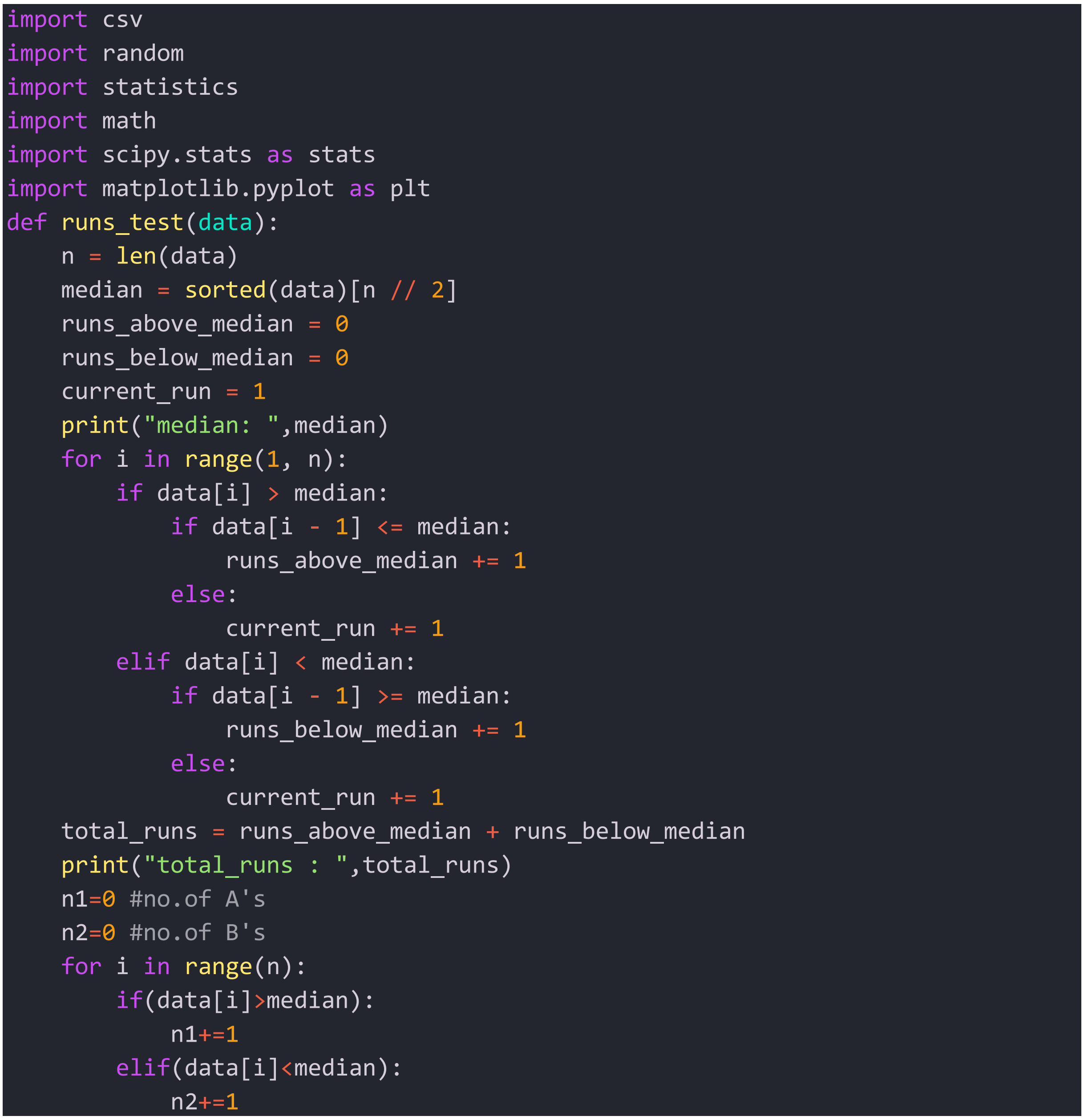
Maths DataSet Analysis-3

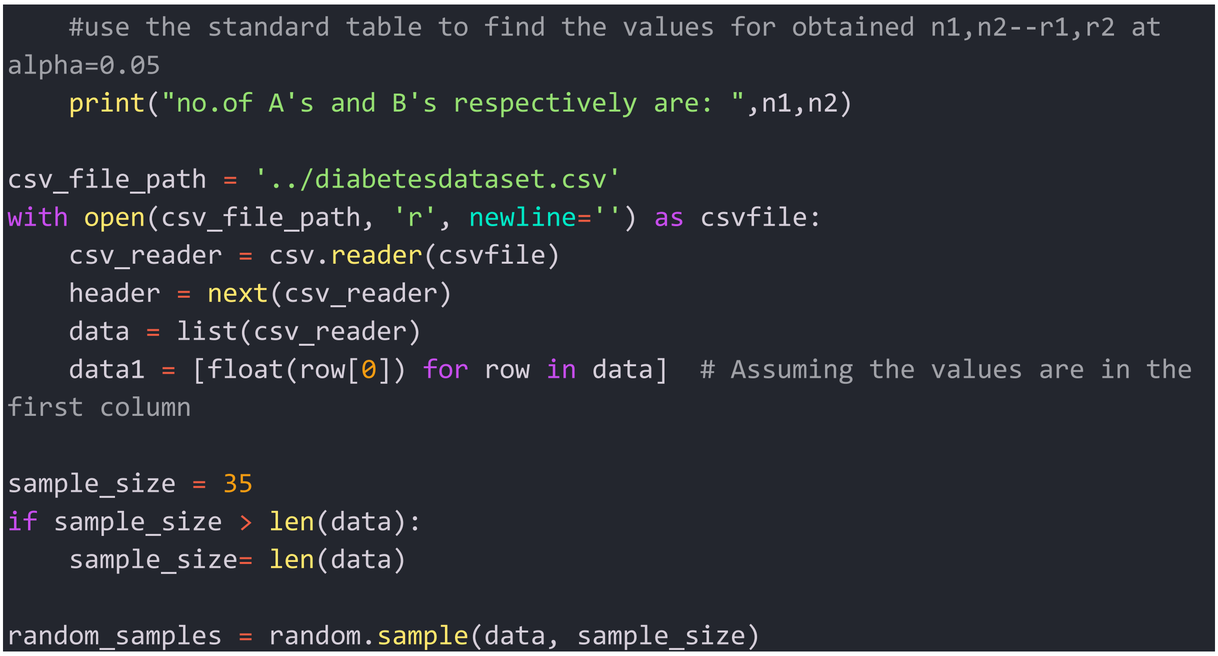
Team:

1. N.V.Saran Kumar - 21bcs075
2. Parth Pawar - 21bcs083
3. Pranay - 21bcs084

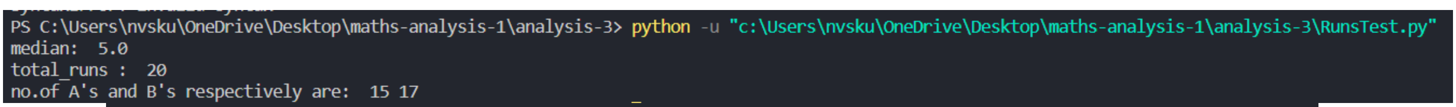
1.Runs Test:



