Name: Adwait Hegde
Roll No: 2019130019
TE Comp (Batch-A)

# EXPERIMENT 3

## Aim:

Hypothesis testing: Perform Hypothesis testing using f-test with SAS

## Problem Statement:

This analysis is aimed at those who are interested in statistics related to species of Iris Flowers. The dataset contains various attributes associated the flowers such as Sepal Length, Sepal Width, Petal Length, Petal Width. The main aim is to determine if the mean of magnitude Petal Length and Width of all species of flower is same or different.

## Implementation:

Dataset used: SASHELP.IRIS

Null hypothesis: The mean of Petal Length and Width of all species of flower is same

Alternate hypothesis: The mean of Petal Length and Width of all species of flower is different

## Code:

```
PROC SQL;
CREATE TABLE WORK.query AS
SELECT Species , SepalLength , SepalWidth , PetalLength , PetalWidth FROM
SASHELP.IRIS;
RUN;
QUIT;

PROC DATASETS NOLIST NODETAILS;
CONTENTS DATA=WORK.query OUT=WORK.details;
RUN;

PROC ANOVA DATA = WORK.query;
CLASS Species;
MODEL PetalWidth = Species;
MEANS Species / tukey lines;
RUN;

PROC PRINT DATA=WORK.details;
RUN;
```

The ANOVA Procedure

| Class Level Information | | |
|---|---|---|
| Class | Levels | Values |
| Species | 3 | Setosa Versicolor Virginica |

| | |
|---|---|
| Number of Observations Read | 150 |
| Number of Observations Used | 150 |

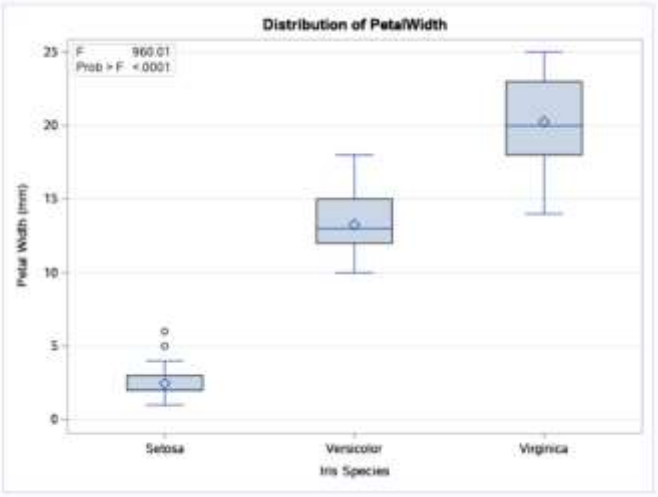
The ANOVA Procedure

Dependent Variable: PetalWidth Petal Width (mm)

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 2 | 8041.333333 | 4020.666667 | 960.01 | <.0001 |
| Error | 147 | 615.660000 | 4.188163 | | |
| Corrected Total | 149 | 8656.993333 | | | |

| R-Square | Coeff Var | Root MSE | PetalWidth Mean |
|---|---|---|---|
| 0.928883 | 17.06365 | 2.046500 | 11.99333 |

| Source | DF | Anova SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Species | 2 | 8041.333333 | 4020.666667 | 960.01 | <.0001 |

