ASM Assignment

Study on effect of facilities on Student Satisfaction in B-Schools

Kartheek Akella 2016A7PS0015H
Shristy Kapoor 2016B2A70688H
Abhishek Pandey 2016A7PS0081H
Rashi Jain 2017A7PS0082H
Agam Sachan 2016A2TS0484H
Anshuman Pati 2016B4A70470H

Contents

Dataset Description

The given data consists of responses to a pilot questionnaire about business schools. Several insights can be derived from the given data which will be mentioned in detail in the further sections. For an overall understanding, some basic insights, explanations and assumptions about the data are enumerated below.

1. The data consists of responses from people of three colleges - VS, ASB and OT.
2. The yearly intake of the colleges is assumed to be of the particular branch of the program (HR, Finance, etc.) which is not explicitly present in the dataset, to which the respondent belongs.
3. The location of the colleges cannot be ascertained with confidence due to no clear majority in the responses for that column. This might be due to the fact that perception of whether a place is a village, town, city or metro-city greatly varies from person to person.
4. The program enrolled for the faculty is assumed to be the MBA program the faculty was enrolled in, not necessarily in the same college.

1 Demographic Analysis

1.1 Categorical Features

1.1.1 Type of Institution

Name of Institute	Type
VS	Public
ASB	Private
OT	Public/Private

The mixed responses for OT most likely indicate that when an institute is private but is still aided by the government, or was taken over by the government causing a blur between what type of institute it would be called at that point. It might also be a result of the survey being taken at different points of time.

