

## Table of Contents

<b>CHAPTER ONE</b> .....	<b>3</b>
<b>Introduction</b> .....	<b>3</b>
1.1 Introduction .....	3
1.2 Problem statement .....	3
1.3 Objectives.....	4
<b>CHAPTER TWO</b> .....	<b>4</b>
<b>Literature Review</b> .....	<b>4</b>
<b>CHAPTER THREE</b> .....	<b>5</b>
<b>Research Methodology</b> .....	<b>5</b>
3.1 Data source .....	5
3.2 Analysis.....	11
3.3 Bootstrap.....	12
3.4 Bootstrap confidence intervals .....	13
3.5 Simulation .....	14
<b>CHAPTER FOUR</b> .....	<b>15</b>
<b>Result and Analysis</b> .....	<b>15</b>
4.1 Logistic regression.....	15
4.2 Bootstrapping.....	16
4.3 Comparing Classical and Bootstrap Parameter Estimates.....	21
4.4 Simulation .....	22
<b>CHAPTER FIVE</b> .....	<b>24</b>
<b>Conclusions and Recommendation</b> .....	<b>24</b>
<b>References</b> .....	<b>24</b>
<b>Appendices</b> .....	<b>25</b>

# **CHAPTER ONE**

## **Introduction**

### **1.1 Introduction**

The economy of Germany is a highly developed social market economy. It has the largest national economy in Europe, the fourth-largest by nominal GDP in the world, and fifth by GDP. In 2017, the country accounted for 28% of the euro area economy according to the IMF. Germany is a founding member of the European Union and the Eurozone.

A credit risk is the risk of default on a debt that may arise from a borrower failing to make required payments. In the first resort, the risk is that of the lender and includes lost principal and interest, disruption to cash flows, and increased collection costs. The loss may be complete or partial. In an efficient market, higher levels of credit risk will be associated with higher borrowing costs. Because of this, measures of borrowing costs such as yield spreads can be used to infer credit risk levels based on assessments by market participants.

Bootstrapping is a practice that depends on random sampling with replacement. Bootstrap is often used when the distribution of a statistic is in doubt or unknown, or where the sample size requires complicated formulas for the calculation of standard errors. (Varian, 2005; Weisstein, n.d.). So, the bootstrap is a tool for making statistical inference when standard parametric assumptions are questionable. Bootstrapping is any test or metric that relies on random sampling with replacement. Bootstrapping allows assigning measures of accuracy defined in terms of bias, variance, confidence intervals, prediction error or some other such measure to sample estimates. This technique allows estimation of the sampling distribution of almost any statistic using random sampling methods. Generally, it falls in the broader class of resampling methods.

### **1.2 Problem statement**

Realizing the distribution of test statistic of random sample extracting from population provides clues as to which methods to be applied in analysing such data. Researchers generally must decide on the nature of the distribution of the sample obtained. A good assumption of the nature will lead to powerful test. Unfortunately, high price is paid if the assumption is wrong. It is vital to consider other methods for analysing data, and for this case bootstrap method was produced.

### **1.3 Objectives**

The aim of this project is to:

1. Fit a logistic regression model
2. To perform bootstrap on the coefficients of logistic regression
3. Construct 95% confidence intervals for the parameters of our bootstrap model
4. To perform simulations of coefficients
5. To calculate the predictive probabilities of 1000 simulations

## **CHAPTER TWO**

### **Literature Review**

According to Rochowicz (2010), bootstrapping is a numerical sampling technique where the data sampled are resampled with replacement. Various descriptive statistics such as mode, median, mean, variance and correlation can be bootstrapped. For our study, we will bootstrap the coefficients parameters. The advantages of using the bootstrapping analysis in this study is there is no need to determine the underlying sampling distribution for any population quantity and interpretations and results are based upon many observations. However, an error might occur when analyzing the nonparametric data by using bootstrapping analysis due to computer have built-in error (Rochowiz, 2010).

Bootstrap implicate inferring the variability in an unknown distribution from the drawn data that have been resampling is the basic idea of the bootstrap. The present exposition is more evaluative and promising than introductory by abstracting relevant bootstrap method into a framework that suits to covariance structure. Efron stated that bootstrapping method has been used and applied to many fields in statistics and a lot of research. This method has been diffused into behavioural sciences field although it in slow pace (Efron, 1988). The usefulness of the bootstrap apply to the correlation coefficient has been a hot issue in the psychological literature.

## CHAPTER THREE

### Research Methodology

#### 3.1 Data source

The dataset was obtained from Eberly College of Science. The German Credit Data is used to provide analysis for the minimization of risk and maximization of profit on behalf of the bank. The German Credit Data contains data on 20 variables and the classification whether an applicant is considered a Good or a Bad credit risk for 1000 loan applicants. Table 1 shows the description of the breast cancer dataset and Table 2 shows the attributes of the dataset.

**Table 1: Description of Credit dataset**

Variable	Description	Categories	Score
Creditability (y)	Creditability: 1: credit-worthy 0: not credit-worthy		
Account.Balance	Balance of current account IN "Deutsche Mark (DM)"	No balance or debit	2
		0 <= ... < 200 DM	3
		... >= 200 DM or checking account for at least 1 year	4
		No account	1
Duration.of.Credit..month	Duration in months (categorized)	<=6	10
		6 < and <= 12	9
		12 < and <= 18	8
		18 < and <= 24	7
		24 < and <= 30	6

