UNIVERSITY OF VAASA

Faculty of Technology

Report on Analysis of Empirical Work using SAS EG

STAT2110 STATISTICAL DATA PROCESSING SAS EG

Name: Onyebuchi Dalbert Zimuzochukwu
Student Number: Z109554
Email Address: Z109554@student.uwasa.fi

| Course Title : STAT2110|

Date: 11.12.2016

INTRODUCTION

The aim of this project is to study the statistics of the presented data. Here a

For this project, I have chosen the following variables:

a. Age (Quantitative)
b. Education (Categorical: Ordinal)
c. Employment (Quantitative)
d. Income (Quantitative)
e. Debt to Income Ratio (Quantitative : Ratio level)
f. Credit Card Debit (Quantitative)
g. Default on Loans. (Categorical: Nominal)
h. Residence (Categorical: Nominal)
i. Other debts (Quantitative)

My reason for chosen all these is because I want to evaluate the empirical data holistically (As a Whole) to get a complete picture about the population.

In this work, the following hypotheses will be examined:

Hypothesis: Younger people have more defaults on loans due to lifestyle preferences.

Hypothesis: College students have more loan debts (default on loans) due to college expenses.

Hypothesis: More Income leads to less debt, as income earners have the means to pay back.
Hypothesis: There is a strong relationship between Debit to income ratio and Credit Card Debt and a positive relationship exist between the two variables.

Hypothesis: More income leads to more Credit Card Debt as high-income earners take out more credit card loan from the banks.

Hypothesis: More income leads to less default on loans as income earners have the means to pay back.

Hypothesis: More income leads to less other debts.

2.0. STATISTICAL DESCRIPTION OF DATA

a. AGE (QUANTITATIVE DATA)

From SECTION 1 A (See Appendix 1), the age population appears to be a young population as the median is 34 years old.

In addition, people of 29 years appears to be the highest occurrence in the population (MODE).

From SECTION 1 B (See Appendix 1), here we notice that the distribution of age is multimodal.

The distribution is positively skewed and platykurtic. Skewness = 0.322, Kurtosis = -0.73

By the way, the number of cases is about 498, so the standard error for skewness is about 0.109 and the standard error for kurtosis is about 0.2195.)

Because of the skewness, I compared interquartile range (IQR) and range to figure out the situation of variability. The IQR is quite small (13) compared with range (36), so the dispersion of the variable values is quite weak.

From SECTION 1 C (See Appendix 1)

A greater number of the population lie between 29 and 42 years old. There are no outliers.

b. EDUCATION (CATEGORICAL DATA: Ordinal)

From Section 2 B (SEE APPENDIX 2) and Section 2 C ((SEE APPENDIX 2), the percentage of people who completed a high school degree is the highest, followed by those who did not complete a high school degree.
Those who went ahead to complete further higher education degrees appears to be smaller, With those with a Master of PHD degree the lowest.
It can be deduced that people represented by this statistic may not be motivated to pursue higher degrees as their basic high school degree can get them a form of comfortable employment. (As is noticed when compared to the employment distribution).
The distribution is unimodal, and the distribution is left-skewed and is platykurtic.
Skewness = 0.87, Kurtosis = -0.36

By the way, the number of cases is about 498, so the standard error for skewness is about 0.109 and the standard error for kurtosis is about 0.2195.)

c. EMPLOYMENT (Quantitative Data)

From Section 3 A (SEE APPENDIX 3), the Mean = 8.17. Median = 7, Standard Deviation = 6.
Skewness = 0.87, Kurtosis = -0.313

By the way, the number of cases is about 498, so the standard error for skewness is about 0.109 and the standard error for kurtosis is about 0.2195.)

From Section 3 B (SEE APPENDIX 3), the distribution is unimodal, mesokurtic and positively skewed.
Because of the skewness, I compared interquartile range (IQR) and range to figure out the situation of variability. The IQR is quite small (10) compared with range (31), so the dispersion of the variable values is quite weak.
From Section 3 C (SEE APPENDIX 3), we notice that the mean number of years spent with employer is 8 and also a few outliers. This could also indicate that job mobility is high as only few stay longer at a particular work place.

d. INCOME (Quantitative Data)
From Section 4 A (SEE APPENDIX 4), Mean = 48.13, Median = 36 (Which shows the economy is fine when compared to modern day). Standard Deviation: 40.32 (very Large)

From Section 4 B (SEE APPENDIX 4), the distribution is unimodal, leptokurtic and positively skewed.
Skewness = 3.84, Kurtosis = 24.29

By the way, the number of cases is about 498, so the standard error for skewness is about 0.109 and the standard error for kurtosis is about 0.2195.)

Because of the skewness, I compared interquartile range (IQR) and range to figure out the situation of variability. The IQR is quite small (32) compared with range (432), so the dispersion of the variable values is quite weak.

From Section 4 C (SEE APPENDIX 4), we have a few outliers, People with very high incomes.

e. DEBT TO INCOME RATIO (Quantitative : Ratio Level)
From Section 5 A (SEE APPENDIX 5), Mean = 10.35, Median = 8.5, Standard Deviation: 6.89

From Section 5 B (SEE APPENDIX 5), the distribution is unimodal, Leptokurtic and positively skewed.
Skewness = 1.09, Kurtosis = 1.36

By the way, the number of cases is about 498, so the standard error for skewness is about 0.109 and the standard error for kurtosis is about 0.2195.)

Because of the skewness, I compared interquartile range (IQR) and range to figure out the situation of variability. The IQR is quite small (9.5) compared with range (40.7), so the dispersion of the variable values is quite weak.

From Section 5 C (SEE APPENDIX 5), there are outliers, people who are in serious debt.

f. CREDIT CARD DEBT (Quantitative Data)
From Section 6 A (SEE APPENDIX 6), Mean = 1.60, Median = 0.90, Mode = 0.086, Standard Deviation: 2.10.

From Section 6 B (SEE APPENDIX 6), the distribution is unimodal, Leptokurtic and positively skewed.
Skewness = 3.41, Kurtosis 15.8

By the way, the number of cases is about 498, so the standard error for skewness is about 0.109 and the standard error for kurtosis is about 0.2195.)

Because of the skewness, I compared interquartile range (IQR) and range to figure out the situation of variability. The IQR is quite small (16.01) compared with range (1.59), so the dispersion of the variable values is quite weak.

From Section 6 C (SEE APPENDIX 6), we have many outliers, which shows that many people in the sample are in debt.

g. DEFAULT ON LOANS (Categorical Data: Nominal)

From Section 7 A (SEE APPENDIX 7), Mean = 0.29, Median = 0.00, Mode = 0.00, Standard Deviation: 0.45.

From Section 7 B (SEE APPENDIX 7), the distribution is unimodal, platykurtic and positively skewed.

Skewness = 0.922, Kurtosis = -1.15

By the way, the number of cases is about 498, so the standard error for skewness is about 0.109 and the standard error for kurtosis is about 0.2195.)

Because of the skewness, I compared interquartile range (IQR) and range to figure out the situation of variability. The IQR (1) is equal with range (1), so the dispersion of the variable values is quite Strong. Of course, it should be as the variables are nominal.

The frequency of people who defaulted on loans were smaller than those who did not.

About 29% of the population defaulted on loans.

From Section 7 C (SEE APPENDIX 7), there are no outliers.

h. RESIDENCE (Categorical: Nominal)

From Section 8 A (SEE APPENDIX 8), Mean = 1.96, Median = 2.00, Mode = 2.00, Standard Deviation: 0.71.

From Section 8 B (SEE APPENDIX 8), the distribution is unimodal, platykurtic and symmetric.
Skewness = 0.059, Kurtosis = -0.99

By the way, the number of cases is about 498, so the standard error for skewness is about 0.109 and the standard error for kurtosis is about 0.2195.)

Because of the skewness, I compared interquartile range (IQR) and range to figure out the situation of variability. The IQR (2) is equal with range (1), so the dispersion of the variable values is quite Strong. Of course, it should be as the variables are nominal. The frequency of people who live in downtown is more than that of other places.

From Section 8 C (SEE APPENDIX 8), there are no outliers.

i. <u>OTHER DEBTS (Quantitative)</u>

From Section 9 A (SEE APPENDIX 9), Mean = 3.24, Median = 2.10, Mode = 7.82, Standard Deviation: 3.485.

From Section 8 B (SEE APPENDIX 9), the distribution is the distribution is unimodal, Leptokurtic and positively skewed.

Skewness = 2.72, Kurtosis = 10.093

By the way, the number of cases is about 498, so the standard error for skewness is about 0.109 and the standard error for kurtosis is about 0.2195.)

Because of the skewness, I compared interquartile range (IQR) and range to figure out the situation of variability.

The IQR (3.07) is equal with range (26.998), so the dispersion of the variable values is quite weak.

From Section 9 C (SEE APPENDIX 9), there are so many outliers.

3.0. ANALYZES OF STATISTICAL DEPENDENCY

a. Age in Years (Quantitative) Versus Residence (Nominal)

From APPENDIX 10, Likelihood Ratio Chi-square probability = 0.3490, Phi Coefficient = 0.379, Contingency Coefficient = 0.3542 and Cramer's V = 0.2678.
All the measures show that there is a moderate relationship between the variables. The percentage distribution of Age is different for different places of residence.
There is also a negative correlation (-0.024) and the P-Value = 0.593. There is a very weak negative linear relationship between the variables: Age and Residence.
The Correlation is not SIGNIFICANT.

b. Age in Years (Quantitative) Versus Years in Current Employer (Quantitative)

From APPENDIX 11, Likelihood Ratio Chi-square probability = 0.99, Phi Coefficient = 1.73 Contingency Coefficient = 0.87 and Cramer's V = 0.32.

All the measures show that there is a strong relationship between the variables.

H0: The population distribution of the variable is independent.
H1: The population distribution is NOT independent.

We can accept the alternate Hypothesis H1 that in the population, the variables are NOT independent.

There is also a positive correlation (0.5515) and the P-Value < 0.0001. The correlation is significant at 0.01 % significance level.
There seems to be moderate positive linear relationship between the variables: Age and Employment.

c. Age in Years (Quantitative) Versus Education (Ordinal)

From APPENDIX 12, Likelihood Ratio Chi-square probability = 0.0018, Phi Coefficient = 0.59, Contingency Coefficient = 0.5 and Cramer's V = 0.297.

All the measures show that there is a moderate relationship between the variables.

There is also a positive correlation (0.00335) and the P-Value = 0.94. The Correlation is not SIGNIFICANT.

There is a very weak linear relationship between the variables: Age and Education.

d. Age in Years (Quantitative) Versus Income (Quantitative)

From APPENDIX 13, Likelihood Ratio Chi-square probability = 1, Phi Coefficient = 3.19, Contingency Coefficient = 0.95 and Cramer's V = 0.53.

All the measures show that there is a strong relationship between the variables.

Ho: The population distribution of the variable is independent.

H1: The population distribution is NOT independent.

We can accept the alternate Hypothesis H1 that in the population, the variables are NOT independent.

There is also a positive correlation (0.45) and the P-Value < 0.001. The correlation is significant at 0.01 % significance level.

There is a moderate positive linear relationship between the variables: Age and Income.

e. Age in Years (Quantitative) Versus default on Loans (Nominal)

From APPENDIX 14, Likelihood Ratio Chi-square probability = 0.0051, Phi Coefficient = 0.33, Contingency Coefficient = 0.32 and Cramer's V = 0.33.

All the measures show that there is a moderate relationship between the variables.

There is also a negative correlation (-0.16) and the P-Value < 0.0002. The correlation is significant at 0.02 % significance level.

Ho: Younger people do NOT have more defaults on loans.

H1: Younger people have more defaults on loans.

We can accept the alternate Hypothesis H1 that Younger people do have more defaults on loans. There are prevailing factors peculiar with the young: fashion, compulsive Disorder e.g.: shopaholics, little or no stable employment, loan seeking for education, misplaced or misguided priorities, and social factors.

f. Education (Ordinal) Versus Residence (Nominal)

From APPENDIX 15, Likelihood Ratio Chi-square probability = 0.22, Phi Coefficient = 0.15, Contingency Coefficient = 0.15 and Cramer's V = 0.1.

All the measures show that there is weak relationship between the variables.

There is also a negative correlation (-0.06) and the P-Value 0.1825.

g. Education (Ordinal) Versus Default on Loans (Nominal)
H_1 = College students have more loan debts (default on loans)
H_o = College students have less loan debts (default on loans)
From APPENDIX 16, Likelihood Ratio Chi-square probability = 0.17, Phi Coefficient = 0.12, Contingency Coefficient = 0.12 and Cramer's V = 0.12.
All the measures show that there is weak relationship between the variables.
There is also a positive correlation (0.0883) and the P-Value = 0.0489. Since the P value is small, the result is significant. The correlation is significant at 5 % significance level. The null Hypothesis H_o is rejected and the alternative Hypothesis is accepted.

We also noticed that the number of people who defaulted on loans were more among those with a college degree (22 out of 30: 73%).

Moreover, this is expected, as there is currently a rise in loan debts among college students.

The Cocharan-Armitage Trend test is as follows: Statistics (Z) = -1.9702, One-sided Pr < Z = 0.0244, Two-sided Pr < |Z| = 0.0488.

h. Education (Ordinal) Versus Debt to income Ratio (Quantitative : Ratio level)

Ho: The population distribution of the variable is independent.
H1: The population distribution is NOT independent.

From APPENDIX 17, Likelihood Ratio Chi-square probability = 0.98, Phi Coefficient = 1.26, Contingency Coefficient = 0.78 and Cramer's V = 0.63.
All the measures show that there is very strong relationship between the variables.

There is also a positive correlation (0.0217) and the P-Value = 0.6284.

An unusually high p value indicates the data match the null model suspiciously well, suggesting perhaps… the data were fabricated.

We can accept the alternate Hypothesis H1 that in the population, the variables are NOT independent.

i. Debt to Income Ratio (Quantitative : Ratio Level) Vs Income (Quantitative)

Ho: More Income leads to more debt
H1: More Income leads to less debt

From APPENDIX 18, Likelihood Ratio Chi-square probability = 1, Phi Coefficient = 6.43, Contingency Coefficient = 0.99 and Cramer's V = 0.62.

