

Statistica Applicata
Corso di Laurea in Scienze Naturali
a. a. 2017/2018

prof. Federico Plazzi

22 Gennaio 2018

Nome: _____

Cognome: _____

Matricola: _____

Alcune indicazioni:

- La prova è costituita da cinque esercizi; dopo ogni esercizio c'è lo spazio in cui scrivere la risposta o le risposte. In caso questo spazio non sia sufficiente, si può continuare a rispondere sul retro del foglio, avendo cura di indicare il numero dell'esercizio a fianco della continuazione della risposta.
- Alcuni esercizi richiedono semplici calcoli, per i quali è consentito l'uso di una calcolatrice ed eventualmente la consultazione di una o più delle tabelle allegate.
- Altri esercizi richiedono invece la lettura dei dati: verrà valutata in questo caso l'argomentazione che giustifica l'interpretazione fornita.
- La durata massima della prova è di 60 minuti.
- Si prega di non scrivere nulla sulle tabelle allegate.

1 Dati

Si sa dalla letteratura che una certa malattia è correlata a un diverso tasso di *metilazione* del genoma: gli individui malati sarebbero caratterizzati da una percentuale più alta di posizioni del genoma connesse a una proteina, chiamata istone H3, in cui, in una precisa posizione, un apposito enzima ha aggiunto tre gruppi metile. Per verificare questa ipotesi, si eseguono una serie di esperimenti in cui l'istone metilato viene legato da un anticorpo specifico, che dà una certa colorazione che può essere letta da uno spettrofotometro. La tabella 1 mostra i risultati dell'esperimento: nella prima colonna ci sono le letture effettuate allo spettrofotometro su campioni di controllo, il cui livello di metilazione è noto; nelle altre quattro colonne sono elencate le letture effettuate su due diversi gruppi di pazienti, A e B, divisi in sani e malati. In sostanza, a un'assorbanza più alta letta allo spettrofotometro corrispondono livelli più alti di metilazione, espressi in $\text{ng}/\mu\text{L}$.

Tabella 1: Assorbanza misurata dallo spettrofotometro per i controlli e per gli individui sani e malati di due gruppi di pazienti.

Controllo		Gruppo A		Gruppo B	
Conc.	Ass.	Sani	Malati	Sani	Malati
100 $\text{ng}/\mu\text{L}$	0,816	0,132	0,066	0,081	0,080
50 $\text{ng}/\mu\text{L}$	0,672	0,426	0,757	1,463	1,790
25 $\text{ng}/\mu\text{L}$	1,249	0,061	0,054	0,078	0,072
12 $\text{ng}/\mu\text{L}$	0,398	0,564	0,717	0,792	0,823
6 $\text{ng}/\mu\text{L}$	0,187	0,096	0,063	0,235	0,078
3 $\text{ng}/\mu\text{L}$	0,092	0,457	0,120	1,281	0,303
1,5 $\text{ng}/\mu\text{L}$	0,067	0,047	0,376	0,272	0,206
0 $\text{ng}/\mu\text{L}$	0,071	0,302	0,846	2,065	0,741

2 Esercizi

2.1 Statistiche di base

Calcolare media, devianza, varianza e deviazione standard dell'assorbanza media misurata negli individui malati del gruppo A.

<i>Media</i>	<i>Devianza</i>	<i>Varianza</i>	<i>Deviazione standard</i>
0,375	0,846	0,121	0,348

2.2 Curva di controllo

Per prima cosa è necessario verificare che la calibrazione dello strumento sia corretta: in altre parole, l'assorbanza nella seconda colonna deve aumentare linearmente con la concentrazione indicata nella prima colonna. Per verificare questo, si esegue un test di correlazione lineare, di cui sono mostrati di seguito i risultati. La correlazione è verificata?

Tabella 2: Correlazione lineare tra concentrazione di istone H3 trimetilato e assorbanza.

	Stima	p-value	r	R^2
Intercetta	0,140035			
Pendenza	0,007668	0,00222	0,9321973	0,8689918

Esiste una correlazione significativa (il p-value è infatti minore di 0,05) tra le due variabili. Tale correlazione è positiva (il coefficiente r è positivo) e in grado di spiegare quasi tutta la varianza come covarianza ($R^2 = 0,87$).

2.3 Distribuzione dell'assorbanza

Si esegue a questo punto un test di Shapiro e Wilk su tutti i dati delle quattro colonne riguardanti gli individui sotto studio; cosa possiamo concludere sulla distribuzione dell'assorbanza? Perché?

Shapiro-Wilk test

```
data: absorbance.data
W = 0.7896, p-value = 2.614e-05
```

La distribuzione dei dati di assorbanza non appare normale, visto che possiamo rifiutare l'ipotesi nulla di normalità del test ($p = 2,614 \times 10^{-5}$).

2.4 Confronto tra sani e malati

Si vogliono ora confrontare in ogni gruppo gli individui sani con quelli malati, per stabilire se effettivamente l'assorbanza è significativamente maggiore per i malati in entrambi i casi. Di seguito sono riportati i risultati di quattro possibili test statistici per ciascuno dei due confronti: qual è l'approccio corretto? Perché? Cosa è possibile concludere?

Tabella 3: Test t a campioni indipendenti

	Gruppo A	Gruppo B
t	-0,80272	0,80425
df	11,282	13,345
p-value	0,4387	0,4353

Tabella 4: Test t a campioni appaiati

	Gruppo A	Gruppo B
t	-1,151	1,354
df	7	7
p-value	0,2875	0,2178

Tabella 5: Test di Mann e Whitney

	Gruppo A	Gruppo B
W	27	40,5
p-value	0,6454	0,4005

Tabella 6: Test di Wilcoxon

	Gruppo A	Gruppo B
V	13	27
p-value	0,5469	0,25

Dato che la variabile non è a distribuzione normale e che non c'è ragione per ritenere i campioni appaiati (i sani e i malati di ogni gruppo sono individui diversi), l'approccio corretto è il test di Mann e Whitney; essendo maggiore di 0,05, il p-value non è significativo, per cui non si può rifiutare l'ipotesi nulla: l'idea che i livelli di metilazione negli individui malati siano più elevati non è verificata in nessuno dei due gruppi.

2.5 Incidenza della malattia

Sebbene ne siano stati sottoposti a esperimento soltanto 8 sani e 8 malati per gruppo, le osservazioni sono state ben più ampie. In particolare, il gruppo A era composto di 53 individui, di cui 26 erano sani e i restanti malati; il gruppo B era composto di 49 individui, di cui 13 sani e i restanti malati. Cosa si può dire a riguardo della proporzione di malati nei due gruppi? Ci sono significativamente più malati nel gruppo B? Quale test bisogna eseguire? Qual è il risultato?

L'approccio corretto per le variabili qualitative è il test del χ^2 . Si costruisce una tabella 2×2 con un solo grado di libertà; il valore di χ^2 risulta 4,5583, che è significativo al 5%: il p-value esatto sarebbe 0,03276. Si può perciò concludere che nel gruppo B c'è una quota di malati significativamente maggiore.