

# Elements of Statistics



# Elements of Statistics:

*A Hands-on Primer*

By

Raghubar D. Sharma

Cambridge  
Scholars  
Publishing



Elements of Statistics: A Hands-on Primer

By Raghubar D. Sharma

This book first published 2017

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2017 by Raghubar D. Sharma

Cover Image *From Chaos to Statistics* © Jaspal Singh Cheema, 2017

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-0034-9

ISBN (13): 978-1-5275-0034-1

For Anjana, Vivek, Shilpa and Pratik

Dedicated to: My High School Mathematics Teacher  
Mr. C.L. Sharma whose gentle teaching style instilled  
love for mathematics in me and my brothers.



# TABLE OF CONTENTS

List of Illustrations .....	xv
Introduction .....	xix
Chapter One.....	1
Basic Math and Symbols Used in Statistics	
Learning Objectives .....	1
Introduction.....	1
Basic Math Needed to Learn Statistics .....	2
The Order of Operations .....	2
Decimals and Fractions.....	3
Truncation .....	6
Exponents.....	6
Square Root.....	7
Logarithms .....	7
Some Common Symbols Used in Statistics .....	7
Summation ( $\Sigma$ ) and Constant (c).....	8
Summation ( $\Sigma$ ), Constant (c), and a Variable .....	8
Exercises for Practice.....	9
Chapter Two .....	11
Statistical Thinking: Levels of Measurement and Variables	
Learning Objectives .....	11
Introduction.....	11
Types of Numbers and Levels of Measurement.....	12
Types of Variables .....	15
Continuous Variables .....	15
Discrete or Qualitative or Categorical Variables.....	15
Why Different Levels of Measure Matter? .....	16
Exercises for Practice.....	18
Chapter Three .....	19
Graphs and Charts: Use and Misuse of Visuals	
Learning Objectives .....	19
Introduction.....	19
Grouped Data.....	20

Cumulative Frequency .....	22
Cumulative Percentage .....	22
Graphs and Charts .....	22
Rules for Creating Graphs and Charts .....	23
Pie Chart and Double-counting .....	27
Misuse of Graphic Representation .....	28
Eliminating Zero from Y-axis .....	28
Stretching or Shrinking the Axis .....	30
Avoiding Misuse of Graphic Representation .....	31
Exercises for Practice .....	32
 Chapter Four .....	 36
Central Tendency: Average tends to be a Central Number	
Learning Objectives .....	36
Introduction .....	36
Central Tendency .....	37
Measures of Central Tendency .....	37
Mode .....	37
Bimodality and Multimodality .....	40
Median .....	40
Median of Grouped Data .....	41
Some Limitations of the Median .....	43
Mean .....	43
Mean of Grouped Data .....	43
Weighted Mean: The Mean of Means .....	45
Skewness, Mean and Median .....	46
Symmetrical Distribution .....	47
Positively or Right-Skewed Distribution .....	48
Negatively or Left-Skewed Distribution .....	50
Summary Table: Level of Measurement and Type of Central Tendency .....	51
Exercises for Practice .....	51
 Chapter Five .....	 53
Variability: Measures of Dispersion	
Learning Objectives .....	53
Introduction .....	53
Dispersion or Variability .....	54
The Range .....	54
The Deviation .....	55
The Mean or Average Deviation .....	56



Variance and Standard Deviation .....	56
Variance .....	57
Calculating Variance .....	57
Standard Deviation .....	58
Standard Deviation of Grouped Data .....	58
An Alternate Method .....	60
Coefficient of Variation .....	62
Practical Use of Coefficient of Variation .....	63
Exercises for Practice .....	63
 Chapter Six .....	 65
Probability .....	
Learning Objectives .....	65
Introduction .....	65
The Concept of Randomness .....	66
Moving beyond the Descriptive Statistics .....	66
Probability Statement .....	67
Rules of Probability .....	67
Theoretical and Empirical Probability .....	68
Theoretical Probability .....	68
Empirical Probability .....	68
Law of Large Numbers and Probability .....	69
The Multiplication Rule .....	69
Independent Events .....	69
Dependent Events .....	70
The Addition Rule .....	71
Mutually Exclusive Events .....	71
Non-Mutually Exclusive Events .....	72
Odds .....	72
Odd Ratios .....	73
Calculating Probability from Odds .....	73
Probability Ratio .....	74
Role of Probability in Statistical Inference .....	74
Exercises for Practice .....	75
 Chapter Seven .....	 77
Sampling .....	
Learning Objectives .....	77
Introduction .....	77
A Sample .....	78
Sampling Frame .....	78

