

## ASSIGNMENT 4 – ANALYSIS OF VARIANCE USING MATLAB

The steps followed in MATLAB and results have been summarised below:

### 1. DATA IMPORT

The data from excel is imported in the form of a table.

```
% Import the data
```

```
Grades = readtable("/MATLAB Drive/PGStudents/Grades.xlsx", opts, "UseExcel", false);
```

### 2. CONVERSION OF DATA INTO ARRAY

The data imported as table is first converted into array and then plotted using box plot.

```
Grades1 = Grades{:,2:7};  
boxplot(Grades1);title('Grades');ylabel('Marks');xlabel('Faculty');
```

The marks of Engineering and Non-Engineering Students are defined separately.

```
>> Grades2 = Grades{11:end,2:7};  
>> Engg=Grades2
```

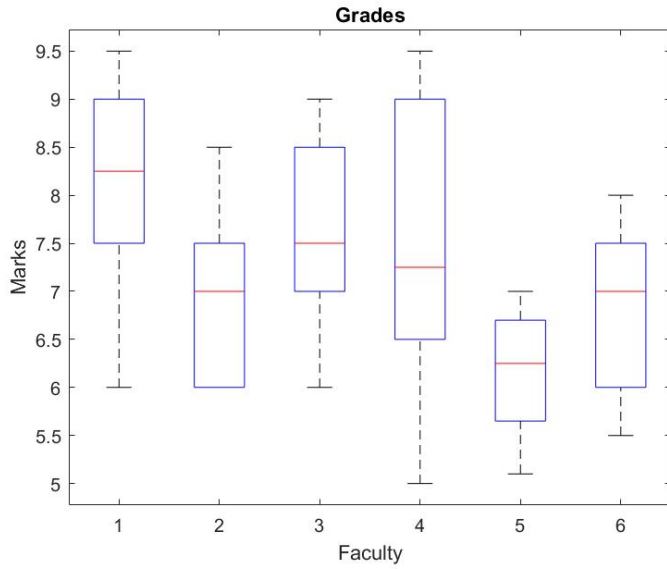
Engg =

9.5000	8.5000	8.5000	9.5000	6.8000	8.0000
7.0000	6.0000	8.0000	7.0000	5.4000	6.0000
8.5000	6.0000	7.0000	6.5000	5.4000	6.0000
7.5000	6.0000	7.0000	6.0000	5.5000	6.0000
8.0000	6.0000	7.5000	6.5000	5.9000	7.5000
8.0000	7.0000	7.0000	6.0000	5.8000	7.0000
9.0000	8.5000	9.0000	9.0000	6.9000	8.0000
9.0000	7.0000	9.0000	9.5000	6.9000	7.0000
8.0000	6.0000	7.0000	8.0000	7.0000	8.0000
9.0000	8.5000	8.5000	8.0000	6.5000	7.0000

```
>> Grades3 = Grades{1:10,2:7};  
>> NonEngg=Grades3
```

NonEngg =

8.5000	7.0000	7.0000	7.0000	6.5000	8.0000
8.5000	8.0000	7.0000	9.0000	6.6000	6.5000
9.0000	7.0000	9.0000	9.5000	6.9000	7.0000
8.5000	7.5000	8.0000	7.5000	5.8000	7.5000
7.0000	7.5000	7.5000	6.5000	5.8000	6.5000
6.0000	7.5000	7.5000	5.0000	5.1000	5.5000
9.0000	7.5000	8.0000	8.0000	6.5000	7.0000
7.5000	7.5000	7.0000	6.5000	6.0000	6.5000
7.0000	6.8000	6.0000	9.0000	6.5000	6.0000
8.0000	6.0000	8.5000	7.0000	5.4000	6.0000

