```python
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```python
!wget --no-check-certificate https://drive.google.com/uc?id=1n9268qSj_5Q7KS_ntpv-wFPQC1UyJh6f -O aero
```

```
--2024-07-03 16:01:58--  https://drive.google.com/uc?id=1n9268qSj_5Q7KS_ntpv-wFPQC1UyJh6f
Resolving drive.google.com (drive.google.com)... 74.125.195.113, 74.125.195.139, 74.125.195.101, ...
Connecting to drive.google.com (drive.google.com)|74.125.195.113|:443... connected.
HTTP request sent, awaiting response... 303 See Other
Location: https://drive.usercontent.google.com/download?id=1n9268qSj_5Q7KS_ntpv-wFPQC1UyJh6f [following]
--2024-07-03 16:01:58--  https://drive.usercontent.google.com/download?id=1n9268qSj_5Q7KS_ntpv-wFPQC1UyJh6f
Resolving drive.usercontent.google.com (drive.usercontent.google.com)... 173.194.203.132, 2607:f8b0:400e:c0c::84
Connecting to drive.usercontent.google.com (drive.usercontent.google.com)|173.194.203.132|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 7461 (7.3K) [application/octet-stream]
Saving to: 'aerofit.csv'

aerofit.csv          100%[===================>]   7.29K  --.-KB/s    in 0s

2024-07-03 16:01:58 (54.4 MB/s) - 'aerofit.csv' saved [7461/7461]
```
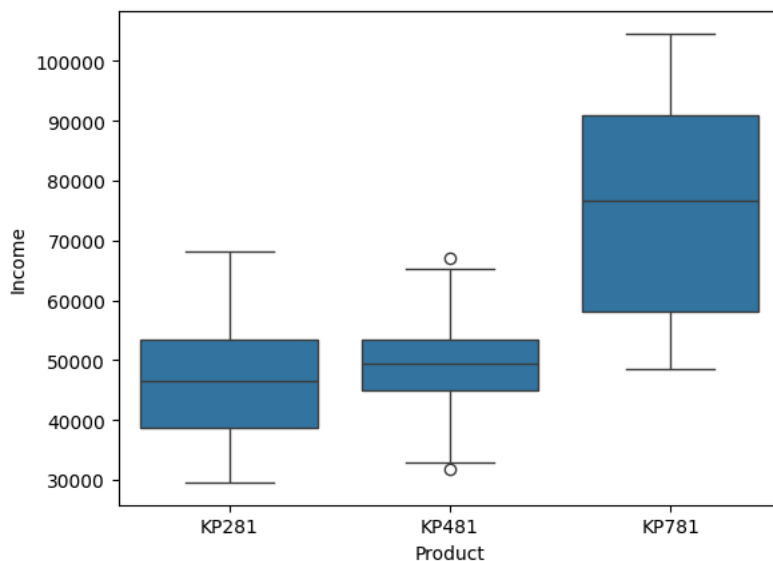
```python
df_aerofit = pd.read_csv('/content/aerofit.csv')
df_aerofit.head()
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
0	KP281	18	Male	14	Single	3	4	29562	112
1	KP281	19	Male	15	Single	2	3	31836	75
2	KP281	19	Female	14	Partnered	4	3	30699	66
3	KP281	19	Male	12	Single	3	3	32973	85
4	KP281	20	Male	13	Partnered	4	2	35247	47

```python
sns.boxplot(df_aerofit, x='Product', y='Income')
```

```
<Axes: xlabel='Product', ylabel='Income'>
```



```python
np.random.seed(42)
df_aerofit["random_group"] = np.random.choice(["g1", "g2", "g3"],size=len(df_aerofit))

df_aerofit.head()
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	random_group
0	KP281	18	Male	14	Single	3	4	29562	112	g3
1	KP281	19	Male	15	Single	2	3	31836	75	g1
2	KP281	19	Female	14	Partnered	4	3	30699	66	g3
3	KP281	19	Male	12	Single	3	3	32973	85	g3
4	KP281	20	Male	13	Partnered	4	2	35247	47	g1

```python
sns.boxplot(x='random_group', y='Income', data=df_aerofit)
```

