

Statistica Applicata

prof. Federico Plazzi

Corso di Laurea in Scienze Naturali

a. a. 2015/2016

PROVA DEL 8 FEBBRAIO 2016

Nome: _____

Cognome: _____

ALCUNE INDICAZIONI:

- La prova consiste in 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.

1 Descrizione dei dati

Le proteine possono svolgere diverse funzioni (“Tipo”): alcune hanno una funzione enzimatica (“Enzima”), altre sono proteine “strutturali”, altre ancora funzionano da trasportatori di informazioni, come certi ormoni (“Segnale”).

Le proteine sono costituite da catene di aminoacidi più o meno lunghe, le quali poi si ripiegano per formare la struttura definitiva. Alcuni aminoacidi tendono a disporsi sulla superficie della proteina (“Esterni”), altri all’interno (“Interni”), mentre altri non hanno tendenze evidenti (“Ambivalenti”).

In tabella 1 è riportato, per ogni proteina prodotta da un certo gene, il numero di aminoacidi per ogni categoria e la lunghezza totale della catena.

2 Dati

Tabella 1: Conteggio degli aminoacidi per tipo e lunghezza totale di 25 proteine.

| Gene | Tipo | Esterni | Interni | Ambivalenti | Lunghezza |
|----------|-------------|---------|---------|-------------|-----------|
| Gene 700 | Enzima | 20 | 87 | 81 | 188 |
| Gene 443 | Enzima | 101 | 226 | 247 | 574 |
| Gene 532 | Enzima | 10 | 53 | 28 | 91 |
| Gene 884 | Enzima | 33 | 177 | 146 | 356 |
| Gene 205 | Enzima | 76 | 285 | 213 | 574 |
| Gene 211 | Enzima | 40 | 106 | 113 | 259 |
| Gene 544 | Enzima | 24 | 118 | 93 | 235 |
| Gene 654 | Enzima | 14 | 46 | 39 | 99 |
| Gene 799 | Strutturale | 33 | 83 | 73 | 189 |
| Gene 320 | Strutturale | 79 | 168 | 151 | 398 |
| Gene 804 | Strutturale | 52 | 210 | 173 | 435 |
| Gene 492 | Strutturale | 83 | 220 | 220 | 523 |
| Gene 126 | Strutturale | 42 | 109 | 108 | 259 |
| Gene 312 | Strutturale | 48 | 88 | 90 | 226 |
| Gene 313 | Strutturale | 135 | 159 | 167 | 461 |
| Gene 801 | Strutturale | 19 | 78 | 66 | 163 |
| Gene 230 | Strutturale | 11 | 57 | 30 | 98 |
| Gene 369 | Segnale | 83 | 243 | 249 | 575 |
| Gene 398 | Segnale | 47 | 88 | 91 | 226 |
| Gene 755 | Segnale | 12 | 50 | 52 | 114 |
| Gene 408 | Segnale | 30 | 180 | 142 | 352 |
| Gene 394 | Segnale | 84 | 222 | 250 | 556 |
| Gene 364 | Segnale | 37 | 152 | 126 | 315 |
| Gene 651 | Segnale | 84 | 284 | 226 | 594 |
| Gene 484 | Segnale | 17 | 62 | 39 | 118 |

3 Esercizi

3.1 Statistiche di base

Calcolare devianza, varianza e deviazione standard delle lunghezze delle proteine segnale.

3.2 Distribuzione dei dati

Per avere qualche indicazione sulla distribuzione delle lunghezze di tutte e 25 le proteine, si costruisce il QQ-Plot.

