

# Statistical hypothesis testing in Biomedical Engineering: Report tutorial – Elsa Doukhan

## A. Galton's stature data: descriptive statistics and Shapiro-Wilk's normality test

The aim of this study is to compare the average heights of 205 women with those of 205 men, using statistical tests. But before doing so, let's start by evaluating certain parameters, such as the distribution of the male and female samples.

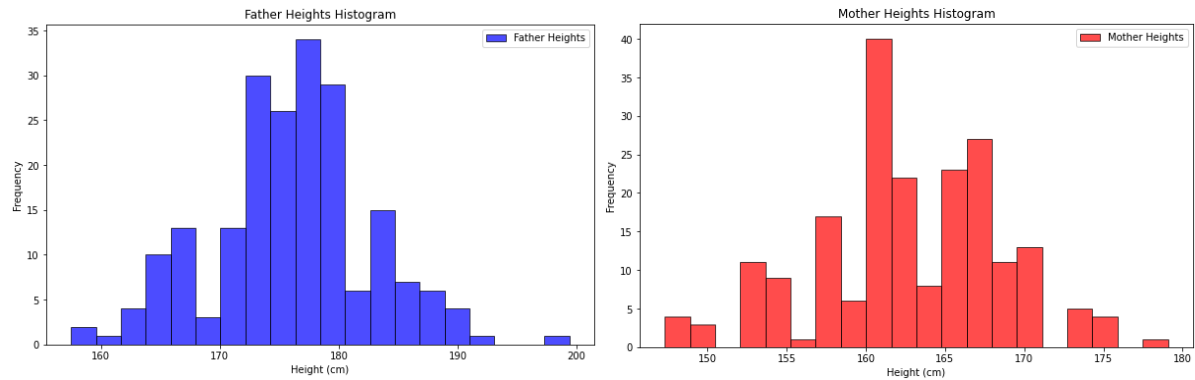


Fig. 1 : Histogram of the father and mothers heights.

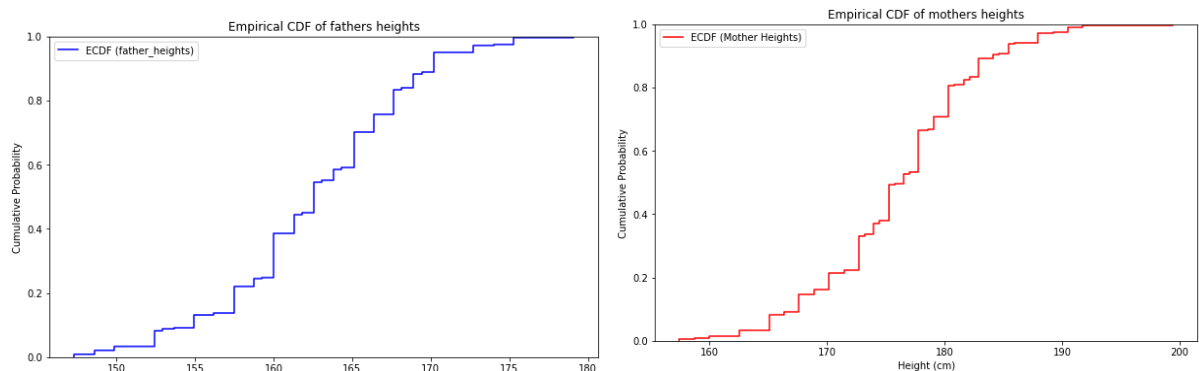


Fig. 2 : Cumulative distribution function for the fathers and mothers.

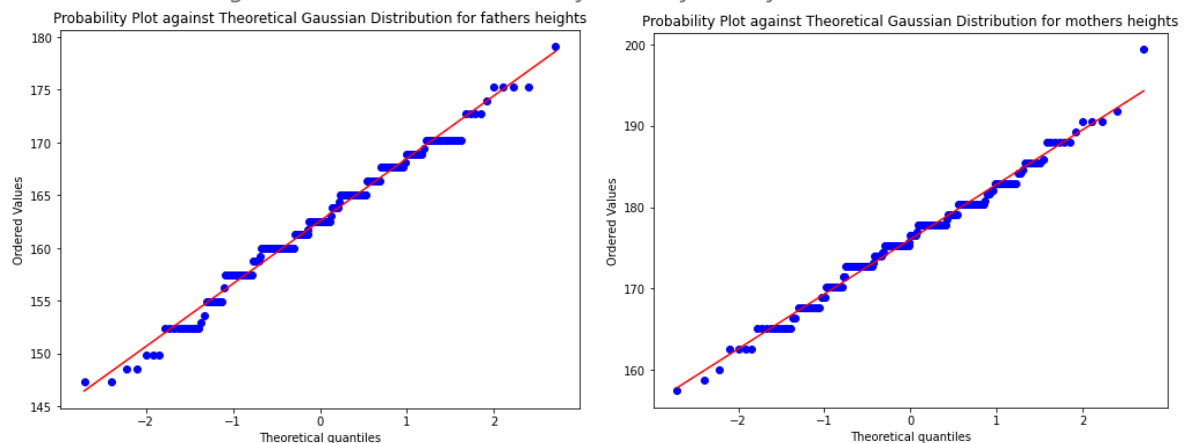


Fig. 3 : Probability plot against the theoretical Gaussian distribution.

The data points closely follow the straight line, it suggests that the data is normally distributed. To confirm this hypothesis, we must do a Shapiro-Wilk test on the two distributions.