All the measures show that there is very strong relationship between the variables.

There is also a negative correlation (-0.053) and the P-Value = 0.2078.

This means that as Income increases, the debit to income ratio decreases. It therefore means that people with larger incomes have lower debts. So we accept the alternate hypothesis H1.

j. Credit Card Debt (Quantitative) vs Debt to Income Ratio (Quantitative : Ratio Level)
Ho: More Debit to income ratio leads to more Credit Card Debt
H1: More Debit to income ratio leads to less Credit Card Debt

From APPENDIX 19, Likelihood Ratio Chi-square probability = 1, Phi Coefficient = 10.78, Contingency Coefficient = 0.99 and Cramer's V = 0.76.
All the measures show that there is very strong relationship between the variables.
There is also a positive correlation (0.485) and the P-Value = < 0.0001. The correlation is significant at 0.01%.

This means that, as debit to income ratio increases, there is a tendency for credit card debt to rise too. Therefore, we accept the null hypothesis H0.

k. Income (Quantitative) vs Credit Card Debt (Quantitative)

H0: More income leads to less Credit Card Debt
H1: More income leads to more Credit Card Debt

From APPENDIX 20, Likelihood Ratio Chi-square probability = 1, Phi Coefficient = 8.29, Contingency Coefficient = 0.99 and Cramer's V = 0.8.
All the measures show that there is strong relationship between the variables.
There is also a positive correlation (0.586) and the P-Value = < 0.0001. The correlation is significant at 0.01%.
This means that, income increases, there is a tendency for credit card debt to rise too. Therefore, we reject the null hypothesis H0, and accept the alternate hypothesis H1.
People with more money tend to spend more with credit cards, as the bank places a higher limit on the amount of money they can spend on their credit cards given the fact that their higher income is a guarantee they will pay back.

l. Income (Quantitative) vs default on loans (Nominal)
H0: More income leads to less default on loans
H1: More income leads to more default on loans
From APPENDIX 21, Likelihood Ratio Chi-square probability = 0.02, Phi Coefficient = 0.48, Contingency Coefficient = 0.43 and Cramer's V = 0.48.
All the measures show that there is moderate relationship between the variables.
There is also a negative correlation (-0.085) and the P-Value = 0.059. The correlation is significant at 5%.
This means that, income increases, there is a decrease in the default on loans. Therefore, we accept the null hypothesis H0.

m. Income (Quantitative) vs other debts (Quantitative)
H0: More income leads to less other debts
H1: More income leads to more other debts
From APPENDIX 22, Likelihood Ratio Chi-square probability = 1, Phi Coefficient = 9.04, Contingency Coefficient = 0.99 and Cramer's V = 0.87.
All the measures show that there is strong relationship between the variables.
There is also a positive correlation (0.6045) and the P-Value = < 0.0001. The correlation is significant at 0.01%.
This means that, as income increases, there is an increase in other debts. Therefore, we accept the alternate hypothesis H1. More income earners spend more and incur more other debts.

4.0. SUMMARY

From the analysis above, the following were observed:

1. The population is not ageing as a lot of the population is between 29 and 42.
2. Only few persons pursued higher education degrees beyond the high school degree.
3. Very few members of the population earned higher than the rest of the population.
4. Job mobility is high as only few stay longer at a particular work place.
5. Many people in the population are in debt.
6. About 29% of the population defaulted on loans.
7. Younger people had more defaults on loans. There are prevailing factors peculiar with the young: fashion, compulsive Disorder e.g. : shopaholics, little or no stable employment, loan seeking for education, misplaced or misguided priorities, and social factors.
8. We also noticed that the number of people who defaulted on loans were more among those with a college degree (22 out of 30: 73%). Moreover, this was expected, as there is currently a rise in college loan debts among college students.
9. People with larger incomes had lower debts (negative correlation between income and debt to income ratio).
10. As debit to income ratio increased, there was a tendency for credit card debt to rise too.
11. As income increased, there is a tendency for credit card debt to rise too. People with more money rend to spend more with credit cards, as the bank places a higher limit on the amount of money they can spend on their credit cards given the fact that their higher income is a guarantee they will pay back.
12. As income increased, there is a decrease in the default on loans.
13. As income increased, there is an increase in other debts. More income earners spend more and incurs more debts.

APPENDIX 1
Distribution Analysis of AGE

SECTION 1a:

Moments			
N	498	Sum Weights	498
Mean	34.8293173	Sum Observations	17345
Std Deviation	8.13399998	Variance	66.1619557
Skewness	0.32283585	Kurtosis	-0.7294384
Uncorrected SS	636997	Corrected SS	32882.492
Coeff Variation	23.3538887	Std Error Mean	0.36449325

Basic Statistical Measures			
Location		Variability	
Mean	34.82932	Std Deviation	8.13400
Median	34.00000	Variance	66.16196
Mode	29.00000	Range	36.00000
		Interquartile Range	13.00000

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.971848	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.094534	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.663991	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3.991907	Pr > A-Sq	<0.0050

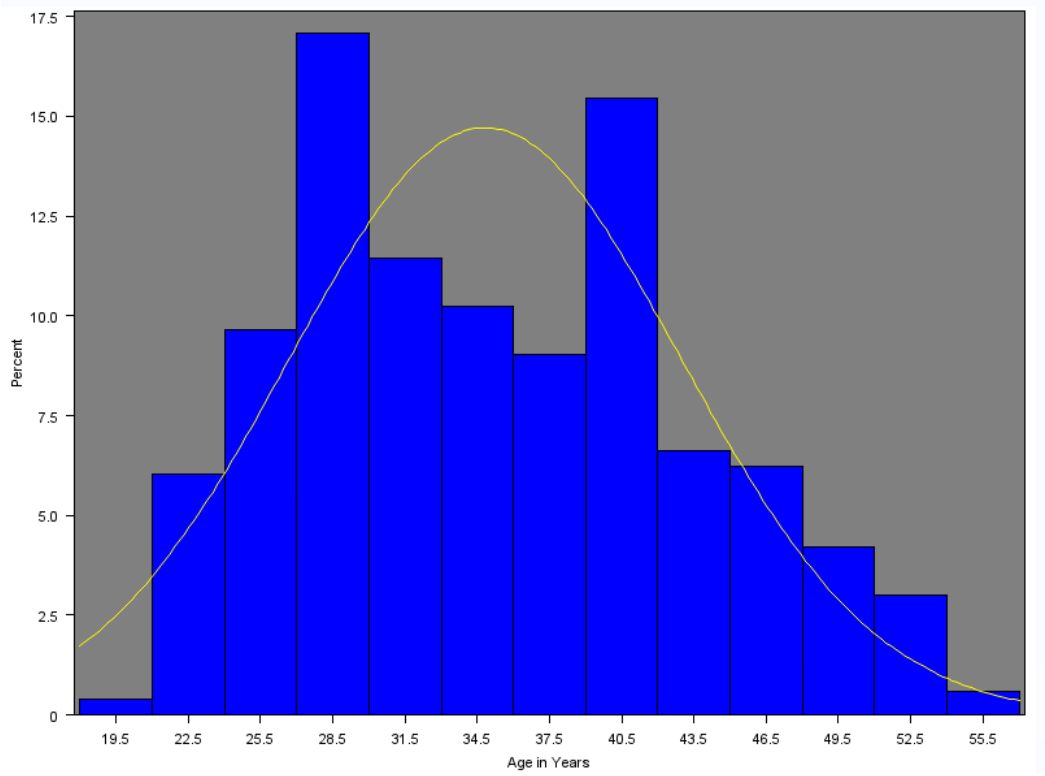
Generated by the SAS System ('Local', X64_SRV12) on 16. joulukuuta 2016 at 1:02:52 PM

Page Break

Distribution analysis of: age

The UNIVARIATE Procedure

SECTION 1 B:



Fitted Normal Distribution for age (Age in Years)

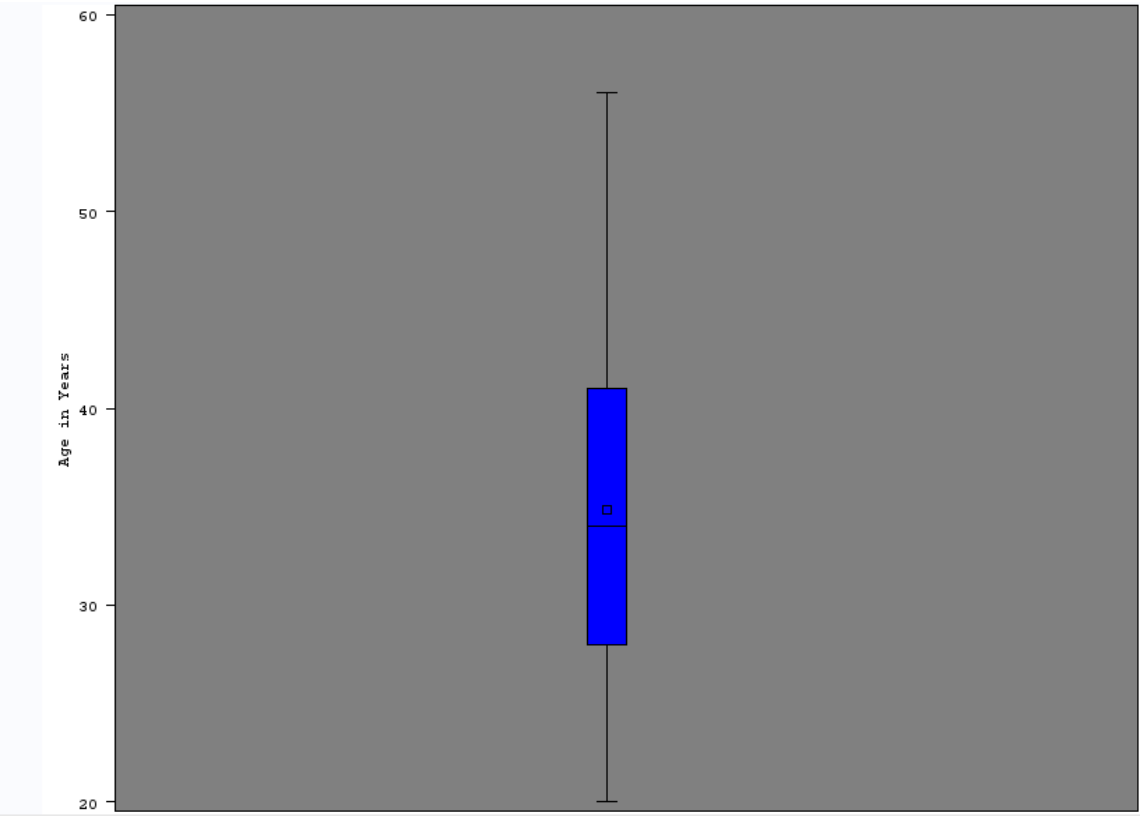
Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	34.82932
Std Dev	Sigma	8.134

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.09453353	Pr > D	<0.010
Cramer-von Mises	W-Sq	0.66399065	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	3.99190700	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	21.0000	15.9068
5.0	23.0000	21.4501
10.0	24.0000	24.4052
25.0	28.0000	29.3430
50.0	34.0000	34.8293
75.0	41.0000	40.3156
90.0	46.0000	45.2535
95.0	49.0000	48.2086
99.0	53.0000	53.7518

Generated by the SAS System ('Local', X64_SRV12) on 16. joulukuuta 2016 at 11:36:43 AM

SECTION 1 C: Box and Whiskers Plot



APPENDIX 2
Distribution Analysis of EDUCATION
SECTION 2 A:

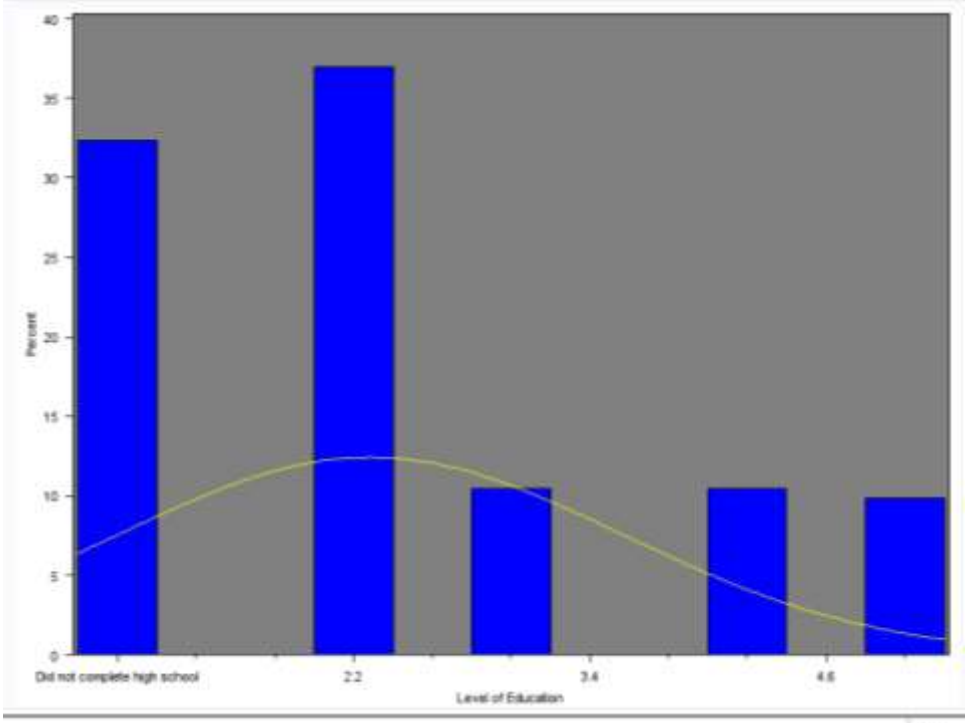
Moments			
N	498	Sum Weights	498
Mean	2.28514056	Sum Observations	1138
Std Deviation	1.28566625	Variance	1.65293771
Skewness	0.87049825	Kurtosis	-0.3605535
Uncorrected SS	3422	Corrected SS	821.51004
Coeff Variation	56.2620204	Std Error Mean	0.05761208

Basic Statistical Measures			
Location		Variability	
Mean	2.285141	Std Deviation	1.28567
Median	2.000000	Variance	1.65294
Mode	2.000000	Range	4.00000
		Interquartile Range	2.00000

