

### Assignment-03

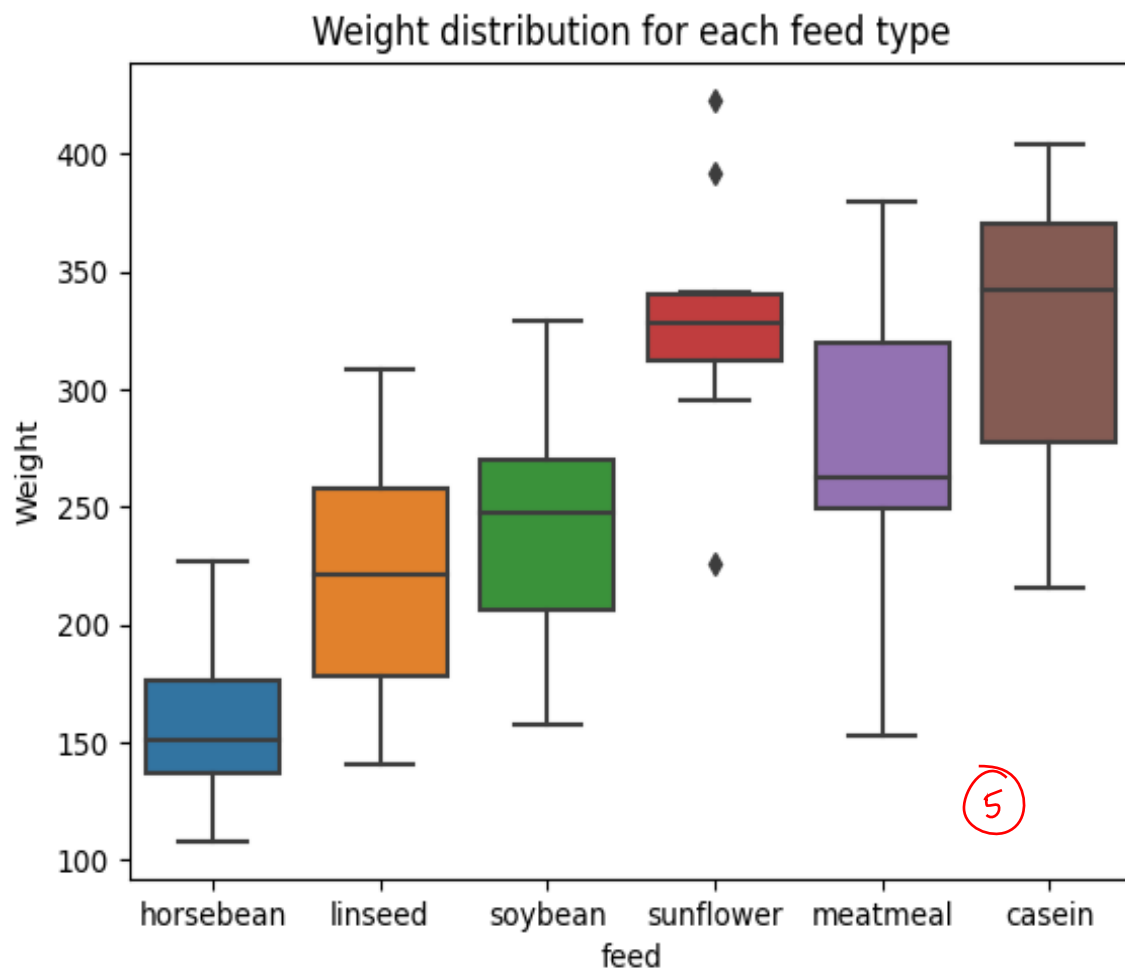
#### Statistical Inference in Bioengineering