1.1.2 Program Offered

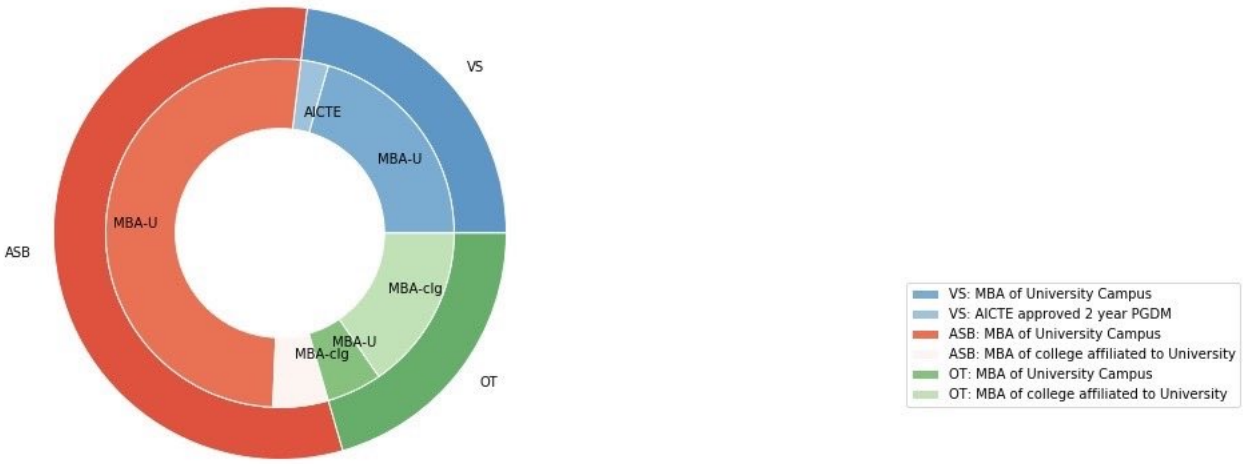

Figure 1: Respondents' Program Offered

1.1.3 Role of Respondent

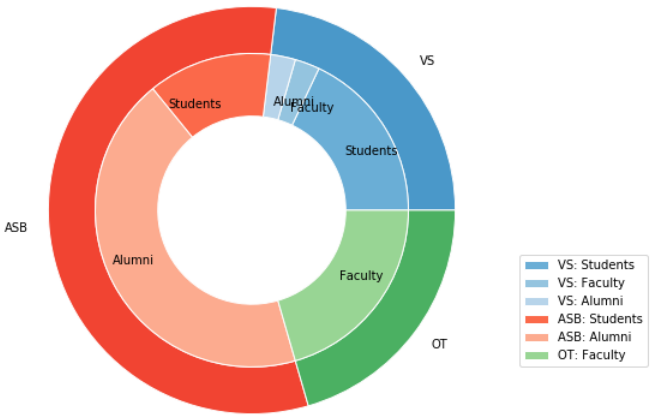

Figure 2: Role of Respondents

1.1.4 Gender

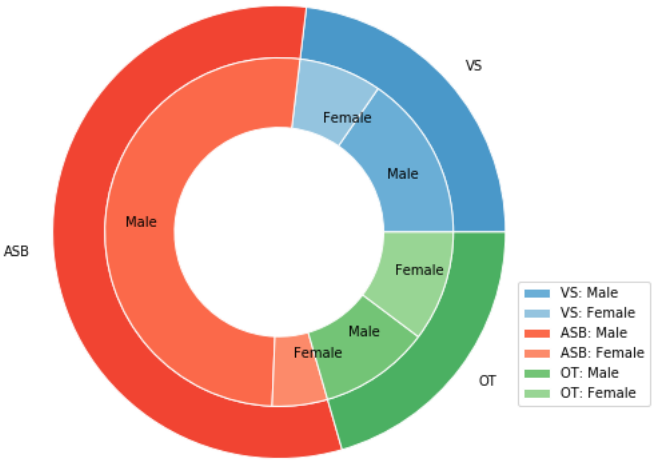

Figure 3: Male and Female Respondents

1.1.5 Native Place

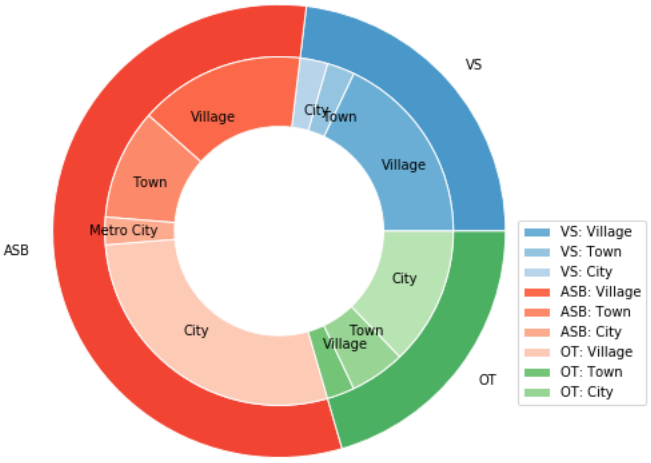

Figure 4: Native Places of Respondents

1.1.6 Stream

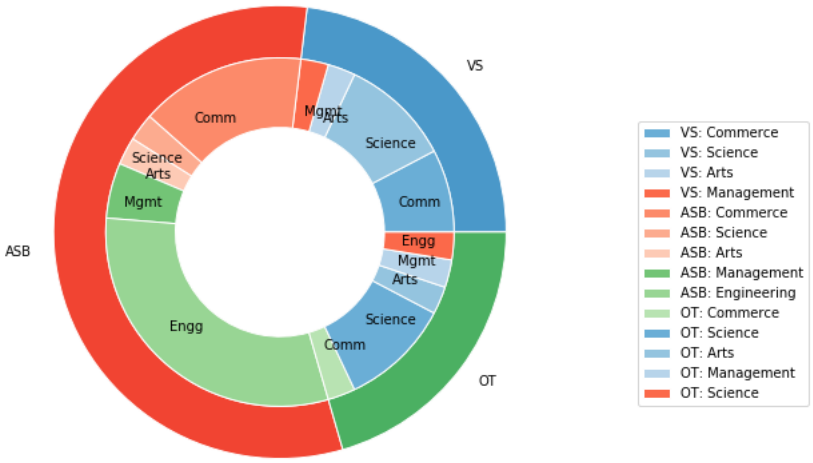

Figure 5: Streams

1.1.7 Highest Qualification

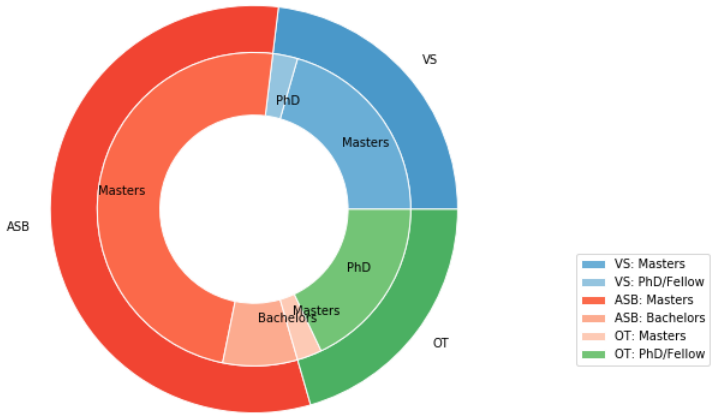

Figure 6: Highest Qualifications

1.2 Quantitative Features

We received multiple responses for placements of the same institute which we can assume as sampling values to the same parameter of the population. For quantitative variables we have calculated interval estimates with 95% confidence. The results are listed below.

