Febbraio 2020

Esercizio 0

Sia X una variabile casuale che segue una legge bernoulliana di parametro p.

1. Quali valori può assumere X?

X=1 successo, X=0 insuccesso

2. Quali valori può assumere il parametro p?

$0 \leq p \leq 1$

3. Esprimete, in funzione di p, il valore atteso E(X)

$E(X) = p$

4./7. Grafici E(X) Var(X)

E(X) diagonale con p max 1 e E(X) max 1

Var(X) parabola tra 0 e 1, con max $\frac{1}{4}$

5. Quali valori può assumere E(X)?

E(X) è tra 0 e 1 in quanto E(X) = p

6. Esprimete, in funzione di p, la varianza Var(X)

$Var(X) = p(1-p)$

8. Quali valori può assumere Var(X)?

var(X) è tra 0 e $\frac{1}{2}$

Esercizio 1

Sia $\overline{X_n}$ media campionaria di un campione di una popolazione bernoulliana di parametro p.

1. Esprimete $E(\overline{X_n})$ in funzione di p

$(\overline{X_n}) = E(X) = p$

2. Esprimete $Var(\overline{X_n})$ in funzione di p e n

$Var(\overline{X_n}) = \frac{Var(X)}{n} = \frac{p(1-p)}{n}$

3. Controllate che $Var(\overline{X_n}) \leq \frac{1}{4n}$

Nel punto massimo Var(X) è uguale a \frac{1}{4} perchè p = \frac{1}{2} perciò Var(X) = \frac{1}{2}-\frac{1}{4}.

Quindi,nel punto massimo, $Var(\overline{X_n}) = \frac{Var(X)}{n} = \frac{1}{4n}$ quindi sarà sempre minore di quel valore.

4. Controllate che, per ogni $\epsilon > 0$, vale la disuguaglianza $P(|\overline{X_n} - p| \leq \epsilon) \geq 1 - \frac{1}{4n\epsilon^2}$

Applico la disuguaglianza di Chebyshev $P(|\overline{X_n} - E(X)| \leq \epsilon) \geq 1 - \frac{Var(\overline{X_n})}{\epsilon^2}$

So che E(X) = p e che $Var(\overline{X_n}) = \frac{Var(X)}{n}$ quindi ottengo

$P(|\overline{X_n} - p| \leq \epsilon) \geq 1 - \frac{Var(X)}{n\epsilon^2}$

Poichè Var(X) nel punto massimo è $\frac{1}{4}$, allora ottengo

$P(|\overline{X_n} - p| \leq \epsilon) \geq 1 - \frac{1}{4n\epsilon^2}$

Esercizio 2

1. $\overline{X_n}$ è uno stimatore non distorto di p?

$$E(\overline{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{2} E(X_i) = \frac{n}{n}E(X) = E(X) = p$$

Ottengo questo risultato per la linearità del valore atteso e risolvendo la sommatoria.

2. Esprimete 1-p in funzione di E(X)

Poichè E(X) = p allora 1-p = 1-E(X)

3. Determinare uno stimatore S_n del parametro 1-p

$$S_n = 1 - \overline{X_n}$$

4. Lo stimatore trovato al punto precedente è non distorto?

Si, infatti

$$E(S_n) = E(1 - \overline{X_n}) = E(1 - E(X)) = E(1-p)$$

Esercizio 3

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:

```
car = pd.read_csv("mtcars2.txt",sep="\t",decimal=",")
car.columns
```

Out[2]:

```
Index(['modello', 'consumo', 'cilindrata', 'peso', 'test400metri',
       'trasmissione', 'marce', 'cilindri'],
      dtype='object')
```

1. Quanti casi contiene il dataset?

2. Tracciare il boxplot del carattere cilindrata

3. Qual è e quali sono i modelli di auto che posssono essere considerati degli outlier rispetto alla cilindrata

4. Calcolate i quartili del carattere cilindrata

5. Calcolate la distanza interquartile del carattere cilindrata

In [3]:

