

EXHIBIT 2

APPLIED STATISTICS FOR PUBLIC POLICY

BRIAN P. MACFIE and PHILIP M. NUFRIO

M.E. Sharpe
Armonk, New York
London, England

Brief T

Copyright © 2006 by M.E. Sharpe, Inc.

All rights reserved. No part of this book may be reproduced in any form
without written permission from the publisher, M.E. Sharpe, Inc.,
80 Business Park Drive, Armonk, New York 10504.

Library of Congress Cataloging-in-Publication Data

Macfie, Brian P., 1955–

Applied statistics for public policy / by Brian P. Macfie and Philip M. Nufrio.
p. cm.

Includes bibliographical references and index.

ISBN 0-7656-1239-9 (cloth : alk. paper)

1. Social sciences— Statistical methods. 2. Political statistics. I. Nufrio, Philip M. II. Title.

HA29.M185 2005

519.5 —dc22

2004023626

Printed in the United States of America

The paper used in this publication meets the minimum requirements of
American National Standard for Information Sciences
Permanence of Paper for Printed Library Materials.
ANSI Z 39.48-1984.

∞

BM (c) 10 9 8 7 6 5 4 3 2 1

Preface

1. Introduct

2. Using PO

3. Presentat

4. Summari

5. Basic Pro

6. Sampling

7. The Cent

8. Introduct

9. Estimatin

10. Validatin

11. Validatin

12. Validatin

13. Validatin

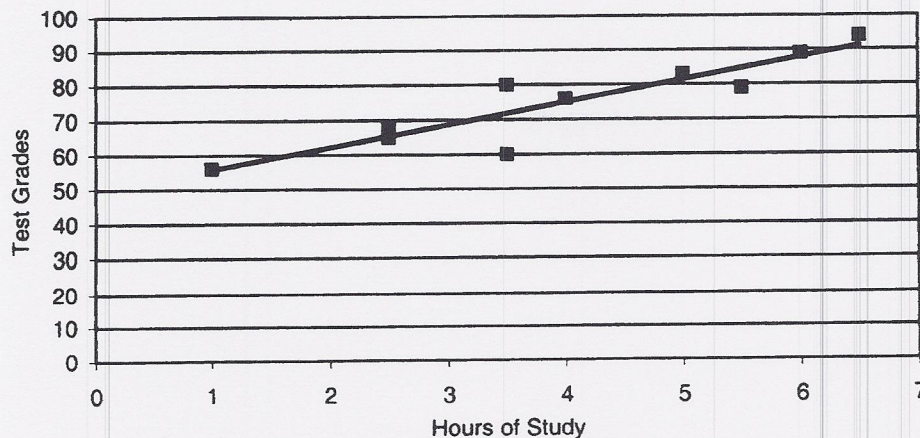


Figure 17.1 Grades and Study Times

31, the director may use some factors or variables that are associated with rainfall (e.g., today's barometric pressure, humidity, cloud cover, etc.) to predict how much rain might be expected. Although correlation analysis will tell us if rainfall is related to barometric pressure, humidity, or cloud cover, it does not predict the amount of rainfall at a given barometric pressure, humidity, or cloud cover. Regression analysis will provide such a prediction.

How to Conduct Single Variable (Simple) Linear Regression

To see how we are going to do this, start by going back to the two hypotheses stated in Chapter 16 (with regard to the relationship between test grades and the number of hours studied, and test grades and the number of hours worked). Recall that in the first case (of hours studies), we said that this was a positive relationship because $r = +0.903$ and that the two variables also had a strong association because $r^2 = 0.815$. In the latter case (of hours worked), we said that this was a negative relationship because $r = -0.914$ and that two variables had a strong association because $r^2 = 0.835$. Although this is certainly important information, these statistics are not going to be much help with the following question. What is the expected test grade for an individual who studies 5 hours a week *or* what is the expected test grade for an individual who works 15 hours a week?

The XY scatter plots in Figures 17.1 and 17.2 duplicates Figures 16.1 and 16.2, respectively, with one addition—we have added a trend line to describe the data. This trend line is used to summarize or model the linear relationship between the two variables using all the data points. This trend line is created using regression analysis.