Shapiro-Wilk Test for Fathers' Heights:

Statistic\_Shapiro\_Wilk\_test: 0.9865325093269348

p-value\_Shapiro\_Wilk\_test: 0.048770397901535034

Shapiro-Wilk Test for Mothers' Heights:

Statistic\_Shapiro\_Wilk\_test: 0.9888627529144287

p-value\_Shapiro\_Wilk\_test: 0.11137823760509491

The Shapiro-Wilk test reveals that the size distribution of women is normal ( $p\text{-value} = 0.11 > 0.05$ ), but that of men is not ( $p\text{-value} = 0.049 < 0.05$ ). However, the  $p\text{-value}$  of 4.8% is well within the rejection zone, but very close to the  $\alpha = 5\%$  limit. As a precaution, we question the result and decide to carry out statistical tests, considering the following two cases:

Case 1, parametric test : we consider that both tests follow a normal distribution. We then perform a Fisher test to determine whether the variances of the two groups differ significantly. The result of this test indicates that no, we then proceed to a T-test.

Case 2 non-parametric test : As indicated in the Shapiro-Wilk test, the men sample does not follow a normal distribution. We then perform a Wilcoxon Test.

Average, standard deviation, kurtosis, skewness of both heights :

Fathers' Heights:	Mothers' Heights:
Average (mean): 162.56 cm	Average (mean): 176.06 cm
Standard Deviation: 5.91 cm	Standard Deviation: 6.71 cm
Kurtosis: -0.05	Kurtosis: 0.37
Skewness: -0.18	Skewness: 0.01

The standard deviations of both women's and father's heights are close. Nevertheless, to check if the variances statistically differ from each other, we perform the Fisher Test.

Case 1 : At a significance level of 0.05, we fail to reject the null hypothesis. There is no significant evidence to suggest that the variances are different. Therefore, we do a T-test.  
At a significance level of 0.05, we reject the null hypothesis. **There is significant evidence to suggest that the means are different.**

Test Statistic_Fisher_test:	1.9181809860773904
P-value_Fisher_test:	0.16681413789677893
T-statistic_t_test:	21.563245576932317
P-value_t_test:	2.2011851880635784e-69

Case 2 : At a significance level of 0.05, we reject the null hypothesis. **There is significant evidence to suggest a difference in medians.**

Test Statistic_Wilcoxon_test:	0.0
P-value_Wilcoxon_test:	6.669828880443431e-36

The results of the two tests (the T-test and the Wilcoxon test) converge towards the same conclusion: there is a statistically significant difference between the heights of mothers and fathers. This suggests that, in the population from which these samples were taken, the heights of mothers and fathers are very likely to be different.

## **B. Material properties of polymerized NDGA–collagen composite fibers: development of biologically based tendon constructs**

The aim of this study is to compare synthetic fibers with different concentration of NDGA and to find the one that best approximates the stiffness of native tendon fibers, using statistical tests. But before doing so, let's start by creating a boxplot to visually compare all data series and make a table with the appropriate descriptive statistics for each data series.

	0	0.05	0.1	0.25	0.5	1
count	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000
mean	10.136704	53.666782	92.912810	105.658179	146.957743	153.655819
std	22.059740	13.991044	19.053888	9.695555	10.647473	21.553136
min	-21.966931	35.687183	71.115129	97.820490	131.799997	132.186668
25%	0.892078	41.428609	83.906946	101.447231	142.179840	136.481494
50%	15.989700	62.533799	92.384316	101.581784	146.926277	155.719959
75%	18.768206	63.036987	94.351559	105.053851	156.062314	157.299935
max	37.000465	65.647334	122.806100	122.387538	157.820288	186.591038

	2	3	3+0.1	Native tendon fibers
count	5.000000	5.000000	5.000000	5.000000
mean	285.655813	325.889270	500.057812	331.758758
std	21.259405	15.264479	11.030055	14.101664
min	261.869593	303.421956	485.854479	310.709865
25%	274.932750	322.224003	491.725339	325.329523
50%	282.482568	323.579816	504.405303	335.190200
75%	290.231298	338.697031	505.203833	341.530386
max	318.762856	341.523545	513.100108	346.033814

Fig. 4 : Table with the appropriate descriptive statistics for each data series.

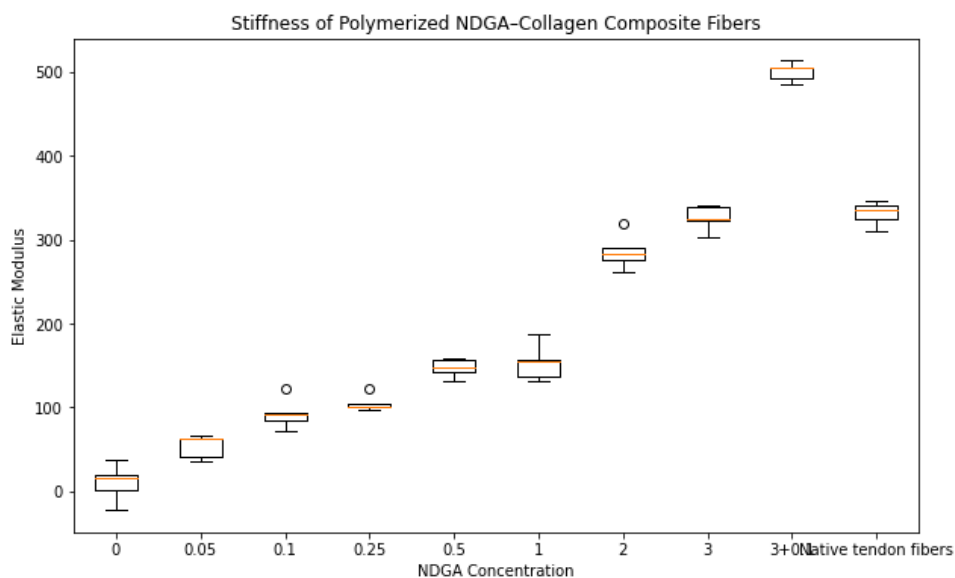


Fig. 5 : Boxplot to visually compare all data series

To compare all data series, we need to perform an analysis of variance (ANOVA) test. Before conducting the ANOVA, we need to check two important assumptions: normality by a Shapiro-Wilk test and homogeneity of variance by a Fisher's test.

