Esame di Laboratorio di Statistica

Prova di laboratorio

Tavano Matteo

24/01/2024

```
#Import librerie
library(readr)

## Warning: replacing previous import 'lifecycle::last_warnings' by
## 'rlang::last_warnings' when loading 'hms'

## Warning: replacing previous import 'lifecycle::last_warnings' by
## 'rlang::last_warnings' when loading 'tibble'

## Warning: replacing previous import 'lifecycle::last_warnings' by
## 'rlang::last_warnings' when loading 'pillar'

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(tidyr)
library(ggplot2)
library(ggthemes)
library(moments)

# Usiamo readr per leggere il dataset
Studenti = read.csv("StudentGPA.csv")
#View(Studenti)
```

Introduzione

- Effetuiamo il test su una variabile è qualitativa (sex) e l'altra quantitativa (GPA): analisi di dipendenza in media;

Analisi variabile qualitativa

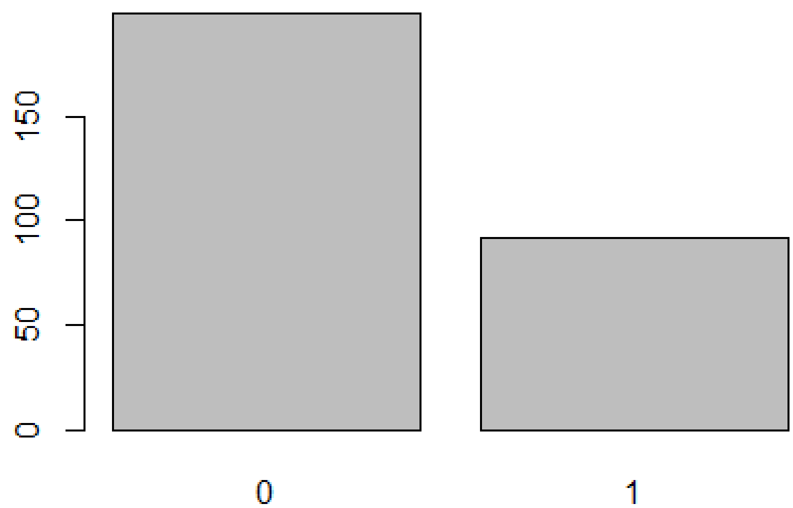
- Otteniamo prima di tutto le frequenze assolute assunte dalle due modalità.

- Ci aiutiamo con la funzione summary.

- Mostriamo le frequenze attraverso un barplot.

```
summary(as.factor(Studenti$sex)) # frequenze

##   0   1
## 199  92
```

```
barplot(table(as.factor(Studenti$sex))) # barplot frequenze
```



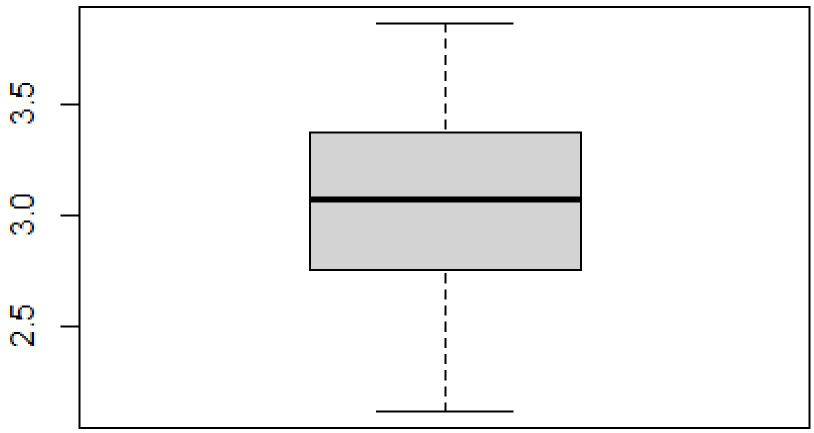
Risultati

- Evinciamo che ci sono più femmine rispetto ai maschi in questo Dataset.

Analisi variabile quantitativa

- Possiamo osservare info sulle differenze tra media e mediana.
- Utilizziamo un boxplot, al fine di evidenziare la mediana e i quartili.

