

Statistica Applicata
Corso di Laurea in Scienze Naturali
a. a. 2018/2019

prof. Federico Plazzi

22 Gennaio 2019

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

La tabella 1 riporta le temperature mensili di due siti marini del Mar Mediterraneo, soggetti perciò a una variazione stagionale più o meno intensa delle temperature. Per ogni sito, è stata valutata anche la copertura macroalgale in termini di biomassa secca nell'unità di superficie.

Tabella 1: Temperature e biomasse algali in due siti del Mar Mediterraneo.

Mese	Sito 1		Sito 2	
	T (°C)	Biomassa (g/100 cm ²)	T (°C)	Biomassa (g/100 cm ²)
Gennaio	13.78	8.46	13.25	21.15
Febbraio	14.91	9.73	15.35	22.18
Marzo	17.35	10.56	18.01	22.62
Aprile	20.58	12.75	20.42	23.40
Maggio	20.64	14.74	20.61	28.54
Giugno	22.82	23.19	23.80	29.18
Luglio	23.93	24.49	30.79	26.90
Agosto	20.61	16.80	26.75	26.75
Settembre	17.19	11.94	21.61	25.93
Ottobre	17.03	10.64	20.85	25.46
Novembre	15.63	9.80	19.72	19.79
Dicembre	14.61	7.43	16.46	18.17

2 Esercizi

2.1 Statistiche di base

Calcolare media, devianza, varianza e deviazione standard della temperatura registrata nel sito 1.

<i>Media</i>	<i>Devianza</i>	<i>Varianza</i>	<i>Deviazione standard</i>
18,26	124,53	10,38	3,22

2.2 Distribuzione dei dati

Per prima cosa, è necessario indagare la distribuzione dei dati. A questo scopo viene eseguito il test di Shapiro e Wilk su tutte le 24 temperature e su tutte le 24 stime di biomassa. I risultati sono indicati in tabella 2. Che cosa se ne può concludere?

Tabella 2: Test di Shapiro e Wilk.

Variabile	W	p-value
Temperatura	0,946	0,2216
Biomassa	0,91766	0,05186

Entrambe le variabili sono a distribuzione normale. Il p-value, infatti, risulta in entrambi i casi (sebbene di molto poco nel caso della biomassa) superiore a 0,05, il che non consente di rifiutare l'ipotesi nulla di normalità.

2.3 Differenze in temperatura

I due siti sono stati scelti nel Mediterraneo perché dovrebbero presentare temperature significativamente diverse. In particolare, il sito 1 dovrebbe essere più freddo.

Si rende quindi necessario effettuare un test per verificare se effettivamente il sito 1 abbia temperature significativamente inferiori al sito 2 nel corso dell'anno. Quale dei seguenti è il test corretto? Perché? Cosa se ne può concludere?

- Test t a due code a campioni indipendenti:

Two Sample t-test

$t = -1.3949$, $df = 19.587$, $p\text{-value} = 0.1787$

alternative hypothesis: true difference in means is not equal to 0

- Test t a una coda a campioni indipendenti:

Two Sample t-test

$t = -1.3949$, $df = 19.587$, $p\text{-value} = 0.08933$

alternative hypothesis: true difference in means is less than 0

- Test t a due code a campioni appaiati:

Paired t-test

$t = -3.1969$, $df = 11$, $p\text{-value} = 0.008502$

alternative hypothesis: true difference in means is not equal to 0

- Test t a una coda a campioni appaiati:

Paired t-test

$t = -3.1969$, $df = 11$, $p\text{-value} = 0.004251$

alternative hypothesis: true difference in means is less than 0

- Test di Mann e Whitney a due code:
Wilcoxon rank sum test

W = 52.5, p-value = 0.2726
alternative hypothesis: true location shift is not equal to 0
- Test di Mann e Whitney a una coda:
Wilcoxon rank sum test

W = 52.5, p-value = 0.1363
alternative hypothesis: true location shift is less than 0
- Test di Wilcoxon a due code:
Wilcoxon signed rank test

V = 7, p-value = 0.009277
alternative hypothesis: true location shift is not equal to 0
- Test di Wilcoxon a una coda:
Wilcoxon signed rank test

V = 7, p-value = 0.004639
alternative hypothesis: true location shift is less than 0

