

Uplink

April 1997

A newsletter for the National Institute for Computer-Assisted Reporting

The stats on CAR Uplink Update

Staying in tune with the May 4-9 advanced CAR seminar in Chapel Hill, N.C., this month's issue hits you with a heavy dose of statistics.

As the world of numbers continues to descend upon reporters at an often dizzying rate, having a little knowledge about histograms or medians is becoming increasingly essential.

With the help of an esteemed collection of journalists who have "been there and done that," this issue attempts to make the load a little easier to bear.

Dave Davis and Joan Mazzolini of *The Plain Dealer* show how they used stats to find disturbing trends in the organ donor business. And NICAR's training director, Sarah Cohen, debuts her new monthly column devoted to stats.

Feeling like this issue may be a bit too techy? Don't worry, stats guru Neill Borowski helps ease us into the topic with a story on the basics of statistics.

Inside

NEW

Monthly Statistics Column

Page 6

Tech Tip

Page 8

IRE/NICAR Notes

Page 11

Handout of the Month

Page 12

Statistics and Medicine

Lonely Hearts

By Dave Davis and Joan Mazzolini
The Plain Dealer

Imagine you're so sick you can't walk from the bedroom to your kitchen and you gasp for every breath. Imagine you go to the doctor, and she tells you your heart is damaged, so damaged you'll need a transplant or you'll die within six months.

Then imagine traveling to the hospital every week, even every day for tests — blood tests, lung capacity tests, and every known heart test.

Imagine you've spent thousands of dollars, and your insurer has spent tens of thousands of dollars, and that your family is praying you get on the waiting list and get a donor organ.

Imagine you go through all of this, but there is really no chance you'll ever get a transplant. Your hospital has stopped performing them, but doctors didn't tell you or any of the other patients for fear you might go elsewhere.

It sounds horrible, but this is exactly what happened to patients awaiting heart transplants at University of Kansas Medical Center in Kansas City and William Beaumont Hospital in suburban Detroit.

In each case, doctors at the hospitals turned away for nonmedical reasons donor organs they were offered over the course of about a year. But they didn't tell patients.

In fact, they let patients believe it was the national organ shortage, and not their own stalled programs, that was causing the unending wait.

In a series published in February by

The Plain Dealer, we used computer-assisted reporting techniques to examine in detail transplant programs that turned away organs that were matched to waiting patients.

We also followed a 37-year-old Pennsylvania woman through the ordeal of a heart transplant, and entered waiting times into FoxPro to examine the great disparity in waiting times at different

Continued on page two

For the Novice Getting Back to Basics

By Neill A. Borowski
The Philadelphia Inquirer

At the outset, we have to agree on something: Much of the time we really aren't using "statistics." We're mostly in the business of measuring populations.

This distinction is important.

Many statistical tools are designed to help the researcher use samples to make inferences about the population.

In computer-assisted reporting and analysis, we usually have the population in front of us. There's no guessing what the full universe of data looks like. We have it all: Every teacher's salary in our state, every fatal traffic accident in the

Continued on page three

Uplink

April 1997

Volume 9, Number 4
A newsletter for the
National Institute for
Computer-Assisted
Reporting

Editor

Brant Houston
brant@nicar.org

Managing Editor

Stan Dorsey
stan@nicar.org

Senior Contributing

Editors

Jo Craven
jo@nicar.org
Andrew Lehren
andy@nicar.org
Richard Mullins
richard@nicar.org

Staff

Sarah Cohen
Seth Hemmelgarn
Daniel Kim
Justin Mayo
Jody Sowell
John Sullivan
Alistair White
Paul Walmsley
Wallace Winfrey
Copy Editors
Wendy Charron
Marilyn Lehren

Uplink is published every month by the National Institute for Computer-Assisted Reporting, 138 Neff Hall Annex, Columbia, MO 65211. (573) 882-0684.

Subscription price is \$40 for IRE members, \$60 for nonmembers.

Postmaster: Please send address changes to NICAR. Send e-mail to nicar@muccmail.missouri.edu.

NICAR is a joint effort of Investigative Reporters and Editors and the University of Missouri School of Journalism. NICAR services include hands-on newsroom training on computer-assisted reporting, special academic and advanced training in data analysis. NICAR is supported by a grant from The Freedom Forum and other foundations intended to help the institute deliver its services nationwide to news organizations and associations.

Continued from page one:

Heart transplants

centers, and with the help of SPSS analyzed the relationship between the average annual number of transplants centers performed and their mortality rate.

Looking for patterns

Volume, or the average annual number of transplants a center performed, was one common thread which ran through each of the stories. For example, centers that performed a small number of transplants each year almost always had fewer resources in terms of medical staff experienced in transplantation. We found they tended to be the centers that would most often turn away donor organs for non-medical reasons.

We were especially intrigued by this question of program volume because, for each type of major organ transplant, half the centers in the country had very small annual transplant numbers. For example, about half the centers in the country did fewer than nine heart transplants a year.

Researching the question

Since donor organs are scarce, we wondered if lives were being lost because a limited number of organs were being spread among too many transplant centers. It was this question, among others, that lured us into a statistical analysis.

Using the National Library of Medicine's Medline database we searched the medical literature online to see what had been published concerning transplant center volume and patient mortality. (You can get a Medline user ID online at <http://igm.nlm.nih.gov> or call 1-800-638-8480.)

We didn't find much, just one article from the *Journal of the American Medical Association* in which the authors found that heart transplant centers doing fewer than nine transplants a year had significantly higher one-year mortality rate.

Honing in with stats

With this information in hand, we sought help from Phil Meyer and John Bare at the University of North Carolina-Chapel Hill. It was Bare who did the heavy statistical lifting, while Meyer watched over our progress with

the big picture in mind: Credible results.

We decided on a two-prong analysis. First, we wanted to determine whether low-volume transplant centers had a higher observed one-year mortality rate than high-volume centers. If so, we wanted to find out if the increased rate of death could be explained by differences in patient risk factors and donor characteristics, or whether it was partly attributable to transplant center volume.

The first part was straightforward data analysis. The second part, however, was trickier. It required using a statistical method known as logistic regression, which is commonly used in medical research to adjust data for certain patient risks.

We had patient-level data on transplant donors and recipients — one record for each of the 55,000 transplants that had been performed from 1987 through 1991. The records also contained a key field identifying the transplant center.

For each organ type, we computed the average annual number of transplants the centers performed. We then labeled each center either "high volume" or "low volume" depending upon whether they were above or below the average volume.

We found that the observed 12-month mortality rate was higher at all times for low-volume centers. In other words, patients who underwent a transplant at a low-volume center had a greater chance of dying in the first year following the transplant.

This held true for all six major types of organs.

Next, using logistic regression, we adjusted for medical differences in donors and recipients and found that center volume was a significant predictor of mortality at one-year.

