

## ASSIGNMENT-04

## STATISTICAL INFERENCE IN BIOENGINEERING

P TEJA VENKATA RAMANA KUMAR

BM23MTECH11006

## Q1:-

```
import pandas as pd

df=pd.read_excel('Assignment4 data.xlsx',sheet_name='Cats')
df
```

	Sex	Bwt	Hwt
0	F	2.0	7.0
1	F	2.0	7.4
2	F	2.0	9.5
3	F	2.1	7.2
4	F	2.1	7.3
...	...	...	...
139	M	3.7	11.0
140	M	3.8	14.8
141	M	3.8	16.8
142	M	3.9	14.4
143	M	3.9	20.5

```
[144 rows x 3 columns]
```

```
# Summary statistics for numerical columns
df.describe()
```

	Bwt	Hwt
count	144.000000	144.000000
mean	2.723611	10.630556
std	0.485307	2.434636
min	2.000000	6.300000
25%	2.300000	8.950000
50%	2.700000	10.100000
75%	3.025000	12.125000
max	3.900000	20.500000

```

# Data understanding
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144 entries, 0 to 143
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype  
---  -
 0   Sex      144 non-null     object  
 1   Bwt      144 non-null     float64  
 2   Hwt      144 non-null     float64  
dtypes: float64(2), object(1)
memory usage: 3.5+ KB

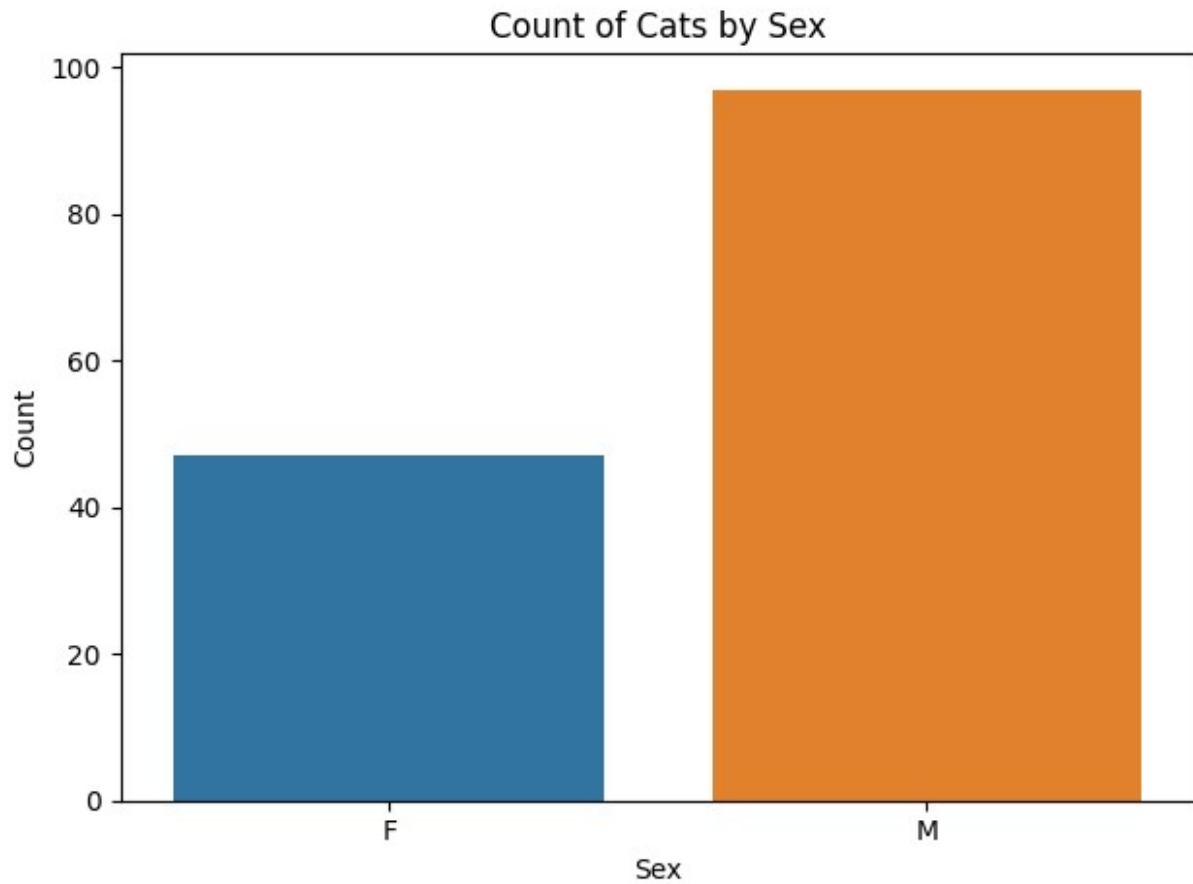
# No Null values
# 1 is describe 2 is continous

import matplotlib.pyplot as plt
import seaborn as sns

# Bar plot for Sex (categorical variable)
plt.figure()
sns.countplot(x='Sex', data=df)
plt.title('Count of Cats by Sex')
plt.xlabel('Sex')
plt.ylabel('Count')
plt.tight_layout()

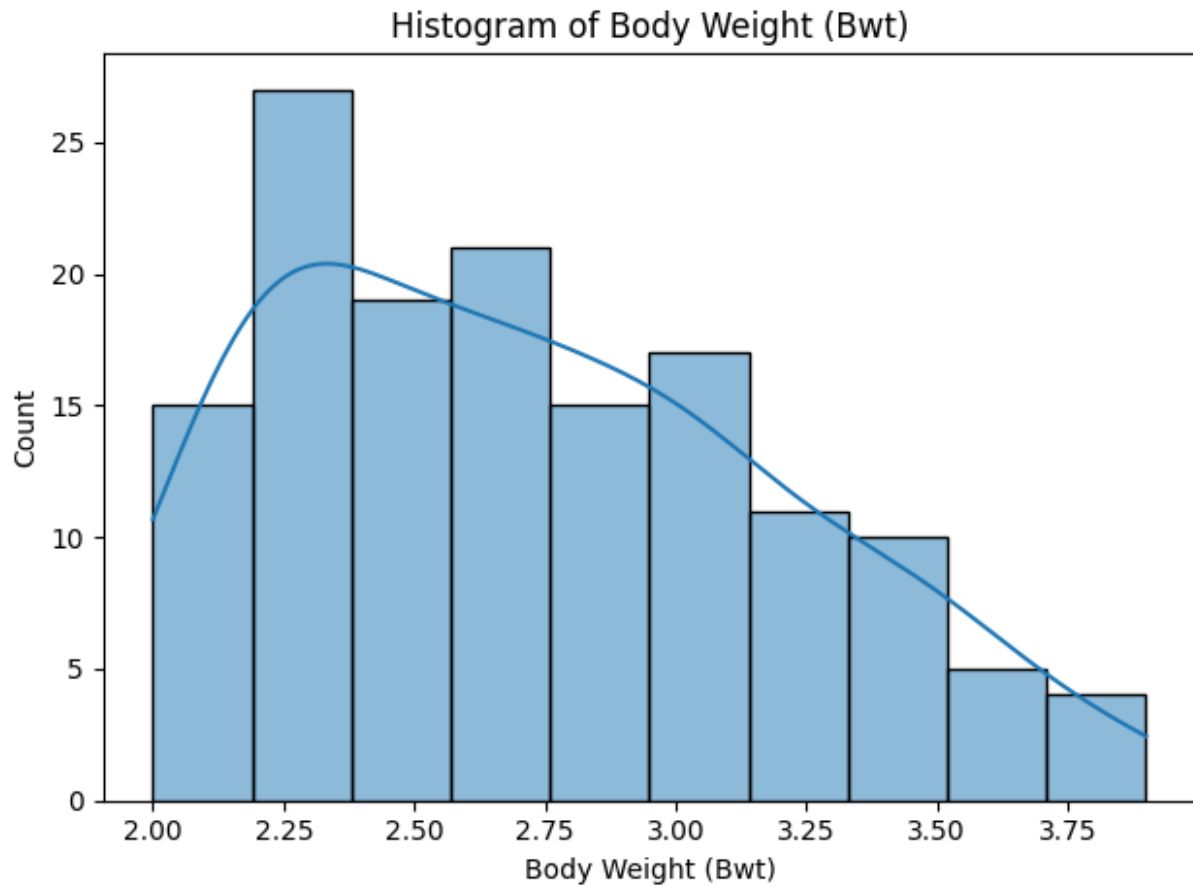
plt.show()

