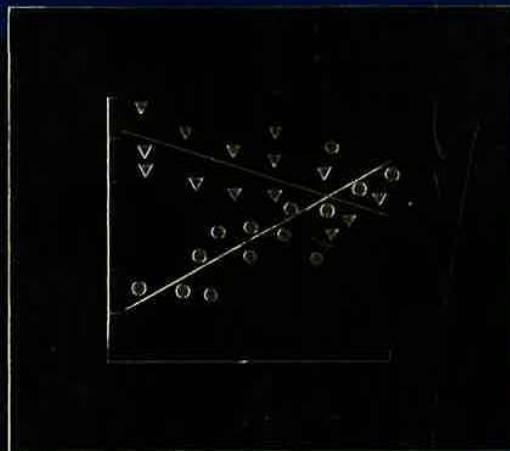


APPLIED LINEAR STATISTICAL MODELS

F O U R T H E D I T I O N



NETER KUTNER NACHTSHEIM WASSERMAN

To
Dorothy, Ron, David,
Nancy, Michelle, Allison,
Maureen, Abigail, Andrew,
Cathy, Christopher, Timothy, Randall, Erin, Fiona



IRWIN Concerned about Our Environment

In recognition of the fact that our company is a large end-user of fragile yet replenishable resources, we at IRWIN can assure you that every effort is made to meet or exceed Environmental Protection Agency (EPA) recommendations and requirements for a "greener" workplace.

To preserve these natural assets, a number of environmental policies, both companywide and department-specific, have been implemented. From the use of 50% recycled paper in our textbooks to the printing of promotional materials with recycled stock and soy inks to our office paper recycling program, we are committed to reducing waste and replacing environmentally unsafe products with safer alternatives.

WCB/McGraw-Hill

A Division of The McGraw-Hill Companies

© The McGraw-Hill Companies, Inc., 1974, 1983, 1990, and 1996

All rights reserved. No part of this publication 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 written permission of the publisher.

Trademark Acknowledgments:

IBM and IBM PC are registered trademarks of International Business Machines Corporation.

DISCLAIMER

Richard D. Irwin, Inc., makes no warranties, either expressed or implied, regarding the enclosed computer software package, its merchantability or its fitness for any particular purpose. The exclusion of implied warranties is not permitted by some states. The exclusion may not apply to you. This warranty provides you with specific legal rights. There may be other rights that you may have which may vary from state to state.

Irwin Book Team

Publisher: Tom Casson
Senior **sponsoring** editor: Richard T. Hercher, Jr.
Senior **developmental** editor: Gail Korosa
Senior Marketing manager: Brian Kibby
Project editor: Lynne Basler
Production supervisor: Laurie Kersch

Designer: Matthew Baldwin
Cover designer: Stuart Paterson
Assistant Manager, Graphics: Charlene R. Perez
Compositor/Art Studio: Graphic Composition, Inc.
Typeface: 10/12 Times **Roman**
Printer: R. R. Donnelley & Sons Company

Library of Congress Cataloging-in-Publication Data

Applied linear statistical models / John Neter . . . [et al.]. — 3rd ed.
p. cm.

Includes index.

ISBN 0-256-11736-5

1. Regression analysis. 2. Analysis of variance. 3. Experimental design. 4. Linear models (Statistics)

QA278.2.A66 1996

519.5'35—dc20

95-37447

Printed in the United States of America

13 14 15 16 17 DOW /DOW 0 9 8 7 6 5 4 3

Multiple Regression—I

Multiple regression analysis is one of the most widely used of all statistical methods. In this chapter, we first discuss a variety of multiple regression models. Then we present the basic statistical results for multiple regression in matrix form. Since the matrix expressions for multiple regression are the same as for simple linear regression, we state the results without much discussion. We conclude the chapter with an example, illustrating a variety of inferences and residual analyses in multiple regression analysis.

6.1 Multiple Regression Models

Need for Several Predictor Variables

When we first introduced regression analysis in Chapter 1, we spoke of regression models containing a number of predictor variables. We mentioned a regression model where the response variable was direct operating cost for a branch office of a consumer finance chain, and four predictor variables were considered, including average number of loans outstanding at the branch and total number of new loan applications processed by the branch. We also mentioned a tractor purchase study where the response variable was volume of tractor purchases in a sales territory, and the nine predictor variables included number of farms in the territory and quantity of crop production in the territory. In addition, we mentioned a study of short children where the response variable was the peak plasma growth hormone level, and the 14 predictor variables included gender, age, and various body measurements. In all these examples, a single predictor variable in the model would have provided an inadequate description since a number of key variables affect the response variable in important and distinctive ways. Furthermore, in situations of this type, we frequently find that predictions of the response variable based on a model containing only a single predictor variable are too imprecise to be useful. We noted the imprecise predictions with a single predictor variable in the Toluca Company example in Chapter 2. A more complex model, containing additional predictor variables, typically is more helpful in providing sufficiently precise predictions of the response variable.

In each of the examples just mentioned, the analysis was based on observational data because the predictor variables were not controlled, usually because they were not susceptible to direct control. Multiple regression analysis is also highly useful in experimental