Department of Computer Science and Engineering (Data Science)
B.Tech. Sem: III Subject: Statistics for Data Science

Experiment 10

Name: Rishabh Singhvi **SAP ID: 60009210206**

Date:	Experiment Title: Analysis of Variance using F-distribution
Aim	To implement one-way and two-way ANOVA
Software	Google Colab
Theory	**1. Can we conclude that treatments are effective on plant growth from the given data? Also check the significance using Tukey's Test. Code:**

```
import numpy as np import pandas as pd import
statsmodels.api as sm from statsmodels.formula.api
import ols from statsmodels.stats.multicomp import
pairwise_tukey hsd
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
path= "/content/drive/MyDrive/SDS/Plant_Growth_New.csv
" df =
pd.read_csv(path) df
```

```
print("Null hypothesis is that there is no significant
difference between the treatment on plants with  thei
r growth.") print("Alternate hypothesis is that there
is a signifi cant difference between the treatment on
plants with  their growth.") alpha = 0.05
mod = ols('Height~Treatment', data = df). fit() aov
= sm.stats.anova_lm(mod) print('One-way anova table
is given below:\n', aov)
```

```
Null hypothesis is that there is no significant
difference between the treatment on plants with  their
growth.
Alternate hypothesis is that there is a significant
difference between the treatment on plants with  their
growth.
One-way anova table is given below:
              df     sum_sq   mean_sq      F    PR(>F)
Treatment    5.0  35.866667  7.173333  13.45  0.000003
```

```
Residual    24.0  12.800000  0.533333     NaN        NaN
```

```
f_statistics = aov['F']['Treatment']
p_value = aov['PR(>F)']['Treatment'] if
p_value> alpha:
  print("Failed to reject null hypothesis for level of
significance = " + str(alpha)) else:
  print("Null hypothesis is rejected for level of sign
ificance = " + str(alpha))
```

```
Null hypothesis is rejected for level of significance
= 0.05
```

```
tukey = pairwise_tukeyhsd(endog =
df['Height'], groups
 = df['Treatment'], alpha = 0.05) print('Tukey Posthoc
analysis for the given dataset is given below: \n'
, tukey)
```

```
Tukey Post-hoc analysis for the given dataset is given
below:
                                  Multiple Comparison of
Means - Tukey HSD, FWER=0.05
=============================================================
=============================================================
====                  group1
group2                  meandiff p-adj   lower    upper
reject
------------------------------------------------------------
------------------------------------------------------------
----     high sun exposure and water daily    high
sun exposure and water weekly      -0.4     0.9  -
1.8282
1.0282  False
    high sun exposure and water daily     low sun
exposure and water daily      -1.0   0.29 -2.4282
0.4282  False
    high sun exposure and water daily     low sun
exposure and water weekly      -2.8  0.001 -4.2282 -
1.3718   True
    high sun exposure and water daily medium sun
exposure and water daily      -1.8 0.0079 -3.2282 -
0.3718   True
    high sun exposure and water daily medium sun
exposure and water weekly      -2.8  0.001 -4.2282 -
1.3718   True
   high sun exposure and water weekly     low sun
exposure and water daily     -0.6 0.7576 -2.0282
0.8282  False
   high sun exposure and water weekly     low sun
exposure and water weekly      -2.4  0.001 -3.8282
0.9718   True
```

```
high sun exposure and water weekly medium sun
```

```
exposure and water daily      -1.4  0.057 -2.8282
0.0282
False
  high sun exposure and water weekly medium sun
exposure and water weekly     -2.4  0.001 -3.8282 -
0.9718
True
    low sun exposure and water daily    low sun
exposure and water weekly     -1.8 0.0079 -3.2282 -
0.3718
True
    low sun exposure and water daily medium sun
exposure and water daily      -0.8 0.5222 -2.2282
0.6282
False
    low sun exposure and water daily medium sun
exposure and water weekly     -1.8 0.0079 -3.2282 -
0.3718
True
   low sun exposure and water weekly medium sun
exposure and water daily       1.0   0.29 -0.4282
2.4282
False
   low sun exposure and water weekly medium sun
exposure and water weekly      0.0    0.9 -1.4282
1.4282
False
medium sun exposure and water daily  medium sun
exposure and water weekly     -1.0   0.29 -2.4282
0.4282
False
```

2.

