



FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES

STA610

SAS PROGRAMMING

PROJECT TITLE:

**ACADEMIC AND EMPLOYABILITY FACTORS INFLUENCING
PLACEMENT IN XYZ CAMPUS**

GROUP: N4CS2415T3

NO	GROUP MEMBERS	STUDENT ID
1.	NURUL AMIRA BINTI ROHI'SHAM	2020996921
2.	NURUL HAZLIN BINTI ZULKIFLI	2020995029
3.	SITI HAJAR BINTI MUHAMAD JOHAN	2020963691

PREPARED FOR:

NOOREZATTY MOHD YUSOP

SUBMISSION DATE:

16 JULY 2021

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1-3
1.1 Background of Study	1
1.2 Description of Data Set	2
1.3 Research Objectives	3
CHAPTER 2: METHODOLOGY	4-8
2.1 Introduction	4
2.2 Data Management	
2.2.1 Assigning a Libref	4
2.2.2 Reading Excel Spreadsheets	4
2.2.3 The Descriptor Portion	4
2.2.4 Printing Selected Variables	5
2.2.5 Defining Titles	5
2.3 Descriptive Statistics	6
2.3.1 Bar and Pie Chart	6
2.3.2 Frequency Report	6
2.3.3 Summary Statistics	6
2.4 Inferential Statistics	
2.4.1 Test for Normality	7
2.4.2 Logistic Regression Analysis	7
2.4.3 Mann-Whitney U Test	8
2.4.4 Kruskal-Wallis Test	8
CHAPTER 3: RESULT AND ANALYSIS	10-20
3.1 Introduction	10
3.2 Descriptive Statistics	
3.2.1 Summary Statistics	10
3.2.2 One-way Frequency Report	11
3.2.3 The Bar Chart of Student Placement by Status	11
3.2.4 Pie Chart of Percentage of Work Experience	12
3.3 Inferential Statistics	
3.3.1 Normality Test	13
3.3.2 Logistic Regression Analysis	18
3.3.3 Mann-Whitney U Test	19
3.3.4 Kruskal-Wallis Test	20

CHAPTER 4: CONCLUSION AND DISCUSSION	21
CHAPTER 5: REFERENCES	22
CHAPTER 6: APPENDICES	23

LIST OF TABLES

TABLE		2-20
1.1	The Description of Data Set	2
3.1	The Summary Statistics of Means Procedure	10
3.2	The Frequency of Post Graduation (MBA)- Specialization	11
3.3	Normality Test for Salary with No Experience	14
3.4	Normality Test for Salary with Experience	14
3.5	Normality Test for Employability Test Percentage in Arts	15
3.6	Normality Test for Employability Test Percentage in Commerce	16
3.7	Normality Test for Employability Test Percentage in Science	16
3.8	Analysis of Maximum Likelihood Estimates for Logistic Regression Analysis	18
3.9	The Median of Salary that being Offered among Candidates whether has a Work Experience or not	19
3.10	The Mean Score of Employability Test Percentage by Specialization in Higher Secondary Education	20

LIST OF FIGURES

FIGURE		12-17
3.1	Frequency of Student Placements by Status	12
3.2	The Percentage of Work Experience	13
3.3	Normal P-P Plot of Salary with No Experience and with Experience	14
3.4	Normal P-P Plot of Employability Test Percentage in Arts and Commerce	16
3.5	Normal P-P Plot of Employability Test Percentage in Science	17

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Learning is one of the important thing that everyone does in their daily life because people can get more knowledge of what happen in this world and learning something new. High school life is not the end of learning; people can continue their studies to a higher institution. The higher the level of institution, the more experience that need to be qualified if people want to continue their studies. People also can get a special treatment which can help them or ease the burden of the students. It is quite difficult for people to continue their studies since there are a lot of people want to learning in a higher institution unfortunately the institution cannot afford all the people to get into the campus. According to Nauta et al. (2009), students with higher level of employability are more likely to meet the demands of employers in terms of productivity.

Employability was viewed by the universities as a skill from the students as it allows them to adapt with the change and be more employable (Hager, 1996). Placement in university are hard to get. Hence, it might be a failure if there is someone who want to get the placement without a specialty that the higher institution looks for (Smith et al., 2018). Therefore, there are a few factors that need to be look into so that it can increase the percentage of people to get into the campus. A past study had done listed several factors that might be the reason of the placement, the factors are percentage education of secondary and higher secondary. It will increase the opportunity to be choose because the people had a knowledge of what they had already learn and it will make it easier for them to learn in higher institution. Other than that, boards of education for secondary and higher secondary, specialization in higher secondary education also are the factors that might influence the placement in higher institution. Next, percentage of degree and field of degree education are the other factors be looking for. There are a few field of degree that hard to get since there are many people chose but if the field of degree is not quietly familiar for the people, therefore it will have less competition. Work experience, salary offers, percentage of employability test and post graduation for specialization are the last factors be study.

1.2 Description of Data Set

This study is used secondary data that obtain from Kaggle website, <https://www.kaggle.com/benroshan/factors-affecting-campus-placement>. This data set consists of placement data of students in XYZ campus. This data is included secondary and higher secondary school percentage and specialization also includes degree specialization, type and work experience and salary offers to the placed students. About the data in this study, there is 215 data details of placements students in XYZ campus. The total of variables in this data is 14 with 5 continuous variables and 9 categorical variables.

Table 1.1: The Description of Data Set

Variable	Variable Names	Description	Level of measurement
Y	Gender	Gender: M = Male F = Female	Binary
X	ssc_p	Secondary Education percentage - 10th Grade	Interval
	ssc_b	Board of Education Central / Others	Nominal
	hsc_p	Higher Secondary Education percentage - 12th Grade	Interval
	hsc_b	Board of Education Central / Others	Nominal
	hsc_s	Specialization in Higher Secondary Education Commerce, Science and Other	Nominal
	degree_p	Degree percentage	Interval
	degree_t	Under Graduation (Degree type)- Field of degree education	Nominal

		Comm & Mgmt, Sci & Tech and Other	
	workex	Work Experience	Binary
	etest_p	Employability test percentage (conducted by college)	Interval
	specialization	Post Graduation (MBA)- Specialization Mkt & Fin and Mkt & HR	Nominal
	mba_p	Post Graduation (MBA) percentage	Interval
	status	Status of placement Placed/Not placed	Nominal
	salary	Salary offered by corporate to candidates	Interval

1.3 Research Objectives

The study purpose is to investigate factors influencing placement in campus. Thus, several objectives of the study are listed as follows:

1. To identify the factors that influenced a candidate in getting placed.
2. To determine whether there is difference in salary between those with work experience or not.
3. To determine the employability test percentage differ by specialization in higher secondary education.

