Table of Contents

Chapter Title Page

စာရေးသူ၏ အမှာစာ VII

1 Introduction to R 1
- 1.1 What is R? 1
- 1.2 Obtaining and Installing R 1
- 1.3 Working with R 2
  - 1.3.1 The Command Prompt 2
  - 1.3.2 The Menu Bar 3
  - 1.3.3 The Toolbar 4
- 1.4 Variable Names and Operators 5
  - 1.4.1 Variable Names 5
  - 1.4.2 Operator 6
  - 1.4.3 Exit from R 8

2 Data Structure and Workspace in R 9
- 2.1 Data Structure 9
  - 2.1.1 Scalar 9
  - 2.1.2 Vector 10

  - 2.1.3 Matrix 11
  - 2.1.4 Array 12
  - 2.1.5 Data Frame 13
  - 2.1.6 List 15
  - 2.1.7 Factor 16
- 2.2 Data Entering 20
  - 2.2.1 Read Data into a Vector or List from the Console 20
  - 2.2.2 Read a line from a Console 21
  - 2.2.3 Using c( ) function 22
  - 2.2.4 Using read.table( ) function 22
  - 2.2.5 Using Data Editor 23
  - 2.2.6 Importing the data 24
- 2.3 Saving and Loading R Workspace 25
  - 2.3.1 Save and Load Command History 25
  - 2.3.2 Save and Load workspace 26
  - 2.3.3 Saving Output/Graph to a file 28

3 Basic Data Management 31
- 3.1 Creating New Variable 31
- 3.2 Recoding Data 31
- 3.3 Renaming Variable 32
- 3.4 Missing Values 33
- 3.5 Sorting, Ranking and Ordering 34
  - 3.5.1 Rank 34
  - 3.5.2 Sort 35
  - 3.5.3 Order 35
- 3.6 Control Structures: Conditional Execution 37
  - 3.6.1 if Statement 37
  - 3.6.2 if...else Statement 37
  - 3.6.3 ifelse Statement 38
  - 3.6.4 Switch Statement 39
- 3.7 Control Structures: Looping 40
  - 3.7.1 for, while and repeat Statement 40

  - 3.7.2 next and break Statements 41
- 3.8 Alternative to Loop 42
  - 3.8.1 apply( ) function 42
  - 3.8.2 lapply( ) function 43
  - 3.8.3 sapply() function 44
  - 3.8.4 vapply() function 44
  - 3.8.5 mapply() function 45
  - 3.8.6 rapply() function 45
  - 3.8.7 tapply() function 45
- 3.9 User Defined Function 46
- 3.10 Set Operations 48
  - 3.10.1 Union of Two Sets 49
  - 3.10.2 Intersection of Two Sets 49
  - 3.10.3 Difference of Two Sets 50
  - 3.10.4 Equal Sets 50
  - 3.10.5 Completement Set 51
- 3.11 Linear Algebra 52
  - 3.11.1 Vector Operation 52
  - 3.11.2 Matrix Operation 54
  - 3.11.3 Solving Systems of Linear Equations 57
- 3.12 Sequences and Repetitions 58
  - 3.12.1 : (colon) Notation 58
  - 3.12.2 seq( ) Function 58
  - 3.12.3 Repetitions 59

4 Graphing Techniques in R 60
- 4.1 Graphic Window 60
- 4.2 Graphical Parameters 60
- 4.3 Line Graph 65
- 4.4 Bar Graph 69
- 4.5 Histogram 72
- 4.6 Pie Graph 72
- 4.7 Box-plot 73

- 4.8 Scatter plot 73
- 4.9 Stem and Leaf Plot 74

5 Summary Statistics 75
- 5.1 What is statistics? 75
- 5.2 Scales of Measurement 76
  - 5.2.1 Nominal Scale 76
  - 5.2.2 Ordinal Scale 77
  - 5.2.3 Interval Scale 77
  - 5.2.4 Ratio Scale 78
- 5.3 Population and Sample 78
  - 5.3.1 Random Sampling without Replacement 79
  - 5.3.2 Random Sampling with Replacement 80
- 5.4 Why Important Statistics? 80
- 5.5 Frquency Distribution 81
  - 5.5.1 Construction of Frequency Distribution 81
  - 5.5.2 Joint Frequency Distribution 87
- 5.6 Measures of Central Tendency 90
  - 5.6.1 Mean 90
  - 5.6.2 Median 91
  - 5.6.3 Mode 92