Sampling Error .....	78
Non-Sampling Error .....	79
Types of Samples .....	79
Probability Samples .....	79
Advantages of Probability Sampling .....	82
Sample Size .....	82
Non-Probability Samples .....	83
Exercise for Practice .....	85
 Chapter Eight .....	 87
The Sampling Distribution and the Normal Curve: Generalizing from a Sample to the Population	
Learning Objectives .....	87
Introduction .....	87
Sampling Distribution of Means .....	88
Central Limit Theorem .....	88
Normal Curve .....	89
Salient Features of Sampling Distribution of Means .....	91
Standard Error of the Sample Mean .....	91
Confidence Intervals .....	92
Calculating Confidence Interval .....	93
Exercises for Practice .....	94
 Chapter Nine .....	 97
Normal Distribution and its Relationship with the Standard Deviation and the Standard Scores	
Learning Objectives .....	97
Introduction .....	97
Standardized Scores .....	99
Conversion of a Raw Score to a Standard Score (z-Scores) .....	99
Calculating the z-Score .....	100
Conversion of a z-Score into a Raw Score .....	102
Finding Probability of an Event using z-Score and the Normal Curve .....	104
Tails of a Curve .....	105
Exercises for Practice .....	106
 Chapter Ten .....	 108
Examining Relationship	
Learning Objectives .....	108
Introduction .....	108

Cross-tabulation .....	109
Univariate Tables .....	109
Bivariate Tables .....	109
Multivariate Tables .....	110
Test of Significance .....	111
Hypothesis Testing .....	112
Assumptions .....	113
Independence .....	114
Normality .....	114
Randomness .....	114
One-tailed and Two-tailed Tests .....	114
Level of Significance .....	115
Level of Confidence .....	115
p-Values .....	115
Calculating p-Value .....	115
Use of p-Values .....	116
The Number of Degrees of Freedom .....	116
Type 1 Error .....	117
Type 2 Error .....	118
Four Steps for Testing a Hypothesis .....	118
Step 1: Hypothesis Statement .....	119
Step 2: The Test of Significance .....	119
Step 3: Calculations of Test Statistic .....	119
Step 4: Decision Rule .....	120
Exercises for Practice .....	120
Chapter Eleven .....	122
Tests of Significance for the Nominal Level Variables	
Learning Objectives .....	122
Introduction .....	122
Research Question, the Level of Measurement and Choice of a Test .....	123
Visual Evaluation of Relationship between two Variables .....	124
The Chi-square ( $\chi^2$ ): A Test of Significance .....	126
Uses of the Chi-square ( $\chi^2$ ) Test .....	126
Some Requirements for a Chi-square ( $\chi^2$ ) Test .....	127
Calculation Steps for the Chi-square ( $\chi^2$ ) .....	127
The Chi-square ( $\chi^2$ ): A Test Dependency .....	127
Expected Frequencies .....	129
The Chi-square ( $\chi^2$ ): A Test of “Goodness of Fit” .....	130
Expected Frequencies .....	132

Measuring the Strength of Association .....	133
Phi ( $\phi$ ).....	133
Interpretation of Phi .....	135
Cramer's V .....	136
Lambda ( $\lambda$ ).....	137
Exercises for Practice.....	139
Chapter Twelve .....	144
Tests of Significant for the Ordinal Level of Variables	
Learning Objectives .....	144
Introduction.....	144
Kruskal's Gamma ( $\gamma$ ).....	145
Spearman's Rho ( $\rho_s$ ) or Spearman's Rank Correlation Coefficient ..	147
Significance Level of Spearman's Rho ( $\rho_s$ ).....	149
Somers' <i>D</i> .....	150
Kendall's Tau-b ( $\tau_b$ ).....	154
Significance Level of Somers' <i>D</i> and Kendall's Tau-b ( $\tau_b$ ).....	156
Exercises for Practice.....	157
Chapter Thirteen.....	163
Tests of Significant for the Interval Level of Variables	
Learning Objectives .....	163
Introduction.....	163
Generalizing from Sample to Its Population .....	164
The <i>t</i> -Distribution .....	164
The <i>t</i> -test .....	165
Standard Error .....	165
Calculating <i>t</i> value.....	166
The <i>t</i> -test for Comparing Two Sample Means .....	168
Exercises for Practice.....	171
Chapter Fourteen .....	172
Measuring Relationship between two Interval Level Variables	
Learning Objectives .....	172
Introduction.....	172
Coefficient of Correlation Pearson's <i>r</i> ).....	173
Calculation of Coefficient of Correlation.....	174
The Significance of Sample Size.....	178
Summary: Interpretation of the Correlation Coefficient .....	179
Coefficient of Determination ( <i>R-Squared</i> ).....	180
Linearity.....	180

