EXHIBIT I

APPLIED STATISTICS FOR PUBLIC POLICY

BRIAN P. MACFIE and PHILIP M. NUFRIO

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

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

Brief Ta

Random Sampling

Most of the book has dealt with samples that were assumed to be randomly drawn. Random sampling is probably one of life's most misunderstood concepts. A random sample is one in which each element in the population theoretically has an equal chance of being selected for study. The world is replete with examples of samples that are not random. A random sample is not a group of readers who mail in a clip-out survey to a local newspaper. It is not the individuals that call into a radio station to voice their opinion. It is not the people who register their response to a question on some Internet site. It is also not the group of people who a TV station interviews on the street. Part of the problem with each of these examples is that individuals can *self-select* themselves to participate. For a sample to be truly random, the probability that each member of the population can be chosen has to be equal. Moreover, no one in the population may be systematically excluded.

Thus, if an analyst is conducting an opinion survey and is trying to get a random sample of the population of property owners in the local county, they cannot simply go to the local mall to select adults as they walk by (actually this is called convenience sampling; that is, the selected participant is at the right place at the right time). In addition to not knowing whether the people selected live in the county or even own property, this would exclude property owners not choosing to go to the mall that day.

Unless the entire population is available for selection, the sample is not random. In the case of trying to obtain a random sample of property owners in the county, it would be more random if the analyst started with a list of all property owners recorded at the county clerk's office. The analyst could then clip out each name separately, place them all in a box, and then blindly select each name until a large enough sample is collected (what constitutes a sufficiently sized sample was covered in Chapter 9). The analyst would then have go out and find each property owner selected. This is extremely important, for a sample can never be random if the subjects are allowed to select themselves. For example, if the analyst simply mailed a survey to the entire list of property owners (which could be costly) and 100 residents returned completed surveys, this is not a random sample because, in effect, the residents self-selected themselves on the basis of which ones felt duty-bound enough to return the surveys. Those property owners who did return the survey may have differed from the rest of the property owners who chose to ignore the survey.

Perhaps the most famous example of a true random sample is the one conducted by the U.S. Selective Service System for the draft (see sheet 40 on the data disk). Every date of the year was put in a fishbowl, and a prespecified number of dates were pulled out. If the birthdate for the draft-eligible individuals was selected, they were drafted. Since everyone's birth data was included, the selection was random. (On the lighter side, also note that the government did not wait for individuals to self-select themselves and come to them; if the individual did not come in, the government went looking.)

Another efficient random sample that is conducted every year is the one used by the Internal Revenue Service (IRS). On an annual basis, the IRS randomly selects about 2 percent of all income tax returns for further review. The selection is done using an algorithm with a random number generator against social security numbers. This is also strictly random because a copy of every taxpayer's return is coded into the computer. Again, note that the IRS also does not wait for the sample to self-select itself.

Stratifi

Another po
sampling. B
sample is th
way that, on
This is then
know in adv
these traits i
$175,000, th
Political
cide what cl
income, age,
als falling w
percentages.
predict the b

System

The last prob
ordered listir
available. Th
as the value c
sampling qua
every item h

Conven

As mention e
bilistic. As su
person who w
the right plac

Quota S

Quota samplii
selecting ever
the understan
pling). For ex
of 200 is need
slight improve