Here are the Shapiro-Wilk test results for each fiber sample :

For each sample in our database, we have a p-value greater than 5%, suggesting that there is insufficient evidence to reject the null hypothesis ( $H_0$ : the sample comes from a normally distributed population). In other words, the data can be considered as coming from a normal distribution.

Shapiro-Wilk test for 0: p-valuevalue\_shapiro\_test = 0.8905  
Shapiro-Wilk test for 0.05: p-valuevalue\_shapiro\_test = 0.0913  
Shapiro-Wilk test for 0.1: p-valuevalue\_shapiro\_test = 0.6409  
Shapiro-Wilk test for 0.25: p-valuevalue\_shapiro\_test = 0.0628  
Shapiro-Wilk test for 0.5: p-valuevalue\_shapiro\_test = 0.6650  
Shapiro-Wilk test for 1: p-valuevalue\_shapiro\_test = 0.4846  
Shapiro-Wilk test for 2: p-valuevalue\_shapiro\_test = 0.7581  
Shapiro-Wilk test for 3: p-valuevalue\_shapiro\_test = 0.5370  
Shapiro-Wilk test for 3+0.1: p-valuevalue\_shapiro\_test = 0.6496  
Shapiro-Wilk test for Native tendon fibers: p-valuevalue\_shapiro\_test = 0.6873

Similarly, the Fisher's (Levene) test indicates that there are no significant differences between the variances of the groups. Levene's test for homogeneity of variance: p-value\_levene\_test = 0.9213

Once these conditions have been met, we can apply the ANOVA test. Since we're only comparing means, we apply the 1-factor ANOVA test. According to the results, the p-value is so small that our integrated development environment rounds it down to 0. Therefore, we reject the H0 hypothesis and we accept the H1 hypothesis with a risk  $\alpha = 5\%$  of being wrong. Which means that There is significant evidence to suggest that at least one group mean differs from the others.

One-way ANOVA F-statistic: 435.3457  
P-value\_ANOVA\_test: 0.0000

Thanks to the boxplot, we can easily notice that groups "2 mg/ml" and "3 mg/ml" differ from each other, same for "3 mg/ml" and "3+1 mg/ml". The other way around, "0.5 mg/ml" and "1 mg/ml" are quite close, same for "3 mg/ml" (mean elastic modulus = 325.889270) and "native tendon fibers" (mean elastic modulus = 331.758758). After an ANOVA test, the Tukey test is used to compare the means of each sample in pairs. Applying this test will enable us to confirm or reject our hypotheses based on the boxplot.

group1	group2	meandiff	p-adj	lower	upper	reject
2	3	40.2335	0.0133	5.2953	75.1716	True
3	3+0.1	174.1685	0.0	139.2304	209.1067	True
0.5	1	6.6981	0.9997	-28.2401	41.6363	False
3	Native tendon fibers	5.8695	0.9999	-29.0687	40.8077	False

According to our results after a Tukey test, we confirm our previous hypothesis. Furthermore, the fact that the H0 hypothesis of the Tukey test is rejected for the comparison of samples "3 mg/ml" and "native tendon fibers" shows that sample "3 mg/ml" is the one that best approximates the stiffness of native tendon fibers.

However, the Tukey test is generally used to simultaneously compare multiple groups and determine which groups have equal means. **It controls the overall error rate**, meaning it takes into account the risk of type I error (false discovery) for all comparisons.

The t-test, on the other hand, specifically compares two groups at a time. **It does not control the overall error rate**; it focuses on the comparison between the two groups you specified. Which make sense to conduct t-test especially between sample "3 mg/ml" and "native tendon fibers" in order to do a more thorough analysis.

According to the Shapiro-Wilk test previously conducted, the 3mg/ml and the Native Tendon

Test Statistic\_Fisher\_test: 0.010841516648103856  
P-value\_Fisher\_test: 0.919635113238099

Fibers groups follow a normal distribution, Additionally, for both groups, the Fisher test shows that there is no significant evidence to suggest that the variances are different. Therefore, we do a T-test.

p value > 0.05, it isn't in the rejection zone, it means the means of 3 and native tendon fibers are not significantly different. Which matches with the results of the Tukey test.

T-statistic\_t\_test: -0.6315577178051993  
P-value\_t\_test: 0.545294961152947

Our hypothesis (based on the boxplots) that sample "3 mg/ml" is the one that best approximates the stiffness of native tendon fibers is validated not only by the Tukey test but also by the T-test, which reinforces the robustness of our conclusion. In other words, the concordant results of the two tests reinforce the validity of the conclusion that the means do not differ significantly between the two samples and that this equality is probably not due to chance.

### C. Muscle fitness-based predictors of bone quality

The aim of this study is to find the best predictor of bone quality between two variables : hand grip strength and vertical jump height. For that, we need to test the correlation between trabecular bone mineral content and these two variables. The second step of the study is to evaluate if there is any difference in the correlation of the variables when we lead the study by divide the population into 2 groups : women and men.

The "Trabecular bone mineral content [mg/mm]" and "GreatestHeightDiff [cm]" groups follow a normal distribution, but not the "CombinedGripStrength [N]" group. However, the Pearson correlation between "Trabecular bone mineral content [mg/mm]" and "CombinedGripStrength [N]" is 0.6417. While the one between "Trabecular bone mineral content [mg/mm]" and "GreatestHeightDiff [cm]" is 0.5708.

Shapiro-Wilk test for Trabecular bone mineral content [mg/mm]:  
Test Statistic = 0.9881, p-value = 0.3537  
The data appears to be normally distributed (Fail to reject H0)

Shapiro-Wilk test for CombinedGripStrength [N]:  
Test Statistic = 0.9454, p-value = 0.0001  
The data does not appear to be normally distributed (Reject H0)

Shapiro-Wilk test for GreatestHeightDiff [cm]:  
Test Statistic = 0.9821, p-value = 0.0970  
The data appears to be normally distributed (Fail to reject H0)

Pearson Correlation between Trabecular Bone Mineral Content and Hand Grip Strength: 0.6417, p-value: 0.0000

Pearson Correlation between Trabecular Bone Mineral Content and Vertical Jump Height: 0.5708, p-value: 0.0000

Pearson's correlation applies between two samples following a normal distribution (case for TBMC and GreatestHeightDiff). In the case of TBMC and CombinedGripStrength, the Spearman correlation is more appropriate.