<Axes: xlabel='random_group', ylabel='Income'>
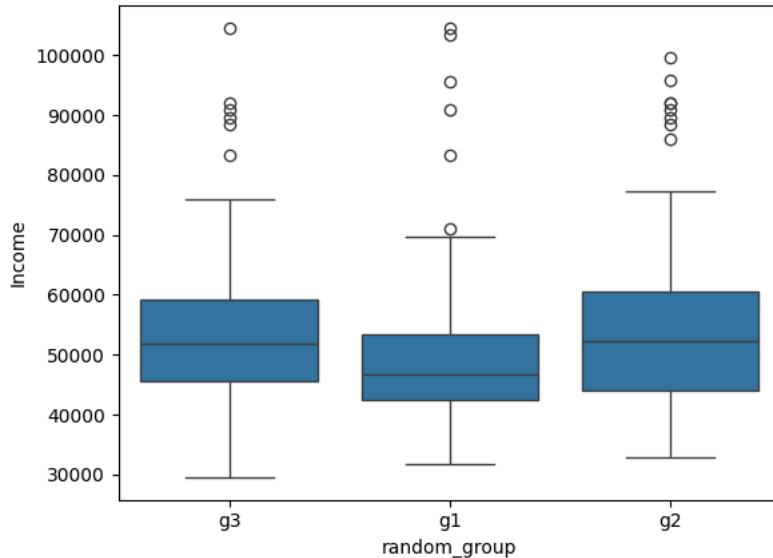

```python
from scipy.stats import f_oneway # Numeric Vs categorical for many categories
from scipy.stats import ttest_ind # Numeric Vs categorical

from statsmodels.graphics.gofplots import qqplot

income_g1 = df_aerofit[df_aerofit["random_group"]=="g1"]["Income"]
income_g2 = df_aerofit[df_aerofit["random_group"]=="g2"]["Income"]
income_g3 = df_aerofit[df_aerofit["random_group"]=="g3"]["Income"]

income_g1.mean(), income_g2.mean(), income_g3.mean()
```

(51840.44827586207, 55812.69642857143, 53594.954545454544)

```python
f_oneway(income_g1, income_g2, income_g3)
```

F_onewayResult(statistic=0.8263259710788367, pvalue=0.43933541801191467)

```python
df_aerofit.Product.unique()
```

array(['KP281', 'KP481', 'KP781'], dtype=object)

```python
income_g1 = df_aerofit[df_aerofit["Product"]=="KP281"]["Income"]
income_g2 = df_aerofit[df_aerofit["Product"]=="KP481"]["Income"]
income_g3 = df_aerofit[df_aerofit["Product"]=="KP781"]["Income"]

f_oneway(income_g1, income_g2, income_g3)
```

F_onewayResult(statistic=89.25903546601671, pvalue=1.5644991316342494e-27)

apply paired t-tests between all products. 2 at a time

```python
ttest_ind(income_g1, income_g2)
```

```
TtestResult(statistic=-1.6817688139914835, pvalue=0.09487529747264932, df=138.0)
```

```python
ttest_ind(income_g2, income_g3)
```

```
TtestResult(statistic=-9.628009470104809, pvalue=7.770191623361236e-16, df=98.0)
```

```python
ttest_ind(income_g1, income_g3)
```

```
TtestResult(statistic=-11.551787854717519, pvalue=4.116508528877672e-21, df=118.0)
```

```python
from scipy.stats import kruskal
```

```python
kruskal(income_g1, income_g2, income_g3)
```

```
KruskalResult(statistic=61.43670384567185, pvalue=4.562357014275808e-14)
```