```



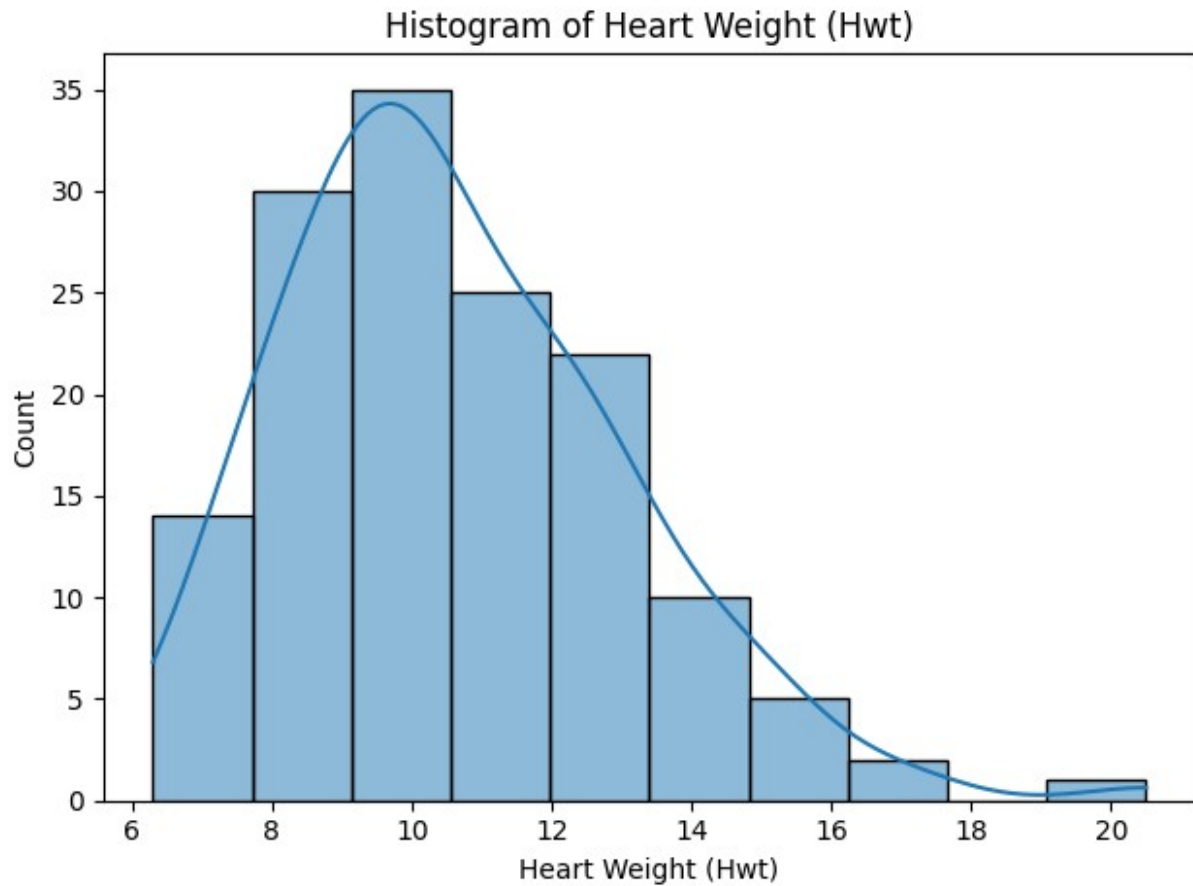
```
# Histogram for Body Weight (Bwt)
plt.figure()
sns.histplot(df['Bwt'], bins=10, kde=True)
plt.title('Histogram of Body Weight (Bwt)')
plt.xlabel('Body Weight (Bwt)')
plt.ylabel('Count')
plt.tight_layout()

plt.show()
```



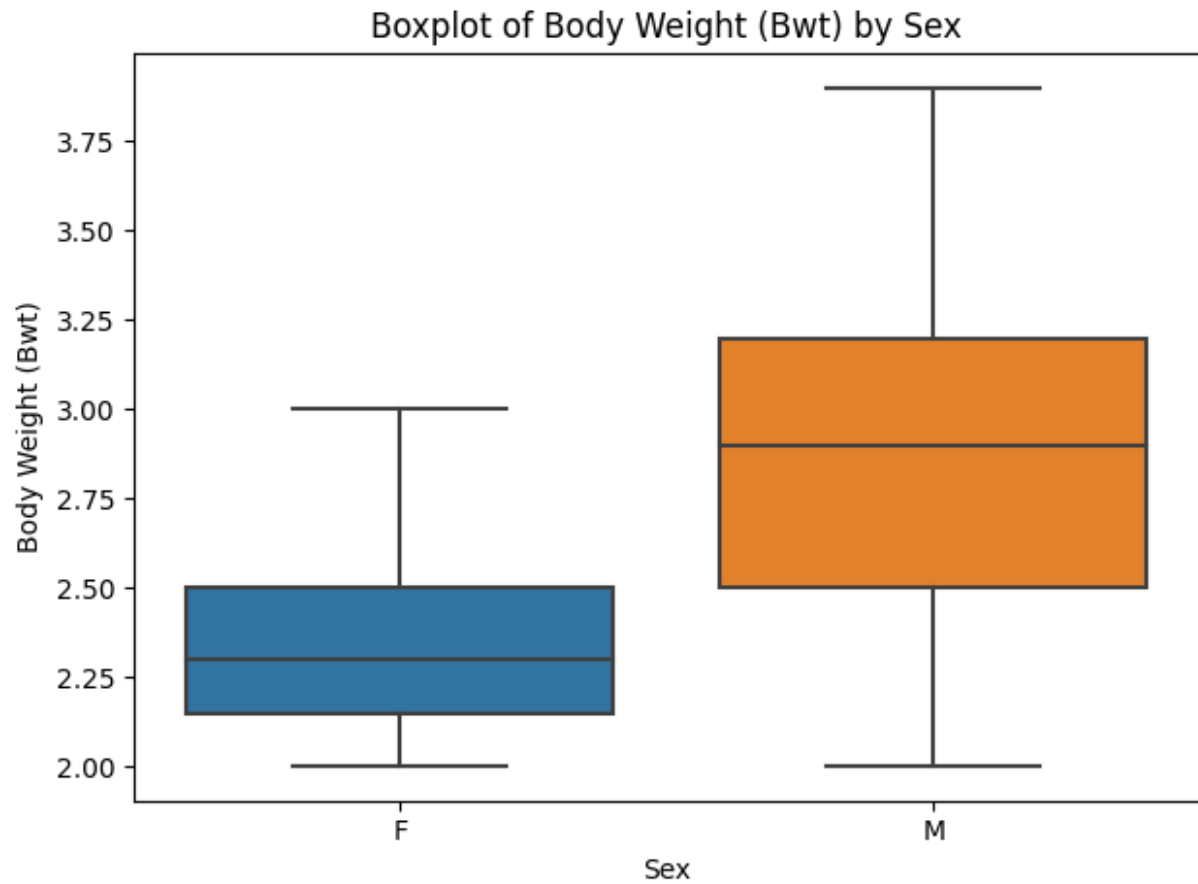
```
# Histogram for Heart Weight (Hwt)
plt.figure()
sns.histplot(df['Hwt'], bins=10, kde=True)
plt.title('Histogram of Heart Weight (Hwt)')
plt.xlabel('Heart Weight (Hwt)')
plt.ylabel('Count')
plt.tight_layout()

plt.show()
```