Assumptions Made:

- The response received is an unbiased estimator of the population parameter.
- Responses follow nearly **Normal Distribution**

1.2.1 Previous Year Placement

The 95% Confidence Interval is found to be (66.95, 84.39). From Figure 8 we can observe that there are two outliers.

Descriptives

			Statistic	Std. Error
PrevYearPlacement	Mean		75.6692	4.30809
	95% Confidence Interval for Mean	Lower Bound	66.9480	
		Upper Bound	84.3905	
	5% Trimmed Mean		78.2507	
	Median		80.0000	
	Variance		723.824	
	Std. Deviation		26.90398	
	Minimum		.00	
	Maximum		100.00	
	Range		100.00	
	Interquartile Range		37.40	

Figure 7: Few Statistics on Previous Year Placements

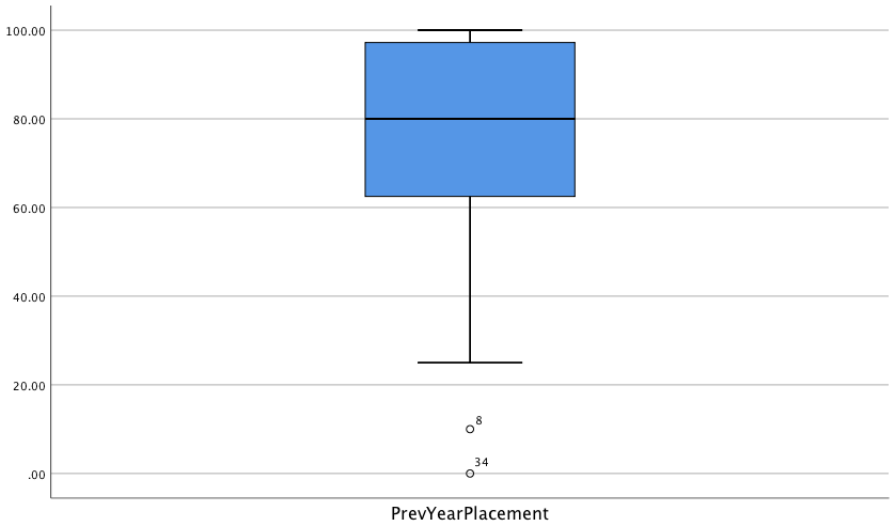

Figure 8: Box Plot of Previous Year Placements

1.2.2 Current Year Placement

The 95% Confidence Interval is found to be (21.87, 41.84).

Descriptives

			Statistic	Std. Error
CurrYearPlacement	Mean		31.6555	5.03123
	95% Confidence Interval for Mean	Lower Bound	21.4703	
		Upper Bound	41.8407	
	5% Trimmed Mean		29.6173	
	Median		10.0000	
	Variance		987.218	
	Std. Deviation		31.42002	
	Minimum		.00	
	Maximum		100.00	
	Range		100.00	
	Interquartile Range		56.00	

Figure 9: Few Statistics on Current Year Placements

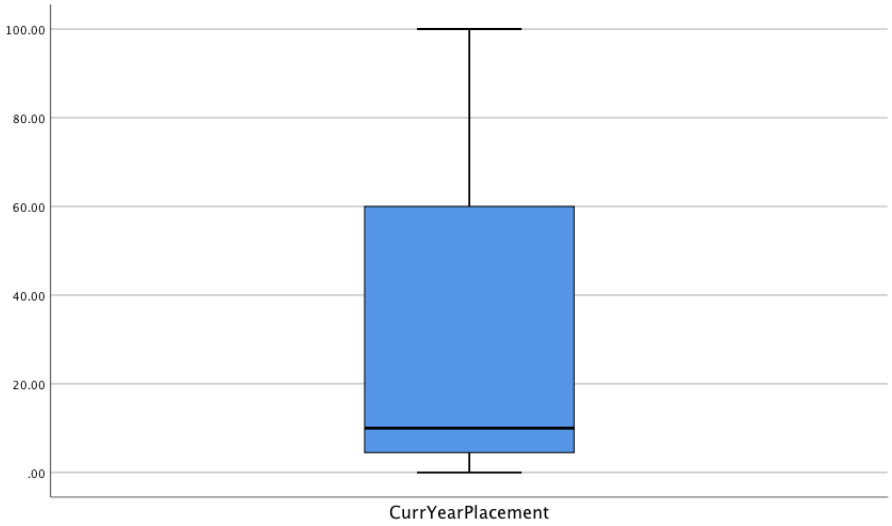

Figure 10: Box Plot of Current Year Placements

1.2.3 Age

Professors
The 95% Confidence Interval is found to be (42.10, 58.78). From the Figure 12 we can observe that there is an outlier.