Check for normality

```python
!wget --no-check-certificate https://drive.google.com/uc?id=1BbOJnKXUEPWrhjjLlZS_2-HoGvtZeubb -O we:
```

```
--2024-07-03 17:23:11--  https://drive.google.com/uc?id=1BbOJnKXUEPWrhjjLlZS_2-HoGvtZeubb
Resolving drive.google.com (drive.google.com)... 74.125.195.102, 74.125.195.101, 74.125.195.139, ...
Connecting to drive.google.com (drive.google.com)|74.125.195.102|:443... connected.
HTTP request sent, awaiting response... 303 See Other
Location: https://drive.usercontent.google.com/download?id=1BbOJnKXUEPWrhjjLlZS_2-HoGvtZeubb [following]
--2024-07-03 17:23:11--  https://drive.usercontent.google.com/download?id=1BbOJnKXUEPWrhjjLlZS_2-HoGvtZeubb
Resolving drive.usercontent.google.com (drive.usercontent.google.com)... 142.250.99.132, 2607:f8b0:400e:c0c::84
Connecting to drive.usercontent.google.com (drive.usercontent.google.com)|142.250.99.132|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 428120 (418K) [application/octet-stream]
Saving to: 'weight-height.csv'

weight-height.csv   100%[===================>] 418.09K  --.-KB/s    in 0.005s

2024-07-03 17:23:12 (84.2 MB/s) - 'weight-height.csv' saved [428120/428120]
```

```python
path = '/content/weight-height.csv'
df_hw = pd.read_csv(path)
```
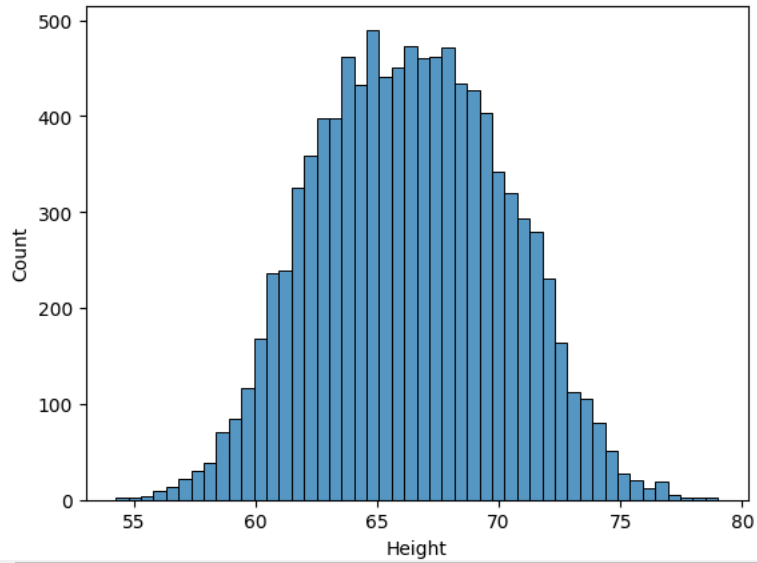
```python
df_hw.head()
```

	Gender	Height	Weight
0	Male	73.847017	241.893563
1	Male	68.781904	162.310473
2	Male	74.110105	212.740856
3	Male	71.730978	220.042470
4	Male	69.881796	206.349801

```python
sns.histplot(df_hw.Height)
```
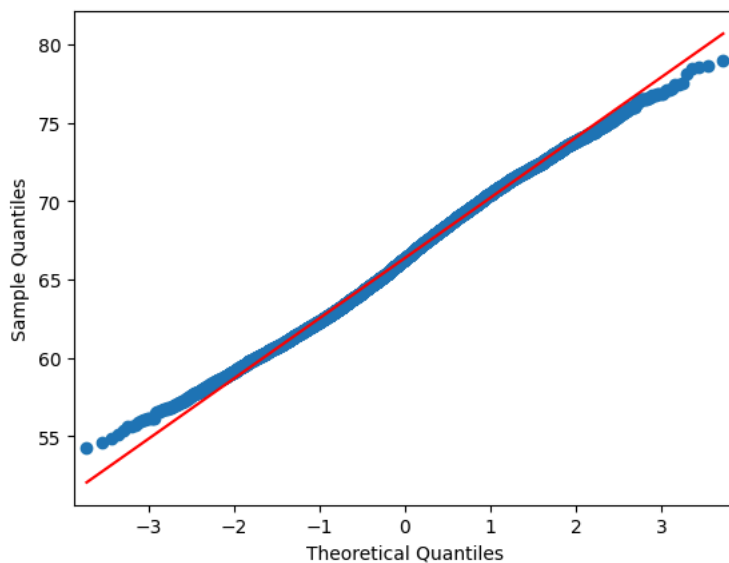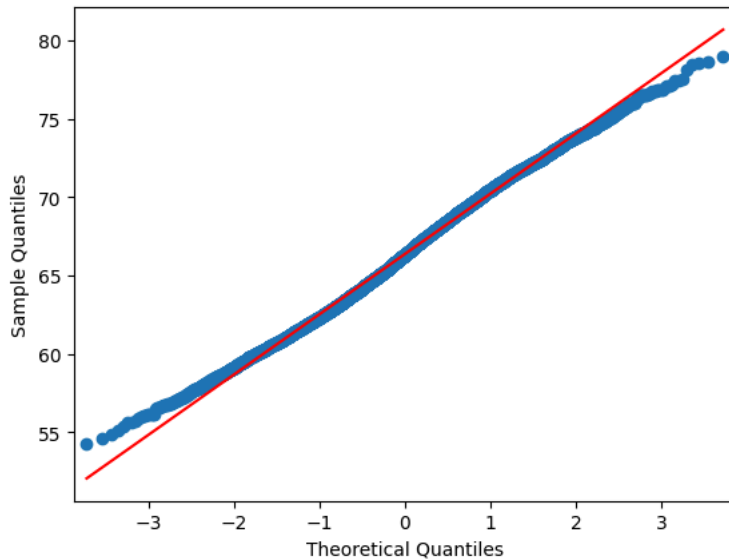
<Axes: xlabel='Height', ylabel='Count'>


```
df_hw.shape
```