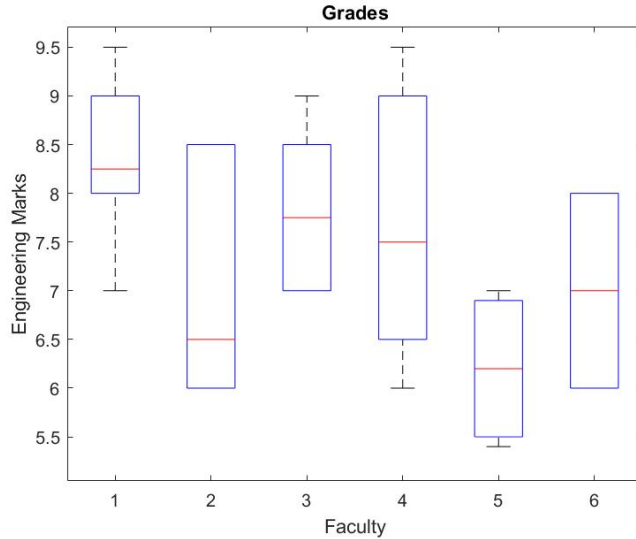


### 3. ANALYSIS OF MARKS USING BOX PLOT

The marks of each domain are plotted.

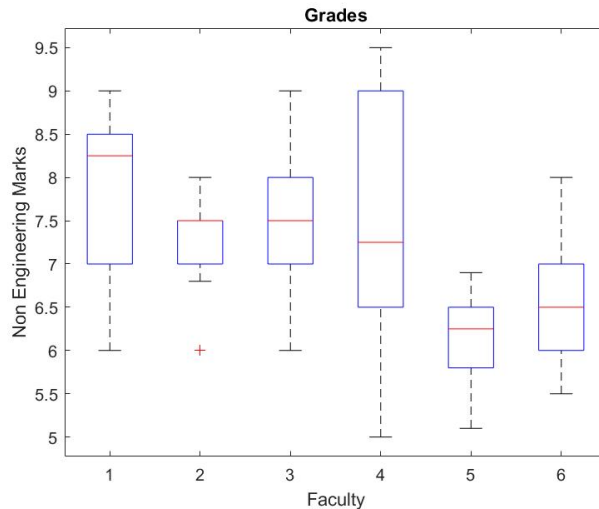
a) Engineering Students:

```
boxplot(Engg);title('Grades');ylabel('Engineering Marks');xlabel('Faculty');
```



b) Non- Engineering Students:

```
boxplot(NonEngg);title('Grades');ylabel('Non Engineering Marks');xlabel('Faculty');
```



It is observed that Engineering students have better average score than Non-Engineering students except by faculty 2 (F2).

#### 4. TWO WAY ANALYSIS IN MATLAB BY USING ANALYSIS OF VARIANCE (ANOVA)

MATLAB has built in ANOVA function. The analysis of given dataset using ANOVA. The significance level is 0.05.

```
[~,~,stats] = anova2(Grades1,10);
```

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Columns	48.634	5	9.72688	11.01	0
Rows	0.954	1	0.95408	1.08	0.3011
Interaction	1.8	5	0.36008	0.41	0.8427
Error	95.449	108	0.88379		
Total	146.838	119			

**From your analysis of the dataset, do you find any statistically significant difference between the marks given by different faculty to the students? What would you comment on the marks received by the students of engg. vs. non-engg. orientation? Is there any difference between the marks given by faculty to engg. vs. non-engg. students?**

From the above table, it is observed that:

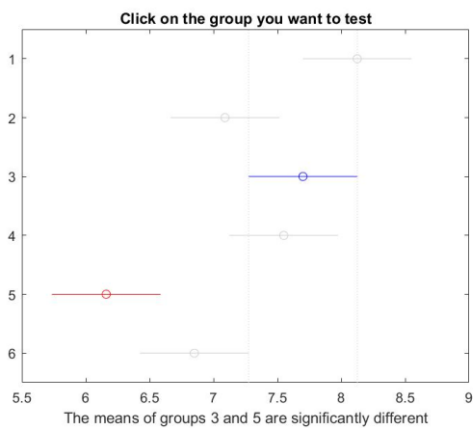
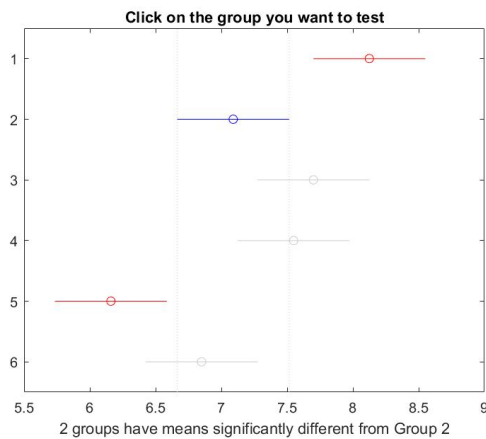
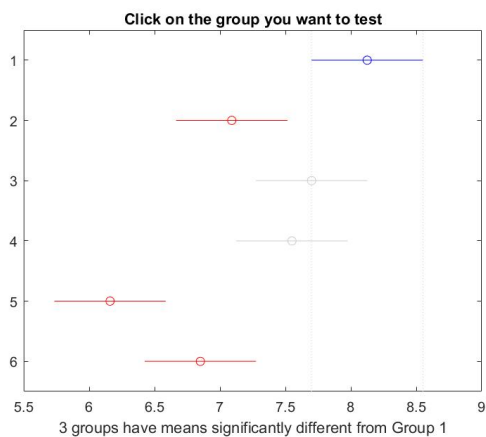
- The p-value of Interaction between Faculties & is **0.8427**. It is not statistically significant at alpha level 0.05. This indicates that there isn't much statistically significant difference in the marks given to students of both the background.
- The p-value (Rows) is **0.3011** for Engineering and Non-Engineering Students. It is not statistically significant at alpha level 0.05. This indicates that there isn't much significant difference in marks given to Engineering and Non-Engineering Students the Students.
- The p-value for marks given by faculty is **0**. This is statistically significant at alpha level 0.05. This indicates that there is statistically significant difference in the marks given by the faculty.

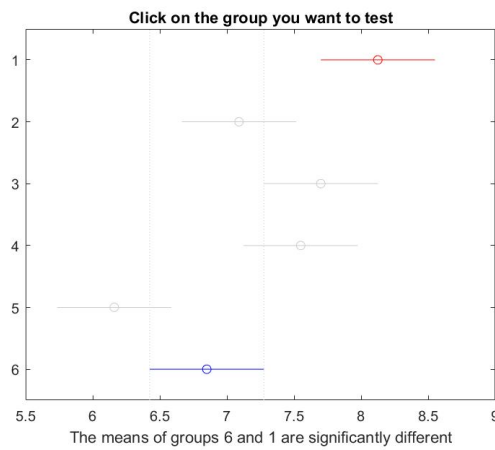
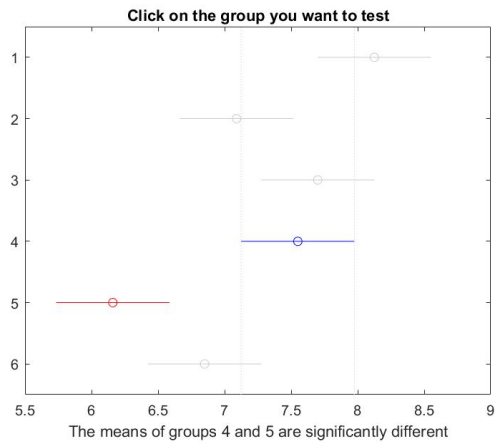
## 5. COMPARISON OF FACULTY MEMBERS

*Can you compare two faculty members at a time to find out for which faculty members you find statistically significant differences in grading the students? Please also compute the Power (probability of avoiding a Type II error) of any of these comparisons and find out the sample size for which the Power  $\geq 90\%$ .*