CHAPTER 2

METHODOLOGY

2.1 Introduction

In order to achieve the three objectives in this study, there are a few methods that will be used. The original data in excel with xlsx format will be convert into SAS dataset form. After convert the file, apply all the suitable SAS procedure and step to the data to obtain the results.

2.2 Data Management

2.2.1 Assigning a Libref

The LIBNAME statement is use to assign a library reference(libref) to the SAS data library and the Libref for this study is 'places'.

General form of the LIBNAME statement:

```
LIBNAME libref 'SAS-library';
```

2.2.2 Reading Excel Spreadsheets

From the original data in excel spreadsheets, the IMPORT procedure is use to create a SAS dataset.

General form of the IMPORT statement:

```
PROC IMPORT DATAFILE='external-file-name'  
            OUT=<libref>'SAS-data-set'  
            DBMS='identifier';  
RUN;
```

2.2.3 The Descriptor Portion

CONTENTS procedure is to display the description of the dataset. The descriptor portion contains the following metadata:

- General properties about the dataset such as dataset name (place.placements), number of observations (215 observations) and variables in the dataset (14 variables).
- Variable properties such as variable name, type, length, format, informat, label).

General form of the CONTENTS procedure:

```
PROC CONTENTS DATA='SAS-data-set';  
RUN;
```

2.2.4 Printing Selected Variables

PRINT procedure is to display all observations and variables in the dataset while VAR statement use when to select variables that need to be included in the report.

General form of PRINT procedure:

```
PROC PRINT DATA='SAS-data-set';  
RUN;
```

General form of VAR statement:

```
PROC PRINT DATA='SAS-data-set';  
    VAR variable(s);  
RUN;
```

2.2.5 Defining Titles

Appropriate titles will be assign to some of the SAS outputs. It will appear at the top of the page in the report. Other than that, the title can specify the color, font and height that needed in the report.

General form of the TITLE statement:

```
TITLEn 'text';
```

2.3 Descriptive Statistics

2.3.1 Bar and Pie Chart

To produce a chart, PROC GCHART statement will be used. Bar and pie chart can be use in the statement to display the suitable variables such as the distribution, average value and percentage of the variables. For this study, vertical bar chart is use since it suitable to display the frequency of student placements by based on status. The y-axis and x-axis are representing by frequency and status respectively. Besides that, pie3d also is use to display the percentage of work experience among students.

General form of the PROC GCHART statement:

```
PROC GCHART DATA='SAS-data-set';  
    VBAR chart-variable...</options>;  
    PIE3D chart-variable...</options>;  
RUN;
```

2.3.2 Frequency Report

Frequency report provides the detail about the categorical and numerical variables. By default, PROC FREQ will display the frequency, percentage, row percentage and column percentage of the variables. In this study, one-way frequency report will be produce. It will display the frequency, percentage, cumulative frequency and cumulative percentage of post_graduation on Specialization.

General form of the PROC FREQ statement:

```
PROC FREQ DATA='SAS-data-set';  
    TABLES variable(s);  
RUN;
```

2.3.3 Summary Statistics

Summary statistics display all the detail of the numerical variables. In this study, it will display a simple descriptive statistic such as number of observations, mean, standard deviation, minimum and maximum values of the variables.

General form of the PROC MEANS statement:

```
PROC MEANS DATA='SAS-data-set';  
RUN;
```

2.4 Inferential Statistics

2.4.1 Test for Normality

The distribution of the data must be normally distributed to use the data for this study. These hypothetical testing related to differences are classified as parametric and non-parametric test. Several requirements must be met to ensure the result of parametric test is reliable. To test whether the data is normal or not normal, the data will be tested using Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling. If the data set do not meet this assumption, non-parametric test analysis will be used. If the tests are greater than significance value 0.05, indicates that the data are normal.

General form of the PROC UNIVARIATE statement:

```
PROC UNIVARIATE DATA = SAS-data-set NORMAL;  
    CLASS variable(s);  
    VAR variable(s);  
    PROBPLOT/NORMAL (variable(s));  
RUN;
```

Hypothesis Statement:

H_0 : Data is normally distributed

H_1 : Data is not normally distributed

Decision rule:

If the p-value < 0.05, reject H_0

Conclusion:

Conclude that the data is not normally distributed.

2.4.2 Logistic Regression Analysis

Logistic regression is used to describe data and explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level

independent variables. This method is appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Also, used to obtain odds ratio in the presence of more than one explanatory variable. In this study, logistic regression analysis is used to identify the factors that influenced a candidate in getting placed.

```
PROC LOGISTIC DATA = SAS-data-set DESCENDING;
    CLASS variable(s);
    MODEL variable(s);
RUN;
QUIT;
```

The logistic function:

$logit = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m$ where p indicates the probability of an event, β_0 is made up of those individuals that presenting the reference level of each and every variable $X_1 \dots m$. While β_i are the regression coefficients associated with the reference group, the X_i is explanatory variables.

2.4.3 Mann-Whitney U Test

The Mann-Whitney U test, sometimes called the Mann Whitney Wilcoxon test is the non-parametric test which the data set is not normally distributed also equivalent of the two-sample t-test. The t-test to make an assumption about the distribution of a population, and the Mann-Whitney U test makes no such assumption. The purpose of this test is used to test whether two samples are likely to derive from the same population.

```
PROC NPAR1WAY DATA = SAS-data-set WILCOXON;
    CLASS variable(s);
    VAR variable(s);
RUN;
```

2.4.4 Kruskal-Wallis Test

Kruskal-Wallis test is non-parametric test which the data set is not normally distributed. The purpose of this test is used to determine if there are statistically

significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable. Kruskal-Wallis test is another alternative to the one-way ANOVA. Which indicates when the p-value less than or equal to the significance level (0.05), we reject the null hypothesis and conclude that not all of population means are equal.

```
PROC NPAR1WAY DATA = SAS-data-set WILCOXON;  
  CLASS variable(s);  
  VAR variable(s);  
RUN;
```

CHAPTER 3

RESULT AND ANALYSIS

3.1 Introduction

This section discussed the result and analysis of this study. There are a few things that need to do the data before proceed to come out with the results which is to check the normality distribution. In this study are using several procedures to manage the data set of academic and employability. There are 3 method that use to achieve the objective of this study which is logistic regression, Mann-Whitney and Kruskal-Wallis. The data was analyzed by using the SAS program which is SAS Studio. The SAS code and some of the outputs were attached in the appendices.

3.2 DESCRIPTIVE STATISTICS

3.2.1 Summary Statistics

In summary statistics, the MEAN Procedure (refer **Appendix 5**) was used to display the population (N), mean, standard deviation, minimum and maximum value.