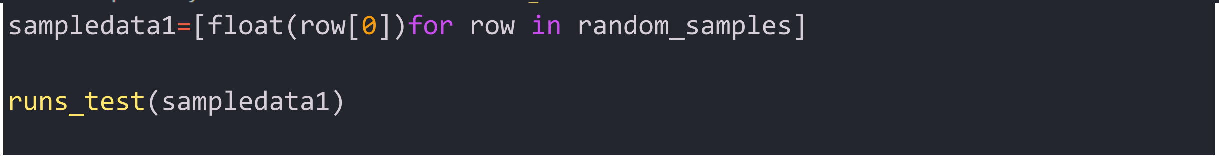
```python
import csv
import random
import statistics
import math
import scipy.stats as stats
import matplotlib.pyplot as plt
def runs_test(data):
    n = len(data)
    median = sorted(data)[n // 2]
    runs_above_median = 0
    runs_below_median = 0
    current_run = 1
    print("median: ",median)
    for i in range(1, n):
        if data[i] > median:
            if data[i - 1] <= median:
                runs_above_median += 1
            else:
                current_run += 1
        elif data[i] < median:
            if data[i - 1] >= median:
                runs_below_median += 1
            else:
                current_run += 1
    total_runs = runs_above_median + runs_below_median
    print("total_runs : ",total_runs)
    n1=0 #no.of A's
    n2=0 #no.of B's
    for i in range(n):
        if(data[i]>median):
            n1+=1
        elif(data[i]<median):
            n2+=1
```



```
    #use the standard table to find the values for obtained n1,n2--r1,r2 at alpha=0.05
    print("no.of A's and B's respectively are: ",n1,n2)

csv_file_path = '../diabetesdataset.csv'
with open(csv_file_path, 'r', newline='') as csvfile:
    csv_reader = csv.reader(csvfile)
    header = next(csv_reader)
    data = list(csv_reader)
    data1 = [float(row[0]) for row in data]  # Assuming the values are in the first column

sample_size = 35
if sample_size > len(data):
    sample_size= len(data)

random_samples = random.sample(data, sample_size)
```



```
PS C:\Users\nvsku\OneDrive\Desktop\maths-analysis-1\analysis-3> python -u "c:\Users\nvsku\OneDrive\Desktop\maths-analysis-1\analysis-3\RunsTest.py"
median:  5.0
total_runs :  20
no.of A's and B's respectively are:  15 17
```



```
sampledata1=[float(row[0])for row in random_samples]

runs_test(sampledata1)
```

output:

Use critical values of r table to find the values r1 and r2.

Here n1=15 and n2=17

From the table;