Using multcompare function in MATLAB, we can now compare every data

```
[c,m,h,nms] = multcompare(stats)
```





From the above charts, it is evident that following significant differences exist:

1. **F1 & F5**, ← Comparing these two faculties for Type II error
2. F1 & F6,
3. F1 & F2,
4. F2 & F5,
5. F3 & F5,
6. F4 & F5

**Comparison of two faculties and computation of Type 2 error.**

```
>> pwrt = sampsizepwr ("t",[0 4.88],1.97, [],20)

pwrt =

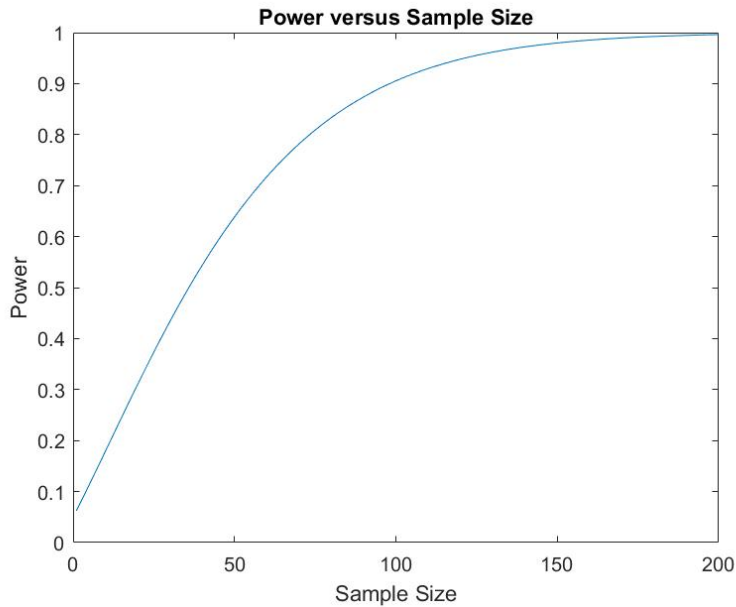
    0.4030

>> n=sampsizepwr ("z",[0 4.88],1.97,0.9,[])

n =

    65

>> pwROUT=sampsizepwr ("z",[0 4.88],1.97, [],n);
figure;
>> plot(n,pwROUT);
>> title('POWER VS SAMPLE SIZE');
>> ylabel('POWER');
>> xlabel('SAMPLE SIZE');
```



Power (probability of avoiding a Type II error) = 40.30%; Sample Size = 65. It can be concluded from above analysis that if the sample size is increased to 65, the probability of avoiding Type II error will be increased to 90%.

***Please share any other insights that you have for this analysis. What could be other unknown factors, if any, if considered could have made the analysis more meaningful?***

In case the sample distributions are not normally distributed and the sample sizes are small ( $n < 30$ ), Mann-Whitney U test (sometimes called the Wilcoxon rank-sum test) can be used to compare the differences between two independent samples. It is considered to be the nonparametric equivalent to the two-sample independent t-test. [Source: [Mann-Whitney U Test - Statology](#)]

The data can be validated in MS Excel as well.

#### **Analysis in MS Excel:**

<b>Student</b>	<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>	<b>F5</b>	<b>F6</b>
Non-Engg.	8.5	7	7	7	6.5	8
Non-Engg.	8.5	8	7	9	6.6	6.5
Non-Engg.	9	7	9	9.5	6.9	7
Non-Engg.	8.5	7.5	8	7.5	5.8	7.5
Non-Engg.	7	7.5	7.5	6.5	5.8	6.5
Non-Engg.	6	7.5	7.5	5	5.1	5.5
Non-Engg.	9	7.5	8	8	6.5	7
Non-Engg.	7.5	7.5	7	6.5	6	6.5
Non-Engg.	7	6.8	6	9	6.5	6
Non-Engg.	8	6	8.5	7	5.4	6
Engineering	9.5	8.5	8.5	9.5	6.8	8
Engineering	7	6	8	7	5.4	6
Engineering	8.5	6	7	6.5	5.4	6
Engineering	7.5	6	7	6	5.5	6