Table 3.1: The Summary Statistics of Means Procedure

The MEANS Procedure						
Variable	Label	N	Mean	Std Dev	Minimum	Maximum
sl_no	sl_no	215	108.0000000	62.2093241	1.0000000	215.0000000
ssc_p	ssc_p	215	67.3033953	10.8272054	40.8900000	89.4000000
hsc_p	hsc_p	215	66.3331628	10.8975092	37.0000000	97.7000000
degree_p	degree_p	215	66.3701860	7.3587433	50.0000000	91.0000000
etest_p	etest_p	215	72.1005581	13.2759564	50.0000000	98.0000000
mba_p	mba_p	215	62.2781860	5.8333846	51.2100000	77.8900000
salary	salary	215	198702.33	154780.93	0	940000.00

The table 3.1 shows the mean value for salary offered by corporate to candidates is the highest mean which is 198702.33 while number of Post Graduation (MBA) percentage is lowest mean which 62.2782. The student of degree percentage and Employability test percentage shows that, the number of minimum values is same which is 50 while the maximum value of these 2 variables is 91 and 98 respectively.

3.2.2 One-way Frequency Report

In a one-way frequency report, the FREQ procedure (refer **Appendix 6**) was used to display the frequency of Post Graduation (MBA) on Specialization. The TABLE statement was used to limit the variables included in the frequency counts is given below:

Table 3.2: The Frequency of Post Graduation (MBA)- Specialization

The FREQ Procedure				
specialisation				
specialisation	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Mkt&Fin	120	55.81	120	55.81
Mkt&HR	95	44.19	215	100.00

Table 3.2 shows that the frequency of Post Graduation on specialization between Marketing & Finance and Marketing & Human Resources. Since this study want to determine the employability test percentage differ by specialization in higher secondary education, so this study need to know the frequency on specialization between Marketing & Finance and Marketing & Human Resources. Based on the result, the frequency of Marketing & Finance is the highest value which is 120, while the values of Marketing & Human Resources are 95 with 44.19 percent.

3.2.3 The Bar Chart of Student Placement by Status

The bar chart was used to graphically display the frequency of the status of placement. In this study, the data shows that students who getting placed and not placed in a campus. **Appendix 7** shown the SAS code for the result below.

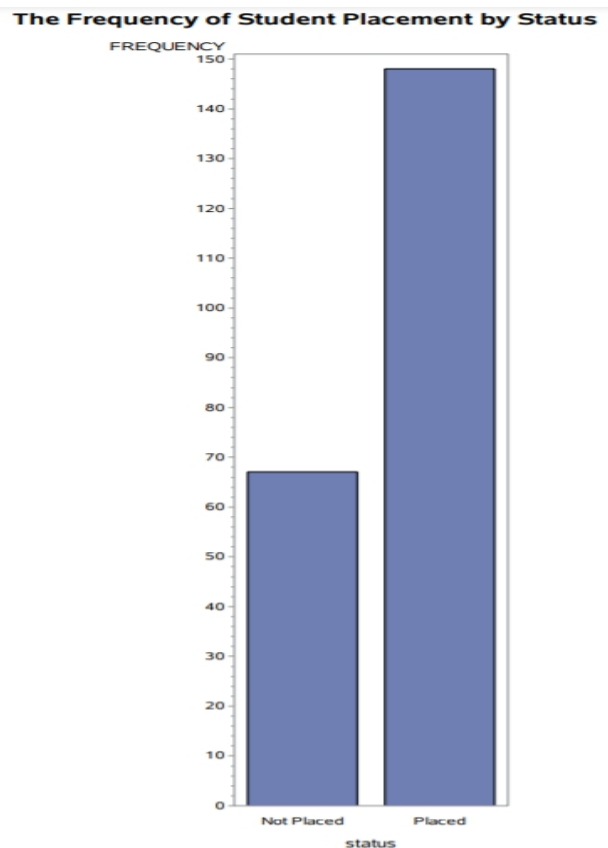


Figure 3.1: Frequency of Student Placement by Status

Figure 3.1 shows the frequency of status of students' placement. From this bar chart, the results are about the placement which happens at the end of post-graduation. Most of the student are getting placed on a campus, as a shown in the table above, the bar chart of placed is the highest which is 148 students on XYZ campus. While, the other student which is 66 students are not placed in the campus. Based on the result placement data of student, most of jobs offers more to students who get a place other than not placed in a campus. However, the salaries that offering is higher to students who get a place.

3.2.4 Pie Chart of Percentage of Work Experience

The Pie Chart was used to graphically display the percentage of students by work experience. In this study, based on second objective want to determine the difference between those with work experience or not. So, we want to know the percentage

either the student have the work experience or not. The result of SAS Code (refer **Appendix 8**) as shown below.

The Percentage of Work Experience

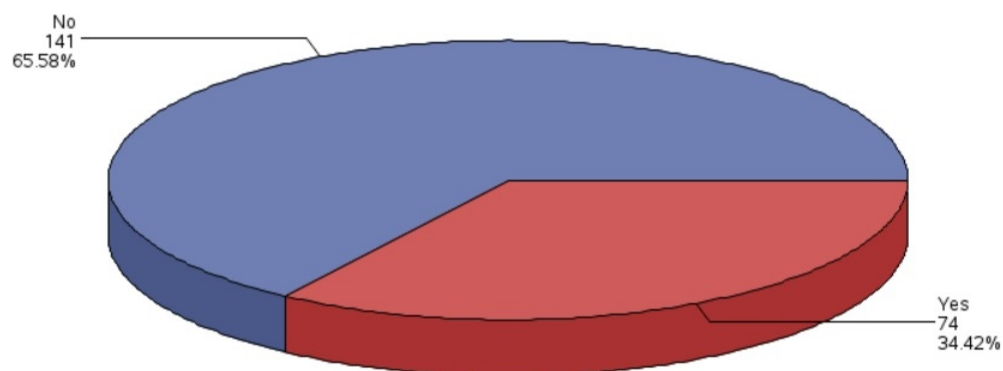


Figure 3.2: The Percentage of Work Experience

Figure 3.2 indicates that the most of student in XYZ campus does not have work experience which is 65.58% even though some of that is students that placed in a campus. Meanwhile, 34.42% is students who have work experience. This is shown that a few students, there are work after they complete their studied. But the job that offering to students are based on student experience in work. This is how, the salary and work experience are important for student to get a better job.

3.3 Inferential Statistics

3.3.1 Normality Test

Refer **Appendix 9** for the SAS code.

a. Salary

Table 3.3: Normality Test for Salary with No Experience

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	.	Pr < W	.
Kolmogorov-Smirnov	D	.	Pr > D	.
Cramer-von Mises	W-Sq	.	Pr > W-Sq	.
Anderson-Darling	A-Sq	.	Pr > A-Sq	.