		30 < and <= 36	5
		36 < and <= 42	4
		42 < and <= 48	3
		48 < and <= 54	2
		> 54	1
Payment.Status.of.Previous.Credit	Payment of previous credits	No previous credits / paid back all previous credits	2
		Paid back previous credits at this bank	4
		No problems with current credits at this bank	3
		Hesitant payment of previous credits	0
		Problematic running account / there are further credits running but at other banks	1
Purpose	Purpose of credit	new car	1
		used car	2
		items of furniture	3

		radio / television	4
		household appliances	5
		repair	6
		education	7
		vacation	8
		retraining	9
		business	10
		other	0
Credit.Amount	Amount of credit in DM (categorized)	<=500	10
		500 < and <= 1000	9
		1000 < and <= 1500	8
		1500 < and <= 2500	7
		2500 < and <= 5000	6
		5000 < and <= 7500	5
		7500 < and <= 10000	4
		10000 < and <= 15000	3
		15000 < and <= 20000	2
		> 20000	1

Value.Savings.Stocks	Value of savings or stocks	< 100,- DM	2
		100,- <= and < 500,- DM	3
		500,- <= .and< 1000,- DM	4
		>= 1000,- DM	5
		Not available / no savings	1
Length.of.current.employment	Employed by current employer for	unemployed	1
		<= 1 year	2
		1 <= and< 4 years	3
		4 <= and< 7 years	4
		>= 7 years	5
Instalment.per.cent	Instalment in % of available income	>= 35	1
		25 <= and < 35	2
		20 <= and< 25	3
		< 20	4
Sex...Marital.Status	Marital Status / Sex	Male: divorced / living apart	1
		Male: single	2
		Male: married / widowed	3
		Female:	4

Guarantors	Further debtors / Guarantors	None	1
		Co-Applicant	2
		Guarantor	3
Duration.in.Current.address	Living in current household for	< 1 year	1
		1 <= and < 4 years	2
		4 <= and < 7 years	3
		>= 7 years	4
Most.valuable.available.asset	Most valuable available assets	Ownership of house or land	4
		Savings contract with a building society / Life insurance	3
		Car / Other	2
		Not available / no assets	1
Age..years.	Age in years (categorized)	0 <= and <= 25	1
		26 <= and <= 39	2
		40 <= and <= 59	3
		60 <= and <= 64	5
		>= 65	4
Concurrent.Credits	Other running credits	Other banks	1

		At department store or mail order house	2
		No further running credits	3
Type.of.apartment	Type of apartment	Free apartment	1
		Rented flat	2
		Owner-occupied flat	3
No.of.Credits.at.this.Bank	Number of previous credits at this bank (including the running one)	One	1
		Two or three	2
		Four or five	3
		Six or more	4
Occupation	Occupation	Unemployed / unskilled with no permanent residence	1
		Unskilled with permanent residence	2
		Skilled worker / skilled employee / minor civil servant	3
		Executive / self-employed / higher civil servant	4
No.of.dependents	Number of persons entitled to maintenance	0 to 2	2
		3 and more	1

Telephone	Telephone	no	1
		yes	2
Foreign.Worker	Foreign worker	yes	1
		no	2

### 3.2 Analysis

Data, the summary and the description will be shown by using R-studio, we also will check if there are any missing values, which is shown by the codes below:

```
>file.choose()
> credit <- read.csv("/Users/zahier/Desktop/stock.csv", header = TRUE)
>attach(credit)
>str(credit)
>summary(credit)
>head(credit)
>tail(credit)
>anyNA(credit)
```

#### 3.2.1 Logistic regression model

The idea of using Logistic regression is to investigate the relationship between the independent variables and dichotomous dependent variable (Kleinbaum and Klein, 2010). Fitrianto and Cing (2014) stated that logistic regression is a popular statistical method to be used in modelling categorical dependent variables. To achieve the aim of the study, the binary logit model is employed. The choice of this statistical technique is based on the nature of the response variable (whether an applicant is credit-worthy or not credit-worthy).

1= Credit worthy

0= Not credit worthy

We will be using base function of `glm()` to fit the logistic regression. The parameters are to be estimated with 95% confidence intervals and a test of hypothesis will also be performed, in which the p-value will need to be lower than 0.05. Codes are shown below:

```
> logit<- glm(Creditability~., data = credit, family = binomial)

>significant1<- summary(logit)$coeff[-1,4] < 0.05
>relevant.1 <- names(significant1)[significant == TRUE]
```

We will be using the significant variables from “relevant.1” and fit into our new logistic model. Code is shown below:

```
logit2<- glm(Creditability~ I(Account.Balance==2)+ I(Account.Balance==3)+
  I(Account.Balance==4)+ Duration.of.Credit..month.+
  I(Payment.Status.of.Previous.Credit==2)+
I(Payment.Status.of.Previous.Credit==3)+
  I(Payment.Status.of.Previous.Credit==4)+ I(Purpose==1)+
  I(Purpose==2)+ I(Purpose==3)+ I(Purpose==9)+
  I(Purpose==10)+ Credit.Amount+ I(Value.Savings.Stocks==4)+
  I(Value.Savings.Stocks==5)+ I(Instalment.per.cent==4)+
  I(Sex...Marital.Status==3)+ I(Guarantors==3)+
  I(Duration.in.Current.address==2)+ Age..years.+
  I(Concurrent.Credits==3)+ I(Type.of.apartment==2)+
  I(Type.of.apartment==3)+ I(No.of.Credits.at.this.Bank==2)+
  I(Occupation==3)+ I(Occupation==2)+ I(Occupation==4)+
  I(No.of.dependents==2)+ I(Telephone==2)+ I(Foreign.worker==2),
  data = credit, family = binomial("logit"))
```

Process is repeated until we have all **significant variables** in a model. We also calculate the confidence interval of our final logistic model.

```
> summary(final_logit)
> confint(final_logit, level = 0.95) #calculate confidence interval
```