P. Teja Venkata Ramana Kumar  
Bm23mtech11006

Q1:-

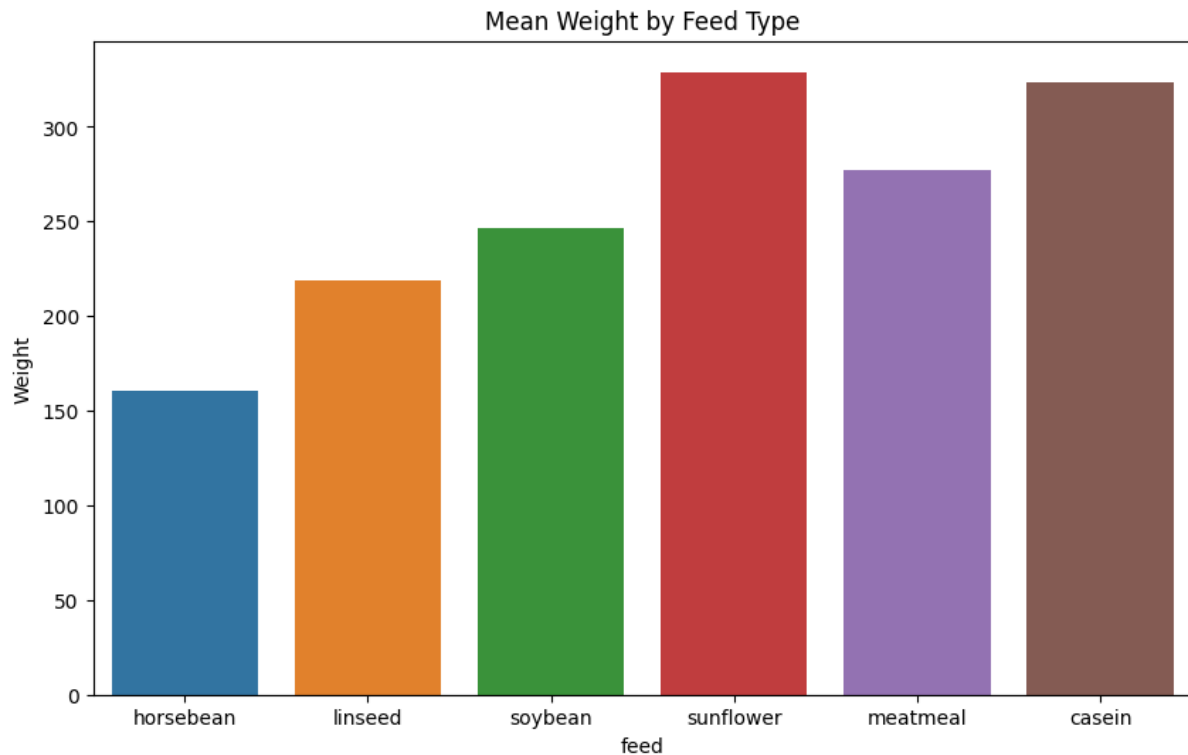
```
import pandas as pd
data=pd.read_excel('Assignment3Data.xlsx',sheet_name='Q1')
data
import matplotlib.pyplot as plt
import seaborn as sns
sns.boxplot(x=data['feed'],y=data['Weight'])
plt.title('Weight distribution for each feed type')

plt.show()
```



5

```
plt.figure(figsize=(10, 6))
sns.barplot(x='feed', y='Weight', data=data, ci=None, estimator=lambda
x: sum(x) / len(x))
plt.title('Mean Weight by Feed Type')
plt.show()
```



```
#Levene's test for homogeneity of variances
from scipy import stats
```

```
stat, p_value = stats.levene(*[data[data['feed'] ==
feed_type]['Weight'] for feed_type in data['feed'].unique()])
print(f"Levene's Test: p-value = {p_value:.4f}")
```

```
if p_value > 0.05:
    print("We do not reject the null hypothesis")
else:
    print("Reject the Null Hypothesis")
```

```
Levene's Test: p-value = 0.5896
```

```
We do not reject the null hypothesis
```

```
# (c) Perform ANOVA
```

```
#H0 (null hypothesis):  $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$  (It implies that the means
of all the population are equal)
```

```
#H1 (null hypothesis): It states that there will be at least one
population mean that differs from the rest
```

```
from statsmodels.formula.api import ols
import statsmodels.api as sm
```

*Normality?*

2

```

anova_Weight = stats.f_oneway(*[data[data['feed'] ==
feed_type]['Weight'] for feed_type in data['feed'].unique()])
print('\nANOVA for weight:')
print('F-statistic:', anova_Weight.statistic)
print('p-value:', anova_Weight.pvalue)

```

5

```

ANOVA for weight:
F-statistic: 15.364799774712534
p-value: 5.936419853471331e-10
# (d) If ANOVA shows a difference, perform post-hoc analysis (Tukey's
HSD)
from statsmodels.stats.multicomp import pairwise_tukeyhsd

if anova_Weight.pvalue < 0.05:
    print(f"The P value is ={anova_Weight.pvalue} < 0.05. So H0 can be
rejected")

    tukey_result = pairwise_tukeyhsd(data['Weight'], data['feed'])
    print("\nTukey's HSD Post-hoc Test:")
    print(tukey_result)
else:
    print(f"The P value is ={anova_Weight.pvalue} < 0.05. So H0 cannot be
rejected")