5.7.2 Standard Deviation(SD) 94
  - 5.7.3 Coefficient of variation (CV) 95
  - 5.7.4 Quantiles 96
  - 5.7.5 Inter-quartile range(IQR) 97
  - 5.7.6 z-scores 97
  - 5.7.7 Summary Measures 98
- 5.8 Moment, Skewness and Kurtosis 100
  - 5.8.1 Moment 100
  - 5.8.2 Skewness 102
  - 5.8.3 Kurtosis 104
- 5.9 Discriptive Statistics by Levels of Measurement 105

  - 5.9.1 Misuse of Statistics 106
  - 5.9.2 Misuse of Graph 107
  - 5.9.3 Misuse of Likert Scale 108

6 Probability Distributions in R 111
- 6.1 Discrete Distributions 111
  - 6.1.1 Bernoulli Distribution 112
  - 6.1.2 Binomial Distribution 114
  - 6.1.3 Poisson Distribution 117
- 6.2 Continuous Distribution 120
  - 6.2.1 Uniform Distribution 120
  - 6.2.2 Normal Distribution 123
  - 6.2.3 Exponential Distribution 127
- 6.3 t-Distribution 130
- 6.4 Other Probability Distributions 133

7 Statistical Inference 134
- 7.1 Point Estimation 134
- 7.2 Interval Estimation 137
  - 7.2.1 Confidence Interval for the Population Mean with σ Known 137
  - 7.2.2 Confidence Interval for the Population Mean with σ Unknown 139
  - 7.2.3 Confidence Interval for the Population Proportion 140
  - 7.2.4 Confidence Interval for the Population Variance 141
  - 7.2.5 Confidence Interval for the Different between Two Means 142
- 7.3 Testing Hypothesis 147
- 7.4 Type I Error and Type II Error 149
- 7.5 One sample, two-sample, and paired-sample Test 150
  - 7.5.1 One-Sample Test 151
  - 7.5.2 Two-Sample Test 151
  - 7.5.3 Paired-Sample Test 152
- 7.6 Normal test 153
  - 7.6.1 One-sample z-test 153
  - 7.6.2 Two samples z-test 163

  - 7.6.3 z-test for Population Proportion 165
  - 7.6.4 z-test for Two Population Proportions 166
- 7.7 t - Test 167
  - 7.7.1 One Sample t-Test 167
  - 7.7.2 Two-sample t-Test 170
  - 7.7.3 Paired-Samples t-Test 172
- 7.8 Chi-Square (χ^2) Test 174
  - 7.8.1 Chi-square Test for Single Variance 174
  - 7.8.2 Chi-square Test for Goodness of Fit 177
  - 7.8.3 Chi-square Test for Independent 180
- 7.9 F Test for Compare Two Variances 185

8 Correlation and Regression 187
- 8.1 Scatter Diagram 187
- 8.2 Correlation Coefficient 189
- 8.3 Spearman's Rank Correlation Coefficient 192
- 8.4 Simple Linear Regression 193
  - 8.4.1 Check for Linearity 195
  - 8.4.2 Check for Ourtliers in Data 196
  - 8.4.3 Coumuting the Pearson's Correlation Coefficient 197
  - 8.4.4 Fitting the Assumed Regression Model 197
  - 8.4.5 Summary Results for Fitted Regression Line 198
  - 8.4.6 Estimated Coefficients and their Statistics 199
  - 8.4.7 Linear regression Diagnostic 201
  - 8.4.8 Residual Diagnostic 202
- References 204

