

**A CONCISE, NON-MATHEMATICAL  
BEGINNER'S GUIDE TO  
PRINCIPAL COMPONENTS &  
CLUSTER ANALYSIS WITH EXCEL**

*Ready-to-use Excel Templates Included*



---

**A Concise, Non-Mathematical  
Beginner's Guide to  
Principal Components and  
Cluster Analysis with Excel**

*Ready-to-use Excel Templates Included*

---

**Kilem L. Gwet, Ph.D.**

AgreeStat Analytics  
Maryland, USA

Copyright © 2024 by Kilem Li Gwet, Ph.D. All rights reserved.

**Published by AgreeStat Analytics; in the United States of America.**

No part of this book may be reproduced or transmitted in any form or by any means, whether electronic or mechanical, including photocopying, recording, or via any information storage and retrieval system, without prior written permission from the publisher. Exceptions are made for reviewers who may quote brief passages in a review for publication in a magazine or newspaper. For permissions and other inquiries, please contact:

**AgreeStat Analytics**

Maryland 20886-4379

e-mail: gwet@agreestat.com

This publication is intended to provide accurate and authoritative information on the subject matter covered. However, it is sold with the understanding that the publisher assumes no responsibility for any errors, inaccuracies, or omissions. The publisher is not engaged in providing professional services. Readers are advised to seek the assistance of a competent professional for expert advice.

**Publisher's Cataloguing in Publication Data:**

Gwet, Kilem Li

*A Concise, Non-Mathematical Beginner's Guide to Principal Components and Cluster Analysis with Excel*

Ready-to-use Excel Templates Included / By Kilem Li Gwet

p. cm.

Includes bibliographical references and index.

1. Statistics
2. Statistical Analysis
3. Data Analysis
4. Statistics - Study - Learning. I. Title.

---

# Preface

---

This book explores two distinct yet related multivariate data analysis techniques: Principal Component Analysis (PCA) and Cluster Analysis. Both methods have found extensive applications in fields like machine learning and artificial intelligence. PCA enables you to simplify complex, multidimensional problems by reducing them to a few manageable dimensions, often allowing for visual representation of the data. On the other hand, Cluster Analysis lets you divide a heterogeneous set of units into more homogeneous groups, known as clusters, based on several attributes associated with these units. These techniques are particularly effective when applied to datasets containing numerous units with multiple numeric attributes.

In this book, I provide a high-level overview of these two techniques, with a focus on helping you develop a deep understanding of the concepts. I also guide you on how to use existing Excel-based tools to perform these analyses, even if your knowledge of Excel is basic.

Why cover both PCA and Cluster Analysis in one book? In my view, a robust multivariate analysis should begin with PCA before employing other techniques. PCA can greatly enhance the implementation of subsequent methods, including Cluster Analysis. I will demonstrate how performing PCA beforehand can significantly improve the efficiency of the  $k$ -Means procedure. By working with the reduced dimensions provided by PCA, you can make an informed estimate of the number of clusters needed and optimize the clustering algorithm.

The literature on Principal Component Analysis is extensive. Introduced by Pearson in 1901 (see [Pearson, 1901](#)) and later rediscovered by Hotelling in 1933 (see [Hotelling, 1933](#)), PCA was initially known as the Hotelling transformation. Although many computer scientists are just now discovering this technique in

the era of machine learning, it has been a cornerstone of statistical analysis for over a century. Similarly, Cluster Analysis was introduced early in the 20th century, primarily in anthropology, and has since become a staple in marketing for customer segmentation. Today, it is routinely used by computer scientists in artificial intelligence.

## About the Author

I earned my Ph.D. in Mathematics from Carleton University's School of Mathematics and Statistics in Ottawa, Canada, in 1997, with a specialization in the design and analysis of statistical surveys. In recent years, I have dedicated considerable time and energy to the field of inter-rater reliability analysis, publishing several papers and books on the subject (you can find some of my works at [https://www.researchgate.net/profile/Kilem\\_Gwet](https://www.researchgate.net/profile/Kilem_Gwet)). My successful use of PCA in analyzing multivariate inter-rater reliability data inspired me to write this book. My goal was to create a step-by-step guide on how to calculate principal components, understand what they represent, how to use them, and what their limitations are. I hope you find this book valuable.