```

The P value is =5.936419853471331e-10 < 0.05. So H0 can be rejected

10

Tukey's HSD Post-hoc Test:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

=====
group1    group2    meandiff p-adj    lower    upper    reject
-----
casein horsebean -163.3833    0.0 -232.3469 -94.4198    True
casein  linseed -104.8333  0.0002 -170.5875 -39.0792    True
casein meatmeal -46.6742  0.3325 -113.9062  20.5577    False
casein  soybean -77.1548  0.0084 -140.5171 -13.7925    True
casein sunflower  5.3333  0.9999 -60.4208  71.0875    False
horsebean linseed  58.55  0.1413 -10.4135 127.5135    False
horsebean meatmeal 116.7091 0.0001  46.3351 187.0831    True
horsebean soybean  86.2286 0.0042  19.5417 152.9155    True
horsebean sunflower 168.7167  0.0   99.7531 237.6802    True
linseed meatmeal  58.1591 0.1277  -9.0729 125.3911    False
linseed soybean  27.6786 0.7933 -35.6837  91.0409    False
linseed sunflower 110.1667 0.0001  44.4125 175.9208    True
meatmeal soybean -30.4805 0.7391 -95.3751  34.4141    False
meatmeal sunflower  52.0076 0.2207 -15.2244 119.2395    False
soybean sunflower  82.4881 0.0039  19.1258 145.8504    True

```

""In this case, "horsebean" has the highest mean difference in weight gain compared to several other feeds.

The largest mean difference is between "horsebean" and "sunflower" (mean difference = 168.72, p < 0.05)

""

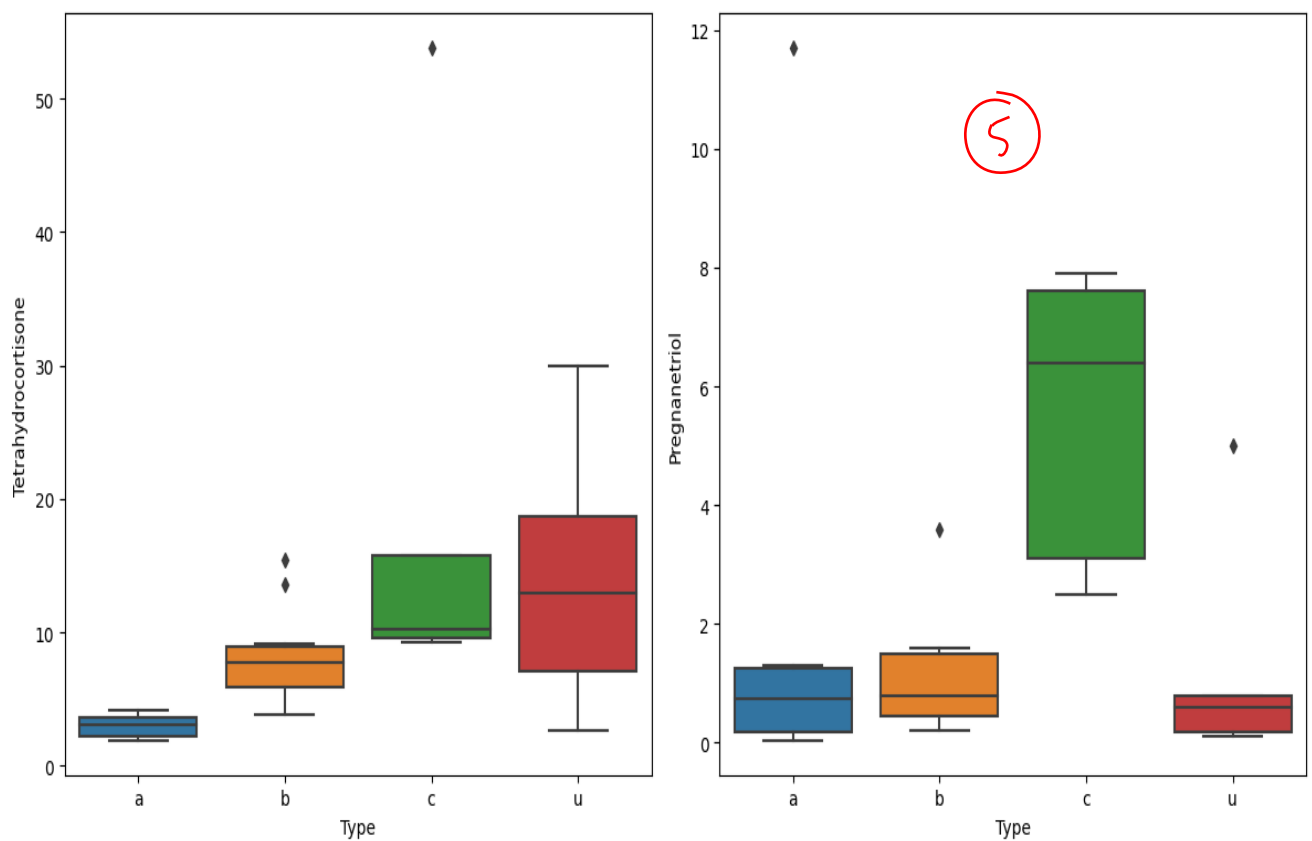
**Q2:-**

```
data1=pd.read_excel('Assignment3Data.xlsx',sheet_name='Q2')
data1
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)

sns.boxplot(x=data1['Type'],y=data1['Tetrahydrocortisone'])

plt.subplot(1, 2, 2)
sns.boxplot(y=data1['Pregnanetriol'],x=data1['Type'])
plt.tight_layout()

