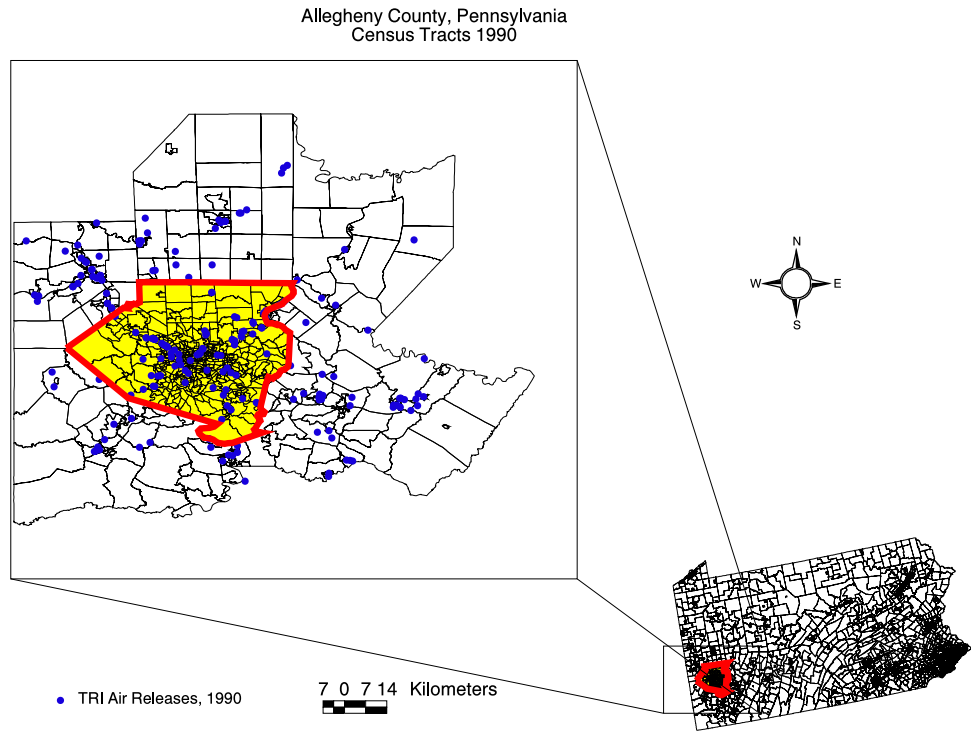


STATS

STATS STATS STATS



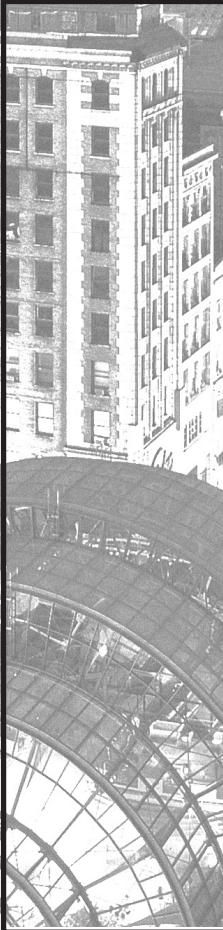
INSIDE:

**How Soil Composition
Affects Density and
Water Capacity – A
Science Fair Project
Using Mixture Experiment**

**Question: Who Chomped
Their Way to the 1999
College Bowl Championship?**

**Statistical Issues in
Assessments of
Environmental Justice**





STUDENT OPPORTUNITIES IN INDIANAPOLIS

- Reduced registration fees
- Reduced Continuing Education fees (advanced only)
- Student mixer
- Placement Service (with reduced fees)
- Cox Scholarship Fun Run/Walk
- College Bowl with student teams
- State of the art Exhibit Hall
- Mix and mingle with renowned statisticians and other students
- Technical presentations



www.amstat.org/meetings/jsm/2000

Continuing Education Program Seeks Student Monitors

The 2000 Joint Statistical Meetings - Continuing Education Program seeks student monitors for its educational presentations. CE presentations will be held at the Westin Hotel—Indianapolis from Saturday, August 12 to Wednesday, August 16.

Student monitors are expected to:

- Attend a brief mandatory monitor training session on-site where important and last minute information will be explained.
- Be available for the entire presen-

tation.

What are the responsibilities of a Student Monitor?

- Pick up required materials, textbooks, etc. for presentation he or she will monitor.
- Handle registrants — check them into the course, hand out required materials and respond to inquiries
- Evaluations — Distribute, collect, tabulate

If you are interested in participating, provide your name, complete mailing

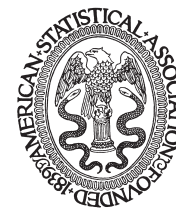
address (if different during the summer), email and phone number to Patricia Hayden, ASA's CE Coordinator. Send your information either by email patricia@amstat.org or by fax to 703-684-3768. Notification will be sent to the Student Monitor when the CE presentations go on-line. The student monitor will be asked to select the presentation he or she wishes to monitor. A student monitor receives materials from the course he or she monitors. ■



STATS

The Magazine For Students Of Statistics

Spring 2000 • Number 28



Editor

Jerome P. Keating
email:
jkeating@utsa.edu

College of Business
University of Texas at San Antonio
6900 North Loop 1604 West
San Antonio, TX 78207

Editorial Board

Beth Chance
E-mail:
bchance@calpoly.edu

Department of Statistics
California Polytechnic State University
San Luis Obispo, CA 93407

E. Jacquelin Dietz
email:
dietz@stat.ncsu.edu

Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

Rudy Guerra
email:
rguerra@mail.smu.edu

Department of Statistical Science
Southern Methodist University
Dallas, TX 75275

Robert L. Mason
email:
rmason@swri.edu

Statistical Analysis Section
Southwest Research Institute
San Antonio, TX 78228

Allan J. Rossman
email:
rossman@dickinson.edu

Department of Mathematics
Dickinson College
P.O. Box 1773
Carlisle, PA 17013

W. Robert Stephenson
email:
wrstephe@iastate.edu

Statistics Department
Iowa State University
327 Snedecor Hall
Ames, IA 50011-1210

Production

Carol Edwards
email:
caroled@amstat.org

American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3415

STATS: The Magazine for Students of Statistics (ISSN 1053-8607) is published three times a year, in the winter, spring, and fall, by the American Statistical Association, 1429 Duke St., Alexandria, Virginia 22314-3415 USA; (703) 684-1221; fax: (703) 684-2036; Web site: www.amstat.org

STATS is published for beginning statisticians, including undergraduate and graduate students who have a special interest in statistics, and is distributed to student members of ASA as part of the annual dues. Subscription rates for others: \$13.00 a year to members; \$20.00 a year to nonmembers.

Ideas for feature articles and material for departments should be sent to the Editor, Jerome P. Keating. Addresses of the Editor and Editorial Board are listed above.

Requests for membership information, advertising rates and deadlines, subscriptions, and general correspondence should be addressed to *STATS* at the ASA office.

Copyright © 2000 American Statistical Association.

Produced by Oakland Street Publishing • Printed in U.S.A.

Features

3 LULU? NIMBY! Statistical Issues in Assessment of Environmental Justice
Lance A. Waller and Erin M. Conlon

15 How Soil Composition Affects Density and Water Capacity — A Science Fair Project Using Mixture Experiment Methods
Mark G. Piepel and Greg F. Piepel

22 Question: Who Chomped Their Way to the 1999 College Bowl Championship
Mark Payton

Departments

2 Editor's Column

24 AP STATS
W. Robert Stephenson and Hal Stern

28 Student Voices
On Becoming A Statistics Educator
Jackie Miller

Cover image fl Figure 2. Map of Allegheny County, Pennsylvania and surrounding counties showing the census tract boundaries from the 1990 United States Census, and sites reporting toxic releases to the Toxic Release Inventory in 1990. (see Lance Waller and Erin Conlon, page 3)

Editor's Column



Jerome P. Keating

Celebrating Diversity

Dear STATS Readers:

I am encouraging everyone to attend this year's JSM, which will be held in Indianapolis, Indiana from August 13, 2000 through August 17, 2000. The title of this year's theme is "*Celebrate Diversity in Statistics*." There are many activities directly arranged for students, such as the Student Mixer, the College Bowl, the Video Theater, The Gertrude Cox Scholarship Run, and much more. It is a great place to meet fellow students, who share your interests in data analysis, computing, math, etc. The College Bowl is a unique opportunity for you and your classmates to pit your skills against those of others across the country. Will the University of Florida repeat as this year's College Bowl Champions or will a perennial power such as the University of Iowa or Iowa State University reclaim the title? See Mark Payton's wonderful article on last year's College Bowl competition in this issue.

Can't afford the trip? Notice the advertisement on page 11 of the preceding issue of *STATS* for the *Data Challenge by Capital One*, which offers a grand prize of \$2,500 in cash and an all-expense paid trip to JSM to receive the award. The Continuing Education Courses may allow you to obtain a stipend and attend an informative session on a current topic free. See Michelle Larson's article on this little known opportunity for students in *STATS*, Issue 25, 1999, pg. 14 and the ad on the inside of the front cover of this issue..

■ Feature Articles

This issue echoes the JSM theme, *Celebrate Diversity in Statistics*, through both new and timeless articles. Perhaps you have heard of the emerging topic of *Environmental Justice*. Lance Waller of Emory University and Erin Conlon of the University of Washington provide you with a lively discussion of many prominent issues related to environmental risk assessment. Certainly your

community has a landfill or a sewage treatment facility. They are not likely to be located in affluent parts of town. If they are in less affluent areas, has the community committed *environmental racism*? See how important the nation's Geographical Information Systems (GIS) are to this area of research. The data obtained from the GIS are used to measure *race* and *exposure* in an effort to determine equity or fairness in environmental risk.

In my office sits a *quincunx*, which was originally developed by Sir Francis Galton to illustrate the normal approximation to the binomial distribution. My son, Greg, and I built one in 1987 as part of his science fair project. In this issue, Mark and Greg Piepel recount Mark's science fair project in this timeless father-son tradition. You will be introduced to a very clear discussion of *mixture experiments* as they relate to the density and water capacity of soil composition.

■ Column Articles

In this edition of AP *STATS*, Bob Stephenson and Hal Stern, the Editor of *Chance*, discuss the important topic of *randomization* in designed experiments. Follow their discussion on systematic versus randomized assignment in an agricultural experiment involving two hybrids of corn. We hope that our pre-college teachers will make use of this enlightening example.

In the Column, Student Voices, Jackie Miller of Ohio State University gives you a glimpse at the process of becoming a statistics educator. She discusses many pedagogical techniques and explains their places in our discipline. She refers to the learning theories of giants, such as Piaget, who have profoundly effected quantitative disciplines. Her one-of-a-kind program speaks eloquently for diversity in statistics education.

Outlier...s will be back next fall. Stay tuned and we hope to see you in Indianapolis.

A handwritten signature in black ink that reads "Jerome P. Keating". The signature is fluid and cursive, with the first letters of the first and last names being capitalized and prominent.

LULU? NIMBY!

Statistical Issues in Assessments of Environmental Justice

“Environmental justice” describes the equitable sharing (and reducing) of burdens due to environmental hazards across all sociodemographic groups. The issue has seen considerable debate in government, policy, and legal circles over the last decade. A presidential Executive Order requires assessments of environmental justice by all federal agencies, but operational definitions and methods for assessment remain subject to debate. We briefly review the development of the issue, definitions, and recent policy decisions relating to environmental justice, then focus on several statistical issues involved. The issue is not purely a statistical one, yet development of accurate statistical summaries and inferences plays a key role in improving assessments of environmental justice.

■ 1. What is Environmental Justice?

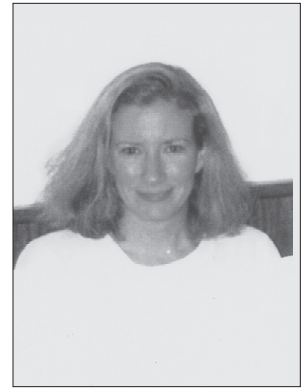
The question of siting of a “locally unwanted land use” (LULU), particularly a source of environmental hazard, usually results in a response of “not in my backyard” (NIMBY). In determining where to place the nation’s waste, “somewhere else” is an answer with strong local support but limited global applicability.

The phrases “environmental justice” (EJ) and “environmental equity” appear increasingly in environmental policy, environmental regulations, legal proceedings and the press. The phrases refer to a situation where no population subgroup carries an undue excess burden due to environmental hazards. Of particular interest are subpopulations defined by racial groups or socioeconomic status, in order to determine if historically disenfranchised populations are bearing the brunt of ill effects due to environmental hazards. In other words, do we as a nation tend to dump our waste in neighborhoods with high proportions of minority racial groups

Lance A. Waller Department of Biostatistics, Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322 and Erin M. Conlon Division of Medical Genetics, University of Washington School of Medicine 1705 NE Pacific Street, Box 357720, Seattle, WA 98195



Lance A. Waller



Erin M. Conlon

and/or the poor?

The Institute of Medicine, a branch of the National Academy of Sciences, recently released a report by its Committee on Environmental Justice summarizing many of the issues associated with environmental justice (Institute of Medicine, 1999). We will briefly review some issues outlined in the report, but quickly move to statistical issues associated with environmental justice assessments. We refer interested readers to the report for more details and a list of related references.

Racial and socioeconomic disparities in the prevalence of health effects are well-known and well-documented (see for example, the May 1997 issue of *American Journal of Public Health*, pp. 740-838 for a series of recent articles). While the distribution of socioeconomic status differs between racial groups, health differences are observed even after adjusting for socioeconomic status (Kingston and Smith 1997). One reason postulated for this difference is disproportionate exposure to environmental hazards by minority racial groups.

Following an investigation of this hypothesis, the Commission for Racial Justice of the United Church of Christ published *Toxic Wastes and Race in the United States* (United Church of Christ 1987). The authors report increased proportions of racial minorities in ZIP codes containing one or more toxic waste sites compared to those containing no waste sites. The report focused national attention on “environmental racism”, and the United States Environmental Protection Agency formed its Office of Environmental Equity (now Office of Environmental Justice) in November 1992 (Sexton, Olden and Johnson 1993).