An Ingenious Utility of Correlation .....	182
Exercises for Practice .....	183
Chapter Fifteen .....	185
Power of a Statistical Test	
Learning Objectives .....	185
Introduction .....	185
Power of a Test .....	186
Calculating Power of a Test .....	186
Factors Affecting the Power of a Test .....	187
The Effect of Sample Size on Power .....	188
The Effect of the Significance Level ( $\alpha$ ) on Power .....	189
The Effect of the Directional Nature of the Alternate Hypothesis .....	190
Parametric and Non-parametric test and Power .....	191
Type 1 Error, Type 2 Error and Power .....	191
Exercises for Practice .....	192
Chapter Sixteen .....	194
Analysis of Variance (ANOVA) Learning Objectives	
Introduction .....	194
Analysis of Variance (ANOVA) .....	195
One-way Analysis of Variance .....	195
Total Sum of Squares (SST) .....	197
Sum of Square within Groups (SSW) .....	198
Sum of Square between Groups (SSB) .....	198
ANOVA Table .....	199
Simpler Method .....	199
Total Sum of Squares (SST) .....	200
Sum of Square within Groups (SSW) .....	200
Sum of Square between Groups (SSB) .....	201
Limitation of ANOVA .....	201
Exercises for Practice .....	202
Chapter Seventeen .....	204
Regression Analysis: A Prediction and Forecasting Technique	
Learning Objectives .....	204
Introduction .....	204
Regression Equation and its Application for Forecasting .....	206
Regression Coefficients .....	206
Variance Explained: The Second Major Use of Regression .....	208

Calculation of Coefficient of Correlation ( $R^2$ ) .....	208
Interpretation of R-Square.....	210
The Correlation Coefficient ( $r$ ) and the Regression Coefficient ( $b$ )...	210
Exercises for Practice.....	211
Solutions to Exercises for Practice .....	213
Appendices .....	228
Tables .....	
References .....	238
Statistical Procedure and Tests by Appropriate Level of Measurement .....	239
Index.....	240

# LIST OF ILLUSTRATIONS

## Figures

Figure 3-1: Vertical Bar Graph

Figure 3.2: Horizontal Bar Graph

Figure 3-3: Line Graph

Figure 3-3A: Vertical Bar Graph

Figure 3-4: Pie Chart

Figure 3-5: Vertical Bar Graph with Zero on *Y*-Axis

Figure 3-6: Vertical Bar Graph without Zero on *Y*-Axis

Figure 3-7: Vertical Bar Graph without Stretched Axis

Figure 3-8: Vertical Bar Graph with Stretched *X*-Axis

Figure 4-1: Mode in a Pie Chart

Figure 4-2: Mode in a Vertical Bar Graph

Figure 4-3: Mode in a Line Graph

Figure 4-4: Bimodal Graph

Figure 4-5: Symmetrical Distribution

Figure 4-6: Positively or Right-Skewed Distribution

Figure 4-7: Individual Total Income, Canada, 2010

Figure 4-8: Negatively or Left-Skewed Distribution

Figure 4-9: Death Rates by Age, Canada, 2000-02

Figure 8-1: Normal Curve or Gaussian Curve

Figure 8-2: Normal Distribution of Observations Between 1, 2, 3, and 4 Standard Deviations

Figure 9-1: Normal Curve with Mean = 30 and Standard Deviation = 4

Figure 9-2: Mary's Score (166)

Figure 9-3: John's Score (145)

Figure 9-4: Area under the Normal Curve between two Scores

Figure 14-1: Wages Earned by Number of Hours Worked

Figure 14-2: Death Rates by Age, Canada, 2005

Figure 17-1: Regression Line

## Tables

Table 1-1 Calculations for a. Sum of Squared Deviations of  $X$  and  $Y$ , b. Sum of Products of  $X$  and  $Y$ , and c. Sum of Product of Summations of  $X$  and  $Y$