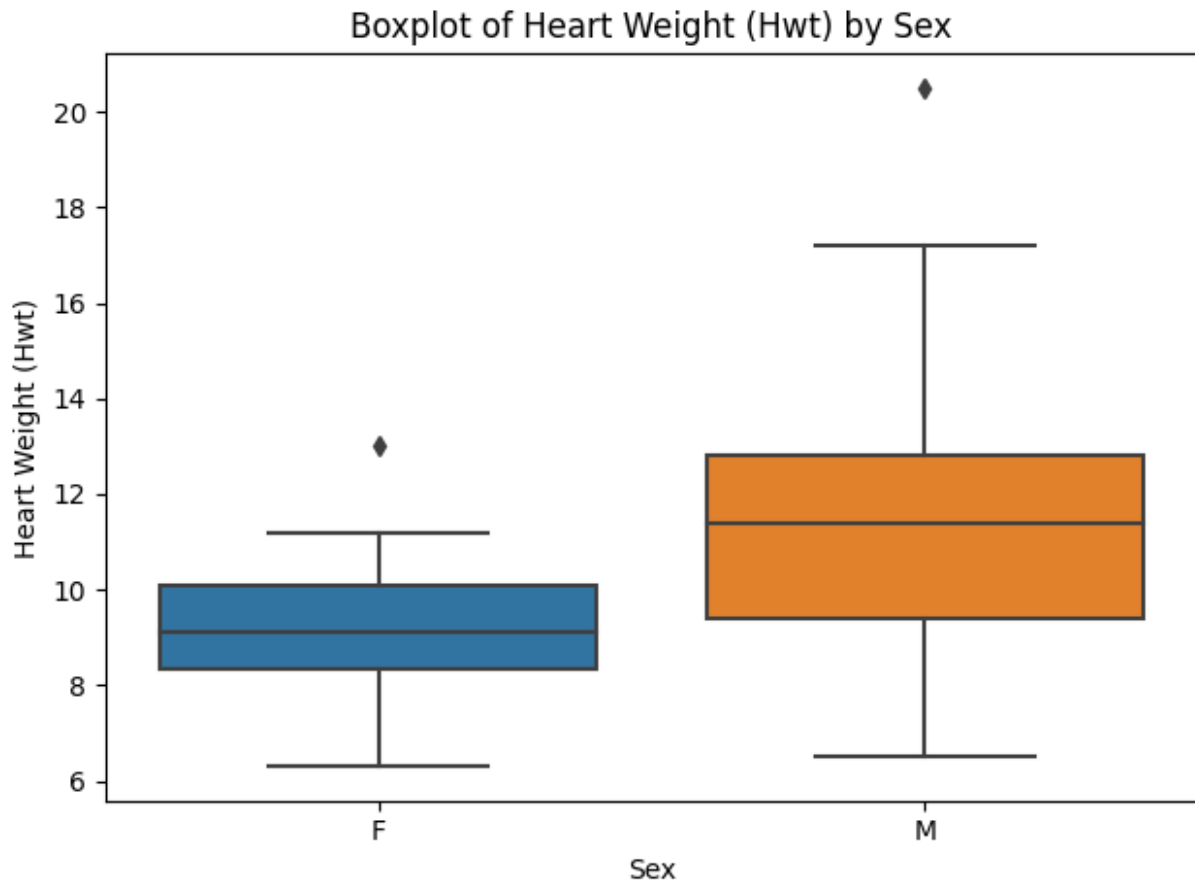
```
# Boxplot for Body Weight (Bwt) by Sex
plt.figure()
sns.boxplot(x='Sex', y='Bwt', data=df)
plt.title('Boxplot of Body Weight (Bwt) by Sex')
plt.xlabel('Sex')
plt.ylabel('Body Weight (Bwt)')
plt.tight_layout()

plt.show()
```



```
# Boxplot for Heart Weight (Hwt) by Sex
plt.figure()
sns.boxplot(x='Sex', y='Hwt', data=df)
plt.title('Boxplot of Heart Weight (Hwt) by Sex')
plt.xlabel('Sex')
plt.ylabel('Heart Weight (Hwt)')
plt.tight_layout()

plt.show()
```

