

# Preface

(အမှာစာ)

ယနေ့ခေတ်သည် ကိန်းဂဏန်းအချက်အလက်များအားသုံး၍ စနစ်တကျလေ့လာစိစစ်ပြီး အထောက်အထားနှင့်ဘာကြော့အချိန်ချေးထားတင်ပြလာကြသည်။ ထိုကြောင့် စာရင်းအင်းပညာ နည်းလမ်းတစ်ခုဖြစ်သော အချက်အလက်များအား လေ့လာစိစစ်ပတ်ဝန်းကျင်တွင် Exploratory data analysis သည်လည်း အရေးပါလာသည်။ အသုံးများလာသော အချိန်နည်းလမ်းသည် သုတေသန နယ်ပယ်အသီးသီးတွင်လည်းကောင်း၊ data science, data mining စသည့် ဘာသာရပ်နယ်ပယ်ထိ များအဖြစ်လည်းကောင်း တွင်ကျယ်စွာ အသုံးပြုလာကြသည်။ ကျယ်ဝန်းသော လေ့လာမှုနယ်ပယ် ဖြစ်သည်နှင့်အညီ အသုံးပြုသည့် စာရင်းအင်းပညာနည်းလမ်းများမှာလည်း ရေပန်းစားမှုများစွာ ရှိလာခဲ့ရာအလိုက် အနည်းငယ်ကွဲပြားမှုရှိကြသည်။ ဤစာအုပ်သည် သုတေသနပြုသူများအတွက် လက်တွေ့တွင်အသုံးများသော စာရင်းအင်းပညာ နည်းလမ်းများပေါင်းစုံထားသည့် Exploratory data analysis ကို ဖြည့်စွက်အသွယ်တက်အတွက် လေ့လာနိုင်စေရန် တင်ပြထားပါသည်။

အခန်း(၁)သည် R အသုံးပြုရာတွင် ပိုမိုလွယ်ကူစေရန် R ကို integrated ပြုလုပ်ထားသည့် RStudio ကိုဦးစွာတင်ပြထားသည်။ မိမိနှစ်သက်ရာ R သို့မဟုတ် RStudio ကိုအသုံးပြု၍စာတမ်း များ အစီရင်ခံစာများရေးသားရာတွင်လွန်စွာအရေးပါသော frequency table, cross-table စသည့် အဖွဲ့အစည်းမျိုးလွယ်ကူစွာတင်ဆောင်ပုံကိုအခန်း(၂)တွင်တင်ပြထားပြီး HTML format အနေဖြင့် အရည်အသွေးမြင့်စွာ ထုတ်ယူနိုင်ပုံကိုပါ တင်ပြထားပါသည်။ အခန်း(၃)သည် ကိန်းရှင်တစ်လုံးတည်း ကို စူးစမ်းလေ့လာသောနည်းလမ်းများကိုဥပမာများနှင့်တင်ပြထားပြီး တွက်ချက်ရယူသည့် statistics များကို ဝေဖန်၍ စနစ်တကျဖွဲ့စည်းထုတ်ယူနိုင်ပုံကို တင်ပြထားပါသည်။ ၎င်းအပြင် ကိန်းရှင်တစ်ခု သည် အမျိုးသော စာရင်းအင်းပညာနည်းလမ်းများအတွက် လိုအပ်ချက်ဖြစ်သော ခေတ်မယ်ပြန်ချက် မရှိပါက ပြုပြင်ပြောင်းလဲနိုင်သော Turkey's Ladder Transformation နည်းကိုပါထည့်သွင်းတင်ပြ ထားသည်။ အခန်း(၄)သည် numerical variables များဖြစ်သော ဦးကိန်းရှင် (bivariate data)များ အတွက် ဆက်စပ်မှုပြုကိန်းအမျိုးမျိုးတွက်ချက်မှုနှင့်အတူ cross-table ဝေဖန်မှုများနှင့်ဆက်စပ်လျက် ရှိသော  $\chi^2$  စစ်ဆေးကိန်းနှင့် အဆိုပါစစ်ဆေးကိန်းမှတွက်ယူသည့် categorical variables များ အတွက် ဆက်စပ်မှုပြုကိန်းတို့တွက်ယူနိုင်ပုံကို တင်ပြထားသည်။ အခန်း(၅)သည် multivariate data များဆက်စပ်မှုကို စူးစမ်းလေ့လာနိုင်သည့် နည်းများဖြစ်သော correlation matrix, partial correlation, canonical correlation တို့အပြင် data reduction နည်းလမ်းများ ဖြစ်သည့် principal component method နှင့် factor analysis တို့ကိုပါတင်ပြထားသည်။ အခန်းတိုင်းတွင် statistics များနှင့်တွက်ချက်လေ့လာခြင်းအပြင် graphic နည်းလမ်းများကိုပါအသုံးပြု၍ လေ့လာ စိစစ် နိုင်ပုံကို တင်ပြထားသည်။