r1=11

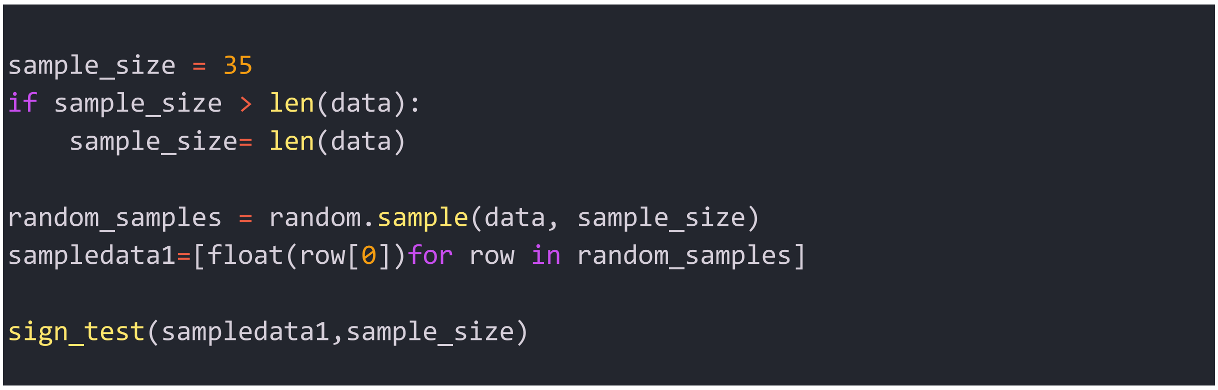
r2=23

And obtained r=total_runs=20

As r1<r<r2 ; accept Ho: Data is random

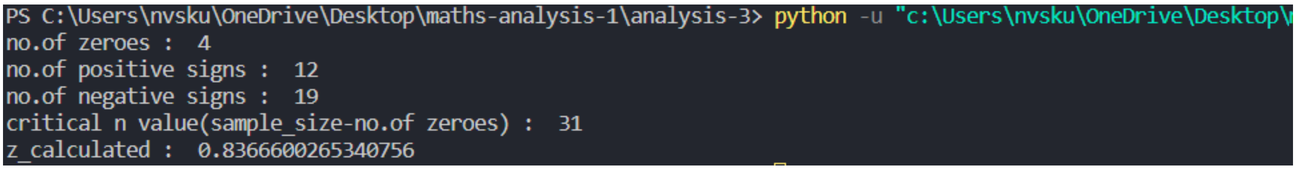
2.Sign test:

```
import csv
import random
import scipy.stats as stats
import matplotlib.pyplot as plt
import math
# H0: Median =5.0
# Ha: Median!=5.0


def sign_test(data,sample_size):

    median = 5.0 #assuming median be 5 in H0
    positives = 0
    negatives = 0
    zeroes=0
    for i in range(0, sample_size):
        if data[i] > median:
            positives+=1
        elif data[i] < median:
            negatives+=1
        else:
            zeroes+=1

    n=(sample_size-zeroes)
    print("no.of zeroes : ",zeroes)
    print("no.of positive signs : ",positives)
    print("no.of negative signs : ",negatives)
    print("critical n value(sample_size-no.of zeroes) : ",n)

    z_cal=abs((min(positives,negatives))-(n/2))/math.sqrt(sample_size/2)
    print("z_calculated : ",z_cal)



csv_file_path = '../diabetesdataset.csv'
with open(csv_file_path, 'r', newline='') as csvfile:
    csv_reader = csv.reader(csvfile)
    header = next(csv_reader)
    data = list(csv_reader)
    data1 = [float(row[0]) for row in data]  # Assuming the values are in the
first column
```



```python
sample_size = 35
if sample_size > len(data):
    sample_size= len(data)

random_samples = random.sample(data, sample_size)
sampledata1=[float(row[0])for row in random_samples]

sign_test(sampledata1,sample_size)
```

output:



```
PS C:\Users\nvsku\OneDrive\Desktop\maths-analysis-1\analysis-3> python -u "c:\Users\nvsku\OneDrive\Desktop\
no.of zeroes :  4
no.of positive signs :  12
no.of negative signs :  19
critical n value(sample_size-no.of zeroes) :  31
z_calculated :  0.8366600265340756
```

Conclusion:

- Find z_tab value from the standard table and compare with z_cal.