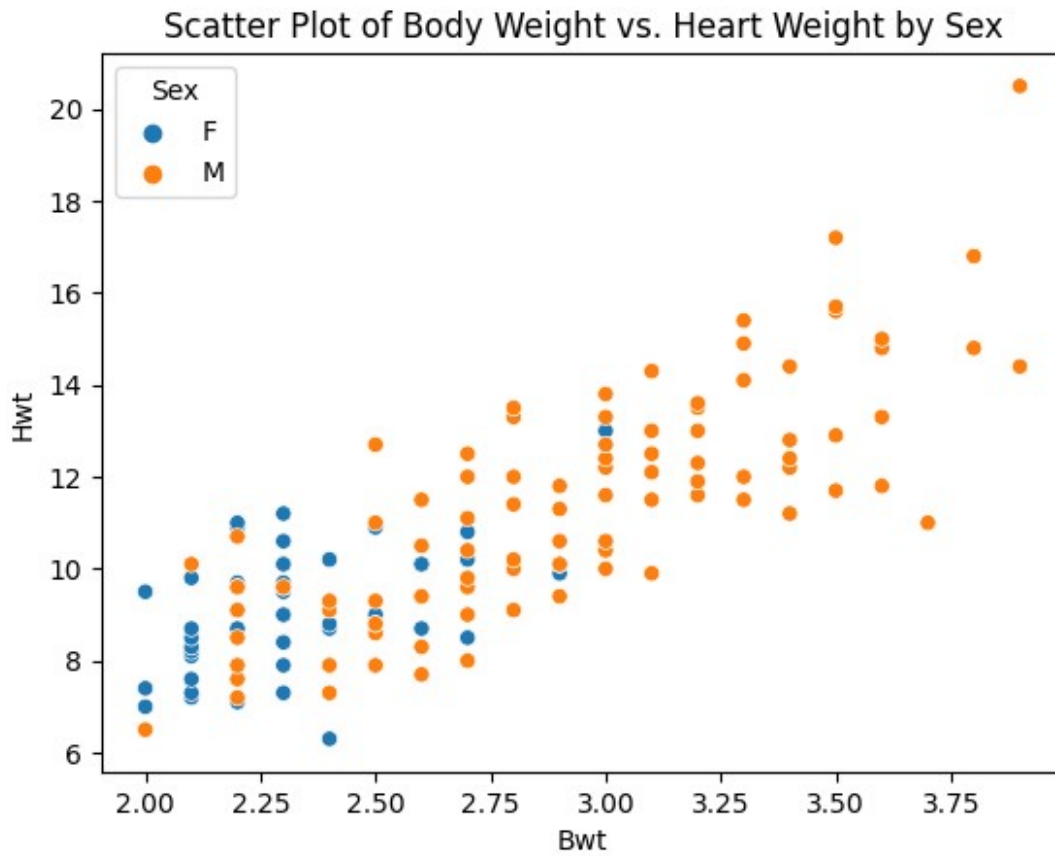


```
# Plot 4: Violin plot of Heart Weight (Hwt) by Sex
plt.figure(figsize=(8, 6))
sns.violinplot(data=df, x='Sex', y='Hwt', inner='quart')
plt.title('Violin Plot of Heart Weight (Hwt) by Sex')
plt.xlabel('Sex')
plt.ylabel('Heart Weight (kg)')
plt.show()
```



```
# Scatter plot of Body Weight (Bwt) vs. Heart Weight (Hwt) colored by Sex
sns.scatterplot(x=df['Bwt'], y=df['Hwt'], hue=df['Sex'])
plt.title('Scatter Plot of Body Weight vs. Heart Weight by Sex')
plt.show()
```





```

"""Null Hypothesis (H0): There is no significant difference in mean
body weight (Bwt)
        between male and female cats.
Alternative Hypothesis (H1): There is a significant difference in
mean body weight (Bwt)
        between male and female cats.

```

```

Statistical Test: Two-sample t-test for independent samples

```

```

    """

```

```

from scipy.stats import ttest_ind

male_bwt = df[df['Sex'] == 'M']['Bwt']
female_bwt = df[df['Sex'] == 'F']['Bwt']

# Perform a two-sample t-test to compare means
t_stat_bwt, p_value_bwt = ttest_ind(male_bwt, female_bwt)
print("t-Test for Body Weight (Bwt):")
print("t-Statistic:", t_stat_bwt)
print("p-Value:", p_value_bwt)

```

```
t-Test for Body Weight (Bwt):  
t-Statistic: 7.330666826537409  
p-Value: 1.590445378625505e-11
```

```
"""Null Hypothesis (H0): There is no significant difference in mean  
heart weight (Hwt)
```

```
    between male and female cats.
```

```
    Alternative Hypothesis (H1): There is a significant difference in  
mean heart weight (Hwt)
```

```
    between male and female cats.
```

```
    Statistical Test: Two-sample t-test for independent samples
```

```
    """
```

```
    {"type": "string"}
```

```
male_hwt = df[df['Sex'] == 'M']['Hwt']  
female_hwt = df[df['Sex'] == 'F']['Hwt']
```

```
# Perform a two-sample t-test to compare means
```

```
t_stat_hwt, p_value_hwt = ttest_ind(male_hwt, female_hwt)
```

```
print("\nt-Test for Heart Weight (Hwt):")  
print("t-Statistic:", t_stat_hwt)  
print("p-Value:", p_value_hwt)
```

```
t-Test for Heart Weight (Hwt):  
t-Statistic: 5.353924104107808  
p-Value: 3.379785752600854e-07
```

```
# conclusion
```

```
if p_value_hwt < 0.05:  
    print(f"Null hypotheses rejected.\nThere is a significant  
difference in mean heart weight (Hwt) between male and female  
cats ." )  
else:  
    print(f"Null hypotheses can not be rejected.\nThere is no  
significant difference in mean heart weight (Hwt) between male and  
female cats." )
```

```
if p_value_bwt < 0.05:  
    print(f"Null hypotheses rejected.\nThere is a significant  
difference in mean body weight (Bwt) between male and female cats ." )  
else:  
    print(f"Null hypotheses can not be rejected.\nThere is no
```

significant difference in mean body weight (Bwt) between male and female cats." )

Null hypotheses rejected.  
.There is a significant difference in mean heart weight (Hwt) between male and female cats .  
Null hypotheses rejected.  
.There is a significant difference in mean body weight (Bwt) between male and female cats .

**The statistical test indicates a dependency between the body weight and heart weight of the cat on gender. Therefore, it can be inferred that there is a positive correlation between body weights and heart weight.**

30

**Q2:-**

*what about association between body & heart weight?*

```
data=pd.read_excel('Assignment4 data.xlsx',sheet_name='Labo')
data
```

	Unnamed: 0	Lab	Spc	Bat	Conc
0	1	L1	S1	B1	0.29
1	2	L1	S1	B1	0.33
2	3	L1	S1	B2	0.33
3	4	L1	S1	B2	0.32
4	5	L1	S1	B3	0.34
...	...	...	...	...	...
247	248	L6	S7	B1	1.50
248	249	L6	S7	B2	1.30
249	250	L6	S7	B2	1.40
250	251	L6	S7	B3	1.50
251	252	L6	S7	B3	1.80

[252 rows x 5 columns]

```
data.set_index('Unnamed: 0',inplace=True)
data
```

	Lab	Spc	Bat	Conc
Unnamed: 0				
1	L1	S1	B1	0.29
2	L1	S1	B1	0.33
3	L1	S1	B2	0.33
4	L1	S1	B2	0.32
5	L1	S1	B3	0.34
...	...	...	...	...
248	L6	S7	B1	1.50

```
249      L6  S7  B2  1.30
250      L6  S7  B2  1.40
251      L6  S7  B3  1.50
252      L6  S7  B3  1.80
```

```
[252 rows x 4 columns]
```

```
# Summary statistics for numerical columns
data.describe()
```

```
          Conc
count  252.000000
mean    1.921508
std     2.473207
min     0.110000
25%     0.467500
50%     1.060000
75%     1.700000
max     9.900000
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 252 entries, 1 to 252
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype
---  -
 0   Lab     252 non-null    object
 1   Spc     252 non-null    object
 2   Bat     252 non-null    object
 3   Conc    252 non-null    float64
dtypes: float64(1), object(3)
memory usage: 9.8+ KB
```

```
# Bar plot for Lab
```