plt.show()
```

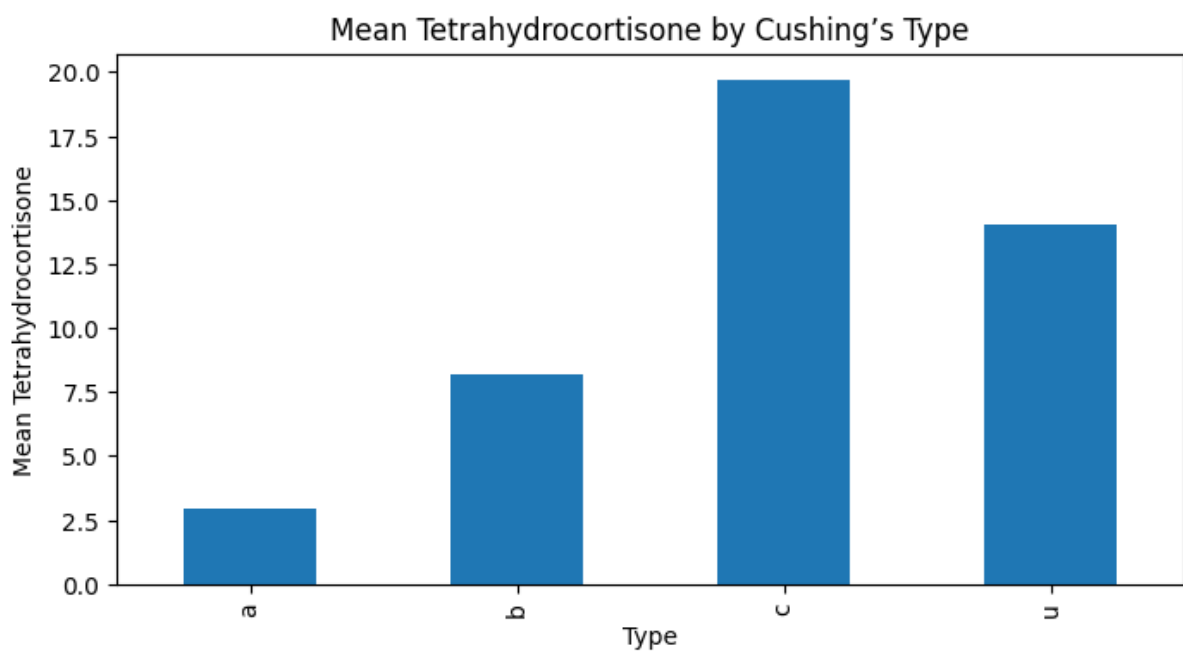
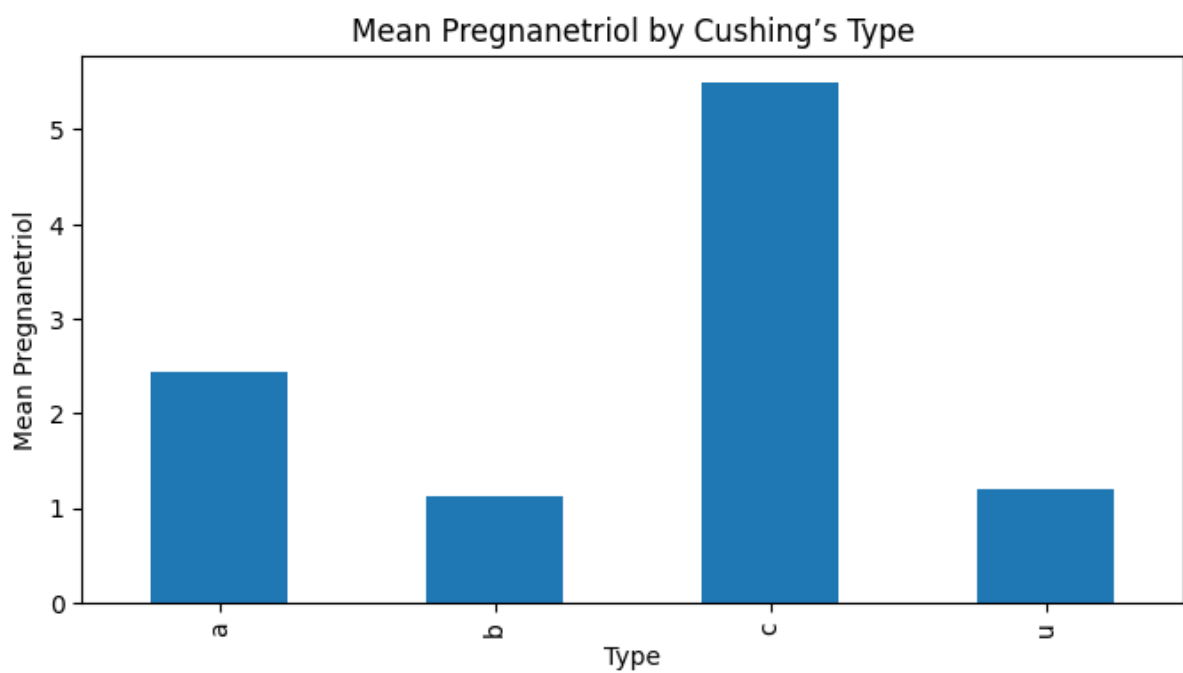


```
# Plot of means for Pregnanetriol by Type
plt.figure(figsize=(8, 4))
means_pregnanetriol = data1.groupby('Type')['Pregnanetriol'].mean()
means_pregnanetriol.plot(kind='bar')
plt.xlabel('Type')
plt.ylabel('Mean Pregnanetriol')
plt.title('Mean Pregnanetriol by Cushing's Type')
plt.show()
```