Spearman Correlation Coefficient between Trabecular Bone Mineral Content and Hand Grip Strength: 0.6331, P-value: 0.0000

With Spearman, the correlation coefficient is 0.6331, which is very close to 0.6417 but still lower. We conclude that CombinedGripStrength is the best predictor of bone quality

The scatter plot between mineral density and the best predictor shows a dispersion of points concentrated towards the  $y=x$  curve, highlighting not only that CombinedGripStrength is the best predictor compared to the other, but also that independently of the other, it is good at all.

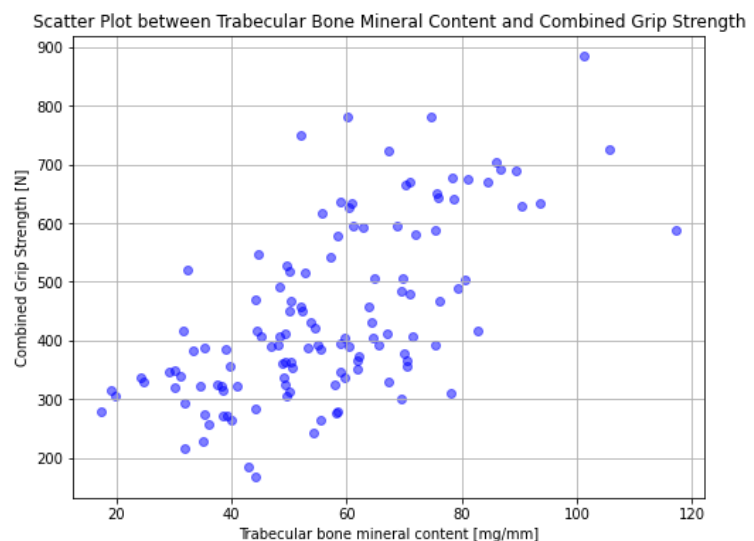


Fig. 6 : Scatter plot between mineral density and the best predictor

However, males and females have different bone quality and risk of osteoporosis. Let's check the correlation between bone mineral content and the best predictor remains significant in the two sexes, and if one sex presents a better prediction.

let's check first the normality of the two groups divided by sexes

Shapiro-Wilk for Trabecular bone mineral content in women - Statistic: 0.9888, p-value: 0.7886  
 Shapiro-Wilk for CombinedGripStrength in women - Statistic: 0.9785, p-value: 0.2714  
 Shapiro-Wilk for Trabecular bone mineral content in men - Statistic: 0.9842, p-value: 0.6803  
 Shapiro-Wilk for CombinedGripStrength in men - Statistic: 0.9786, p-value: 0.4321

Every sample has a normal distribution, we can therefore do a Pearson correlation :

Pearson Correlation between Trabecular Bone Mineral Content and Hand Grip Strength for women: 0.3189, p-value: 0.0071  
 Pearson Correlation between Trabecular Bone Mineral Content and Hand Grip Strength for men: 0.3929, p-value: 0.0030

The correlation is much weaker when subjects are separated by sex, although the correlation for men is slightly higher than for women.

Note : Not better with the other predictor. When the population is separated by sex, the other predictor remains less good.

Test correlation avec Vertical Jump Height :  
 Shapiro-Wilk for GreatestHeightDiff in women - Statistic: 0.9826, p-value: 0.4412  
 Shapiro-Wilk for GreatestHeightDiff in men - Statistic: 0.9892, p-value: 0.9020  
 Pearson Correlation between Trabecular Bone Mineral Content and GreatestHeightDiff for women: 0.2876, p-value: 0.0158  
 Pearson Correlation between Trabecular Bone Mineral Content and GreatestHeightDiff for men: 0.3548, p-value: 0.0079

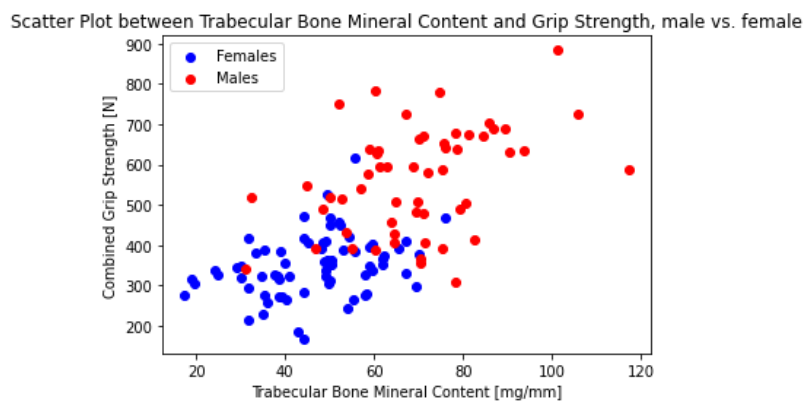


Fig. 7 : Scatter between plot mineral density and the best predictor separating males and females.

According to the scatter plot, men tend to have a higher mineral content than women, however we can note that the men's scatter is more dispersed than for women.

Finally, bone mineral content should decrease with age, but it does not in this dataset. Let's plot a scatter plot between age and mineral content hypothesize why.

As we can see on the scatter plot, our sample is not homogeneously distributed in terms of age, with 108 subjects under 30 years old and 17 over 30. Therefore, our sample is unsatisfactory for assessing the Trabecular bone mineral content as a function of age.