Though she did not use CAR, Cheryl Thompson of the *Kansas City Star* uncovered similarly troubling trends at KUMC in a series of investigative stories published in May of 1995. That series, which revealed the hospital had stopped performing transplants and was turning away healthy hearts while continually adding patients to their waiting list, is available in the IRE Resource center.

For a copy of the volume analysis technical documentation or a reprint of the entire transplant series, call Davis at 1-800-688-4802 or send email to ddavis00@aol.com.

Continued from page one:

Getting started

nation, or every sentence imposed by a judge.

By over-testing our data, we may be complicating and confusing our analysis. Avoid using statistical tools “because they’re there.” Always keep it simple.

With that said, here are some fundamental steps to measure your data. Let’s say our database is all of the teacher salaries in a state:

Always, always plot your data

This probably should be your first step after making sure the database is clean. Plot a histogram (for a description of a histogram, see the page margin to the right). It will show how your data are distributed. A normal distribution will be mound-shaped. Look for oddities in your data. Are there a large number of salaries far off to the right of your histogram? This could be a highly paid district.

Calculate the ‘basics’

These are the range, arithmetic mean, the median and the mode (remember “ramm”). Lets say we’re looking at salaries. The mean — often called the average — is the sum of the salaries divided by their count. The median is the salary in the middle after the salaries are ranked (when there is an even count of salaries, there will be two salaries in the middle; the mean of these two salaries is the median). The mode is the salary that appears most often. The range is the highest salary minus the lowest salary.

Rank, rank and compare

Who gets the most? Group the districts and calculate their averages and medians. Which districts pay the most? How do these compare with the state average?

Use the total experience (in years) of each teacher to calculate the salary per experience year. Rank these. Do you have the salary from five years ago? Calculate the percentage change and rank the change. What district salaries changed the most? The least? For me, one of the handiest tools in software is the =rank function in MS Excel. It will tell you where things stand in a flash.

Look at the spread

Two sets of data could have the same means, but would look far different when they are

plotted. The average may be \$45,000 for two sets of salaries. However, one may have a low salary of \$30,000 and a high salary of \$65,000. The other data set may have a low of \$8,000 and a high of \$90,000.

The range will help. But the standard deviation is a better tool to help with this. Keep in mind that there are two standard deviations — the sample and the population standard deviations. (In Excel, the sample standard deviation function is =stdev and the population standard deviation function is =stdevp).

The standard deviation offers a quick view of how the data are distributed around the mean. Rule of thumb: The mean plus and minus one standard deviation includes about 68% of the data and the mean plus and minus two standard deviation includes about 95% of the data. What are the extreme values or outliers? Check out how many salaries are beyond three standard deviations. How many go beyond two standard deviation? These may be newsworthy.

Study the scatterplot

We’re not explaining regression analysis in this article (regression looks at the impact one variable, such as experience level, has on another variable, such as salaries). However, you can quickly look at the “big picture” through a scatterplot.

The “x” (horizontal) axis is the independent variable — in this case, total years of experience for each teacher. The “y” (vertical) axis is the dependent variable — the salary of each teacher. A scatterplot will show you the regression line and the data points scattered about the line. The points farthest from the line are outliers.

A regression line in a scatterplot of teacher salaries would be upward sloping — it would go from the lower left (low experience/low salary) to the upper right (high experience/high salary). If there are points in the upper-left quadrant of your plot far from the line, it might indicate teachers who are paid a lot with very little experience. And if there are points in the lower-right quadrant, it might indicate teachers with a lot of experience but at low pay.

Neill Borowski can be reached at (609) 779-3884, or send email to neill.borowski@phillynews.com

A histogram looks like a column chart. However, a column in a column chart might, for example, reflect average teacher salaries (y axis) by school district (x axis). A histogram portrays the frequencies of the variable intervals we’re looking at — in this case, teacher salaries. So the x axis would be intervals of salaries, such as \$10,001 to \$15,001 to \$20,000 and so on. The y axis — which governs the height of the column — simply counts the number of salaries that fall within each of the intervals. If there are 900 teachers making \$10,001 to \$15,000 and 1,200 making \$15,001 to \$20,000 it would begin a histogram “stairstep.” In a normal distribution, the histogram will rise to a peak in the middle and fall off (mound shaped).

A trip to the bookstore

Nora Paul has dedicated this month's On the Internet (page 10) column to finding data sets and statistics guides on the Internet. Be sure to check it out to complete your statistics resource guide.

Many of these resources can be ordered online through the Ernie Pyles Memorial Bookstore on the web at <http://www.duff.net/ernie/>. For a non-framed version go to <http://www.duff.net/ernie/stats.html>. The site also contains numerous other journalism-related books for order.

Recently on the NICAR-L listserv, a series of messages arose pertaining to the subject of statistics resources for reporters. Many of the list members shared their knowledge and experiences with certain texts. Following are some of those books, along with members' comments and how to email them:

The *SPSS Guide to Statistical Analysis* by Maria Norusis may be the best primer around. My two additional references are: *PDQ Statistics* by Norman and Streiner, and *Statistics: A Spectator Sport* by Richard Jaeger. Both attempt to make the link between this subject and the layman, or at least the lay nerd.

Ford Fessenden, *Newsday*
fordfess@newsday.com

The Jessica Utts text, *Seeing Through Statistics...* is the perfect book for CAR types. It teaches you why to use stats, without much math. This book is clearly written, including lots of examples (many taken from news media). It will work well for those who want to learn this on their own, outside of a classroom setting.

I also recommend a program called Data Desk. It's built around the idea of "exploratory data" analysis, which uses some slick data visualization tools such as rotating 3-d plots, brushing, color ranges, et al.

Stephen K. Doig
steve.doig@asu.edu

For the lighter side of statistical reading, try these:

•*How to Lie With Statistics*, the classic little book first published in 1954, by Darrell Huff. I've still seen it in college bookstores. The pictures alone are worth the price.

•*200% of Nothing: From Percentage Pumping to Irrational Ratios, An Eye-Opening Tour through the Twists and Turns of Math Abuse and Innumeracy*, by A.K. Dewdney, John Wiley and Sons. This one isn't really statistics per se, but it's good reading for reporters who are finding themselves working with numbers of any kind more often.

Sarah Cohen, NICAR training director
sarah@nicar.org

I like Victor Cohn's *News & Numbers*. Cohn was a *Washington Post* science and medical reporter. Much of the book is devoted to deciphering statistical claims by scientists, doctors, environmentalists and others in the health arena, but he also addresses basic concepts too like variance, rates, significance, and so forth. Very readable.

Rick Linsk, *Asbury Park Press*
linsk@app.com

Philip Meyer's book *The New Precision Journalism* is an excellent resource. It has a statistics section, and deals with poll and survey design, data collection and analysis.