The question is, where did this trend line come from? It is actually a straight-line equation estimated from a sample of data. Recall from first-year algebra in high school, a straight-line equation is written as:

$$\hat{Y} = b_0 + b_1 X_i \quad (17.1)$$

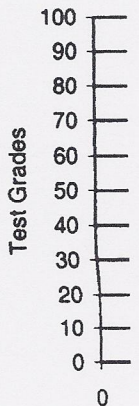


Figure 17.2

where $\hat{Y} = p$
 $b_0 = i$
 $b_1 = s$
 $X_i = v$

The straight-line trended, the slope represents the c/Y will change 1. is defined as how way to express t change in X .

The b_0 coefficient value of Y when y -axis (hence, it

Finally, \hat{Y} (ca for any value of or may not be ec

There is one straight-line equation of either coefficient dependent variable

Least Squares

The general idea best fits the data reasons. First, p

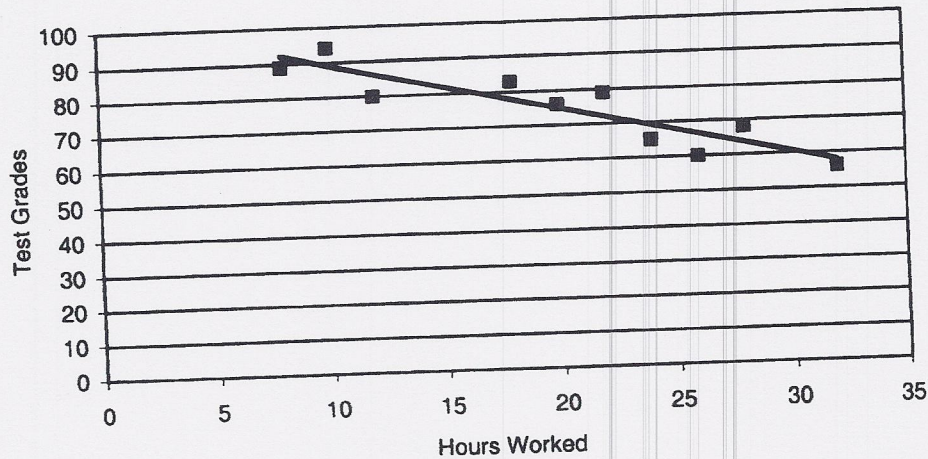


Figure 17.2 Grades and Hours Worked

where \hat{Y} = predicted value of the dependent variable (Y),
 b_0 = intercept coefficient of a linear equation,
 b_1 = slope coefficient of a linear equation, and
 X_i = values for the independent variable (X).

The straight-line equation (Equation (17.1)) fits the trend lines in Figures 17.1 and 17.2. As illustrated, the slope or "pitch" of the line is denoted by the b_1 coefficient in the equation. The slope represents the change in Y per a one-unit change in X . For example, if $b_1 = 1.5$, we are saying that Y will change 1.5 units for each one-unit change in X . Stated mathematically, the slope of any line is defined as how much the line rises or falls relative to the distance it travels horizontally. Another way to express this formula is that the slope of a line is equal to the ratio of the change in Y given change in X .

The b_0 coefficient in the equation denotes the intercept. The intercept represents the average value of Y when X equals 0. Mathematically, this is the point where the straight line intercepts the y -axis (hence, it is called the intercept).

Finally, \hat{Y} (called "Y hat") is the symbol for the predicted value of the dependent variable (Y) for any value of X that is used in the linear equation with the intercept b_0 and the slope b_1 . It may or may not be equal to the actual value of Y .

There is one thing to keep in mind about the regression coefficients (b_0 and b_1). Although the straight-line equation is written in Equation (17.1) with all positive (+) signs, the estimated values of either coefficient (b_0 or b_1) could actually be negative, depending on the association with the dependent variable (\hat{Y}).

Least Squares Estimation

The general idea in linear regression analysis is to estimate a straight-line equation model that best fits the data. Linear functions have great appeal in empirical social science work for two reasons. First, previous research has shown that many functions in the social sciences are linear