

In [ ]:

```
#S07 T01: Tasca del test d'hipòtesis
```

In [ ]:

```
#Nivell 1
```

In [62]:

```
#Exercici 1
#Agafa un conjunt de dades de tema esportiu que t'agradi i selecciona un atribut del co
njunt de dades.
#Calcula el p-valor i digues si rebutja la hipòtesi nul·la agafant un alfa de 5%.

#Llibreries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import plotly.express as px
import random
import seaborn as sns
import warnings
import statistics

#Prenem el conjunt de dades de tema esportiu de la pàgina web
#https://www.kaggle.com/abecklas/fifa-world-cup/version/5

#Dataset FIFA World Cup
#Fitxer: Python/WorldCupPlayers.csv
#Fitxer: Python/WorldCups.csv
#Fitxer: Python/WorldCupMatches.csv

wcupplay_df = pd.read_csv('Python/WorldCupPlayers.csv', engine="python", error_bad_line
s=False, warn_bad_lines=False, sep=',')
wcup_df = pd.read_csv('Python/WorldCups.csv', engine="python", error_bad_lines=False, w
arn_bad_lines=False, sep=',')
wcupmatch_df = pd.read_csv('Python/WorldCupMatches.csv', engine="python", error_bad_lin
es=False, warn_bad_lines=False, sep=',')
```

In [63]:

```
wcup_df
```

Out[63]:

	Year	Country	Winner	Runners-Up	Third	Fourth	GoalsScored	Qual
0	1930	Uruguay	Uruguay	Argentina	USA	Yugoslavia	70	
1	1934	Italy	Italy	Czechoslovakia	Germany	Austria	70	
2	1938	France	Italy	Hungary	Brazil	Sweden	84	
3	1950	Brazil	Uruguay	Brazil	Sweden	Spain	88	
4	1954	Switzerland	Germany FR	Hungary	Austria	Uruguay	140	
5	1958	Sweden	Brazil	Sweden	France	Germany FR	126	
6	1962	Chile	Brazil	Czechoslovakia	Chile	Yugoslavia	89	
7	1966	England	England	Germany FR	Portugal	Soviet Union	89	
8	1970	Mexico	Brazil	Italy	Germany FR	Uruguay	95	
9	1974	Germany	Germany FR	Netherlands	Poland	Brazil	97	
10	1978	Argentina	Argentina	Netherlands	Brazil	Italy	102	
11	1982	Spain	Italy	Germany FR	Poland	France	146	
12	1986	Mexico	Argentina	Germany FR	France	Belgium	132	
13	1990	Italy	Germany FR	Argentina	Italy	England	115	
14	1994	USA	Brazil	Italy	Sweden	Bulgaria	141	
15	1998	France	France	Brazil	Croatia	Netherlands	171	
16	2002	Korea/Japan	Brazil	Germany	Turkey	Korea Republic	161	
17	2006	Germany	Italy	France	Germany	Portugal	147	
18	2010	South Africa	Spain	Netherlands	Germany	Uruguay	145	
19	2014	Brazil	Germany	Argentina	Netherlands	Brazil	171	



In [73]:

```
#Calcula el p-valor i digues si rebutja la hipòtesi nul·la agafant un alfa de 5%.
#p-valor: nivell de significació observat
#alpha: nivell de significació d'error
import scipy.stats as stats

alpha = 0.05

wcup_df['Attendance'] = wcup_df['Attendance'].astype(str).str.replace('.', '').astype(int)
wcup_df['Attendance'].dropna() #Esborrem els valors Nan

pvalue = stats.bartlett(wcup_df['Attendance'], wcup_df['Year']).pvalue
print('P-value:', round(pvalue, 10))
#P-value: 0.0 té tendència a valor 0 i és menor a 0.05. Es rebutja la hipòtesi nul·la.

if (pvalue < alpha):
    print('Es rebutja la hipòtesi nul·la')
else:
    print('Hipòtesi nul·la acceptada')
```

P-value: 0.0

In [ ]:

```
#Nivell 2
```

In [74]:

```
#Exercici 2
#Continua amb el conjunt de dades de tema esportiu que t'agradi i selecciona dos altres atributs del conjunt de dades.
#Calcula els p-valors i digues si rebutgen la hipòtesi nul·la agafant un alfa de 5%.
#p-valor: nivell de significació observat
#alpha: nivell de significació d'error
import scipy.stats as stats

alpha = 0.05

pvalue = stats.bartlett(wcup_df['GoalsScored'], wcup_df['QualifiedTeams']).pvalue
print('P-value:', round(pvalue, 8))
#P-value: 1e-08 té tendència a valor 0 i és menor a 0.05. Es rebutja la hipòtesi nul·la.

if (pvalue < alpha):
    print('Es rebutja la hipòtesi nul·la')
else:
    print('Hipòtesi nul·la acceptada')
```

P-value: 1e-08

In [ ]:

```
#Nivell 3
```

In [78]:

```
#Exercici 3
#Continua amb el conjunt de dades de tema esportiu que t'agradi i selecciona tres atributs del conjunt de dades.
#Calcula el p-valor i digues si rebutja la hipòtesi nul·la agafant un alfa de 5%.
#p-valor: nivell de significació observat
#alpha: nivell de significació d'error
import scipy.stats as stats

alpha = 0.05

pvalue = stats.bartlett(wcup_df['GoalsScored'], wcup_df['QualifiedTeams'], wcup_df['MatchesPlayed']).pvalue
print('P-value:', round(pvalue, 8))
#P-value: 3e-08 té tendència a valor 0 i és menor a 0.05. Es rebutja la hipòtesi nul·la.

if (pvalue < alpha):
    print('Es rebutja la hipòtesi nul·la')
else:
    print('Hipòtesi nul·la acceptada')
```

P-value: 3e-08

Es rebutja la hipòtesi nul·la