Bill Doskoch
bdoskoch@compusmart.ab.ca

It never hurts to collect as many useful stats books as possible. Some specific suggestions:

•*Statistical Concepts: A Basic Program* by Brown, Amos and Mink. You'll probably have to go to a good used bookstore for this one. It does a terrific job of teaching basic stats. It starts with explaining terms such as variable, frequency, distribution, etc., in definitions that go a sentence or two.

•*Dictionary of Statistics and Methodology*, by W. Paul Vogt., from Sage Publications.

•*The Practice of Social Research*, by Earl Babbie is a good introductory methods textbook. It goes through theory/research, measurement, operationalization, etc.

•*Statistical Reasoning for the Behavioral Sciences*, by Richard Shavelson is another textbook used in graduate stats introductory courses.

•*Political Science Research Methods*, by Johnson and Joslyn, is a book someone passed along to me. The writing is generally clear, and the book is short on Greek letters. I always like to find stats books that explain things in sentences, not just equations, and this one does that.

•*Understanding Political Variables*, by Buchanan, is a book I don't use a lot, but it's generally readable.

•*Basic Statistics: An Introductory Survey for Students of Mathematics, Social Sciences, and Business*, by Parl. Again, unless there's a new edition, check the dusty shelves of a local used bookstore.

John Bare
jbare@interpath.com

Power to the people

By Janet Roberts

Wilmington (N.C.) Morning Star

Janet Roberts attended NICAR's beginner bootcamp in March 1996, when she was the newspaper's assistant city editor. Since then, she attended the intermediate bootcamp in January and was named computer resources editor.

When an incumbent Republican school board member slugged the husband of another GOP incumbent in the parking lot after a candidates' forum, we knew last May's Republican primary election was like none we had seen before.

One faction of the New Hanover County Republican Party was calling for a redistricting plan based on neighborhood schools. So what if that would mean a return to segregation and the loss of federal money for our school system?

The county's more mainstream Republicans said they wanted neighborhood schools, too, but they weren't willing to throw out 20 years of civil rights gains to achieve them.

The Morning Star's news columns were full of daily stories about the debate. But none of those stories shed much light on the key question: How was the fight playing out with voters?

Often, when newspapers want the answer to that question, they turn to old-fashioned shoe-leather journalism. You knock on doors.

For a small newspaper, the problem with that approach is the time it takes. You knock on lots of doors and you don't always find registered voters or people who are giving the election much thought. The crapshoot doesn't always result in an enlightening story.

Here's where the computer came in. We thought we could use it to take some of the chance out of the process.

Who we found

How did it work?

In four days' time, we found and interviewed a few dozen voters who had been following the school board race closely and had strong opinions about which side was right.

We interviewed and photographed one of two black Republicans registered in one of our target precincts. Imagine the likelihood of finding him had we just gone out randomly knocking on doors.

The story got beyond the candidates' rhetoric and painted the issues in real terms, telling of

real parents with real kids in school. Their opinions of what course the school system should take likely helped other voters decide how they wanted to vote. We gave it showcase play on our Sunday cover, two days before the primary.

How we did it

This was by no means a scientific process. Rather, we used the computer to help sharpen our natural reporting instincts.

Excel helped us zero in on precincts where we wanted to interview people. Paradox pinpointed voters who would make good interviews.

We decided to look for Republican precincts where voters tend to be independent-minded: They don't just flip the Republican lever; they sometimes vote for Democrats. Our theory was that these independent-minded voters probably give their decisions a lot of thought, were likely following the party dispute closely and would help turn the race for one faction or another. We also wanted precincts with a lot at stake, where kids would have to be bused to maintain racial balance in the schools.

Using simple sorts in Excel, we zeroed in on a handful of precincts that matched our criteria. We sorted first by demographics, finding the Republican-dominated precincts. Then we looked for voting trends, finding one precinct, for example, where Republicans had helped a black Democrat win the last general election.

From the few interesting precincts we found, we decided to focus on three — a blue-collar neighborhood, a white upper-class neighborhood, and a suburb where a lot of kids are bused.

This is where Paradox came in. We had a database of all registered voters in New Hanover County. It lists names, addresses, party affiliation, date of birth, sex, race. Even more helpful to our story: It tells if they voted in the past several elections and even lists some telephone numbers. For each of the three precincts, we asked

Correction

There was an error in the formula for February's lead story about the Gini coefficient.

The correct formula is:

```
Gini coefficient=1 - SUM
((Xi-Xj)*(Yi+Yj))
where:
X is the cumulative
proportion of households;
Y is the cumulative
proportion of income;
i is a particular income
category;
j is i-1, or the preceding
income category.
*Xi, Xj, Yi, Yj must equal
one or less
```

Continued on page seven

Stats from the road

By Sarah Cohen
NICAR Training Director

statistics: numerical data assembled and classified so as to present significant information

— Webster's New World Dictionary

When scores of otherwise street-wise reporters show up at conference panels with titles like "significance testing" and "regression techniques", it's clear that something is happening within the world of journalism.

And when an Associated Press bureau chief actually asks for a lecture on the use of numbers in the newsroom, it's yet another sign that numbers — yes, statistics — are becoming an important element in the beat reporter's toolbox.

Most statistics courses and books begin with a simple premise that is just not true for most reporters. They assume the bulk of our work comes from surveys, not from comprehensive reports that are collected from everyone.

We don't take a sample of hospital discharges; we look at all 40 million of them. Instead of surveying borrowers to find out if they feel discriminated against, reporters usually look to existing reports that compile information on all loan applications in their areas. In other words, instead of designing the perfect survey, reporters must decide whether to attempt a study at all, or if they should fight for records that may be of use.

We also need some simple measures that don't bog down our readers or our stories. Sure, experts will claim there's no analysis that will compensate for all the flaws in medical, education, or lending data. Fair enough. But our job is to use present information, not hide it, even when it's somewhat flawed. Our challenge is to decide when it's fatally flawed, or how to present it most fairly.

This month I'll focus on one of the tools that has become a staple in journalism — the humble crosstab. Crosstabs can reveal many relationships without clouding them through complicated formulas.

How? By giving you an apples-to-apples comparison and by providing you with information that conforms to the way reporters

think — in neat categories like high and low, big and small, yes and no.

A crosstab will do the job quickly and cleanly.

We're going to use Philip Meyer's rule for crosstabs to get started: Put the "independent variable" in your crosstab as a column, and put the "dependent variable" as a row. Now generate the percents of the column, and you have a neat little table that you can describe easily to readers, editors and even your next-door neighbors.

Step 1: Figure out your categories

Some categories are natural: race, sex or vote. Others require a little work. One common category is high- versus low-income. You can make this sensitive to your area or your numbers by using the quartiles in your data as cutoffs for your categories. Another approach is to split detailed categories, like those found in Census data, into natural groups. An example is the highest level of school people in an area completed: No high school, high school, some college, college, graduate school. Now you have six groups where you began with 20 or so possible values for the number of years of school completed.

Make sure every category is included in your analysis. This may mean doing a little arithmetic, or thinking of your data in a new way.