Number of people below 30 years old : 108  
 Number of people above 30 years old : 17

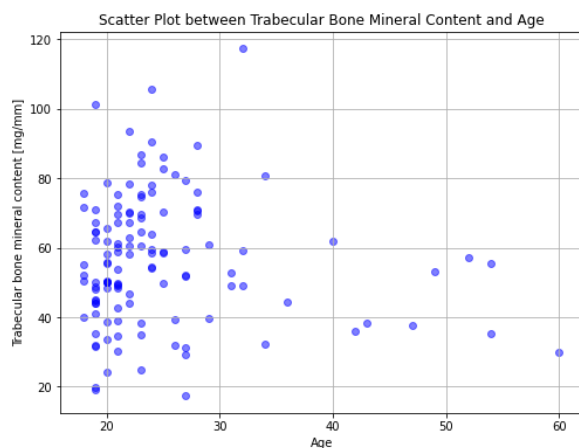


Fig. 8 : Scatter plot between mineral density and the best predictor separating males and females.

## Annexe :

### Exercise A : Python program

```
""" Tutorial for statistics in Python.
```

```
    You can use your own Python install or got to https://trinket.io
```

```
    You will need matplotlib, pandas and scipy packages
```

```
    Good luck!
```

```
"""
```

```
import matplotlib.pyplot as plt
```

```
import pandas as pd
```

```
import scipy as sp
```

```
# Load data
```

```
file = open('./dataA.txt')
```

```
data = pd.read_csv(file, sep = '\t')
```

```
#Question 2
```

```
data['heightFathers'] = data['heightFathers'] * 2.54
```

```
data['heightMothers'] = data['heightMothers'] * 2.54
```

```
#df=pd.DataFrame(data)
```

```
#df
```

```
#Question 3
```

```
# Plot the histogram for father heights
```

```
plt.figure(figsize=(10, 6))
```

```
plt.hist(data['heightFathers'], bins=20, color='blue', edgecolor="black",  
alpha=0.7, label='Father Heights')
```

```
plt.xlabel('Height (cm)')
```

```
plt.ylabel('Frequency')
```

```
plt.title('Father Heights Histogram')
```

```
plt.legend()
```

```
# Plot the histogram for mother heights
```

```
plt.figure(figsize=(10, 6))
```

```
plt.hist(data['heightMothers'], bins=20, color='red', edgecolor="black", alpha=0.7,  
label='Mother Heights')
```

```
plt.xlabel('Height (cm)')
```

```
plt.ylabel('Frequency')
```

```
plt.title('Mother Heights Histogram')
```



```

plt.legend()

# Show the plots
plt.show()

##Question 4 : cumulative function

import matplotlib.pyplot as plt
import numpy as np

# Sort the data
father_heights = np.sort(data['heightMothers'])
mother_heights = np.sort(data['heightFathers'])

new_data = data[['heightMothers', 'heightFathers']]
print(new_data)

import matplotlib.pyplot as plt
import seaborn as sns

# Plot the empirical CDF of mothers heights
plt.figure(figsize=(10, 6))
sns.ecdfplot(mother_heights, color='red', label='ECDF (Mother Heights)')
plt.xlabel('Height (cm)')
plt.ylabel('Cumulative Probability')
plt.title('Empirical CDF of mothers heights')
plt.legend()

# Plot the empirical CDF of fathers heights
plt.figure(figsize=(10, 6))
sns.ecdfplot(father_heights, color='blue', label='ECDF (father_heights)')
plt.xlabel('Height (cm)')
plt.ylabel('Cumulative Probability')
plt.title('Empirical CDF of fathers heights')
plt.legend()

# Show the plot
plt.show()

#Question 5

import scipy.stats as stats

# Create the probability plot
plt.figure(figsize=(8, 6))
res = stats.probplot(mother_heights, plot=plt, dist='norm')
plt.title('Probability Plot against Theoretical Gaussian Distribution for mothers heights')

# Create the probability plot

```



```
plt.figure(figsize=(8, 6))
res = stats.probplot(father_heights, plot=plt, dist='norm')
plt.title('Probability Plot against Theoretical Gaussian Distribution for fathers heights')
```

```
# Show the plot
plt.show()
```

*#the data points closely follow the straight line, it suggests that the data is normally distributed*

*#Question 6*

```
# Perform the Shapiro-Wilk test for mothers' heights
stat_mother, p_value_mother_Shapiro_Wilk_test = stats.shapiro(mother_heights)
print(f'Shapiro-Wilk Test for Mothers\' Heights:\n'
      f'Statistic_Shapiro_Wilk_test: {stat_mother}\n'
      f'p-value_Shapiro_Wilk_test: {p_value_mother_Shapiro_Wilk_test}\n')
```

```
# Perform the Shapiro-Wilk test for fathers' heights
stat_father, p_value_father_Shapiro_Wilk_test = stats.shapiro(father_heights)
print(f'Shapiro-Wilk Test for Fathers\' Heights:\n'
      f'Statistic_Shapiro_Wilk_test: {stat_father}\n'
      f'p-value_Shapiro_Wilk_test: {p_value_father_Shapiro_Wilk_test}\n')
```

*#pour Les hommes,  $0.1 > 0.05$  donc notre hypothèse n'est pas rejetée, Les valeurs suivent une loi normale.*

*#Pour Les femmes,  $0.048 < 0.05$  donc notre hyp est rejetée mais on émet un doute. On fait donc un paramétrique et un non paramétrique teste car*

*#Parametric tests are preferred when the data meets the assumptions. They tend to be more powerful, which means they're better at detecting true effects.*

*#Non-parametric tests are preferred when assumptions are violated (e.g., non-normality, ordinal data) or when you want to be more conservative in your analysis.*