Table 3.4: Normality Test for Salary with Experience

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.675876	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.262503	Pr > D	<0.0100
Cramer-von Mises	W-Sq	2.06753	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	11.29513	Pr > A-Sq	<0.0050

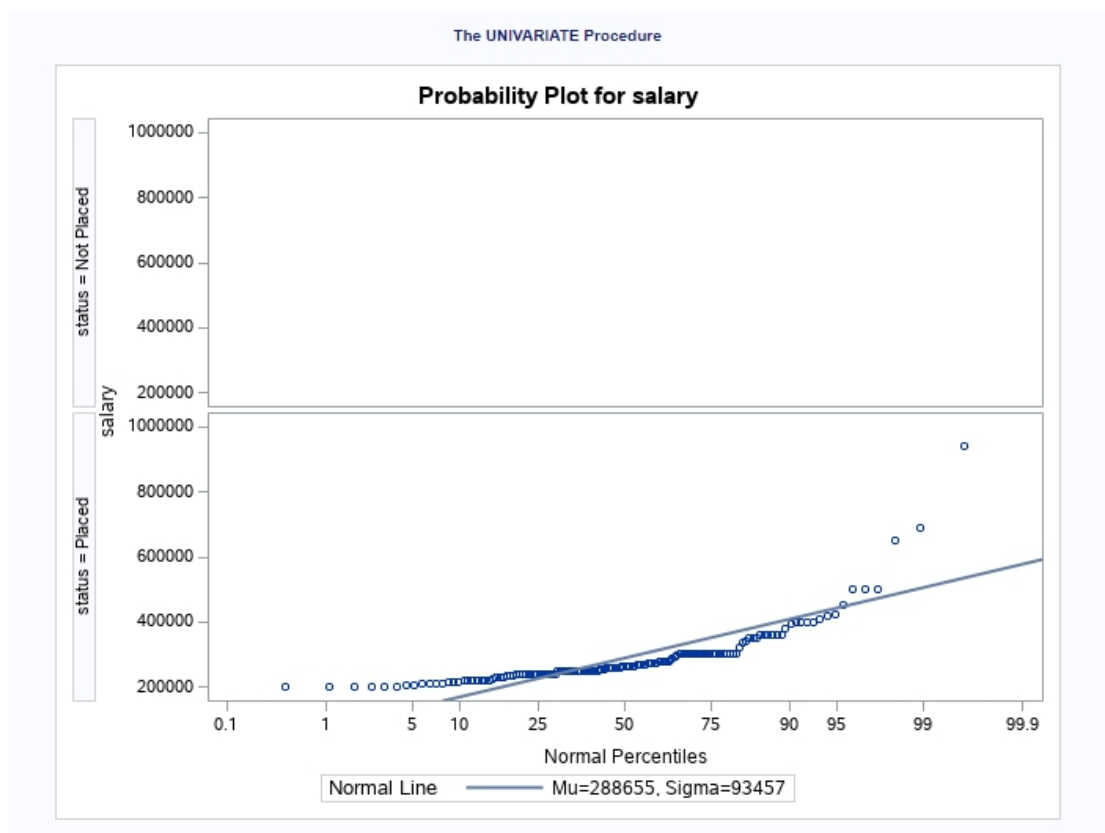


Figure 3.3: Normal P-P Plot of Salary with No Experience and with Experience

Hypothesis Statement:

H_0 = The salary with no experience and with experience is normally distributed.

H_1 = The salary with no experience and with experience is not normally distributed.

Test Statistics:

p-value = 0 and 0.0100

Critical Value:

$\alpha = 0.05$

Decision:

Reject H_0 since p-value = 0 and $0.0100 < \alpha = 0.05$

Conclusion:

The salary with no experience and with experience is not normally distributed.

Interpretation: Table 3.3 and table 3.4 show the probability of the Kolmogorov-Smirnov are 0 and 0.0100 less than alpha value 0.05. Figure 3.3 also shows that the distribution are not normal in the probability plot. Hence, there are enough evidence to conclude that the salary with no experience and with experience are not normally distributed.

b. Employability Test Percentage

Table 3.5: Normality Test for Employability Test Percentage in Arts

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.919086	Pr < W	0.3111
Kolmogorov-Smirnov	D	0.211516	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.0678	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.398535	Pr > A-Sq	>0.2500

Table 3.6: Normality Test for Employability Test Percentage in Commerce

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.943918	Pr < W	0.0001
Kolmogorov-Smirnov	D	0.101265	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.267602	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1.81516	Pr > A-Sq	<0.0050

Table 3.7: Normality Test for Employability Test Percentage in Science

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.942358	Pr < W	0.0005
Kolmogorov-Smirnov	D	0.094127	Pr > D	0.0457
Cramer-von Mises	W-Sq	0.199	Pr > W-Sq	0.0052
Anderson-Darling	A-Sq	1.399112	Pr > A-Sq	<0.0050