How can you get these groups? If you have SPSS, it's easy — just run Transform, Ranks, then choose quartiles instead of rank. It will give you a 1 for the lowest quartile, through 4 for the highest, as a new variable.

If you're using Excel, it takes a few more steps. The simplest way is simply to sort the values and assign them in order. You can also use the PERCENTRANK function, and round it to the next lowest quartile by using the FLOOR function. This won't be exact, but it will be close.

Step 2: Decide which field, or variable, is independent, and which is dependent.

One clue is to look for which one came first in time. The choice of insurance (or the lack of it) probably came before the birth of your newborn. The ethnic group membership (black, white or Hispanic) obviously came before income. And the income probably came before the decision to apply for a mortgage. When you have this kind of obvious sequence, the later event could depend

Continued on page seven

Have a problem you would like tackled in this space? Perhaps you know of a technique that you've found valuable in your work? Maybe there is something you've heard of, but would like to see discussed in greater detail? Let us know about it. Beginning this month Uplink will publish a monthly column intended to help reporters sort out the numbers they are facing every day. Written by NICAR's on-the-road training director, Sarah Cohen, the column will focus on the most basic definition of statistics — the attempt to present numbers in some meaningful way.

Continued from page six:

Using the crosstab

on the earlier event. It's dependent. Remember, statistics can't tell you this. So it involves some thinking before you get started.

Step 3: Produce the crosstab (in SPSS) or a PivotTable (in Excel)

This will count up the number of people in each cell of a square. If you've set up your crosstab with the independent variable in columns, then you just need to ask for the column percents in a crosstab or PivotTable. In SPSS, the option is under the "Cell Statistics" button. In Excel, it's under "Options" when you double-click on the data field. Here is an example of a crosstab, using an Excel PivotTable, for 2,800 mortgage applications in 1995 from high-income home-buyers in Nashville, Tenn.

	A	B	C	D	E	F
1	IncGrp	High				
2						
3			ETHNIC1			
4	GIVEN	Data	White	Black	Grand Total	
5	Denied	#	314	41	355	
6		% Col	12%	31%	13%	
7	Approved	#	2332	93	2425	
8		% Col	88%	69%	87%	
9	Total #		2646	134	2780	
10	Total % Col		100%	100%	100%	
11						

It doesn't take a lot of sophistication to see that the rate of denials for black applicants in this high-income group is about three times higher than the rate for white applicants. That's because you simply have to glance across the line you care about to see the difference, adjusted for the number of applications from each group.

Is this legitimate? Did it happen by chance?

We'll tackle the use of significance testing in another column. But the short answer is: It happened. For the high-income applicants in Nashville in 1995, it may not matter whether it was a fluke. We haven't tried to make anything bigger of this analysis than it is. It's a description of the facts for that year and that place, not an indication of something bigger.

Step 4 (or is it step 1?): Don't leave your brain behind

You can make all kinds of comparisons in tables like this, but please don't leave behind what you already know. For instance, think about the relationship between income and education. If you test it, you'll probably find big news: People with less education have higher incomes! The reason is the "lurking variable" that so many people forget about. In this case, it's age. People reach their prime earning power in their 40s and 50s. But people are also becoming more educated.

So just like we had to control this analysis for income, we also have to control other analyses for items like age, education, poverty or a whole rash of other factors. That's why this last step might easily be your first: If you don't know what you're looking for, then read the academic literature, talk to other reporters who have covered the issue and talk with local experts before you even start.

Sarah Cohen can be reached at (301) 942-2199 or e-mail her at sarah@nicar.org.

NICAR and IRE have many books available for purchase.

Among those available are the Reporter's Handbook (3rd edition) by Steve Weinberg and The New Precision Journalism by Philip Meyer.

For information on other books and ordering and pricing information call Wendy Charron at (573) 882-0684 or send her email at wendy@nicar.org

Continued from page five:

CAR trips

Paradox to give us a list of all the registered Republicans who had voted in the '94 primary. That eliminated the apathetic voters from the lists and gave us a better chance of landing decent interviews. I printed out these lists, and reporters Scott Whisnant, Tricia Vance and Phil Hervey took it from there, blitzing on the phones to conduct the interviews.

The follow-up

We used the same technique the day after the

election for our post-primary analysis. We used Excel to find the swing precincts, and then culled lists of voters from our voter registration database in Paradox. This all resulted in a story where voters explained the outcome of the race rather than tired pundits.

What I liked most about these stories is that they relied heavily on number crunching and data but were vastly devoid of numbers.

Janet Roberts can be reached at (910) 343-2389, or send e-mail to starnews@wilmington.net.

FoxPro wrote me a letter

By Jo Craven
NICAR staff

For our recent conference in Nashville, NICAR gathered the country's top journalists in computer-assisted reporting and asked them to jot down their ideas, tips, and practical anecdotes. The result is a binder full of handouts and tipsheets that among other things cover math and stats, suggest CAR stories for sports reporters, describe how to diagnose hospitals, and detail how to investigate immigration issues. The sheets are currently available through the IRE Resource Center. They can be ordered individually, or you can purchase the entire set for \$75. Eventually, this resource will be available electronically.

Why work if you can get a computer to do the job for you?

Recently, when I needed to send a form letter to 133 different businesses, I thought of NICAR data guru Richard Mullins, who teaches this ethic. I could sit and type 133 different addresses on top of 133 different copies of the same form letter — or I could get a computer to do it for me. I decided to let the computer do the job.

Here's what I had available: a FoxPro database that included, among other things, the names and addresses of the businesses that I wanted to contact.

I knew I could write a scanloop in FoxPro to pull out the addresses. And I found out that I could use the TEXT/ENDTEXT function to insert a text block — in this case, my form letter — after each scanned address. Later, I found I could get FoxPro to use the same database to print out mailing labels. What would have taken me a couple of hours, FoxPro did in minutes. (Incidentally, you could accomplish this same task with Microsoft Word's Mail Merge without writing a FoxPro program).

To create the FoxPro program, which I called "letters," in the command window, I typed:

```
modify command letters

In the .prg, or program, window that appeared, I wrote the following program (an explanation follows):

CLEAR
CLEAR ALL
CLOSE ALL

SET CENTURY ON
SET TALK OFF
SET STATUS BAR OFF

USE c:\business.dbf

files = " "

SCAN

files = alltrim(str(business.id))

SET ALTERNATE TO "c:\" + files
SET ALTERNATE ON

?alltrim(cmonth(date())) + " " +
alltrim(str(day(date()))) + ", " +
alltrim(str(year(date())))
?" "
```

```
?name
?address
?alltrim(city)+", "+state+" "+zip
?" "
```

TEXT

To Whom It May Concern:

This is where I typed my letter.

Sincerely,
Jo Craven

ENDTEXT

CLOSE ALTERNATE

ENDSCAN

SET TALK ON

SET STATUS BAR ON

My database, business.dbf, had 133 records. This program uses an alias called "files" to create 133 different text files, each containing the form letter topped with the different business addresses.

