STATISTICS for People Who (*Think They*) HATE STATISTICS

2 EDITION

NEIL J. SALKIND
University of Kansas

SAGE Publications
International Educational and Professional Publisher
Thousand Oaks ▪ London ▪ New Delhi

Copyright © 2004 by Sage Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

SPSS® software screens are reproduced with the permission of SPSS, Inc.

Cartoons copyright 2004 by Jason Love. For permission to reprint, contact the artist at: www.jasonlove.com

---

For information:

Sage Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

Sage Publications Ltd.
6 Bonhill Street
London EC2A 4PU
United Kingdom

Sage Publications India Pvt. Ltd.
B-42 Panchsheel Enclave
Post Box 4109
New Delhi 110017
India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Salkind, Neil J.
Statistics for people who (think they) hate statistics / Neil J. Salkind--2nd ed.
p. cm.
Includes index.
ISBN 0-7619-2788-3 (cloth) — ISBN 0-7619-2776-X (paper)

1. Statistics. I. Title.

HA29.S2365 2004
519.5--dc22

2003019122

05 10 9 8 7 6

Acquiring Editor: Lisa Cuevas Shaw
Editorial Assistant: Margo Crouppen
Production Editor: Diana E. Axelsen
Copy Editor: Liann Lech
Typesetter/Designer: Tim Giesen/Straight Line Design
Indexer: Mary Mortensen
Cover Designer: Michelle Lee Kenny

Yippee!

1. Sta

Σigma F

2. Me
3. *Viv*
4. A F
5. Ice

Taking

6. Hyp
7. Are

19 The Ten (or More) Best Internet Sites for Statistics Stuff

In the first edition of *Statistics for People Who (Think They) Hate Statistics*, we told readers like you that if you're not yet using the Internet as a part of your learning and research activities, you are missing out on an extraordinary resource. Today, more than ever, students, researchers, and others certainly are taking advantage of this vast resource, but there's still some reluctance on the part of newbies.

We all should recognize that what's on the Internet will not make up for a lack of studying or motivation—nothing will do that—but you can certainly find a great deal of information that will enhance your whole college experience. And this doesn't even begin to include all the fun you can have!

So, now that you're a certified novice statistician, here are some Internet sites that you might find very useful should you want to learn more about statistics. Some are the same from last time, and some are new—have fun.

Although the locations of Web sites on the Internet are more stable than ever, they still can change frequently. The URL (uniform resource locator) that worked today might not work tomorrow. It's for this reason that you can find all of these Web sites (corrected and up-to-date) on the Web site for *Statistics for People Who (Think They) Hate Statistics*, which you can find at http://www.soe.ku.edu/faculty/Salkind/stats_fpwhs. Just go there and look for the resources link.

TONS AND TONS OF RESOURCES

Here's the mother lode. Pages and pages of every type of statistical resource you can want has been creatively assembled by Professor

David W. Stockburger at http://www.psychstat.smsu.edu/scripts/dws148f/statisticsresourcesmain.asp. This site receives the gold medal of statistics sites. Don't miss it.

For example, take a look at Berrie's page (at http://www.huizen.dds.n~berrie/) and see some QuickTime (short movies) of the effects of changing certain data points on the value of the mean and standard deviation. Or, look at the different home pages that have been created by instructors for courses offered around the country. Or, look at all of the different software packages that can do statistical analysis (I consulted this list as I wrote Chapter 18).

CALCULATORS GALORE!

Want to draw a histogram? How about a table of random numbers? A sample-size calculator? The Statistical Calculators page at http://www.stat.ucla.edu/calculators/ has just about every type (more than 15) of calculator and table you could need. Enough to carry you through any statistics course that you might take and even more.

For example, you can click on the Random Permutations link and complete the two boxes (as you see in Figure 19.1 for 2 random permutations of 100 integers), and you get the number of permutations you want. This is very handy when you need a table of random numbers for a specific number of participants so you can assign them to groups.