```

# Plot of means for Tetrahydrocortisone by Type
plt.figure(figsize=(8, 4))
means_tetrahydrocortisone =
data1.groupby('Type')['Tetrahydrocortisone'].mean()
means_tetrahydrocortisone.plot(kind='bar')
plt.xlabel('Type')
plt.ylabel('Mean Tetrahydrocortisone')
plt.title('Mean Tetrahydrocortisone by Cushing's Type')
plt.show()

```



```

from scipy.stats import levene

# Assumption: Levene's test for homogeneity of variances ✓
levene_test_pregnanetriol = levene(data1['Pregnanetriol'][data1['Type']
== 'a'],
                                data1['Pregnanetriol'][data1['Type']
== 'b'],
                                data1['Pregnanetriol'][data1['Type']
== 'c'],
                                data1['Pregnanetriol'][data1['Type']
== 'u'])[1]

levene_test_tetrahydrocortisone =
levene(data1['Tetrahydrocortisone'][data1['Type'] == 'a'],

data1['Tetrahydrocortisone'][data1['Type'] == 'b'],

data1['Tetrahydrocortisone'][data1['Type'] == 'c'],

data1['Tetrahydrocortisone'][data1['Type'] == 'u'])[1] Normality?

print(f'Levene\'s test for homogeneity of variances (Pregnanetriol): p-
value = {levene_test_pregnanetriol:.4f}')
print(f'Levene\'s test for homogeneity of variances
(Tetrahydrocortisone): p-value =
{levene_test_tetrahydrocortisone:.4f}')
Levene's test for homogeneity of variances (Pregnanetriol): p-value =
0.5196
Levene's test for homogeneity of variances (Tetrahydrocortisone): p-value =
0.2505

# (c) Perform ANOVA for Pregnanetriol
#H0 (null hypothesis):  $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$  (It implies that the means
of all the population are equal)
#H1 (null hypothesis): It states that there will be at least one
population mean that differs from the rest

anova_pregnanetriol =
stats.f_oneway(data1['Pregnanetriol'][data1['Type'] == 'a'],
                data1['Pregnanetriol'][data1['Type'] ==
'b'],
                data1['Pregnanetriol'][data1['Type'] ==
'c'],
                data1['Pregnanetriol'][data1['Type'] ==
'u'])

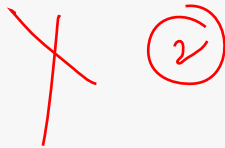
print('\nANOVA for Pregnanetriol:')
print('F-statistic:', anova_pregnanetriol.statistic)
print('p-value:', anova_pregnanetriol.pvalue)

```

ANOVA for Pregnanetriol:  
F-statistic: 3.5392626216260634  
p-value: 0.030507062483867444

```
# (d) Perform ANOVA for Tetrahydrocortisone
anova_tetrahydrocortisone =
stats.f_oneway(data1['Tetrahydrocortisone'][data1['Type'] == 'a'],
data1['Tetrahydrocortisone'][data1['Type'] == 'b'],
data1['Tetrahydrocortisone'][data1['Type'] == 'c'],
data1['Tetrahydrocortisone'][data1['Type'] == 'u'])

print('\nANOVA for Tetrahydrocortisone:')
print('F-statistic:', anova_tetrahydrocortisone.statistic)
print('p-value:', anova_tetrahydrocortisone.pvalue)
ANOVA for Tetrahydrocortisone:
F-statistic: 3.2257394791378426
p-value: 0.0412182793672776
```




```
# (d) Perform ANOVA for Tetrahydrocortisone
anova_tetrahydrocortisone =
stats.f_oneway(data1['Tetrahydrocortisone'][data1['Type'] == 'a'],
data1['Tetrahydrocortisone'][data1['Type'] == 'b'],
data1['Tetrahydrocortisone'][data1['Type'] == 'c'],
data1['Tetrahydrocortisone'][data1['Type'] == 'u'])

print('\nANOVA for Tetrahydrocortisone:')
print('F-statistic:', anova_tetrahydrocortisone.statistic)
print('p-value:', anova_tetrahydrocortisone.pvalue)
ANOVA for Tetrahydrocortisone:
F-statistic: 3.2257394791378426
p-value: 0.0412182793672776
```

```
# (e) Perform post-hoc analysis
# Post-hoc analysis for Pregnanetriol
if anova_pregnanetriol.pvalue < 0.05:
    print(f"The P value is {anova_pregnanetriol.pvalue} < 0.05. So H0
can be rejected")

    tukey_pregnanetriol =
pairwise_tukeyhsd(endog=data1['Pregnanetriol'], groups=data1['Type'],
alpha=0.05)
    print('\nPost-hoc analysis for Pregnanetriol:')
    print(tukey_pregnanetriol.summary())
else:
```



```
print(f"The P value is ={anova_pregnanetriol.pvalue} < 0.05. So H0  
can be rejected")
```

```
# Post-hoc analysis for Tetrahydrocortisone
```

```
if anova_tetrahydrocortisone.pvalue < 0.05:  
    print(f"The P value is ={anova_tetrahydrocortisone.pvalue} < 0.05.  
So H0 can be rejected")
```