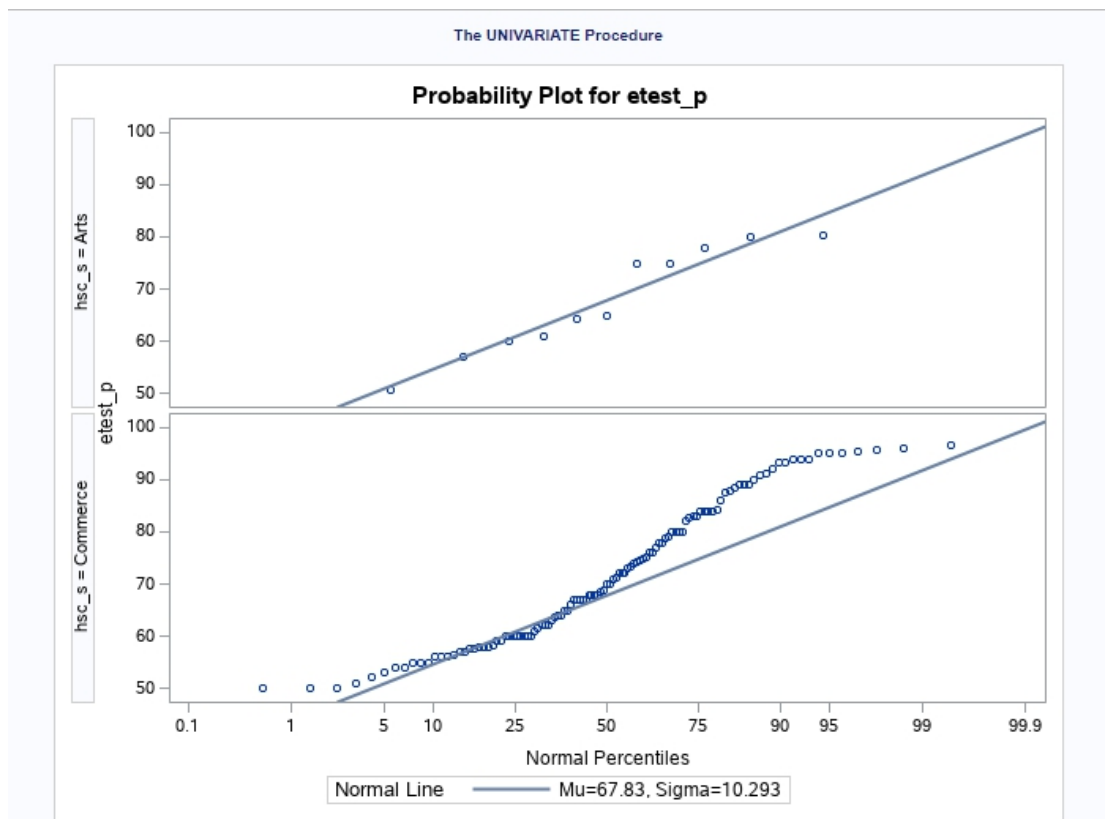


Figure 3.4 : Normal P-P Plot of Employability Test Percentage in Arts and Commerce

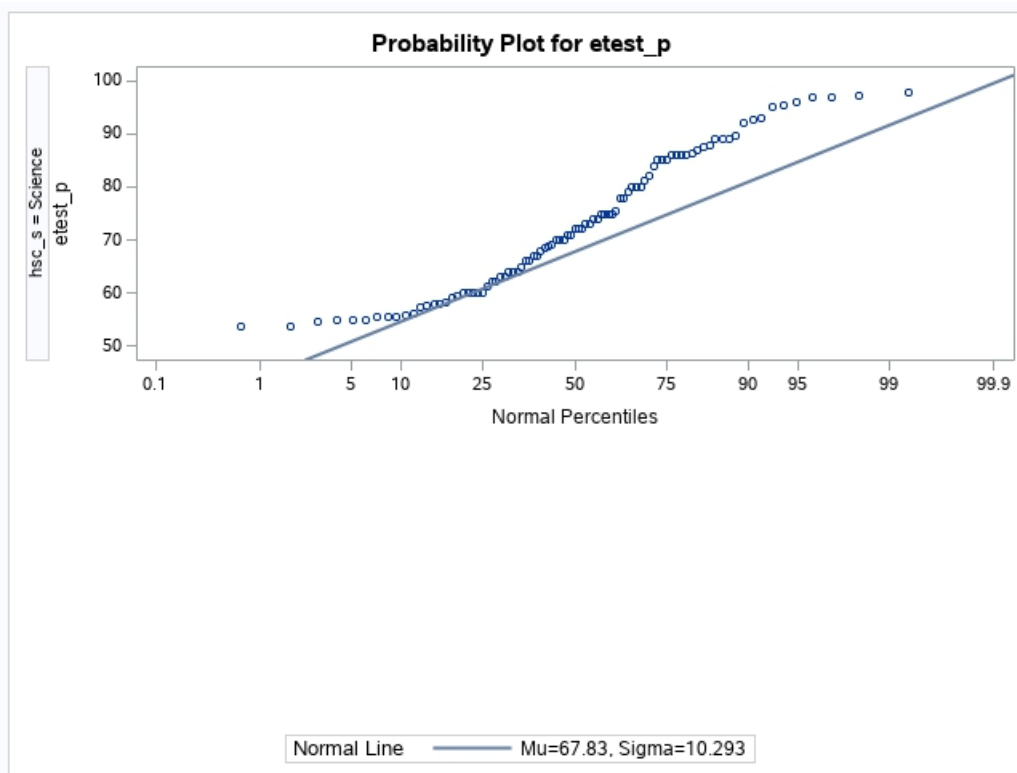


Figure 3.5: Normal P-P Plot of Employability Test Percentage in Science

Hypothesis Statement:

H_0 = The employability test percentage in Arts, Commerce and Science are normally distributed.

H_1 = The employability test percentage in Arts, Commerce and Science are not normally distributed.

Test Statistics:

p-value = 0.1500, 0.0100 and 0.0457

Critical Value:

$\alpha = 0.05$

Decision:

Reject H_0 since only one p-value = 0.1500 > $\alpha = 0.05$ and there are two p-value = 0.0100 and 0.0457 < $\alpha = 0.05$

Conclusion:

The employability test percentage in Arts, Commerce and Science are not normally distributed.

Interpretation: Table 3.5 shows Kolmogorov-Smirnov value is 0.1500 where more than 0.05 while based on table 3.6 and table 3.7, it show that the probability of the Kolmogorov-Smirnov are 0.0100 and 0.0456, less than alpha value 0.05. Hence, there are significance to conclude that the employability test percentage in Arts, Commerce and Science are not normally distributed. Figure 3.4 shows that only employability test percentage in Arts is normal while in Commerce shows it is not normally distributed. Figure 3.5 also shows that the probability plot are not normal in Science.

3.3.2 Logistic Regression Analysis

Refer **Appendix 10** for the SAS code.

Table 3.8: Analysis of Maximum Likelihood Estimates for Logistic Regression Analysis

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-19.0372	5.0436	14.2468	0.0002
gender	1	1	0.6370	0.3382	3.5475	0.0596
ssc_p		1	0.2315	0.0470	24.2463	<.0001
ssc_b	1	1	-0.0836	0.3537	0.0558	0.8132
hsc_p		1	0.1064	0.0376	8.0079	0.0047
hsc_b	1	1	-0.2009	0.3597	0.3119	0.5765
hsc_s	1	1	-0.5088	0.4895	1.0802	0.2986
hsc_s	2	1	-0.00561	0.6059	0.0001	0.9926
degree_p		1	0.1931	0.0550	12.3470	0.0004
degree_t	1	1	-0.8392	0.3925	4.5729	0.0325
workex	1	1	-1.0118	0.3504	8.3397	0.0039
etest_p		1	-0.0174	0.0221	0.6194	0.4313
specialisation	1	1	-0.1467	0.2781	0.2783	0.5978
mba_p		1	-0.2106	0.0569	13.6796	0.0002

Estimated regression equation:

$$\text{logit}(p) = -19.0372 + 0.6370\text{Male} + 0.2315\text{ssc_p} - 0.0836\text{Central} + 0.1064\text{hsc_p} - 0.2009\text{Central} - 0.5088\text{Commerce} - 0.00561\text{Science} +$$

0.1931degree_p - 0.8392Sci&Tech - 1.0118No - 0.0174etest_p - 0.1467Mkt&HR
- 0.2106mba_p

Interpretation: Table 3.8 shows there are twelve independent variables which are gender, secondary education percentage, board of education, higher secondary education percentage, board of education, specialization in higher secondary education, degree percentage, degree type of under graduation, work experience, employability test percentage, post graduation and post graduation percentage who are contributed the value to the status of placement in XYZ Campus. However, there are only six independent variables who are significant for status of placement in XYZ Campus since the p-value are less than 0.05. Therefore, only secondary education percentage, higher secondary education percentage, degree percentage, degree type of under graduation in Communication and Management, no work experience and post graduation percentage that influenced a candidate in getting placed in XYZ Campus.