Table 3-1: A Frequency Distribution of Books by Subject

Table 3-2: A Frequency Distribution of Number of Persons by Income

Table 3-3: A Frequency Distribution of Number of Persons by Income

Table 3-4: Workforce by Employment Equity Status

Table 4-1: Responses for Perceiving Pollution in a City

Table 4- 2: Number of Persons by Age Interval

Table 4- 3: Frequency, Midpoint and Estimated Total Years by Age Interval

Table 4-4: Calculation of Weighted Mean

Table 4-5: Symmetrical, Positively Skewed and Negatively Skewed Distributions

Table 4.6: Measure of Central Tendency and Level of Measurement

Table 5-1: Calculation of the Mean Deviation

Table 5-2: Calculation of Variance

Table 5-3: Calculation of Standard Deviation of Grouped Data

Table 5-4: Standard Deviation of Grouped Data by Simpler Method

Table 7-1: Calculation of the Sample Size for each Stratum

Table 7-2: Calculation of Sample Size from the Percentage Distribution of Population by Age in each Stratum

Table 10-1: Number of Families by Type

Table 10-2: Income by Sex

Table 10-3: Income by Sex and Age

Table 11-1: Number of Voters by Party Preference

Table 11-2: Number of Voters by Party Preference and Sex of Respondents

Table 11-3: Percentage of Voters by Party Preference and Sex of Respondents

Table 11-4: Party Affiliation by Religion

Table 11-4a: Party Affiliation by Religion, Calculation of Expected Frequencies



- Table 11-4b: Party Affiliation by Religion, Expected Frequencies  
Table 11-4c: Party Affiliation by Religion, Calculation of Chi-square  
Table 11-5: Sales Patterns before the Introduction of a New Cereal  
Table 11-5a: Calculation of Expected Frequencies after the Introduction of a New Cereal  
Table 11-5b: Expected Sales Patterns after the Introduction of a New Cereal  
Table 11-5c: Sales Patterns before and after the Introduction of a New Cereal, Calculations of the Chi-Square  
Table 11-6: Took Remedial Course and Did Well on the Test, Observed Frequencies ( $f_o$ )  
Table 11-6a: Took Remedial Course and Did Well on the Test, Calculations for Expected Frequencies ( $f_e$ )  
Table 11-6b: Took Remedial Course and Did Well on the Test, Expected Frequencies ( $f_e$ )  
Table 11-6c: Calculation for Chi-Square  
Table 11-7 Attitude toward Abortion by Religion
- Table 12-1: Support for Abortion ( $y$ ) by Religiosity ( $x$ )  
Table 12-2: The Gross Domestic Product and the Human Development Index of Selected Nations  
Table 12-3: Support for Charitable Giving by Church Attendance  
Table 12-3a: Calculations of Concordant Cells  
Table 12-3b: Calculation of Discordant Cells  
Table 12-3c: Calculation of Ties ( $T_y$ ) on the Dependent Variable  
Table 12-3d: Calculation of Ties ( $T_x$ ) on the Independent Variable
- Table 14-1 Number of Kilometres Walked ( $x$ ) per Day in a Month and Number of Kilograms Reduced ( $y$ )  
Table 14-2: Wages Earned by Number of Hours Worked
- Table 15-1: Critical Value of  $z$  to Reject  $H_0$  at a Given  $\alpha$ -Level  
Table 15-2: Decision on  $H_0$  and Type of Error
- Table 16-1: Reduction in Blood Pressure by Treatment  
Table 16-2: ANOVA Table  
Table 16-3: Reduction in Blood Pressure in groups Using medication (Group 1), Diet (Group 2), and Exercise (Group 3)
- Table 17-1 Number of Kilometres Walked ( $x$ ) per Day in a Month and Number of Kilograms of Weight Reduced ( $y$ )

Table 17.2: Number of Kilometres Walked ( $x$ ), Actual Number of Kilograms Reduced ( $y$ ), and Predicted Number of Kilograms Reduced ( $\hat{y}$ )

# INTRODUCTION

Statistics has become a form of logic or rhetoric that everyone needs to learn to navigate the modern world. Though this book is primarily aimed at undergraduate students who are required to take at least one compulsory statistics course before graduation, it is also a valuable guide for anyone who has little or no background in statistics and wants to become statistics literate. Without the pretensions of the famous book that the learning of statistics can be without tears or that you don't need to understand symbols, formulae, and equations, this book will prepare you to understand basic statistics and to complete your statistics course without anxiety. This book has been written with the conviction that you don't need to be a mathematician to learn statistics. It is a crucial resource for students taking a required statistics course who are intimidated by statistical symbols, formulae, and daunting equations.