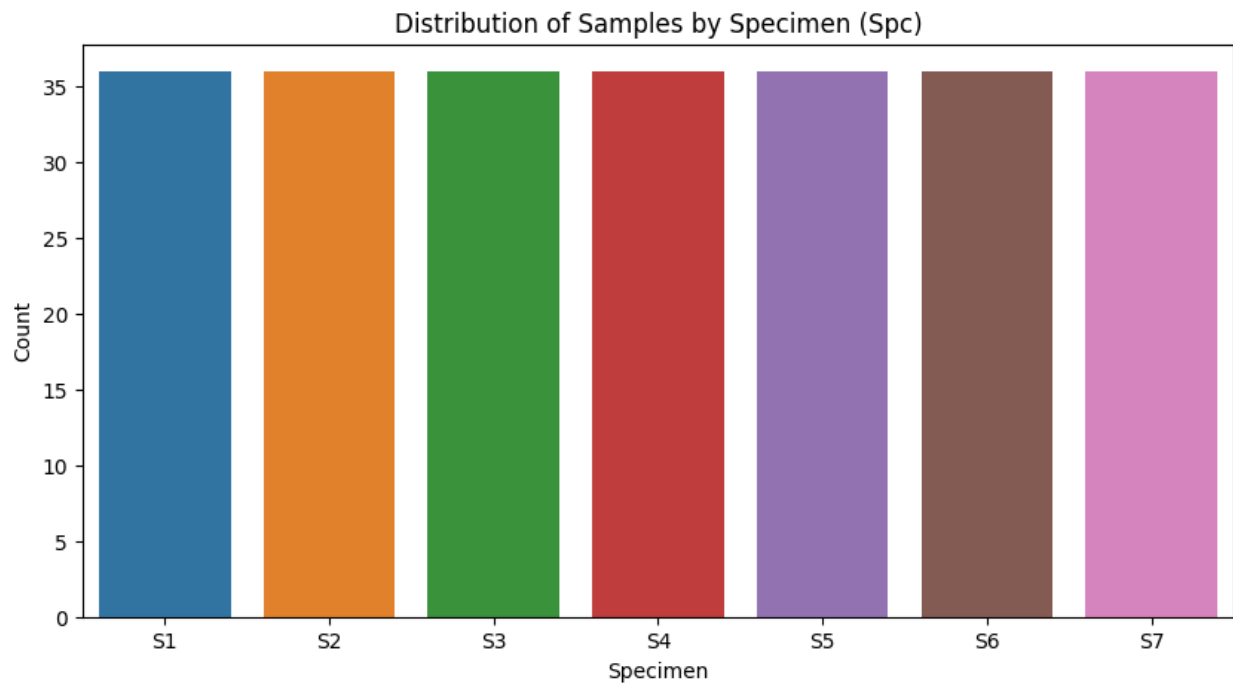
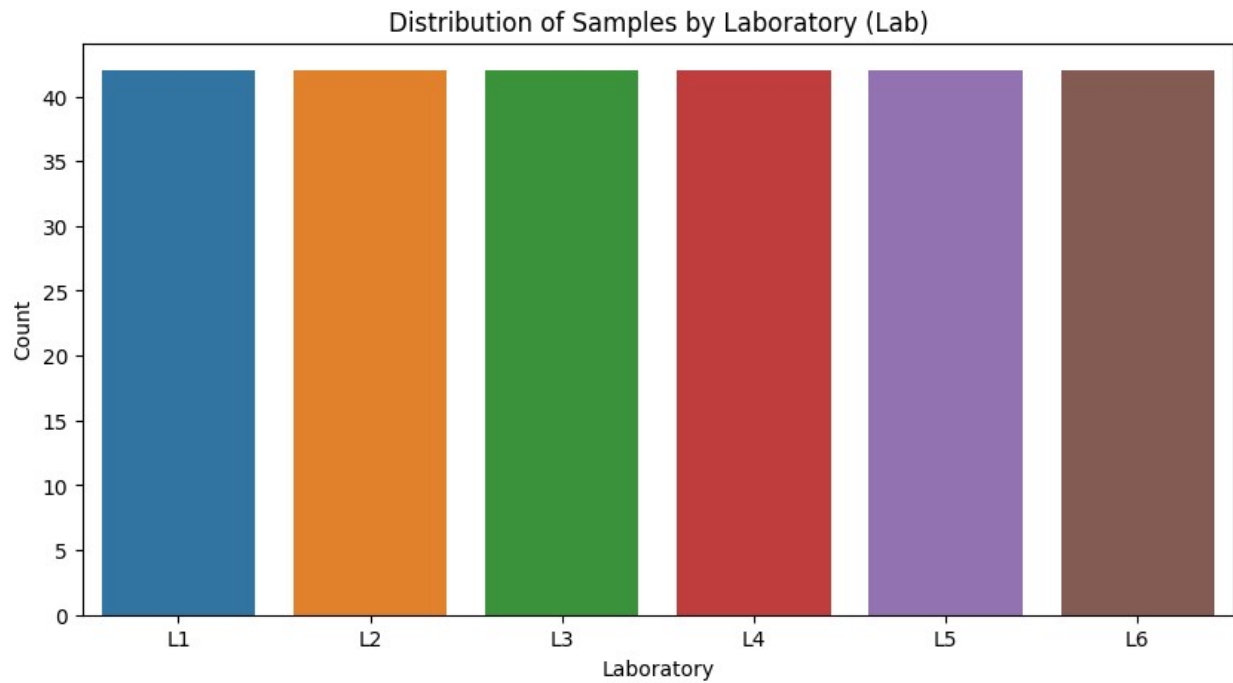
```
plt.figure(figsize=(10, 5))
sns.countplot(x='Lab', data=data)
plt.title('Distribution of Samples by Laboratory (Lab)')
plt.xlabel('Laboratory')
plt.ylabel('Count')
plt.show()
```

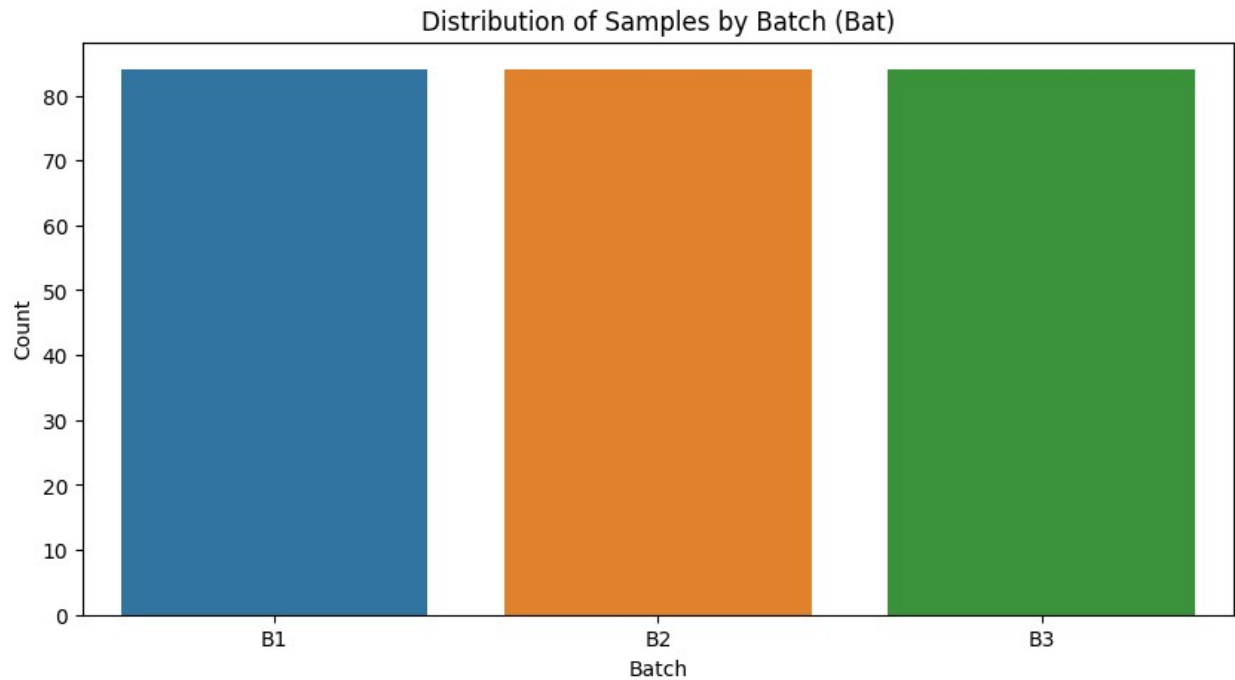
```
# Bar plot for Spc (Specimen)
```

```
plt.figure(figsize=(10, 5))
sns.countplot(x='Spc', data=data)
plt.title('Distribution of Samples by Specimen (Spc)')
plt.xlabel('Specimen')
plt.ylabel('Count')
plt.show()
```

```
# Bar plot for Bat (Batch)
```

```
plt.figure(figsize=(10, 5))
sns.countplot(x='Bat', data=data)
plt.title('Distribution of Samples by Batch (Bat)')
plt.xlabel('Batch')
plt.ylabel('Count')
plt.show()
```