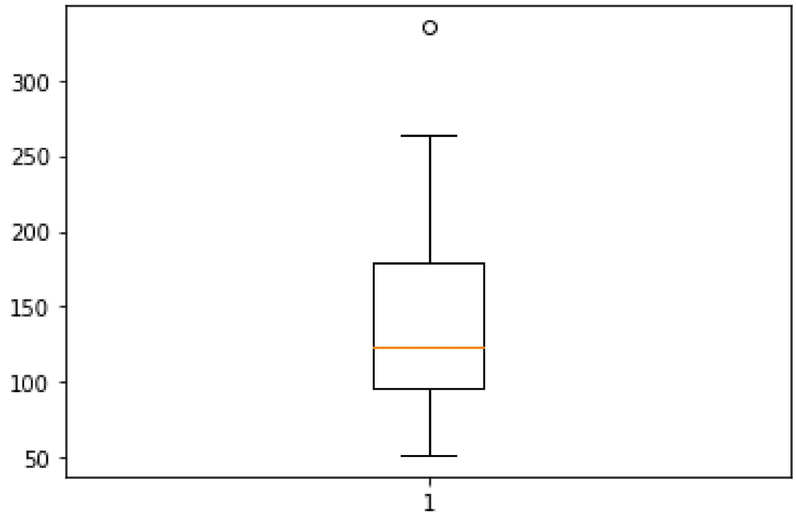
```
len(car.dropna())
```

Out[3]:

```
32
```

In [4]:

```
plt.boxplot(car.cilindrata)
plt.show()
```



In [5]:

```
car[car.cilindrata == max(car.cilindrata)].modello
```

Out[5]:

```
30    Maserati Bora
Name: modello, dtype: object
```

In [6]:

```
print(car.cilindrata.quantile(0),car.cilindrata.quantile(0.25),car.cilindrata.qu
antile(0.5),car.cilindrata.quantile(0.75),car.cilindrata.quantile(1))
```

```
52.0 96.5 123.0 180.0 335.0
```

In [7]:

```
car.cilindrata.quantile(0.75)-car.cilindrata.quantile(0.25)
```

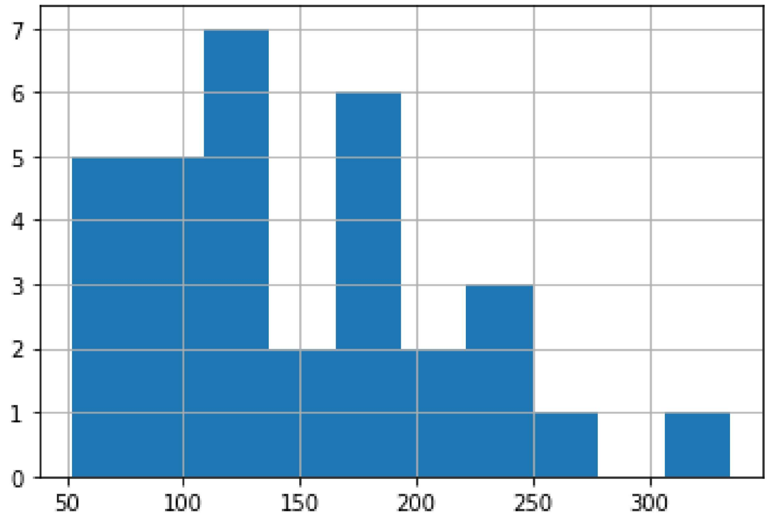
Out[7]:

```
83.5
```

6. Tracciate un grafico diverso dal boxplot

In [8]:

```
car.cilindrata.hist()
plt.show()
```

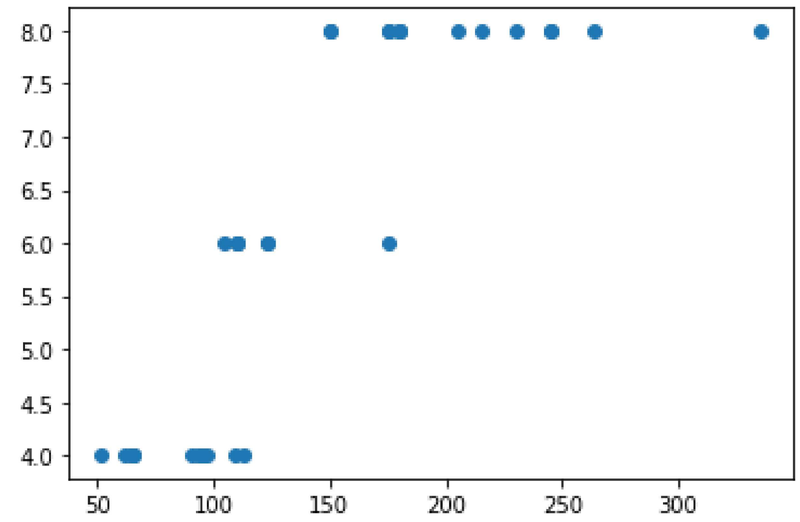


Esercizio 4

1. Tracciate un grafico per controllare se c'è una relazione tra il numero di cilindri e la cilindrata

In [9]:

```
plt.scatter(car.cilindrata, car.cilindri)
plt.show()
```



2. Individuate nel grafico una relazione tra il numero di cilindri e la cilindrata

Tre fasce di valori di cilindrata con cilindri uguali.