In 1994, President Clinton issued Executive Order 12898: “Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations”. The order requires each federal agency to “analyze the environmental effects, including human health, economic, and

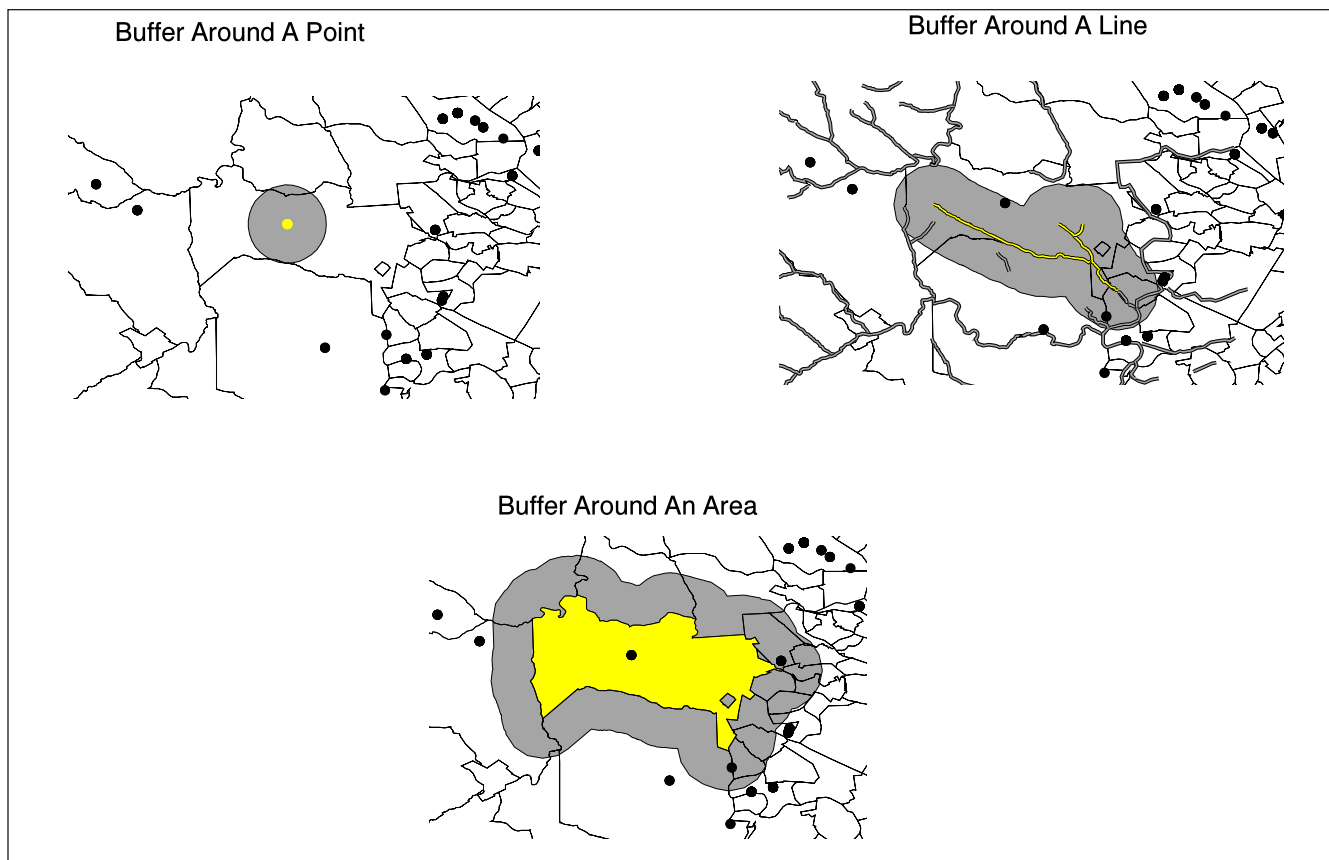


Figure 1. Examples of GIS buffers around a point, a line, and an area.

social effects” of federally-sponsored programs. One result of the executive order is that all agencies receiving federal funds must perform environmental justice assessments of the impact of their work. In addition to the siting of hazardous waste sites, this applies to the building of roads, and any other project receiving part or all of its funding from the U.S. government.

As you might expect, the issue is controversial and can be quite political. The issue falls under the general category of “mandated science” or research conducted primarily for decision-making purposes (Salter 1988). While most will agree that the concept of environmental justice is a worthwhile goal for society, little agreement exists on how to operationalize, evaluate, and quantify the issue. Politics aside, there are many interesting statistical issues involved in assessing environmental justice. We will explore some of these issues below. The key question we wish to address is: Are subpopulations subjected to disproportionate environmental exposures and their effects?

■ 2. GIS and EJ

Wagener and Williams (1993) suggest that assessing environmental justice involves comparisons of the distribution of three elements throughout the population. First, what is the

distribution of *exposure* across the study area? Second, how are the population’s *demographics* distributed across the study area? Third, how are any *health or other effects* distributed across the study area? All three questions involve the geographic distribution of values across space. In effect, we may think of an environmental justice assessment requiring the combination and comparison of a map of exposures, a map of population demographics, and a map of outcomes possibly due to the exposures. Geographic Information Systems (GISs) provide a computational way to combine and display spatially referenced information, and we focus our attention on the use of GISs in assessments of environmental justice.

A geographic information system (GIS) is a collection of computer software routines for capturing, storing, analyzing and displaying spatially-referenced data. A computer system is considered a GIS if it can perform spatial operations on data and use latitude and longitude and other spatial information to answer geographic questions such as: “what proportion of the population within 1 kilometer of a toxic waste site has a certain disease?”, and “what are the locations of pollution monitoring stations in a state?” (ESRI, 1990-1995.)

GIS databases consist primarily of two file formats, vector (line-type) and raster (image-type).

The vector format stores geographic features such as points, lines and areas as a series of coordinates with their associated observed or measured values referred to as “attributes”. Point data include location coordinates and an attribute value. Line data include connected segments such as streets, rivers and administrative boundaries and an attribute value. Area data include polygons such as census tracts, counties and states and an associated attribute value for the area. In contrast, the raster format stores geographic information as a grid structure, or file of pixels (picture cells). Examples of raster files include remotely sensed data from satellites such as vegetation cover or weather data; other applications using raster files include magnetic resonance imaging, radar, and electron microscope imaging. Most GISs are based on the vector file format, but allow for use of raster data (Croner et al. 1996).

A GIS stores databases for many attribute values, often from different sources; examples include street names and addresses, exposure sites, population counts, residences of disease cases, demographic data such as age, race, and socioeconomic variables, and geographic features such as rivers and lakes. The GIS links each of these georeferenced databases and their attribute values through an internal relational database that associates values from common points, lines or areas. The maps of attribute values are considered “layers” by the GIS and can be displayed simultaneously. The relational database manager allows selection on the basis of attributes so that features of the data can be highlighted or summarized. One method of highlighting is “buffering”, which allows the user to find features within a given distance (e.g. 1 kilometer) of a point, line, or area anywhere on a given map. The user can select features from one or more attribute layers that fall partly or entirely within the buffer (ESRI 1996, Clarke et al. 1996). Figure 1 provides examples of point, line, and area buffers.

The layering and buffering features of GIS are important tools in assessing environmental justice. The distributions of the three elements of environmental justice of (exposure, population demographics and outcomes) define three data layers which can be linked through a GIS. The buffering feature allows the user to locate specific areas of interest and to calculate exposures and outcomes for the chosen subregions using the various data layers, and is in wide use in environmental justice assessments as outlined below.

■ 3. Statistical Issues

3.1 Types of Data

As with most statistical applications, the available data often define the applicable methods

in assessments of environmental justice. As mentioned above, the key components of an EJ assessment involve the definition of population subgroups, measurement of exposures, and the measurement of outcomes.

Population: The first component involves demographic data primarily from the United States Census regarding race, ethnicity, and socioeconomic status. Due to confidentiality restrictions, such data are not available by household, but rather as summaries for various sets of small areas. Counties partition states, census tracts partition counties, census block groups partition tracts, and blocks partition block groups. Figure 2 shows the boundaries of the 499 census tracts (1990 census) for Allegheny County, Pennsylvania. Census summaries are also available for ZIP codes, as defined by the United States Postal Service. However, ZIP codes may change at any time depending on the needs of the postal system, while tracts, block groups, and blocks are generally fixed at least for the 10 years between each census.

Most EJ assessments treat census values as fixed, measured, and static values. In reality, census data are combinations of enumeration and sampled data (see Billard 1999 for an overview). Immigration and emigration within and between census regions certainly occur and provide further uncertainty in the data. Few EJ assessments explicitly consider the dynamical nature of the population’s measured sociodemographic structure. Such considerations using population projections and local samples provide ground for further statistical developments.

Exposure: The second component involves measurement of environmental exposures received by individuals in the study area. Such individual-level data typically are only available for particular exposures in particular individuals (e.g. in an occupational cohort of workers in a nickel refinery), and not regularly measured in the general population. Environmental monitoring of ambient levels of potential environmental hazards and developments in the statistical design and analysis of environmental monitoring networks increase each year, but widely available, widescale information on the distribution of environmental exposures remains rare. Note that accurate measurement of the ambient level of a contaminant at a given location may not correspond directly to the exposure received by an individual at that location. The dose received depends on factors such as the individual’s respiration rate (for inhalation), hand-mouth contact (for ingestion), and many others. As an example, Wallace et al. (1985) report that the best predictors of an

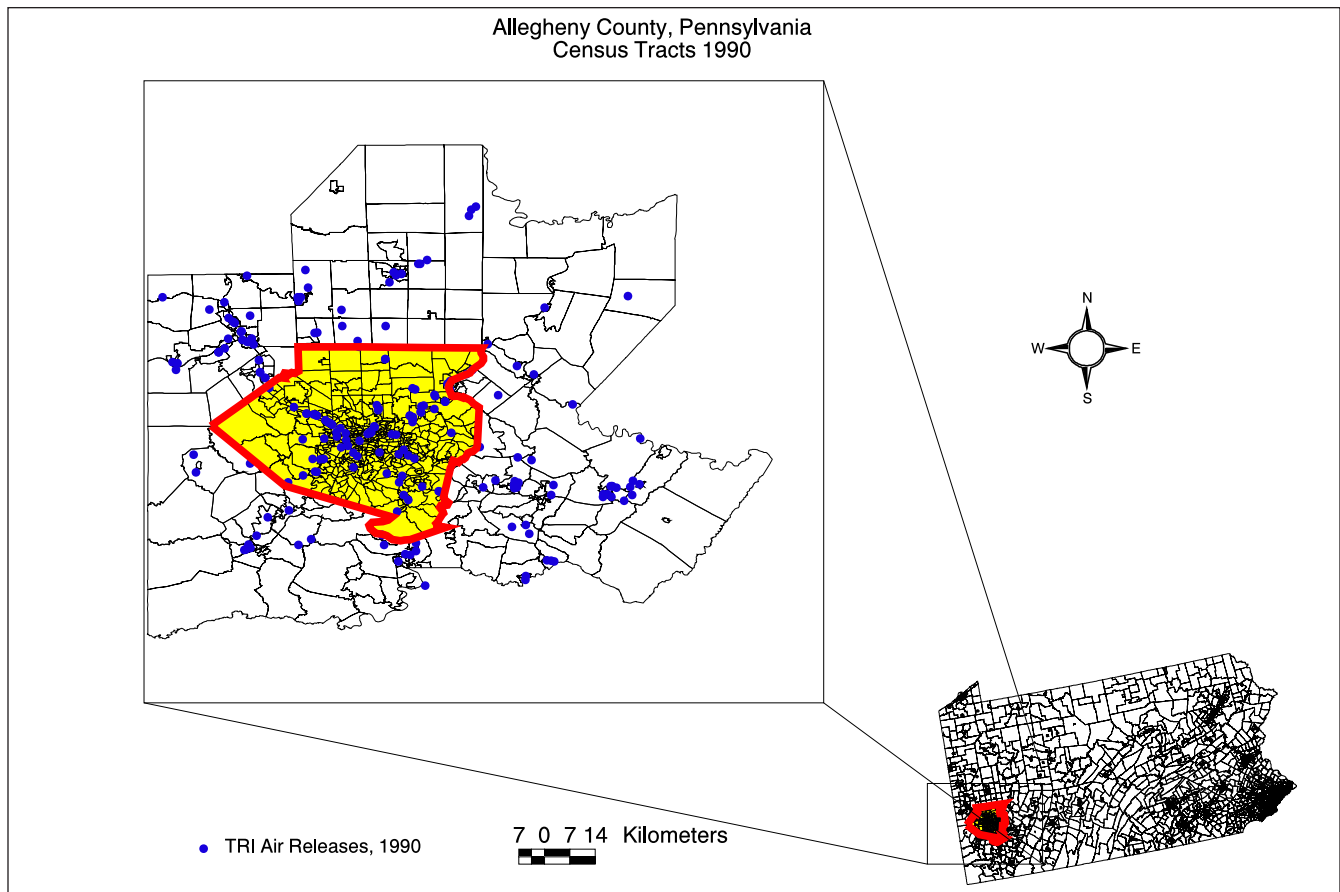


Figure 2. Map of Allegheny County, Pennsylvania and surrounding counties showing the census tract boundaries from the 1990 United States Census, and sites reporting toxic releases to the Toxic Release Inventory in 1990.

individual's exposure to benzene are whether he or she smokes, and how often he or she drives or fuels a car, not estimated ambient exposures or proximity to releases.

Despite results such as those of Wallace et al. (1985), researchers often base EJ assessments on proximity to hazardous waste facilities or industrial releases of toxic chemicals and compounds, due primarily to the lack of widely-available individual exposure measurements and comprehensive ambient exposure measurements. Many EJ assessments use the United States' Environmental Protection Agency's (EPA's) Toxic Chemical Release Inventory (TRI). A company releasing or transferring any of approximately 300 chemicals during the course of a year generally must report the release or transfer to the EPA. The TRI is not a complete listing of toxic releases as small businesses and government facilities are currently exempt from reporting. The TRI does not report exposure, merely annual releases in wide ranges of thousands of pounds (e.g. 10,000-99,999 pounds per year). As a result, many EJ assessments ignore the reported release amounts and simply use the locations of TRI sites, defining "exposure" in terms of proximity to the release locations. TRI sites

reporting releases in Allegheny County for 1990 appear in Figure 2.