3.3.3 Mann-Whitney U Test

Refer **Appendix 11** for the SAS code.

Table 3.9: The Median of Salary that being offered among Candidates whether has a Work Experience or not

Variable	Median		Chi-Square	P-value
	No Experience	With Experience		
Salary	262000.00	267500.00	15.5310	0.0001

Interpretation: Based on table 3.9, there are 215 candidates were observed the median of salary amount that being offered by corporate from work experience. The assumption of normality was checked and found that the data violated the assumption. Thus Mann-Whitney analysis was employed for the second objective to test the difference between median salary that being offered by corporate to candidates between candidates with no work experience and with work experience. The results show that there is a significant difference on median salary that being offered among the candidates between the candidate with no work experience and with work experience since p-value was 0.0001 because of the sample size is large

which is more than 100 and chi-square is 15.5310. Furthermore, it was found that the mean salary that being offered to the candidates with no work experience is 262000.00 was slightly lower compared to salary that being offered to the candidates with work experience, 267500.00. Hence, we can say that there is differ in salary between candidates with no work experience and work experience.

3.3.4 Kruskal-Wallis Test

Refer **Appendix 13** for the SAS code.

Table 3.10: The Mean Score of Employability Test Percentage by Specialization in Higher Secondary Education

Variable	Mean			Chi-Square	P-value
	Commerce	Science	Arts		
Employability	106.469027	111.846154	91.909091	1.1529	0.5619

Interpretation: Table 3.10 shows the chi-square statistic for the mean score of employability test percentage by specialization in higher secondary education. Kruskal-wallis test is being used for performing the third objective since the assumption of normality is violated. Other than that, it is used to test whether mean employability test percentage differ by specialization in higher secondary education whether in Commerce, Science and Arts. The mean value for specialization in Science is the highest which is 111.846154 compared to Commerce, 106.469027 and Arts, 91.909091 while chi-square value is 1.1529. However, the p-value is 0.5619 where it is greater than alpha value, 0.05. Therefore, there is no enough evidence to reject the null hypothesis and conclude that the differences between the means employability test percentage are not statistically significant. Thus, we can say that there is no differ in employability test percentage with specialization in higher secondary education whether in Commerce, Science or Arts.

CHAPTER 4

CONCLUSION AND DISCUSSION

In this project, the data was analyzed by using SAS Studio. There are several figures and charts that determined the value such as frequency and percentage. For the descriptive statistics, there are vertical bar chart to represent student placements in XYZ Campus by status. By that, 148 students who are getting placed while the other student which is 66 students are not getting placed in a campus. Next, pie chart also had been used to find the percentage of work experience of the student that getting placed in a campus. 65.58% of that is students that getting placed in XYZ Campus did not have work experience and 34.42% is students who have work experience.

To achieve the first objective, logistic regression analysis was used to identify factors that influenced a candidate in getting placed. There are only six independent variables out of twelve independent variables are being tested who are contributed the value to the status of placement in XYZ Campus since all the p-value are less than 0.05. Therefore, secondary education percentage, higher secondary education percentage, degree percentage, degree type of under graduation in Communication and Management, no work experience and post graduation percentage that influenced a candidate in getting placed in XYZ Campus.

Based on the normality test, the salary with no experience and with experience are not normally distributed since p-value are 0 and 0.0100 where it is less than 0.05. Hence, Mann-Whitney test is being used in this study to perform the second objective. By using Mann-Whitney test, it can analyse whether the work experience can be differed in salary amount that being offered by corporate. The results show that p-value is 0.0001 and can be conclude that there is differ in salary that being offered by corporate between candidates with no work experience and work experience.

Lastly, Kruskal-Wallis test is being used to achieve the third objective since the employability test percentage in Arts, Commerce and Science are also not normally distributed because only one out of three p-value is more than alpha value, 0.05. Thus, it makes to reject the null hypothesis. The statistics table by Kruskal-Wallis test shows p-value is 0.5619 where it is greater than 0.05. There is enough evidence to conclude that there is no differ in employability test percentage with specialization in higher secondary education whether in Commerce, Science or Arts.

CHAPTER 5

REFERENCES

- Eom, J. (2021). *ScholarlyCommons Easily Importing Excel Files To SAS*.
- Nauta, A., van Vianen, A., van der Heijden, B., van Dam, K., & Willemsen, M. (2009). Understanding the factors that promote employability orientation: The impact of employability culture, career satisfaction, and role breadth self-efficacy. *Journal of Occupational and Organizational Psychology*, 82(2), 233–251. <https://doi.org/10.1348/096317908X320147>
- Pa. (1967). 済無No Title No Title No Title. *Angewandte Chemie International Edition*, 6(11), 951–952., 25–29.
- Smith, S., Taylor-Smith, E., Smith, C. F., & Webster, G. (2018). The impact of work placement on graduate employment in computing: Outcomes from a UK-based study. *International Journal of Work-Integrated Learning*, 19(4), 359–369.

CHAPTER 6

APPENDICES

Appendix 1: Assigning a Libref

SAS Data Library

```
LIBNAME places '/home/u50035588/mydata/project';
```

Appendix 2: Reading Excel Spreadsheets

PROC IMPORT

```
PROC IMPORT OUT = places.placements
  DATAFILE = '/home/u50035588/mydata/project/New_Placement_Data_Full_Class.xlsx'
  DBMS = xlsx;
RUN;

PROC IMPORT OUT = places.placement
  DATAFILE = '/home/u50035588/mydata/project/New_Placement_Data_Full_Class.xlsx'
  DBMS = xlsx;
RUN;
```

Appendix 3: The Descriptor Portion

```
PROC CONTENTS DATA = places.placements;
RUN;
```