- If sampleSize > 26:
  - If z_cal <= z_tab: Accept Ho
  - Else: reject Ho

  If sampleSize<=26:

  - If z_cal <= ztab: Reject Ho
  - Else: accept Ho

From the table:

Z_tab=1.9600

Z_cal=0.8366

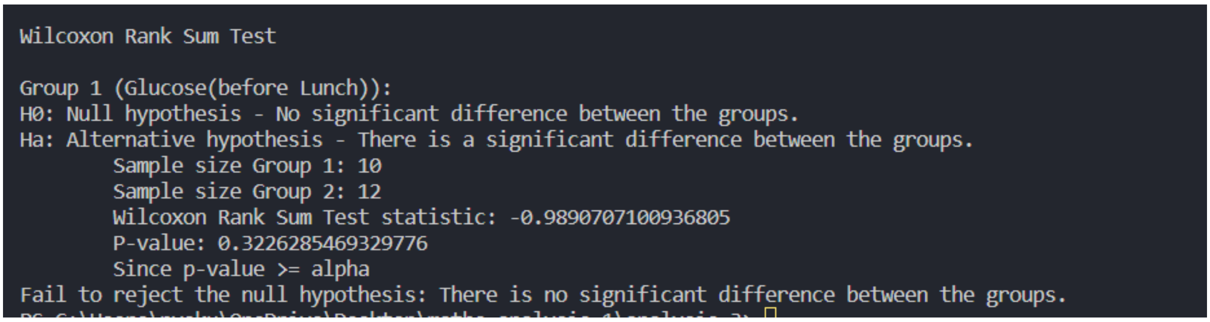
z_cal < z_tab and samplesize > 26:

Accept Ho

3.Wilcoxon Rank Sum Test:

```python
import pandas as pd
import scipy.stats as stats
import random

print("\nWilcoxon Rank Sum Test\n")
df = pd.read_csv('../diabetesdataset.csv')

# Separate your data into two groups (e.g., group1 and group2)
group1 = df['Glucose(before Lunch)']
group2 = df['Glucose(after Lunch)']

print("Group 1 (Glucose(before Lunch)):")
print(f"H0: Null hypothesis - No significant difference between the groups.")
print(f"Ha: Alternative hypothesis - There is a significant difference between the groups.")

sample_size_group1 = 10  # Choose the sample size for group 1
sample_size_group2 = 12  # Choose the sample size for group 2

print("\tSample size Group 1: " + str(sample_size_group1))
sample_group1 = random.sample(group1.tolist(), sample_size_group1)
print("\tSample size Group 2: " + str(sample_size_group2))
sample_group2 = random.sample(group2.tolist(), sample_size_group2)

# Perform the Wilcoxon Rank Sum Test
statistic, p_value = stats.ranksums(sample_group1, sample_group2)

# Output the results
print("\tWilcoxon Rank Sum Test statistic:", statistic)
print("\tP-value:", p_value)

# Interpret the results
alpha = 0.05  # significance level
if p_value < alpha:
    conclusion = (
        f'''\tSince p-value < alpha\n'''
        f"Reject the null hypothesis: There is a significant difference between the groups.")
else:
    conclusion = (
        f'''\tSince p-value >= alpha\n'''
        f"Fail to reject the null hypothesis: There is no significant difference between the groups.")

print(conclusion)
```