Outcomes: The third component involves possible adverse outcomes including health events such as disease incidence or mortality, and measures of healthcare utilization (e.g. the number of asthma visits to a particular emergency room). See Carlin and Xia (1999) for examples of both types of outcomes. Death certificates record the primary cause(s) of death, health surveys report health outcomes in a sample of individuals, and disease registries record the incident (new) and prevalent (existing) cases of "reportable" diseases (diseases health care workers are required to report to state or federal agencies responsible for maintaining registries). However, registries tend to vary from state to state as to which diseases are reported and in what format. For the purposes of this paper, we will ignore particular outcomes, but discuss possibilities for moving from exposure-based to risk-based EJ assessments in Section 4 below.

Note that all three components of an EJ assessment involve data originally collected for other purposes. Most EJ assessments do not follow traditional patterns of statistical design, rather they

involve merging data collected by different agencies for different purposes. As such we are limited in the conclusions we can draw, especially with respect to any notion of causation. The data are observational rather than experimental, and demographic data are “ecologic” meaning that they are summaries over groups rather than specific measurements on individuals (see Morgenstern 1998, for an overview of the analysis of ecologic data). For example, we may know the proportion of individuals in each population subgroup for a particular county, but may not know the geographical distribution of the subgroups within the county. If this county contains a source of environmental hazard and has a high proportion of a particular subgroup, we may be tempted to conclude evidence of environmental injustice. However, from such data, we could not distinguish between a situation where the subgroup of interest was tightly clustered around the hazard (indicating even stronger inequity than we would observe at the county level), a situation where the subgroups were entirely integrated (indicating little evidence of inequity within the county, but evidence of inequity at the county level), or a situation where the subgroup of interest is actually clustered away from the source (indicating a different direction of inequity within the county than at the county

level). To minimize the impact of this so-called “ecologic fallacy”, assessors often use the smallest geographic units available that still provide accurate data. While issues relating to observational and ecologic data apply to most data components, other issues are specific to individual components, as outlined below.

3.2 Measuring race

The most common subpopulations of interest in EJ assessments are groups defined by race. Measurement of race largely depends on self-identification with a choice from a set of possible responses. The 1990 U.S. Census data bases race on respondents choosing the race they most closely identified with from the options provided (e.g. “White”, “Black”, “American Indian”, “Asian or Pacific Islander”, or “Other”). A separate question in the census determines Hispanic origin, so respondents identifying themselves as being of Hispanic origin were also requested to choose a race category from the list above. The 2000 Census will allow respondents to pick more than one race category, further complicating the data structure and format.

3.3 Measuring exposure

As mentioned above, EJ assessments often

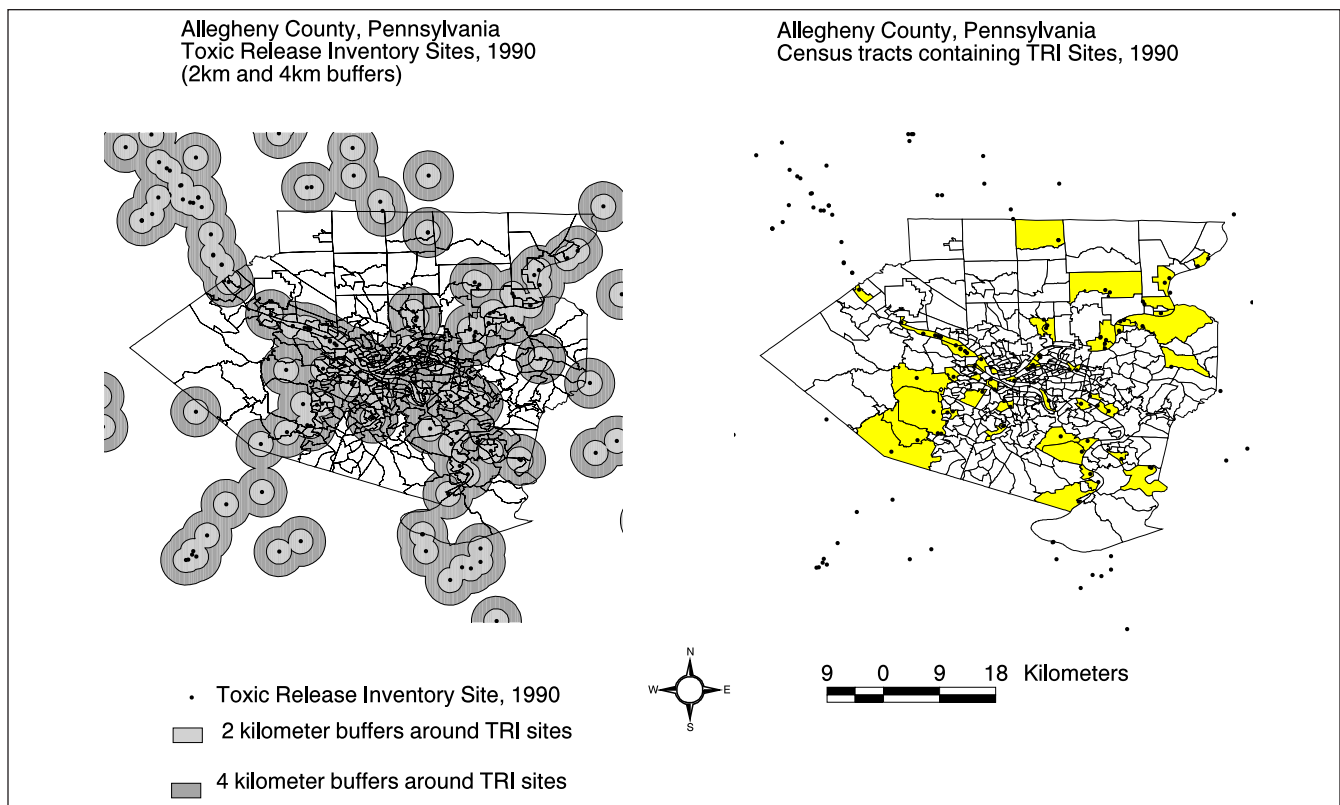


Figure 3. On the left, an illustration of two-, and four-kilometer buffers around sites reporting releases to the Toxic Release Inventory (TRI) in 1990. On the right, exposed tracts based on the “container model” (see text), i.e. census tracts in Allegheny County, Pennsylvania containing TRI sites.

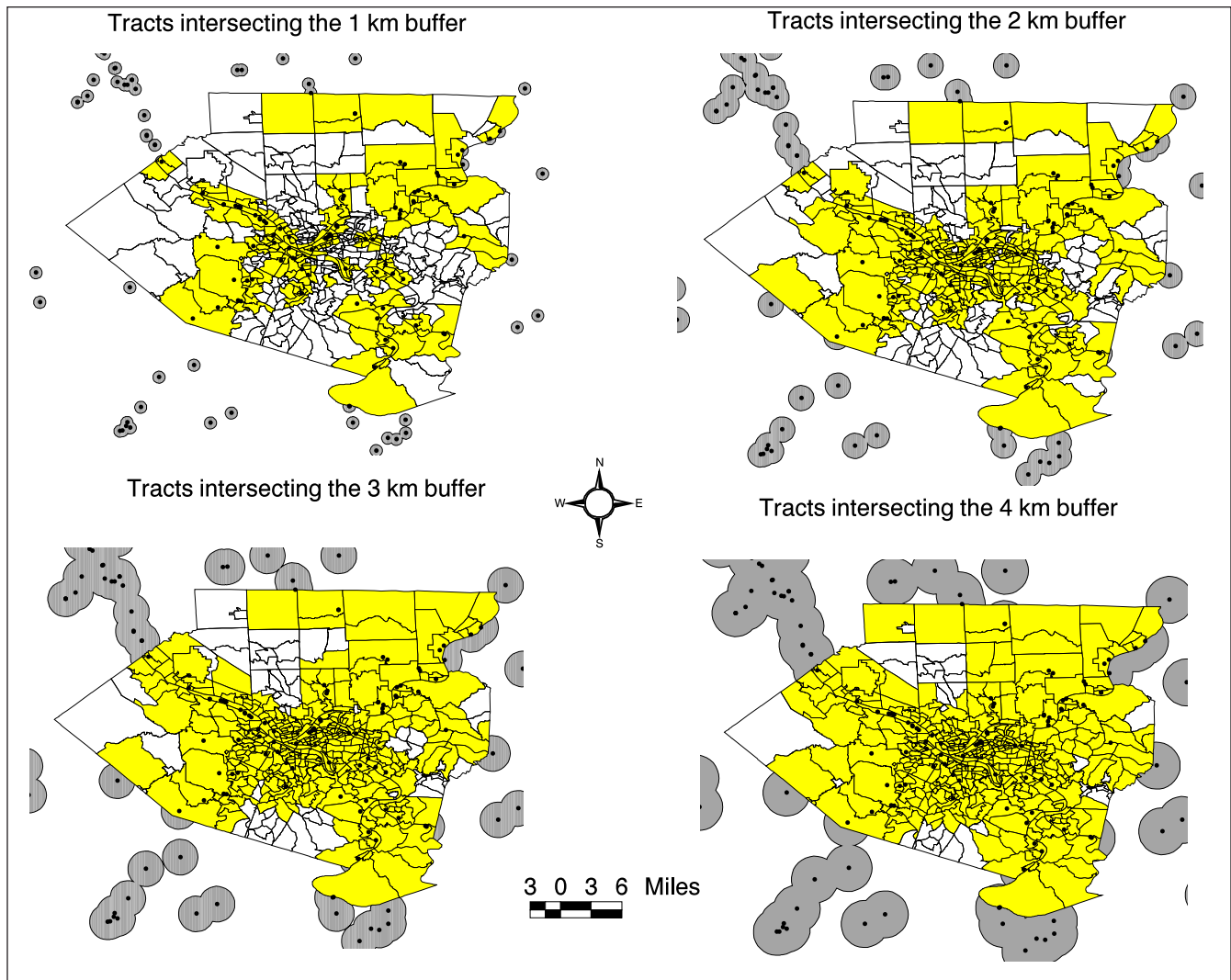


Figure 4. Census tracts intersecting various distance buffers around sites reporting releases to the 1990 Toxic Release Inventory in Allegheny County, Pennsylvania. Considering these tracts “exposed” represents an extension of the container model.

utilize location of hazards to define surrogates of ambient exposures due to the hazards, which in turn are surrogates of individual exposures received from the hazard. Waller et al. (1997) use the term “exposure potential” to refer to such surrogates, to clearly differentiate from measured exposures. The particular form of such exposure potentials generally falls into one of two categories. The first is the “container” model (Talen and Anselin 1998, referred to as a “coincidence” structure by Sheppard et al. 1999) where individuals are considered “exposed” if they reside in a census region (or ZIP code) containing any source of hazard (e.g. a TRI site). The map to the right in Figure 3 illustrates the tracts labeled “exposed” in Allegheny County for 1990 under the container model. The second approach is a proximity or distance based model where exposure potential is defined by the distance from a hazard. The simplest proximity approach is based on GIS

defined “buffers” where individuals living within a certain distance of a hazard location are defined as “exposed” and those outside the buffer are not. Such measures are very common in EJ assessments; one compares the proportion of individuals in each subgroup for individuals within buffer regions to the proportions for individuals outside buffer regions. The map to the left in Figure 3 illustrates 2 and 4 km buffers around the TRI sites in Allegheny County for 1990. To determine the number of individuals residing within the buffer regions, one has to account for the fact that buffer boundaries do not match census boundaries. Therefore, one either includes all tracts having any portion within the buffer in an extension of the container model (an example of this appears in Figure 4), or one assumes populations are uniformly distributed within census regions and allocates a proportion of the population corresponding to the area of the region falling within the buffer. Other proximity-

based surrogates for exposure involve distance-decay functions where one defines exposure potential as a decreasing function of distance.

Both the container and proximity models assume accurate location data. Scott et al. (1997) investigate the accuracy of 620 TRI (1987-1992) locations from South Carolina using GIS to locate the reported street address, and global positioning systems (GPS's) to find the latitude and longitude of each site. They find that 271 of the 620 (48%) of the reported latitude and longitude values placed the site in the wrong census block group. While most adjustments were relatively small, Scott et al. (1997) report substantial effects of location on environmental justice assessments for one site whose true location was more than 10,500 meters from its reported location. The issue of data quality is rather sobering and reinforces the need for inclusion of some measure of accuracy for both attribute and location data.

3.4 Measuring equity/fairness

Once we have data for demographics, exposures (or some proxy measure), and (possibly) outcomes, we next address the issue of equity or fairness in exposures to hazards among our demographically defined subgroups. What is fair? Should hazards be equally distributed among population subgroups? How should exposures be adjusted as demographics change in a particular region? Should all parties have equal say in siting decisions? Is an equitably decided location still fair when new research reveals excess risk of disease due to a substance previously thought to be safe? Phillips and Sexton (1999) provide an interesting and thorough discussion of various definitions of "fair" in EJ from an environmental policy standpoint.

For the purposes of this paper, we will consider our assessment of equity confined to assessment of differences in subpopulations' exposure potential. There are two approaches we could consider. First, we could compare summaries of exposure potential experienced by each subgroup. For example, is the mean exposure potential experienced within subgroup A different from that experienced within subgroup B? Inference typically follows a two-sample *t*-test, or

similar comparison of sample means. Second, we could compare the proportion of each subgroup experiencing a common exposure potential value. For example, what proportion of subgroup A receives exposure *x* or higher? How does this proportion compare to the overall proportion of subgroup A across the study area? In this case, inference follows a comparison of proportions, adjusting for the fact that the exposed individuals represent a subset of the total population in subgroup A. The first approach is similar to diversity indices utilized in econometrics to compare incomes between population subgroups (Gastwirth 1989, Nayak and Gastwirth 1989). Waller et al. (1997) argue that the second approach may be more appropriate for EJ assessments, particularly if one is interested in using observed inequities in exposure (potential) to predict resulting inequities in risk of particular outcomes. In this case, a particular exposure value (say *x*) translates to a risk via a dose-response relationship. The proportion of each subpopulation experiencing a given exposure potential is translated to the risks induced by that level of exposure potential. Naturally, the accuracy of such an approach critically depends on the accuracy of the model linking dose and response.