To make this work, I needed each record to have a unique ID of no more than eight characters. My data did not have such a field, so I created one using the recno() function: I modified the database by adding a field called "id" and told FoxPro to:

replace all id with recno()

This numbered each record, providing the necessary unique ID. Here is the program again, including explanations of its components:

```
CLEAR
CLEAR ALL
CLOSE ALL
```

- These three lines clear the slate, making sure that no other files are open when the program starts. They are standard in any FoxPro program.

```
SET CENTURY ON
```

- This sets the dateline that starts each letter.

```
SET TALK OFF
```

```
SET STATUS BAR OFF
```

- These turn off FoxPro's bells and whistles in order to make the program run faster.

```
USE c:\business.dbf
```

- This tells the computer which database to use — the one with the business names and addresses

```
files = " "
```

Continued on page nine

Continued from page eight:

Tech tip for letters

• This reveals that an alias called "files" will be used. It is defined and plugged into the program later. This line can be omitted; including it is a matter of form.

SCAN

• This begins the scanloop. In FoxPro, a scanloop tells the computer to look at every record in a database and perform whatever functions are specified within the scanloop — in this case, setting up separate text files, typing in the various names and addresses from the database, and inserting the text block written within the scanloop.

```
files = alltrim(str(business.id))
```

• This defines the alias "files." It says that, for this program, "files" equals "id," which is the field in business.dbf containing the unique IDs. "Alltrim" makes sure blank spaces don't confuse the machine. And "str" tells the computer to consider the numeric, unique IDs as character data.

```
SET ALTERNATE TO "c:\\" + files
SET ALTERNATE ON
```

• This is where the individual text files are created. "Set Alternate To" tells the computer that everything it does within this scanloop should be copied to another file on the c:\ drive. "Files" has already been defined as "id," so for every unique ID, a new text file will be created on the c:\ drive. "Set Alternate On" tells the computer to act on the "Set Alternate To" command. "Set Alternate To" won't work without "Set Alternate On."

```
?alltrim(cmonth(date())) + " " +
alltrim(str(day(date()))) + ", " +
alltrim(str(year(date())))
```

• This makes the date. If your computer's date and time are set correctly, this command will produce the current date. The question mark (?) tells FoxPro to type the information.

```
? " "
```

• This produces a blank line between the date and the address.

```
?name
?address
?alltrim(city)+", "+state+" "+zip
?" "
```

• This tells FoxPro to pull out the information in the fields "name," "address," "city," "state," and "zip," inserting the comma where appropriate, and spacing between the state and ZIP code.

(Remember, we already told the computer which database to use — business.dbf) The last line leaves a blank space between the address and the letter below.

TEXT

To Whom It May Concern:

This is where I typed my letter. Your letter can be as long, or short, as you like.

Sincerely,
Jo Craven

ENDTEXT

• TEXT tells FoxPro a text block is coming and should be inserted in the scanloop after the address. Anything can be typed into the text block. A tip: Use carriage returns after each line to ensure proper line breaks. ENDTEXT must be typed at the end of the text block.

CLOSE ALTERNATE

• Remember, FoxPro was typing everything from the scanloop into a text file named after the unique ID. This closes that file, so that when the scanloop begins again, Foxpro will name the next alternate file after the next unique ID and thus create the different text files. (When I ran this program, I ended up with 133 text files with names called 1.txt, 2.txt, on through 133.txt)

ENDSCAN

• This tells FoxPro the scan is done, and should be started anew on the next record in the database, until there are no more records left.

```
SET TALK ON
SET STATUS BAR ON
```

• These turn the bells and whistles back on. (Remember, I turned them off in the beginning to make the program run faster.)

After typing this program into the .prg window, to save and exit, type ctrl W. To run the program, in the FoxPro command window, type:

do letters

Making mailing labels for envelopes is even easier because FoxPro offers a wizard that will draw the addresses from the same database. From the Run pulldown menu, choose Wizard... and then "labels" and follow the steps. Tip: Standard mailing labels are size 5160.

Jo Craven can be reached at (573) 882-0684,
or send e-mail to jo@nicar.org

Here are some useful

FOI sites:

• Freedom of

Information Center:

<http://>

www.missouri.edu/

[~foiwww/](http://www.foiwww/)

Call (573) 882-4856

or e-mail Kathleen

Edwards at

[jourke@muccmail.](mailto:jourke@muccmail.missouri.edu)

missouri.edu

• Reporters

Committee for

Freedom of the

Press: <http://>

www.rcfp.org

Surfing for stats

Want to be an IRE member? It just got easier. You can now join IRE via our web server, or you can download a copy of our membership form in Adobe Acrobat format from the web site and submit it by fax to (573) 884-5544. For more information, check out <http://www.ire.org/membership.html>. If you have any questions, e-mail Web Master Wally Winfrey at wally@nicar.org.

by Nora Paul

The Poynter Institute

"He uses statistics as a drunken man uses lamp-posts — for support rather than illumination."

—Andrew Lang

Now, you don't want to be accused of that, right? Learning how to correctly find, interpret, and use statistics for illuminated reporting is one of the most challenging of all journalistic tasks.

The creative and proper use of databases and statistics can help us move, as godfather of precision journalism, Phil Meyer, says, "from event reporting, to putting the event into context."

Event coverage says, "There was a car accident at 3rd and Maple." Context journalism will go on to say, "This is the thirteenth accident at that intersection in the past six months."

Thorough context journalism will go even further to say, "This is the third most dangerous intersection in the city." And investigative journalism will find out why. Statistics and analysis of databases will be the fuel for this kind of journalism.

But there is a danger. As British Prime Minister Benjamin Disraeli, in a quote attributed to him in Mark Twain's memoirs, said, there are three kinds of lies, "lies, damn lies and statistics."

The use of statistics for context setting, "truth-telling" and reality checking in this age of easily accessed data, and misinformation, is a challenge and an opportunity for reporters.

This column will give you some places to hunt for statistics on the Internet, some guides you can read on how to interpret statistics and some resources for locating experts who can help you understand the numbers you crunch yourself.

Looking for statistics on the net

I hate reinventing the wheel, and there has already been a great one created on this topic by information consultant Aggi Raeder (aggi@netcom.com) in the September 1996 (page 16 *Searcher* magazine at <http://www.infotoday.com/searcher/srchrtop.htm>).

She lists meta-sites, and then breaks down

statistics sources into the subject categories of Agriculture, Banking, Business, Crime, Economics, Education, Energy, Government, Health, International, Local Government, Politics, and Population. The article itself isn't on the net, but the information about how to order reprints of the article is.

One of the sites she lists at the top as an example of a great meta-site (a compilation of the sites which are compilations of sites) for statistical information are Larry Schankman's pages. Schankman, a librarian at Mansfield University's Pennsylvania Library, has two pages, both of which should be given medals for organization, thoroughness and description of the resources he links to.