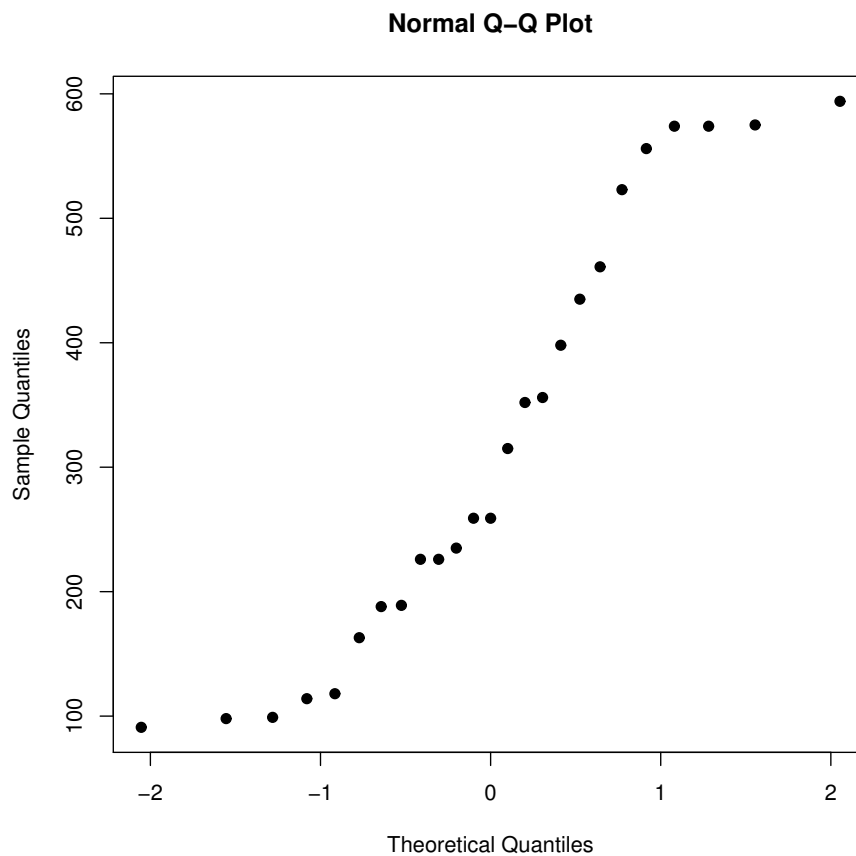


Figura 1: QQ-Plot delle lunghezze delle 25 proteine.

Cosa si può concludere sulla distribuzione dei dati? Quale dei seguenti potrebbe essere il risultato di un eventuale test di Shapiro e Wilk? Perché?

1. Shapiro-Wilk test

data: Lunghezza
W = 0.9895, p-value = 0.12008

2. Shapiro-Wilk test

data: Lunghezza
W = 0.9054, p-value = 0.02412

3. Shapiro-Wilk test

data: Lunghezza
W = 0.3694, p-value = 0.85013

3.3 Correlazioni

Secondo un'ipotesi, il numero di aminoacidi interni influenza la lunghezza della proteina: più aminoacidi ci sono che tendono a disporre all'interno, più la catena si deve allungare per contenerli. Di seguito sono riportati alcuni dati ottenuti dal calcolo di un modello di correlazione lineare tra numero di aminoacidi interni e lunghezza totale della proteina. L'ipotesi è confermata o smentita? Perché?

Tabella 2: Correlazione lineare tra aminoacidi interni e lunghezza della proteina.

| | Stima | p-value | r | R^2 |
|------------|--------|-----------|--------|--------|
| Intercetta | 7.6212 | | | |
| Pendenza | 0.4212 | 3.684e-16 | 0.9731 | 0.9469 |

3.4 Confronti tra diversi tipi di proteine

Può essere interessante capire se i diversi tipi di proteine hanno lunghezza diverse. Qui sotto sono riportati i risultati di tre possibili diversi approcci.

1. Una possibilità è di effettuare un test t di appartenenza di ogni gruppo alla popolazione generale.

Tabella 3: Risultati del test t di appartenenza di ogni campione alla popolazione generale.

| Campione | t | p-value |
|-------------|---------|---------|
| Enzima | -0.3273 | 0.753 |
| Strutturale | 0.5266 | 0.615 |
| Segnale | -0.2652 | 0.798 |

2. Una seconda possibilità è l'effettuazione di tutti i possibili confronti a coppie usando un test di Mann e Whitney.

Tabella 4: Risultati del test di Mann e Whitney su tutte le possibili coppie di campioni.

| Primo campione | Secondo campione | W | p-value |
|----------------|------------------|------|---------|
| Enzima | Strutturale | 25 | 0.4945 |
| Enzima | Segnale | 34.5 | 0.9232 |
| Strutturale | Segnale | 42.5 | 0.5635 |

3. La terza possibilità è la One-Way ANOVA.

Tabella 5: One-Way ANOVA.

| | g.l. | Devianza | Varianza | F | p-value |
|--------------|------|-----------|----------|--------|---------|
| <i>entro</i> | 22 | 716387.06 | 32563.05 | | |
| <i>tra</i> | 2 | 16545.58 | 8272.79 | 0.2541 | 0.7779 |

Cosa si può concludere? Ci sono differenze in lunghezza tra i diversi tipi di proteine o no? Quale test è il più appropriato per dirlo?