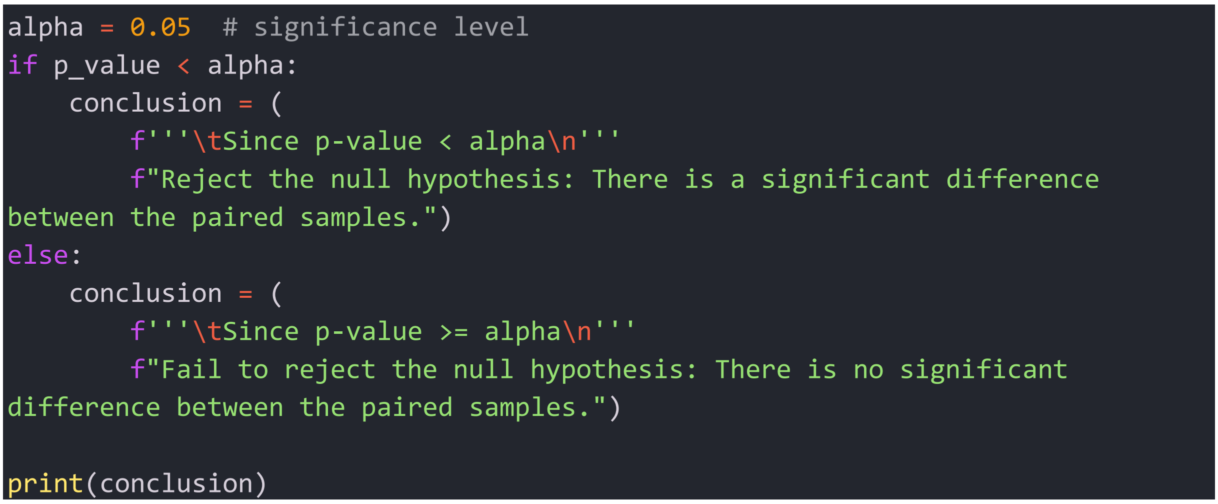
Output:



```
Wilcoxon Rank Sum Test

Group 1 (Glucose(before Lunch)):
H0: Null hypothesis - No significant difference between the groups.
Ha: Alternative hypothesis - There is a significant difference between the groups.
        Sample size Group 1: 10
        Sample size Group 2: 12
        Wilcoxon Rank Sum Test statistic: -0.9890707100936805
        P-value: 0.3226285469329776
        Since p-value >= alpha
Fail to reject the null hypothesis: There is no significant difference between the groups.
```