Engineering	8	6	7.5	6.5	5.9	7.5
Engineering	8	7	7	6	5.8	7
Engineering	9	8.5	9	9	6.9	8
Engineering	9	7	9	9.5	6.9	7
Engineering	8	6	7	8	7	8
Engineering	9	8.5	8.5	8	6.5	7
<b>Mean</b>	8.125	7.09	7.7	7.55	6.16	6.85
<b>Standard Deviation</b>	0.901	0.885	0.849	1.356	0.615	0.796
<b>Variance</b>	0.813	0.783	0.721	1.839	0.378	0.634

<b>Student</b>	<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>	<b>F5</b>	<b>F6</b>
Engineering	9.5	8.5	8.5	9.5	6.8	8
Engineering	7	6	8	7	5.4	6
Engineering	8.5	6	7	6.5	5.4	6
Engineering	7.5	6	7	6	5.5	6
Engineering	8	6	7.5	6.5	5.9	7.5
Engineering	8	7	7	6	5.8	7
Engineering	9	8.5	9	9	6.9	8
Engineering	9	7	9	9.5	6.9	7
Engineering	8	6	7	8	7	8
Engineering	9	8.5	8.5	8	6.5	7
<b>Mean</b>	8.35	6.95	7.85	7.6	6.21	7.05
<b>Standard Deviation</b>	0.784	1.141	0.851	1.390	0.674	0.832
<b>Variance</b>	0.614	1.303	0.725	1.933	0.454	0.692
<b>Student</b>	<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>	<b>F5</b>	<b>F6</b>
Non-Engg.	8.5	7	7	7	6.5	8
Non-Engg.	8.5	8	7	9	6.6	6.5
Non-Engg.	9	7	9	9.5	6.9	7
Non-Engg.	8.5	7.5	8	7.5	5.8	7.5
Non-Engg.	7	7.5	7.5	6.5	5.8	6.5
Non-Engg.	6	7.5	7.5	5	5.1	5.5
Non-Engg.	9	7.5	8	8	6.5	7
Non-Engg.	7.5	7.5	7	6.5	6	6.5
Non-Engg.	7	6.8	6	9	6.5	6
Non-Engg.	8	6	8.5	7	5.4	6
<b>Mean</b>	7.9	7.23	7.55	7.5	6.11	6.65
<b>Standard Deviation</b>	0.994	0.554	0.864	1.394	0.582	0.747
<b>Variance</b>	0.989	0.307	0.747	1.944	0.339	0.558

**Anova: Two-Factor With Replication**

SUMMARY	F1	F2	F3	F4	F5	F6	Total
<i>Non-Engg.</i>							
Count	10	10	10	10	10	10	60
Sum	79	72.3	75.5	75	61.1	66.5	429.4
Average	7.9	7.23	7.55	7.5	6.11	6.65	7.156666667
Variance	0.9889	0.3068	0.7472	1.9444	0.3388	0.5583	1.1150

*Engineering*

Count	10	10	10	10	10	10	60
Sum	83.5	69.5	78.5	76	62.1	70.5	440.1
Average	8.35	6.95	7.85	7.6	6.21	7.05	7.335
Variance	0.6139	1.3028	0.7250	1.9333	0.4543	0.6917	1.3576

*Total*

Count	20	20	20	20	20	20
Sum	162.5	141.8	154	151	123.2	137
Average	8.125	7.09	7.7	7.55	6.16	6.85
Variance	0.8125	0.7831	0.7211	1.8395	0.3783	0.6342

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Sample	0.9541	1	0.9541	1.0795	0.3011	3.9290
Columns	48.6344	5	9.7269	11.0059	0.0000	2.2984
Interaction	1.8004	5	0.3601	0.4074	0.8427	2.2984
Within	95.4490	108	0.8838			