The results obtained depend on the particular form of exposure potential used. To illustrate, consider the Allegheny County data. Table 1 summarizes the raw population proportions for two race groups from the 1990 census for the container model, and various proximity models. For the buffer analysis, we summarize the proportion of each population subgroup residing in tracts having any portion within the buffer (see Figure 4). We limit the summary to tracts within Allegheny County (even though tracts in the neighboring counties intersect some buffers). Note the differing amounts of inequity suggested by the crude proportions in Table 1. Also note that since most tracts are more than 1 km across, there is considerable overlap in the sets of tracts associated with each buffer radius. Even though results change with buffer radius, the magnitude of the change is small due to the large amount of overlap.

Waller et al. (1997) consider the use of cumulative distribution functions (CDF's) of

Table 1. Percent of residents (U.S. Census) identifying selves as "black" or "white" for tracts containing Toxic Release Inventory sites or intersecting buffers around TRI sites in Allegheny County, Pennsylvania.

Race category	Total county proportions	Site in Tract	Distance to nearest TRI site			
			1 km	2 km	3km	4km
"Black"	11.20%	10.64%	11.28%	12.59%	12.63%	12.05%
"White"	87.56%	88.51%	87.75%	86.19%	86.07%	86.66%

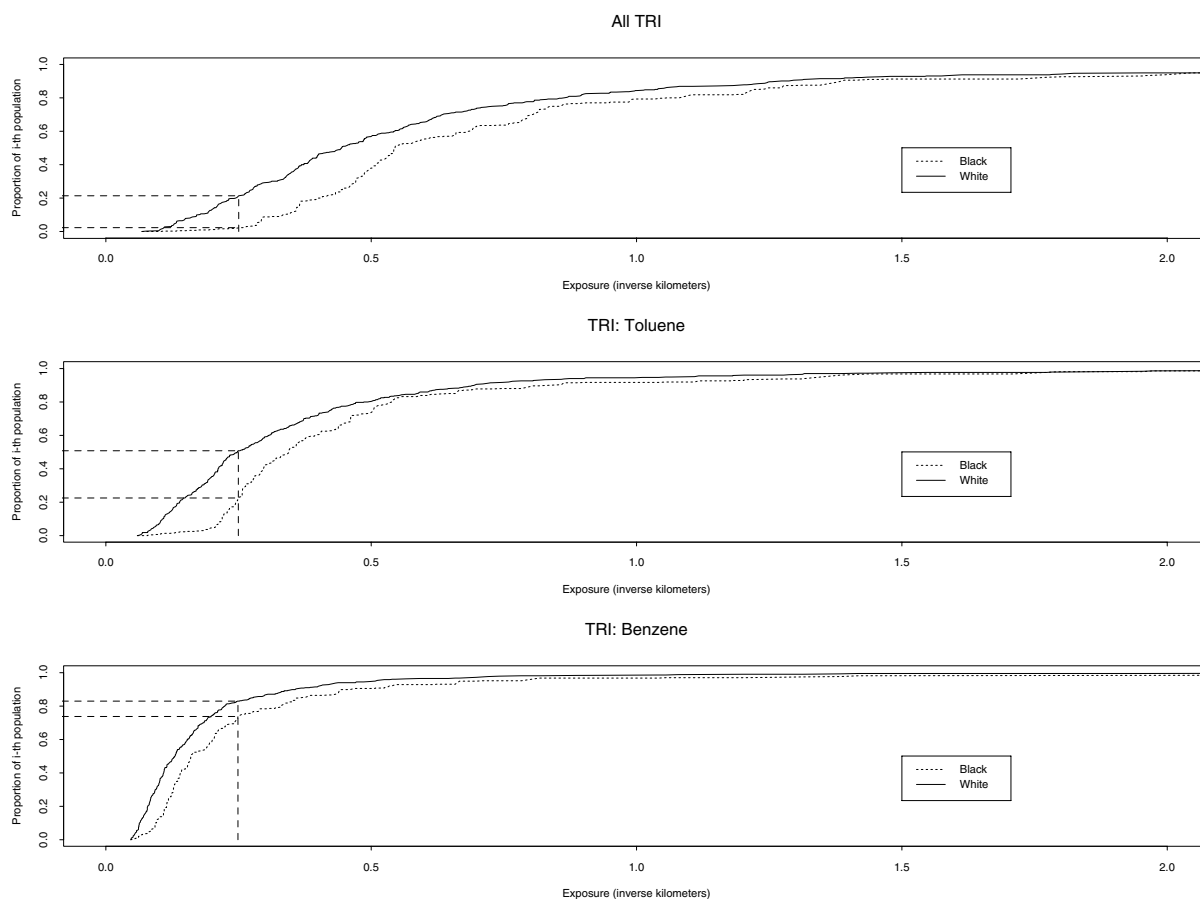


Figure 5. Cumulative distribution functions of an exposure potential (see text) defined by the inverse of the distance between the nearest TRI site and the centroid of the census tract of residence for persons identifying themselves as “black” or “white” in the 1990 Census for Allegheny County, Pennsylvania. The vertical line at $x = 0.25 \text{ km}^{-1}$ provides reference to the proportion of each subpopulation residing farther than 4 km from the nearest TRI site.

exposure potential values within each subpopulation as a way to summarize differences across potential buffer cutoffs. Specifically, suppose we wish to compare the “black” and “white” subgroups with respect to an exposure potential defined by the inverse distance of the center of each census tract to the nearest TRI site. For each tract, we find the minimum distance to any TRI site, then we order tracts by the inverse of these distances (our exposure potential) within each subpopulation. For any value of exposure potential, x , we find the proportion of the total black population in Allegheny County residing in tracts with exposure potential less than or equal to x . Plotting this proportion versus a wide range of exposure potential values results in a graph of the CDF for this subpopulation. We perform a similar exercise for the white population. Inference and descriptive statistics comparing two CDF’s include Kolmogorov-Smirnov tests, Cramér-Von Mises statistics, and familiar QQ and PP plots.

The top graph in Figure 5 illustrates the two CDF curves for Allegheny County, based on the

inverse distance to the nearest TRI site. We also consider exposure potentials based on the inverse distance to the nearest TRI site emitting certain chemicals. Figure 6 shows the locations of all TRI sites with a 4 km buffer, the locations of TRI sites reporting releases of toluene and/or benzene (all 8 benzene sites also reported releases of toluene), and the proportion of each tract population responding “black” to the 1990 Census race question. For all three measures of exposure potential, we see a clear difference between the curves. As indicated by the vertical line in Figure 5, slightly over 20% of the white population experiences an exposure potential less than $x = 0.25$ for all TRI sites (i.e. they live more than 4 km from the nearest TRI site), while only 2.2% of the black population experiences corresponding exposure potentials. That is, only 2.2% of the county’s black population resides outside tracts intersecting the 4 km buffers shown in Figure 6. For sites reporting releases of toluene, approximately 50% of the white population resides further than 4 km away, corresponding to only

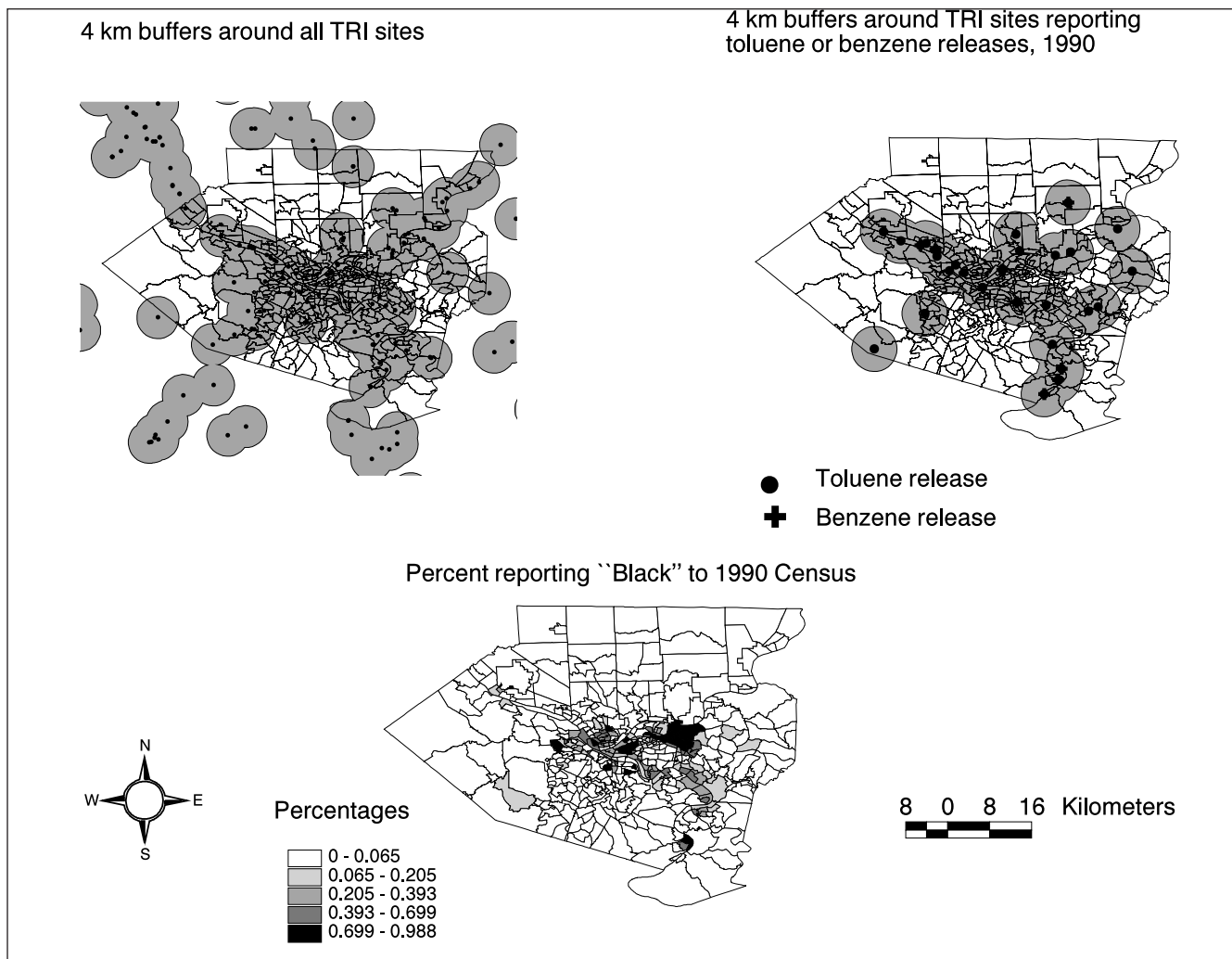


Figure 6. Four-kilometer buffers around all Toxic Release Inventory sites (1990 TRI), and around those sites reporting releases of toluene and benzene. Also, the proportion residing in each census tract responding “black” to the race question in the 1990 Census. The buffer regions allow comparison to the exposure $x = 0.25 \text{ km}^{-1}$ in Figure 5 (see text).

22.5% of the black population. There is less disparity in sites reporting benzene releases, and less exposure potential in the general population. (Since there are fewer sites, distances between tract centroids and sites are generally larger). Note that the “top” CDF in each graph in Figure 5 corresponds to the subpopulation experiencing relatively higher exposure potential values (the exposure potential density would be shifted toward higher values for the associated subpopulation). Also, note that the CDF’s reveal comparisons for any choice of x .

Rather than seeing the different results from different methods as “right” or “wrong”, it is important to realize that each approach addresses only particular aspects of our overall research question. That is, we want to answer: “Are subpopulations subjected to disproportionate environmental exposures?” Using the container model, we answer: “Do subpopulation proportions

differ between tracts containing TRI sites, and those without TRI sites?” Using a buffer-based approach, we answer: “Do subpopulation proportions differ between people residing within the specified radius, and those residing further away?” We can also think of this issue in terms of conditional probabilities. For example, suppose p_B and p_W represent the proportion “exposed” in the black and white subpopulations, respectively. (Contrast these proportions to the proportion of the total “exposed” in each subpopulation as reported in Figure 1). Then it is certainly possible that $Pr[p_B > p_W | \text{container “exposures”}] \neq Pr[p_B > p_W | \text{buffer “exposures”}] \neq Pr[p_B > p_W | \text{proximity “exposures”}]$. The differences are subtle, but crucial for making sense of seemingly differing results. An example involves the United Church of Christ (1987) report which essentially uses a container-based approach and National Priority List Superfund sites (rather than TRI sites), and a

follow-up analysis by Anderton et al. (1994) using census tracts and a general list of Toxic Storage and Disposal Sites (compiled from the TRI and other sources). The UCC report finds evidence of increased percent minority population (non-whites and white Hispanics) in ZIP codes containing the sites of interest. In contrast, Anderton et al. (1994) find little consistent evidence for a general pattern of environmental inequity in the tract level data. Anderton et al. (1994) provide a thorough discussion of reasons for the differences, and conclude that tract level analysis is more appropriate. The level of data aggregation is a key element of any EJ assessment, and clearly provides a context within which one must interpret results. While Anderton et al. (1994) may argue that tract-level analysis provides a more accurate picture of underlying inequities than ZIP code-level analysis, it is important to realize that one is merely exchanging one set of limitations for another. The value placed on any set of limitations inevitably includes some subjectivity and is likely to vary between the different parties interested in EJ issues.

■ 4. Opportunities: where do we go from here?

There remain many areas for future statistical work in the area of environmental justice, and we describe two such areas here. First, one may wish to include outcome data to move from assessments of exposure inequities to risk inequities. Second, the language of “environmental justice” is often particular to the political structure of the United States, however, many related areas of study occur throughout the scientific literature and the integration of such methods and theories may improve environmental justice assessments.

Including outcome data in environmental justice assessments seems a natural goal. However, there is some debate over whether it is absolutely necessary. First of all, including particular outcomes necessarily makes the assessment more restrictive, in that an observed difference in exposures (or potentials) would be discounted if one could not link exposure differences to observable differences in the occurrence of the outcome of interest. Second, it is difficult to observe small increases in risk due to environmental factors, even in the best of data. The effect needs to be strong and consistent to be detectable in observational data. As mentioned above, environmental justice assessments typically involve data collected for other purposes linked through a GIS. In effect, demanding that an assessment make the link to a particular outcome before positive identification of environmental injustice may require more evidence than the data

will be able to provide. In statistical terms, the power to detect the effect may be so small that one rarely observes a “statistically significant” result, even if a practically significant difference exists.