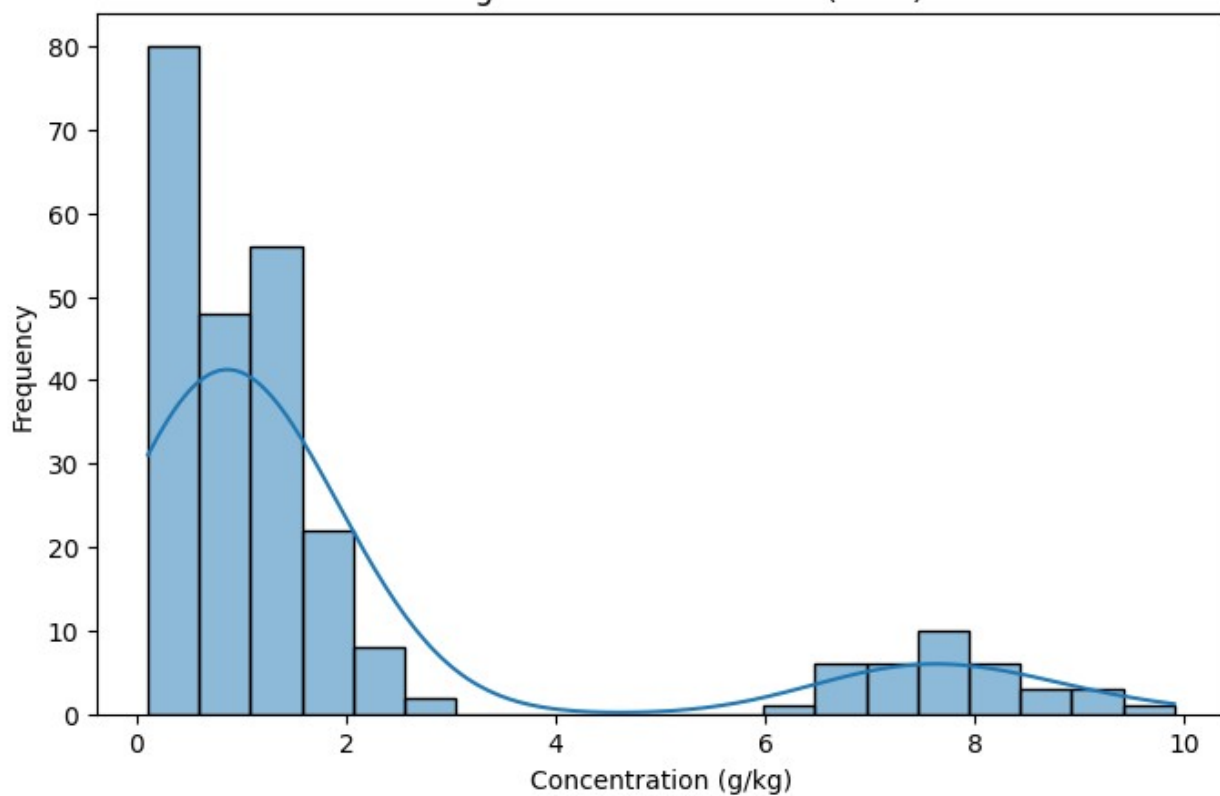


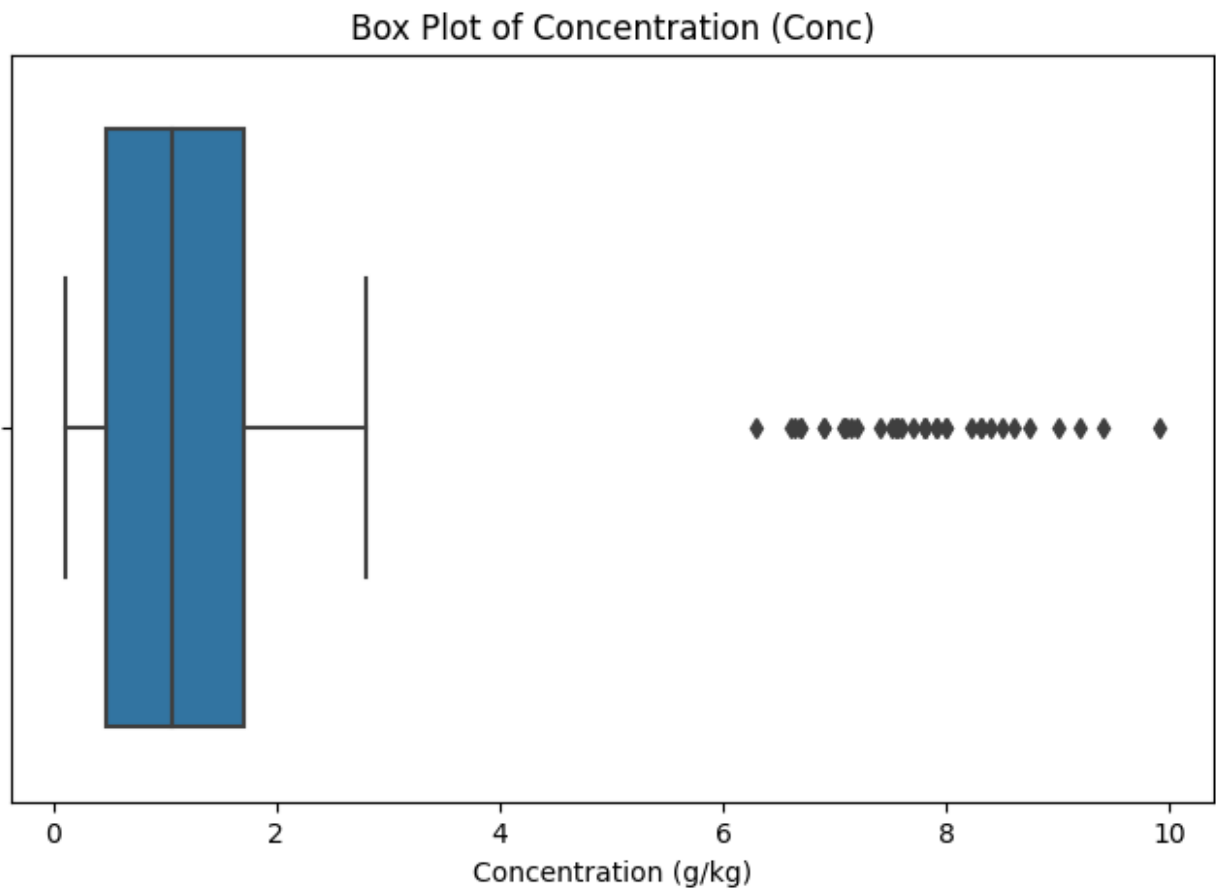
```
# Histogram for Conc
plt.figure(figsize=(8, 5))
sns.histplot(data['Conc'], bins=20, kde=True)
plt.title('Histogram of Concentration (Conc)')
plt.xlabel('Concentration (g/kg)')
plt.ylabel('Frequency')

# Box plot for Conc
plt.figure(figsize=(8, 5))
sns.boxplot(x=data['Conc'])
plt.title('Box Plot of Concentration (Conc)')
plt.xlabel('Concentration (g/kg)')

plt.show()
```