The CONTENTS Procedure

Data Set Name	PLACES.PLACEMENTS	Observations	215
Member Type	DATA	Variables	15
Engine	V9	Indexes	0
Created	06/19/2021 20:13:27	Observation Length	120
Last Modified	06/19/2021 20:13:27	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1168
Obs in First Data Page	215
Number of Data Set Repairs	0
Filename	/home/u50035588/mydata/project/placements.sas7bdat
Release Created	9.0401M8
Host Created	Linux
Inode Number	133572513
Access Permission	rw-r--r--
Owner Name	u50035588
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
8	degree_p	Num	8	BEST.		degree_p
9	degree_t	Char	9	\$9.	\$9.	degree_t
11	etest_p	Num	8	BEST.		etest_p
2	gender	Char	1	\$1.	\$1.	gender
6	hsc_b	Char	7	\$7.	\$7.	hsc_b
5	hsc_p	Num	8	BEST.		hsc_p
7	hsc_s	Char	8	\$8.	\$8.	hsc_s
13	mba_p	Num	8	BEST.		mba_p
15	salary	Num	8	BEST.		salary
1	sl_no	Num	8	BEST.		sl_no
12	specialisation	Char	7	\$7.	\$7.	specialisation
4	ssc_b	Char	7	\$7.	\$7.	ssc_b
3	ssc_p	Num	8	BEST.		ssc_p
14	status	Char	10	\$10.	\$10.	status
10	workex	Char	3	\$3.	\$3.	workex

```
PROC CONTENTS DATA = places.placement;
RUN;
```

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
16	P	Char	1	\$1.	\$1.	P
17	Q	Char	61	\$61.	\$61.	Q
8	degree_p	Num	8	BEST.		degree_p
9	degree_t	Num	8	BEST.		degree_t
11	etest_p	Num	8	BEST.		etest_p
2	gender	Num	8	BEST.		gender
6	hsc_b	Num	8	BEST.		hsc_b
5	hsc_p	Num	8	BEST.		hsc_p
7	hsc_s	Num	8	BEST.		hsc_s
13	mba_p	Num	8	BEST.		mba_p
15	salary	Num	8	BEST.		salary
1	sl_no	Num	8	BEST.		sl_no
12	specialisation	Num	8	BEST.		specialisation
4	ssc_b	Num	8	BEST.		ssc_b
3	ssc_p	Num	8	BEST.		ssc_p
14	status	Num	8	BEST.		status
10	workex	Num	8	BEST.		workex

Appendix 4:

PROC PRINT

```
PROC PRINT DATA = places.placements;
RUN;
```

Obs	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
1	1	M	67	Others	91	Others	Commerce	58	Sci&Tech	No	55	Mkt&HR	58.8	Placed	270000
2	2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	200000
3	3	M	65	Central	68	Central	Arts	64	Comm&Mgmt	No	75	Mkt&Fin	57.8	Placed	250000
4	4	M	56	Central	52	Central	Science	52	Sci&Tech	No	66	Mkt&HR	59.43	Not Placed	.
5	5	M	85.8	Central	73.6	Central	Commerce	73.3	Comm&Mgmt	No	96.8	Mkt&Fin	55.5	Placed	425000
6	6	M	55	Others	49.8	Others	Science	67.25	Sci&Tech	Yes	55	Mkt&Fin	51.58	Not Placed	.
7	7	F	46	Others	49.2	Others	Commerce	79	Comm&Mgmt	No	74.28	Mkt&Fin	53.29	Not Placed	.
8	8	M	82	Central	64	Central	Science	66	Sci&Tech	Yes	67	Mkt&Fin	62.14	Placed	252000
9	9	M	73	Central	79	Central	Commerce	72	Comm&Mgmt	No	91.34	Mkt&Fin	61.29	Placed	231000
10	10	M	58	Central	70	Central	Commerce	61	Comm&Mgmt	No	54	Mkt&Fin	52.21	Not Placed	.
11	11	M	58	Central	61	Central	Commerce	60	Comm&Mgmt	Yes	62	Mkt&HR	60.85	Placed	280000
12	12	M	69.6	Central	68.4	Central	Commerce	78.3	Comm&Mgmt	Yes	60	Mkt&Fin	63.7	Placed	250000
13	13	F	47	Central	55	Others	Science	65	Comm&Mgmt	No	62	Mkt&HR	65.04	Not Placed	.
14	14	F	77	Central	87	Central	Commerce	59	Comm&Mgmt	No	68	Mkt&Fin	68.63	Placed	218000
15	15	M	62	Central	47	Central	Commerce	50	Comm&Mgmt	No	76	Mkt&HR	54.96	Not Placed	.
16	16	F	65	Central	75	Central	Commerce	69	Comm&Mgmt	Yes	72	Mkt&Fin	64.66	Placed	200000
17	17	M	63	Central	66.2	Central	Commerce	65.6	Comm&Mgmt	Yes	60	Mkt&Fin	62.54	Placed	300000
18	18	F	55	Central	67	Central	Commerce	64	Comm&Mgmt	No	60	Mkt&Fin	67.28	Not Placed	.
19	19	F	63	Central	66	Central	Commerce	64	Comm&Mgmt	No	68	Mkt&HR	64.08	Not Placed	.
20	20	M	60	Others	67	Others	Arts	70	Comm&Mgmt	Yes	50.48	Mkt&Fin	77.89	Placed	236000
21	21	M	62	Others	65	Others	Commerce	66	Comm&Mgmt	No	50	Mkt&HR	56.7	Placed	265000
22	22	F	79	Others	76	Others	Commerce	85	Comm&Mgmt	No	95	Mkt&Fin	69.05	Placed	393000
23	23	F	69.8	Others	60.8	Others	Science	72.23	Sci&Tech	No	55.53	Mkt&HR	68.81	Placed	360000
24	24	F	77.4	Others	60	Others	Science	64.74	Sci&Tech	Yes	92	Mkt&Fin	63.62	Placed	300000
25	25	M	76.5	Others	97.7	Others	Science	78.86	Sci&Tech	No	97.4	Mkt&Fin	74.01	Placed	360000
26	26	F	52.58	Others	54.6	Central	Commerce	50.2	Comm&Mgmt	Yes	76	Mkt&Fin	65.33	Not Placed	.
27	27	M	71	Others	79	Others	Commerce	66	Comm&Mgmt	Yes	94	Mkt&Fin	57.55	Placed	240000
28	28	M	63	Others	67	Others	Commerce	66	Comm&Mgmt	No	68	Mkt&HR	57.69	Placed	265000
29	29	M	76.76	Others	76.5	Others	Commerce	67.5	Comm&Mgmt	Yes	73.35	Mkt&Fin	64.15	Placed	350000
30	30	M	62	Central	67	Central	Commerce	58	Comm&Mgmt	No	77	Mkt&Fin	51.29	Not Placed	.
31	31	F	64	Central	73.5	Central	Commerce	73	Comm&Mgmt	No	52	Mkt&HR	56.7	Placed	250000
32	32	F	67	Central	53	Central	Science	65	Sci&Tech	No	64	Mkt&HR	58.32	Not Placed	.
33	33	F	61	Central	81	Central	Commerce	66.4	Comm&Mgmt	No	50.89	Mkt&HR	62.21	Placed	278000

```
PROC PRINT DATA = places.placement;
RUN;
```