### 3.3 Bootstrap

To perform bootstrap, we will be using R package called ‘boot’. To perform using the boot package, we need to create a statistics function. The beauty of it is the statistics can be anything. For this study, we will create a function of our model’s coefficient. Next the statistics function will be applied to the ‘boot’ function with R equals to 1000 bootstrap replicates of the regression coefficients. The ‘boot’ function returns the sample from original values for each

component of the statistic and the bootstrap estimates of bias. Next, using 'plot' will produce a histogram and normal quantile- comparison plot of the bootstrap replications of each coefficient. Thus, the formula code is shown below:

```
> set.seed(123)

> boot.data <- function(data, indices, formula){
+   data <- data[indices,]
+   modelfit <- glm(formula, data=data, family = binomial("logit"))
+   coefficients(modelfit)}

> coef.boot <- boot(data=credit, statistic=boot.data, R=1000,
+                   formula= logit8$formula)

> plot(coef.boot, index = 2) #Account Balance category 2 coef
> plot(coef.boot, index = 2) #Account Balance category 2 coef
> plot(coef.boot, index = 3) #Account Balance category 3 coef
> plot(coef.boot, index = 4) #Account Balance category 4 coef
> plot(coef.boot, index = 5) #Duration of credit coef
> plot(coef.boot, index = 6) #Payment status of previous credit category 2
coef
> plot(coef.boot, index = 7) #Payment status of previous credit category 3
coef
> plot(coef.boot, index = 8) #Payment status of previous credit category 4
coef
> plot(coef.boot, index = 9) #Purpose category 1 coef
> plot(coef.boot, index = 10) #Purpose category 3 coef
> plot(coef.boot, index = 11) #Sex and Marital Status category 3 coef
> plot(coef.boot, index = 12) #Guarantors category 3 coef
> plot(coef.boot, index = 13) #Duration in current address category 2 coef
> plot(coef.boot, index = 14) #Type of apartment category 2
```

### 3.4 Bootstrap confidence intervals

For this study, we will use Normal-Theory interval to calculate our bootstrap confidence intervals. Normal theory interval assumes that statistic T is normally distributed and uses the estimate of bootstrapped sampling variance to construct a 100-(1-alpha) percent confidence interval. To obtain bootstrap confidence intervals, we will use boot.ci function. Code is shown below:

```
> boot.ci(coef.boot, index=2, type = c("norm", "perc", "bca") )
```

Boot.ci function is repeated for the other 13 independent variables.

### 3.5 Simulation

In this section, we would like to:

1. Fit a logistic model with only the variable Duration of Credit as explanatory variable
2. Run 1000 simulations of the coefficients; using b\_sim function from package of coreSim
3. Plot the intercept and the predictor of the coefficients
4. To calculate predicted probabilities from logistic regression models by creating a function.

```
> set.seed(123)
> logit_duration<- glm(Creditability~Duration.of.Credit..month.,
+                      family = binomial("logit"),
+                      data=credit)

> library(coreSim)
> sim1<- b_sim(logit_duration, nsim = 1000) #run 1000 simulations

> plot(sim1, xlab="Intercept", ylab= "Duration")
> points(coef(logit_duration), col=2)

> pred_prob <- function(x){                      #predicted probability
+   1 / (1 + exp(-x))
+ }

> pred_prob(coef(logit_duration))                #pred prob of glm model
> summary(pred_prob(sim1))                      #pred prob of 1000 sims
```

## CHAPTER FOUR

### Result and Analysis

#### 4.1 Logistic regression

Our goal here is to create a logistic regression model where all variables are significant. The summary of the logistic model is shown below while table 2 shows all the variables that were included in the final model:

```

Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                   -0.707364   0.331297  -2.135 0.032750 *
I(Account.Balance == 2)TRUE      0.516398   0.192611   2.681 0.007339 **
I(Account.Balance == 3)TRUE      1.160238   0.343547   3.377 0.000732 ***
I(Account.Balance == 4)TRUE      1.884966   0.213086   8.846 < 2e-16 ***
Duration.of.Credit..month.     -0.039694   0.006716  -5.910 3.41e-09 ***
I(Payment.Status.of.Previous.Credit == 2)TRUE  0.860972   0.269435   3.195 0.001396 **
I(Payment.Status.of.Previous.Credit == 3)TRUE  0.971918   0.359956   2.700 0.006932 **
I(Payment.Status.of.Previous.Credit == 4)TRUE  1.418825   0.299150   4.743 2.11e-06 ***
I(Purpose == 1)TRUE               1.224486   0.316320   3.871 0.000108 ***
I(Purpose == 3)TRUE               0.432245   0.190764   2.266 0.023460 *
I(Sex...Marital.Status == 3)TRUE  0.412176   0.162863   2.531 0.011380 *
I(Guarantors == 3)TRUE          1.027863   0.405112   2.537 0.011173 *
I(Duration.in.Current.address == 2)TRUE  -0.449322   0.176255  -2.549 0.010795 *
I(Type.of.apartment == 2)TRUE     0.492079   0.175864   2.798 0.005141 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1221.73  on 999  degrees of freedom
Residual deviance:  975.78  on 986  degrees of freedom
AIC: 1003.8

```

*Figure 1: Summary of final logistic regression model*

**Table 2: Significant variables**

Variables	Category
Account.Balance	2
	3
	4

Duration.of.Credit..month.	
Payment.Status.of.Previous.Credit	2
	3
	4
Purpose	1
	3
Sex...Marital.Status	3
Guarantors	3
Duration.in.Current.address	2
Type.of.apartment	2

## 4.2 Bootstrapping

In this section, we will be applying bootstrapping technique to approximate the distribution. To perform bootstrap, we created a function that returns the statistic. We performed the ‘boot’ with, R equals to 1000 bootstrap replicates of the regression coefficients. ‘Coef.boot’ returns the sample from original values for each component of the statistic and also the bootstrap estimates of bias.

