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

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
0	1	56	F	1	Y	1	62.0	61.0	79
1	2	26	F	1	Y	1	62.0	60.0	70
2	3	33	F	1	Y	1	66.0	64.0	85
3	4	39	F	1	N	0	64.0	63.0	87
4	5	27	M	2	N	0	73.0	75.0	72

```
cwdata = df["CWDistance"]
```

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

```
(25, 82.48, 15.058552387264855)
```

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

```
(0.8234523266982029, 0.20512540845395266)
```

By using mathematical functions

```
h0_mean = 80   #Given
```

```
h1_mean = cwdata.mean()                          #z=(x-mu)/sig/sqrt(n)
h1_mean                                                  #z=(x-mu)/sigma
```

```
82.48
```

```
sigma = cwdata.std()/np.sqrt(len(cwdata))
sigma
```

```
3.0117104774529713
```

```
z = (h1_mean - h0_mean)/sigma
```

```
p_val = (1 - stats.norm.cdf(abs(z))) #ONE TAIL  I.E LOWER/UPPER TAIL PART
# p_val = (1 - stats.norm.cdf(abs(z)))*2   #TWO TAIL           #pval or prob value
p_val
```

```
0.20512540845395266
```

Since p value is greater than the significance level, we accept the null hypothesis

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_201
da = pd.read_csv(url)
da.tail(24)
```

	SEQN	ALQ101	ALQ110	ALQ130	SMQ020	RIAGENDR	RIDAGEYR	RIDRETH1	DMDCITZN
5711	93659	1.0	NaN	1.0	1	1	62	1	2.0

5712	93661	1.0	NaN	2.0	2	2	27	2	1.0
5713	93663	1.0	NaN	4.0	2	1	43	1	2.0
5714	93664	2.0	1.0	2.0	2	1	39	1	2.0
5715	93665	NaN	NaN	NaN	2	2	34	3	1.0
5716	93668	1.0	NaN	NaN	1	2	73	1	1.0
5717	93670	1.0	NaN	3.0	1	1	32	3	2.0
5718	93671	1.0	NaN	3.0	2	1	45	4	1.0
5719	93672	2.0	2.0	NaN	1	2	63	2	1.0
5720	93675	1.0	NaN	1.0	2	1	38	5	2.0
5721	93676	1.0	NaN	2.0	2	2	35	4	1.0
5722	93677	1.0	NaN	1.0	2	2	34	3	1.0
5723	93679	2.0	1.0	NaN	1	2	72	4	1.0
5724	93682	NaN	NaN	NaN	2	2	41	5	1.0
5725	93684	2.0	2.0	NaN	2	1	34	4	1.0
5726	93685	1.0	NaN	2.0	1	1	53	1	2.0
5727	93689	2.0	1.0	NaN	2	2	69	1	1.0

```
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

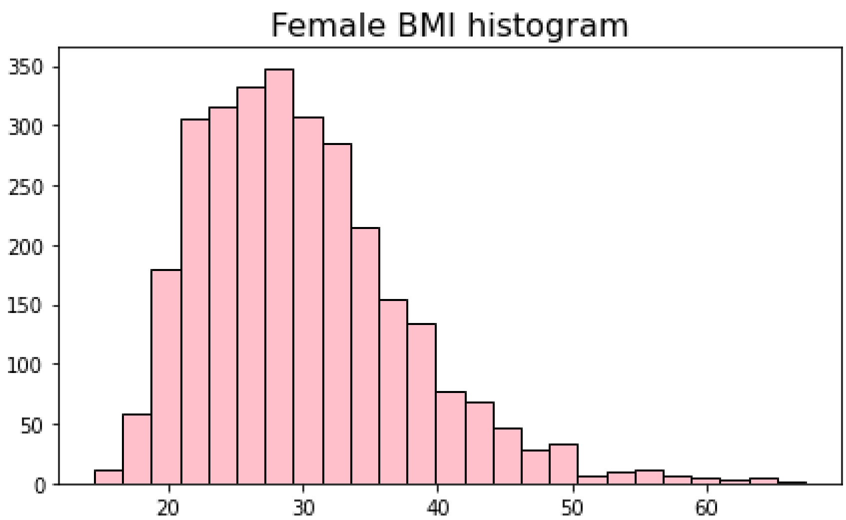
Since the p-value (6.59e-10) is extremely small, we can reject the Null hypothesis that the mean

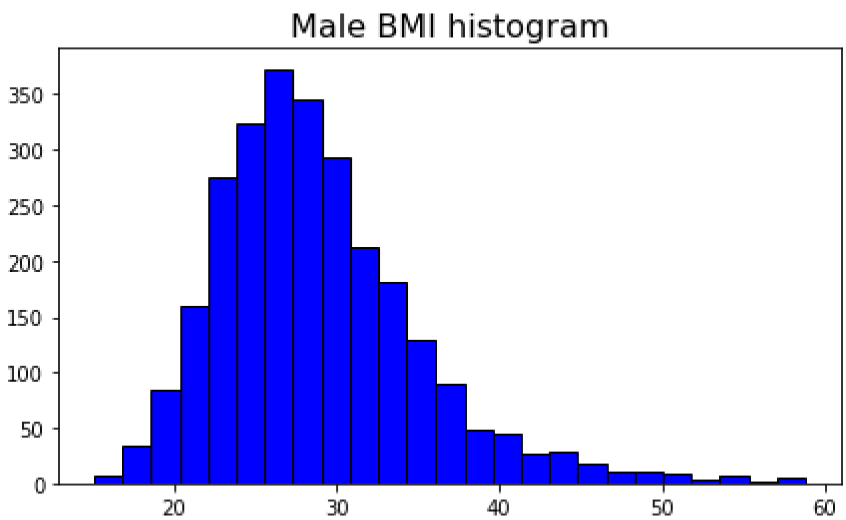
BMI of males is same as that of females. Note, we used `alternative="two-sided"` in the z-test because here we are checking for inequality.

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()
```





** Do males and females have equal smoking rates(SMQ020x)? Considering alpha=0.05 **

Population: Adults in the NHANES data. Parameter of Interest: Smoking Rates.

Null Hypothesis: Smoking rates of male and female in the polulation is same.

Alternative Hypthosis: Smoking rates among the male and female is different.

```
h0_mean = male['SMQ020'].mean()                             #z=(x-mu)/sig/sqrt(n)
h0_mean                                                     #z=(x-mu)/sigma
```

```
1.5016310257339616
```

```
h1_mean = females['SMQ020'].mean()                             #z=(x-mu)/sig/sqrt(n)
h1_mean                                                     #z=(x-mu)/sigma
```

```
1.70497311827957
```

```
sigma = females['SMQ020'].std()/np.sqrt(len(females))
sigma
```

```
0.009757648807358701
```

```
z = (h1_mean - h0_mean)/sigma
z
```

```
20.83925098762097
```

```
p_val = (1 - stats.norm.cdf(abs(z)))*2   #TWO TAIL            #pval or prob value
p_val
```

```
0.0
```

Since p-value is less than the significance level we accept the null hypothesis

✓ 0s completed at 2:41 AM