```
    tukey_tetrahydrocortisone =  
pairwise_tukeyhsd(endog=data1['Tetrahydrocortisone'],  
groups=data1['Type'], alpha=0.05)  
    print('\nPost-hoc analysis for Tetrahydrocortisone:')  
    print(tukey_tetrahydrocortisone.summary())
```

```
else:  
    print(f"The P value is ={anova_tetrahydrocortisone.pvalue} < 0.05.  
So H0 can be rejected")
```

The P value is =0.030507062483867444 < 0.05. So H0 can be rejected

Post-hoc analysis for Pregnanetriol:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

```
=====
```

group1	group2	meandiff	p-adj	lower	upper	reject
a	b	-1.32	0.7627	-5.0519	2.4119	False
a	c	3.06	0.2416	-1.316	7.436	False
a	u	-1.24	0.8433	-5.4124	2.9324	False
b	c	4.38	0.0264	0.4217	8.3383	True
b	u	0.08	0.9999	-3.6519	3.8119	False
c	u	-4.3	0.0553	-8.676	0.076	False

```
-----
```

The P value is =0.0412182793672776 < 0.05. So H0 can be rejected

Post-hoc analysis for Tetrahydrocortisone:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

```
=====
```

group1	group2	meandiff	p-adj	lower	upper	reject
a	b	5.2133	0.722	-8.5182	18.9448	False
a	c	16.7533	0.0394	0.6517	32.855	True
a	u	11.05	0.2201	-4.3023	26.4023	False
b	c	11.54	0.1552	-3.0245	26.1045	False
b	u	5.8367	0.6474	-7.8948	19.5682	False
c	u	-5.7033	0.7619	-21.805	10.3983	False

```
-----
```



### Q3):-

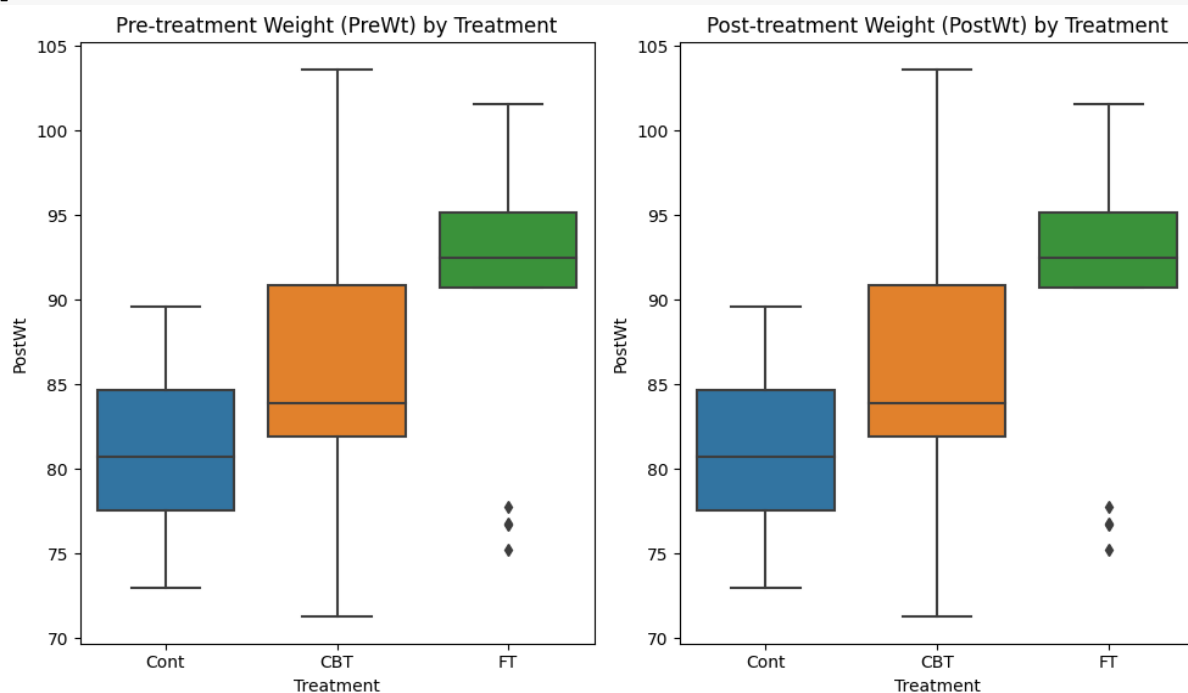
```
data3=pd.read_excel('Assignment3Data.xlsx',sheet_name='Q3')
data3
# (a) Visualize the data using boxplots and a plot of means

plt.figure(figsize=(10, 6))
plt.subplot(1, 2, 1)
sns.boxplot(x='Treatment', y='PostWt', data=data3)
plt.title('Pre-treatment Weight (PreWt) by Treatment')

plt.subplot(1, 2, 2)
sns.boxplot(x='Treatment', y='PostWt', data=data3)
plt.title('Post-treatment Weight (PostWt) by Treatment')