influenced b **st wants to know whether or not plant growth is exposure and wate**
frequency.
ds and lets them grow for two months under different conditions fo
osure and watering frequency. After two months, she records the
h plant, in inches. Perform a two-way ANOVA to determine if
quency and sunlight exposure have a significant effect on plant
to determine if there is any interaction effect between watering
ıd sunlight

were given low sun exposure and water daily ts were given medium sun
posure and water daily
were given high sun exposure and water daily plants were given low su
e and water weekly plants were given medium sun exposure and water
plants were given high sun exposure and water weekly

- 5
- 5 plan
- 5
- 5
- 5
- 5

6, 6, 6, 5, 6, 5, 5, 6, 4, 5, 6, 6, 7, 8, 7, 3, 4, 4, 4, 5, 4,
8

```
"/content/drive/MyDrive/SDS/Plant_Growth_New_2.c

(path)
```

4, 4, 4, 4, 5, 6,
7,

Code:

```
path= sv" d
pd.read_csv
```

```
print('Null hypothesis H_01 is that there is no signif
icant effect of sunlight exposure on plant growth.')
print('Alternate hypothesis H_a1 is that there is a si
gnificant effect of sunlight exposure on plant growth.
') print('Null hypothesis H_02 is that there is no
signif icant effect of Watering frequency on plant
growth.') print('Alternate hypothesis H_a2 is that
there is a si gnificant effect of Watering frequency
on plant growth
.') print('Null hypothesis H_03 is that there is no
signif icant interaction effect of watering frequency
and sun light exposure on plant growth.')
print('Alternate hypothesis H_a3 is that there is a si
gnificant interaction effect of watering frequency and
sunlight exposure on plant growth.')
```

```
Null hypothesis H_01 is that there is no significant
effect of sunlight exposure on plant growth. Alternate
hypothesis H_a1 is that there is a significant effect
of sunlight exposure on plant growth.
Null hypothesis H_02 is that there is no significant
effect of Watering frequency on plant growth.
Alternate hypothesis H_a2 is that there is a
significant effect of Watering frequency on plant
growth.
Null hypothesis H_03 is that there is no significant
interaction effect of watering frequency and sunlight
exposure on plant growth.
Alternate hypothesis H_a3 is that there is a
significant interaction effect of watering frequency
and sunlight exposure on plant growth.
```

```
mod = ols('Height~Sun_Exposure*Water', data = df).fit(
)
aov = sm.stats.anova_lm(mod) print('Two-way anova
table is given below: \n', aov)
```

```
Two-way anova table is given below:
                    df      sum_sq    mean_sq
F    PR(>F)
Sun_Exposure       2.0  24.866667  12.433333
23.3125  0.000002
Water              1.0   8.533333   8.533333
16.0000  0.000527
Sun_Exposure:Water   2.0   2.466667   1.233333
2.3125  0.120667
Residual           24.0  12.800000   0.533333
NaN       NaN
```

```
f_statistics_1 = aov['F']['Sun_Exposure'] p_value_1
= aov['PR(>F)']['Sun_Exposure']
```

	`` print('f_statistics_1 = ', f_statistics_1, 'and p_valu e_1=', p_value_1) if p_value_1> alpha: print("Failed to reject null hypothesis H_01 for lev el of significance = " + str(alpha)) else: print("Null hypothesis H_01 is rejected for level of significance = " + str(alpha)) ``` f_statistics_1 = 23.312500000000004 and p_value_1= 2.371555925858255e-06 Null hypothesis H_01 is rejected for level of significance = 0.05 ``` f_statistics_2 = aov['F']['Water'] p_value_2 = aov['PR(>F)']['Water'] print('f_statistics_2 = ', f_statistics_2, 'and p_valu e_2 =', p_value_2) if p_value_2 > alpha: print("Failed to reject null hypothesis H_02 for lev el of significance = " + str(alpha)) else: print("Null hypothesis H_02 is rejected for level of significance = " + str(alpha)) ``` f_statistics_2 = 16.0 and p_value_2 = 0.0005269080727817035 Null hypothesis H_02 is rejected for level of significance = 0.05 ``` f_statistics_3 = aov['F']['Sun_Exposure:Water'] p_value_3 = aov['PR(>F)']['Sun_Exposure:Water'] print('f_statistics_3 = ', f_statistics_3, 'and p_valu e_3 =', p_value_3) if p_value_3 > alpha: print("Failed to reject null hypothesis H_03 for lev el of significance = " + str(alpha)) else: print("Null hypothesis H_03 is rejected for level of significance = " + str(alpha)) ``` f_statistics_3 = 2.3125000000000013 and p_value_3 = 0.12066712248670274 Failed to reject null hypothesis H_03 for level of significance = 0.05
Conclusion	Hence we have studied and implemented analysis of variance using F-distribution.

Signature of Faculty