

**Statistica Applicata**  
**Corso di Laurea in Scienze Naturali**  
**a. a. 2016/2017**

prof. Federico Plazzi

14 Giugno 2017

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

In Tabella 1 sono mostrati i dati relativi alle misurazioni effettuate su un gruppo di 20 individui della specie *Testudo hermanni* (Testudines: Cryptodira). Gli individui considerati provengono tutti dalla stessa zona e sono stati catturati e marcati in diversi momenti dell'anno, in modo da poterli riconoscere in seguito.

Tabella 1: Misurazioni relative a 20 esemplari di *T. hermanni*.

| Numero dell'esemplare | Peso (g) | Lunghezza (mm) |
|-----------------------|----------|----------------|
| 13                    | 904      | 170            |
| 27                    | 521.5    | 134.5          |
| 43                    | 1168     | 180.5          |
| 70                    | 860.5    | 168            |
| 81                    | 116      | 80             |
| 82                    | 727      | 152            |
| 112                   | 622      | 141.5          |
| 116                   | 1322     | 192            |
| 210                   | 305      | 109            |
| 221B                  | 155      | 88             |
| 228                   | 930.5    | 169            |
| 241                   | 1233     | 184            |
| 514                   | 962.5    | 172.5          |
| 699                   | 1031.5   | 175            |
| 779                   | 1296     | 192.3          |
| 785                   | 578.5    | 136.5          |
| 799                   | 908      | 163            |
| 811                   | 913      | 172.5          |
| 827                   | 1004.5   | 170            |
| 899                   | 585.5    | 141            |

# 2 Esercizi

## 2.1 Statistiche di base

Calcolare media, devianza, varianza e deviazione standard del peso dei 20 esemplari.

$mean(PES) = 807,2$ ;  $SS(PES) = 2356688$ ;

$variance(PES) = 117834,4$ ;  $standard.deviation(PES) = 343,2702$

## 2.2 Distribuzione dei risultati

La figura 1 mostra, in alto, la distribuzione di lunghezza e peso dei 20 esemplari e, in basso, i relativi Q-Q Plot. Dati i risultati del test di Shapiro e Wilk per le due variabili, quale di essi è compatibile con i grafici relativi alla lunghezza e quale, invece, è compatibile con i grafici relativi al peso?

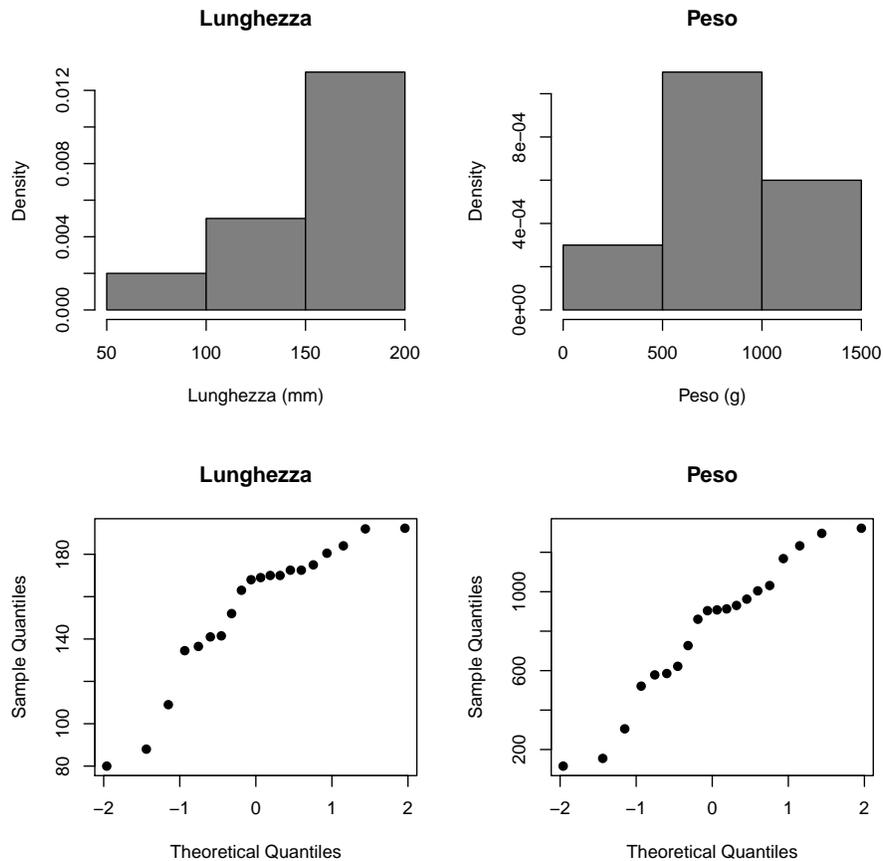


Figura 1: Istogrammi e Q-Q Plot.