Descriptives

			Statistic	Std. Error
Age	Mean		50.44444	3.617516
	95% Confidence Interval for Mean	Lower Bound	42.10244	
		Upper Bound	58.78645	
	5% Trimmed Mean		50.54938	
	Median		51.00000	
	Variance		117.778	
	Std. Deviation		10.852547	
	Minimum		33.000	
	Maximum		66.000	
	Range		33.000	
	Interquartile Range		17.500	
	Skewness		-.112	.717
	Kurtosis		-.514	1.400

Figure 11: Few Statistics on Age of Professors

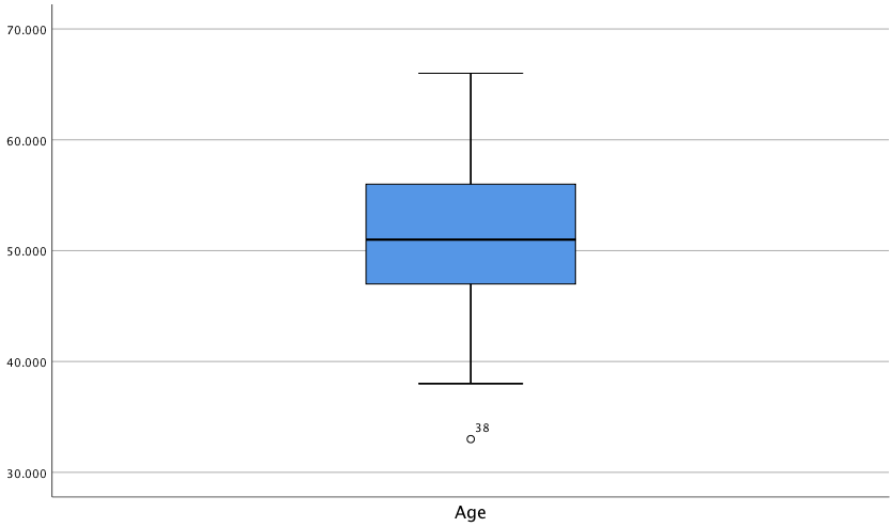

Figure 12: Box Plot of Age of Professors

Students

The 95% Confidence Interval is found to be (42.10, 58.78). From the Figure 14 we can observe that there are three outliers.

Descriptives

			Statistic	Std. Error
Age	Mean		22.50000	.484612
	95% Confidence Interval for Mean	Lower Bound	21.43338	
		Upper Bound	23.56662	
	5% Trimmed Mean		22.44444	
	Median		22.00000	
	Variance		2.818	
	Std. Deviation		1.678744	
	Minimum		20.000	
	Maximum		26.000	
	Range		6.000	
	Interquartile Range		1.500	
	Skewness		.968	.637
	Kurtosis		.704	1.232

Figure 13: Few Statistics on Age of Students

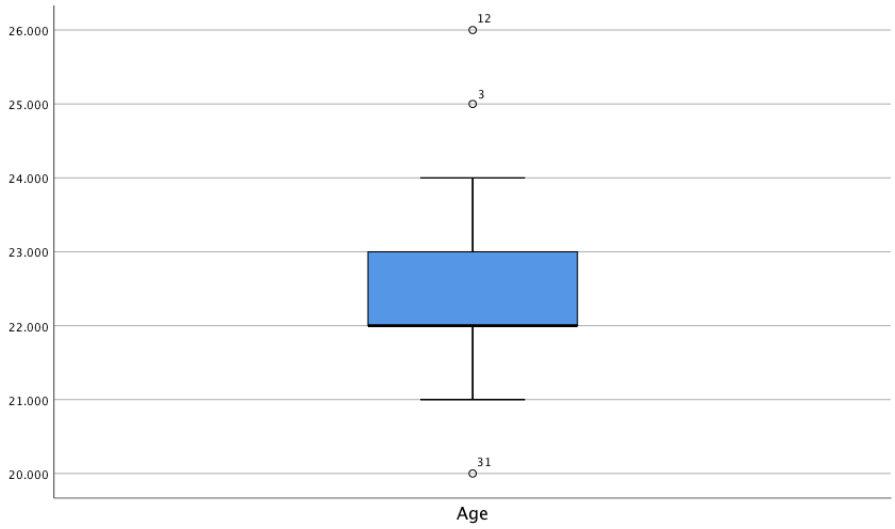

Figure 14: Box Plot of Age of Students

1.2.4 Intake

We are assuming that the intake reported by the respondents are the intakes of their particular branch. Thus, the different branches can have different intakes.