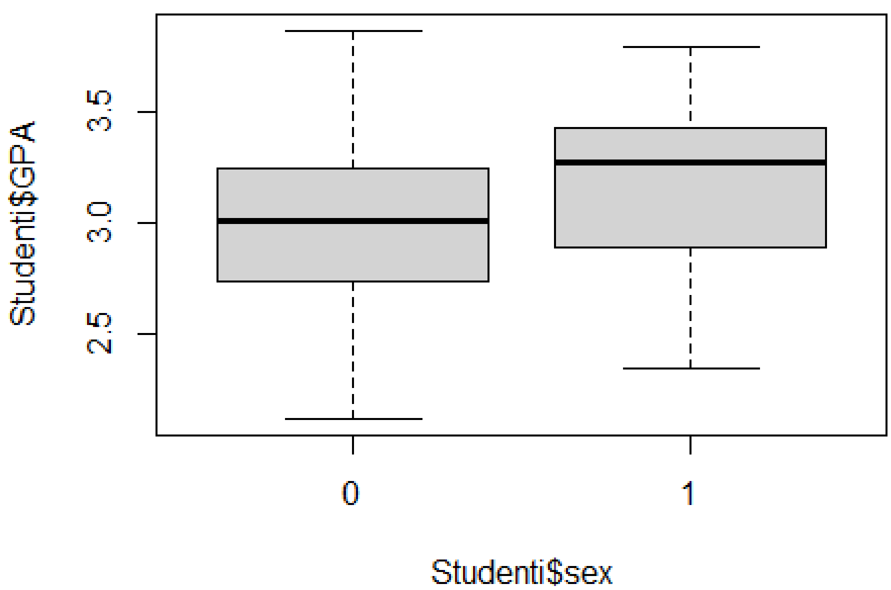
```
summary(Studenti$GPA)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.120   2.760   3.080   3.048   3.380   3.870

boxplot(Studenti$GPA) #boxplot
```



Dividere nei vari casi qualitativi la variabile quantitativa

- Costruiamo due boxplot, divisi a seconda delle modalità 0,1 assunti dalla variabile sex.

```
boxplot(Studenti$GPA~Studenti$sex)
```



```
#Dividere nei vari casi qualitativi la variabile quantitativa
```

Simmetria, mediana, media, quartili

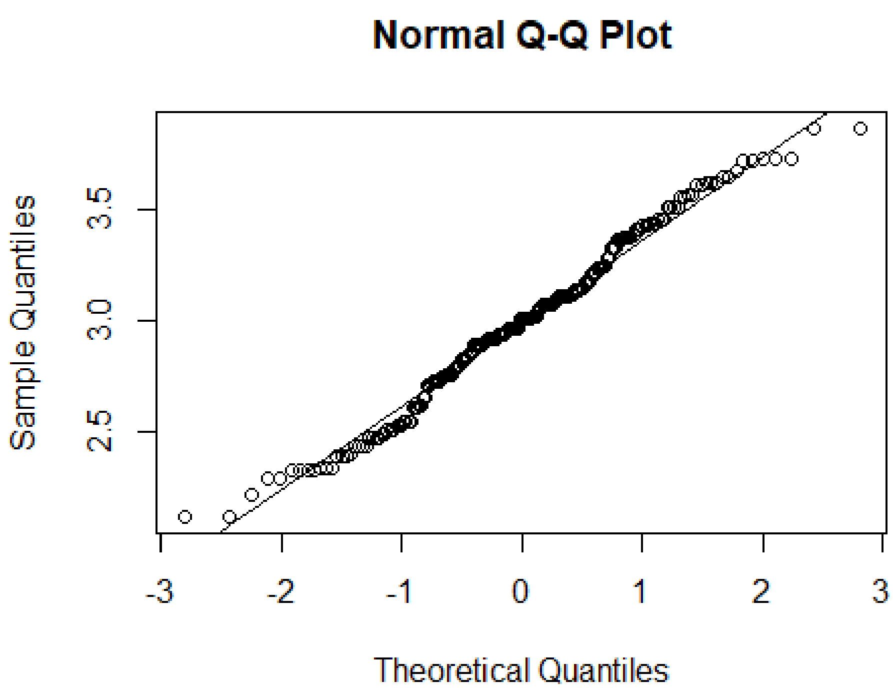
- Osservazioneper tutte le variabili qualitative in rapporto a quella quantitativa, traendo conclusioni sulla simmetria e la differenze tra le medie dei valori qualitativi.

```
tapply(Studenti$GPA, Studenti$sex, summary)

## $`0`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   2.120   2.740   3.010   2.995   3.245   3.870 
## 
## $`1`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   2.350   2.917   3.280   3.163   3.430   3.800
```