*#Question 7*

```
from scipy.stats import kurtosis, skew
```

```
# Compute the average (mean)
average_mother = np.mean(mother_heights)
average_father = np.mean(father_heights)
```

```
# Compute the standard deviation
std_deviation_mother = np.std(mother_heights)
std_deviation_father = np.std(father_heights)
```

```
# Compute the kurtosis
kurtosis_mother = kurtosis(mother_heights)
kurtosis_father = kurtosis(father_heights)
```

```

# Compute the skewness
skewness_mother = skew(mother_heights)
skewness_father = skew(father_heights)

# Print the results
print("Mothers' Heights:")
print(f"Average (mean): {average_mother:.2f} cm")
print(f"Standard Deviation: {std_deviation_mother:.2f} cm")
print(f"Kurtosis: {kurtosis_mother:.2f}")
print(f"Skewness: {skewness_mother:.2f}\n")

print("Fathers' Heights:")
print(f"Average (mean): {average_father:.2f} cm")
print(f"Standard Deviation: {std_deviation_father:.2f} cm")
print(f"Kurtosis: {kurtosis_father:.2f}")
print(f"Skewness: {skewness_father:.2f}")

# perform a Fisher test to find out if the variances should be equal
#they will be, so we perform a T-test
#on the other hand, wilcon test, both will lead to the same conclusion

#Question 8

#Fisher Test
# Perform Levene's test for homogeneity of variances
statistic, p_value_Fisher_test = stats.levene(mother_heights, father_heights)

# Print the results
print(f"Test Statistic_Fisher_test: {statistic}")
print(f"P-value_Fisher_test: {p_value_Fisher_test}")

# Interpret the results
if p_value_Fisher_test > 0.05:
    print("\nAt a significance level of 0.05, we fail to reject the null hypothesis.")
    print("There is no significant evidence to suggest that the variances are different.")
else:
    print("\nAt a significance level of 0.05, we reject the null hypothesis.")
    print("There is significant evidence to suggest that the variances are different.")

#T-Test
# Perform independent two-sample t-test
t_statistic, p_value_t_test = stats.ttest_ind(mother_heights, father_heights)

# Print the results
print(f"T-statistic_t_test: {t_statistic}")
print(f"P-value_t_test: {p_value_t_test}")

```

```

# Interpret the results
if p_value_t_test > 0.05:
    print("\nAt a significance level of 0.05, we fail to reject the null hypothesis.")
    print("There is no significant evidence to suggest that the means are different.")
else:
    print("\nAt a significance level of 0.05, we reject the null hypothesis.")
    print("There is significant evidence to suggest that the means are different.")

# Wilcoxon test
# Perform Wilcoxon signed-rank test (paired)
statistic, p_value_Wilcoxon_test = stats.wilcoxon(mother_heights, father_heights)

# Print the results
print(f"Test Statistic_Wilcoxon_test: {statistic}")
print(f"P-value_Wilcoxon_test: {p_value_Wilcoxon_test}")

# Interpret the results
if p_value_Wilcoxon_test > 0.05:
    print("\nAt a significance level of 0.05, we fail to reject the null hypothesis.")
    print("There is no significant evidence to suggest a difference in medians.")
else:
    print("\nAt a significance level of 0.05, we reject the null hypothesis.")
    print("There is significant evidence to suggest a difference in medians.")

```

## Exercise B : Python program

*#Exercise B*

*#Question 1*

```

import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

# Load the data
file_path = './dataB.txt'
data = pd.read_csv(file_path, sep='\t')

# Create a boxplot
plt.figure(figsize=(10, 6))
plt.boxplot(data.values, labels=data.columns)
plt.title('Stiffness of Polymerized NDGA-Collagen Composite Fibers')
plt.xlabel('NDGA Concentration')
plt.ylabel('Elastic Modulus')

```

```

plt.show()

# Generate descriptive statistics
descriptive_stats = data.describe()
print(descriptive_stats)

#Question 2

import scipy.stats as stats

# Checking normality
for column in data.columns:
    stat, p_value_shapiro_test = stats.shapiro(data[column])
    print(f"Shapiro-Wilk test for {column}: p-value={p_value_shapiro_test} = {p_value_shapiro_test:.4f}")
# Checking homogeneity of variance
statistic, p_value_levene_test = stats.levene(*[data[column] for column in data.columns])
print(f"\nLevene's test for homogeneity of variance: p-value={p_value_levene_test} = {p_value_levene_test:.4f}")

import scipy.stats as stats

# Perform one-way ANOVA
f_statistic, p_value_ANOVA_test = stats.f_oneway(*[data[column] for column in data.columns])

# Print the results
print(f"One-way ANOVA F-statistic: {f_statistic:.4f}")
print(f"P-value ANOVA test: {p_value_ANOVA_test:.4f}")

# Interpret the results
if p_value_ANOVA_test > 0.05:
    print("\nAt a significance level of 0.05, we fail to reject the null hypothesis.")
    print("There is no significant evidence to suggest that at least one group mean differs from the others.")
else:
    print("\nAt a significance level of 0.05, we reject the null hypothesis.")
    print("There is significant evidence to suggest that at least one group mean differs from the others.")

#Question 3
from statsmodels.stats.multicomp import pairwise_tukeyhsd

# Reshape data for Tukey's test
df = data.melt(var_name='NDGA Concentration', value_name='Elastic Modulus')
# Perform Tukey's test
tukey_results = pairwise_tukeyhsd(endog=df['Elastic Modulus'], groups=df['NDGA Concentration'], alpha=0.05)

```

```

# Print the results
print(tukey_results)

#Question 4

# Calculate mean stiffness for each concentration of NDGA
mean_stiffness_by_concentration = data.mean()
print(mean_stiffness_by_concentration)

#We note that the mean elastic modulus for the group receiving
#3mg/ml NDGA (325.889270) is closest to the mean for native tendon fibers
(331.758758).
#It would therefore be interesting to statistically compare the averages of these
two groups.
#To determine the most appropriate test, we need to check several criteria:
#According to the Shapiro-Wilk test, the 3mg/ml groups follow a normal
distribution, so we apply either the t-test or the welch test.
#To determine which of the two tests to apply, we look at the variances. For both
groups, the variances are close
# std: 15264479 for 3mg/ml group and std: 14.101664 for native tendon fibers, but
to confirm that there are no significant differences, we perform a Fisher test.

#Fisher Test
# Perform Levene's test for homogeneity of variances
group_3mgml = data['3']
native_tendon_fibers = data['Native tendon fibers']

statistic, p_value_Fisher_test = stats.levene(group_3mgml, native_tendon_fibers)

# Print the results
print(f"Test Statistic_Fisher_test: {statistic}")
print(f"P-value_Fisher_test: {p_value_Fisher_test}")

# P-value_Fisher_test: 0.919635113238099, There is no significant evidence to
suggest a difference between the variances.
# we can undertake a t-test.

#T-Test
# Perform independent two-sample t-test
t_statistic, p_value_t_test = stats.ttest_ind(group_3mgml, native_tendon_fibers)

# Print the results
print(f"T-statistic_t_test: {t_statistic}")
print(f"P-value_t_test: {p_value_t_test}")

# Interpret the results
if p_value_t_test > 0.05:
    print("\nAt a significance level of 0.05, we fail to reject the null
hypothesis.")
    print("There is no significant evidence to suggest that the means are
different.")