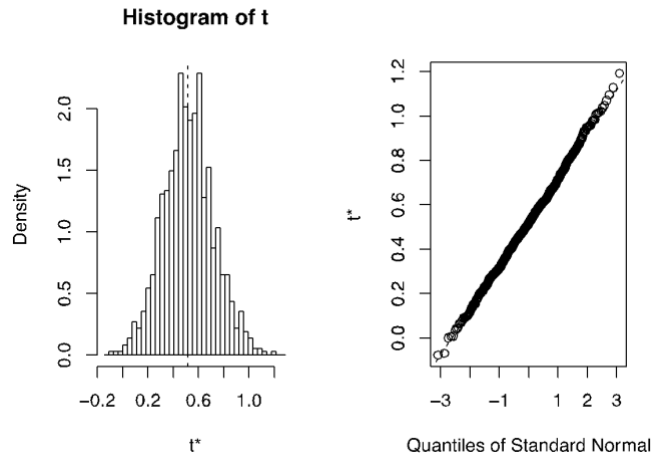


Figure 2: Account Balance category 2 coefficient

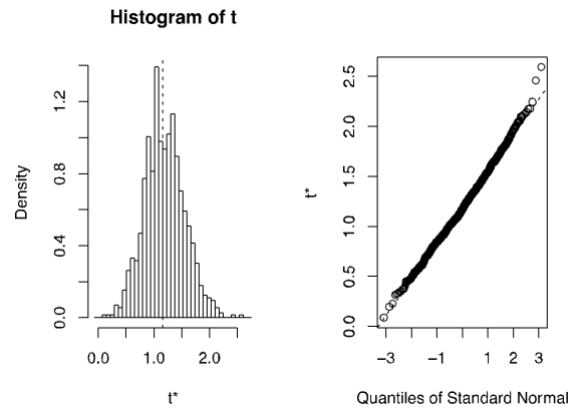


Figure 3: Account Balance category 3 coefficient

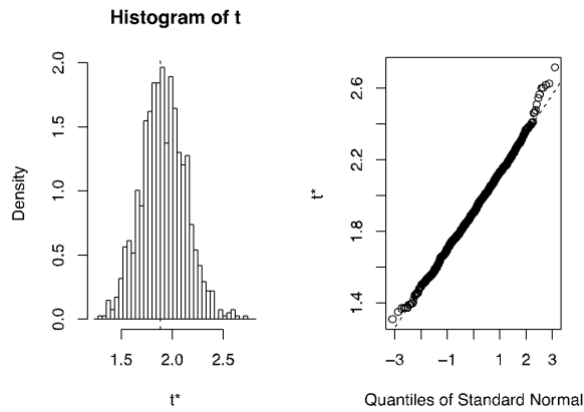


Figure 4: Account Balance category 4 coefficient

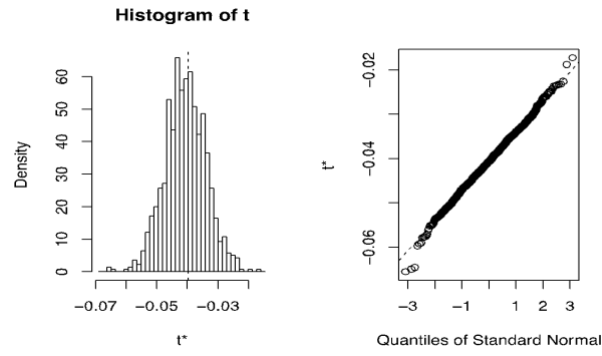


Figure 5: Duration of credit coefficient

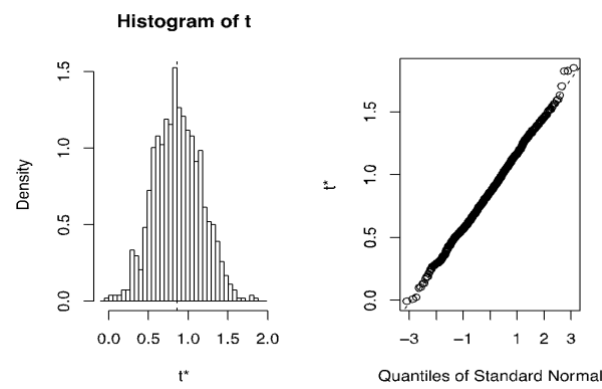


Figure 6: Payment status of previous credit category 2 coefficient

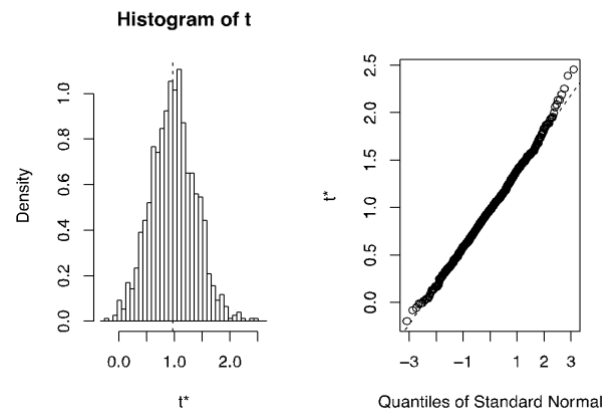


Figure 7: Payment status of previous credit category 3 coefficient

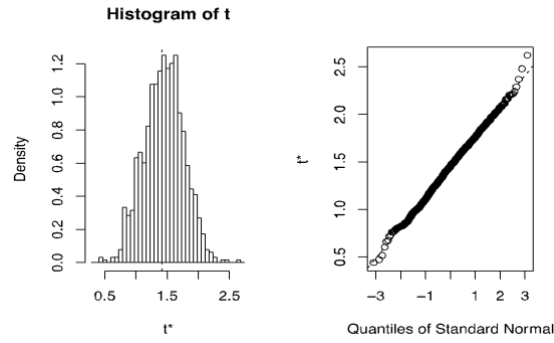


Figure 8: Payment status of previous credit category 4 coef

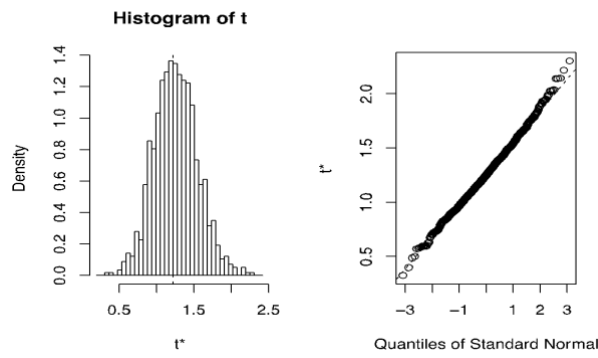


Figure 9: Purpose category 1 coefficient

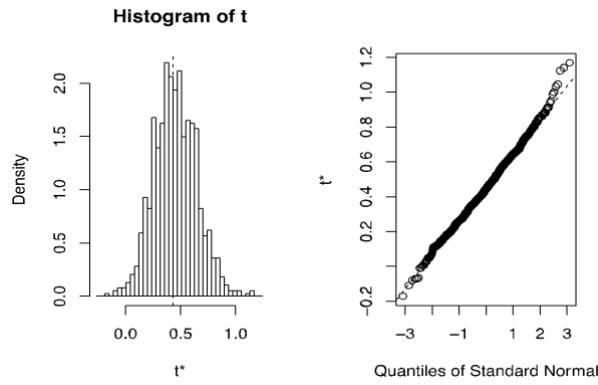


Figure 10: Purpose category 3 coefficient

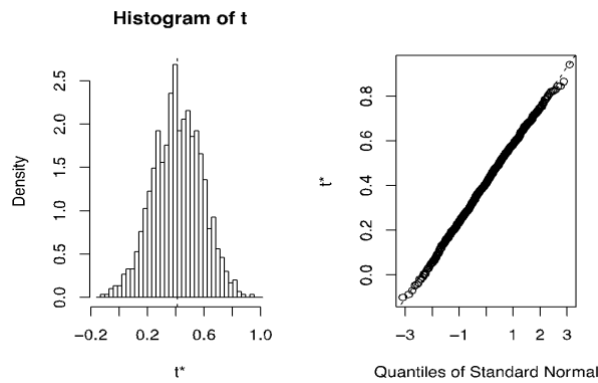


Figure 11: Sex and Marital Status category 3 coefficient