If you have any comments or questions, feel free to contact me at the email address [gwet@agreestat.com](mailto:gwet@agreestat.com). I will do my best to respond as promptly as I possibly can.

---

Kilem Li Gwet, Ph.D.

# Contents

<b>Acknowledgment</b>	<b>ix</b>
<b>Part I: Principal Component Analysis</b>	<b>xi</b>
<b>1 An Overview of Principal Components</b>	<b>2</b>
1.1 <i>Introduction</i>	3
1.2 <i>Illustrating Composite Score Variables</i>	5
1.3 <i>Geometric Framework for Principal Component Analysis</i>	11
1.4 <i>Concluding Remarks</i>	14
<b>2 Computing Principal Components</b>	<b>16</b>
2.1 <i>Introduction</i>	17
2.2 <i>Description of PCA Results</i>	21
2.3 <i>Using the Excel Template Multivariate.xlsm</i>	25
2.4 <i>Concluding Remarks</i>	31
<b>3 Using Principal Components</b>	<b>38</b>
3.1 <i>Introduction</i>	39
3.2 <i>Interpreting Principal Components</i>	40
3.2.1 <i>Optimal Number of Principal Components</i>	40
3.2.2 <i>Principal Component Loadings</i>	43
3.3 <i>Applications of Principal Components</i>	46
3.3.1 <i>Identifying Relevant Variables with Principal Components</i>	50
3.3.2 <i>Identifying Outliers with Principal Components</i>	53
3.4 <i>Concluding Remarks</i>	59

---

<b>Part II: k-Means Cluster Analysis with Excels</b>	<b>60</b>
<b>4 Introduction to Cluster Analysis</b>	<b>62</b>
4.1 <i>Introduction</i>	63
4.1.1 <i>When to Use Cluster Analysis?</i>	64
4.1.2 <i>Applications of Cluster Analysis</i>	66
4.1.3 <i>Types of Clustering Methods</i>	68
4.2 <i>Clustering and the Notion of Proximity</i>	69
4.3 <i>k-Means Clustering: How Does it Work?</i>	72
4.4 <i>Clustering with Excel</i>	76
4.4.1 <i>Iris15 Dataset &amp; Manual Implementation of Cluster Analysis</i>	77
4.4.2 <i>Iris Dataset &amp; Cluster Analysis Automation</i>	77
4.5 <i>Concluding Remarks</i>	78
<b>5 Implementing k-Means Clustering in Excel</b>	<b>80</b>
5.1 <i>Introduction</i>	81
5.2 <i>Manual k-Means Clustering of the Iris15 Dataset</i>	82
5.2.1 <i>Description of the Iris15 Dataset</i>	83
5.2.2 <i>Initial Cluster Centers</i>	85
5.2.3 <i>The Iterative Process</i>	89
5.2.4 <i>Visual Exploration of Clusters</i>	93
5.3 <i>Cluster Analysis with Multivariate.xlsm</i>	97
5.4 <i>Concluding Remarks</i>	109
<b>A Working with Excel Data Tables</b>	<b>110</b>
A.1 <i>What is an Excel Data Table</i>	111
A.2 <i>How to Create a Data Table in Excel?</i>	112
A.3 <i>Creating a Named Range in Excel</i>	116
<b>Bibliography</b>	<b>122</b>
<b>Author Index</b>	<b>124</b>
<b>Subject Index</b>	<b>125</b>



# Acknowledgment

I want to express my heartfelt gratitude to my family for their unwavering support throughout the creation of this book. I am especially thankful to my wife, Suzy, and our three daughters, Mata, Lelna, and Addia. They have graciously tolerated my endless hours at the computer, the long workdays, and the busy weekends over the past few years. Without their patience and understanding, this book would not have been possible.

Kilem Li Gwet, Ph.D.  
Maryland, USA: August 2024



# PART I

## PRINCIPAL COMPONENT ANALYSIS (PCA)

### List of Part I Chapters

---

<i>Chapter#</i>	<i>Title</i>	<i>Page#</i>
1	An Overview of Principal Components.....	2
2	Computing Principal Components.....	16
3	Using Principal Components.....	38