```

```

else:
    print("\nAt a significance level of 0.05, we reject the null hypothesis.")
    print("There is significant evidence to suggest that the means are
different.")

```

## Exercise C : Python program

*#Exercise C*

*#Question 1*

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

file_path = './dataC.txt'
data = pd.read_csv(file_path, sep='\t')

# Assuming you have a DataFrame 'data' with columns:
# 'Trabecular bone mineral content [mg/mm]', 'CombinedGripStrength [N]',
# 'GreatestHeightDiff [cm]'

# Check for normality using histograms
plt.figure(figsize=(15, 5))

plt.subplot(1, 3, 1)
sns.histplot(data['Trabecular bone mineral content [mg/mm]'], kde=True)
plt.title('Trabecular bone mineral content [mg/mm]')

plt.subplot(1, 3, 2)
sns.histplot(data['CombinedGripStrength [N]'], kde=True)
plt.title('CombinedGripStrength [N]')

plt.subplot(1, 3, 3)
sns.histplot(data['GreatestHeightDiff [cm]'], kde=True)
plt.title('GreatestHeightDiff [cm]')

plt.tight_layout()
plt.show()

from scipy.stats import shapiro

# Assuming you have a DataFrame 'data' with the specified columns

columns_to_test = ["Trabecular bone mineral content [mg/mm]",
"CombinedGripStrength [N]", "GreatestHeightDiff [cm]"]

for column in columns_to_test:
    stat, p_value = shapiro(data[column])
    print(f"Shapiro-Wilk test for {column}:")

```

```

print(f"Test Statistic = {stat:.4f}, p-value = {p_value:.4f}")

if p_value > 0.05:
    print("The data appears to be normally distributed (Fail to reject H0)")
else:
    print("The data does not appear to be normally distributed (Reject H0)")

print("\n")

# Check for outliers using box plots
plt.figure(figsize=(15, 5))

plt.subplot(1, 3, 1)
sns.boxplot(data['Trabecular bone mineral content [mg/mm]'])
plt.title('Trabecular bone mineral content [mg/mm]')

plt.subplot(1, 3, 2)
sns.boxplot(data['CombinedGripStrength [N]'])
plt.title('CombinedGripStrength [N]')

plt.subplot(1, 3, 3)
sns.boxplot(data['GreatestHeightDiff [cm]'])
plt.title('GreatestHeightDiff [cm]')
plt.tight_layout()
plt.show()

#Ces graphiques mettent en évidence une corrélation entre Trabecular bone mineral
content [mg/mm] et GreatestHeightDiff [cm]

from scipy.stats import pearsonr

# Calculate Pearson correlation coefficients
corr_hand_grip = pearsonr(data['Trabecular bone mineral content [mg/mm]'],
data['CombinedGripStrength [N]'])
corr_jump_height = pearsonr(data['Trabecular bone mineral content [mg/mm]'],
data['GreatestHeightDiff [cm]'])

print(f"Pearson Correlation between Trabecular Bone Mineral Content and Hand Grip
Strength: {corr_hand_grip[0]:.4f}, p-value: {corr_hand_grip[1]:.4f}")
print(f"Pearson Correlation between Trabecular Bone Mineral Content and Vertical
Jump Height: {corr_jump_height[0]:.4f}, p-value: {corr_jump_height[1]:.4f}")

#The "Trabecular bone mineral content [mg/mm]" and "GreatestHeightDiff [cm]"
groups follow a normal distribution,
#but not the "CombinedGripStrength [N]" group.
#However, the Pearson correlation between "Trabecular bone mineral content
[mg/mm]" and "CombinedGripStrength [N]" is 0.6417.
#while that between "Trabecular bone mineral content [mg/mm]" and
"GreatestHeightDiff [cm]" is 0.5708.
#Pearson's correlation applies between two samples following a normal distribution
(case for TBMC and GreatestHeightDiff).

```



*#In the case of TBMC and CombinedGripStrength, the Spearman correlation is more appropriate.*

*# Calculate Spearman correlation coefficients (to compare correlation between a normal and an anormal distribution)*

```
from scipy.stats import spearmanr
```

*# Assuming 'data' is your DataFrame and you want to correlate 'Trabecular bone mineral content [mg/mm]' and 'CombinedGripStrength [N]'*

*# 'Trabecular bone mineral content [mg/mm]' follows a normal distribution, while 'CombinedGripStrength [N]' does not*

*# Calculate Spearman correlation coefficient and p-value*

```
corr_coefficient, p_value = spearmanr(data['Trabecular bone mineral content [mg/mm]'], data['CombinedGripStrength [N]'])
```

```
print(f"Spearman Correlation Coefficient between Trabecular Bone Mineral Content and Hand Grip Strength: {corr_coefficient:.4f}, P-value: {p_value:.4f}")
```

*#With Spearman, the correlation coefficient is 0.6331, which is very close to 0.6417 but still lower.*

*#We conclude that CombinedGripStrength is the best predictor of bone quality*

*#Question 2*

```
import matplotlib.pyplot as plt
```

*# Scatter plot between Trabecular bone mineral content and the best predictor*