စာရေးသူ၏ အမှာစာ

လေ့လာခဲ့သောပညာရပ်နှင့် ဖြတ်သန်းခဲ့သော အတွေ့အကြုံများအရ ဘဝ၏ပုံရိပ် တစ်ခုသဖွယ် အခြေခံစာရင်းအင်းပညာရပ်အတွက် R ကိုရေးသားခဲ့သည်။ ၁၉၈၆ခုနှစ်က အစပြု၍ ဆရာငယ်ဘဝမှာ ကျောင်းသားများအား သင်ကြားပေးခဲ့သော ကွန်ပျူတာ ဘာသာစကားများဖြစ်သည့် Fortran, Basic, Pascal နှင့် C တို့သည် စာရင်းအင်းပညာ အဓိကဆရာတစ်ဦးအတွက် စာရင်းအင်းပညာနှင့် ကွန်ပျူတာဘာသာစကားတို့ကို တွဲလျက် ယှဉ်ပါးလာခဲ့သည်။ ၁၉၈၇ ရောက်ပြန်တော့ စာရင်းအင်းပညာအတွက် ကွန်ပျူတာ ဆော့ဖ်(လ်)တစ်ခုဖြစ်သည့် SPSS(Statistical Package for Social Sciences) သည် စာရင်းအင်းပညာဆရာတစ်ဦးအတွက် ကောင်းမွန်တဲ့ အဖော်မွန်ဖြစ်လာရသည်။ စတင်လေ့လာစဉ်က SPSS သည် DOS version သာရှိခဲ့သည်။ နောက်ပိုင်း Windows version 6.0 နှင့်အတူ မိမိရဲ့ဌာနအတွက် Lecturer notes ကိုဖန်တီးရင်း ကျန်းမာရေးနှင့် စာရင်းအင်းဦးစီးဌာနနှင့် နိုင်ငံခြားရင်းနှီးမြှုပ်နှံမှုကော်မရှင်ရုံးတို့တွင် သင်ကြားပေးခဲ့ရသည်။ ထိုစဉ်က SPSS ကို မြန်မာလိုရေးချင်ခဲ့ပေမယ့် များပြားလှတဲ့တာဝန်များကြား ဖြတ်သန်းရင်း အကောင်အထည် မဖော်နိုင်ခဲ့ပေ။ ၁၉၉၉ ခုနှစ်ကာလ နိုင်ငံရပ်ခြား ပါရဂူကျောင်းသား ဘဝမှာ မိမိရဲ့သုတေသနအတွက် SPlus Software ကို သုံးခဲ့ရသည်။ သူရဲ့ ကောင်းကွက်တွေကို တွေ့လာခဲ့ရသည်။ Software တစ်ခုလိုသာမက Programming language အနေဖြင့်ပါသုံးခဲ့ရသည်။ လွယ်ကူစွာနဲ့ အဆင်သင့် သုံးရန်အတွက် user friendly အဖြစ်ဖန်တီးထားတဲ့ software တွေရဲ့ အကန့်အသတ်ကို နားလည်ခဲ့ရသလို နက်ရှိုင်းသောလေ့လာမှုများအတွက် မလုံလောက်ကြောင်း သိရှိခဲ့ရသည်။ ဤတွင် လူပြောများလာတဲ့ R, အသုံးများလာတဲ့ R, နိုင်ငံခြားတက္ကသိုလ်များမှာ သင်ထောက်ကူအဖြစ် အသုံးပြုနေပြီဖြစ်တဲ့ R သည် ရင်ထဲသို့တိုးဝင်လာခဲ့သည်။ စူးစမ်းလေ့လာကြည့်တော့ မိမိအသုံးပြုခဲ့တဲ့ SPlus နှင့် အားလုံးနီးပါး ဆင်တူမှုနဲ့အတူ အခမဲ့ရရှိခြင်းက စာရင်းအင်းပညာလေ့လာသူများ၊ သုတေသီများနဲ့ သင်ကြားရေးဌာနများအတွက် များစွာ လိုအပ်သည်ဟု ထင်မြင်လာသည်။ R သည် Software တစ်ခုဖြစ်သလို programme Language တစ်ခုလည်းဖြစ်နေခြင်းက ၎င်း၏ အားသာချက်ဖြစ်နေသည်။

နဂိုရှိရင်းစွဲ ဆန္ဒအငှတ်ကလေးက ပြန်ပေါ်လာခဲ့ရသည်။ အားလုံးနားလည်လွယ်အောင် မြန်မာလိုမိတ်ဆက်မည်။ လက်တွေ့အသုံးကျတဲ့ စာရင်းအင်းပညာ၏ နက်ရှိုင်းမှု၊ ကွန်ပျူတာ ဘာသာစကား၏ နက်နဲမှု၊ အင်္ဂလိပ်လိုဖြစ်နေတဲ့ အဟန့်အတားများကို ကျော်ဖြတ်ရန် မြန်မာလိုရေးသင့်သည်ဟု ဆုံးဖြတ်ခဲ့သည်။ စတင်ရေးသားစဉ်က R program ကိုသာ ဖောင်းပေးခဲ့

သော်လည်း စာရင်းအင်းပညာလေ့လာသူ ကျောင်းသားများနှင့် သုတေသီများအတွက် စာရင်းအင်းပညာ၏ အခြေခံသဘောတရားများကို R နှင့်တွဲ၍ အကျရာတင်ခဲ့သည်။ အတွေ့အကြုံများကို အခြေခံ၍ နေ့စဉ်တွေ့ကြုံနေရသည့် ပြဿနာများကို R နှင့်တွဲပြီး ဖြေရှင်းပုံကို စာရင်းအင်းဆိုင်ရာ ကောက်ချက်ချခြင်းများနှင့်အတူ ရေးသားခဲ့သည်။