The application of statistics in social research has become imperative. A gap usually exists between the time when students take their first statistics course and when they engage in their first serious research project (typically during an internship, or a research methods course, or their final year project/thesis). Because of this gap, students often don't remember basic statistics well enough to apply it effectively in their research. Hence, there is a need for a “desk reference,” “refresher,” or “core concept” text—an Elements of Statistics for burgeoning researchers, à la Strunk and White's Elements of Style. This book will serve as an excellent desk reference, refresher, or core concept text for the budding researcher. It will also be helpful while interning or working as a research assistant or research associate.

Those who feel left out when their colleagues, supervisors, or bosses use statistics will benefit from this book. This particular group of people has been on my mind for a long time. When I was employed as a workforce data analyst for an employment equity office in the early 1990s, I routinely prepared reports and presentations for senior management. The basic education of my manager was grade 11, and her boss, an assistant deputy minister (ADM), had grade 12 with an accounting certification. They both had missed the opportunity to learn statistics in school. One day, my manager returned from her boss's office after discussing a report I had prepared and told me that the ADM preferred “circles” not “hills” in

the report. After some pondering, I realized that her boss preferred pie charts over bar graphs. This report was on employment equity designated groups, which included Aboriginal peoples, women, and persons with disabilities—where a respondent could be counted more than once. For example, the same person could be counted as a woman, an Aboriginal person, and a person with a disability. In Chapter 3, I explain that a pie chart is not appropriate where there is double-counting of respondents or observations.

Because my manager had little to no statistical literacy, it would have been futile to explain to her why pie charts are not a good idea for this kind of data. But since then, I have felt that I should write a book on statistics that could help people like my manager and her boss. Today, even if you work in non-statistical areas such as policy, communications, and journalism, you need to have some knowledge of statistics. Nowadays, statistical literacy is as important as literacy itself. This book is written in a self-help, hands-on learning style so the reader can easily attain the skills needed to achieve a basic understanding of statistics and be comfortable with presentations loaded with statistics.

This book follows an easy-to-comprehend format. It gives a strong foundation in the basics, while calculations elaborate on the basics in sequences designed for students and general readers who have never taken a statistics course. Simply put, it is a hands-on primer. The idea is that when you're reading it, you won't need a calculator or a computer.

The book contains 17 chapters along with statistical tables in the appendix. Chapter 1 provides a refresher on basic math and statistical symbols, while Chapter 2 discusses the levels of data measurement and types of variables. Chapter 3 deals with the visual representation of data. It also cautions researchers on the use and misuse of graphics. Chapters 4 and 5 discuss the measures of central tendency and variability, respectively. Chapter 6 familiarizes the reader with the basic concepts of probability. As researchers are always required to work with samples, Chapter 7 is devoted to methods for selecting appropriate samples. The next three chapters deal with important concepts that a researcher must know before embarking on applying a statistical technique to data. Chapter 8 discusses the sampling distribution and the normal curve, particularly with respect to generalizing from a sample to the population. Chapter 9 elucidates the relationship between the normal distribution and standard scores. It also includes conversion of raw scores into standard scores, an essential requirement for comparative research. Chapter 10 examines relationships between variables as well as the procedure and essential concepts for testing a hypothesis.

Chapters 11, 12, and 13 are devoted to tests of significance for the nominal-, the ordinal-, and the interval/ratio-level variables, respectively. Chapter 14 discusses the correlation coefficient, and Chapter 15 is devoted to the statistical power of a statistical test. Chapter 16 provides the basics on analysis of variance (ANOVA), and Chapter 17 focuses on regression. Thus, the book ends with a comprehensive survey of applied statistics and fills the lacunae left by the majority of statistics books.

All exercises and examples in the book have been developed by the author. Due care has been taken to credit the sources used in the book. Any omission in referencing is, of course, unintentional and once pointed out will be rectified in the next edition.



# CHAPTER ONE

## BASIC MATH AND SYMBOLS USED IN STATISTICS

### Learning Objectives

In order to learn statistics, you need some knowledge of basic mathematics, which most of you have already acquired during your grade and high school years. Because memories tend to fade over time, this chapter serves as a refresher of basic mathematical operations. It also introduces some commonly used statistical symbols. Specifically, you will learn about:

- the order of operations;
- fractions, decimals, exponents, and logarithms; and
- the most frequently used statistical symbols.