Histogram of Concentration (Conc)

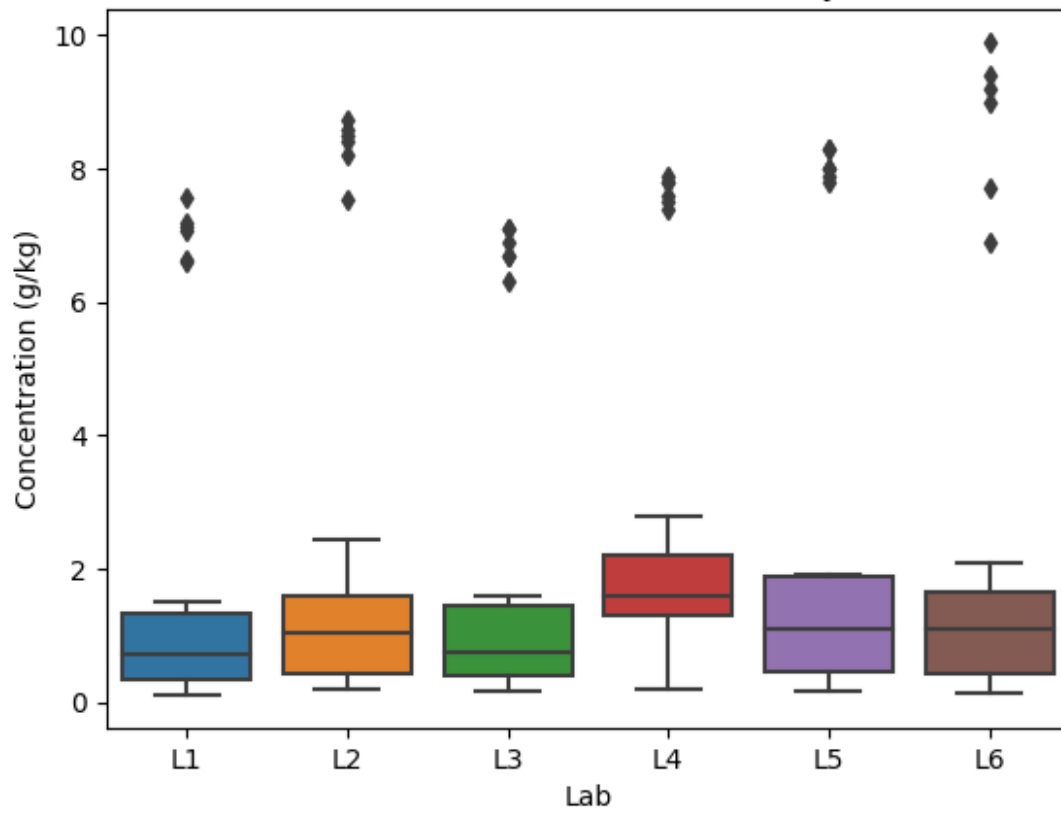


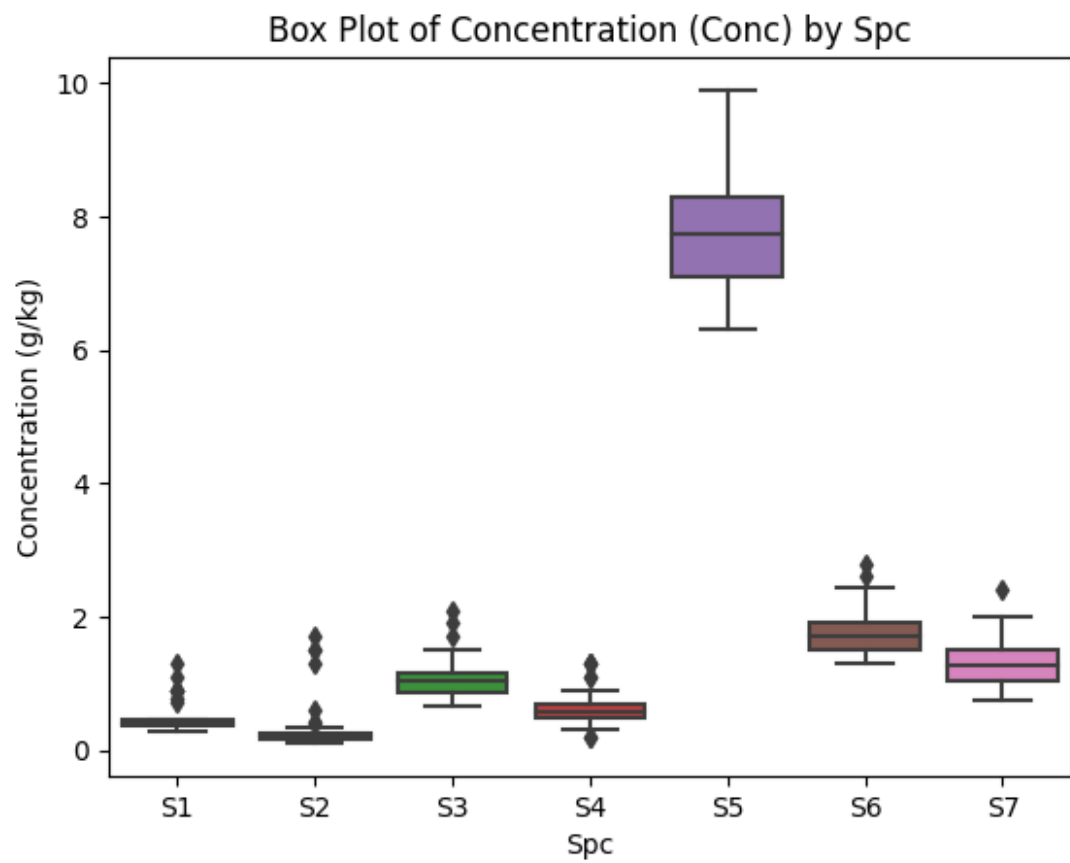


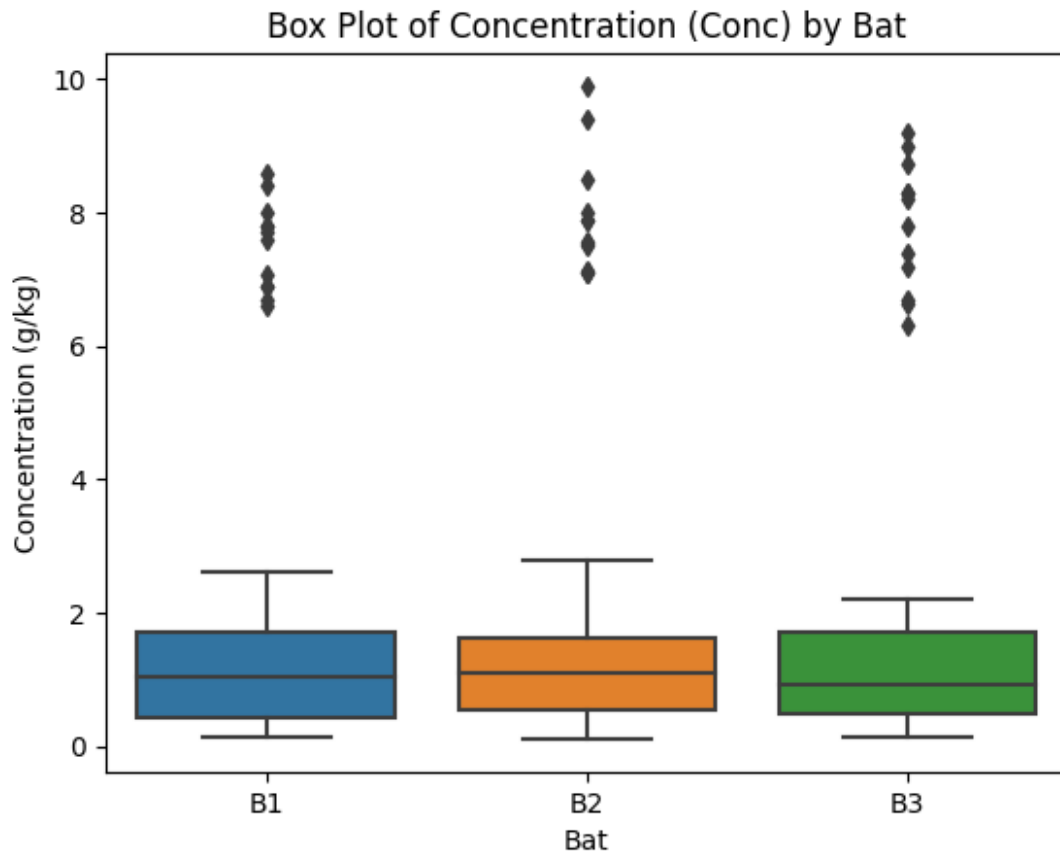
```
# Box plot for Conc by Lab
for i in ['Lab', 'Spc', 'Bat']:
    plt.figure()
    sns.boxplot(x=i, y='Conc', data=data)
    plt.title(f'Box Plot of Concentration (Conc) by {i}')
    plt.xlabel(f'{i}')
    plt.ylabel('Concentration (g/kg)')
    plt.show()
```



Box Plot of Concentration (Conc) by Lab







```
"""
```

Comparing Concentration (Conc) Across Laboratories (Lab):

Null Hypothesis ( $H_0$ ): There is no significant difference in mean concentration (Conc)

across laboratories.

Alternative Hypothesis ( $H_1$ ): There is a significant difference in mean concentration (Conc)

across laboratories.

Statistical Test: One-way Analysis of Variance (ANOVA)

```
"""
```

```
{"type": "string"}
```

```
from scipy.stats import f_oneway
```

```
# Perform one-way ANOVA to test for differences in concentration by laboratory
```

```
anova_result_lab = f_oneway(*[data[data['Lab'] == lab]['Conc'] for lab in data['Lab'].unique()])
```

```
# Display the ANOVA result
```

I really like this clear presentation of tests.

+5 marks

```

print("ANOVA for Concentration by Laboratory:")
print("F-statistic:", anova_result_lab.statistic)
print("p-Value:", anova_result_lab.pvalue)

#conclusion
if anova_result_lab.pvalue <0.05:
    print(f"Null hypotheses rejected.\nThere is a significant
difference in mean concentration (Conc) across laboratories." )
else:
    print(f"Null hypotheses can not be rejected.\nThere is no
significant difference in mean concentration (Conc) across
laboratories" )

```

```

ANOVA for Concentration by Laboratory:
F-statistic: 0.6031212722776087
p-Value: 0.6975980669553611
Null hypotheses can not be rejected.
There is no significant difference in mean concentration (Conc) across
laboratories

```

## There is no significant difference in mean concentration (Conc) across laboratories

---

```

"""Comparing Concentration (Conc) Across Specimens (Spc):

Null Hypothesis (H0): There is no significant difference in mean
concentration (Conc) across specimens.
Alternative Hypothesis (H1): There is a significant difference in mean
concentration (Conc) across specimens.
"""