This is not to say that including outcomes is impossible or undesirable. Waller et al. (1997) describe an approach where they find the area between the exposure potential CDF's as a summary of environmental justice across all buffer areas. To emphasize exposure differences relating to the largest increases in disease risk, they weight the integrand by the slope of a fitted dose response relationship. They embed the approach in a hierarchical modeling framework and use Bayesian methods to provide inference. The Bayesian approach allows proper accounting of the multiple levels of uncertainty inherent in the assessment, and provides posterior inference regarding the index of injustice defined by the weighted, integrated exposure difference. They illustrate the approach using Allegheny County leukemia data. The method shows promise, but a thorough evaluation of its performance requires more complete data.

While the environmental justice movement has been a growing grassroots phenomenon in the United States since the mid-to-late 1980's, the issue has links to several areas of related research in other fields. The geography literature contains a considerable number of references addressing social equity in terms of access to public assets such as libraries or parks (see Talen and Anselin 1998 for a nice description of the spatial and statistical issues involved). Such studies assess equity in access to a perceived “good” feature, very similar to the environmental justice assessment of equity in avoidance to a perceived “bad” feature. Of particular interest in the access equity literature is the notion of a “spatial constraint” (Hodge and Gatrell 1976, McLafferty and Ghosh, 1982, and McLafferty 1984), i.e. the current structure of the population limits the amount of equity that can be achieved. For various historical reasons, the current distribution of population subgroups is fairly segregated, so that there are very few locations which provide equal access (or avoidance) to all subgroups. Waller et al. (1999) illustrate this for Allegheny County and show that potential sites for equitably locating a new environmental hazard correspond to a single contour line winding through the county. The precise location of the contour depends on the particular measure of inequity used.

International research investigating the public health impacts of income disparities also bears similarity to methods for environmental justice assessments. As one example, Szwarcwald et al.

(1999) use the Gini coefficient (a measure of uncertainty popular in econometrics) to investigate the impact of income inequality on homicide rates in Rio de Janeiro, Brazil. Similar measures could explore differences in environmental exposures between income groups. Other approaches use “deprivation indices” to summarize differences in socioeconomic status. See Jolley et al. (1992) for an example addressing differences in health outcomes with respect to such indices. Similar methods may prove useful in future environmental justice assessments in the United States.

In summary, we find the issue of environmental justice provides a wealth of interesting statistical opportunities. In the framework of mandated science, assessments must be done. The statistician’s goal is to improve the implementation and interpretation of these assessments. Unlike many standard statistical problems where there is some unknown “true” value waiting to be discovered either through estimation or modeling, the “true” amount of environmental inequity depends on the substance(s) under consideration, the aggregation level of the data, and the particular type of “inequity” under consideration. Hampering any efforts for accurate assessment are issues in data availability and data quality. Numerical results cannot be interpreted outside of the full context of these issues, and the challenge for the statistician is to include such constraints in the analysis, and clearly communicate all results conditional on the operational definitions made in the process.

Acknowledgements

This research was supported in part by the National Institute of Environmental Health Sciences grant R01 1-R01-ES07750 (LAW and EMC), and by the Environmental Protection Agency Science to Achieve Results Fellowship No. U-915240-01-0 (EMC). The views are those of the authors and are not necessarily those of NIH, NIEHS or EPA.

References

Anderton, D.L., Anderson, A.B., Oakes, J.M., and Fraser, M.R. (1994) “Environmental equity: the demographics of dumping”. *Demography*, 31, 229–248.

Billard, L. (1999) “Sampling and the Census 2000”. *STATS*, 25, 6–11.

Clarke, K.C., McLafferty, S.L and Tempalski, B.J. (1996) “On epidemiology and geographic information systems: a review and discussion of future directions”. *Emerging Infectious Diseases*. 2, 85–92.

Croner, C.M., Sperling, J. and Broome, F.R. (1996)

“Geographic information systems (GIS): new perspectives in understanding human health and environmental relationships”. *Statistics in Medicine*. 15, 1961–1977.

Environmental Systems Research Institute, Inc. (1996) “ArcView GIS, The Geographic Information System for Everyone”. Redlands, CA: ESRI.

Environmental Systems Research Institute, Inc. (1990–1995) “Understanding GIS, The ARC/INFO Method”. Redlands, CA: ESRI and New York: Wiley.

Gastwirth, J.L., Nayak, T.K., and Wang, J.-L. (1989) “Statistical properties of measures of between-group income differentials”. *Journal of Econometrics*, 42, 5–19.

Hodge, D., and Gatrell, A. (1976) “Spatial constraint and the location of urban public facilities”. *Environment and Planning A*, 8, 215–230.

Institute of Medicine (1999) *Toward Environmental Justice: Research, Education, and Policy Needs*. Washington, D.C.: National Academy Press.

Jolley, D.J., Jarman, B., and Elliott, P. (1992) “Socio-economic confounding” in *Geographical and Environmental Epidemiology: Methods for Small-Area Studies*, P. Elliott, J. Cuzick, D. English, and R. Stern (eds). Oxford University Press, London.

Kingston, R.S. and Smith, J.P. (1997) “Socioeconomic status and racial and ethnic differences in functional status associated with chronic diseases”. *American Journal of Public Health*. 87, 805–810.

Morgenstern, H. (1998) “Ecologic studies” in *Modern Epidemiology*, K. Rothman and S. Greenland (eds). Lippencott-Raven Publishers, Philadelphia.

McLafferty, S. (1984) “Constraints on distributional equity in the location of public services”. *Political Geography Quarterly*, 3, 33–47.

McLafferty, S.L., and Ghosh, A. (1982) “Issues in measuring differential access to public services”. *Urban Studies*, 19, 383–389.

Nayak, T.K., and Gastwirth, J.L. (1989) “The use of diversity analysis to assess the relative influence of factors affecting the income distribution”. *Journal of Business and Economic Statistics*, 7, 453–460.

Phillips, C.V., and Sexton, K. (1999) “Science and policy implications in defining environmental justice”. *Journal of Exposure Analysis and Environmental Epidemiology*. 9, 9–17.

Continued from page 13

Salter, L. (1988) *Mandated Science: Science and*

continued on page 28

How Soil Composition Affects Density and Water Capacity

A Science Fair Project

■ 1. Introduction

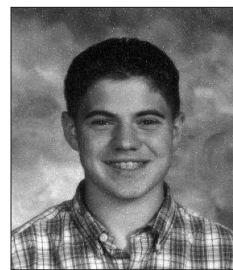
For a middle school science fair project, Mark conducted a mixture experiment to study how soil composition affects soil density and water capacity. Mark mixed different proportions of sand, silt, and clay to simulate different soil compositions. The idea to study density and water capacity of soil came from searching lists of science fair topics on the Internet. The idea of soil being made of sand, silt, and clay came from a science catalog. We combined these two ideas for Mark's science fair project. Through research, Mark learned that soil density and water capacity are important for many reasons. For example, the water capacity of soil is important for irrigation of crops and gardens, and for absorption of rain water or melting snow. Following the science fair, Greg (Mark's dad) applied mixture experiment data analysis methods, which were beyond Mark's middle school capabilities.

A *mixture experiment* involves mixing two or more components in various proportions, and observing the resulting values of one or more response variables. In a basic mixture experiment, a response is assumed to depend only on the relative proportions of the components, and not on the total amount. The component proportions are subject to the constraints

$$0 \leq x_i \leq 1, i = 1, 2, \dots, q \text{ and } \sum x_i = 1, \quad (1)$$

and may be subject to additional constraints on component proportions or linear combinations of

Mark G. Piepel, 649 Cherrywood Loop, Richland, WA 99352, mpiepel@aol.com; Greg F. Piepel, Pacific Northwest National Laboratory, P.O. Box 999, Richland, Washington 99352, greg.piepel@pnl.gov



Mark G. Piepel and Greg F. Piepel

component proportions. Mixture experiments can be augmented to also investigate the effect of a total amount variable, or the effects of one or more process variables (Cornell 1990, Piepel and Cornell 1994).

In the rest of the article, we explain how Mark collected data using a mixture experiment design and analyzed the data using simple graphs. We also present the results of Greg's work to develop and validate mixture experiment models, and to study how the soil components affected soil density and water capacity.

■ 2. Experimental Design

Mark mixed sand, silt, and clay in different proportions to make 10 soil mixtures corresponding to an *augmented simplex centroid (ASC) design* (Cornell 1990, p. 73). As shown in Figure 1, the three-component ASC design consists of the three vertices, three edge midpoints, three interior points, and overall centroid of the three-component simplex. Mark repeated the experiment (described in the next section) four times using a randomized complete block arrangement.

■ 3. Materials and Experimental Procedure

We got free sand from a local concrete company and bought the clay from a ceramics store. We dug the silt from the Yakima River delta. Because the sand and river silt were wet, we spread them in thin layers on two large plastic sheets to dry. Mark used a small shovel to turn the sand and silt, and a spatula to smash silt clods.

Mark used a measuring cup to measure the sand, silt, and clay to make soils of the 10 different compositions shown in Figure 1. He used a knife to level each full measuring cup so the volumes

would be accurate. Mark then put the measured volumes of sand, silt, and clay in heavy-duty, zip-top plastic bags and shook them until evenly mixed. Each bag was marked with the number of the soil mixture given in Figure 1.

Mark measured and mixed the sand, silt, and clay to form the 10 soils only once (not four times) because of time constraints. Thus, the four replicates (blocks) do not include variations from measuring and mixing the soils. However, preliminary testing showed that variations in measuring and mixing soils was small compared to the variations in the remaining steps of the process. Hence, we decided this “limitation” in replicating the experiment would have little consequence.

The steps of the procedure Mark used to measure the density and water capacity of each soil mixture are listed in the appendix. The steps were adapted and combined from procedures described by Campbell (1992) and Cassel and Nielsen (1986). For each of the four replicates (blocks), Mark tested the 10 soil mixtures and measured the responses in a different random order. A photo during Step 7 of the procedure is shown in Figure 2.

The responses of interest are soil density, water capacity by volume (WCV), and water capacity by mass (WCM). The formulas for calculating values of these responses from measured quantities are given in Equations (A.1), (A.2), and (A.3) of the appendix.

■ 4. Results

Mark wrote the results of the procedure steps on data sheets in a notebook. The density, WCV, and WCM measurements from the four replicates of the experiment are given in Table 1. Several values in Table 1 are enclosed in parentheses, denoting outlying data values that were not used in subsequent data summaries and analyses. Mark decided which observations to declare outliers by comparing response values for replicate tests, and by referring to observations he made on data/observation sheets during the experiments. In a few cases, the wet soil in the cylinder separated into two parts, with an air pocket in between. The air pocket may have slowed the water from moving lower in the cylinder, which would cause the water capacity of the soil to be too high. In a few other cases, the soil components were noticeably segmented in the top portion of the cylinder, which was judged to have yielded outlying WCM values.

Evidence of a block effect was inconsistent across the three responses, so we decided to treat the data as if there were no block effects. Mark used Figures 3, 4, and 5 to show the averages (over the four replicates) of density, WCV, and WCM for

each soil mixture. The outliers shown by parentheses in Table 1 were not used to calculate the averages.

Figure 3 shows that when the proportion of sand goes from 0 to 2/3, the density goes up, and then the density goes down when the proportion goes to 1. When the proportions of silt and clay go from 0 to 1/3, the density goes up, and then the density goes down when the proportions go to 1. The density is highest for the soil mixture with 2/3 sand and 1/6 each of silt and clay. The density is lowest for the soil mixture with all clay.

Figure 4 shows that when the proportion of sand goes from 0 to 1, the WCV goes down. When the proportion of silt goes from 0 to 1, the WCV stays about the same. When the proportion of clay goes from 0 to 1, the WCV goes up. The WCV is highest for the soil mixture with all clay. The WCV is lowest for the soil mixture with all sand.

Figure 5 shows that when the proportion of sand goes from 0 to 1, the WCM goes down. When the proportion of silt goes from 0 to 1/3, the WCM goes down, and then the WCM goes up when the proportion goes to 1. When the proportion of clay goes from 0 to 1, the WCM goes up. The WCM is highest for the soil mixture with all clay. The WCM is lowest for the soil mixture with all sand.

■ 5. Fitted Mixture Experiment Models

We observed some trends in the density, water capacity by mass, and water capacity by volume that appear to be related to the proportions of sand, silt, and clay in the mixture. These observations made us wonder whether we could use mixture experiment models to predict the density and water capacity of soil as a function of the proportions of sand, silt, and clay in the soil. Such models, along with analytical procedures for determining the proportions of sand, silt, and clay in a particular soil, could be used to predict the irrigation needs or water absorption properties of a particular soil. We present fitted models for soil density and WCM below. We do not present models for WCV because of space considerations and the complications of having “less than” data for soil #1.