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.824352	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.28053	Pr > D	<0.0100
Cramer-von Mises	W-Sq	5.61323	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	32.83038	Pr > A-Sq	<0.0050

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:14:13 PM

SECTION 2 B:



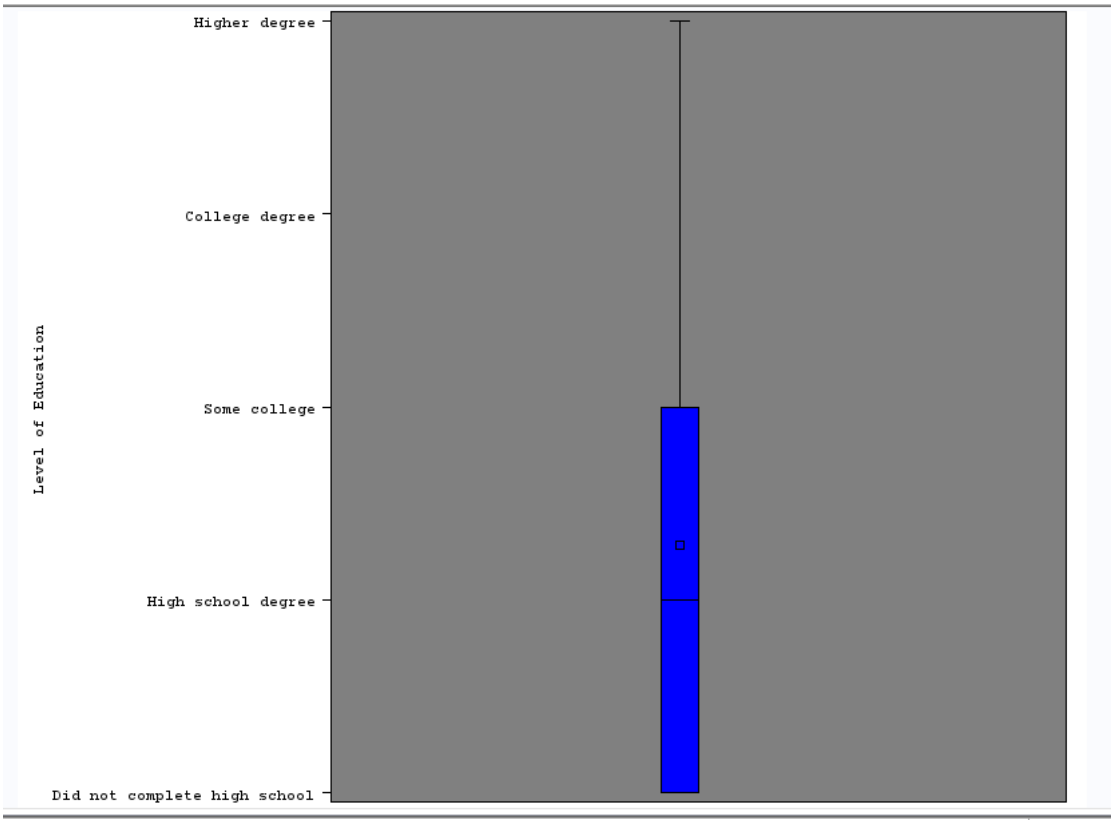
Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	2.285141
Std Dev	Sigma	1.285666

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.2805302	Pr > D	<0.010
Cramer-von Mises	W-Sq	5.6132302	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	32.8303828	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	1.00000	-0.70577
5.0	1.00000	0.17041
10.0	1.00000	0.63749
25.0	1.00000	1.41797
50.0	2.00000	2.28514
75.0	3.00000	3.15231
90.0	4.00000	3.93279
95.0	5.00000	4.39987
99.0	5.00000	5.27605

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:14:13 PM

SECTION 2 C: BOX and Whiskers Plot



APPENDIX 3
Distribution Analysis of EMPLOYMENT
SECTION 3 A:

Moments			
N	498	Sum Weights	498
Mean	8.16666667	Sum Observations	4067
Std Deviation	6.75455978	Variance	45.6240778
Skewness	0.8697534	Kurtosis	0.31316477
Uncorrected SS	55889	Corrected SS	22675.1667
Coeff Variation	82.7088952	Std Error Mean	0.30267906

Basic Statistical Measures			
Location		Variability	
Mean	8.166667	Std Deviation	6.75456
Median	7.000000	Variance	45.62408
Mode	0.000000	Range	31.00000
		Interquartile Range	10.00000

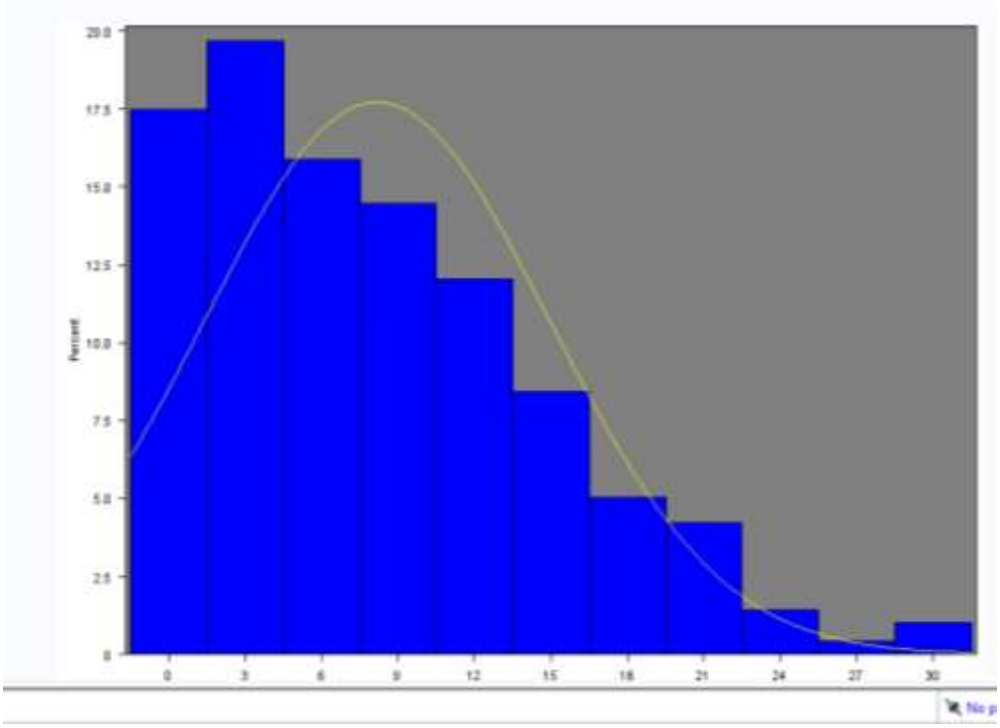
Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	8.16667	7.57198	8.76135
Std Deviation	6.75456	6.35951	7.20234
Variance	45.62408	40.44333	51.87374

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	26.98127	Pr > |t|	<.0001
Sign	M	223	Pr >= |M|	<.0001
Signed Rank	S	49840.5	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.923651	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.117776	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.445612	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	9.623536	Pr > A-Sq	<0.0050

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:11:52 PM

SECTION 3 B:



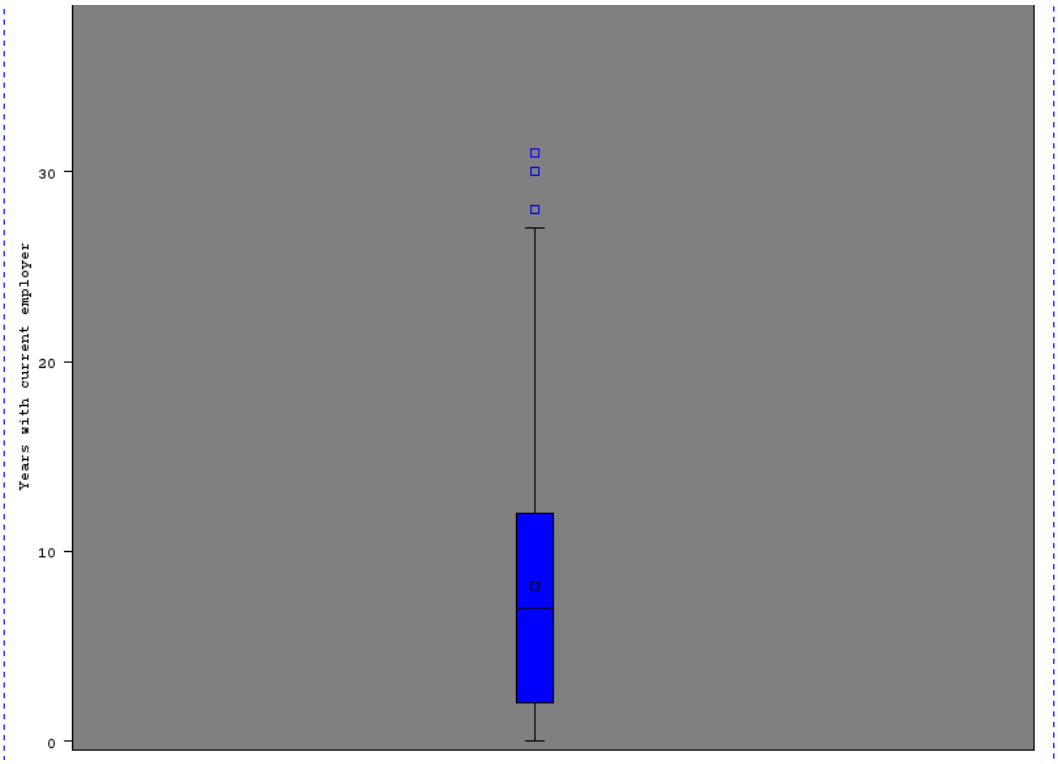
Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	8.166667
Std Dev	Sigma	6.75456

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.11777589	Pr > D	<0.010
Cramer-von Mises	W-Sq	1.44561232	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	9.62353639	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	0.0000	-7.54679
5.0	0.0000	-2.94360
10.0	0.0000	-0.48965
25.0	2.0000	3.61079
50.0	7.0000	8.16667
75.0	12.0000	12.72255
90.0	18.0000	16.82298
95.0	22.0000	19.27693
99.0	30.0000	23.88012

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:11:52 PM

SECTION 3 C: BOX and Whiskers Plot


Years with current employer
30
20
10
0

APPENDIX 4
Distribution Analysis of INCOME
SECTION 4 A:

Variable: income (Household income in thousands ($))

Moments			
N	498	Sum Weights	498
Mean	48.1365462	Sum Observations	23972
Std Deviation	40.3282985	Variance	1626.37166
Skewness	3.84197548	Kurtosis	24.2861021
Uncorrected SS	1962236	Corrected SS	808306.715
Coeff Variation	83.7789615	Std Error Mean	1.80715427

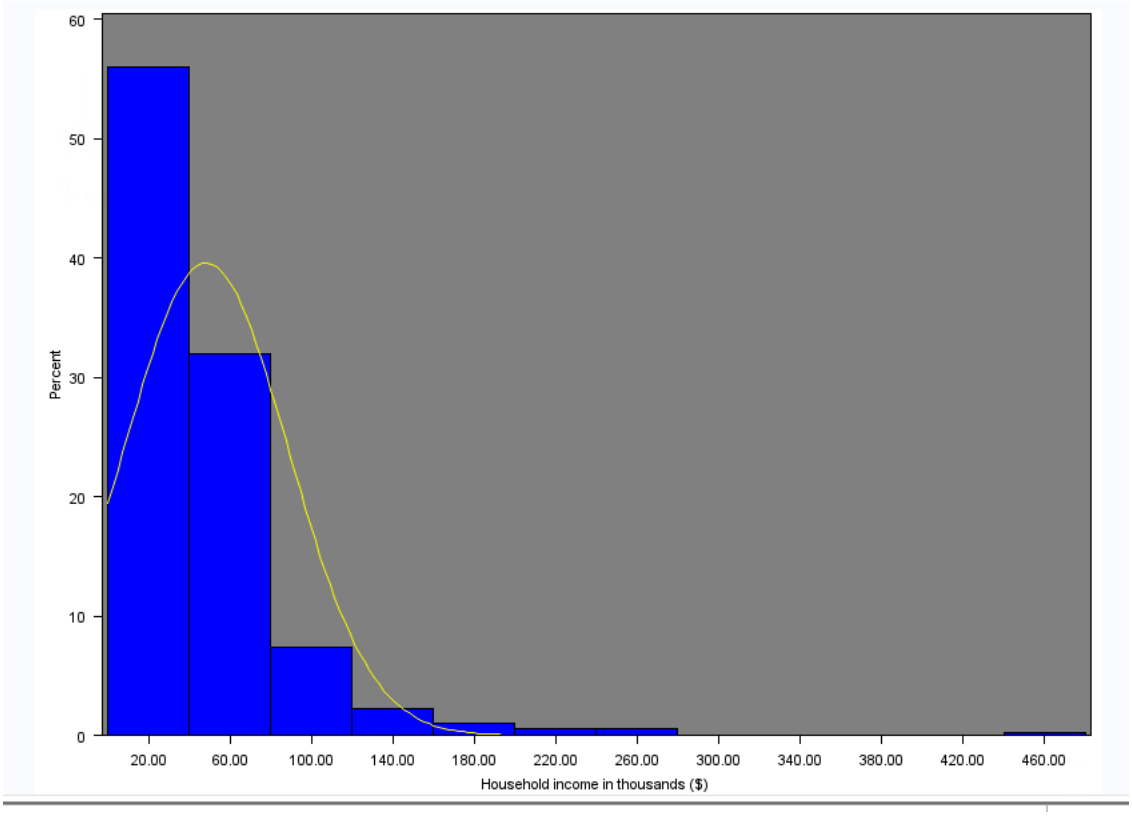
Basic Statistical Measures			
Location		Variability	
Mean	48.13655	Std Deviation	40.32830
Median	36.00000	Variance	1626
Mode	21.00000	Range	432.00000
		Interquartile Range	32.00000

Note: The mode displayed is the smallest of 3 modes with a count of 16.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.662338	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.198646	Pr > D	<0.0100
Cramer-von Mises	W-Sq	6.987985	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	39.12007	Pr > A-Sq	<0.0050