- Business and Economic Numeric Data

<http://www.clark.net/pub/lischank/web/ecostats.html>

From health care expenditures, to housing starts, to lottery funds by State, this page lists all manner of statistical resources from credible sources.

- Be especially sure to check out his "Library Guides Series: Statistical Resources" at <http://www.clark.net/pub/lischank/web/govstats.html>

The site is broken down by categories to help you find web sites and books and reports that are *not* on the web.

- Electronic Data Dissemination of Census Publications

<http://www.clark.net/pub/lischank/web/cendoc.html>

He takes the mother of all statistics sources, the U.S. Census Bureau, and links to its resources, with clear explanations and descriptions of what you will find.

These pages are fascinating to read. Just looking at the types of statistics that are gathered can get story ideas flowing. Check these pages out, bookmark them, then send a note of thanks to Schankman (lschank@clark.net) for his great work.

Guides for using statistics

- Statistics Every Writer Should Know at <http://nilesonline.com/stats/> is the website every writer should bookmark. Robert Niles, *Rocky Mountain News* webmaster and former reporter, has put together the resource guide for under-

Continued on page thirteen

IRE gets Joyce grant

By Brant Houston
IRE Executive Director

By this time next year, there should be a new center on campaign financing open at the headquarters of Investigative Reporters and Editors and the National Institute for Computer-Assisted Reporting.

IRE has just received the first half of a \$342,000 two-year grant from the Joyce Foundation to create and operate the on-line Campaign Finance Information Center.

The center will serve as a repository of federal, state, and municipal databases on campaign contributions and spending. The center will also keep other related public records and offer training to journalists both on-line and at seminars.

Although IRE envisions a nationwide archive, the initial effort with the Joyce Foundation's funds will concentrate on the Great Lakes region, which includes the states of Illinois, Indiana, Michigan, Minnesota, Ohio, Iowa and Wisconsin.

The center will collect the databases of campaign finances and make them easy to use and search on the Internet. The center will not be in the business of doing data entry.

In addition to this electronic data repository the center will offer links to other pertinent Web sites, provide on-line tutorials to teach reporters how to access, analyze and work with campaign data, and publish a newsletter on the subject.

As usual, IRE and NICAR will coordinate its work with other campaign finance study groups to prevent duplication of efforts and to expand the number of resources available.

IRE and NICAR expect the center to begin operations in the early fall once staff and equipment are in place in the offices at the Missouri School of Journalism.

While the grant is a large one, the new center is an ambitious project and we're hoping that we may get additional funds.

NICAR already has an extensive database library of government information, including campaign information that will become a part of the center's archives. The center hopes to collect databases of state and municipal campaign reports that have been created by news organizations and other groups.

The Joyce Foundation is a private foundation based in Chicago with assets of \$650 million. Its

Money and Politics Program supports efforts to reform the campaign finance system to reduce the influence of special interest and enhance the public's participation in elections.

Some databases have already been created with the help of the Joyce Foundation.

Up and coming

While we gear up for the creation of the new center on campaign finance, we're working hard to keep our on-the-road training program up to date and useful.

We will continue to offer our traditional basic one to two-day regional seminars, but we plan to add more seminars on particular topics.

Late this summer we hope to focus on education reporting and databases at a three-day seminar in Philadelphia.

We also plan to offer workshops in other areas such as medicine, business, and managing CAR.

In addition, you will see more seminars that combine workshops on better reporting techniques that IRE has traditionally offered with NICAR's hands-on training.

If you are interested in helping host one of these seminars or have ideas about workshops, we should do, please email brant@ire.org

More CAR stories in Resource Center

IRE's recent contest attracted dozens of new CAR stories and each of those has been indexed and catalogued in IRE's Resource Center. You can search the indexes by going to www.ire.org and choosing the resource center pages. From there you can search the indexes for new and old CAR stories by key words.

Handouts from our computer conferences are also available as abstracts and sometimes in pdf format, which can be easily downloaded and printed out.

A complete set of the handouts, is available through the IRE resource center for \$75, including shipping.

Also, to order a set of abstracts from any of NICAR's national computer conferences, call the resource center at (573) 882-3364 or email Bruce Moores at bruce@nicar.org

Brant Houston can be reached at (573) 882-0684, or email him at brant@nicar.org

Check out the
IRE-L and NICAR-L
mailing list archives
on our websites at
<http://www.ire.org>
and <http://www.nicar.org>
You can see posts
to both lists
organized by
thread, author and
date. The list
archives are
available in html or
in plain text format.

The complete set of
handouts from
CAR Rock 1995 in
Cleveland are
available on-line at
<http://www.nicar.org/cleveland/handouts/index.html>

Census profiles communities

**This is an excerpt
from a handout
provided by Paul
Overberg of U.S.A.
Today at NashCAR,
the NICAR national
conference held
March 6-9 in
Nashville.
To order the full
handout, call the IRE
Resource Center at
(573)882-0684.**

By Paul Overberg
U.S.A Today

The 1990 Census remains a good reference, but it's more history than news. Cheer up! Here are several reliable government sources of 1990s demographic data on counties or even smaller areas. All of them are free or nominal, clean and electronic. On first glance, what they offer seems pretty obvious. The secret lies in analyzing and comparing the pieces to find patterns.

Yes, this data is estimated or incomplete. But remember: Only one person in seven got the long Census form, the source of most of what we write about the 1990 Census. In all cases below, read the documentation to understand the data's limits.

- Annual county-level population estimates and components

<http://www.census.gov/population/www/estimates/countypop.html>

The Census Bureau calculates these based on July 1 and releases the data the following March. Even better than getting the estimate is getting the underlying data used to make the estimate: births, deaths, elderly (65+), under 65, net domestic migration, net international immigration, federal citizen movement, group quarters population.

Of these, the single most valuable is net domestic migration, especially when you line up several consecutive years. That's because it's the collective verdict of people voting with their feet. People usually don't move far enough to leave a county. When they do, often it's to go to college, retire or take a better job, meaning many of them won't come back.

- Annual county-level estimates for race/Hispanic origin, sex, and age

<http://www.census.gov/population/www/estimates/co9094.html> (Totals only)

<http://www.census.gov/population/estimates/county/casrh/> (Details by five-year age group)

The Census Bureau calculates these periodically on contract and releases them about a year after the estimate date. (Current data is 1990-94; 1990-96 data is due out in July).

These differ in methods and timing from the main estimate process above, so expect differences. You can use the race/Hispanic origin/sex totals for each state and county to watch for trends.

The same data, broken down by five-year age groups, lets you see changes in a county's age structure versus the state and nation.