```
plt.figure(figsize=(8, 6))
```

```
plt.scatter(data['Trabecular bone mineral content [mg/mm]'],
```

```
data['CombinedGripStrength [N]'], alpha=0.5, color='blue')
```

```
plt.title('Scatter Plot between Trabecular Bone Mineral Content and Combined Grip Strength')
```

```
plt.xlabel('Trabecular bone mineral content [mg/mm]')
```

```
plt.ylabel('Combined Grip Strength [N]')
```

```
plt.grid(True)
```

```
plt.show()
```

*#Question 3*

```
import pandas as pd
```

*# Separate datas in two groups : female (F) and male (M)*

```
women_data = data[data['Gender'] == 'F']
```

```
men_data = data[data['Gender'] == 'M']
```

*# Let's check the normality of the two groups divided by sexes*

*# For women*

*# Shapiro-Wilk test for 'Trabecular bone mineral content [mg/mm]'*

```

statistic, p_value = shapiro(women_data['Trabecular bone mineral content
[mg/mm]'])
print(f"Shapiro-Wilk for Trabecular bone mineral content in women - Statistic:
{statistic:.4f}, p-value: {p_value:.4f}")

# Shapiro-Wilk test for 'CombinedGripStrength [N]'
statistic, p_value = shapiro(women_data['CombinedGripStrength [N]'])
print(f"Shapiro-Wilk for CombinedGripStrength in women - Statistic:
{statistic:.4f}, p-value: {p_value:.4f}")
# For men
# Shapiro-Wilk test for 'Trabecular bone mineral content [mg/mm]'
statistic, p_value = shapiro(men_data['Trabecular bone mineral content [mg/mm]'])
print(f"Shapiro-Wilk for Trabecular bone mineral content in men - Statistic:
{statistic:.4f}, p-value: {p_value:.4f}")

# Shapiro-Wilk test for 'CombinedGripStrength [N]'
statistic, p_value = shapiro(men_data['CombinedGripStrength [N]'])
print(f"Shapiro-Wilk for CombinedGripStrength in men - Statistic: {statistic:.4f},
p-value: {p_value:.4f}")

#Every groups are normally distributed, we can calculate the Pearson correlation
coefficients

# Calculate Pearson correlation coefficients
correlation_women = pearsonr(women_data['Trabecular bone mineral content
[mg/mm]'], women_data['CombinedGripStrength [N]'])
correlation_men = pearsonr(men_data['Trabecular bone mineral content [mg/mm]'],
men_data['CombinedGripStrength [N]'])

print(f"Pearson Correlation between Trabecular Bone Mineral Content and Hand Grip
Strength for women: {correlation_women[0]:.4f}, p-value:
{correlation_women[1]:.4f}")
print(f"Pearson Correlation between Trabecular Bone Mineral Content and Hand Grip
Strength for men: {correlation_men[0]:.4f}, p-value: {correlation_men[1]:.4f}")

# Test correlation avec Vertical Jump Height
print("Test correlation avec Vertical Jump Height :")
#for women
# Shapiro-Wilk test for 'CombinedGripStrength [N]'
statistic, p_value = shapiro(women_data['GreatestHeightDiff [cm]'])
print(f"Shapiro-Wilk for GreatestHeightDiff in women - Statistic: {statistic:.4f},
p-value: {p_value:.4f}")
# => normal distribution

#for men
# Shapiro-Wilk test for 'CombinedGripStrength [N]'
statistic, p_value = shapiro(men_data['GreatestHeightDiff [cm]'])
print(f"Shapiro-Wilk for GreatestHeightDiff in men - Statistic: {statistic:.4f},
p-value: {p_value:.4f}")
# => normal distribution
# Calculate Pearson correlation coefficients

```

```
correlation_women = pearsonr(women_data['Trabecular bone mineral content
[mg/mm]'], women_data['GreatestHeightDiff [cm]'])
correlation_men = pearsonr(men_data['Trabecular bone mineral content [mg/mm]'],
men_data['GreatestHeightDiff [cm]'])
```

```
print(f"Pearson Correlation between Trabecular Bone Mineral Content and
GreatestHeightDiff for women: {correlation_women[0]:.4f}, p-value:
{correlation_women[1]:.4f}")
print(f"Pearson Correlation between Trabecular Bone Mineral Content and
GreatestHeightDiff for men: {correlation_men[0]:.4f}, p-value:
{correlation_men[1]:.4f}")
```

*#Question 4*

*# Scatter plot for females*

```
plt.scatter(women_data['Trabecular bone mineral content [mg/mm]'],
women_data['CombinedGripStrength [N]'], label='Females', color='blue')
```

*# Scatter plot for males*

```
plt.scatter(men_data['Trabecular bone mineral content [mg/mm]'],
men_data['CombinedGripStrength [N]'], label='Males', color='red')
```

*# Labels and title*

```
plt.xlabel('Trabecular Bone Mineral Content [mg/mm]')
plt.ylabel('Combined Grip Strength [N]')
plt.title('Scatter Plot between Trabecular Bone Mineral Content and Grip Strength,
male vs. female')
```

*# Legend*

```
plt.legend()
```

*# Show plot*

```
plt.show()
```

*#5) Bone mineral content should decrease with age, but it does not in this dataset. Looking at the*

*#scatter plot between age and mineral content, can you hypothesize why?*

*# Initialisation des compteurs*

```
below_30 = 0
```

```
above_30 = 0
```

*# Boucle pour parcourir les données et compter les personnes*

```
for age in data['Age']:
```

```
    if age < 30:
```

```
        below_30 += 1
```

```
    else:
```

```
        above_30 += 1
```

*# Affichage des résultats*

```
print(f"Number of people below 30 years old : {below_30}")  
print(f"Number of people above 30 years old : {above_30}")
```

*#Answer, Because our sample is nothomogeneously distributed in terms of age, with 108 subjects under 30 and 17 over 30.*

*#Our sample is therefore unsatisfactory for assessing the Trabecular bone mineral content as a function of age.*