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:10:47 PM

SECTION 4 B:



Fitted Normal Distribution for income (Household income in thousands ($))

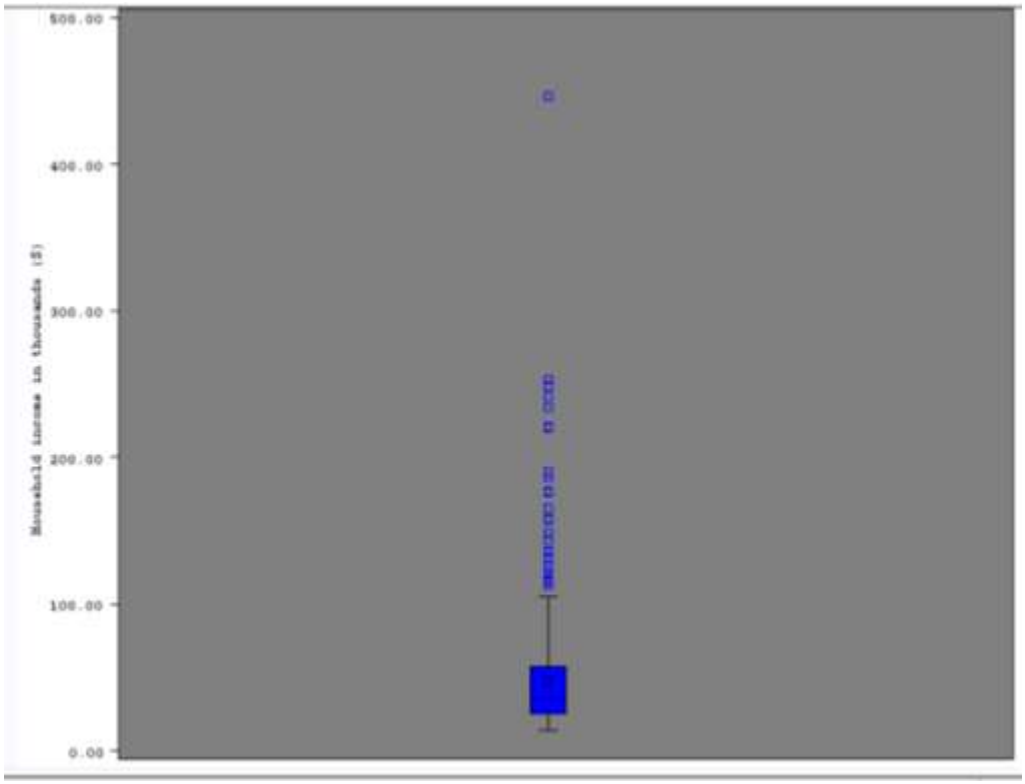
Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	48.13655
Std Dev	Sigma	40.3283

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.1986463	Pr > D	<0.010
Cramer-von Mises	W-Sq	6.9879849	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	39.1200689	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	15.0000	-45.68111
5.0	17.0000	-18.19760
10.0	20.0000	-3.54625
25.0	25.0000	20.93552
50.0	36.0000	48.13655
75.0	57.0000	75.33757
90.0	86.0000	99.81934
95.0	116.0000	114.47069
99.0	234.0000	141.95420

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:10:47 PM

SECTION 4 C: BOX and Whiskers Plot



APPENDIX 5
Distribution Analysis of Debt – To - Income
SECTION 5 A:

The UNIVARIATE Procedure
Variable: debtinc (Debt to income ratio (x100)

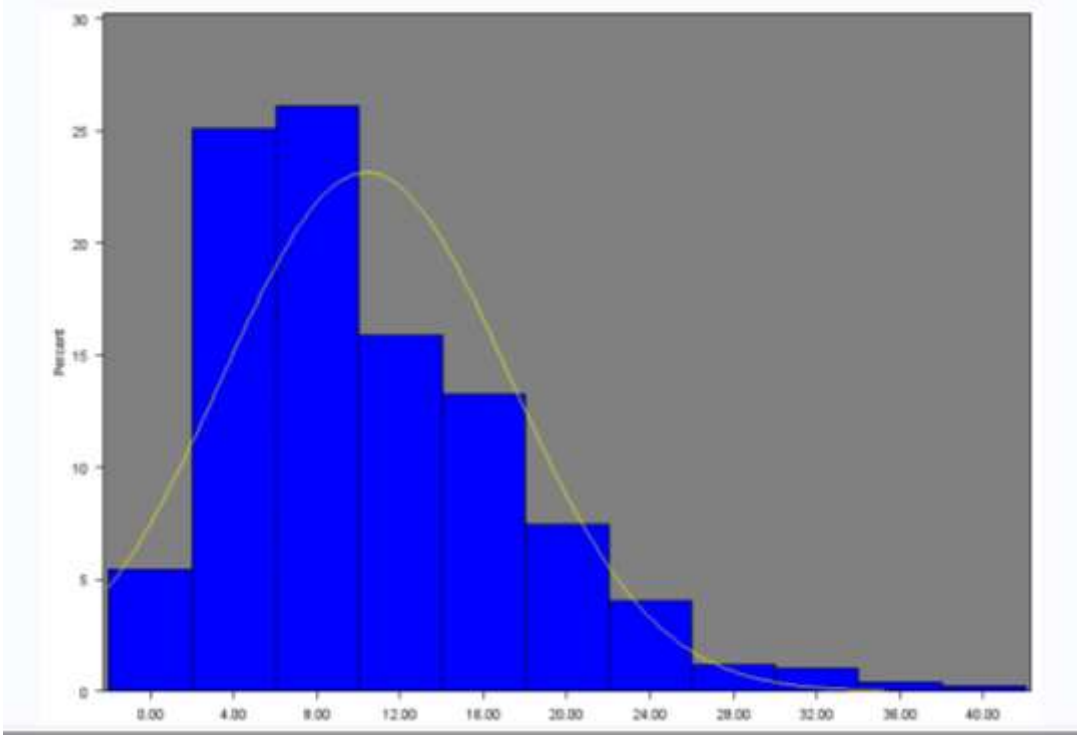
Moments			
N	498	Sum Weights	498
Mean	10.3564257	Sum Observations	5157.5
Std Deviation	6.89632107	Variance	47.5592443
Skewness	1.0921039	Kurtosis	1.36364264
Uncorrected SS	77050.21	Corrected SS	23636.9444
Coeff Variation	66.5897798	Std Error Mean	0.30903154

Basic Statistical Measures			
Location		Variability	
Mean	10.35643	Std Deviation	6.89632
Median	8.55000	Variance	47.55924
Mode	4.50000	Range	40.70000
		Interquartile Range	9.50000

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.923605	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.106543	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.519668	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	9.227837	Pr > A-Sq	<0.0050

Generated by the SAS System ('Local', X64_SRV12) on 16. joulukuuta 2016 at 1:58:42 PM

SECTION 5 B:



Fitted Normal Distribution for debtinc (Debt to income ratio (x100)

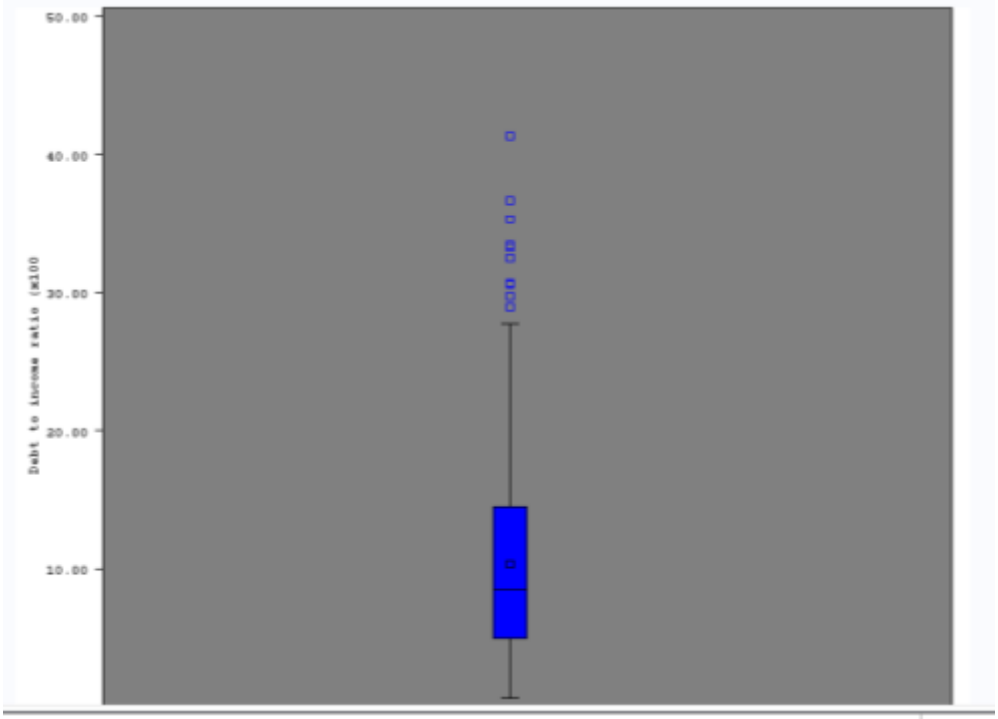
Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	10.35643
Std Dev	Sigma	6.896321

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.10654292	Pr > D	<0.010
Cramer-von Mises	W-Sq	1.51966790	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	9.22783693	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	0.90000	-5.68682
5.0	1.90000	-0.98701
10.0	2.80000	1.51843
25.0	5.00000	5.70493
50.0	8.55000	10.35643
75.0	14.50000	15.00792
90.0	19.80000	19.19442
95.0	23.30000	21.69986
99.0	33.30000	26.39967

Generated by the SAS System ('Local', X64_SRV12) on 16. joulukuuta 2016 at 1:58:42 PM

SECTION 5 C: Box and Whiskers Plot



APPENDIX 6
Distribution Analysis of CREDIT CARD DEBTS
SECTION 6 A:

Variable: creddebt (Credit card debt in thousands ($))

Moments			
N	498	Sum Weights	498
Mean	1.59614901	Sum Observations	794.882206
Std Deviation	2.10488737	Variance	4.43055086
Skewness	3.41227448	Kurtosis	15.8527337
Uncorrected SS	3470.73422	Corrected SS	2201.98378
Coeff Variation	131.872862	Std Error Mean	0.09432226

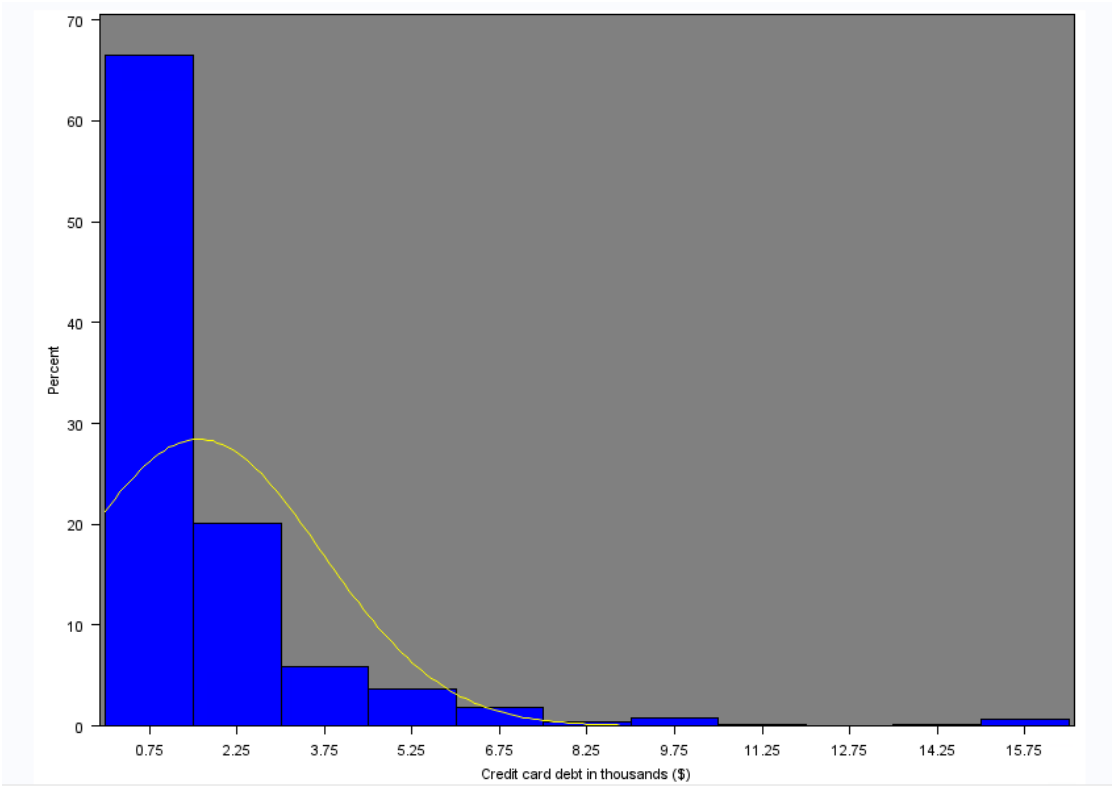
Basic Statistical Measures			
Location		Variability	
Mean	1.596149	Std Deviation	2.10489
Median	0.904983	Variance	4.43055
Mode	0.085785	Range	16.01664
		Interquartile Range	1.59271

Note: The mode displayed is the smallest of 2 modes with a count of 2.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.647737	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.226249	Pr > D	<0.0100
Cramer-von Mises	W-Sq	8.39778	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	45.88088	Pr > A-Sq	<0.0050

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 5:01:47 PM

SECTION 6 B:



Fitted Normal Distribution for creddebt (Credit card debt in thousands ($))

