7	-0.2	-1	0.2
0.1			0.1

Non è stata data nessuna info sulla gaussianità dei dati. D'altronde abbiamo poche osservazioni e anche un test di normalità non sarebbe significativo. Quindi usiamo un test non parametrico.
Siamo nel caso di dati indipendenti, possiamo usare un test di Kruskal Wallis. Distribuiamo i ranghi su tutte le osservazioni.

Rabbia	*Stato Neutro*	*Noia*	*Felicità*
1.2 ***16***	*-0.4* ***3***	*-0.3* ***4***	*1.4* ***17***
0.3 ***9***	*0.5* ***12***	*0.4* ***10.5***	*1.6* ***18***
0.9 ***15***	*0.4* ***10.5***	*-0.7* ***2***	*0.7* ***13.5***
0.7 ***13.5***	*-0.2* ***5***	*-1* ***1***	*0.2* ***8***
0.1 ***6.5***			*0.1* ***6.5***

R1=12 R2=7.625 R3=4.375 R4=12.6
Rtot=9.5
D=5(12-9.5)^2+4*(7.625-9.5)^2+4*(4.375-9.5)^2+5*(12.6-9.5)^2=198.425 H=12/(18*19)*D=6.96*
Il valore del chi quadro critico (siamo nel caso di alta numerosità) con 3 gdl abbiamo 7.815
Si accetta ipotesi nulla di assenza di effetto dello stato emotivo sul parametro osservato.

Es 4.2

In uno studio relativo all'analisi della qualità dell'aria viene monitorato in 6 le concentrazioni (in g/m^3) di biossido di zolfo. Le misure medie giornaliere vengono confrontate su giorni. Ci chiediamo se esita una differenza significativa tra i diversi quartieri($\alpha = 0.05$)

	Quartiere					
Giorno	A	B	C	D	E	F
1	55	75	32	60	36	48
2	57	70	69	65	41	52
3	60	24	43	68	98	54
4	52	85	51	58	33	64
5	57	69	72	56	28	53

Le misurazioni dei diversi giorni non possono essere considerate indipendenti. Il test più adatto è il test di Friedman.

	Quartiere					
Giorno	*A*	*B*	*C*	*D*	*E*	*F*
1	*55 4*	*75 6*	*32 1*	*60 5*	*36 2*	*48 3*
2	*57 3*	*70 6*	*69 5*	*65 4*	*41 1*	*52 2*
3	*60 4*	*24 1*	*43 2*	*68 5*	*98 6*	*54 3*
4	*52 3*	*85 6*	*51 2*	*58 4*	*33 1*	*64 5*
5	*57 4*	*69 5*	*72 6*	*56 3*	*28 1*	*53 2*
	R1=18	*R2=24*	*R3=16*	*R4=21*	*R5=11*	*R6=15*

Abbiamo quindi 5 soggetti (giorni) e 6 trattamenti (quartieri). La statistica vale 6.02 e il chi quadro critico con 5 gdl è 11.07. Si accetta l'ipotesi nulla di uguaglianza dei quartieri.

Es 5.1

Si vuole verificare se 4 diverse temperature di cottura abbiano o meno effetto sulla densità dei mattoni prodotti. Si faccia l'ipotesi che le misure derivino da variabili supposte gaussiane e si utilizzi una significatività pari a ($\alpha = 0.05$).

Trattamento	Densità				
T1	21.8	21.9	21.7	21.6	21.7
T2	21.7	21.4	21.5	21.4	
T3	21.9	21.8	21.8	21.6	21.5
T4	21.9	21.7	21.8	21.4	

Test anova.
Le misure (densità) sono raggruppate in funzione della temperatura. Quindi si hanno 4 gruppi.
Si ottiene un F pari a 2.024. il valore critico di F (3,14)=3.11. Quindi si accetta l'ipotesi nulla di non effetto della temperatura sulla densità dei mattoni.

Es 5.2

Si vuole verificare se 4 diverse temperature di cottura abbiano o meno effetto sulla densità dei mattoni prodotti. Si faccia l'ipotesi che le misure derivino da variabili supposte gaussiane e si utilizzi una significatività pari a ($\alpha = 0.05$). Le misure sono effettuate su gruppi a pari numerosità, uguale a 6.

	T1	T2	T3	T4
Media	21.7	21.5	22.0	23.0
Varianza	0.11	0.14	0.11	0.09

Test anova. Abbiamo un F pari a 23.64. Il valore dello F critico F(3,20)=3.1. Si rifiuta l'ipotesi nulla di uguaglianza dell'effetto dei trattamenti sulla densità.

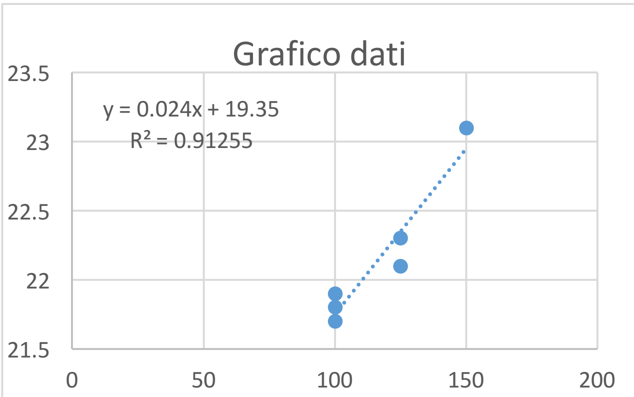
Es 6 (2016-2017)
Nella tabella seguente sono mostrate le densità di mattoni prodotti a due diverse temperature. Si stimi il modello di regressione che lega densità alla temperatura e se ne verifichi la significatività ($\alpha = 0.05$).

Temperatura	Densità		
100	21.8	21.9	21.7
125	22.1	22.3	
150	23.1	22.7	

Variabile dipendente: densità. Variabile indipendente: temperatura.

OUTPUT RIEPILOGO

Statistica della regressione	
R multiplo	*0.950577581*
R al quadrato	***0.903597737***
R al quadrato corretto	*0.884317284*
Errore standard	***0.174220955***
Osservazioni	*7*



ANALISI VARIANZA

	gdl	*SQ*	*MQ*	*F*	*Significatività F*
Regressione	*1*	*1.422521008*	*1.422521008*	***46.86600221***	*0.001015471*
Residuo	*5*	*0.151764706*	*0.030352941*		
Totale	*6*	*1.574285714*			

	Coefficienti	*Errore standard*	*Stat t*	*Valore di significatività*
Intercetta	*19.6*	*0.389569899*	*50.31189543*	*5.86292E-08*
Variabile X 1	***0.021647059***	*0.003162059*	***6.845874832***	***0.001015471***

D2 (tutti)
Si voglia analizzare la variabilità cardiaca in funzione dello umore del soggetto. Si supponga di avere due gruppi di 5 soggetti ciascuno: il primo gruppo è in condizione di normalità, mentre il secondo presenta dei sintomi depressivi. Si fanno per ogni soggetto 10 misure in un giorno. Supponendo gaussiana la distribuzione delle misure, si commenti se è corretto usare un test t per confrontare i 50 valori ottenuti in corrispondenza dello stato normale, con i 50 ottenuti nello stato depresso.

Non è possibile usare un t test: le misure sono ripetute e quindi dipendenti tra loro. (è stato valutato esclusivamente questo aspetto).

Altri dettagli

Si potrebbe:
1)fare una media delle misure in ciascun soggetto
2)utilizzare un test che prevede l'uso di misure ripetute
3) utilizzare metodi "gerarchici" (non incluso nel programma)