

NEASE LIBRARY EASTERN NAZARENE COLLEGE



3 2999 00208 6713

MULTIVARIATE DATA ANALYSIS

FIFTH
EDITION



HAIR ■ ANDERSON ■ TATHAM ■ BLACK

Acquisitions Editor: Whitney Blake
Assistant Editor: John Larkin
Editorial Assistant: Rachel Falk
Vice President/Editorial Director: James Boyd
Marketing Manager: John Chillingworth
Marketing Director: Brian Kibby
Production Editor: Aileen Mason
Production Coordinator: Carol Samet
Managing Editor: Dee Josephson
Associate Managing Editor: Linda DeLorenzo
Manufacturing Supervisor: Arnold Vila
Manufacturing Manager: Vincent Scelta
Design Manager: Pat Smythe
Interior Design: Brian Deep
Cover Design: Jayne Conte
Illustrator (Interior): ElectraGraphics, Inc.
Composition: York Graphic Services, Inc.
Cover Photo: William Whitehurst/The Stock Market

Copyright © 1998, 1995, 1992, 1987, 1984 by Prentice-Hall, Inc.
Upper Saddle River, New Jersey 07458

All rights reserved. No part of this book may be reproduced, in any form or by
any means, without written permission from the Publisher.

Library of Congress Cataloging-in-Publication Data

Multivariate data analysis / Joseph F. Hair, Jr. . . . [et al.].
p. cm.

Rev. ed. of: **Multivariate data analysis with readings.** 4th ed. c1995.

Includes **bibliographical** references (p. -) and index.

ISBN 0-13-894858-5

1. **Multivariate** analysis. I. Hair, Joseph F. II. **Multivariate data analysis**
with **readings.**

QA278.M85 1998

519.5'35—dc21

97-47031

CIP

Prentice-Hall International (UK) Limited, London
Prentice-Hall of Australia Pty. Limited, Sydney
Prentice-Hall Canada, Inc., Toronto
Prentice-Hall Hispanoamericana, S.A., Mexico
Prentice-Hall of India Private Limited, New Delhi
Prentice-Hall of Japan, Inc., Tokyo
Prentice-Hall Asia Pte. Ltd., Singapore
Editora **Prentice-Hall** do Brasil, Ltda., Rio de Janeiro

Printed in the United States of America

15 14 13 12 11

without theoretical support. In those instances the researcher is running a much greater risk of finding results with little or no generalizability. Instead, in using these additional variables, the researcher must be guided by theory that is supported by empirical analysis. In this manner, both practical and statistical significance can hopefully be met.

Stage 3: Assumptions in Multiple Regression Analysis

We have shown how improvements in prediction of the dependent variable are possible by adding independent variables and even transforming them to represent aspects of the relationship that are not linear. But to do so we must make several assumptions about the relationships between the dependent and independent variables that affect the statistical procedure (least squares) used for multiple regression. In the following sections we discuss testing for the assumptions and corrective actions to take if violations occur.

Assessing Individual Variables versus the Variate

The assumptions underlying multiple regression analysis apply both to the individual variables (dependent and independent) and to the relationship as a whole. Chapter 2 examined the available methods for assessing the assumptions for individual variables. But in multiple regression, once the variate has been derived, it acts collectively in predicting the dependent variable. This necessitates assessing the assumptions not only for individual variables but also for the variate itself. This section focuses on examining the variate and its relationship with the dependent variable for meeting the assumptions of multiple regression. These analyses actually must be performed *after* the regression model has been estimated in stage four. Thus, testing for assumptions must occur not only in the initial phases of the regression, but also after the model has been estimated.

The basic issue is whether, in the course of calculating the regression coefficients and predicting the dependent variable, the assumptions of regression analysis have been met. Are the errors in prediction a result of an actual absence of a relationship among the variables, or are they caused by some characteristics of the data not accommodated by the regression model? The assumptions to be examined are as follows:

- Linearity of the phenomenon measured
- Constant variance of the error terms
- Independence of the error terms
- Normality of the error term distribution

The principal measure of prediction error for the variate is the **residual**—the difference between the observed and predicted values for the dependent variable. Plotting the residuals versus the independent or predicted variables is a basic method of identifying assumption violations for the overall relationship. When examining residuals, some form of standardization is recommended, as it makes the residuals directly comparable. (In their original form, larger predicted values naturally have larger residuals.) The most widely used is the **studentized resid-**