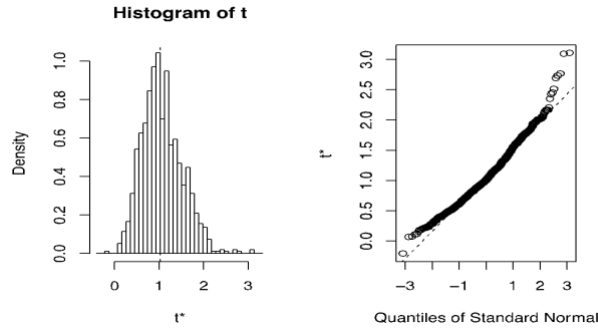


Figure 12: Guarantors category 3 coefficient

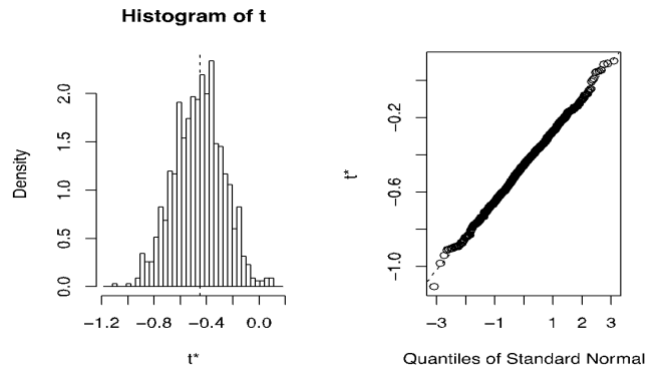


Figure 13: Duration in current address category 2 coefficient

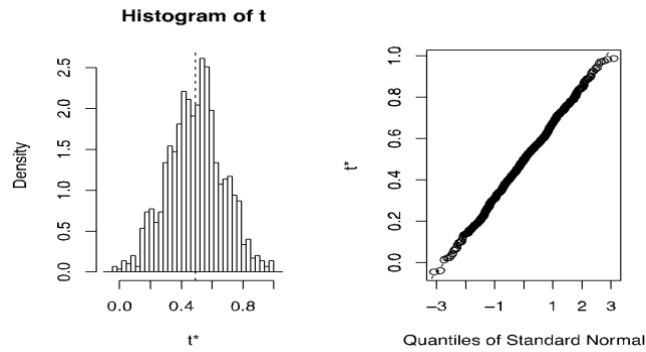


Figure 14: Type of apartment category 2

Based on Figure 2 until Figure 14, The bootstrap distributions for all coefficients are roughly symmetric. The broken line in each histogram displays the location of the regression coefficient for the model fit to the original sample. On the other hand, from Quantiles of Standard Normal of Q-Q plots, we can safely assume that sample of every variable comes from a population that is normally distributed.

### 4.3 Comparing Classical and Bootstrap Parameter Estimates.

Based on Table 3 and Table 4, it can be observed that the original and bootstrap parameter estimates are very close. The standard errors of estimates for bootstrap are slightly higher than the values obtained from original logistic model.