စာရင်းအင်းပညာ ဝေါဟာရများကို မြန်မာလိုပြန်ရာတွင် စာရင်းအင်းပညာ ဘာသာရပ်ကို သင်ကြားပေးခဲ့သော ဆရာကြီး၊ ဆရာမကြီးများ၏ သင်ကြားခဲ့မှုများကို ပြန်ပြောင်းသတိရရင်း သူတို့ရေးသားခဲ့သော မြန်မာလိုစာအုပ်များကို မှီငြမ်းမှုပြုခဲ့ပါသည်။ မြန်မာလိုအသုံးအနှုန်းမျိုးစုံကို တတ်နိုင်သမျှစုစည်း၍ လက်ရှိမြန်မာစာလုံးပေါင်းသတ်ပုံကျမ်းအရ ကိုက်ညှိခဲ့ပါသည်။ နောင်လူများ စာရင်းအင်းပညာဝေါဟာရများကို မြန်မာလို မှန်ကန်စွာသုံးစွဲတတ်စေရန် ရည်ရွယ်ခဲ့ပါသည်။ အကယ်၍ ချို့ယွင်းအားနည်းမှုများရှိပါက စာရေးသူ၏ချို့ယွင်းမှုသာဖြစ်ကြောင်း တင်ပြအပ်ပါသည်။ ဤစာအုပ်ပါ ပုစ္ဆာအများစုမှာ ဆရာကြီး၊ ဆရာမကြီးများနှင့် ယခင်စီးပွားရေးဘာသာအထူးပြု စာပေးစာယူသင်တန်းများ၏ စာအုပ်များမှထုတ်နှုတ်ဖော်ပြ၍ R ကိုသုံးပြီး တွက်ချက်တင်ပြထားပါသည်။

ဤစာအုပ်တွင် အခန်း(၁)နှင့် (၂) သည် R အကြောင်း အခြေခံအုတ်မြစ်ကို ပေးထားသည်။ အခန်း(၃)သည် data ကို အခြေခံစီမံခန့်ခွဲခြင်း(basic data management) ကိုတင်ပြထားသည်။ အခန်း(၄)သည် ဂရပ်များရေးဆွဲပုံကို အကျဉ်းချုံး ဖော်ပြထားသည်။ အခန်း(၃)နှင့်(၄)သည် R ကို အသေးစိတ်လေ့လာလိုသူများအတွက်ကိုသာ ရည်ရွယ်၍ ထည့်သွင်းရေးသားထားခြင်းဖြစ်သည်။ အခန်း(၅)သည် သရုပ်ဖော်စာရင်းအင်း (descriptive statistics) ကိုရှင်းပြထားသည်။ အခန်း(၆) သည် ဖြစ်တန်စွမ်းပြန့်ချက် (probability distribution)တို့ကိုတွက်ချက်ခြင်း၊ ဂရပ်များရေးဆွဲခြင်းနှင့် ကျပန်းကိန်းရှင် (random variable) ထုတ်ယူပုံတို့ ရှင်းပြထားသည်။ အခန်း(၇)သည် အစက်ခန့်မှန်းခြင်း (point estimatation) နှင့် ကြားခန့်မှန်းခြင်း (interval estimation) တို့ပါဝင်သော ကောက်ချက်ချစာရင်းအင်းပညာ (statistical inference) ကို ဖော်ပြထားသည်။ အခန်း(၈) သည် ကိန်းရှင်နှစ်ခုတို့၏ ဆက်စပ်မှုများအတွက် ဆက်စပ်မှုပြကိန်း (correlation coefficient) နှင့် မျဉ်းဖြောင့်ဆက်သွယ်မှု (linear regression) တို့ကို ဥပမာများဖြင့် တင်ပြထားပါသည်။

ဤစာအုပ်သည် R-programme နှင့် R-software ကို လေ့လာသင်ယူသူများအတွက်ရည်ရွယ်သလို စာရင်းအင်းပညာလေ့လာသူများ၊ သုတေသနပြုလုပ်နေသူများအတွက် အခြေခံကောင်းများ ပေးလိမ့်မည်ဟု မျှော်လင့်မိပါသည်။

ဒေါက်တာစိုးဝင်း, PhD
စာရင်းအင်းပညာ

ရက်စွဲ။ ။ ၁၀-၁-၂၀၁၉