1. **VS**

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Intake_VS	9	98.000	2.000	100.000	50.44444	30.834685	950.778
Valid N (listwise)	9						

Figure 15: Intake of VS Institute

2. **ASB**

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Intake_ASB	22	198.000	2.000	200.000	164.50000	55.251589	3052.738
Valid N (listwise)	22						

Figure 16: Intake of ASB Institute

3. OT

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Intake_OT	8	1160.000	40.000	1200.000	243.75000	390.235735	152283.929
Valid N (listwise)	8						

Figure 17: Intake of OT Institute

We can see that the standard deviation for OT institute is very high compared to that of VS and ASB.

2 T-tests

We compared the previous year and current year placements of the institutes using paired t-tests. We first checked for a global trend in placements followed across all institutes followed by analysis of the trend institute wise.

2.1 Overall Trend

2.1.1 Framing Hypothesis

H_o: The placements of previous year and current year do not differ significantly.
H_a: The placements of previous year and current year **do** differ significantly.

2.1.2 Test and Distribution

We shall use paired t-test as the respondents are same.

2.1.3 Significance Level

We use $\alpha = 0.05$ level of significance

2.1.4 Decision Rule

- $p > 0.05$; fail to reject H_o
- $p < 0.05$; reject H_o

2.1.5 Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Previous Year Placements	75.669	39	26.9040	4.3081
	Current Year Placements	30.80221	39	31.824064	5.095929

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Previous Year Placements & Current Year Placements	39	-.556	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Previous Year Placements - Current Year Placements	44.867026	51.845135	8.301866	28.060777	61.673274	5.404	38	.000

Figure 18: Paired t-test to check if placements differ significantly overall

2.1.6 Conclusion

As observed from the table. the p-value has a value $p = 0.000$. Since $p < \alpha = 0.05$, we reject our null hypothesis. Hence, a statistically significant difference between placements of current and previous year exists.

2.2 Trend in each individual institute

2.2.1 Framing Hypothesis

H_o : The placements of previous year and current year do not differ significantly.
H_a : The placements of previous year and current year **do** differ significantly.

2.2.2 Test and Distribution

We shall use paired t-test as the respondents are same.

2.2.3 Significance Level

We used $\alpha = 0.05$ level of significance

2.2.4 Decision Rule

- $p > 0.05$; fail to reject H_o
- $p < 0.05$; reject H_o

2.2.5 Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Previous year placements VS	55.444	9	22.8589	7.6196
	Current Year Placement VS	66.7778	9	6.99603	2.33201
Pair 2	Previous year Placement ASB	92.8227	22	9.83313	2.09643
	Current year placement ASB	15.6948	22	29.12891	6.21030
Pair 3	Previous year placements OT	51.2500	8	30.16502	10.66495
	Current year placements OT	31.8750	8	22.66802	8.01436

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Previous year placements VS & Current Year Placement VS	9	.758	.018
Pair 2	Previous year Placement ASB & Current year placement ASB	22	-.375	.085
Pair 3	Previous year placements OT & Current year placements OT	8	-.904	.002

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Previous year placements VS - Current Year Placement VS	-11.33333	18.13836	6.04612	-25.27571	2.60904	-1.874	8	.098
Pair 2	Previous year Placement ASB - Current year placement ASB	77.12791	34.06095	7.26182	62.02613	92.22969	10.621	21	.000
Pair 3	Previous year placements OT - Current year placements OT	19.37500	51.58055	18.23648	-23.74742	62.49742	1.062	7	.323

Figure 19: Pair t-test to check if differ significantly for each individual institute

2.2.6 Conclusions

- **VS:** As observed from the table, the p-value has a value $p = 0.098$. Since $p > \alpha = 0.05$, we fail to reject the null hypothesis. Hence, there is no statistically significant difference between placement of previous and current year.

- **ASB:** As observed from the table, the p-value has a value $p = 0.000$. Since $p < \alpha = 0.05$, we reject our null hypothesis. Hence, a statistically significant difference exists between placement of the institute in previous and current year.

- **OT:** As observed from the table, the p-value has a value $p = 0.323$. Since $p > \alpha = 0.05$, we fail to reject the null hypothesis. Hence, placement for both the years do not show statistically significant difference.