	Estimate	Std. Error	95% C. I	
(Intercept)	-0.7074	0.3313	-1.3641	-0.0632
Account.Balance = 2	0.5164	0.1926	0.1404	0.8961
Account.Balance = 3	1.1602	0.3435	0.5094	1.8640
Account.Balance = 4	1.8850	0.2131	1.4740	2.3105
Duration.of.Credit..month.	-0.0397	0.0067	-0.0530	-0.0267
Payment.Status.of.Previous.Credit = 2	0.8610	0.2694	0.3370	1.3956
Payment.Status.of.Previous.Credit = 3	0.9719	0.3600	0.2740	1.6876
Payment.Status.of.Previous.Credit = 4	1.4188	0.2992	0.8377	2.0124
Purpose = 1	1.2245	0.3163	0.6262	1.8716
Purpose = 3	0.4322	0.1908	0.0616	0.8104
Sex.Marital.Status = 3	0.4122	0.1629	0.0939	0.7328
Guarantors = 3	1.0279	0.4051	0.2709	1.8719
Duration.in.Current.address = 2	-0.4493	0.1763	-0.7955	-0.1038
Type.of.apartment = 2	0.4921	0.1759	0.1468	0.8368

*Table 3: Parameter estimates of logistic model*

	Estimate	Std. Error	95% C. I	
(Intercept)	-0.7031	0.3316	-1.3615	-0.0618
Account.Balance = 2	0.5214	0.1987	0.1220	0.9009
Account.Balance = 3	1.1925	0.3606	0.4210	1.8350
Account.Balance = 4	1.9140	0.2170	1.4310	2.2810
Duration.of.Credit..month.	-0.0406	0.0067	-0.0520	-0.0256
Payment.Status.of.Previous.Credit = 2	0.8707	0.2968	0.2695	1.4330
Payment.Status.of.Previous.Credit = 3	0.9814	0.4021	0.1744	1.7505
Payment.Status.of.Previous.Credit = 4	1.4495	0.3195	0.7620	2.0140
Purpose = 1	1.2608	0.2913	0.6170	1.7590
Purpose = 3	0.4536	0.1936	0.0315	0.7902

Sex.Marital.Status = 3	0.4131	0.1723	0.0737	0.7489
Guarantors = 3	1.0717	0.4531	0.0960	1.8720
Duration.in.Current.address = 2	-0.4595	0.1863	-0.8043	-0.0740
Type.of.apartment = 2	0.4928	0.1793	0.1398	0.8428

*Table 4: Confidence intervals by bootstrapping*

Comparing the length of confidence intervals for original and bootstrap methods, it was observed that the original and bootstrap methods have almost similar interval length. However, it was observed that the interval length of variables Purpose=3 and Guarantors=3 are wider to values obtained from original logistic model.

None of the confidence intervals are going to be the exact as the actual confidence interval is exactly equal to the nominal confidence level (1-alpha). However, as long as the estimator is consistent, all the intervals will also be consistent, which means that the confidence level approaches 1-alpha as the sample size gets larger.

#### 4.4 Simulation

Figure 15 displays the plot coefficients intercept and duration of credit of simulations and original logistic regression model (logit\_duration). The plot shows that most of the coefficients of the simulations are on the range of 1.45 and 1.75 (intercept) while -0.03 and -0.045 for the variable duration of credit. Meanwhile the red point shows the original data with intercept of 1.67 and duration of credit of -0.04. Table 5 and 6 meanwhile shows the comparisons of predicted probabilities of the simulations and original logistic model (logit\_duration). It can be observed that mean predicted probabilities of 1000 simulations are almost identical with the original model.

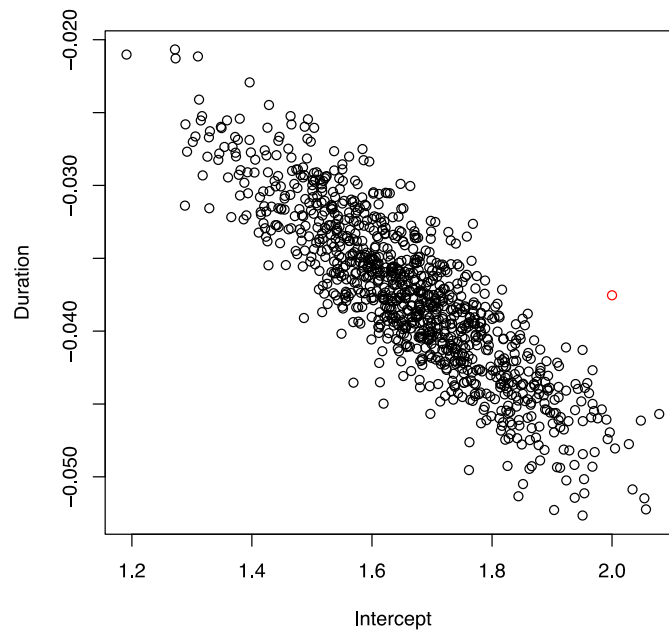


Figure 15: Uncertainty plot

Logit_duration	
Intercept	Duration
0.8410888	0.4906167

Table 5: Predicted probabilities of logit\_duration

1000 simulations	
Intercept	Duration
Min: 0.7670	Min: 0.4868
Mean: 0.8398	Mean: 0.4906
Max: 0.8888	Max: 0.4948

Table 6: Predicted probabilities of 1000 simulations

## **CHAPTER FIVE**

### **Conclusions and Recommendation**

#### **5.1 Conclusion**

The bootstrap technique was used on the coefficients from a logistic model. Confidence intervals (C.I) were also computed, both on the original and bootstrapped logistic models. Most of the results of the estimates and C.I were similar between of the original and bootstrapped. The discussion in this study concluded the sample used in bootstrapping is normally distributed. For simulation, we run 1000 simulations of coefficient a logistic regression of based on duration of credit. We also compare the coefficients between the original and the simulations using a plot. Lastly, we computed the predicted probabilities which the results show almost similar values between the original and simulations.

#### **5.2 Recommendation**

Based on the study, bootstrap can provide correct statistical inference and to be used for forecasting for future researches. Regarding simulation, many complex statistical methods can be applied; for example, to find the relationship between gender of category 2 and payment status of previous credit of category 4 by simulating the original logistic regression.