3. Indice numerico a supporto della vostra risposta

In [10]:

```
car.cilindrata.corr(car.cilindri)
```

Out[10]:

```
0.8324474527218194
```

Esercizio 5

1. media del carattere cilindrata

In [11]:

```
m = car.cilindrata.mean()
m
```

Out[11]:

146.6875

2. dev standard del carattere cilindrata

In [12]:

```
ds = car.cilindrata.std()
ds
```

Out[12]:

68.56286848932059

3. generate un campione casuale di 32 elementi

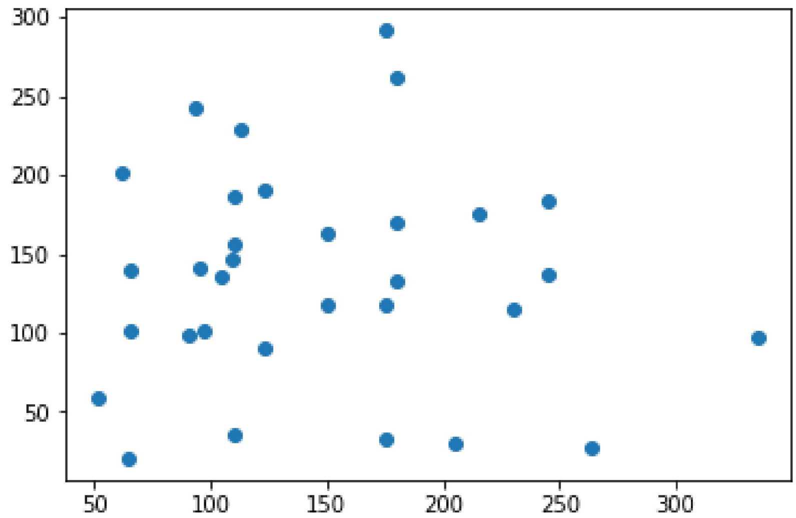
In [13]:

```
import scipy.stats as st
X = st.norm(m,ds)
valoriSimulati = X.rvs(32)
```

4. tracciate il diagramma di dispersione tra cilindrata e valori simulati

In [14]:

```
plt.scatter(car.cilindrata,valoriSimulati)
plt.show()
```



5./6. Ordinare i valori

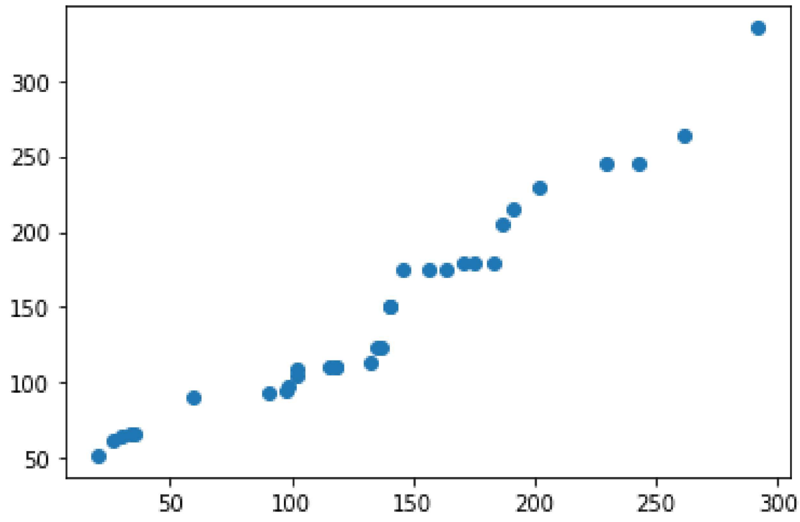
In [15]:

```
cilindrataSorted = car.cilindrata.sort_values()
valoriSimulatiSorted = np.sort(valoriSimulati)
```

7. Diagramma di dispersione tra i valori ordinati

In [16]:

```
plt.scatter(valoriSimulatiSorted,cilindrataSorted)
plt.show()
```



8.