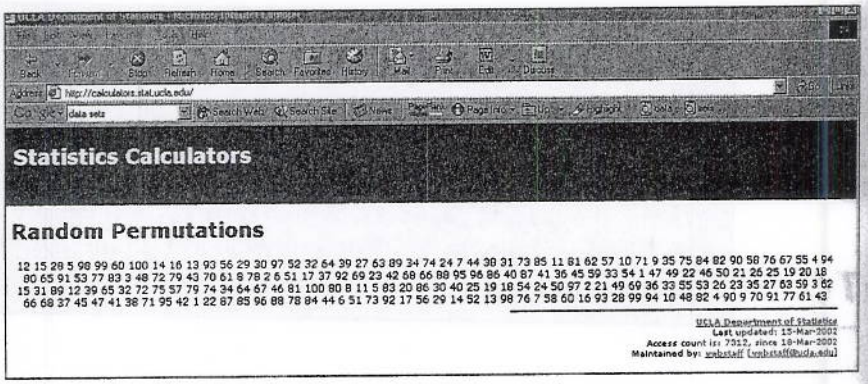

Figure 19.1. Generating a Set of Random Numbers

WHO'S WHO AND WHAT'S HAPPENED

The History of Statistics page located at http://www.Anselm.edu/homepage/jpitocch/biostatshist.html contains portraits and bibliographies of famous statisticians and a time line of important contributions to the field of statistics. So, do names like Bernoulli, Galton, Fisher, and Spearman pique your curiosity? How about the development of the first test between two averages during the early 20th century? It might seem a bit boring until you have a chance to read about the people who make up this field and their ideas—in sum, pretty cool ideas and pretty cool people.

IT'S ALL HERE

SurfStat Australia (at http://www.anu.edu.au/nceph/surfstat/surfstat-home/surfstat.html) is the online component of a basic stat course taught at the University of Newcastle, Australia, but has grown far beyond just the notes originally written by Annette Dobson in 1987, and updated over several years' use by Anne Young, Bob Gibberd, and others. Among other things, SurfStat contains a complete interactive statistics text. Besides the text, there are exercises, a list of other statistics sites on the Internet, and a collection of Java applets (cool little programs you can use to work with different statistical procedures).

HYPERSTAT

This online tutorial with 18 lessons, at http://www.davidmlane.com/hyperstat/index.html, offers nicely designed and user-friendly coverage of the important basic topics. What we really liked about the site was the glossary, which uses hypertext to connect different concepts to one another. For example, in Figure 19.2, you can see the definition of descriptive statistics also linked to other glossary terms, such as mean, standard deviation, and box plot. Click on any of those and zap! you're there.

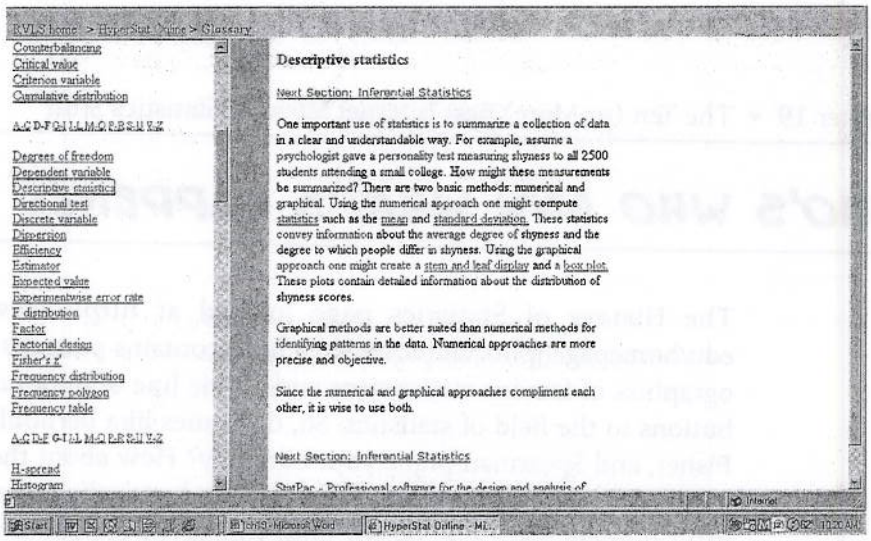

Figure 19.2. Sample HyperStat Screen

DATA? YOU WANT DATA?

There are data all over the place, ripe for the picking. Here are just a few. What to do with these? Download them to be used as examples in your work or as examples of analysis that you might want to do, and you can use these as a model.