### **References**

1. Varian, H. (2005). "Bootstrap Tutorial". Mathematical Journal, 9, 768-775.
2. Weisstein, E. W. Bootstrap. Available at:  
<http://mathworld.wolfram.com/BootstrapMethods.html>
3. Kleinbaum, D.G., Klein, M. (2002). Logistic Regression: A Self- learning Text. Springer, New York.
4. Rochowicz, J. A. (2010). Bootstrapping Analysis, Inferential Statistics and EXCEL. Spreadsheets in Education (eJSiE), Vol. 4: Iss. 3, Article 4.

## Appendices

### *Summary of logistic model with all variables*

	Estimate	Std. Error	z	value	Pr(> z )
(Intercept)	1.264e+01	8.836e+02	0.014	0.988584	
Account.Balance2	4.980e-01	2.296e-01	2.168	0.030124	
*					
Account.Balance3	9.324e-01	3.791e-01	2.460	0.013899	
*					
Account.Balance4	1.807e+00	2.444e-01	7.395	1.41e-13	
***					
Duration.of.Credit..month.5	-8.742e-01	2.557e+03	0.000	0.999727	
Duration.of.Credit..month.6	-1.262e+01	8.836e+02	-0.014	0.988603	
Duration.of.Credit..month.7	5.130e-02	1.358e+03	0.000	0.999970	
Duration.of.Credit..month.8	-1.311e+01	8.836e+02	-0.015	0.988159	
Duration.of.Credit..month.9	-1.387e+01	8.836e+02	-0.016	0.987476	
Duration.of.Credit..month.10	-1.293e+01	8.836e+02	-0.015	0.988323	
Duration.of.Credit..month.11	1.776e+00	1.137e+03	0.002	0.998754	
Duration.of.Credit..month.12	-1.369e+01	8.836e+02	-0.015	0.987642	
Duration.of.Credit..month.13	1.467e+00	1.427e+03	0.001	0.999180	
Duration.of.Credit..month.14	-1.293e+01	8.836e+02	-0.015	0.988322	
Duration.of.Credit..month.15	-1.303e+01	8.836e+02	-0.015	0.988238	
Duration.of.Credit..month.16	-1.676e+01	8.836e+02	-0.019	0.984867	
Duration.of.Credit..month.18	-1.408e+01	8.836e+02	-0.016	0.987284	
Duration.of.Credit..month.20	-1.293e+01	8.836e+02	-0.015	0.988326	
Duration.of.Credit..month.21	-1.378e+01	8.836e+02	-0.016	0.987561	
Duration.of.Credit..month.22	7.446e-01	1.843e+03	0.000	0.999678	
Duration.of.Credit..month.24	-1.379e+01	8.836e+02	-0.016	0.987550	
Duration.of.Credit..month.26	2.341e+00	2.557e+03	0.001	0.999269	
Duration.of.Credit..month.27	-1.398e+01	8.836e+02	-0.016	0.987380	
Duration.of.Credit..month.28	-1.400e+01	8.836e+02	-0.016	0.987363	
Duration.of.Credit..month.30	-1.369e+01	8.836e+02	-0.015	0.987640	
Duration.of.Credit..month.33	-1.497e+01	8.836e+02	-0.017	0.986483	
Duration.of.Credit..month.36	-1.416e+01	8.836e+02	-0.016	0.987211	
Duration.of.Credit..month.39	-1.341e+01	8.836e+02	-0.015	0.987896	
Duration.of.Credit..month.40	-2.924e+01	2.557e+03	-0.011	0.990877	
Duration.of.Credit..month.42	-1.334e+01	8.836e+02	-0.015	0.987957	
Duration.of.Credit..month.45	-1.609e+01	8.836e+02	-0.018	0.985470	
Duration.of.Credit..month.47	4.035e+00	2.557e+03	0.002	0.998741	
Duration.of.Credit..month.48	-1.468e+01	8.836e+02	-0.017	0.986741	
Duration.of.Credit..month.54	-1.345e+01	8.836e+02	-0.015	0.987855	
Duration.of.Credit..month.60	-1.368e+01	8.836e+02	-0.015	0.987652	
Duration.of.Credit..month.72	-3.080e+01	2.557e+03	-0.012	0.990390	
Payment.Status.of.Previous.Credit1	-2.640e-01	5.942e-01	-0.444	0.656755	
Payment.Status.of.Previous.Credit2	6.123e-01	4.704e-01	1.302	0.192980	
Payment.Status.of.Previous.Credit3	9.162e-01	5.044e-01	1.816	0.069317	
.					
Payment.Status.of.Previous.Credit4	1.529e+00	4.709e-01	3.247	0.001165	
**					
Purpose1	1.709e+00	3.979e-01	4.295	1.75e-05	
***					
Purpose2	8.517e-01	2.751e-01	3.096	0.001958	
**					
Purpose3	9.122e-01	2.575e-01	3.542	0.000397	
***					
Purpose4	4.325e-01	7.893e-01	0.548	0.583667	
Purpose5	8.669e-02	5.700e-01	0.152	0.879120	
Purpose6	-5.966e-02	4.070e-01	-0.147	0.883462	
Purpose8	2.132e+00	1.189e+00	1.793	0.072901	
.					
Purpose9	7.345e-01	3.493e-01	2.103	0.035463	
*					