ပေါက်တာဦးဝင်း, PhD  
 စာရင်းအင်းပညာ  
 ရက်စွဲ၊ ၉-၉-၂၀၁၉

# Table of Contents

Title	Page
<b>Acknowledgement</b>	
<b>Preface</b>	
<b>Chapter 1</b>	
<b>Introduction to RStudio</b>	1
1.1 Introduction	1
1.2 Installation of RStudio	1
1.3 Opening and Closing RStudio	2
1.4 Executing R Script in RStudio	6
1.5 Saving R Files	7
1.6 Viewer Window	9
1.6.1 Data viewer	9
1.6.2 Viewer Tab	11
1.7 Importing Data into R	12
1.8 Packages for Exploratory Data Analysis (EDA)	14
<b>Chapter 2</b>	
<b>Frequency Table and Cross-Tabulation</b>	15
2.1 Introduction	15
2.2 Reviewing the Data	15
2.3 Generating Frequency Tables	18
2.3.1 Simple Frequency Table	19
2.3.2 Frequency Tables with Grouped Variable	20
2.3.3 Generating Frequency Tables at Once	21
2.3.4 Generating Frequency Table of Numeric Variable	22
2.4 Generating Cross Tables	23
2.4.1 Simple Cross-Table	23
2.4.2 Cross-Table with Grouped Variable	25
2.5 HTML Format Table	26
<b>Chapter 3</b>	
<b>Univariate Exploratory Data Analysis</b>	28
3.1 Introduction	28
3.2 Univariate Summary Statistics	28
3.2.1 Measures of Central Tendency	28
3.2.2 Measures of Variation	29
3.2.3 Measures of Shape of Distribution	32
3.2.4 Computing Univariate Statistics	33
3.2.5 Transposing, Selecting Univariate Statistics	36
3.2.6 Computing Univariate Statistics with Grouped Variable	37

3.2.7 Data Frame Summaries	38
3.3 Computing Summary Statistics with Sampling Weights	39
3.4 Univariate Visual Presentation	41
3.4.1 Histogram and Normal Curve	41
3.4.2 Density Plot	44
3.4.3 Boxplot	45
3.4.4 Normal Probability Plot	47
3.5 Transformation for Normal	49
<b>Chapter 4</b>	
<b>Bivariate Exploratory Data Analysis</b>	53
4.1 Introduction	53
4.2 Correlation for Numerical/Ordinal Variables	53
4.2.1 Pearson Correlation Coefficient	53
4.2.2 Spearman's Rank Correlation Coefficient	55
4.3 Cross Tabulation and Test of Independence	56
4.3.1 Pearson $\chi^2$ Test	56
4.3.2 Likelihood Ratio $\chi^2$ Test	58
4.4 Correlation for Nominal Variables	58
4.4.1 Contingency Correlation Coefficient	58
4.4.2 Phi Coefficient	59
4.4.3 Cramer's V Coefficient	59
4.5 Graphical Bivariate EDA	61
4.5.1 Scatter Plot	61
4.5.2 Boxplot	62
<b>Chapter 5</b>	
<b>Multivariate Exploratory Data Analysis</b>	64
5.1 Introduction	64
5.2 Correlation of Multiple Variables	64
5.2.1 Correlation Matrix	64
5.2.2 Partial Correlation Matrix	66
5.2.3 Canonical Correlation Coefficient	68
5.3 Multivariate Data Reduction Techniques	71
5.3.1 Principal Component Analysis	71
5.3.2 Exploratory Factor Analysis	78
5.4 Graphical Multivariate EDA	86
5.4.1 Scatter Plot for Multiple Variables	86
5.4.2 Boxplot for Multiple Variables	87
<b>References and Resources</b>	89