(10000, 3)

```
from statsmodels.graphics.gofplots import qqplot
```

```
qqplot(df_hw.Height, line="s")
```

```
!wget --no-check-certificate https://drive.google.com/uc?id=1v48RRtiivufSX4ewrMaxsADK73GnN2kI -O wa:
```

```
--2024-07-03 17:31:26--  https://drive.google.com/uc?id=1v48RRtiivufSX4ewrMaxsADK73GnN2kI
Resolving drive.google.com (drive.google.com)... 74.125.195.100, 74.125.195.139, 74.125.195.138, ...
Connecting to drive.google.com (drive.google.com)|74.125.195.100|:443... connected.
HTTP request sent, awaiting response... 303 See Other
Location: https://drive.usercontent.google.com/download?id=1v48RRtiivufSX4ewrMaxsADK73GnN2kI [following]
--2024-07-03 17:31:26--  https://drive.usercontent.google.com/download?id=1v48RRtiivufSX4ewrMaxsADK73GnN2kI
Resolving drive.usercontent.google.com (drive.usercontent.google.com)... 142.250.99.132, 2607:f8b0:400e:c0c::84
Connecting to drive.usercontent.google.com (drive.usercontent.google.com)|142.250.99.132|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1656272 (1.6M) [application/octet-stream]
Saving to: 'waiting_time.csv'

waiting_time.csv    100%[====================>]   1.58M  --.-KB/s    in 0.04s

2024-07-03 17:31:27 (36.8 MB/s) - 'waiting_time.csv' saved [1656272/1656272]
```
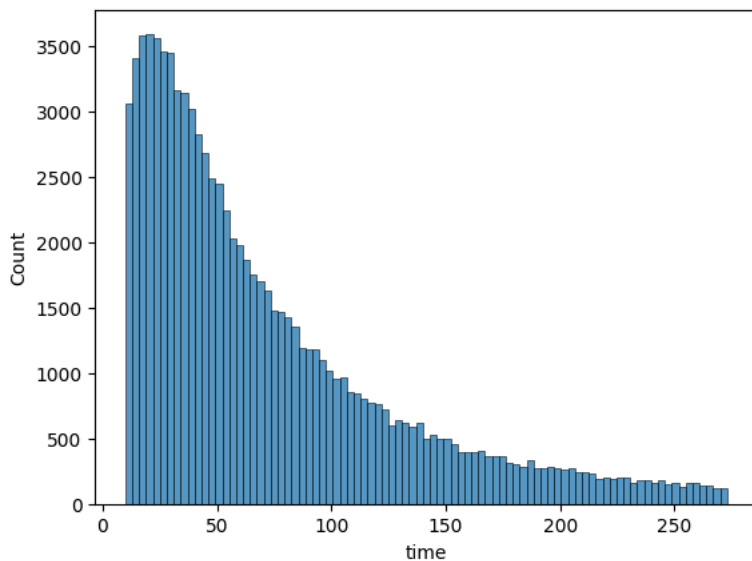
```
path = '/content/waiting_time.csv'
df_wt = pd.read_csv(path)
```

```
df_wt.head()
```

	time
0	184.003075
1	36.721521
2	29.970417

```python
sns.histplot(df_wt["time"])
```

<Axes: xlabel='time', ylabel='Count'>


```python
qqplot(df_wt["time"], line="s")
plt.show()
```