Cornell (1986, 1990) notes the 10-point three-component ASC design supports fitting the Scheffé *special quartic model*:

$$E(y) = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{1123}x_1^2x_2x_3 + b_{1223}x_1x_2^2x_3 + b_{1233}x_1x_2x_3^2 \quad (2)$$

as well as the Scheffé *special cubic model*

$$E(y) = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 \quad (3)$$

Table 1. Replicate Data for the Ten Soil Mixtures ^(a)

Soil#	Density (g/ml)				Water Capacity by Volume (g/ml) ^(b)				Water Capacity by Mass (%) ^(c)			
	Replicate				Replicate				Replicate			
	1	2	3	4	1	2	3	4	1	2	3	4
1	1.64	1.59	1.57	1.63	< 0.130 ^(d)	< 0.105 ^(d)	< 0.100 ^(d)	< 0.100 ^(d)	7.5	8.5	8.0	(4.5)
2	1.38	1.39	1.38	1.53	0.175	0.179	0.182	0.167	12.5	13.0	13.5	10.0
3	0.85	0.88	0.90	0.97	0.227	0.222	0.196	0.217	24.4	24.0	24.5	21.5
4	1.54	1.52	1.60	1.55	(0.312)	0.141	0.128	0.132	9.0	10.0	8.4 ^(e)	8.5
5	1.51	1.61	1.59	1.58	0.172	0.167	0.169	0.169	12.4	11.5	11.0	11.0
6	1.37	1.42	1.44	1.42	0.208	0.196	0.189	0.196	(17.9)	12.5	13.0	13.0
7	1.64	1.64	1.58	1.65	0.175	0.161	0.152	(0.208)	9.5	9.5	10.5	(13.5)
8	1.63	1.68	1.67	1.76	0.143	0.122	0.122	0.119	9.0	8.5	8.5	7.0
9	1.53	1.46	1.54	1.56	0.172	0.156	0.156	0.179	11.5	9.0	11.5	13.5
10	1.19	1.20	1.30	1.41	0.189	0.185	0.196	0.213	17.9	16.5	16.0	15.0

a. Parentheses show data points that Mark identified as outliers by comparing response values for replicate tests and considering observations about unusual appearances of wet soil columns made on data sheets. These data points were excluded from subsequent data analyses.

b. Grams of water divided by milliliters of wet soil.

c. Percentage of wet soil that is water = 100 x (grams of water divided by grams of wet soil).

d. All of the soil (pure sand) in the cylinder was wet, so only less than values can be reported.

and the Scheffé quadratic model

$$E(y) = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 \quad (4)$$

Model (4) contains three linear terms (e.g., b_1x_1) and three quadratic blending terms (e.g., $b_{12}x_1x_2$). Model (3) contains the same six terms as (4), plus the special cubic blending term ($b_{123}x_1x_2x_3$) that provides for modeling response surfaces with humps or valleys. Model (2) contains the same six terms as (4), plus three special quartic terms (e.g., $b_{1123}x_1^2x_2x_3$) that provide for approximating more complicated response surfaces (see Cornell 1990). Over many years, practitioners have found that models of the forms of (2), (3), and (4) adequately approximate the majority of mixture experiment response surfaces. We studied Figures 3, 4, and 5 to decide that models (2), (3), or (4) should provide adequate fits to the soil

mixture data.

Greg fitted Scheffé mixture experiment models of the forms (2), (3), and (4) to the data in Table 1 to determine which was most appropriate (based on lack-of-fit F tests) for each response. Then, he sequentially removed nonsignificant ($p > 0.10$) higher-order terms from the complete Scheffé model form selected for each response. Greg used the Minitab (1998) software to perform the regression work. The resulting fitted models for soil density and WCM are given in Table 2. A reduced form of the special quartic model was obtained for soil density, while a reduced quadratic model was obtained for WCM. Obtaining different mixture model forms for these two responses was slightly surprising to Greg, given that soil density and WCM have a pairwise correlation of -0.962. However, careful study of Figures 3 and 5 (and the data plots in Figures 7 and 8 discussed in the next

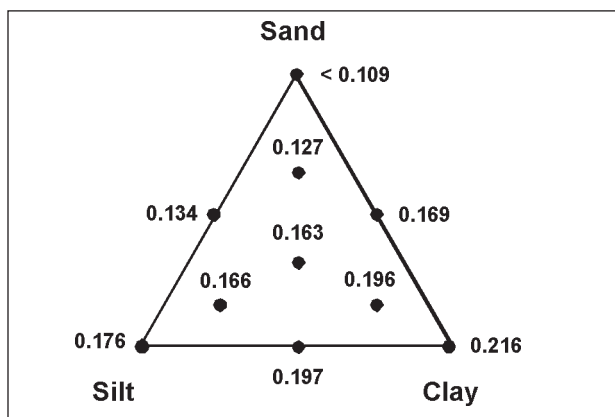


Figure 4. Average Values of Water Capacity by Volume (g/ml)

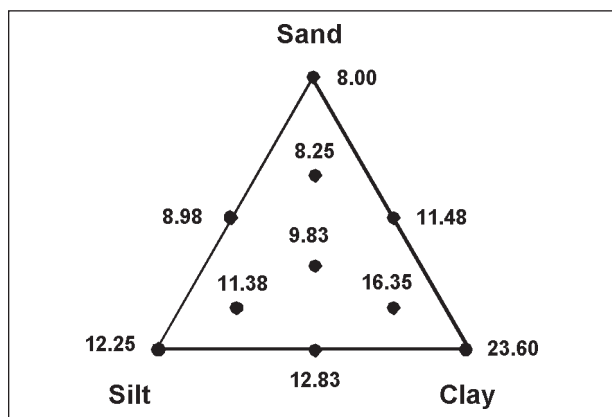


Figure 5. Average Values of Water Capacity by Mass (%)

section) reveals differences in the two responses sufficient to explain why different mixture model forms were obtained.

Table 2 shows the fitted models for soil density and WCM have R^2 values of 0.949 and 0.941, respectively. Further, the fitted models do not have statistically significant lack-of-fits (p -values > 0.10), where “pure error” was estimated using the four replicates on each soil. These findings suggest the fitted models should do an accurate job predicting soil density and water capacity of soil mixtures. Section 7 discusses work performed to validate the models in Table 2 with data not used to develop the models.

■ 6. Effects of Sand, Silt, and Clay

Traditionally, someone determines the effect of a variable on a response by changing the variable while holding all other variables fixed, and observing the changes in the response. This procedure is impossible in mixture experiments, where an increase in one component must be offset by decreases in one or more other components. The most common approach for assessing the effect of a mixture component on a response is to offset an increase in the component with decreases in the remaining components while keeping the remaining components in the same relative proportions as in a reference mixture.

For a simplex mixture region, such as in the soil mixture problem, it is traditional to choose the overall centroid ($\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$) as the reference mixture. Then, the effect of a component is measured along the axis of that component (a line going through the overall centroid and the vertex for that component). Figure 6 displays the component axes for the three-component soil mixture simplex, along with the ASC design points. Figure 6 shows that the ASC design is an *axial design*, since all of the design points fall on the component axes. Hence, we assessed the effects of the components by plotting predicted and measured response values versus component values along the axes.

Figure 7 shows a plot of measured and predicted values of soil density versus values of sand, silt, and clay along the respective component axes. Figure 8 illustrates a similar plot for WCM. The predicted response values are from the fitted models in Table 2.

Figures 7 and 8 suggest the proportion of silt in soil has a small effect on soil density and water capacity compared to the effects of sand and clay. Increasing the proportion of clay in a soil mixture strongly decreases density and increases water capacity. Increasing the proportion of sand in a soil mixture tends to moderately increase density until the soil is made of higher proportions of sand, in

Table 2. Regression Coefficients and Statistics from Fitting and Reducing Mixture Models of the Form (2) and (4) to the Soil Mixture Experiment Data

Mixture Experiment Model Term	Soil Density	Water Capacity by Mass
Sand	1.601	7.288
Silt	1.420	12.029
Clay	0.894	23.855
Sand*Silt	0.156	
Sand*Clay	1.263	-15.911
Silt*Clay	1.010	-18.132
Sand*Silt*Clay		
Sand ² *Silt*Clay	7.529	
Sand*Silt ² *Clay		
Sand*Silt*Clay ²	-6.241	
R^2	0.949	0.941
s	0.056	1.213
LOF p-value	0.138	0.110

which case density decreases. Increasing the proportion of sand tends to decrease water capacity, until the soil is made of higher proportions of sand, at which point the water capacity levels off.

The preceding statements about the effects of sand, silt, and clay on soil density and water capacity summarize general trends. The fitted models in Table 2 indicate the three soil components do have nonlinear blending effects on soil density and water capacity. Hence, general trends do not provide a complete summary.

■ 7. Model Validation

To validate the fitted response models in Table 2, we mixed additional soil samples and tested them using the procedure described in the appendix. The additional soil mixtures are displayed in Figure 9, and their compositions and response values are listed in Table 3. We included the overall centroid (soil #20) in the validation design to provide some basis for verifying that no bias in results exists between the original experiment (data in Table 1) and the validation experiment (data in Table 3).

The results in Table 3 for soil #20 are very close to the results previously obtained for soil #7. This result suggested to us that there was no appreciable bias in the validation experiment results compared to the original experiment results.

Table 4 displays predicted values (from the models in Table 2), measured values, and 95% prediction intervals (95% PI) of soil density and WCM for the 10 validation soil compositions. The measured values of soil density and WCM are within the 95% PI for all 10 of the validation soil

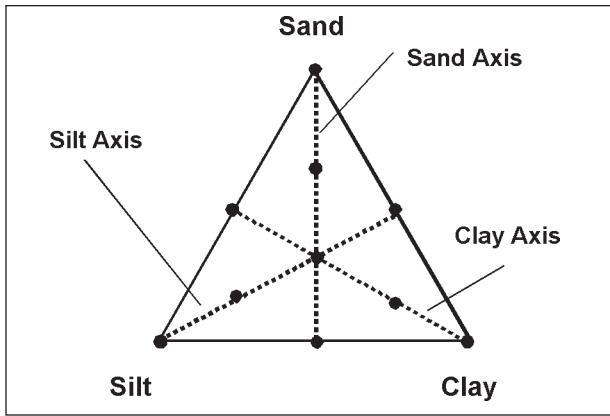


Figure 6. Component Axes Along Which Component Effects are Determined, with Augmented Simplex Centroid Design Points Superimposed

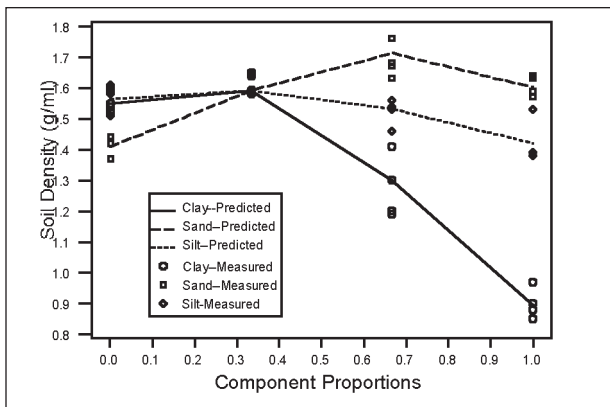


Figure 7. Predicted and Measured Values of Soil Density (g/ml) Along Component Axes

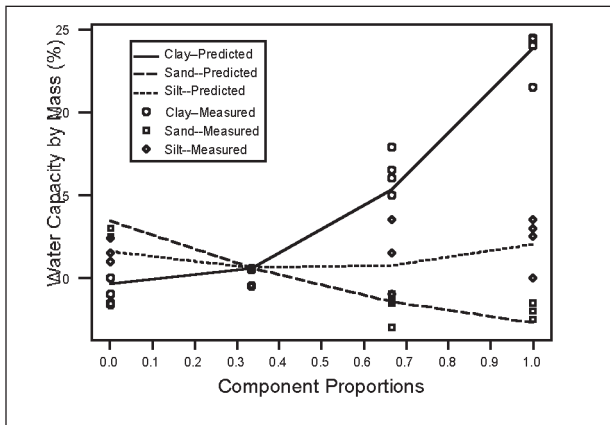


Figure 8. Predicted and Measured Values of Water Capacity by Mass (%) Along Component Axes

compositions. Thus, the Scheffé mixture polynomial models listed in Table 2 should provide excellent predictions of soil density and WCM over the whole three-component simplex composition space for soil mixtures of sand, silt, and clay.

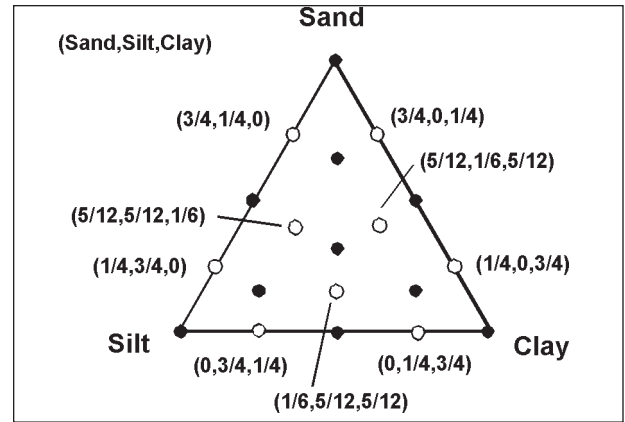


Figure 9. Soil Mixtures for Model Validation (○) Along with ASC Design Soil Mixtures (●)

8. Summary and Recommendations

The density and water capacity of soil depend on its proportions of sand, silt, and clay. Sand increases density and decreases water capacity. Silt decreases density a little, and either increases or decreases water capacity, depending on the proportions of sand and clay. Clay decreases density and increases water capacity. However, sand, silt, and clay have nonlinear blending effects on soil density and water capacity, so that general trends for component effects are not sufficient for predicting results. Scheffé quadratic and special quartic polynomial mixture models provide for adequately predicting soil WCM and density, respectively.

Studying the effects of soil composition on the density and water capacity of soil using mixture experiment methods was a fun and interesting middle school science fair project for Mark. The project won the middle school grand prize at Christ the King School in Richland, Washington, and won second place amongst sixth graders in the Mid-Columbia regional science fair. We note that Mark's science fair project did not include the

Table 3. Soil Compositions and Response Values for Validation Soils

Soil #	Sand	Silt	Clay	Density (g/ml)	WCV (g/ml)	WCM (%)
11	3/4	1/4	0	1.58	0.128	7.0
12	1/4	3/4	0	1.48	0.169	11.0
13	0	3/4	1/4	1.50	0.204	12.5
14	0	1/4	3/4	1.16	0.213	17.5
15	1/4	0	3/4	1.31	0.213	14.5
16	3/4	0	1/4	1.68	0.127	8.0
17	5/12	5/12	1/6	1.65	0.137	8.5
18	1/6	5/12	5/12	1.52	0.179	12.0
19	5/12	1/6	5/12	1.68	0.164	11.0
20	1/3	1/3	1/3	1.62	0.164	9.5

Table 4. Predicted and Measured Density and WCM With 95% Prediction Intervals for Validation Soils

Soil #	Density (g/ml)			WCM (%)		
	Pred.	Meas.	95% PI	Pred.	Meas.	95% PI
11	1.59	1.58	1.46-1.71	8.5	7.0	5.9-11.1
12	1.49	1.48	1.37-1.62	10.8	11.0	8.3-13.4
13	1.48	1.50	1.36-1.60	11.6	12.5	9.0-14.2
14	1.21	1.16	1.09-1.34	17.5	17.5	14.8-20.2
15	1.31	1.31	1.18-1.43	16.7	14.5	14.1-19.4
16	1.66	1.68	1.54-1.78	8.4	8.0	5.8-11.1
17	1.65	1.65	1.53-1.78	9.7	8.5	7.1-12.2
18	1.47	1.52	1.34-1.59	11.9	12.0	9.3-14.5
19	1.59	1.68	1.47-1.71	11.0	11.0	8.4-13.5
20	1.59	1.62	1.49-1.71	10.6	9.5	8.1-13.1

mixture experiment modeling, component effects, and model validation investigations presented in Sections 5, 6, and 7. Those investigations were performed by Greg, with help from Mark, after the science fair was over.