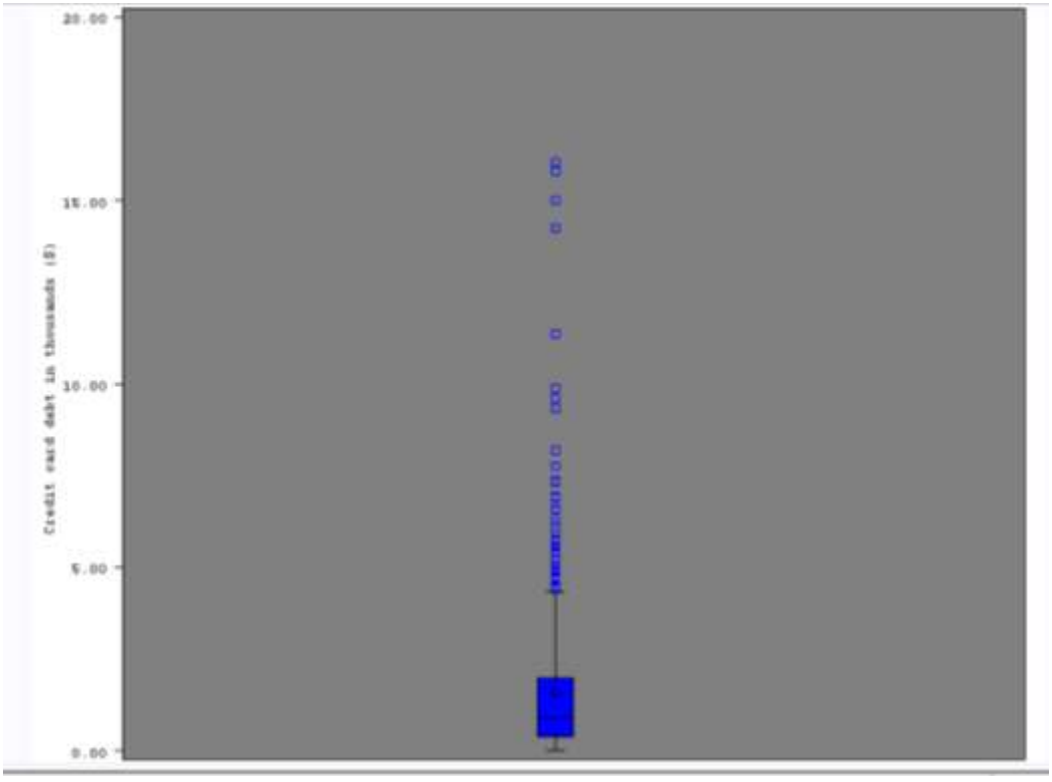
Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	1.596149
Std Dev	Sigma	2.104887

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.2262486	Pr > D	<0.010
Cramer-von Mises	W-Sq	8.3977801	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	45.8808847	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	0.02941	-3.30055
5.0	0.11070	-1.86608
10.0	0.18251	-1.10137
25.0	0.37721	0.17642
50.0	0.90498	1.59615
75.0	1.96992	3.01587
90.0	3.70370	4.29367
95.0	5.54954	5.05838
99.0	11.35939	6.49285

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 5:01:47 PM

SECTION 6 C: Box and Whiskers Plot



APPENDIX 7
Distribution Analysis of DEFAULT ON LOANS
SECTION 7 A:

Variable: default (Previously loan defaulted)

Moments			
N	498	Sum Weights	498
Mean	0.29116466	Sum Observations	145
Std Deviation	0.45475605	Variance	0.20680307
Skewness	0.9221534	Kurtosis	-1.154285
Uncorrected SS	145	Corrected SS	102.781124
Coeff Variation	156.185182	Std Error Mean	0.02037811

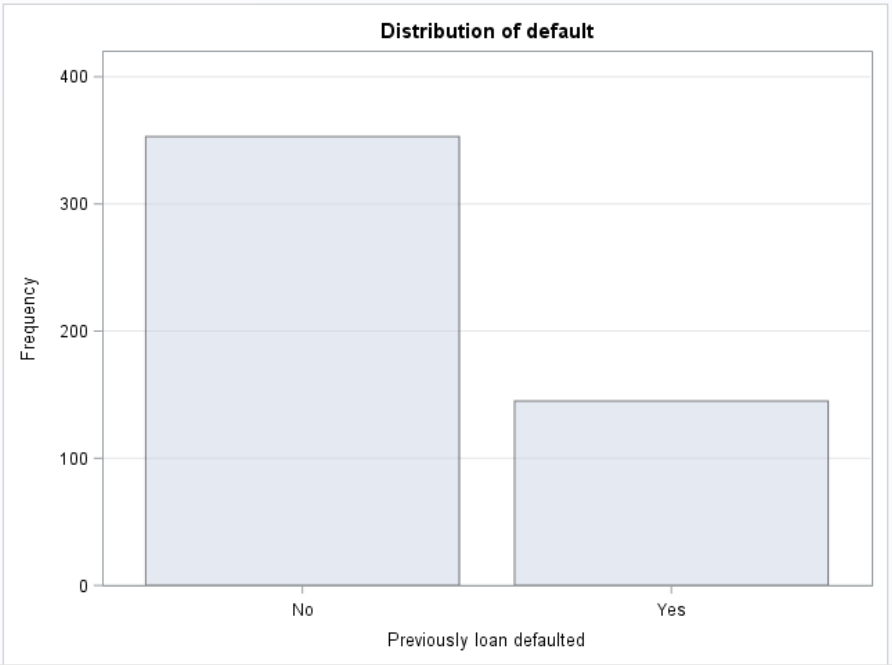
Basic Statistical Measures			
Location		Variability	
Mean	0.291165	Std Deviation	0.45476
Median	0.000000	Variance	0.20680
Mode	0.000000	Range	1.00000
		Interquartile Range	1.00000

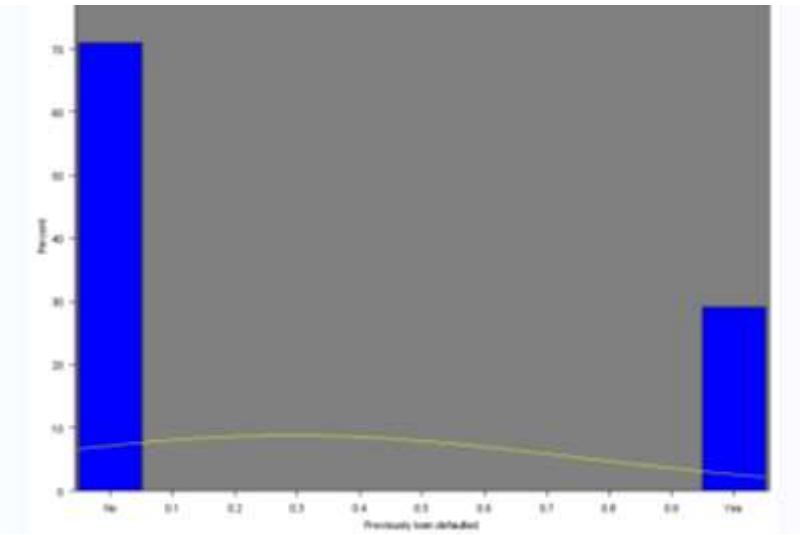
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.56974	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.447835	Pr > D	<0.0100
Cramer-von Mises	W-Sq	19.95898	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	110.275	Pr > A-Sq	<0.0050

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 5:03:07 PM

SECTION 7 B:

Previously loan defaulted				
default	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	353	70.88	353	70.88
Yes	145	29.12	498	100.00





Fitted Normal Distribution for default (Previously loan defaulted)

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	0.291165
Std Dev	Sigma	0.454756

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.447835	Pr > D	<0.010
Cramer-von Mises	W-Sq	19.958982	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	110.275008	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	0.00000	-0.76676
5.0	0.00000	-0.45684
10.0	0.00000	-0.29163
25.0	0.00000	-0.01556
50.0	0.00000	0.29116
75.0	1.00000	0.59789
90.0	1.00000	0.87396
95.0	1.00000	1.03917
99.0	1.00000	1.34909

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 5:03:07 PM

SECTION 7 C: Box and Whiskers Plot

APPENDIX 8
Distribution Analysis of RESIDENCE
SECTION 8 A:

Distribution analysis of: residence

The UNIVARIATE Procedure
Variable: residence (Place of residence)

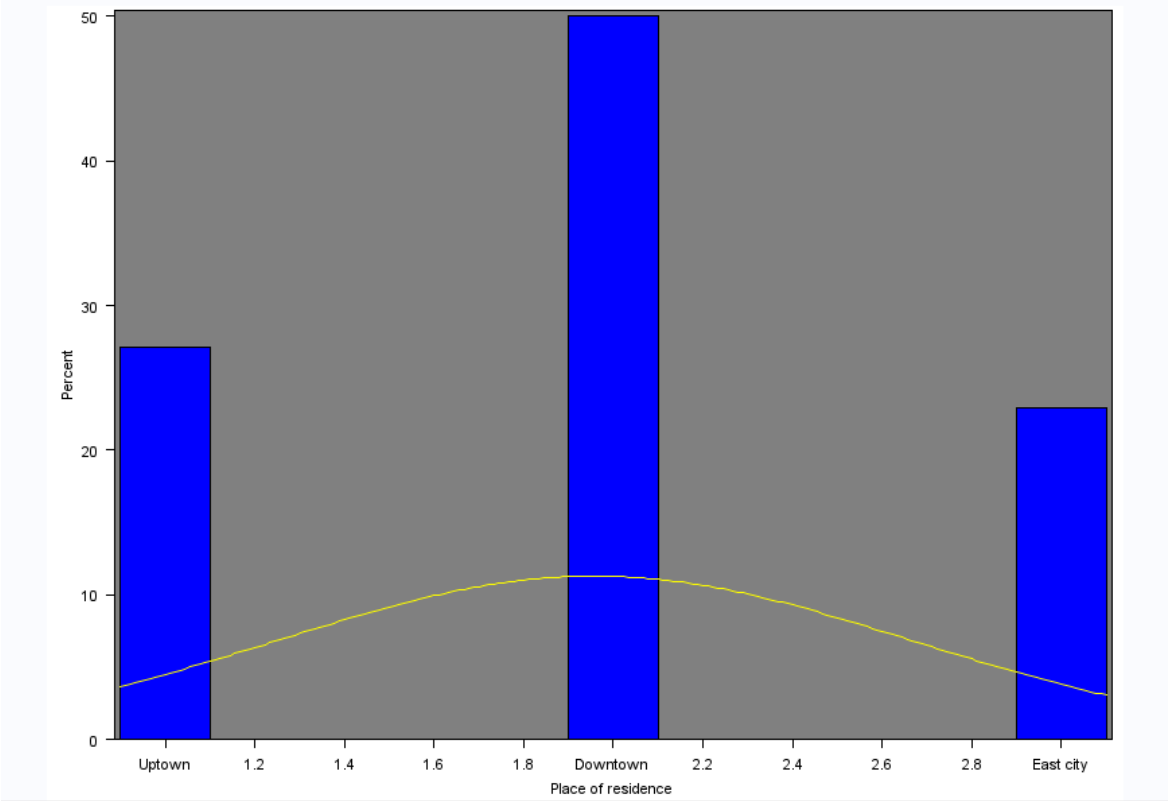
Moments			
N	498	Sum Weights	498
Mean	1.95783133	Sum Observations	975
Std Deviation	0.70655804	Variance	0.49922426
Skewness	0.05970861	Kurtosis	-0.9907958
Uncorrected SS	2157	Corrected SS	248.114458
Coeff Variation	36.0888106	Std Error Mean	0.03166162

Basic Statistical Measures			
Location		Variability	
Mean	1.957831	Std Deviation	0.70656
Median	2.000000	Variance	0.49922
Mode	2.000000	Range	2.00000
		Interquartile Range	1.00000

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.806964	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.252711	Pr > D	<0.0100
Cramer-von Mises	W-Sq	7.048296	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	41.29203	Pr > A-Sq	<0.0050

Generated by the SAS System ('Local', X64_SRV12) on 16. joulukuuta 2016 at 4:12:48 PM

SECTION 8 B:



Fitted Normal Distribution for residence (Place of residence)

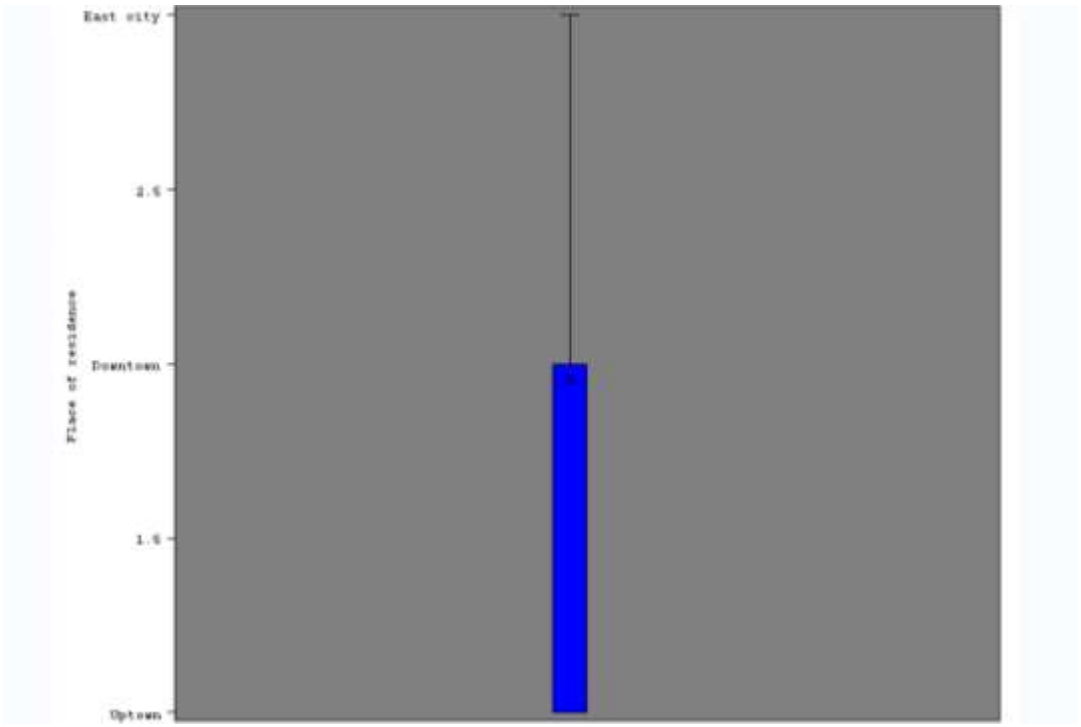
Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	1.957831
Std Dev	Sigma	0.706558