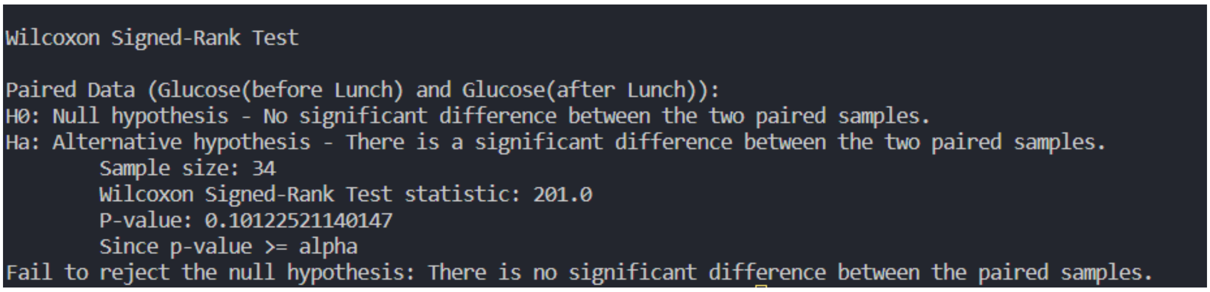
4.Wilcoxon Signed Rank Test(samplesize>30):

```python
import pandas as pd
import scipy.stats as stats
import random

print("\nWilcoxon Signed-Rank Test\n")

df = pd.read_csv('../diabetesdataset.csv')

# Assuming you have a paired dataset with 'Glucose(before Lunch)' and 'Glucose(after Lunch)' columns

print("Paired Data (Glucose(before Lunch) and Glucose(after Lunch)):")
print(f"H0: Null hypothesis - No significant difference between the two paired samples.")
print(f"Ha: Alternative hypothesis - There is a significant difference between the two paired samples.")

sample_size = 34  # Choose the sample size

print("\tSample size: " + str(sample_size))
sample_before = random.sample(df['Glucose(before Lunch)'].tolist(), sample_size)
sample_after = random.sample(df['Glucose(after Lunch)'].tolist(), sample_size)

# Perform the Wilcoxon Signed-Rank Test
statistic, p_value = stats.wilcoxon(sample_before, sample_after)

# Output the results
print("\tWilcoxon Signed-Rank Test statistic:", statistic)
print("\tP-value:", p_value)

# Interpret the results
```



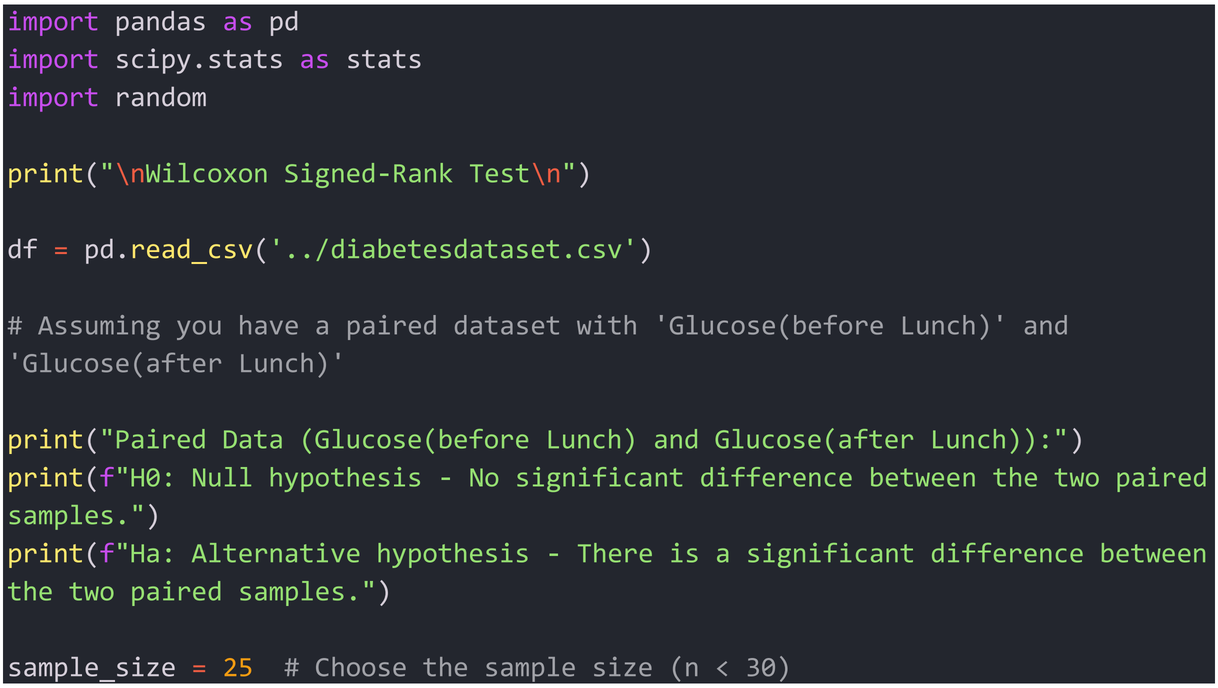
```
alpha = 0.05  # significance level
if p_value < alpha:
    conclusion = (
        f'''\tSince p-value < alpha\n'''
        f"Reject the null hypothesis: There is a significant difference
between the paired samples.")
else:
    conclusion = (
        f'''\tSince p-value >= alpha\n'''
        f"Fail to reject the null hypothesis: There is no significant
difference between the paired samples.")

print(conclusion)
```