plt.tight_layout()
plt.show()
```

5

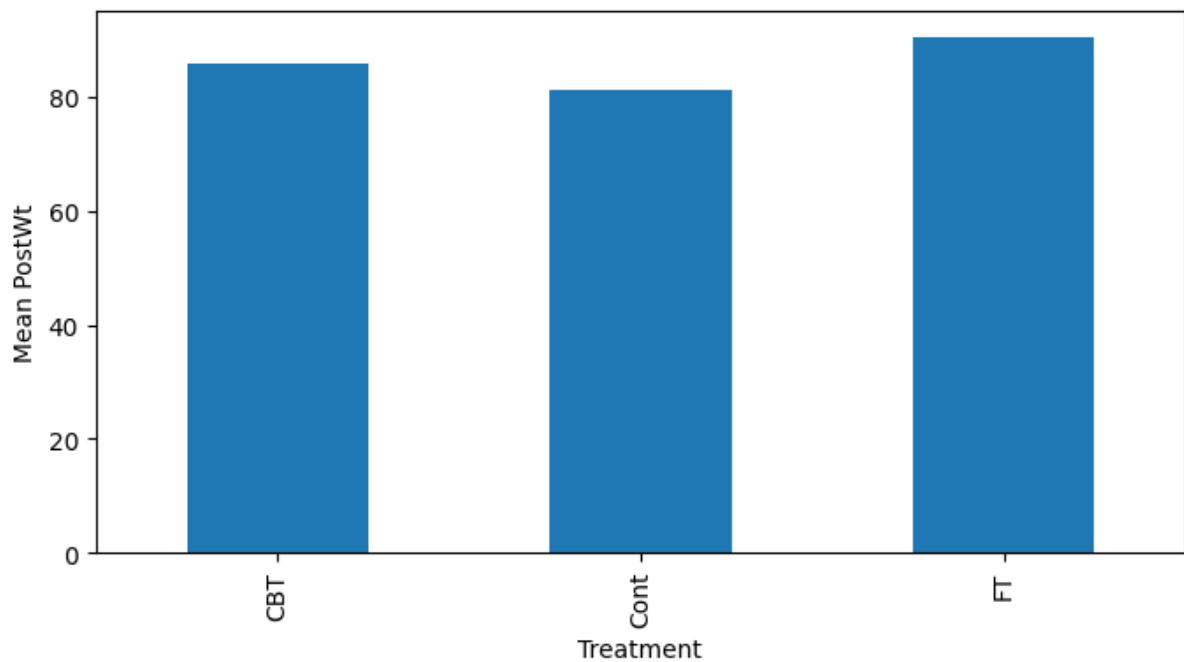
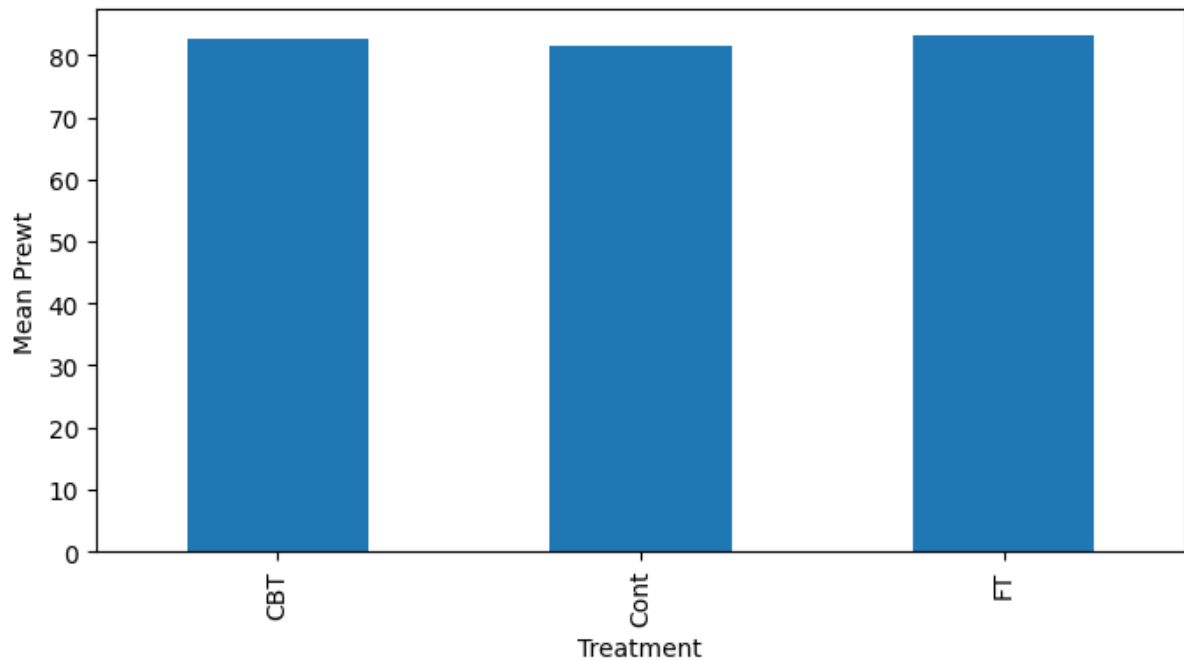