3 Analysis on Categorical Features

3.1 Students v Experience

Previous Experience	7	58.33%
No Previous Experience	5	41.67%

3.2 Gender v Stream

H_0: Gender and Stream of Respondent are independent.
H_a: Gender and Stream of Respondent are dependent.
From the figure 3.2 it is clear that p-value (from likelihood ratio) < 0.05, so we reject H_0 and accept H_a

Gender * Stream Crosstabulation

Count

		Stream					
		Arts	Commerce	Engineer	Manageme	Science	Total
Gender	Fema	1	0	1	1	6	9
	Male	2	10	12	3	3	30
Total		3	10	13	4	9	39

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	14.553[a]	4	.006
Likelihood Ratio	15.310	4	.004
N of Valid Cases	39		

a. 7 cells (70.0%) have expected count less than 5. The minimum expected count is .69.

3.3 Highest Qualification v Native Place

H_0: Qualification and Native Place of Respondent are independent.
H_a: Qualification and Native Place of Respondent are dependent.
From the figure 3.2 it is clear that p-value (from likelihood ratio) < 0.05, so we reject H_0 and accept H_a

Qualification * NativePlace Crosstabulation

Count

		NativePlace				
		City	Metro C	Town	Village	Total
Qualification	Bachelors degr	2	1	0	0	3
	Masters Degree	10	0	5	13	28
	PhD/ Fellow	5	0	2	1	8
Total		17	1	7	14	39

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	17.343[a]	6	.008
Likelihood Ratio	12.300	6	.056
N of Valid Cases	39		

a. 9 cells (75.0%) have expected count less than 5. The minimum expected count is .08.

4 Factor Analysis

Factor analysis is a way to find the underlying components that explain the underlying correlations among a set of variables. It is used to reduce the data into a small number of components that represent the maximum variance observed among all the variables. In this study, the factor analysis is done on the dataset with 60 questions on a Likert scale of 1 to 5. Factor analysis is done to divide the 60 questions into different components and to eliminate inconsistent components.
Extraction Procedure - Principal Component Analysis
Selection procedure - eigenvalue < 1
Rotation method - varimax

4.1 Factor Analysis of B

4.1.1 Descriptive Statistics

REFER TO TABLE 0

4.1.2 Communalities

Communalities indicate how much variance of the initial variables is accounted for by the new components. Initial communalities are the variance estimate before the factor analysis and it is accounted for by all the variables without reduction, whereas the extraction communalities are the communalities after

factor extraction, accounted for by the extracted components. The small values indicate that the variables do not fit well with the factor solution and the questions could very well be neglected. In this study, there are 60 questions and there initial and extraction communalities are shown as follows:
REFER TO TABLE 1

4.1.3 Total Variance Explained

At first there are 60 questions. PCA is used on these 60 questions. The PCA gives us 60 eigenvalues as result which represent 60 components, but only the high eigen-values represent some meaningful component. We have taken the standard eigen-value of 1 and taken all the components having eigen-value greater than 1. The total number of components come to 16 with a variance of 89.305 %.
REFER TO TABLE 2

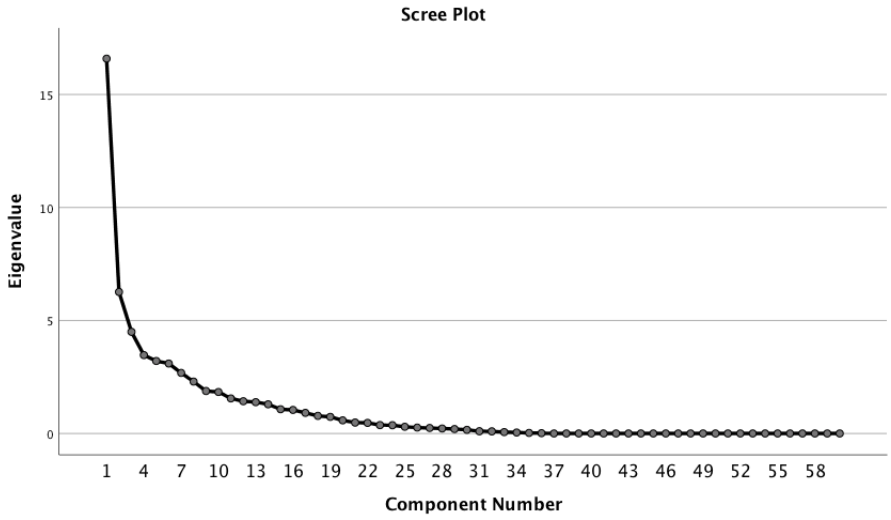

Figure 20: Scree Plot of B Matrix

4.1.4 Component Matrix

Component matrix shows the correlation between the questions and the components found. This component matrix shows multicollinearity between variables and to avoid this we use rotation method (varimax method). This new matrix is called the rotated component matrix. Rotation is used to reduce the number of components in which the variable shows high values.
REFER TO TABLE 3 FOR COMPONENT MATRIX
REFER TO TABLE 4 FOR ROTATED COMPONENT MATRIX