- Annual county-level migration files, with median household income

For county-to-county and most recent state-to-state files, call Emily Gross at the IRS Statistics of Income Division, (202) 874-0944. For previous years' state-to-state files, dial the division's BBS, (202) 874-9574. Or try the often-overloaded Fedworld address: <ftp://ftp.fedworld.gov/pub/irs-soi/>

The IRS uses the Social Security number on each tax return to check county of residence in consecutive years. For each county, it counts how many stayed put from one year to the next, and how many moved between counties. For each county-to-county flow, it reports returns (roughly, households), personal exemptions (roughly, persons), aggregate income and, since 1992-93, median household income.

You can match a county's inbound and outbound files to calculate net flows between nearby counties, but you can also examine the files separately for distinct patterns in where people come from and go to.

- Annual County Business Patterns

<http://www.census.gov/epcd/cbp/view/cbpview.html>

The Census Bureau bases these on a survey conducted in early March. They're released 15 months after the survey year. Data shows employment, payroll and number of establishments by industry. Some data on payroll and employment is masked to prevent disclosure in small categories. But the number of establishments isn't. Neither is their distribution by size.

This data is helpful in following regional trends by industry, especially in the hard-to-track services sector: clinics, churches, day-care centers, hair salons, etc. The data includes basic groups like retailing and construction and smaller groups like "eating and drinking places."

Continued on page thirteen

Continued from page twelve:

Census sources

Other useful data

- Annual estimates for towns and cities
<http://www.census.gov/population/www/estimates/citypop.html>

These are “places” or “minor civil subdivisions,” in Census speak. Basically, they’re any self-governing political entity smaller than a county. (“Places” covers all such things in 30 states.

In 20 states in the Northeast and Midwest, a variety of local political units like townships, towns, boroughs and villages are self-governing, so the bureau lists them separately).

Estimates are based on July 1 and released the following July. The most recent estimate was July 1, 1994. (The bureau skipped the 1995 release to focus on improving its methodology) These aren’t as accurate as county estimates, but are helpful for watching patterns.

- Annual county totals in various Social Security programs
<http://www.ssa.gov/statistics/oasdi94.html>

http://www.ssa.gov/statistics/ssi_st_cty_d95/
<http://www.ssa.gov/statistics/eedata93.html>
(For most current data on a few counties, call Social Security Administration at (410) 965-8904).

The Social Security Administration has placed several annual county-level statistical reports on the web. They’re tabular, so download and import is simple. Currency varies.

Other sources

Finally, don’t forget to check for customized analyses and research done by your state demographer and state or county planning department.

Also, see if your marketing department pays for access to demographic data from a company like Claritas. It’s often overkill for newsrooms and easy to misuse, but it can help portray customized or ZIP code areas, where between-Census data is rare.

Paul Overberg can be reached at (703) 276-5427 or email him at
70534.3647@compuserve.com

The Census Bureau produces several reports including the Statistical Abstract of the United States, the County and City Data Book and USA Counties.

These documents include data ranging from the cost of living to health care. They’re available as books, but the CD-ROM versions include more data. Many of the tables can be downloaded for free from the Internet.

Continued from page ten:

On the Internet

standing statistics using the language and framework of reporting. Read it all and you’ll be much further along in understanding the basics.

- Got those numbers, but don’t know what they really mean? Subject specialists spend years learning how to interpret those numbers. You need some help as you move from one subject to another in the course of your reporting. Check into the Reporting Tools area of The Foundation for American Communications’ website for journalists at http://www.facsnet.org/report_tools/main.html

FACS Guides and Primers, and News and Numbers are great backgrounders on understanding the numbers behind the news. Click on the “Covering Business” link and get an explanation of reading financial statements and doing financial analysis.

- Research Methods Resources from the College of Education at Arizona State University at <http://seamonkey.ed.asu.edu/~behrens/links> to an incredible array of resources about statistical analysis.

The Statistical Instruction Internet Palette (SIIP) includes the Data Gallery with samples of data displayed in different ways and the Equation Gallery that gives you various formulas for statistics (i.e. - click on “variance” and you will get a formula, plus an in-depth explanation of what it means).

Experts explain the numbers

Sometimes the numbers aren’t enough. You have to understand their importance.

St. Petersburg Times researcher Kitty Bennett has pulled together an experts and sources directories, topped with a link to ProfNet’s Expert Database at <http://time.vyne.com/profnet/ped/eg.acgi>

She has found university and think tank expert directories and listed them for you on the Special Libraries Association / News Division website at <http://sunsite.unc.edu/slanews/internet/experts.html>

Nora Paul can be reached at (813) 821-9494, or send email to npaul@poynter.org

Get your training

IRE committees always are on the lookout for good volunteers. Of particular interest to Uplink readers may be the Computer Committee, chaired by IRE Director Shawn McIntosh of *The Dallas Morning News*. To volunteer, send e-mail to McIntosh at dmnproj@airmail.net. For information on other IRE committees, send e-mail to Managing Director Brant Houston at brant@nicar.org

On-the-road training

NICAR and, where indicated, The Associated Press provide specialized training in your newsroom. Learn to transfer data from government files into newsroom PCs. Build spreadsheets for insightful stories on the beat. Comprehend documents with database managers. Navigate the Internet and on-line databases.

Cost varies. For information, call Lisa Barnes at (573) 882-8969, or to register, call the numbers below.

Note: In the following list, "Closed" means the session is open only to members of the host organization.

- Chicago, Ill., *Chicago Tribune*. April 30-May 2. Closed.
- Charleston, W. Va., Marshall University. May 13-16. Closed.
- Denmark, Sweden. May 26-June 6. Closed.
- Reuters. May 29-30. Closed.
- Reuters. June 26-27. Closed.
- Washington, D.C., Freedom Forum. July 14-15. Closed.
- Tempe, Ariz. Reporting and Education seminar. July 25-26. Closed.
- Minneapolis, Minn. August 4. Closed.
- Charleston, W. Va. Aug. 19-22. Marshall University. Closed.
- Medill School of Journalism. Sept. 11-13. Open.
- New York, N.Y. Sept. 26-27. Columbia University. Open.

Conferences

NICAR will offer training and seminars at the following professional conferences, including the IRE and NICAR national conferences. Costs vary. For information or to register, call Lisa Barnes at (573) 882-8969

- IRE National Conference, Phoenix, Ariz., — June 12-15.

Bootcamps

Bootcamps are week-long, intensive training sessions offered at NICAR's headquarters at the Missouri School of Journalism in Columbia, Mo.

As with on-the-road training, you will learn to transfer data from government files into newsroom PCs, build spreadsheets for stories on the beat, comprehend documents with da-

tabase managers, and navigate the 'Net and on-line databases — but you'll be drilled all day, every day for a full week. Tuition ranges from \$500-\$1,000 depending on circulation or market size.

For information, call Wendy Charron at (573) 882-0684.

- May 4-9. Advanced Bootcamp concentrating on stats and maps at the University of North Carolina-Chapel Hill with Philip Meyer, author of *The New Precision Journalism*. The seminar is held in Chapel Hill. For details, call (573) 882-0684.

