

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

prof. Federico Plazzi

27 Settembre 2017

Nome: \_\_\_\_\_

Cognome: \_\_\_\_\_

Matricola: \_\_\_\_\_

**Alcune indicazioni:**

- La prova è costituita da quattro 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.

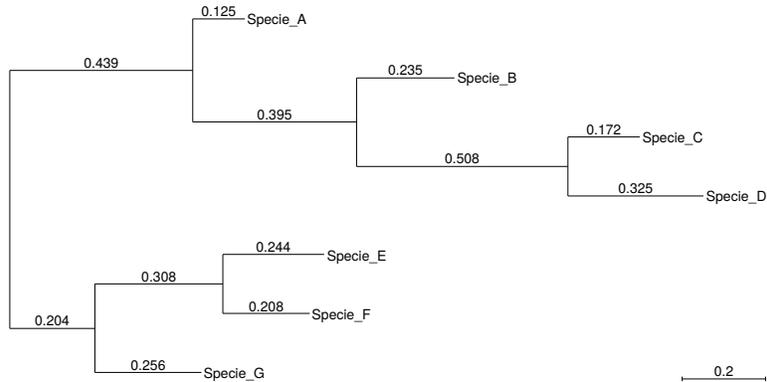


Figura 1: Albero evolutivo di sette specie (lunghezze dei rami)

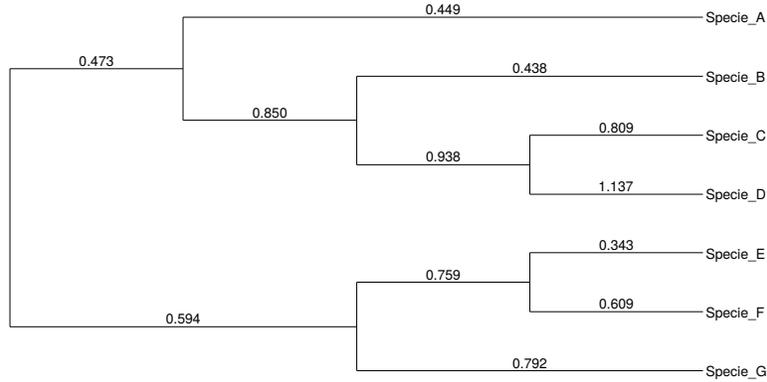


Figura 2: Albero evolutivo di sette specie ( $\omega$ )

## 1 Dati

A seguito di uno studio di biologia evolutiva molecolare, è stato ottenuto l'albero mostrato in Figura 1; le sette specie coinvolte sono indicate con le lettere da A a G. Le lunghezze dei rami indicano il numero di sostituzioni attese per nucleotide nelle basi di DNA analizzate (la scala di riferimento è riportata in basso a destra): in buona sostanza, più è lungo un ramo, più mutazioni sono avvenute durante l'evoluzione che ha portato al gruppo a cui conduce il ramo. Si possono calcolare anche altri parametri, tra cui il cosiddetto "rapporto  $\omega$ ": esso dà un'indicazione sull'intensità e la modalità di selezione naturale. Un valore di  $\omega$  maggiore di 1 indica una selezione direzionale, un valore molto più piccolo di 1 (per esempio, 0,2) indica invece selezione stabilizzante. I valori di  $\omega$  sono indicati, per lo stesso albero, in Figura 2, dove le lunghezze dei rami sono state fissate arbitrariamente soltanto in modo da allineare i nomi delle specie.

## 2 Esercizi

### 2.1 Statistiche di base

Calcola media, varianza e deviazione standard delle lunghezze dei rami mostrate in Figura 1.

```
mean(branch.lengths)
0.2849905
variance(branch.lengths)
0.011964
standard.deviation(branch.lengths)
0.1093801
```

### 2.2 Distribuzione dei parametri relativi ai rami

Di seguito sono mostrati i risultati del test di Shapiro e Wilk eseguito sulle lunghezze dei rami mostrate in Figura 1 e sui valori di  $\omega$  mostrati in Figura 2. Interpreta e commenta i risultati.

Shapiro-Wilk test

```
data: branch.lengths
W = 0.94842, p-value = 0.6139
```

Shapiro-Wilk test

```
data: omegas
W = 0.9584, p-value = 0.7608
```

*Le variabili sono tutte e due a distribuzione normale, visto che entrambi i test hanno un p-value superiore a 0,05.*

### 2.3 Il gruppo (C+D)

Le specie C e D sono strettamente imparentate dal punto di vista evolutivo e sono entrambe specie con caratteristiche particolari rispetto alle altre 5. Il ramo che porta al gruppo (C+D) è il più lungo dell'albero (vedi Fig. 1) e sarebbe interessante capire se è significativamente più lungo degli altri, il che permetterebbe di affermare che l'evoluzione di queste due specie ha richiesto significativamente più mutazioni che negli altri casi.

Quale procedura è indicata in questo caso? Cosa ne risulta?

*Trattandosi di una variabile a distribuzione normale (come risulta dall'esercizio precedente, si può applicare il test Z calcolando la devziata normale.*

$$Z = \frac{0,508 - \mu}{\sigma} \quad (1)$$

*La lunghezza del ramo è infatti 0,508; si ottiene per Z un valore di 2,03885. Usando la tabella del test Z si ottiene un p-value compreso tra 0,9788 e 0,9793 (il risultato esatto, calcolato con R, sarebbe 0,9792675), per cui il test risulta significativo sia a una coda sia a due code: il ramo che porta al gruppo (C+D) ha una lunghezza significativamente maggiore degli altri.*

### 2.4 Correlazione tra lunghezza dei rami e $\omega$

La tabella 1 mostra i risultati ottenuti dal calcolo di un modello di correlazione lineare tra lunghezza dei rami e  $\omega$ . Si può affermare che i rami più lunghi sono caratterizzati da una selezione direzionale, ossia da valori più elevati di  $\omega$ ?

Tabella 1: Correlazione tra lunghezza dei rami e  $\omega$ .

	Stima	p-value	$r$	$R^2$
Intercetta	0,4403			
Pendenza	0,8502	0,1871	0,4087263	0,1670571

*Non esiste alcuna correlazione tra lunghezza dei rami e  $\omega$ : il valore di  $r$  non è significativo ( $p = 0,1871$ ). Di conseguenza, non c'è alcun legame tra il carico di mutazione (la lunghezza dei rami) e le modalità di selezione (stabilizzante o direzionale, indicate dal valore di  $\omega$ ).*