4.1.5 Tabulated Form

This table has the component number and rotated component matrix value of highest magnitude for each question.
REFER TO TABLE 5

4.1.6 Component Score Matrix

This matrix is created by the average of the values of the correlation between the questions and the components. These values are tabulated for each person.

```
COMPUTE k1=(B15+B19+B20+B21+B25+B34+B36+B48+B48+B49+B50+B52+B54+B60)/13.
EXECUTE.
COMPUTE k2=(B2+B3+B4+B8+B17+B23+B24+B37)/8.
EXECUTE.
COMPUTE k3=(B5+B6+B11+B46)/4.
EXECUTE.
COMPUTE k4=(B10+B16+B18)/3.
EXECUTE.
COMPUTE k5=(B44+B45)/2.
EXECUTE.
COMPUTE k6=(B12+B53+B56+B57)/4.
EXECUTE.
COMPUTE k7=(B27+B28+B29)/3.
EXECUTE.
COMPUTE k8=(B31+B35+B38)/3.
EXECUTE.
COMPUTE k9=(B22+B32+B39+B43)/4.
EXECUTE.
COMPUTE k10=(B40+B41+B47+B58)/4.
EXECUTE.
COMPUTE k11=(B13+B14+B55)/3.
EXECUTE.
COMPUTE k12=(B1+B9-B30)/3.
EXECUTE.
COMPUTE k13=(B51+B59)/2.
EXECUTE.
COMPUTE k14=B33.
EXECUTE.
COMPUTE k15=(B26+B42)/2.
EXECUTE.
COMPUTE k16=B7.
EXECUTE.
```

REFER TO TABLE 6

After these values are found, descriptive analysis on these values is done and the mean, standard deviation etc. is found.
REFER TO TABLE 7
Now, the means for the different components are analysed. As we can see the means are more or less consistent. There is only one mean which is out of the place, that is the mean for component 12 (1.63). That means that the questions for the component 12 are not reliable and hence can be dropped. Other than this there can be 2 categories - one with means in the range (4,5) and other means in the range (3,4). Thus, in total we have divided the components into 3 categories. The odd one out is the component 12 which can be neglected.

4.2 Factor Analysis of C

4.2.1 Descriptive Statistics

REFER TO TABLE 8

4.2.2 Communalities

REFER TO TABLE 9

4.2.3 Total Variance Explained

The total number of components come to **13** with a variance of **87.84** %.
REFER TO TABLE 10