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.2527111	Pr > D	<0.010
Cramer-von Mises	W-Sq	7.0482964	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	41.2920269	Pr > A-Sq	<0.005

Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	1.00000	0.31413
5.0	1.00000	0.79565
10.0	1.00000	1.05234
25.0	1.00000	1.48127
50.0	2.00000	1.95783
75.0	2.00000	2.43440
90.0	3.00000	2.86332
95.0	3.00000	3.12002
99.0	3.00000	3.60153

Generated by the SAS System ('Local', X64_SRV12) on 16. joulukuuta 2016 at 4:12:48 PM

SECTION 8 C: Box and Whiskers Plot


East city
2.5
Downtown
1.5
Uptown
Place of residence

APPENDIX 9
Distribution Analysis of OTHER DEBTS
SECTION 9 A:

Variable: othdebt (Other debt in thousands ($))

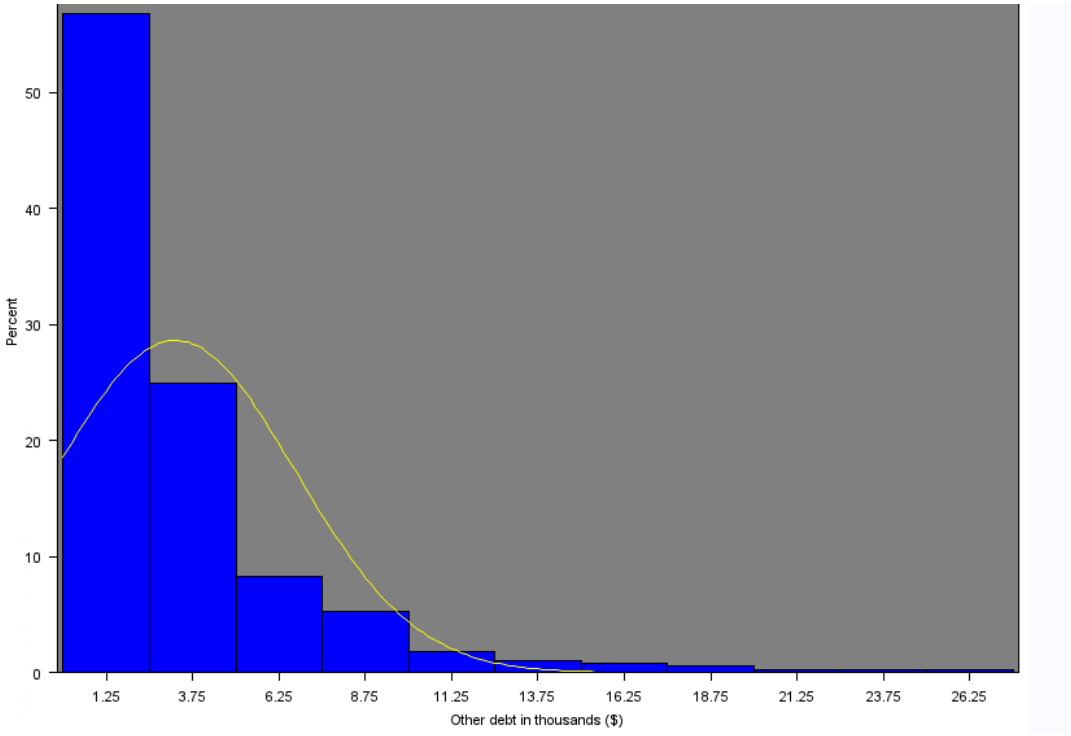
Moments			
N	498	Sum Weights	498
Mean	3.24199356	Sum Observations	1614.51279
Std Deviation	3.48503104	Variance	12.1454413
Skewness	2.7214447	Kurtosis	10.0931169
Uncorrected SS	11270.5244	Corrected SS	6036.28434
Coeff Variation	107.496544	Std Error Mean	0.15616797

Basic Statistical Measures			
Location		Variability	
Mean	3.241994	Std Deviation	3.48503
Median	2.107296	Variance	12.14544
Mode	7.823400	Range	26.98802
		Interquartile Range	3.06546

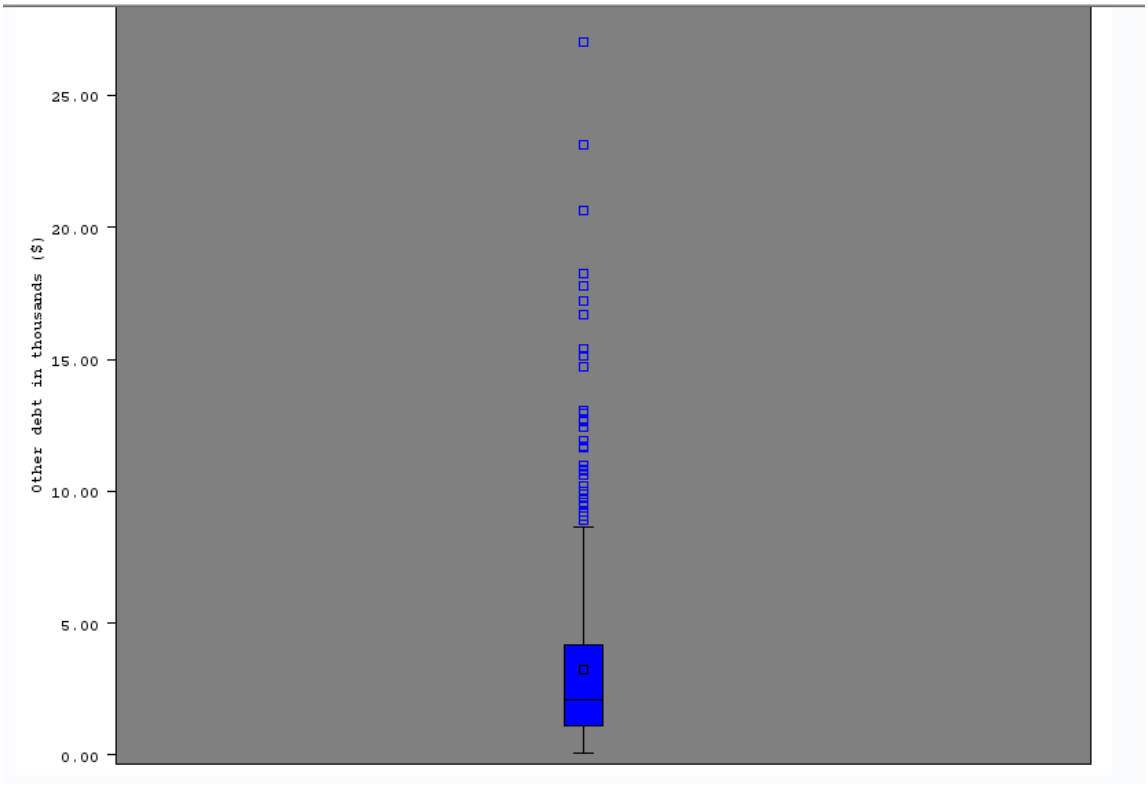
Basic Confidence Limits Assuming Normality			
Parameter	Estimate	95% Confidence Limits	
Mean	3.24199	2.93516	3.54882
Std Deviation	3.48503	3.28120	3.71607
Variance	12.14544	10.76629	13.80914

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.724578	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.180835	Pr > D	<0.0100
Cramer-von Mises	W-Sq	6.484599	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	36.31359	Pr > A-Sq	<0.0050

SECTION 9 B:



SECTION 9 C: Box and Whiskers Plot



APPENDIX 10

Statistic	DF	Value	Prob
Chi-Square	72	71.4422	0.4964
Likelihood Ratio Chi-Square	72	76.0664	0.3490
Mantel-Haenszel Chi-Square	1	0.2865	0.5924
Phi Coefficient		0.3788	
Contingency Coefficient		0.3542	
Cramer's V		0.2678	
WARNING: 64% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 498

Correlation Analysis

The CORR Procedure

1 With Variables:	residence
1 Variables:	age

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
residence	498	1.95783	0.70656	2.00000	1.00000	3.00000	Place of residence
age	498	34.82932	8.13400	34.00000	20.00000	56.00000	Age in Years

Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	age
residence	-0.02401
Place of residence	0.5930

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	age
residence	-0.01606
Place of residence	0.7207

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 3:55:54 PM

APPENDIX 11

Statistics for Table of employ by age

Statistic	DF	Value	Prob
Chi-Square	1044	1495.5976	<.0001
Likelihood Ratio Chi-Square	1044	909.2442	0.9989
Mantel-Haenszel Chi-Square	1	151.1621	<.0001
Phi Coefficient		1.7330	
Contingency Coefficient		0.8661	
Cramer's V		0.3218	
WARNING: 100% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 498

Generated by the SAS System ('Local', X64_SRV12) on 20. joulukuuta 2016 at 11:20:18 AM

Correlation Analysis

The CORR Procedure

1 With Variables:	age
1 Variables:	employ

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
age	498	34.82932	8.13400	34.00000	20.00000	56.00000	Age in Years
employ	498	8.16667	6.75456	7.00000	0	31.00000	Years with current employer

Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	employ
age	0.55150
Age in Years	<.0001

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

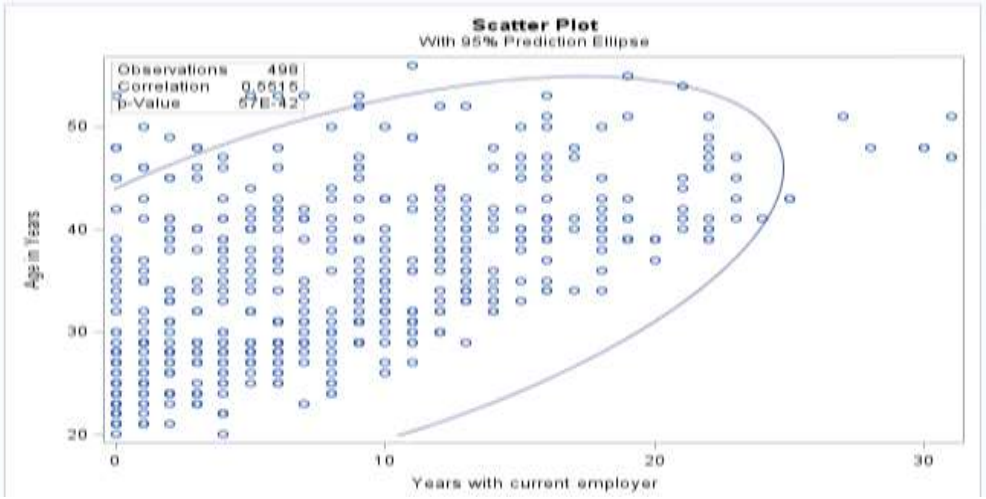
	employ
age	0.55713
Age in Years	<.0001

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 3:55:23 PM

Page Break

Correlation Analysis

The CORR Procedure



APPENDIX 12

Statistics for Table of ed by age

Statistic	DF	Value	Prob
Chi-Square	144	175.2553	0.0391
Likelihood Ratio Chi-Square	144	198.2497	0.0018
Mantel-Haenszel Chi-Square	1	0.1961	0.6579
Phi Coefficient		0.5932	
Contingency Coefficient		0.5102	
Cramer's V		0.2966	
WARNING: 82% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 498

Generated by the SAS System ('Local', X64_SRV12) on 20. joulukuuta 2016 at 11:52:02 AM

The CORR Procedure

1 With Variables:	age
1 Variables:	ed

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
age	498	34.82932	8.13400	34.00000	20.00000	56.00000	Age in Years
ed	498	2.28514	1.28567	2.00000	1.00000	5.00000	Level of Education

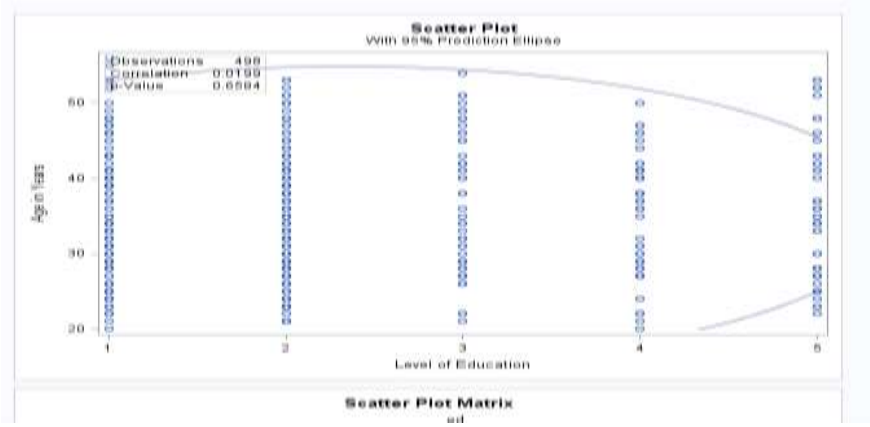
Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	ed
age	0.01986
Age in Years	0.6584

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	ed
age	0.00335
Age in Years	0.9406

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:34:02 PM


Scatter Plot
With 95% Prediction Ellipse
Observations 498
Correlation 0.0159
p-Value 0.6584
Age in Years
50
40
30
20
1
2
3
4
5
Level of Education
Scatter Plot Matrix

APPENDIX 13

Correlation Analysis

The CORR Procedure

1 With Variables:	age
1 Variables:	income

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
age	498	34.82932	8.13400	34.00000	20.00000	56.00000	Age in Years
income	498	48.13655	40.32830	36.00000	14.00000	446.00000	Household income in thousands ($)

Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	income
age	0.46990
Age in Years	<.0001