- May 18-23. Regular Bootcamp on general computer-assisted reporting. For details, call (573) 882-0684

NEW • July 11-13. Mini Bootcamp on general computer-assisted reporting. For details, call (573) 882-0684.

- Aug. 10-15. Regular Bootcamp on general computer-assisted reporting. For details, call (573) 882-0684

South of the Border

Periodistas de Investigacion (IRE-Mexico) offers computer-assisted reporting training in Spanish to journalists in Mexico and Latin America. PI, based in Mexico City, has developed a two-day workshop that covers the basics of investigative reporting and computer-assisted reporting, and a four-to-five day bootcamp that covers CAR intensively.

The workshops can be held in almost any city with prior notice. The courses cover using internet, spreadsheets, database managers, and generally electronic records, and uses examples from the Mexican and Latin American press.

There is one bootcamp scheduled in May:

- Mexico City — May 14-18.

Annual membership to IRE-Mexico is \$25 (or \$200 pesos mexicanos) and is open to journalists or students of journalism living in Mexico and Latin America.

Instructors are Lise Olsen, managing director of IRE-Mexico, and Pedro Enrique Armendares, assistant director. Both have more than seven years of experience in investigative reporting. Olsen at *The Virginian-Pilot* (Norfolk, Virginia) and Armendares at *La Jornada* (in Mexico City). Call Olsen at (52)-5-554-7613 for more information, send email to: lise@ire.org

Growing collection of federal databases

From the NICAR library

NICAR offers a number of federal government databases. Here is a list of our growing collection:

NEW • Federal Aviation Administration's accidents and incidents, including major plane crashes since 1971.

NEW • NASA's air safety reporting system, including anonymous complaints by pilots and air traffic controllers. Useful for finding near misses and problems at local airports, 1988-1996.

- A monthly CD subscription for all 1995-96 Federal Election Commission campaign contributions by individuals and political action committees, plus all presidential matching fund requests.

- The Health Care Financing Administration's 1995 database of all Medicare-funded inpatient work in U.S. hospitals.

- Federal Railroad Administration data for accidents, casualties, and highway crossings. 1991-1995.

- Coast Guard boating accidents, 1969-1994.

- Federal Aviation Administration data, including airplane maintenance work documented in the service difficulty report, pilot licenses and grades, and aircraft registration.

- Home Mortgage Disclosure Act records, for tracking who gets loans and who gets turned down, and finding redlining patterns, 1992-1995.

- Federal procurement data, 1992-1995, includes breakdowns by agency.

- Alcohol, Tobacco and Firearms gun dealer records. 1993-1995.

- National Bridge Inventory System data, includes inspection grades. 1994-1995

NEW • FBI Uniform Crime Reports, a detailed compilation of crime data that includes statisti-

cal breakdowns of individual murders. This includes the new 1995 data.

- Social Security death records, by name and social security number, going back to 1937.

- Occupational Safety and Health Administration violation data includes worker accidents and exposures to hazardous chemicals by companies, 1974-1996.

- U.S. Department of Transportation truck accident and census data. It includes accidents by company and road.

- U.S. Small Business Administration loan guarantees, 1989-1995. This includes the name of the business, address, amount covered by the SBA, and status, including whether the loan went bad.

- U.S. Small Business Administration disaster loan guarantees, 1989-1995. This includes individuals and businesses, the amount covered by the SBA, and the status, including whether the loan went bad.

- U.S. Small Business Administration's list of minority companies certified for SBA assistance in seeking federal contracts. It includes the name of the company, its address, the owner, type of business and phone number.

- The National Inventory of Dams. 1991-1995.

- U.S. Department of Transportation hazardous materials accidents database, a collection of roadway, rail, air and waterway accidents from 1971 to 1995.

- U.S. Department of Transportation fatal accident reporting system. It includes all roadway accidents from 1988 to 1995.

- U.S. Coast Guard directory of U.S. merchant vessels. It includes the name of the ship, the managing owner, home port and various descriptive information.

For up-to-date prices and more information, call (573) 882-0684, or send e-mail to info@nicar.org

NICAR's week-long bootcamps in Columbia, Mo., offer hands-on training in computer-assisted reporting skills, including the use of spreadsheets and database managers, accessing data in various media, such as nine-track tapes, and negotiating for data.

For more information, call NICAR (573) 882-0684, or send e-mail to info@nicar.org

Bits, Bytes and Barks

Help Wanted

- *The Orange County Register* is in search of a technology reporter. Applicants for the position should submit a packet of current work, a resume and a one-page statement of philosophy to: John Doussard, 625 N. Grand Ave., Santa Ana, Calif. 92701. Telephone: (714) 953-2225. Fax: 714-565-3681.

- The *Dayton Daily News* (160,000 daily; 220,000 Sunday) is looking for a computer-assisted reporter who knows FoxPro, can crunch data for daily and long-term computer stories, help other reporters on computer projects, pull information off the Internet and write a story when needed. This is an opportunity to work on a newspaper that is serious about computer-assisted journalism. The *Daily News* has won numerous national awards in the 1990s for computer projects, and was a Pulitzer finalist twice. If interested, send resume, cover letter, clips and references to John Erickson, News Manager, c/o Dayton Daily News, 45 S. Ludlow St., Dayton, Ohio, 45402. Email address: John_Erickson@CoxOhio.com Phone: 937-225-2266.

NICAR Net

Don't forget to keep up with NICAR on the Internet.

Subscribe to our listserve and join in as reporters talk about how to do the job better. Email to LISTSERV@MIZZOU1.MISSOURI.EDU. In the message, on the first line, write: subscribe NICAR-L your name. To join

IRE on the Internet, the instructions are the same except, on the first line, write: subscribe IRE-L your name.

Statistics newsletter

If you are interested in how we use (and sometimes misuse) statistics, there is a newsletter on the Internet that is good and fun.

It's part of the CHANCE Course Project, which operates from Dartmouth and is supported by the National Science Foundation.

The newsletter includes abstracts from recent newspaper and journal articles that have made use of statistics.

It is published in electronic format and can be found on the web at http://www.geom.umn.edu/docs/education/chance/chance_news/news.html

For an email subscription, drop an email message to jlsnell@dartmouth.edu

Get your subscription

A one-year subscription to Uplink is available for \$40 for IRE members and \$60 for nonmembers.

In upcoming months, Uplink will be taking a look at some of this year's Pulitzer-winning stories that involved CAR, issues relating to the coverage of transportation, and CAR developments taking place at smaller newspapers and television stations. To subscribe, call (573) 882-0684, or send email to wendy@nicar.org

NON-PROFIT ORG.
U.S. POSTAGE
PAID
PERMIT NO. 309
COLUMBIA, MO. 65211

Investigative Reporters and Editors, Inc.
100 Neff Hall
University of Missouri
School of Journalism
Columbia, MO 65211