### Introduction

H.G. Wells once said, “Statistical thinking one day will be as necessary for the efficient citizenship as the ability to read and write.”<sup>1</sup> Arguably, that day has arrived, as we are bombarded daily with statistics from television commentators, newspapers, popular multimedia, and advertising billboards. Terms and phrases such as *batting average*, *the chance of winning an election*, *outliers*, and *median income* are all statistics that are routinely used in popular media. Yet many people, including university students, think that statistics is not relevant to their learning. Statistics is not taken seriously because we continuously hear that statistics is not an objective science. Yet, experts in various fields often use statistics to

---

<sup>1</sup> Wells, H.G. 1903. *Mankind in the Making*, London: Chapman & Hall, Page 204.

refute each other's claims (Haan<sup>2</sup>, 2008). Politicians frequently use statistics during election campaign debates to score points over one another. The idea that statistics can be used to lie was made popular by Darrell Huff<sup>3</sup> in 1954. In fact, Huff's book, *How to Lie with Statistics*, does not propagate lying with statistics; it illustrates how statistics can be misused. The following example shows that the *average* is meaningless without reference to the spread (variability) of data. Let's say that five students in a class brought \$5, \$6, \$2, \$2, and \$60 each to buy lunch. If you used only the average, you would say that on average a student brought \$15 for lunch. The use of the average suggests that students brought more-than-sufficient money for lunch, whereas at least two, or 40% of the students had enough money only to buy a soft drink. In statistical terms, the \$60 is an *outlier*, which increases the spread of data. Using an *average* without considering the spread of data around the *mean* value can create a misleading impression.

Another misconception is that to learn basic statistics you need to know advanced mathematics, when, in fact, knowledge of basic math is sufficient to understand most statistics. In the next section, we review the necessary math needed to learn statistics. Though some equations in statistics textbooks may look daunting, you can easily understand and apply basic statistics without attempting to solve these intimidating equations.

## Basic Math Needed to Learn Statistics

You can learn to apply advanced-level statistics without learning advanced-level math. These days, most statistical calculations are done by computer software. You can learn to interpret statistics produced by statistical software such as SPSS and SAS without learning advanced-level mathematics. Even if your basic math is rusty, the following review will be sufficient to learn essential statistics.

## The Order of Operations

The order of operations is the basic principle that governs the sequence of operations, which is bracket (parenthesis), exponent, division, multiplication,

---

<sup>2</sup> Haan, Michael. *Introduction to Statistics for Canadian Social Scientists*, Don Mills, Ontario: Oxford University Press, 2008.

<sup>3</sup> Huff, Darrell, and Irving Geis (illustrator). *How to Lie with Statistics*, New York: W.W. Norton and Company Inc., 1954.



addition, and subtraction. You may be familiar with **BEDMAS**, an acronym you likely learned in grade or high school to memorize the sequence of operations, where:

**B** stands for **Bracket**;  
**E** stands for **Exponent**;  
**D** stands for **Division**;  
**M** stands for **Multiplication**;  
**A** stands for **Addition**; and  
**S** stands for **Subtraction**.

Let's say we have to solve the following equation to find the value of  $x$ :

$$x = 2 \times (3 + 5)^2 \div 4 + 6 - 7$$

Applying BEDMAS, we will first find a solution for the equation in the **bracket**:  $(3 + 5) = 8$ .

Next, we will find the value of the **exponent** of  $8 = 8^2 = 64$ .

The next step will be **division**:  $64 \div 4 = 16$ .

After division, the next operation is **multiplication**:  $2 \times 16 = 32$ .

After multiplication, the next step is **addition**:  $32 + 6 = 38$ .

The final step is **subtraction**:  $38 - 7 = 31$ .

*Note:* If there is more than one operation within the bracket, follow the same sequence. For example, if the equation in the bracket is  $(5 + 6 \div 2 \times 4 - 1)$ , then perform the operations in the following sequence: first, divide 6 by 2 = 3; next, multiply 3 by 4 = 12; then, add 5 to 12 = 17, and then subtract 1 from 17 = 16. The value in the bracket is 16.

## Decimals and Fractions

It is necessary to be comfortable with decimals because many statistics and statistical relationships are expressed in decimals. You can use a calculator to calculate a decimal from a fraction. One way to convert a fraction to a decimal is to divide the *numerator* by the *denominator*. For example,  $\frac{8}{16}$  is equal to 8 divided by 16, which is equal to 0.5. A fraction

can also be expressed as a percentage by converting the fraction to a decimal and then multiplying it by 100:  $0.5 \times 100 = 50\%$ .

