```
import statsmodels.api as sm
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

```
/usr/local/lib/python3.7/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing is deprecated. Use
  import pandas.util.testing as tm
```

Is distance Is the average cartwheel distance (in inches) for adults more than 80 inches?

Population: All adults

Parameter of Interest: μ, population mean cartwheel distance.

Null Hypothesis: μ = 80

Alternative Hypthosis: μ > 80

Data:

25 adult participants.

$\mu = 83.84$

$\sigma = 10.72$

```
df = pd.read_csv("CART.csv")
df.head()
```

ID	Age	Gender	GenderGroup	Glasses	GlassesGroup	Height	Wingspan	CWDistance	Complete	CompleteGroup	Score

```
n = len(cwdata)
mean = cwdata.mean()
sd = cwdata.std()
(n, mean, sd)
```

```
(25, 83.84320000000001, 10.716018932420752)
```

```
sm.stats.ztest(cwdata, value = 80, alternative = "larger")
```

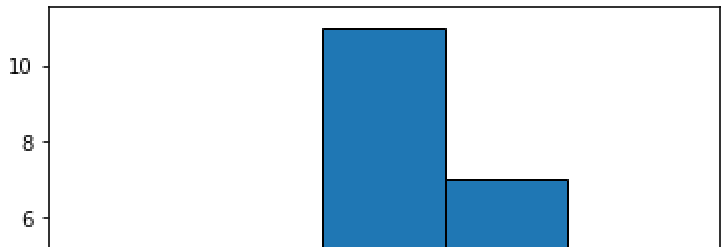
```
(1.756973189172546, 0.039461189601168366)
```

Conclusion of the hypothesis test

Since the p-value (0.0394) is lower than the standard confidence level 0.05, we can reject the Null hypothesis that the mean cartwheel distance for adults (a population quantity) is equal to 80 inches. There is strong evidence in support for the alternatine hypothesis that the mean cartwheel distance is, in fact, higher than 80 inches. Note, we used `alternative="larger"` in the z-test.

We can also plot the histogram of the data to check if it approximately follows a Normal distribution.

```
plt.hist(cwdata,bins=5,edgecolor='k')
plt.show()
```



Difference in Population Means

Research Question

Considering adults in the NHANES data, do males have a significantly higher mean Body Mass Index than females?

Population: Adults in the NHANES data.
Parameter of Interest: $\mu_1 - \mu_2$, Body Mass Index.

Null Hypothesis: $\mu_1 = \mu_2$
Alternative Hypthosis: $\mu_1 \neq \mu_2$

Data:

2976 Females $\mu_1 = 29.94$
$\sigma_1 = 7.75$

2759 Male Adults
$\mu_2 = 28.78$
$\sigma_2 = 6.25$

$\mu_1 - \mu_2 = 1.16$

```
url = "https://raw.githubusercontent.com/kshedden/statswpy/master/NHANES/merged/nhanes_2015_2016.csv"
da = pd.read_csv(url)
da.head()
```

	SEQN	ALQ101	ALQ110	ALQ130	SMQ020	RIAGENDR	RIDAGEYR	RIDRETH1	DMDCITZN	DMDEDUC2	DMDMART
0	83732	1.0	NaN	1.0	1	1	62	3	1.0	5.0	1
1	83733	1.0	NaN	6.0	1	1	53	3	2.0	3.0	3
2	83734	1.0	NaN	NaN	1	1	78	3	1.0	3.0	1
3	83735	2.0	1.0	1.0	2	2	56	3	1.0	5.0	6
4	83736	2.0	1.0	1.0	2	2	42	4	1.0	4.0	3

```
females = da[da["RIAGENDR"] == 2]
male = da[da["RIAGENDR"] == 1]


n1 = len(females)
mu1 = females["BMXBMI"].mean()
sd1 = females["BMXBMI"].std()

(n1, mu1, sd1)
```

```
(2976, 29.939945652173996, 7.75331880954568)
```

```
n2 = len(male)
mu2 = male["BMXBMI"].mean()
sd2 = male["BMXBMI"].std()