Acknowledgments

Sections 1, 2, 3, 4 and 8 contain condensed and revised portions of a science fair report written by Mark in early 1998 as a sixth-grade student. Greg wrote Sections 5, 6, and 7, and condensed and revised the other sections.

We thank Paula Heller for providing the references on measuring water capacity of soil, Toby Landers of Hanford High School for lending several graduated cylinders, and Rebecca Moak (the science teacher) of Christ the King School. We also thank Nancy Foote of Pacific Northwest National Laboratory for her excellent copy-editing work. Last, but certainly not least, we thank the editor, associate editor, and two referees whose very helpful comments led to several improvements in the article.

References

Campbell, G.S. (1992), "Estimating the Water Holding Capacity of Soils," Department of Crop and Soil Sciences, Washington State University, Pullman, Washington.

Cassel, D.K. and Nielsen, D.R. (1986), "Field Capacity and Available Water," in *Methods Soil Analysis, Part 1*, Editor A. Klute, pp. 901-924, American Society of Agronomy, Madison, Wisconsin.

Cornell, J.A. (1986), "A Comparison Between Two Ten-point Designs for Studying Three-component Mixture Systems," *Journal of Quality Technology*, 18, 1-15.

Cornell, J.A. (1990), *Experiments With Mixtures: Designs, Models, and the Analysis of Mixture Data*, Second Edition, John Wiley & Sons, New York.

Minitab (1998), *Minitab Statistical Software, Release 12 for Windows*, Minitab, Inc., State College, Pennsylvania.

Piepel, G.F. and Cornell, J.A. (1994), "Mixture Experiment Approaches: Examples, Discussion, and Recommendations," *Journal of Quality Technology*, 26, 177-196.

Appendix: Steps of the Procedure for Measuring Soil Density and Water Capacity

The steps of the procedure used to measure density, water capacity by volume (WCV), and water capacity by mass (WCM) of a soil are listed below.

1. Weigh an empty 100-ml cylinder and record its mass in grams using a scale that weighs to the nearest 0.1 g.
2. Fill 100-ml cylinder with the soil and pack the soil by tapping the cylinder on the table. Add soil and pack until the packed volume is 85 to 100 ml.
3. Weigh the cylinder plus soil, and record its mass in grams.
4. Subtract the cylinder mass from cylinder-plus-soil mass to get soil mass in grams.
5. Record the volume of the soil in the cylinder to the nearest ml.
6. Calculate soil density in g/ml using the formula

$$\text{Density} = \text{Mass}/\text{Volume.} \quad (\text{A.1})$$

7. Put 10 ml of water on top of the soil sample and cover with a double layer of plastic wrap held on by a rubber band, and let sit for 20 hours.

8. Record the volume of the dry portion of the soil in the cylinder to the nearest ml.

9. Subtract the volume of dry soil from the total soil volume to get the volume of wet soil (in ml) in the cylinder.

10. Calculate WCV in g/ml using the formula

$$WCV = 10/\text{Volume of wet soil.} \quad (\text{A.2})$$

Here 10 is the ml of water that is assumed to weigh 10 g.

11. Weigh a paper plate and record its mass in grams. Tare the scale.

12. Remove approximately 20 g of wet soil from the top of the cylinder and put on the paper plate.

13. Break up wet soil into small pieces to dry.

14. Weigh the wet soil (with scale tared) and record its mass in grams.

15. Put the plate with wet soil on table to dry for approximately 24 hours.

16. Weigh the dry soil and plate, then subtract the mass of the plate to get the mass of the dry soil in grams.

17. Subtract the mass of dry soil from the mass of wet soil to get the mass of water in wet soil, in grams.

18. Calculate WCM in % using the formula

$$WCM = 100 \times \frac{\text{mass of water in wet soil}}{\text{mass of wet soil}} \quad (\text{A.3})$$

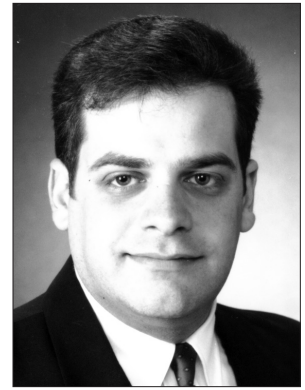
Sidebar by Mark Piepel

The soil mixture experiment was the first science fair project I did, when I was in sixth grade. I'm in eighth grade now. It was very interesting learning about statistics for mixture experiments. I enjoyed working with my dad. It was funny when we went to the river to get silt, and an old man yelled at us for walking across his property. We were suspicious looking wearing winter coats and boots, and carrying a shovel and a bucket!

I was surprised when I won the grand prize at my school's science fair. Also I got out of school for a day to go to the regional science fair where I won second prize for sixth graders. I won ribbons for both prizes and some money at the regional science fair!

Question: Who Chomped Their Way to the 1999 College Bowl Championship?

Answer:



Mark Payton

■ Organize Your Team For College Bowl 2000.

The 1999 College Bowl at the Baltimore Joint Statistical Meetings (JSM), sponsored by the National Statistical Honor Society Mu Sigma Rho and the ASA Section on Statistical Education, was once again successful. Teams representing the graduate programs from the Universities of Florida, Iowa, Maryland and South Carolina participated in the two-day affair. After Tuesday morning's semifinal competition, emceed by Walt Piegorsch of the University of South Carolina, teams from Maryland and Florida advanced to Wednesday's final, emceed by Chuck McCullogh from Cornell University. The loss by the University of Iowa signified at least a lull in the dominance from the Hawkeye State (Iowa or Iowa State have won this competition four out of the last five years). The Maryland team's advancement to the final round continues a College Bowl tradition of local teams doing well (runners-up from previous College Bowls include University of Chicago at the Chicago JSM, UC-Santa Barbara at the Anaheim JSM, and Texas A&M at the Dallas JSM). The Maryland team actually was a conglomeration of students from the campuses at College Park and Baltimore County. Maryland and Florida faced off Wednesday morning, with the Gators from Florida coming away with their first championship. Congratulations Gators!

This year's College Bowl offered some unusual twists and surprises. Due to malfunctioning buzzers, the contestants were forced to strike their water glasses in order to ring in to answer a question. Though it added some subjectivity to the proceedings, many claimed that the Bowl had a nice "down-home feel" and many also preferred the

pleasant "ding" of the water glass to the buzzer. No major controversies erupted due to the lack of electronic buzzers. Another twist to the proceedings occurred since only four teams entered into the competition. A special game was scheduled for Wednesday's action pitting the champion Florida team against a team of faculty members. The Faculty Team consisted of past emcees Bob "Boss" Hogg of the University of Iowa, Linda Young of the University of Nebraska, John Boyer of Kansas State University, Bill Warde of Oklahoma State University. The Faculty got off to a quick start in the game, taking a commanding lead over the Gators who appeared a bit hesitant at the beginning. But soon the Gators woke up and started to gain momentum. The score was tied with one toss up and bonus question left. John Boyer saved the day for the Faculty Team by answering the last toss up question. Special thanks go to the Faculty Team for their willingness to participate.

■ What Happens In College Bowl?

The Bowl is a single elimination tournament played each year at the Joint Statistical Meetings. Teams of four players are used. Each match consists of a fixed number of toss up questions (usually around 15) and usually lasts 15-20 minutes. Each time a team answers the toss up question correctly, they are awarded a bonus question. All questions are worth 10 points. Many subjects are used in College Bowl, including questions regarding the history of statistics, mathematical statistics, statistical methodology, to questions about ASA and its officers and journals. Occasionally a "groaner" question is asked to keep the proceedings light. And often a participant will give a comical answer to a question that gets the audience laughing. It's not necessary to study for the competition, though many of the successful teams in the past (including Florida) have organized study sessions. The important thing to remember is that College Bowl is

Mark Payton is an assistant professor in the Statistics Department at Oklahoma State University.



Florida's championship team, from left, Angel Novikov, Brian Caffo, Ziyad Mahfoud, Philip McGoff and Galin Jones.

meant to be fun and a way to get students involved in ASA and JSM.

■ How Does Participation Benefit Me?

Besides the fun of participation and the knowledge and useful information learned, every student participant in every College Bowl has received an award. The 1999 College Bowl again saw excellent corporate sponsorship, with very generous awards offered by MathSoft, SAS, Minitab, StatXact, Marcel Dekker, John Wiley & Sons, Addison-Wesley and Duxbury, with a total dollar value exceeding \$10,000! Awards ranged from student versions of popular statistics software, to textbooks, to several full professional versions of statistical software. Award selection was by "playground rules": the members of the winning team (in random order, of course) chose first, from all the award vouchers. Next, the second-place team chose from remaining vouchers, and so on. College Bowl is also a great way to meet students and faculty from other institutions.

■ How Do I Get My Team Organized?

The College Bowl is scheduled to return again in 2000 at Indianapolis. It is anticipated that several of the '99 sponsors will return as sponsors in '00, and there will surely again be awards for the 8 teams

participating. Teams consist of 4 players (no alternate). An eligible player must have been a student in good standing (at any level) at some time during the 2000 calendar year. Joint teams from several universities/colleges are welcome. Rules of play and sample questions, as well as information on Mu Sigma Rho, can be found at <http://www.stat.sc.edu/msrnatl.html>.

To reserve a spot for your team, there are no registration fees or forms to fill out. All we ask is that you (1) reserve your spot in good faith (i.e. make a commitment to actually field a team), and (2) submit the names of you team's players. In years past, teams have been asked to provide questions for the competition. This is no longer required of the participating teams. Teams are encouraged, however, to submit questions if they so desire. If you expect to field a team, please notify Mark Payton as soon as possible: mpayton@okstate.edu. Your team is not officially registered, though, until it submits the names of its team members. Teams may register up to the Joint Statistical Meetings. However, only eight teams will be accepted. If byes are necessary in the first round, they will be awarded to registrants in the order they registered.

Hope to see you in Indianapolis!



Faculty team, from left, John Boyer, Linda Young, Bill Warde, and Bob Hogg (not shown). In the background is moderator Chuck McCullough.

■ 1. Question

How important is randomization?

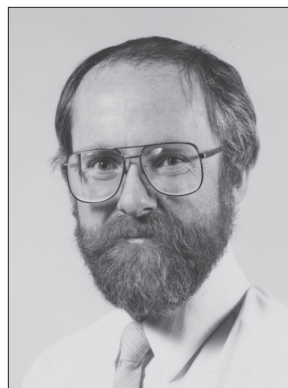
■ 2. Activity

Across the midwest farmers are constantly looking for that competitive edge that will increase profits. Responding to this demand, seed companies have developed, through cross breeding, hybrid varieties of corn with higher and higher yields. More recently, through genetic engineering, there are now corn varieties that are resistant to the affects of herbicide residue and others that can combat pests like the European corn borer. Once a variety of corn is developed, the true test of its values comes in field trials. A field trial is a designed experiment used to compare varieties of corn (or soybeans, or wheat, etc.) in terms of average yield (or some other measure of quality). Sir Ronald Fisher developed many of the methods of applied statistics while analyzing agricultural field experiments at Rothamsted in England. The following activity simulates an agricultural field experiment, or field trial, conducted to compare two varieties of corn, A and B.

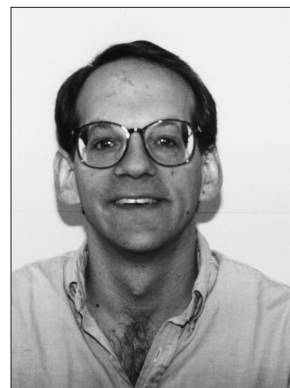
■ Class Activity Introduction

Researchers at a large seed company are planning a field trial to compare two hybrid varieties of corn. The response of interest is the yield, in bushels per acre. The better variety will be the one with the highest yields but the researchers recognize that variation in soil composition, fertility and drainage will have effects on the growth of plants and thus yield. There is a field with 36 plots available for the experiment. On 18 plots variety A will be planted and on the other 18 plots variety B will be planted. The researchers wish to see if the two varieties have equal yields, on average, or if the two varieties differ significantly. If the two varieties really do differ, the researchers would like their experiment and the

W. Robert Stephenson and Hal Stern are University Professors of Statistics at Iowa State University, Ames, IA.



**W. Robert
Stephenson**



Hal Stern

subsequent statistical analysis to detect this true difference. The ability of a statistical procedure to detect a true difference is called the power of the procedure. The researchers must decide how to assign the varieties to the plots.

Convenience Assignment

It is easiest to plant one variety on 18 plots on one side of the field and the other variety on the 18 plots on the other side. Modern machinery can plant up to 18 rows at a time, so planting in this way can be done in one or two passes through the field. Below is a picture of such an assignment and the yields, in bushels per acre, for each plot.

A	A	A	B	B	B
130	149	139	155	137	145
A	A	A	B	B	B
149	133	152	131	147	136
A	A	A	B	B	B
141	156	137	146	132	148
A	A	A	B	B	B
150	142	155	136	152	133
A	A	A	B	B	B
139	155	139	147	137	153
A	A	A	B	B	B
155	138	150	137	145	136

Summary	n	mean	std. dev
A	18	144.9	8.29
B	18	141.8	7.65

Based on this assignment, by convenience, does one variety appear to have a larger mean yield? Is there a significant difference in mean yields between the two corn varieties?

Systematic Assignment

Many people think that an alternating sequence is a random, or at least an unbiased, sequence. For example, when assigning participants to treatment and control, taking every other participant (alternating) for the treatment group appears random. However, if participants are lined up alternating between male and female then all the males will be in one group and all the females in the other. Gender and group would be completely confounded. That is the effects of treatment and control are inseparable from gender effects. In a field, an alternating pattern would be like a checkerboard. Below is a picture of such an alternating pattern and the yields, in bushels per acre, for each plot.