Purpose10	1.432e+00	8.211e-01	1.744	0.081228
•				
Credit.Amount	-1.537e-04	4.778e-05	-3.217	0.001294
**				
Value.Savings.Stocks2	3.370e-01	3.050e-01	1.105	0.269106
Value.Savings.Stocks3	2.869e-01	4.178e-01	0.687	0.492213
Value.Savings.Stocks4	1.337e+00	5.446e-01	2.455	0.014083
*				
Value.Savings.Stocks5	9.245e-01	2.718e-01	3.402	0.000669
***				
Length.of.current.employment2	-1.327e-01	4.501e-01	-0.295	0.768072
Length.of.current.employment3	3.300e-01	4.299e-01	0.768	0.442705
Length.of.current.employment4	7.971e-01	4.711e-01	1.692	0.090672
•				
Length.of.current.employment5	2.049e-01	4.334e-01	0.473	0.636317
Instalment.per.cent2	-1.935e-01	3.264e-01	-0.593	0.553235
Instalment.per.cent3	-5.056e-01	3.587e-01	-1.409	0.158697
Instalment.per.cent4	-8.244e-01	3.212e-01	-2.567	0.010270
*				
Sex...Marital.Status2	2.973e-01	4.006e-01	0.742	0.458066
Sex...Marital.Status3	8.727e-01	3.924e-01	2.224	0.026156
*				
Sex...Marital.Status4	4.065e-01	4.691e-01	0.867	0.386192
Guarantors2	-3.490e-01	4.209e-01	-0.829	0.406928
Guarantors3	9.525e-01	4.542e-01	2.097	0.035971
*				
Duration.in.Current.address2	-7.400e-01	3.081e-01	-2.402	0.016297
*				
Duration.in.Current.address3	-4.797e-01	3.460e-01	-1.386	0.165621
Duration.in.Current.address4	-3.820e-01	3.148e-01	-1.213	0.225045
Most.valuable.available.asset2	-3.117e-01	2.674e-01	-1.166	0.243737
Most.valuable.available.asset3	-8.984e-02	2.457e-01	-0.366	0.714678
Most.valuable.available.asset4	-7.587e-01	4.339e-01	-1.748	0.080413
•				
Age..years.	1.115e-02	9.623e-03	1.158	0.246709
Concurrent.Credits2	-1.984e-02	4.371e-01	-0.045	0.963804
Concurrent.Credits3	3.786e-01	2.600e-01	1.456	0.145317
Type.of.apartment2	4.607e-01	2.474e-01	1.862	0.062564
•				
Type.of.apartment3	7.436e-01	4.962e-01	1.498	0.134013
No.of.Credits.at.this.Bank2	-4.333e-01	2.559e-01	-1.693	0.090381
•				
No.of.Credits.at.this.Bank3	-9.084e-02	6.282e-01	-0.145	0.885036
No.of.Credits.at.this.Bank4	-2.301e-01	1.104e+00	-0.208	0.834907
Occupation2	-7.793e-01	6.946e-01	-1.122	0.261862
Occupation3	-7.307e-01	6.656e-01	-1.098	0.272343
Occupation4	-6.616e-01	6.818e-01	-0.970	0.331847
No.of.dependents2	-3.066e-01	2.605e-01	-1.177	0.239232
Telephone2	3.086e-01	2.123e-01	1.454	0.146049
Foreign.Worker2	1.474e+00	6.888e-01	2.140	0.032346
*				

## Summary for our new model

```

Call:
glm(formula = Creditability ~ I(Account.Balance == 2) + I(Account.Balance ==
==
3) + I(Account.Balance == 4) + Duration.of.Credit..month. +
I(Payment.Status.of.Previous.Credit == 2) +
I(Payment.Status.of.Previous.Credit ==
3) + I(Payment.Status.of.Previous.Credit == 4) + I(Purpose ==
1) + I(Purpose == 3) + I(Sex...Marital.Status == 3) + I(Guarantors ==
3) + I(Duration.in.Current.address == 2) + I(Type.of.apartment ==
2), family = binomial("logit"), data = credit)

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.6838	-0.8619	0.4452	0.7521	1.9258

Coefficients:

	Estimate	Std. Error	z value
Pr(> z )			
(Intercept)	-0.707364	0.331297	-2.135
0.032750 *			
I(Account.Balance == 2)TRUE	0.516398	0.192611	2.681
0.007339 **			
I(Account.Balance == 3)TRUE	1.160238	0.343547	3.377
0.000732 ***			
I(Account.Balance == 4)TRUE	1.884966	0.213086	8.846
< 2e-16 ***			
Duration.of.Credit..month.	-0.039694	0.006716	-5.910
3.41e-09 ***			
I(Payment.Status.of.Previous.Credit == 2)TRUE	0.860972	0.269435	3.195
0.001396 **			
I(Payment.Status.of.Previous.Credit == 3)TRUE	0.971918	0.359956	2.700
0.006932 **			
I(Payment.Status.of.Previous.Credit == 4)TRUE	1.418825	0.299150	4.743
2.11e-06 ***			
I(Purpose == 1)TRUE	1.224486	0.316320	3.871
0.000108 ***			
I(Purpose == 3)TRUE	0.432245	0.190764	2.266
0.023460 *			
I(Sex...Marital.Status == 3)TRUE	0.412176	0.162863	2.531
0.011380 *			
I(Guarantors == 3)TRUE	1.027863	0.405112	2.537
0.011173 *			
I(Duration.in.Current.address == 2)TRUE	-0.449322	0.176255	-2.549
0.010795 *			
I(Type.of.apartment == 2)TRUE	0.492079	0.175864	2.798
0.005141 **			

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1221.73 on 999 degrees of freedom  
Residual deviance: 975.78 on 986 degrees of freedom  
AIC: 1003.8

Number of Fisher Scoring iterations: 5

### Confidence intervals of bootstrapping

```
boot.ci(coef.boot, index=1, type="norm")
boot.ci(coef.boot, index=3, type="norm")
boot.ci(coef.boot, index=4, type="norm")
boot.ci(coef.boot, index=5, type="norm")
boot.ci(coef.boot, index=6, type="norm")
boot.ci(coef.boot, index=7, type="norm")
boot.ci(coef.boot, index=8, type="norm")
boot.ci(coef.boot, index=9, type="norm")
boot.ci(coef.boot, index=10, type="norm")
boot.ci(coef.boot, index=11, type="norm")
boot.ci(coef.boot, index=12, type="norm")
boot.ci(coef.boot, index=13, type="norm")
boot.ci(coef.boot, index=14, type="nor
```

