Linguistica Applicata

*Esercitazione di R per linguisti*

(A/A 2023-24)

## Terzo appello estivo del 15 luglio 2024

**Esercizio 1**

Caricate il file *esercizio\_1.txt*, che contiene vari dati riguardanti la realizzazione del genitivo inglese con *of* o con *s*:

* costruite la distribuzione di frequenza univariata e bivariata delle due variabili GENITIVE e POSSESSOR;
* rappresentate graficamente il rapporto tra le due variabili GENITIVE e POSSESSOR;
* dimostrate se esiste un’associazione statisticamente significativa tra queste variabili.

**Esercizio 2**

Caricate il file *esercizio\_2.txt*, che contiene i valori della frequenza fondamentale F1 raccolti da un gruppo si soggetti maschi e femmine:

* rappresentate graficamente la distribuzione della variabile HZ\_F1 rispetto ai due gruppi di SEX (M e F)
* calcolate i valori di media, deviazione standard e range interquartile della variabile HZ\_F1 per i due gruppi
* dimostrate se esiste una differenza statisticamente significativa nel valore di F1 nei due gruppi di soggetti.

**Svolgimento Esercizio 1:**

Per prima cosa imposto la cartella di lavoro con **setwd()** e successivamente carico esercizio\_1.txt con il seguente comando:

es1 <- read.table('esercizio\_1.txt', header = T, fileEncoding= 'UTF-8', stringsAsFactors=T)

**Costruisco la distribuzione di frequenza univariata e bivariata delle due variabili GENITIVE e POSSESSOR:**

**Frequenza assoluta di POSSESSOR:**

possAss <- table(es1$POSSESSOR)

possAss

abstract animate concrete

139 118 43

**Frequenza relativa di POSSESSOR:**

> poss\_rel <- prop.table(table(es1$POSSESSOR))

> poss\_rel

abstract animate concrete

0.4633333 0.3933333 0.1433333

**Frequenza assoluta di GENITIVE:**

genAss <- table(es1$GENITIVE)

genAss

of s

150 150

**Frequenza relativa GENITIVE:**

> gen\_rel <-prop.table(table(es1$GENITIVE))

> gen\_rel

of s

0.5 0.5

**Costruisco la distribuzione di frequenza bivariata delle due variabili GENITIVE e POSSESSOR:**

freq\_bivar <- table(es1$POSSESSOR, es1$GENITIVE)

freq\_bivar

of s

abstract 92 47

animate 16 102

concrete 42 1

**Frequenza relativa bivariata:**

> prop.table(freq\_bivar)

of s

abstract 0.306666667 0.156666667

animate 0.053333333 0.340000000

concrete 0.140000000 0.003333333

**Rappresento adesso graficamente il rapporto tra le due variabili GENITIVE e POSSESSOR:**

Poiché le variabili in questo caso non sono numeriche posso utilizzare il barplot oppure il mosaic plot.

> mosaicplot(GENITIVE~POSSESSOR, data = es1, main = "rapporto tra GENITIVE e POSSESSOR")

Immagine che contiene testo, schermata, diagramma, Rettangolo

Descrizione generata automaticamente

**Dimostro se esiste un’associazione statisticamente significativa tra queste variabili:**

Poiché si tratta di variabili categoriali per verificare se esiste un’associazione statisticamente significativa utilizzo **il Pearson’s Chi-Squared test**.

Per poter applicare ciò utilizzo la frequenza assoluta bivariata che mi sono ricavato precedentemente.

> chisq.test(freq\_bivar)

Pearson's Chi-squared test

data: freq\_bivar

X-squared = 116.34, df = 2, **p-value < 2.2e-16**

**Osservo che il p-value ottenuto è inferiore al valore soglia di 0.05** dunque possiamo affermare che esiste una associazione statisticamente significativa e non casuale tra queste variabili poiché posso rigettare l’ipotesi nulla di indipendenza.

**Svolgimento Esercizio 2:**

Per prima cosa carico il dataframe con il seguente comando:

es2 <- read.table('esercizio\_2.txt', header = T, fileEncoding= 'UTF-8', stringsAsFactors=T)

**rappresentate graficamente la distribuzione della variabile HZ\_F1 rispetto ai due gruppi di SEX (M e F):**

Per rappresentare graficamente la distribuzione della variabile HZ\_F1 rispetto ai due gruppi di SEX ovvero M e F scelgo di utilizzare un boxplot eseguendo il seguente comando:

> boxplot(HZ\_F1~SEX, data = es2, main = "distribuzione di HZ\_F1 rispetto ai due gruppi di SEX (M e F)")

Ottengo:

Immagine che contiene testo, schermata, diagramma, Rettangolo

Descrizione generata automaticamente

**calcolate i valori di media, deviazione standard e range interquartile della variabile HZ\_F1 per i due gruppi:**

Calcolo **la media** utilizzando la funzione **mean()** per i due gruppi

> mean(es2$HZ\_F1[es2$SEX == 'M'])

[1] 484.274

> mean(es2$HZ\_F1[es2$SEX == 'F'])

[1] 528.8548

Calcolo la **deviazione standard** utilizzando la funzione **sd():**

> sd(es2$HZ\_F1[es2$SEX == 'M'])

[1] 87.90112

> sd(es2$HZ\_F1[es2$SEX == 'F'])

[1] 110.801

Calcolo il **range interquartile** utilizzando la funzione **IQR()**:

> IQR(es2$HZ\_F1[es2$SEX == 'M'])

[1] 124.915

> IQR(es2$HZ\_F1[es2$SEX == 'F'])

[1] 160.875

**dimostrate se esiste una differenza statisticamente significativa nel valore di F1 nei due gruppi di soggetti.**

Per fare tale dimostrazione è necessario utilizzare il **t.test**, tuttavia, prima è necessario verificare se i valori sono distribuiti normalmente e per fare questo utilizzo il test di Shapiro-Wilk:

Preparo i due gruppi:

> Maschi <- es2[es2$SEX == 'M', "HZ\_F1"]

> Femmine <- es2[es2$SEX == 'F', "HZ\_F1"]

A questo punto applico lo Shapiro.test sui due gruppi:

> shapiro.test(Maschi)

Shapiro-Wilk normality test

data: Maschi

W = 0.97239, p-value = 0.1907

> shapiro.test(Femmine)

Shapiro-Wilk normality test

data: Femmine

W = 0.98699, p-value = 0.7723

Osservo che in entrambi i casi il p-value è superiore al valore soglia di 0.05 dunque posso procedere eseguendo il t-test:

> t.test(Maschi, Femmine)

Welch Two Sample t-test

data: Maschi and Femmine

t = -2.4416, df = 112.19, **p-value = 0.01619**

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

95 percent confidence interval:

-80.758016 -8.403651

sample estimates:

mean of x mean of y

484.2740 528.8548

In base al risultato del p-value, inferiore al valore soglia di 0.05, possiamo concludere che esiste una differenza statisticamente significati nel valore di F1 nei due gruppi di soggetti.