You will also learn that the concept of probability is central to statistical prediction. Probability is expressed in decimal points, and a chance of an event happening is expressed in percentage. For example, if there is a 30% chance of catching a fish from a river, you could say that the probability of catching a fish is 0.3. The following is a refresher from grade or high school mathematics on fractions, decimals, and percentages, which you can calculate by hand or with a basic calculator.

### ***Finding the Percent of a Number***

**Example:** *to find out what is 92% of 28.*

- Multiply the number by the percent:  $28 \times 92 = 2576$ .
- Divide the total by 100:  $2576 \div 100$ .
- To find out the answer, move the decimal point two places to the left: **25.76**.

### ***Finding Percentage***

**Example:** *to find out what percent is 28 of 92?*

- Divide the first number by the second:  $28 \div 92 = 0.3043$
- Multiply the answer by 100:  $0.3043 \times 100$
- Move the decimal point two places to the right: **30.43%**.

### ***Converting a Fraction to a Decimal***

**Example:** *Convert  $\frac{1}{3}$  to a decimal.*

- Divide the numerator of the fraction by the denominator:  $1 \div 3 = 0.3333$ .

### ***Converting a Fraction to a Percent***

- After converting a fraction to a decimal, simply multiply by 100 and move the decimal point two places to the right:  $0.3333 \times 100 = 33.33\%$ .

### ***Converting a Percent to a Fraction***

**Example:** Convert 75% to a fraction.

- Remove the percent sign from 75.
- Make a fraction with the percent as the numerator and 100 as the denominator:  $\frac{75}{100} = 75 \div 100 = \mathbf{0.75}$ .

### ***Converting a Decimal to Percent***

**Example:** Convert 0.75 to a percent.

- Multiply the decimal by 100:  $0.75 \times 100 = 75$ .
- Add a percent sign to 75: **75%**.

### ***Converting a Percent to a Decimal***

**Example:** Convert 75% to a decimal.

- Divide the percent by 100:  $75 \div 100 = \mathbf{0.75}$ .

### ***Rounding Decimals***

The rounding of decimal points makes calculation easier and presentation of the numbers clearer. One reason we usually need to round a number is because a decimal may extend endlessly. For example,  $1/3$  results in 0.3333333333. In the rounding of decimals, we need to consider two questions.

1. How many places should we carry the decimal point?
2. How do we decide that the last number reflects the remainder?

The answer to the first question is that it depends on the type of data. Some demographic data, such as survival rates, may extend to sixth decimal point; whereas in many other situations, you might decide to keep only one decimal place. Conventionally, we round to two decimal places.

The answer to the second question is that we retain the value of the last decimal place if the value next to the retained decimal place number is less than 5. For example, 2.344 is rounded to 2.34. We increase the value of the retained decimal place by 1 if the next decimal place is 5 or greater. For example, 3.765 is rounded to 3.77.

Some statisticians suggest that in a dataset, the number ending with 5 after the decimal point should be rounded up for one-half of the time and down for the other half of the time. For example, say the first number ending with 5 after the decimal point is 2.235; it is rounded down to 2.23. Say the second number ending with 5 after the decimal point is 6.475; it is rounded up to 6.48. **However, I would suggest that for numbers ending with 5 after the decimal point, you round up every time.**

### Truncation

When we retain the decimal place just as it is, without changing the value of the decimal place, it is called *truncation*. Some computer programs truncate the numbers without changing the value of the retained decimal place; for example, 2.334 and 2.337 are both retained as 2.33. In other words, both numbers are truncated to 2.33.

### Exponents

In statistics, exponents are used quite routinely; therefore, it is important to familiarize yourself with them. Basically, an exponent indicates the number of times a numeral should be multiplied with another. For example,  $2^3$  means that 2 is multiplied 3 times:  $2 \times 2 \times 2$ . In the expression  $2^3$ , the 2 is called the *base* and the 3 is called the *exponent*. It is also commonly called *2 raised to the power of 3*. Here, the 2 is called a *base* and the 3 a *power*.

#### ***Multiplication and Division of Two Exponents with Identical Bases***

If two exponents with identical bases are multiplied, the rule is to add the exponents. For example, to multiply  $2^4$  by  $2^3$ , you add the exponents 4 and 3. Thus,  $2^4 \times 2^3$  becomes  $2^{(4+3)}$ , or  $2^7$ . Its value is:  $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 128$ .

If two exponents with identical bases are divided, the rule is to subtract the exponents. For example, to divide  $2^5$  by  $2^3$ , you subtract the exponent 3 from 5. Thus,  $2^5 \div 2^3$  becomes  $2^{(5-3)}$ , or  $2^2$ . Its value is:  $2 \times 2 = 4$ .

