Import Libraries

In [14]:

```
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats
import numpy as np
from sklearn.preprocessing import scale
```

In [3]:

```
data = pd.read_csv("blood_pressure.csv")
```

In [4]:

```
data.head()
```

Out[4]:

	patient	sex	agegrp	bp_before	bp_after
0	1	Male	30-45	143	153
1	2	Male	30-45	163	170
2	3	Male	30-45	153	168
3	4	Male	30-45	153	142
4	5	Male	30-45	146	141

Shape

In [5]:

```
print(data.shape)
```

```
(120, 5)
```

Info for each column

In [6]:

```
print(data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 120 entries, 0 to 119
Data columns (total 5 columns):
patient      120 non-null int64
sex          120 non-null object
agegrp       120 non-null object
bp_before    120 non-null int64
bp_after     120 non-null int64
dtypes: int64(3), object(2)
memory usage: 4.8+ KB
None
```

check the null values

In [7]:

```
print(data.isnull().sum())
```

```
patient      0
sex          0
agegrp       0
bp_before    0
bp_after     0
dtype: int64
```

Pair Sample T-test

pair sample t-test is used to check mean of two group at different point in time. Here, we want to know whether the bp of patient before and after taking drug is same or difference. We want to know drugs has any affect on patient bp or not

In [10]:

```
data.describe().T
```

Out[10]:

	count	mean	std	min	25%	50%	75%	max
patient	120.0	60.500000	34.785054	1.0	30.75	60.5	90.25	120.0
bp_before	120.0	156.450000	11.389845	138.0	147.00	154.5	164.00	185.0
bp_after	120.0	151.358333	14.177622	125.0	140.75	149.5	161.00	185.0

Here, We can say that mean before intervention is 156.45 and mean after the intervention 151.35. Clearly, we can say the there is difference between means before and after intervention but is this difference statstically significant?

Assumption

Let's take two Assumption:

1. H0 : variance are same or We accept the null hypothesis, that there is no differnece before and after the drugs
2. H1 : Variance are not same menas we have rejected the null hypothesis, that means medicine has affect on Bp.

In [12]:

```
alpha = 0.05
Stats,pvalue = stats.levene(data['bp_before'],data['bp_after'])
if pvalue>alpha:
    print("Failed to reject null hypothesis or variance are same")
else:
    print("Rejected the null hypothesis or variance are not same")
```

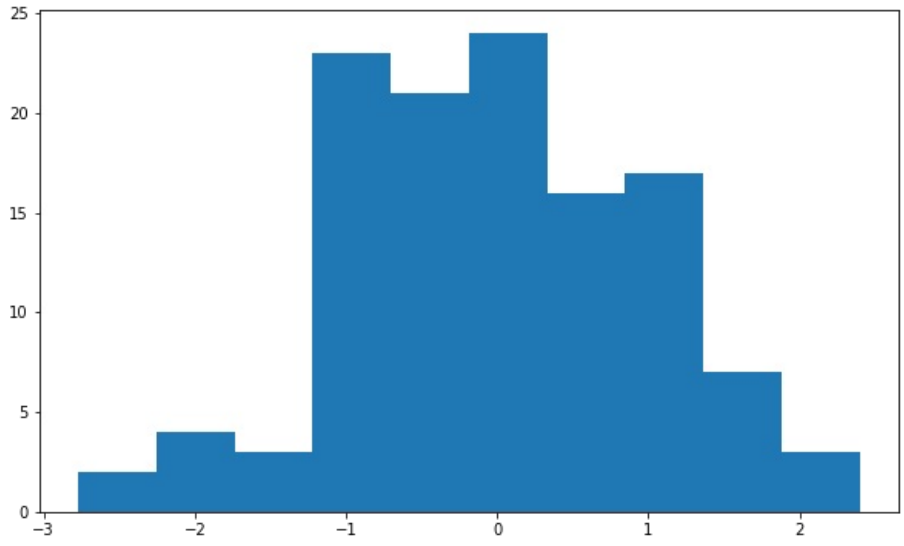
```
Rejected the null hypothesis or variance are not same
```

Check for normality

In [15]:

```
diff = scale((np.array(data['bp_before'])-(np.array(data['bp_after']))))
plt.figure(figsize=(10,6))
plt.hist(diff)
plt.show()
```

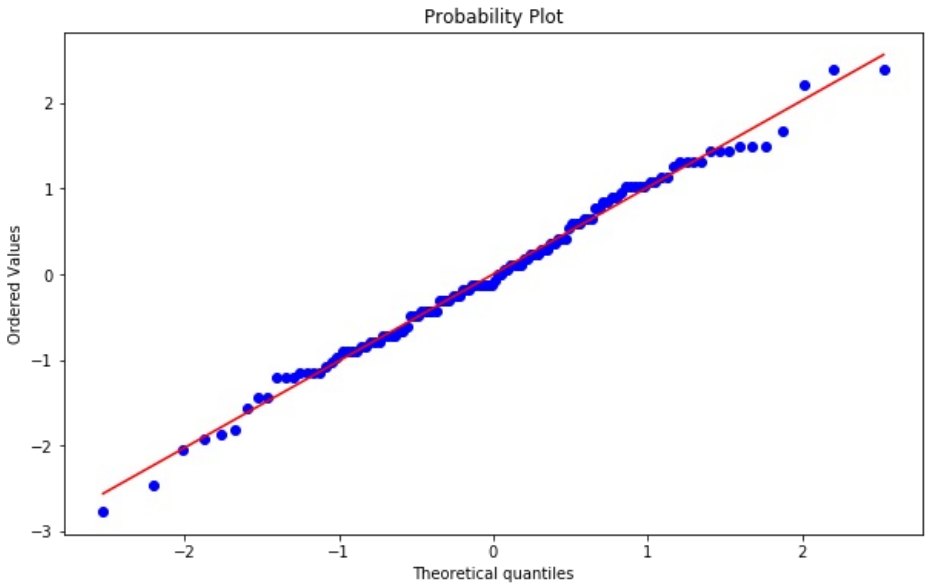
```
C:\Users\Admin\Anaconda3\lib\site-packages\sklearn\utils\validation.py:475: DataConversionWarning: D
ata with input dtype int64 was converted to float64 by the scale function.
  warnings.warn(msg, DataConversionWarning)
```



Check the Q-Q plot to check if data set comes from normal distribution. Q-Q plot generates the probability of sample data against quantiles

In [17]:

```
plt.figure(figsize=(10,6))
stats.probplot(diff,plot=plt,dist="norm")
plt.show()
```



After 1 standard deviation the points are started scattered from the redline.They doesn't follow redline.We saw that most of data close to red line so we accept the assumption of normality

Let's use Shapiro-wilk test to check the normality

1.H0 : Residuals are normally distributed

2.H1 : Residuals are not normally distributed

In [19]:

```
Stats,pvalue = stats.shapiro(diff)
if pvalue>alpha:
    print("Failed to reject Null Hypothesis")
else:
    print("Rejected the Null Hypothesis Testing")
```

```
Failed to reject Null Hypothesis
```

Result shows that we can reject null hypothesis and accept that there is significant difference before and after intervention.