Obs	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
1	1	1	67	2	91	2	1	58	1	1	55	1	58.8	2	270000
2	2	1	79.33	1	78.33	2	2	77.48	1	2	86.5	2	66.28	2	200000
3	3	1	65	1	68	1	3	64	2	1	75	2	57.8	2	250000
4	4	1	56	1	52	1	2	52	1	1	66	1	59.43	1	0
5	5	1	85.8	1	73.6	1	1	73.3	2	1	96.8	2	55.5	2	425000
6	6	1	55	2	49.8	2	2	67.25	1	2	55	2	51.58	1	0
7	7	2	46	2	49.2	2	1	79	2	1	74.28	2	53.29	1	0
8	8	1	82	1	64	1	2	66	1	2	67	2	62.14	2	252000
9	9	1	73	1	79	1	1	72	2	1	91.34	2	61.29	2	231000
10	10	1	58	1	70	1	1	61	2	1	54	2	52.21	1	0
11	11	1	58	1	61	1	1	60	2	2	62	1	60.85	2	260000
12	12	1	69.6	1	68.4	1	1	78.3	2	2	60	2	63.7	2	250000
13	13	2	47	1	55	2	2	65	2	1	62	1	65.04	1	0
14	14	2	77	1	87	1	1	59	2	1	68	2	68.63	2	218000
15	15	1	62	1	47	1	1	50	2	1	76	1	54.96	1	0
16	16	2	65	1	75	1	1	69	2	2	72	2	64.66	2	200000
17	17	1	63	1	66.2	1	1	65.6	2	2	60	2	62.54	2	300000
18	18	2	55	1	67	1	1	64	2	1	60	2	67.28	1	0
19	19	2	63	1	66	1	1	64	2	1	68	1	64.08	1	0
20	20	1	60	2	67	2	3	70	2	2	50.48	2	77.89	2	236000
21	21	1	62	2	65	2	1	66	2	1	50	1	56.7	2	265000
22	22	2	79	2	76	2	1	85	2	1	95	2	69.06	2	393000
23	23	2	69.8	2	60.8	2	2	72.23	1	1	55.53	1	68.81	2	360000
24	24	2	77.4	2	60	2	2	64.74	1	2	92	2	63.62	2	300000
25	25	1	76.5	2	97.7	2	2	78.86	1	1	97.4	2	74.01	2	360000
26	26	2	52.58	2	54.6	1	1	50.2	2	2	76	2	65.33	1	0
27	27	1	71	2	79	2	1	66	2	2	94	2	57.55	2	240000
28	28	1	63	2	67	2	1	66	2	1	68	1	57.69	2	265000
29	29	1	76.76	2	76.5	2	1	67.5	2	2	73.35	2	64.15	2	350000
30	30	1	62	1	67	1	1	58	2	1	77	2	51.29	1	0
31	31	2	64	1	73.5	1	1	73	2	1	52	1	56.7	2	250000
32	32	2	67	1	53	1	2	65	1	1	64	1	58.32	1	0
33	33	2	61	1	81	1	1	66.4	2	1	50.89	1	62.21	2	278000

Appendix 5:

PROC MEANS

```
PROC MEANS DATA = PLACES.PLACEMENTS;
RUN;
```

Appendix 6:

PROC FREQ

```
PROC FREQ DATA = PLACES.PLACEMENTS;
TABLE specialisation;
RUN;
```

Appendix 7:

```
PROC GCHART DATA = PLACES.PLACEMENTS;  
  VBAR STATUS;  
  TITLE 'The Frequency of Student Placement by Status';  
RUN;
```

Appendix 8:

```
TITLE 'The Percentage of Work Experience';  
PROC GCHART data = PLACES.PLACEMENTS;  
  pie3d workex/percent=arrow noheadingtype;  
RUN;
```

Appendix 9:

Normality Test

```
PROC UNIVARIATE DATA = places.placements NORMAL;  
  VAR salary;  
  CLASS status;  
  PROBPLOT/NORMAL (MU=est SIGMA=est);  
RUN;
```

```
PROC UNIVARIATE DATA = places.placements NORMAL;  
  VAR etest_p;  
  CLASS hsc_s;  
  PROBPLOT/NORMAL (MU=est SIGMA=est);  
RUN;
```

Appendix 10:

Logistic Regression

```
PROC LOGISTIC DATA = places.placement DESCENDING;  
  CLASS gender ssc_b hsc_b hsc_s degree_t workex specialisation;  
  MODEL status(event='2') = gender ssc_p ssc_b hsc_p hsc_b hsc_s degree_p  
    degree_t workex etest_p specialisation mba_p;  
RUN;  
QUIT;
```

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	98.1	Somers' D	0.923
Percent Discordant	3.9	Gamma	0.923
Percent Tied	0.0	Tau-a	0.398
Pairs	9916	c	0.961

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	166.5359	13	<.0001
Score	120.2419	13	<.0001
Wald	45.2923	13	<.0001

Appendix 11:

Mann-Whitney U Test

```
PROC NPAR1WAY DATA = places.placements WILCOXON;
  CLASS workex;
  VAR salary;
RUN;
```

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable salary Classified by Variable workex					
workex	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
No	84	5982.0	6258.0	257.513864	71.214286
Yes	64	5044.0	4768.0	257.513864	78.812500
Average scores were used for ties.					

Wilcoxon Two-Sample Test					
Statistic	Z	Pr > Z	Pr > Z	t Approximation	
				Pr > Z	Pr > Z
9672.000	3.9398	<.0001	<.0001	<.0001	0.0001
Z includes a continuity correction of 0.5.					

Appendix 12:

```
PROC MEANS DATA = places.placements MEDIAN;
  VAR salary;
  CLASS workex;
RUN;
```

Appendix 13:

Kruskal-Wallis Test

```
PROC NPAR1WAY DATA = places.placements WILCOXON;  
    CLASS hsc_s;  
    VAR etest_p;  
RUN;
```

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable etest_p Classified by Variable hsc_s					
hsc_s	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Commerce	113	12031.0	12204.0	455.354125	108.469027
Science	91	10178.0	9828.0	450.548458	111.846154
Arts	11	1011.0	1188.0	200.919178	91.909091
Average scores were used for ties.					

Kruskal-Wallis Test

Chi-Square	DF	Pr > ChiSq
1.1529	2	0.5619