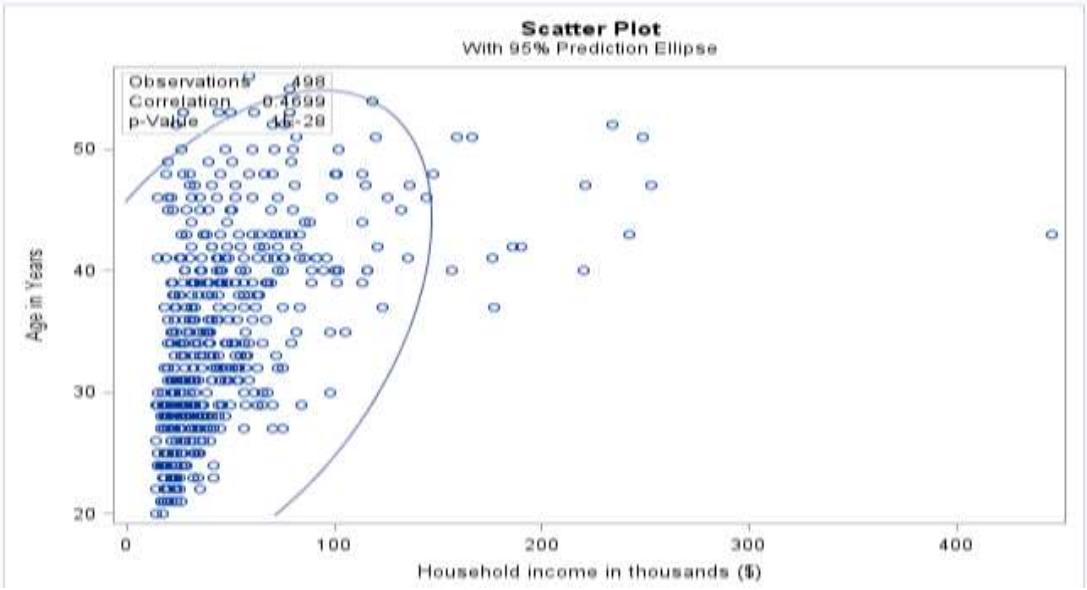
Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	income
age	0.58454
Age in Years	<.0001

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 3:54:51 PM

Correlation Analysis

The CORR Procedure



APPENDIX 14

Correlation Analysis

The CORR Procedure

1 With Variables:	age
1 Variables:	default

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
age	498	34.82932	8.13400	34.00000	20.00000	56.00000	Age in Years
default	498	0.29116	0.45476	0	0	1.00000	Previously loan defaulted

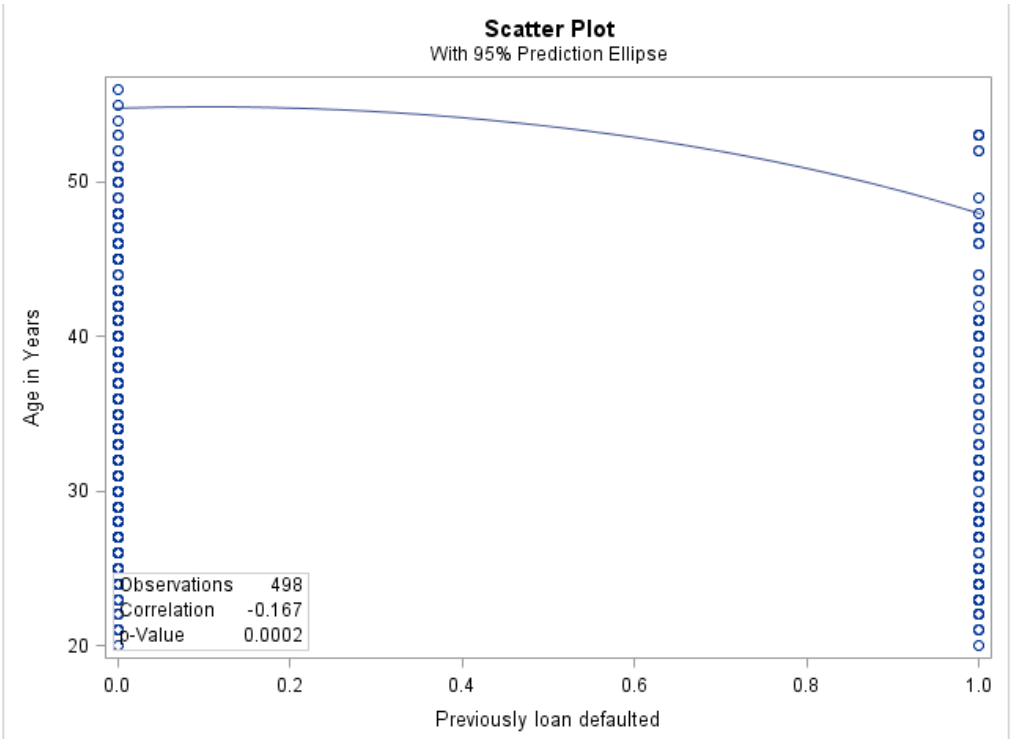
Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	default
age	-0.16713
Age in Years	0.0002

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	default
age	-0.18103
Age in Years	<.0001

Generated by the SAS System ('Local', X64_SRV12) on 16. joulukuuta 2016 at 2:01:56 PM



APPENDIX 15

Correlation Analysis

The CORR Procedure

1 With Variables:	residence
1 Variables:	ed

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
residence	498	1.95783	0.70656	2.00000	1.00000	3.00000	Place of residence
ed	498	2.28514	1.28567	2.00000	1.00000	5.00000	Level of Education

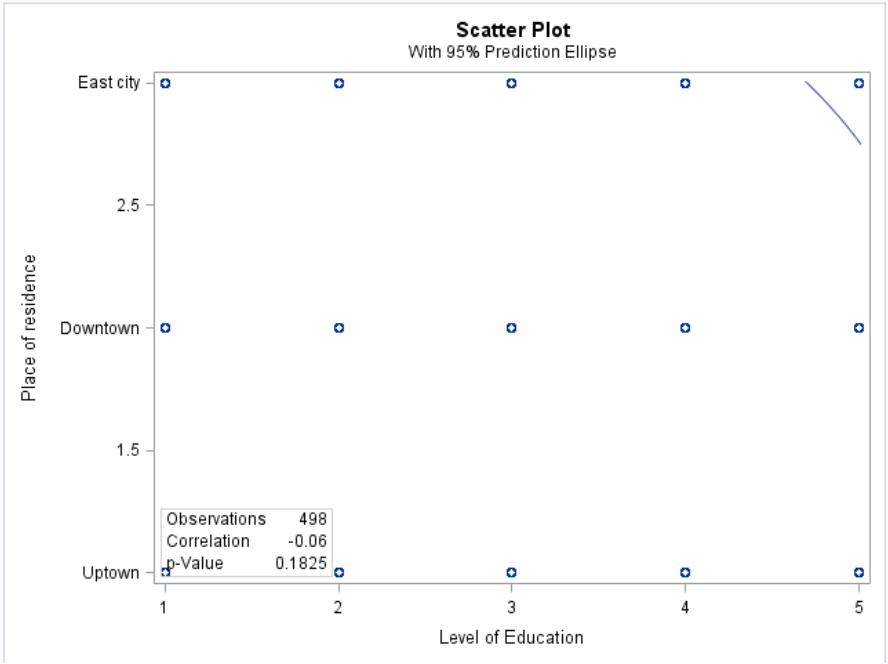
Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	ed
residence	-0.05983
Place of residence	0.1825

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	ed
residence	-0.05453
Place of residence	0.2245

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 3:55:36 PM



APPENDIX 16

1 With Variables:	default
1 Variables:	ed

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
default	498	0.29116	0.45476	0	0	1.00000	Previously loan defaulted
ed	498	2.28514	1.28567	2.00000	1.00000	5.00000	Level of Education

Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	ed
default	0.08829
Previously loan defaulted	0.0489

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	ed
default	0.08923
Previously loan defaulted	0.0466

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:39:50 PM

Table of ed by default

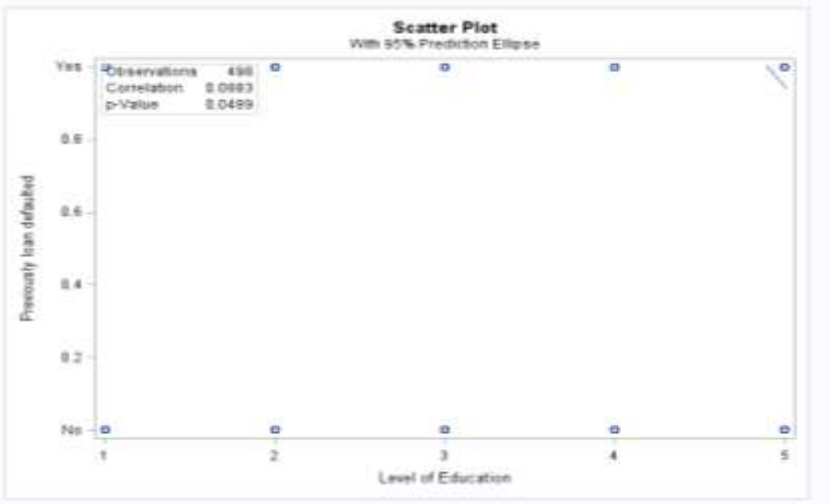
		default(Previously loan defaulted)		
		No	Yes	Total
ed(Level of Education)				
Did not complete high school	Frequency	122	39	161
	Col Pct	34.56	26.90	
High school degree	Frequency	130	54	184
	Col Pct	36.83	37.24	
Some college	Frequency	38	14	52
	Col Pct	10.76	9.66	
College degree	Frequency	30	22	52
	Col Pct	8.50	15.17	
Higher degree	Frequency	33	16	49
	Col Pct	9.35	11.03	
Total	Frequency	353	145	498

Statistics for Table of ed by default

Statistic	DF	Value	Prob
Chi-Square	4	6.6747	0.1541
Likelihood Ratio Chi-Square	4	6.4412	0.1685
Mantel-Haenszel Chi-Square	1	3.8740	0.0490
Phi Coefficient		0.1158	
Contingency Coefficient		0.1150	
Cramer's V		0.1158	

Cochran-Armitage Trend Test	
Statistic (Z)	-1.9702
One-sided Pr < Z	0.0244
Two-sided Pr > |Z|	0.0488

Sample Size = 498


Scatter Plot
With 95% Prediction Ellipse
Observations 498
Correlation 0.0883
p-Value 0.0489
Yes
No
0.8
0.6
0.4
0.2
Previously loan defaulted
1
2
3
4
5
Level of Education

APPENDIX 17

1 With Variables:	debtinc
1 Variables:	ed

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
debtinc	498	10.35643	6.89632	8.55000	0.60000	41.30000	Debt to income ratio (x100
ed	498	2.28514	1.28567	2.00000	1.00000	5.00000	Level of Education

Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	ed
debtinc	0.02174
Debt to income ratio (x100	0.6284

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

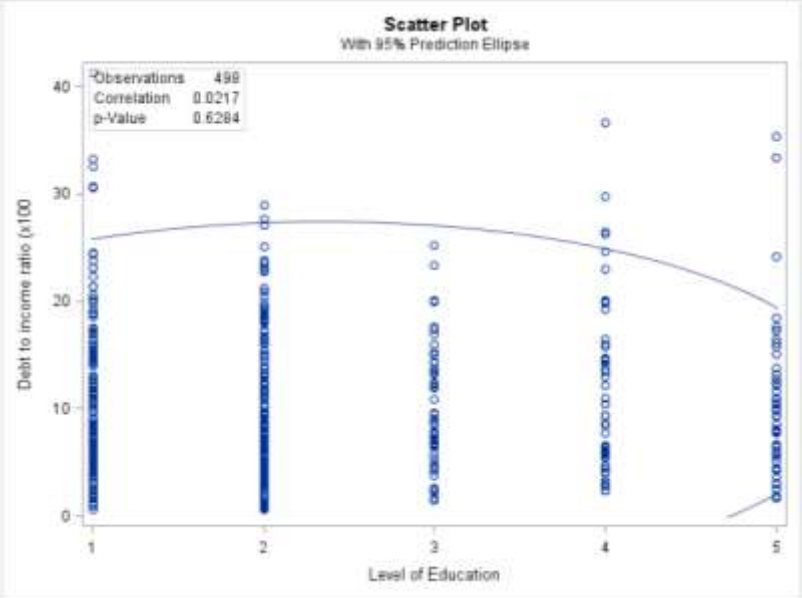
	ed
debtinc	0.02218
Debt to income ratio (x100	0.6214

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:46:29 PM

Statistic	DF	Value	Prob
Chi-Square	816	789.8759	0.7380
Likelihood Ratio Chi-Square	816	735.4215	0.9797
Mantel-Haenszel Chi-Square	1	0.2348	0.6280
Phi Coefficient		1.2594	
Contingency Coefficient		0.7831	
Cramer's V		0.6297	

WARNING: 100% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 498



APPENDIX 18

1 With Variables:	debtinc
1 Variables:	income

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
debtinc	498	10.35643	6.89632	8.55000	0.60000	41.30000	Debt to income ratio (x100
income	498	48.13655	40.32830	36.00000	14.00000	446.00000	Household income in thousands ($)

Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	income
debtinc	-0.05299
Debt to income ratio (x100	0.2378

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

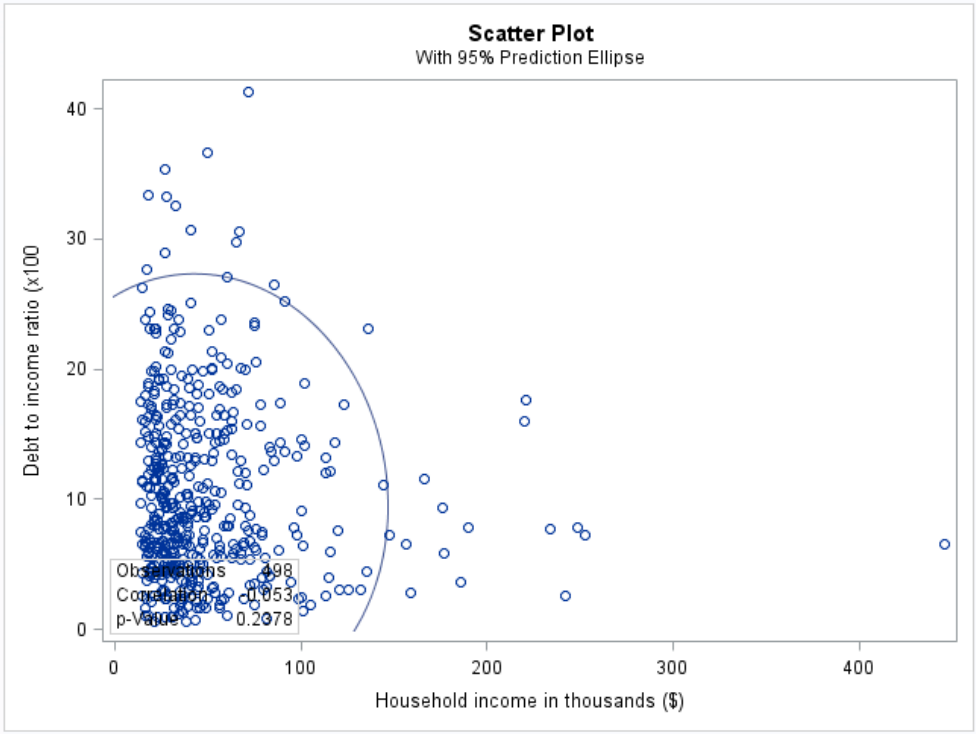
	income
debtinc	-0.04668
Debt to income ratio (x100	0.2985