{"type": "string"}

# Perform one-way ANOVA to test for differences in concentration by
specimen
anova_result_spc = f_oneway(*[data[data['Spc'] == spc]['Conc'] for spc
in data['Spc'].unique()])

# Display the ANOVA result
print("\nANOVA for Concentration by Specimen:")
print("F-statistic:", anova_result_spc.statistic)
print("p-Value:", anova_result_spc.pvalue)

# conclusion
if anova_result_spc.pvalue <0.05:

```

```

    print(f"Null hypotheses rejected.\nThere is a significant
difference in mean concentration (Conc) across specimens." )
else:
    print(f"Null hypotheses can not be rejected.\nThere is no
significant difference in mean concentration (Conc) across
specimens" )

```

ANOVA for Concentration by Specimen:

F-statistic: 1218.3822990397357

p-Value: 2.782678823304331e-179

Null hypotheses rejected.

There is a significant difference in mean concentration (Conc) across specimens.

**There is a significant difference in mean concentration (Conc) across specimens.**

```

"""
Comparing Concentration (Conc) Across Batches (Bat):
Null Hypothesis (H0): There is no significant difference in mean
concentration (Conc) across batches.
Alternative Hypothesis (H1): There is a significant difference in mean
concentration (Conc) across batches.
"""

{"type": "string"}

# Perform one-way ANOVA to test for differences in concentration by
batch
anova_result_batch = f_oneway(*[data[data['Bat'] == batch]['Conc'] for
batch in data['Bat'].unique()])

# Display the ANOVA result
print("\nANOVA for Concentration by Batch:")
print("F-statistic:", anova_result_batch.statistic)
print("p-Value:", anova_result_batch.pvalue)

# conclusion
if anova_result_batch.pvalue < 0.05:
    print(f"Null hypotheses rejected.\nThere is a significant
difference in mean concentration (Conc) across batches." )
else:
    print(f"Null hypotheses can not be rejected.\nThere is no
significant difference in mean concentration (Conc) across batches" )

```

ANOVA for Concentration by Batch:

F-statistic: 0.03361378166030767

p-Value: 0.9669492712999681

Null hypotheses can not be rejected. \There is no significant difference in mean concentration (Conc) across batches

**There is no significant difference in mean concentration (Conc) across batches**

---

*What about interaction effects?* (30)