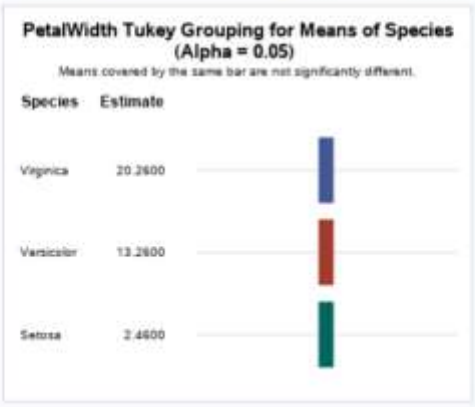



The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for PetalWidth

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

| | |
|---|---|
| Alpha | 0.05 |
| Error Degrees of Freedom | 147 |
| Error Mean Square | 4.188163 |
| Critical Value of Studentized Range | 3.34842 |
| Minimum Significant Difference | 0.9691 |



```
PROC SQL;
CREATE TABLE WORK.query AS
SELECT Species , SepalLength , SepalWidth , PetalLength , PetalWidth FROM
SASHELP.IRIS;
RUN;
QUIT;

PROC DATASETS NOLIST NODETAILS;
CONTENTS DATA=WORK.query OUT=WORK.details;
RUN;

PROC ANOVA DATA = WORK.query;
CLASS Species;
MODEL PetalLength = Species;
MEANS Species / tukey lines;
RUN;

PROC PRINT DATA=WORK.details;
RUN;
```

The ANOVA Procedure

| Class Level Information | | |
|---|---|---|
| Class | Levels | Values |
| Species | 3 | Setosa Versicolor Virginica |

| | |
|---|---|
| Number of Observations Read | 150 |
| Number of Observations Used | 150 |

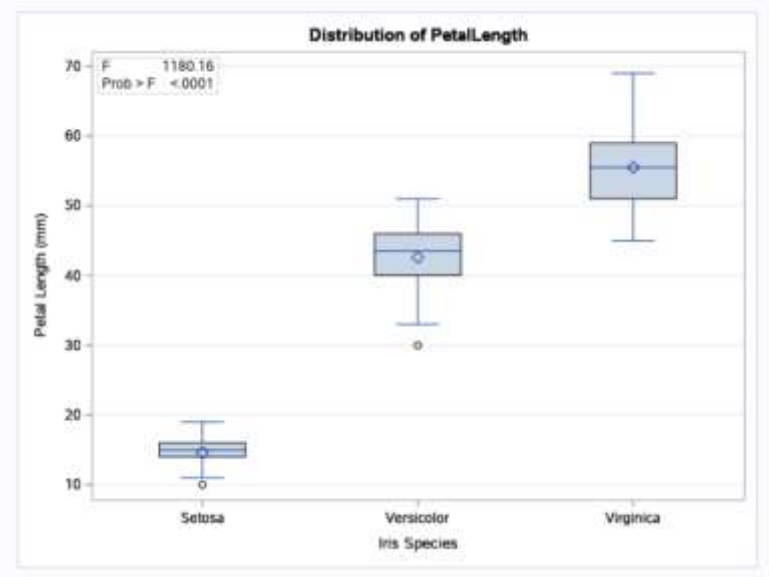
The ANOVA Procedure

Dependent Variable: PetalLength Petal Length (mm)

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 2 | 43710.28000 | 21855.14000 | 1180.16 | <.0001 |
| Error | 147 | 2722.26000 | 18.51878 | | |
| Corrected Total | 149 | 46432.54000 | | | |

| R-Square | Coeff Var | Root MSE | PetalLength Mean |
|---|---|---|---|
| 0.941372 | 11.45116 | 4.303345 | 37.58000 |

| Source | DF | Anova SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Species | 2 | 43710.28000 | 21855.14000 | 1180.16 | <.0001 |

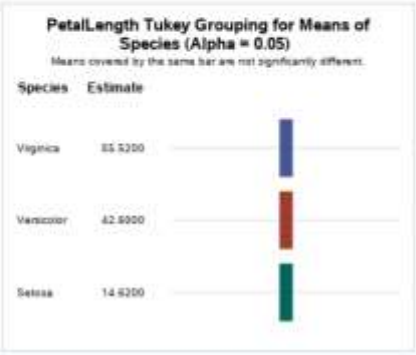



The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for PetalLength

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

| | |
|---|---|
| Alpha | 0.05 |
| Error Degrees of Freedom | 147 |
| Error Mean Square | 18.51878 |
| Critical Value of Studentized Range | 3.34842 |
| Minimum Significant Difference | 2.0376 |




## Conclusion:

Since the p-value obtained for both the case in the table is < 0.001, i.e., the p-value is < 0.01, the null hypothesis can be rejected. Therefore, the Alternate hypothesis is accepted which states that the mean of Petal Length and Width of all species of flower is different