- Statistical Reference Datasets at http://www.itl.nist.gov/div898/strd/
- United States Census Bureau (a huge collection and a gold mine of data) at http://factfinder.census.gov/servlet/DatasetMainPageServlet?_lang=en
- The Data and Story Library (http://lib.stat.cmu.edu/DASL/) with great annotations about the data (look for the stories link)
- Tons of economic data sets at Growth Data Sets (at http://www.bris.ac.uk/Depts/Economics/Growth/datasets.htm)

Then there are all the data sets that are available through the federal government (besides the census). Your tax money supports it, so why not use it? For example, there's FEDSTATS (at http://www.fedstats.gov/), where more than 70 agencies in the U.S. federal government produce statistics of interest to the public. The Federal Inter-

agency Council on Statistical Policy maintains this site to provide easy access to the full range of statistics and information produced by these agencies for public use. Here you can find country profiles contributed by the (boo!) CIA; public school student, staff, and faculty data (from the National Center for Education Statistics); and the Atlas of the United States Mortality (from the National Center for Health Statistics). What a ton of data!

MORE AND MORE AND MORE AND MORE RESOURCES

The University of Michigan's Statistical Resources on the Web (at http://www.lib.umich.edu/govdocs/stats.html) has hundreds and hundreds of resource links, including those to banking, book publishing, the elderly, and, for those of you with allergies, pollen count. Browse, search for what exactly it is that you need—no matter, you are guaranteed to find something interesting.

PLAIN, BUT FUN

At http://mathforum.org/workshops/sum96/data.collections/datalibrary/data.set6.html, you can find a data set including the 1994 National League Baseball Salaries or the data on TV, Physicians, and Life Expectancy. Nothing earth-shaking, just fun to download and play with.

HOW ABOUT STUDYING STATISTICS IN STOCKHOLM?

The World Wide Web Virtual Library: Statistics is the name of the page, but the one-word title is misleading because the site (from the good people at the University of Florida at http://www.stat.ufl.edu/

vlib/statistics.html) includes information on just about every facet of the topic, including data sources, job announcements, departments, divisions and schools of statistics (a huge description of programs all over the world), statistical research groups, institutes and associations, statistical services, statistical archives and resources, statistical software vendors and software, statistical journals, mailing list archives, and related fields. Tons of great information is available here. Make it a stop along the way.

MORE AND MORE AND MORE RESOURCES—AGAIN!

Statistics on the Web at http://www.maths.uq.edu.au/~gks/webguide/datasets.html is another location that's just full of information and references that you can easily access. Here, you'll find information on professional organizations, institutes and consulting groups, educational resources, Web courses, online textbooks, publications and publishers, statistics book lists, software-oriented pages, mailing lists and discussion groups, and even information on statisticians and other statistical people.

ONLINE STATISTICAL TEACHING MATERIALS

If you do ever have to teach statistics, or even tutor fellow students, this is one place you'll want to visit: http://noppa5.pc.helsinki.fi/links.html. It contains hundreds of resources on every topic that was covered in *Statistics for People Who (Think They) Hate Statistics* and more. You name it and it's here: regression, demos, history, Sila (a demonstration of inference), an interactive online tutorial, statistical graphics, handouts to courses, teaching materials, journal articles, and even quizzes! Whew, what a deal. There tends to be a lot of material that may not be suited to what you are doing in this class, but this wide net has certainly captured some goodies.

MORE AND MORE AND MORE STUFF

Statistics.com (www.statistics.com) has it all—a wealth of information on courses, software, statistical methods, jobs, books, and even a homework helper. For example, if you want to know about free Web-based stat packages, click on that link on the left-hand side of the page. Here's one (see Figure 19.3) from Dr. Bill Trochim that is similar to the flowcharts we used at the beginning of several chapters in Part IV of *Statistics for People* . . . . You just click your way through answering questions to get the answer to what type of analysis should be used.