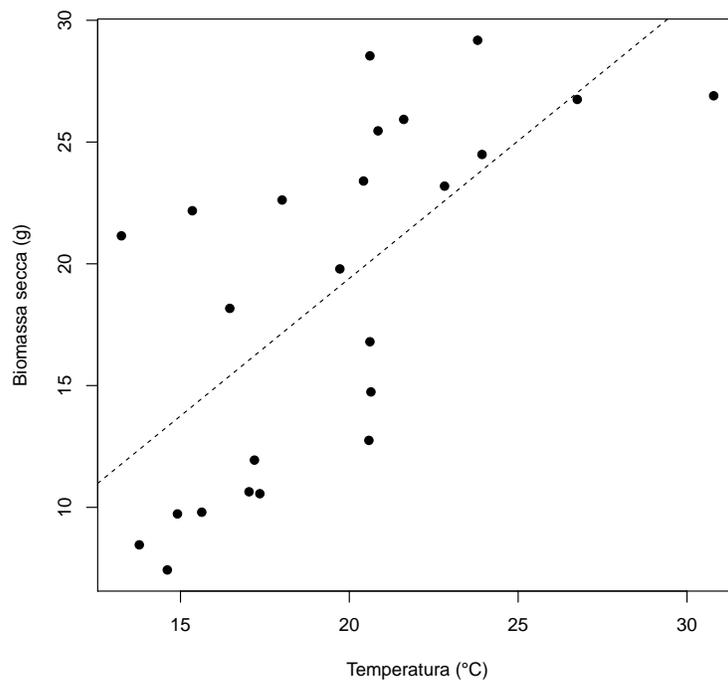
La temperatura è a distribuzione normale e i campioni sono appaiati, perché la temperatura alle latitudini temperate varia in modo stagionale, quindi è opportuno confrontare la temperatura di ogni mese del sito 1 con la temperatura allo stesso mese del sito 2. Il test corretto è perciò il test t a campioni appaiati. Siccome si vuole sapere se il sito 1 è più freddo del sito 2, la scelta migliore sarebbe il test a una coda (la quarta opzione). Siccome il p-value è $< 0,05$ (0,004251) il test è significativo e possiamo rifiutare l'ipotesi nulla: i due campioni non vengono dalla stessa popolazione e la temperatura al sito 1 è effettivamente più bassa di quella al sito 2.

2.4 Legame tra temperatura e biomassa

La figura 2.4 mostra l'andamento della biomassa in funzione della temperatura nei due siti; la linea tratteggiata è la linea di regressione tracciata secondo i parametri del modello di correlazione lineare mostrati in tabella 3. In base alla figura 2.4 e alla tabella 3, cosa si può dire del rapporto tra temperatura e biomassa macroalgale?

Tabella 3: Regressione lineare tra temperatura e biomassa macroalgale

	Stima	p-value	r	R^2
Intercetta	-3,1915			
Pendenza	1,1296	0,000298	0,6748134	0,4554



Nonostante dalla figura emerge una certa debolezza della correlazione, evidenziata da un R^2 modesto (0,4554) la correlazione è presente e significativa: il p-value è infatti inferiore a 0,05 (0,000298). Si può quindi concludere che a temperature maggiori sono correlate biomasse macroalgali maggiori.

2.5 Temperature massime

Date alcune condizioni dell'ecosistema, si intendono per mesi caldi quelli che hanno fatto registrare una temperatura strettamente maggiore di 20°C. Quale approccio statistico si può usare per determinare se il numero di mesi caldi è significativamente diverso nei due siti? Qual è il risultato? Quale conclusione si può trarre?

L'approccio statistico corretto è il χ^2 , perché si ha a che fare con una variabile qualitativa: i 12 mesi vengono ripartiti in caldi e non caldi. Si compila una tabella di contingenza 2×2 che risulta: In base a questa tabella si calcolano le frequenze

	Mesi non caldi	Mesi caldi
Sito 1	7	5
Sito 2	5	7

attese usando i totali di riga e di colonna e il valore di χ^2 , che risulta 0,16667. Il valore soglia al 5% che possiamo trovare in tabella è 3,84; dato che il risultato ottenuto è molto inferiore, non è significativo (il p-value preciso sarebbe 0,6831) e si può rifiutare l'ipotesi nulla: sebbene il sito 2 sia significativamente più caldo del sito 1, non c'è differenza nel numero di mesi caldi nei due siti.