(i)

(ii)

Perchè è importante che la variabile valoriSimulati sia basata su un campione di 32 elementi

Perchè i valori numeri avrebbero differente distribuzione se il numero di elementi del campione fossero differenti. Infatti anche cilindrata ha 32 elementi

Esercizio 6

In [17]:

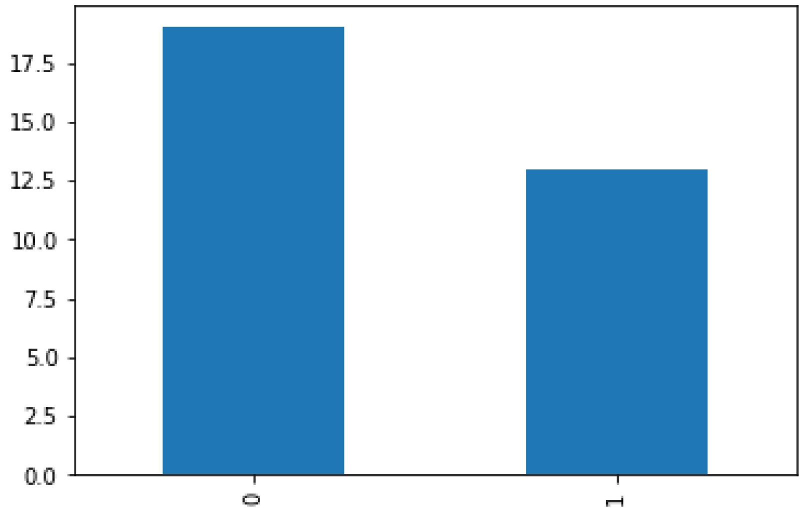
```
tr = car.trasmissione
```

1. Tracciate un grafico opportuno per descrivere il carattere trasmissione

Il plot bar mette in risalto che il carattere trasmissione può assumere solo due valori, 0 e 1, e ne da le frequenze.

In [18]:

```
tr.value_counts().plot.bar()
plt.show()
```



2. Che legge segue la variabile casuale X = "tipo di trasmissione"

Segue una legge bernoulliana in quanto abbiamo solo due tipi di valori 0 e 1, ovvero se la trasmissione è automatica o non lo è.

3. Stima valore atteso di X

In [19]:

```
tr.mean()
```

Out[19]:

```
0.40625
```

4. Lo stimatore è non distorto?

Si, lo stimatore è non distorto perchè è la media campionaria la quale è sempre uno stimatore non distorto per il valore atteso.

5. Quale è la taglia del campione?

In [20]:

```
len(tr.dropna())
```

Out[20]:

32

6. tabella delle frequenze assolute

In [21]:

```
pd.crosstab(index=tr, columns=['Frequenze Assolute'],colnames=[''],margins=True)
```

Out[21]:

trasmissione	Frequenze Assolute	All
0	19	19
1	13	13
All	32	32

7. Stimate la probabilità che un'auto abbia la trasmissione manuale (1)

P(X = 1)

In [22]:

```
car.trasmissione.mean()
```

Out[22]:

0.40625

8. è non distorto

Si, lo stimatore è non distorto perchè è la media campionaria la quale è sempre uno stimatore non distorto per il valore atteso.

9. Stimate la probabilità che un'auto abbia la trasmissione manuale (1)

1- P(X = 1)

In [23]:

```
1-car.trasmissione.mean()
```

Out[23]:

```
0.59375
```

10. è non distorto

1-E(X) è uno stimatore non distorto, vedi esercizio 2.4

11. Dato alpha = 0.85 determinate l'errore massimo commesso con probabilità maggiore o uguale ad alpha, per eccesso o per difetto, nella stima del valore atteso di X.

$$P(|T_n - E(X)| \leq \epsilon) \leq \alpha$$

Standardizzo e applico TLC

$$2\Phi(\frac{\epsilon\sqrt{n}}{\sigma}) - 1 \leq \alpha$$
$$\epsilon \leq \frac{\Phi^{-1}(\frac{\alpha+1}{2})\sigma}{\sqrt{n}}$$

In [24]:

```
import math
a = (0.85+1)/2
Z = st.norm()
phi = Z.ppf(a)
n = len(tr.dropna())
sigma = tr.std()
e = (phi*sigma)/math.sqrt(n)
e
```

Out[24]:

```
0.12698102114579055
```