- Test di Shapiro e Wilk #1:  
Shapiro-Wilk test  
 $W = 0.87674$ ,  $p\text{-value} = 0.01546$
- Test di Shapiro e Wilk #2:  
Shapiro-Wilk test  
 $W = 0.94662$ ,  $p\text{-value} = 0.3186$

*Il primo test si riferisce alla lunghezza, che non ha una distribuzione normale, come si vede molto bene dai grafici, in particolare dall'istogramma. Il secondo test si riferisce invece al peso, che ha una distribuzione normale: da notare, infatti, che nel Q-Q Plot i punti seguono una disposizione molto più simile a una retta che non nel Q-Q Plot relativo alla lunghezza dell'animale.*

### 2.3 Correlazione lunghezza-peso

Si vuole verificare l'attesa correlazione tra lunghezza e peso di un esemplare. La tabella 2 mostra i dati ottenuti. Cosa si può dire riguardo alla previsione? È stata verificata oppure no?

Tabella 2: Correlazione tra lunghezza e peso.

|            | Stima    | p-value                 | $r$       | $R^2$     |
|------------|----------|-------------------------|-----------|-----------|
| Intercetta | -856,452 |                         |           |           |
| Pendenza   | 10,763   | $5,504 \times 10^{-14}$ | 0,9795553 | 0,9595286 |

*La correlazione è positiva ( $r = 0,9795553$ ) e significativa ( $p \lll 0,05$ ): come previsto, al crescere della lunghezza dell'animale ne aumenta il peso ( $R^2 = 0,9595286$ ).*

### 2.4 Nuove osservazioni: l'individuo 302

Viene osservato un nuovo esemplare di *T. hermanni*, quello marcato come "302". Il suo peso è molto piccolo, anche se non il minore in assoluto: l'individuo 302 pesa infatti 121,3 g, mentre l'individuo 81, già registrato, pesa appena 116 g. Possiamo sostenere che l'individuo 302 sia significativamente più piccolo (leggero) degli altri 20? Come si può rispondere a questa domanda? Perché?

*L'individuo 302 è in effetti più piccolo (leggero) degli altri 20. Siccome il peso è a distribuzione normale (vedi sopra), possiamo eseguire un test Z. La media dei pesi dei 20 esemplari è 807,2 g; la deviazione standard 343,2702. La deviatore normale, quindi, è pari a*

$$Z = \frac{X - \mu}{\sigma} = -1,998134 \quad (1)$$

*che in tabella corrisponde a un valore maggiore di 0,9767. Questo significa che il valore è significativamente basso, anche assumendo un test a due code, anziché a una coda come sottinteso nella formulazione della domanda.*

## 2.5 Nuove osservazioni: nella zona dell'individuo 302

Dopo l'individuo 302, vengono catturati altri 17 individui nella stessa zona, tutti abbastanza leggeri rispetto ai 20 iniziali. Si vogliono confrontare i due campioni per sapere se fanno parte della stessa popolazione o meno: di sotto sono elencati i risultati del test di Shapiro e Wilk sui 18 nuovi individui (compreso il 302) e i risultati di quattro diversi test statistici a una coda per il confronto.

### 1. Test di Shapiro e Wilk:

Shapiro-Wilk test

```
data: Nuove osservazioni
W = 0.95228, p-value = 0.4619
```

### 2. Test possibili:

- Test t a due campioni indipendenti:

Two Sample t-test

```
data: Nuove osservazioni and Peso
t = -8.6498, df = 19.098, p-value = 2.474e-08
alternative hypothesis: true difference in means is less
than 0
```

- Test t a due campioni appaiati:

Paired t-test

```
data: Nuove osservazioni and Peso
t = -10.245, df = 17, p-value = 5.42e-09
alternative hypothesis: true difference in means is less
than 0
```

- Test di Mann e Whitney:

Mann-Whitney rank sum test

```
data: Nuove osservazioni and Peso
W = 13, p-value = 1.111e-08
alternative hypothesis: true location shift is less than
0
```

- Test di Wilcoxon:

Wilcoxon signed rank test

data: Nuove osservazioni and Peso

V = 0, p-value = 3.815e-06

alternative hypothesis: true location shift is less than  
0

Qual è il test più indicato? Qual è il risultato del confronto?

*Entrambi i campioni sono a distribuzione normale, per cui l'approccio corretto è il test  $t$  a due campioni indipendenti che, restituendo un  $p$ -value molto basso ( $p = 2,474 \times 10^{-8}$ ), è altamente significativo: i due campioni appartengono a popolazioni diverse e, siccome il test è stato eseguito a una coda, quella a cui appartiene l'individuo 302 ha dimensioni minori.*