## Square Root

The square root of a number is the reverse operation of a square of that number. For example:

Square of  $n = n^2$

If  $n = 3$ , then  $3^2 = 3 \times 3 = \mathbf{9}$ .

The square root of  $n = \sqrt{n}$ .

If  $n = 9$ , then  $\sqrt{9} = \sqrt{3 \times 3} = \mathbf{3}$ .

## Logarithms

You may have learned in high school that a logarithm is a special type of exponent. The base of a logarithm is either 10 or 2.718. When the base is 10, it is called the *common logarithm*; when the base is 2.718, it is called the *natural logarithm*. What this means is that if we raise 10 to the power of 3 ( $10^3$ ), the answer is  $10 \times 10 \times 10 = \mathbf{1000}$ , and, hence, the common log of 1000 is 3. Similarly, if we raise 2.718 to the power of 3 ( $2.718^3$ ), the answer is about 20 ( $2.718 \times 2.718 \times 2.718 \approx 20.01$ ), which means the natural log of 20 will be about 3, or, to be precise, 2.9957. The  $\approx$  stands for “approximately equal to.”

## Some Common Symbols Used in Statistics

The symbols used in statistical equations may intimidate a non-mathematical person. The fear of symbols sometimes disheartens a person to learn statistics. This fear is unnecessary. Once you understand the meaning of symbols, the fear disappears and the learning of statistics becomes easy.

Most statistics textbooks use  $X$  and  $Y$  as symbols for *variables*. Basically, a variable is a characteristic (such as sex or social class) or a quantity (such as age or income). A variable varies between its categories. For example, sex can take a value of a male or a female, and class might vary between the lower, middle, or upper class. Similarly, age might take any value from one day old to 100 years old, and income could vary from 0 dollars to billions of dollars. The symbol  $N$  is used for the number of persons or the number of cases in a population, and the symbol  $n$  is used for the number of persons or the number of cases in a sample. Generally, uppercase letters ( $X, Y, Z$ ) are used to represent population characteristics, and lowercase letters ( $x, y, z$ ) are used to denote sample characteristics.

The most dreaded symbol for a person unfamiliar with statistics is a Greek-alphabet uppercase sigma, which is written as  $\Sigma$  and denotes the adding up or summing up of numbers. For example,  $\Sigma(X_1, X_2, X_3)$  indicates that we are adding quantities represented by the symbols  $X_1, X_2,$  and  $X_3$ . Simply put, if  $X_1 = 2, X_2 = 3,$  and  $X_3 = 4,$  then  $\Sigma(X_1, X_2, X_3) = 2 + 3 + 4 = 9$ . It is that simple. You will see sigma written as follows:

$$\sum_{i=1}^N X_i$$

In the above example,  $N = 3$ .  $X_i$  indicates that  $X$  takes  $i$  values. Because  $X$  takes three values (2, 3, and 4), in this example  $i$  is equal to 3.

The summation sign,  $\Sigma$ , is the most frequently used symbol in statistics. The following rules will be helpful to understand its use.

### Summation ( $\Sigma$ ) and Constant ( $c$ )

Written in symbols:  $\Sigma c = N \times c$ .

$\Sigma c$  means that the sum of constants is equal to the number of times a constant appears in the series multiplied by the value of the constant:

if  $c = 10$  and  $N = 6$ , then  $N \times c = 10 \times 6 = 60$ . It is the same as:  
 $10 + 10 + 10 + 10 + 10 + 10 = 60$ .

### Summation ( $\Sigma$ ), Constant ( $c$ ), and a Variable

Written in symbols:  $\Sigma cX_i$

The symbols above suggest that you first multiply each value of variable  $X$  with the constant and then add them up. If  $c = 10, X_1 = 3, X_2 = 4,$  and  $X_3 = 5,$  then,

$$\Sigma cX_i = 10 \times 3 + 10 \times 4 + 10 \times 5 = 30 + 40 + 50 = 120.$$

### Summation ( $\Sigma$ ) and Two Variables ( $X$ and $Y$ )

- a.  $\Sigma(X - Y)^2$  (Sum of Squared Deviations of  $X$  and  $Y$ )
- b.  $\Sigma XY$  (Sum of Products of  $X$  and  $Y$ )
- c.  $\Sigma X \Sigma Y$  (Product of Summations of  $X$  and  $Y$ )

Table 1-1 provides the calculations for a., b., and c. for two variables  $X$  and  $Y$  with three values.