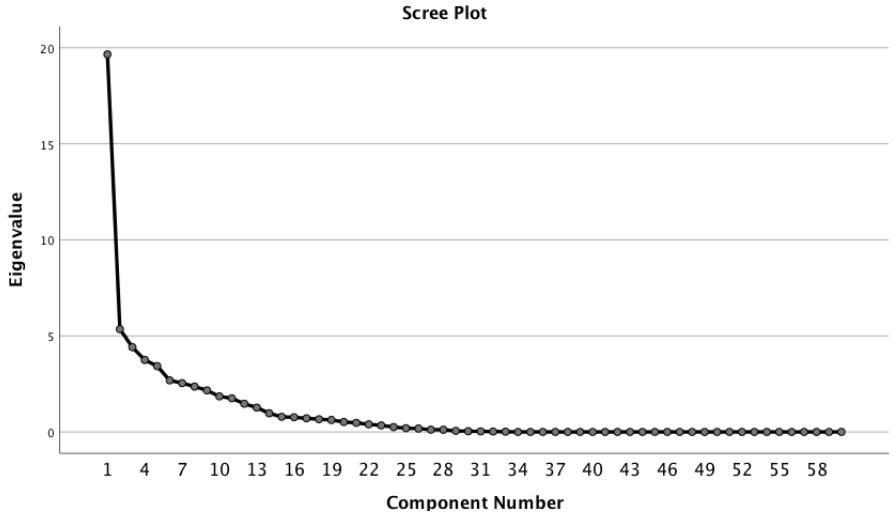

Figure 21: Scree Plot of C Matrix

4.2.4 Component Matrix

REFER TO TABLE 11 FOR COMPONENT MATRIX
REFER TO TABLE 12 FOR ROTATED COMPONENT MATRIX

4.2.5 Tabulated Form

REFER TO TABLE 13

4.2.6 Component Score Matrix

This matrix is created by the average of the values of the correlation between the questions and the components. These values are tabulated for each person.

```
COMPUTE X1=(C3+C28+C32+C36+C42+C47+C48+C49+C52+C53+C55+C57+C58+C59)/14.
EXECUTE.
COMPUTE X2=(C20+C24+C26+C29+C30+C37+C43)/7.
EXECUTE.
COMPUTE X3=(C14+C15+C16+C18+C19+C27+C31+C34)/8.
EXECUTE.
COMPUTE X4=(C33+C35+C38+C39+C41+C46)/6.
EXECUTE.
COMPUTE X5=(C5+C8+C11+C12)/4.
EXECUTE.
COMPUTE X6=(C1+C6+C7+C17+C40)/5.
EXECUTE.
COMPUTE X7=(C4+C44+C45)/3.
EXECUTE.
COMPUTE X8=(C2+C56)/2.
EXECUTE.
COMPUTE X9=(C22+C23+C25)/3.
EXECUTE.
COMPUTE X10=(C13+C21-C60)/3.
EXECUTE.
COMPUTE X11=(C51+C54)/2.
EXECUTE.
COMPUTE X12=(C50)/1.
EXECUTE.
COMPUTE X13=(C9+C10)/2.
EXECUTE.
```

REFER TO TABLE 14

After these values are found, descriptive analysis on these values is done and the mean, standard deviation etc. is found.

REFER TO TABLE 15

Now, the means for the different components are analysed. As we can see the means are more or less consistent. There is only one mean which is out of the place, that is the mean for component 10 (1.39). That means that the questions for the component 12 are not reliable and hence can be dropped. Other than this there can be 2 categories - one with means in the range (4,5) and other means in the range (3,4). Thus, in total we have divided the components into 3 categories. The odd one out is the component 10 which can be neglected.

4.3 Factor Analysis of D

REFER TO TABLE 16

4.3.1 Communalities

REFER TO TABLE 17

4.3.2 Total Variance Explained

The total number of components come to **3** with a variance of **78.28** %.

REFER TO TABLE 18

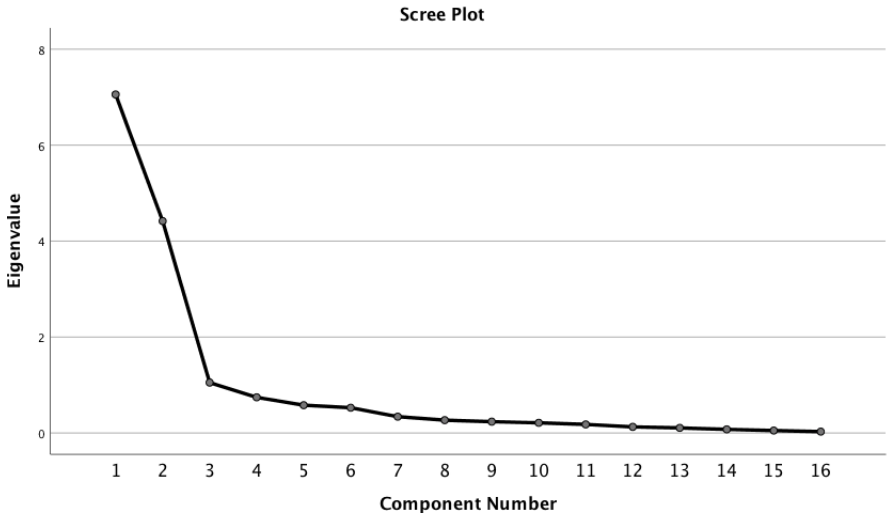

Figure 22: Scree Plot of D Matrix

4.3.3 Component Matrix

REFER TO TABLE 19 FOR COMPONENT MATRIX
REFER TO TABLE 20 FOR ROTATED COMPONENT MATRIX

4.3.4 Tabulated Form

REFER TO TABLE 21

4.3.5 Component Score Matrix

This matrix is created by the average of the values of the correlation between the questions and the components. These values are tabulated for each person.

```
COMPUTE Y1=(D1+D2+D4+D6+D9+D10+D12+D14)/8.
EXECUTE.
COMPUTE Y2=(D3+D7+D8+D11+D15)/5.
EXECUTE.
COMPUTE Y3=(D5+D13+D16)/3.
EXECUTE.
```

REFER TO TABLE 22
After these values are found, descriptive analysis on these values is done and the mean, standard deviation etc. is found.
REFER TO TABLE 23
Now, the means for the different components are analysed. As we can see the means are more or less consistent.