INTER-RATER RELIABILITY ANALYSIS USING SAS®

Other Books by the author:

- HANDBOOK OF INTER-RATER RELIABILITY, (5^{th} Edition)
  The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters
  VOLUME 1: Analysis of Categorical Ratings
  ISBN: 1792354630 / 978-1792354632

- HANDBOOK OF INTER-RATER RELIABILITY, (5^{th} Edition)
  The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters
  VOLUME 2: Analysis of Quantitative Ratings
  ISBN: 1792354649 / 978-1792354649

- Beginner's Guide to Principal Components: Applications with Microsoft Excel
  ISBN: 0970806256 / 978-0970806253

https://agreestat.com/books/

INTER-RATER RELIABILITY ANALYSIS USING SAS®

A Practical Guide for Analyzing, Categorical and Quantitative Ratings

Kilem Li Gwet, Ph.D.

AgreeStat Analytics
P.O. Box 2696
Gaithersburg, MD 20886-2696
USA

Copyright © 2021 by Kilem Li Gwet, Ph.D. All rights reserved.

Published by AgreeStat Analytics. Printed and bound in the United States of America.

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by an information storage and retrieval system – except by a reviewer who may quote brief passages in a review to be printed in a magazine or a newspaper – without permission in writing from the publisher. For information, please contact Advanced Analytics, LLC at the following address:

AgreeStat Analytics
PO BOX 2696,
Gaithersburg, MD 20886-2696
e-mail: contact@agreestat.com

This publication is designed to provide accurate and authoritative information in regard of the subject matter covered. However, it is sold with the understanding that the publisher assumes no responsibility for errors, inaccuracies or omissions. The publisher is not engaged in rendering any professional services. A competent professional person should be sought for expert assistance.

Publisher's Cataloguing in Publication Data:

Gwet, Kilem Li

Inter-Reliability Analysis using SAS®

A Practical Guide for Analyzing Categorical and Quantitative Ratings/ By Kilem Li Gwet

p. cm.

Includes bibliographical references and index.

1. Biostatistics
2. Statistical Methods
3. Statistics - Study - Learning. I. Title.

ISBN 978-1-7923-7488-3

Preface

I wrote this book, primarily to assist researchers and students with the calculation of various inter-rater reliability coefficients using the SAS system. For categorical ratings, you will learn how to use SAS for computing various chance-corrected agreement coefficients (CAC) such as Cohen's kappa, Gwet's AC_1/AC_2, Krippendorff's alpha and many others. For quantitative ratings, you will learn how SAS can be used to compute different variants of the Intraclass Correlation Coefficient (ICC).

The use of SAS in this book is basic. Therefore, you do not need expert-level knowledge of SAS to find this book useful. However, I expect you to have some familiarity with the SAS environment, to be able to run an existing SAS macro and to use function modules included in a SAS/IML library. Even if your knowledge of SAS is limited, you will still be able to download all the programs discussed in this book, and to follow step-b-step instructions so that you can reproduce the same results. If you do not have access to SAS , you may still be able to use the free cloud-based SAS® OnDemand for Academics. For this, you may create a SAS profile to access this very user-friendly platform using the following link:

https://welcome.oda.sas.com/

Whichever version of SAS you are using, you will want to ensure that you have access the IML software. Some of the solutions recommended in this book require the use of SAS/IML functions, as previously mentioned.

This book does not present a rigorous mathematical description of the different inter-rater reliability coefficients discussed. Interested readers may want to see Gwet (2021a, volume 1) for a comprehensive treatment of chance-corrected

agreement coefficients to analyze categorical ratings, or Gwet (2021b, volume 2) for a formal treatment of intraclass correlation coefficients used in the analysis of quantitative ratings. The focus of this book is on computational methods with SAS .

For now, FREQ and SURVEYFREQ are the only built-in SAS procedures with limited capability for analyzing categorical ratings. Most practitioners will only need the FREQ procedure for their analysis. Both procedures produce the exact same agreement coefficients. However, if your ratings come from a survey, which is based on a complex sampling design[1], then only SURVEYFREQ will give you valid standard error estimates. SURVEYFREQ accounts for the complexity of the sample to compute the correct standard error whereas FREQ always assumes that all subjects had the same chance of being selected.

For SAS/STAT® version 15.2 or later, these 2 procedures allow you to compute the unweighted and weighted versions of Cohen's kappa as well as the unweighted Gwet's AC_1 and the unweighted Prevalence-Adjusted Bias-Adjusted Kappa coefficient or PABAK. However, the number of raters that can be analyzed is limited to 2 and the 3 agreement coefficients are used to analyze categorical ratings only. Consequently, if you want to analyze multiple raters, use alternative agreement coefficients other than kappa, PABAK and AC_1, or analyze quantitative ratings that require the use of intraclass correlations, then you will need a special SAS macro function or a SAS/IML function module. In this book, I will present SAS macro functions and SAS/IML libraries that you will be able to use to achieve all these goals.

This book is divided into 5 chapters. Chapter 1 provides a general overview of what SAS has to offer in the area of inter-rater reliability analysis. I will point out some advantages as well as some disadvantages for using the FREQ procedure. Chapter 2 describes different approaches for analyzing categorical ratings from a two-rater inter-rater reliability study. In chapter 3, I discuss

[1] A complex sampling design involves sample selection schemes such stratification or clustering where some subjects have a higher chance of being selected to participate in a study than others.

various approaches for analyzing categorical ratings from multiple raters. SAS macro functions as well as new SAS/IML libraries of function modules will be presented. Chapter 4 focuses on the analysis of quantitative ratings. I will present a SAS macro and other SAS/IML functions that allow you to compute various intraclass correlations. Finally, chapter 5 will review miscellaneous techniques for analyzing categorical ratings. These techniques include comparing the difference of 2 agreement coefficients for statistical significance or the benchmarking of agreement coefficients to qualify the strength of agreement among raters.

Kilem Li Gwet, Ph.D.

Contents