```
# Plot of means for Pregnanetriol by Type
plt.figure(figsize=(8, 4))
means_prewt= data3.groupby('Treatment')['PreWt'].mean()
means_prewt.plot(kind='bar')
plt.xlabel('Treatment')
plt.ylabel('Mean Prewt')

plt.show()

# Plot of means for Tetrahydrocortisone by Type
plt.figure(figsize=(8, 4))
means_postwt = data3.groupby('Treatment')['PostWt'].mean()
means_postwt.plot(kind='bar')
```

```
plt.xlabel('Treatment')
plt.ylabel('Mean PostWt')
plt.show()
```



```
# (b) Calculate the difference in body weight for each subject
data3['WeightDiff'] = data3['PostWt'] - data3['PreWt']
data3
#(c) Assumption : Levene's test for homogeneity of variances
levене_test = levene(data3['WeightDiff'][data3['Treatment'] == 'Cont'],
                      data3['WeightDiff'][data3['Treatment'] == 'CBT'],
                      data3['WeightDiff'][data3['Treatment'] == 'FT'])[1]
```

Normality

②

```
print(f'\nLevene\'s test for homogeneity of variances (WeightDiff): p-  
value = {levене_test:.4f}')
```

Levene's test for homogeneity of variances (WeightDiff): p-value = 0.5964

```
# (d) Perform ANOVA and post-hoc analysis
```

```
#H0 (null hypothesis):  $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$  (It implies that the means  
of all the population are equal)
```

```
#H1 (null hypothesis): It states that there will be at least one  
population mean that differs from the rest
```

```
anova_result = stats.f_oneway(data3['WeightDiff'][data3['Treatment'] ==  
'Cont'],  
                               data3['WeightDiff'][data3['Treatment'] ==  
'CBT'],  
                               data3['WeightDiff'][data3['Treatment'] ==  
'FT'])
```

```
print('\nANOVA for WeightDiff:')  
print('F-statistic:', anova_result.statistic)  
print('p-value:', anova_result.pvalue)
```

```
# Post-hoc analysis
```

```
tukey = pairwise_tukeyhsd(endog=data3['WeightDiff'],  
groups=data3['Treatment'], alpha=0.05)
```

```
print('\nPost-hoc analysis:')  
print(tukey.summary())
```

```
print('\nPairwise comparisons:')  
print(tukey.groupsunique)
```

```
print(tukey.meandiffs)
```

```
print(tukey.pvalues)
```

ANOVA for WeightDiff:

F-statistic: 5.422296868238266

p-value: 0.006498652980699739

Post-hoc analysis:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

```
=====
```

group1	group2	meandiff	p-adj	lower	upper	reject
CBT	Cont	-3.4569	0.2124	-8.3273	1.4135	False
CBT	FT	4.2578	0.1607	-1.2506	9.7662	False
Cont	FT	7.7147	0.0045	2.0901	13.3393	True

```
-----
```

Pairwise comparisons:

```
['CBT' 'Cont' 'FT']
```

```
[-3.45689655  4.25780933  7.71470588]
```

```
[0.21244284  0.16074613  0.00451275]
```

**CBT vs. Cont (Control):**

Mean Difference: -3.4569

p-value: 0.2124

Result: Not statistically significant ( $p > 0.05$ )

Interpretation: There is no statistically significant difference in effectiveness between Cognitive Behavioral Therapy (CBT) and the Control group.

**CBT vs. FT (First Therapy):**

Mean Difference: 4.2578

p-value: 0.1607

Result: Not statistically significant ( $p > 0.05$ )

Interpretation: There is no statistically significant difference in effectiveness between Cognitive Behavioral Therapy (CBT) and the First Therapy (FT).

**Cont (Control) vs. FT (First Therapy):**

Mean Difference: 7.7147

p-value: 0.0045

Result: Statistically significant ( $p \leq 0.05$ )

Interpretation: There is a statistically significant difference in effectiveness between the Control group and the First Therapy (FT). FT appears to be more effective compared to the Control group.

**Overall Interpretation:**

The First Therapy (FT) shows a statistically significant higher effectiveness compared to the Control group (Cont). However, there is no statistically significant difference in effectiveness observed between Cognitive Behavioral Therapy (CBT) and both the Control group (Cont) and the First Therapy (FT).

Q4):-

```
import statsmodels.stats.api as sms

variance = 1000 # Variance in cell numbers
effect_size = 100 # Difference in means
alpha = 0.05 # Significance level
power = 0.80 # Desired power

n=sms.tt_ind_solve_power(effect_size=effect_size, nobs1=None,
alpha=alpha, power=power, ratio=1.0, alternative='two-sided')

print("Sample size for alpha = 0.05:", n)
import statsmodels.stats.api as sms

variance = 1000 # Variance in cell numbers
effect_size = 100 # Difference in means
alpha = 0.05 # Significance level
power = 0.80 # Desired power

n=sms.tt_ind_solve_power(effect_size=effect_size, nobs1=None,
alpha=alpha, power=power, ratio=1.0, alternative='two-sided')

print("Sample size for alpha = 0.05:", n)
```

Sample size for alpha = 0.05: 10