(n2, mu2, sd2)
```

```
(2759, 28.778072111846985, 6.252567616801485)
```

```
sm.stats.ztest(females["BMXBMI"].dropna(), male["BMXBMI"].dropna(),alternative='two-sided')
```

```
(6.1755933531383205, 6.591544431126401e-10)
```

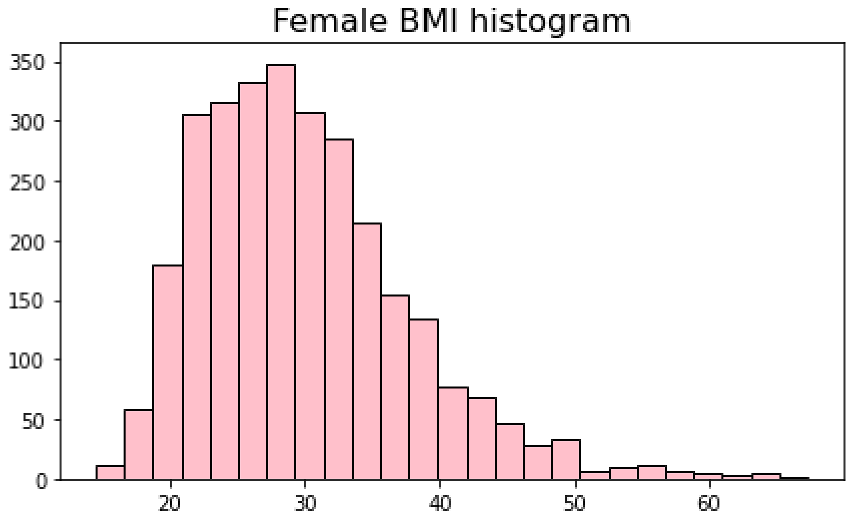
Conclusion of the hypothesis test

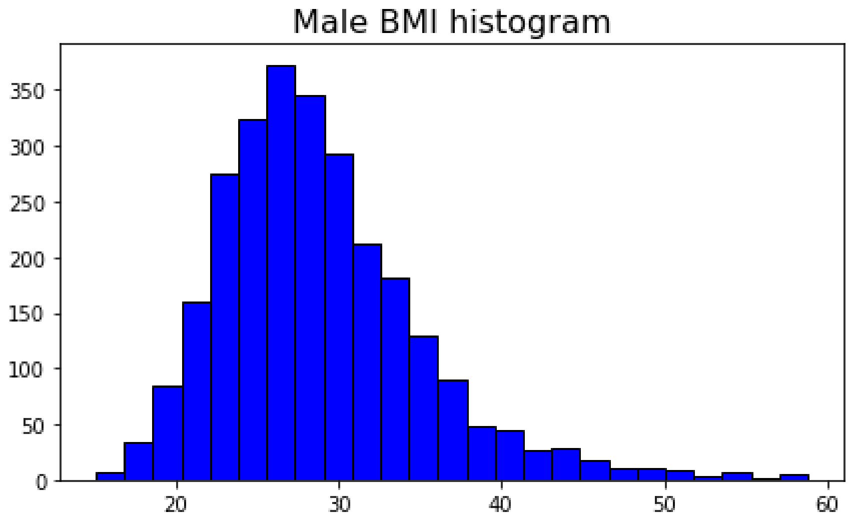
Since the p-value (6.59e-10) is extremely small, we can reject the Null hypothesis that the mean BMI of males is same as that of females. Note, we used `alternative="two-sided"` in the z-test because here we are checking for inequality.

We can also plot the histogram of the data to check if it approximately follows a Normal distribution.

```
plt.figure(figsize=(7,4))
plt.title("Female BMI histogram",fontsize=16)
plt.hist(females["BMXBMI"].dropna(),edgecolor='k',color='pink',bins=25)
plt.show()

plt.figure(figsize=(7,4))
plt.title("Male BMI histogram",fontsize=16)
plt.hist(male["BMXBMI"].dropna(),edgecolor='k',color='blue',bins=25)
plt.show()
```


Female BMI histogram
0
50
100
150
200
250
300
350
20
30
40
50
60


Male BMI histogram
0
50
100
150
200
250
300
350
20
30
40
50
60