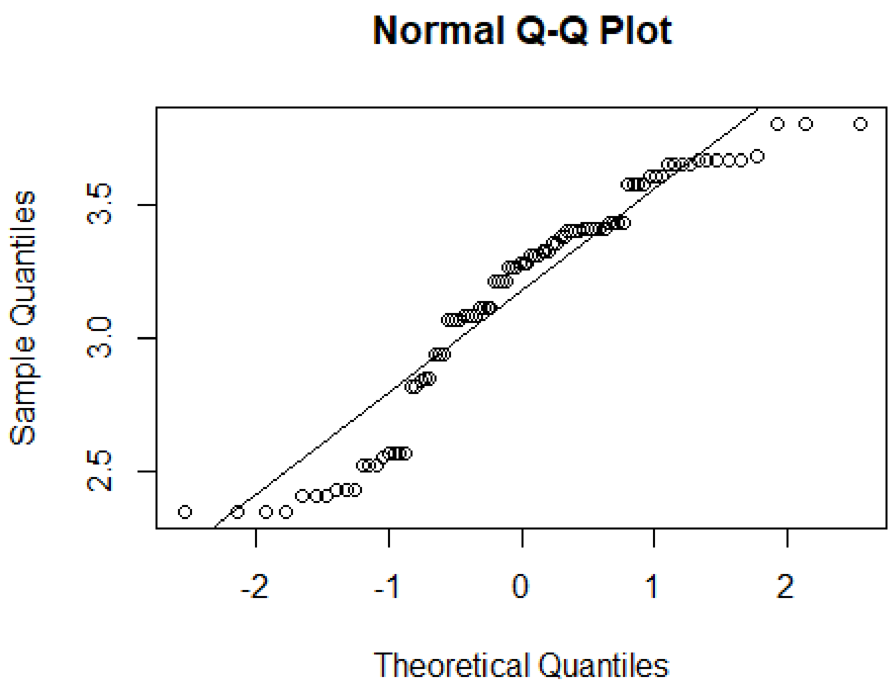
Prendiamo le classi quantitative con specifiche caratteristiche Qualitative

- Otteniamo due diversi q-q plot:

```
NomeCar_Qual1= Studenti$GPA[Studenti$sex=="0"] #femmine
NomeCar_Qual2= Studenti$GPA[Studenti$sex=="1"] #maschio

qqnorm(NomeCar_Qual1)
qqline(NomeCar_Qual1)
```




```
qqnorm(NomeCar_Qual2)
qqline(NomeCar_Qual2)
```



#Skewness e Curtosi

Un istogramma asimmetrico presenta una coda più lunga dell'altra.

Se la coda destra è più lunga, si parla di asimmetria positiva, se la coda sinistra è più lunga si ha asimmetria negativa.

Si noti che:

- se l'asimmetria è positiva allora media > mediana;
- se c'è simmetria allora media = mediana;
- se l'asimmetria è negativa allora media < mediana.

Se la distribuzione di frequenza è:

- simmetrica se l'indice di simmetria = 0;
- asimmetrica negativa se l'indice di simmetria < 0;
- asimmetrica positiva se l'indice di simmetria > 0.

curtosi

La curtosi corrisponde ad un allontanamento dalla distribuzione di frequenza normale (o gaussiana), che viene considerata come riferimento.

```
Una distribuzione platicurtica (ipornormale) presenta un maggiore
appiattimento e code leggere, mentre una distribuzione leptocurtica
(ipernormale) manifesta un maggiore allungamento e code pesanti.
```

Se la distribuzione di frequenza è:

- normocurtica se l'indice di curtosi = 3;
- leptocurtica se l'indice di curtosi > 3;
- se è platicurtica se l'indice di curtosi < 3.
- In entrambi i casi possiamo vedere un indice di curtosi prossimo a 3, lievemente platocurtica.

```
skewness(NomeCar_Qual1)

## [1] -0.005970197

kurtosis(NomeCar_Qual1)

## [1] 2.408791

skewness(NomeCar_Qual2)

## [1] -0.5899758

kurtosis(NomeCar_Qual2)

## [1] 2.247222
```

Test di indipendenza:

- rifiutare o meno l'ipotesi che le varianze siano uguali se le varienze sono uguali

```
#Se sono indipendenti
var.test(NomeCar_Qual1,NomeCar_Qual2)

##
##  F test to compare two variances
##
## data:  NomeCar_Qual1 and NomeCar_Qual2
## F = 0.86033, num df = 198, denom df = 91, p-value = 0.3854
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.5975574 1.2101203
## sample estimates:
## ratio of variances
##          0.8603304
```

#Varienze sono uguali

- Valutiamo qui le medie nell'ipotesi che siano uguali o meno

```
t.test(NomeCar_Qual1,NomeCar_Qual2, var.equal= T) #T-test

##
##  Two Sample t-test
##
## data:  NomeCar_Qual1 and NomeCar_Qual2
## t = -3.3589, df = 289, p-value = 0.0008876
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.26695020 -0.06968952
## sample estimates:
## mean of x mean of y
##  2.994724  3.163043
```