output:



```
Wilcoxon Signed-Rank Test

Paired Data (Glucose(before Lunch) and Glucose(after Lunch)):
H0: Null hypothesis - No significant difference between the two paired samples.
Ha: Alternative hypothesis - There is a significant difference between the two paired samples.
        Sample size: 34
        Wilcoxon Signed-Rank Test statistic: 201.0
        P-value: 0.10122521140147
        Since p-value >= alpha
Fail to reject the null hypothesis: There is no significant difference between the paired samples.
```

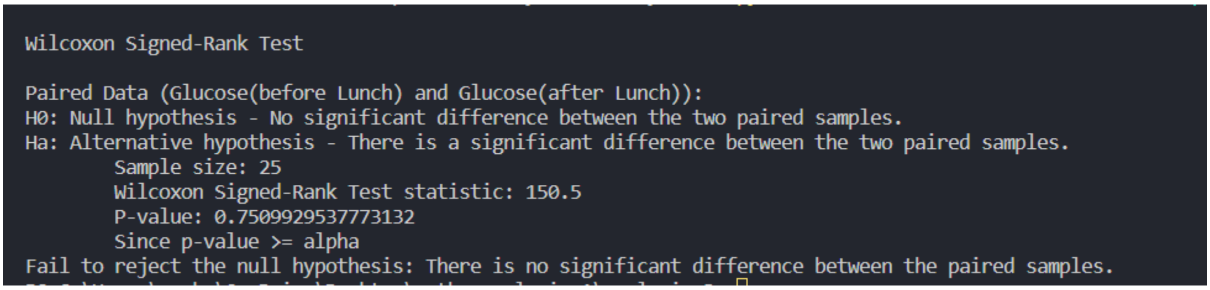
b.Samplesize <= 30:



```
import pandas as pd
import scipy.stats as stats
import random

print("\nWilcoxon Signed-Rank Test\n")

df = pd.read_csv('../diabetesdataset.csv')

# Assuming you have a paired dataset with 'Glucose(before Lunch)' and
'Glucose(after Lunch)'

print("Paired Data (Glucose(before Lunch) and Glucose(after Lunch)):")
print(f"H0: Null hypothesis - No significant difference between the two paired
samples.")
print(f"Ha: Alternative hypothesis - There is a significant difference between
the two paired samples.")

sample_size = 25  # Choose the sample size (n < 30)
```

```
print("\tSample size: " + str(sample_size))

# Sample the data for 'Glucose(before Lunch)' and 'Glucose(after Lunch)'
columns
sample_before = random.sample(df['Glucose(before Lunch)'].tolist(),
sample_size)
sample_after = random.sample(df['Glucose(after Lunch)'].tolist(), sample_size)

# Perform the Wilcoxon Signed-Rank Test
statistic, p_value = stats.wilcoxon(sample_before, sample_after)

# Output the results
print("\tWilcoxon Signed-Rank Test statistic:", statistic)
print("\tP-value:", p_value)

# Interpret the results
alpha = 0.05  # significance level
if p_value < alpha:
    conclusion = (
        f'''\tSince p-value < alpha\n'''
        f"Reject the null hypothesis: There is a significant difference
between the paired samples.")
else:
    conclusion = (
        f'''\tSince p-value >= alpha\n'''
        f"Fail to reject the null hypothesis: There is no significant
difference between the paired samples.")

print(conclusion)
```

output:



```
Wilcoxon Signed-Rank Test

Paired Data (Glucose(before Lunch) and Glucose(after Lunch)):
H0: Null hypothesis - No significant difference between the two paired samples.
Ha: Alternative hypothesis - There is a significant difference between the two paired samples.
        Sample size: 25
        Wilcoxon Signed-Rank Test statistic: 150.5
        P-value: 0.7509929537773132
        Since p-value >= alpha
Fail to reject the null hypothesis: There is no significant difference between the paired samples.
```