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:26:07 PM

Statistic	DF	Value	Prob
Chi-Square	21828	20575.3212	1.0000
Likelihood Ratio Chi-Square	21828	3261.9976	1.0000
Mantel-Haenszel Chi-Square	1	1.3956	0.2375
Phi Coefficient		6.4277	
Contingency Coefficient		0.9881	
Cramer's V		0.6214	

WARNING: 100% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 498


Scatter Plot
With 95% Prediction Ellipse
Debt to income ratio (x100
40
30
20
10
0
0
100
200
300
400
Household income in thousands ($)
498
-0.053
0.2078

APPENDIX 19

1 With Variables:	creddebt
1 Variables:	debtinc

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
creddebt	498	1.59615	2.10489	0.90498	0.01484	16.03147	Credit card debt in thousands ($)
debtinc	498	10.35643	6.89632	8.55000	0.60000	41.30000	Debt to income ratio (x100

Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	debtinc
creddebt	0.48476
Credit card debt in thousands ($)	<.0001

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

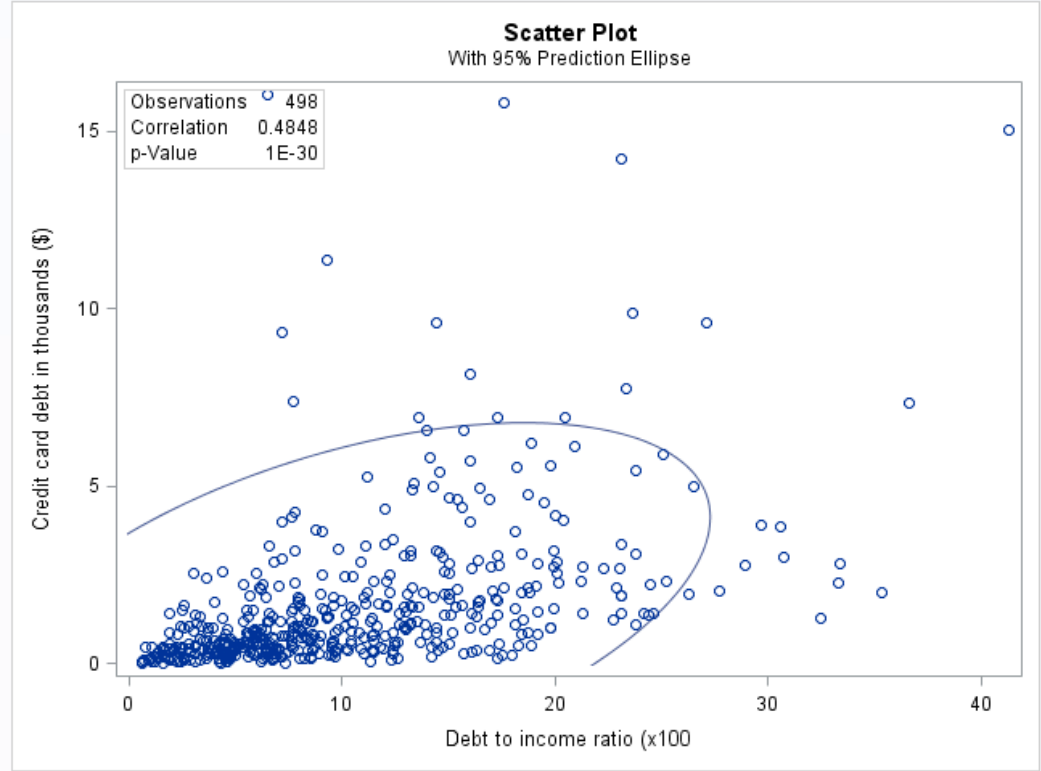
	debtinc
creddebt	0.61709
Credit card debt in thousands ($)	<.0001

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:45:00 PM

Statistic	DF	Value	Prob
Chi-Square	54060	57930.6800	<.0001
Likelihood Ratio Chi-Square	54060	4307.2574	1.0000
Mantel-Haenszel Chi-Square	1	116.8063	<.0001
Phi Coefficient		10.7855	
Contingency Coefficient		0.9957	
Cramer's V		0.7551	

WARNING: 100% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 498


Scatter Plot
With 95% Prediction Ellipse
Observations 498
Correlation 0.4848
p-Value 1E-30
Credit card debt in thousands ($)
Debt to income ratio (x100
0
5
10
15
0
10
20
30
40

APPENDIX 20

1 With Variables:	income
1 Variables:	creddebt

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
income	498	48.13655	40.32830	36.00000	14.00000	446.00000	Household income in thousands ($)
creddebt	498	1.59615	2.10489	0.90498	0.01484	16.03147	Credit card debt in thousands ($)

Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	creddebt
income	0.58605
Household income in thousands ($)	<.0001

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

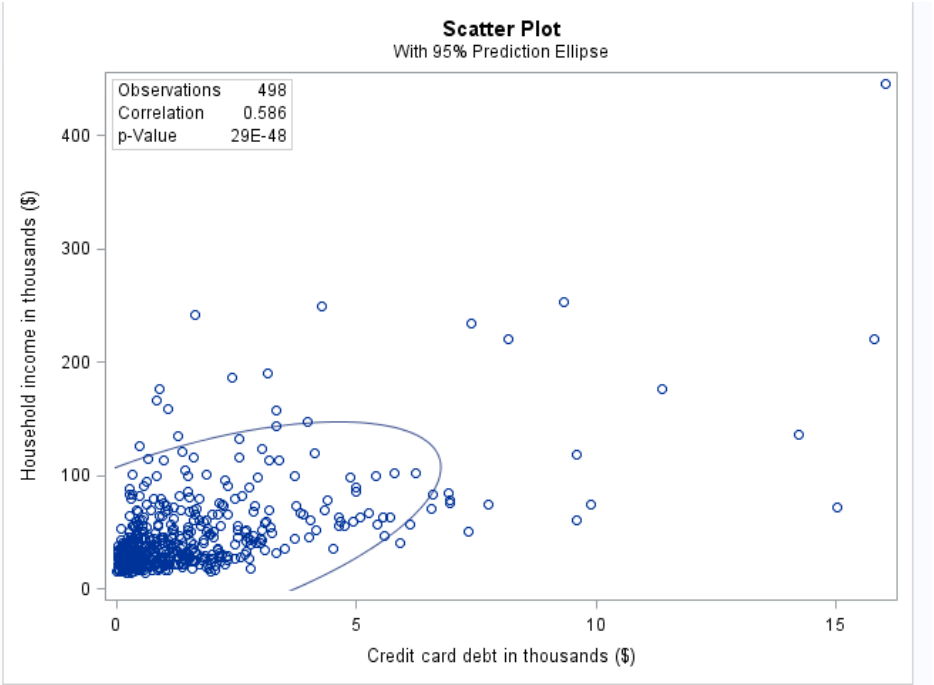
	creddebt
income	0.48674
Household income in thousands ($)	<.0001

Generated by the SAS System ('Local', X64_SRV12) on 12. tammikuuta 2017 at 1:05:03 AM

Statistic	DF	Value	Prob
Chi-Square	28355	34185.5866	<.0001
Likelihood Ratio Chi-Square	28355	3504.5054	1.0000
Mantel-Haenszel Chi-Square	1	170.6952	<.0001
Phi Coefficient		8.2853	
Contingency Coefficient		0.9928	
Cramer's V		0.8010	

WARNING: 100% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 498



APPENDIX 21

1 With Variables:	income
1 Variables:	default

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
income	498	48.13655	40.32830	36.00000	14.00000	446.00000	Household income in thousands ($)
default	498	0.29116	0.45476	0	0	1.00000	Previously loan defaulted

Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	default
income	-0.08457
Household income in thousands ($)	0.0593

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

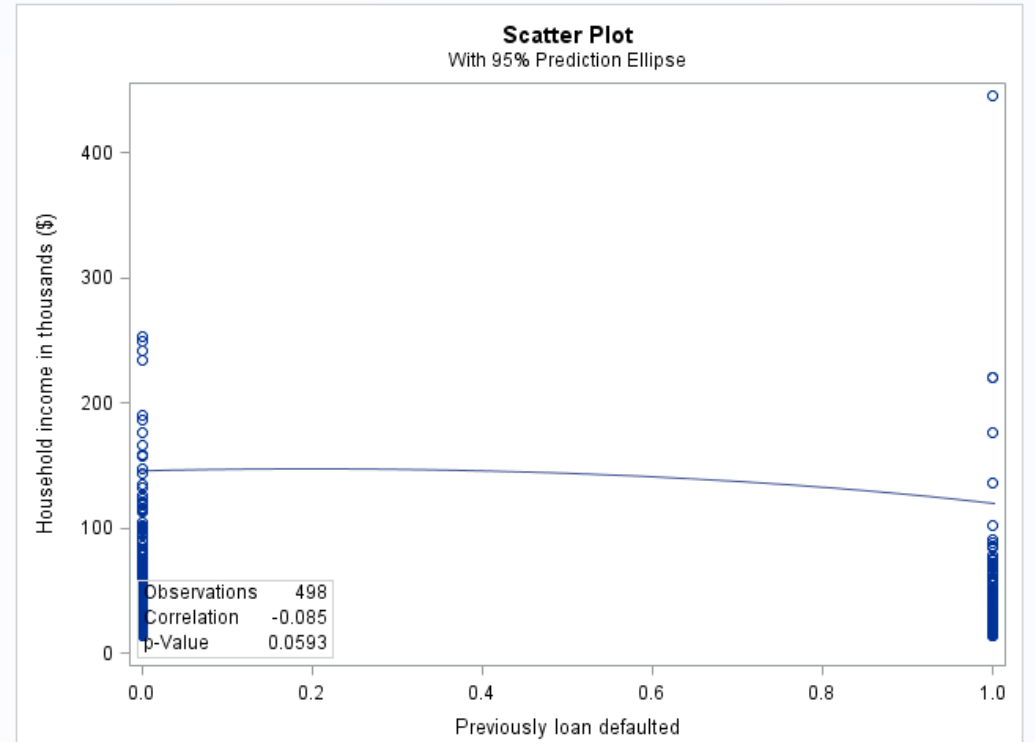
	default
income	-0.17667
Household income in thousands ($)	<.0001

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:37:07 PM

Statistic	DF	Value	Prob
Chi-Square	107	113.5321	0.3145
Likelihood Ratio Chi-Square	107	138.9160	0.0206
Mantel-Haenszel Chi-Square	1	3.5543	0.0594
Phi Coefficient		0.4775	
Contingency Coefficient		0.4309	
Cramer's V		0.4775	

WARNING: 88% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 498



APPENDIX 22

1 With Variables:	othdebt
1 Variables:	income

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
othdebt	498	3.24199	3.48503	2.10730	0.04558	27.03360	Other debt in thousands ($)
income	498	48.13655	40.32830	36.00000	14.00000	446.00000	Household income in thousands ($)

Pearson Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

	income
othdebt	0.60449
Other debt in thousands ($)	<.0001

Spearman Correlation Coefficients, N = 498
Prob > |r| under H0: Rho=0

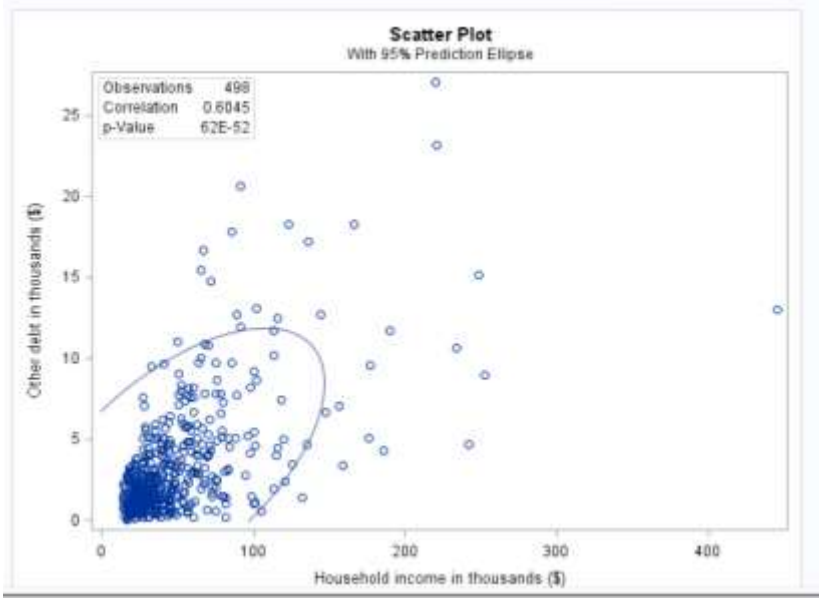
	income
othdebt	0.53640
Other debt in thousands ($)	<.0001

Generated by the SAS System ('Local', X64_SRV12) on 15. joulukuuta 2016 at 4:41:49 PM

Statistic	DF	Value	Prob
Chi-Square	36915	40720.6932	<.0001
Likelihood Ratio Chi-Square	36915	3838.3476	1.0000
Mantel-Haenszel Chi-Square	1	181.5897	<.0001
Phi Coefficient		9.0426	
Contingency Coefficient		0.9939	
Cramer's V		0.8742	

WARNING: 100% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Sample Size = 498



NAME: Onyebuchi Dalbert Zimuzochukwu
Student Number: Z109554 Email Address: Z109554@student.uwasa.fi