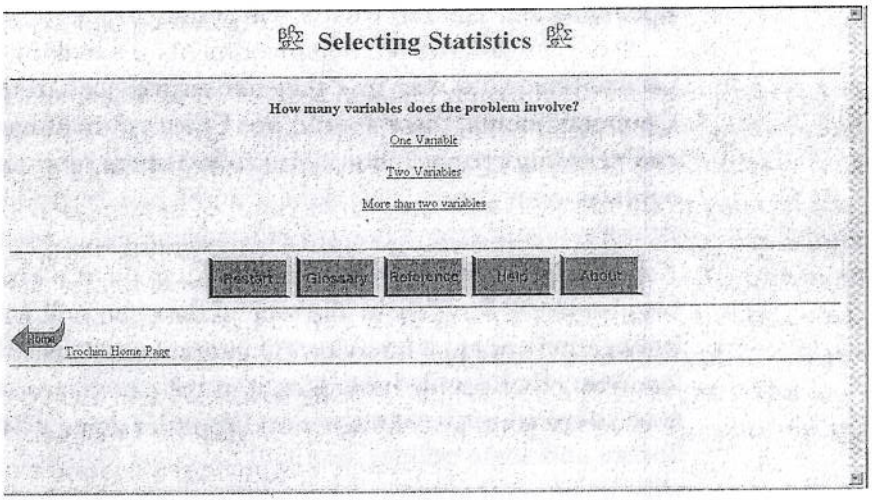

Figure 19.3. Selecting the Correct Stat Technique to Use—Just a Few Clicks Away

20 The Ten Commandments of Data Collection

Now that you know how to analyze data, you would be well served to hear something about collecting them. The data collection process can be a long and rigorous one, even if it involves only a simple, one-page questionnaire given to a group of students, parents, patients, or voters. The data collection process may very well be the most time-consuming part of your project. But as many researchers do, this period of time is also used to think about the upcoming analysis and what it will entail.

Here they are: the ten commandments for making sure your data get collected in a way that they are usable. Unlike the original Ten Commandments, these should not be carved in stone (because they can certainly change), but if you follow them, you can avoid lots of aggravation.

Commandment 1. As you begin thinking about a research question, also begin thinking about the type of data you will have to collect to answer that question. Interview? Questionnaire? Paper and pencil? Find out how other people have done it in the past by reading the relevant journals in your area of interest and consider doing what they did.

Commandment 2. As you think about the type of data you will be collecting, think about where you will be getting the data. If you are using the library for historical data or accessing files of data that have already been collected, such as census data (available through the U.S. Census Bureau and some online), you will have few logistical problems. But what if you want to assess the interaction between newborns and their parents? The attitude of teachers toward unionizing? The age at which people over 50 think they are old? All of these questions involve needing people to provide the answers, and finding people can be tough. Start now.

Commandment 3. Make sure that the data collection forms you use are clear and easy to use. Practice on a set of pilot data so you can make sure it is easy to go from the original scoring sheets to the data collection form.

Commandment 4. Always make a duplicate copy of the data file, and keep it in a separate location. Keep in mind that there are two types of people: those who have lost their data and those who will. Keep a copy of data collection sheets in a separate location. If you are recording your data as a computer file, such as a spreadsheet, be sure to make a backup!

Commandment 5. Do not rely on other people to collect or transfer your data unless you have personally trained them and are confident that they understand the data collection process as well as you do. It is great to have people help you, and it helps keep the morale up during those long data collection sessions. But unless your helpers are competent beyond question, you could easily sabotage all your hard work and planning.

Commandment 6. Plan a detailed schedule of when and where you will be collecting your data. If you need to visit three schools and each of 50 children needs to be tested for a total of 10 minutes at each school, that is 25 hours of testing. That does not mean you can allot 25 hours from your schedule for this activity. What about travel from one school to another? What about the child who is in the bathroom when it is his turn, and you have to wait 10 minutes until he comes back to the classroom? What about the day you show up and Cowboy Bob is the featured guest . . . and on and on. Be prepared for anything, and allocate 25% to 50% more time in your schedule for unforeseen happenings.

Commandment 7. As soon as possible, cultivate possible sources for your subject pool. Because you already have some knowledge in your own discipline, you probably also know of people who work with the type of population you want or who might be able to help you gain access to these samples. If you are in a university community, it is likely that there are hundreds of other people competing for the same subject sample that you need. Instead of competing, why not try a more out-of-the-way (maybe 30 minutes away) school district or social group or civic organization or hospital, where you might be able to obtain a sample with less competition?

Commandment 8. Try to follow up on subjects who missed their testing session or interview. Call them back and try to reschedule. Once you get in the habit of skipping possible participants, it becomes too

easy to cut the sample down to too small a size. And you can never tell—the people who drop out might be dropping out for reasons related to what you are studying. This can mean that your final sample of people is qualitatively different from the sample you started with.

Commandment 9. Never discard the original data, such as the test booklets, interview notes, and so forth. Other researchers might want to use the same database, or you may have to return to the original materials for further information.

And Number 10? Follow the previous 9. No kidding!