A	B	A	B	A	B
130	137	139	155	149	145
B	A	B	A	B	A
137	133	140	143	147	148
A	B	A	B	A	B
141	144	137	146	144	148
B	A	B	A	B	A
138	142	143	148	152	145
A	B	A	B	A	B
139	143	139	147	149	153
B	A	B	A	B	A
143	138	138	149	145	148

Summary	n	mean	std. dev
A	18	142.3	5.75
B	18	144.5	5.37

Based on this assignment, alternating, does one variety appear to have a larger mean yield? Is there a significant difference in mean yields between the two corn varieties?

Discuss the results from the analysis of the convenience assignment data and those from the analysis of the alternating assignment data. Some may find it a bit disturbing that B appears better for one assignment while A appears better for the other. Of course, this could be due to chance variation. It could also be due to a poor assignment of treatments. For example, the right side/left side assignment is vulnerable to bias due to soil fertility, or drainage that is different from one side of the field to the other. The checkerboard assignment is also susceptible to fertility, drainage or other gradients.

Random Assignment

What if chance is used to assign varieties to plots? How, physically, would you randomly assign

varieties to plots? Come up with a randomization scheme to assign variety A to 18 plots and variety B to the remaining 18 plots. Record your

Randomized Assignment

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

assignments in the table below.

Once you have completed your random assignment, ask your instructor for “The Truth” — this sheet gives the yield for each plot using either variety. “The Truth” was used to fill in the yields for the plots in the convenience and alternating patterns you looked at earlier. In general, “The Truth” is not available since it requires knowing what would happen to the same plot of land using each of the treatments.

Write down the yields for your random assignment — if you have an A in the row 1, column 1 plot then you would put down 130 whereas if you have a B in the row 1, column 1 plot you would put down 118 for the yield. Repeat for all squares. This gives you 18 A yields and 18 B yields. Based on this assignment, at random, did you find a significant difference in mean yield between the two corn varieties?

Share and discuss your results. Examine “The Truth” more closely. Which variety appears to have the larger yield? By how much?

3. Suggested Solution

Convenience Assignment

Using a two independent sample analysis to compare the mean yields of the two varieties the value of the t-test statistic is 1.17 with an associated two sided P-value of 0.25. The P-value is the same whether you use the pooled or non-pooled option on the TI-83. If you are using the conservative degrees of freedom, $\min(n_1-1, n_2-1) = 17$, the P-value would be 0.26. Although variety A has a slightly larger mean yield, there is not a statistically significant difference between the sample mean yields for the two varieties.

THE TRUTH

A = 130 B = 118	A = 149 B = 137	A = 139 B = 127	A = 167 B = 155	A = 149 B = 137	A = 157 B = 145
A = 149 B = 137	A = 133 B = 121	A = 152 B = 140	A = 143 B = 131	A = 159 B = 147	A = 148 B = 136
A = 141 B = 129	A = 156 B = 144	A = 137 B = 125	A = 158 B = 146	A = 144 B = 132	A = 160 B = 148
A = 150 B = 138	A = 142 B = 130	A = 155 B = 143	A = 148 B = 136	A = 164 B = 152	A = 145 B = 133
A = 139 B = 127	A = 155 B = 143	A = 139 B = 127	A = 159 B = 147	A = 149 B = 137	A = 165 B = 153
A = 155 B = 143	A = 138 B = 126	A = 150 B = 138	A = 149 B = 137	A = 167 B = 145	A = 148 B = 136

Alternating Assignment

Using a two independent sample analysis to compare the mean yields of the two varieties the value of the t-test statistic is -1.20 with an associated two sided P-value of 0.24. The P-value is the same whether you use the pooled or non-pooled option on the TI-83. If you are using the conservative degrees of freedom, $\min(n_1-1, n_2-1) = 17$, the P-value would be 0.25. Although variety B has a slightly larger mean yield, there is not a statistically significant difference between the sample mean yields for the two varieties.

Random Assignment

How one randomly assigns varieties to plots is a good class discussion question. Some students might suggest flipping a coin for each plot; heads = A and tails = B. This is random but will not assure 18 plots with variety A and 18 with variety B.

One way to randomly assign the varieties to the plots is to use a die.

—Roll the die, this will give the row number for the plot

—Roll the die again, this will give the column number for the plot

—Assign variety A to the plot with the (row,column) numbers from above

—Repeat the steps above until 18 plots have variety A

—Fill in the remaining 18 plots with variety B

Another way to randomly assign the varieties to plots is to use the TI-83 to generate a random assignment. Essentially what we want to do is to select 18 of the 36 plots at random to receive variety A. The remaining 18 plots (they are chosen at random by default) will receive variety B. To do this first label the plots sequentially from 1 to 36. Then using the TI-83 calculator;

—Put the numbers 1, 2, 3,..., 36 in L1.

—Generate 36 uniform random digits in L2.

Math → **PRB** → **1:rand** → **ENTER** rand (36) → **STO** → **L2**

—Arrange L2 in ascending order while carrying the entries from L1 along. 2nd → **LIST** → **OPS** → **1:SortA**(→ **ENTER** SortA(L2, L1) → **ENTER**

—Read off the first 18 numbers in list L1. These plot numbers will receive variety A. The remaining plot numbers will receive variety B.

Example Randomization with yields

A 130	A 149	A 139	A 167	B 137	A 157
A 149	B 121	B 140	B 131	A 159	B 136
A 141	A 156	A 137	A 158	A 144	A 160
A 150	B 130	B 143	B 136	B 152	B 133
B 127	A 155	B 127	A 159	A 149	B 153
B 143	B 126	B 138	A 149	B 145	B 136

Summary	n	mean	std. Dev
A	18	150.4	9.49
B	18	136.3	8.74

Using a two independent sample analysis to compare the mean yields of the two varieties the value of the t-test statistic is 4.64 with an associated two sided P-value that is virtually zero. The P-value is the same whether you use the pooled or non-pooled option on the TI-83. Even using the conservative degrees of freedom, $\min(n_1-1, n_2-1) = 17$, the P-value is virtually zero. Varieties have different mean yields and that difference is statistically significant.

Closer examination of “THE TRUTH” reveals that variety A has a yield that is 12 higher than variety B on every plot. The true difference in yield between variety A and variety B is 12 bushels per acre.

■ 4. Discussion

Assignment by convenience or using an alternating pattern failed to uncover the true difference between the two varieties. "THE TRUTH" was set up in such a way that the convenience pattern and alternating pattern would mislead the experimenter. If you look closely at "THE TRUTH" you will see that there are alternating high/low yield gradients running diagonally across the field. By planting one variety on one side of the field, or in the alternating pattern, the superiority of variety A is hidden by these diagonal yield gradients. In real fields the truth is not known but non-random assignment of varieties to plots can mislead the experimenter in much the same way. The hidden patterns in real fields can confound the effects of the varieties.

Randomization, the random assignment of varieties to plots tends to take hidden patterns (or lurking variables) and spread their effects evenly across the treatment groups. This allows us to see the underlying truth most of the time. This disclaimer, "most of the time," is important. Even with randomization, we are not guaranteed to find a statistically significant difference even when a real difference does exist. In fact, the chance that a test of hypothesis can detect a difference when one exists is called the power of the test. By looking at the results of tests based on many random assignments, this activity can be used to simulate the power of the two sample t-test to detect a difference in mean yield of 12 bushels per acre. When this randomization activity was done by 40 AP statistics teachers at a short course, all but one of the teachers obtained a t-test statistic that was statistically significant. That is, the simulated power was 39 out of 40 or 97.5%.

■ 5. More on Power

Let's look at power in a little more detail. What we would like to know is of all the possible randomizations of varieties to plots how many would produce a significant difference in sample mean yields? There are over 9 billion possible randomizations so enumerating all of them is out of the question. We can tackle this problem theoretically with some simplifying assumptions. For the two sample problem, it is easiest to look at power assuming normally distributed values with a common, and known variance. For the corn yield example we might assume that the yields for variety A are normally distributed with a mean μ_A and variance $\sigma^2=87$. Additionally, let's assume that the yields for variety B are normally distributed with a mean μ_B and variance $\sigma^2=87$. The value 87 for the population variance is obtained from the values reported in "THE TRUTH." We need to first establish what is a statistically significant

difference. To do this we can use the 68-95-99.7 (or empirical) rule. Recall that approximately 95% of normally distributed values are within 2 standard deviations of the mean. So any difference whose absolute value is greater than 2 standard deviations is statistically significant at approximately the 5% level. We have the variances for individual yields but we need the variance (to get the standard deviation) of the difference in sample mean yields.

Sample mean yields ($n=18$) for variety A will be normally distributed with a center at μ_A and a variance

$$\frac{\sigma^2}{n} = \frac{87}{18} = 4.833$$

Similarly, sample mean yields ($n=18$) for variety B will be normally distributed with a center at μ_B and a variance

$$\frac{\sigma^2}{n} = \frac{87}{18} = 4.833$$

The difference in sample mean yields will be normally distributed with a center at $\mu_A - \mu_B$ and a variance of

$$\frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{87}{18} + \frac{87}{18} = 9.667$$

Thus the standard deviation for the difference in two sample mean yields ($n=18$) is

$$\sqrt{9.667} = 3.11$$

Any absolute difference in sample mean yields larger than two standard deviations (6.22) would be considered statistically significant.

To calculate the power all we would need to do is to compute the probability of getting a difference in sample mean yields that is less than -6.22 or greater than 6.22 when we assume the true difference in means $\mu_A - \mu_B = 12$. This is just the probability that a normal random variable with mean 12 and standard deviation 3.11 takes on a value less than -6.22 or greater than 6.22 . We can obtain the standardized values

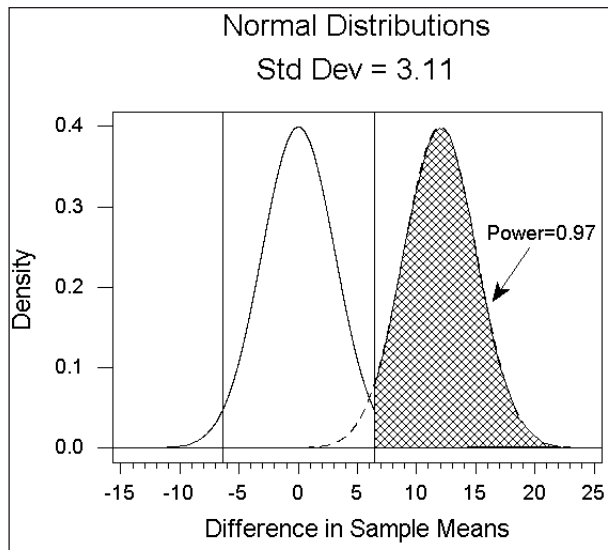
$$z_1 = \frac{-6.22 - 12}{3.11} = -5.86$$

$$z_2 = \frac{6.22 - 12}{3.11} = -1.86$$

The normal cumulative distribution function (cdf) for z_1 is zero and so contributes nothing to the power calculation. The cdf for z_2 is 0.03, so the chance of being greater than $z_2 = -1.86$, and thus the power, is $1 - .03 = .97$. The computation of the power is illustrated in the figure below.

Power is actually a function of how big a

difference you want to detect. In the calculation above, the true difference of 12 will be picked up most of the time by a two independent sample test when randomization is used to assign varieties to plots. The power will be much lower for smaller true differences. You can adjust “THE TRUTH” so that variety A beats variety B by say 6 bushels. You will find that the power as calculated above (think about moving the right hand normal curve in the figure above so that it is



centered at 6 instead of 12) is less than before (around 0.50). Power is clearly a function of the size of the true difference. Procedures have more power to detect large differences than small differences. Power is also affected by sample size. We know that larger sample sizes are good because they reduce the variation in the sample mean. It is nice to know that larger sample sizes also provide more power for much the same reason. Think about how the

Scientists in the Making of Standards. Dordrecht, Holland: Kluwer Academic Publishers.

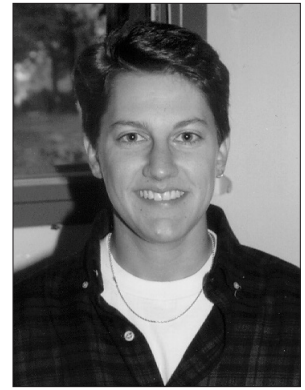
- Scott, M., Cutter, S.L., Menzel, C., and Ji, M. (1997) “Spatial accuracy of the EPA’s environmental hazards databases and their use in environmental equity analysis”. *Applied Geographic Studies*. 1, 45–61.
- Sexton, K. Olden, K., and Johnson, B.L. (1993) “‘Environmental justice’: the central role of research in establishing a credible scientific foundation for informed decision making”. *Toxicology and Industrial Health*, 9, 685–727.
- Sheppard, E., Leitner, H., McMaster, R.B., and Tian, H. (1999) “GIS-based measures of environmental equity: exploring their sensitivity and significance”. *Journal of Exposure Analysis and Environmental Epidemiology*. 9, 18–28.
- Szwarcwald, C.L., Bastos, F.I., Viacava, F., and Andrade, C.L.T. (1999) “Income inequality and homicide rates in Rio de Janeiro, Brazil”. *American Journal of Public Health*. 89, 845–850.
- Talen, E., and Anselin, L. (1998) “Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds”. *Environment and Planning A*. 30, 595–613.
- United Church of Christ (1987) *Toxic Wastes and*

Race in the United States: A National Report on the Racial and Socio-Economic Characteristics of Communities with Hazardous Waste Sites. New York: United Church of Christ.

- Wagener, D.K. and Williams, D.R. (1993) “Equity in environmental health: data collection and interpretation issues”. *Toxicology and Industrial Health*, 9, 775–795.
- Wallace, L.A. (1995) “The exposure of the general population to benzene”. *Cell Biology and Toxicology*. 5, 297–314.
- Waller, L.A., Louis, T.A. and Carlin, B.P. (1997) “Bayes methods for combining disease and exposure data in assessing environmental justice”. *Environmental and Ecological Statistics*. 4, 267–281.
- Waller, L.A., Louis, T.A., and Carlin, B.P. (1999) “Environmental justice and statistical summaries of differences in exposure distributions”. *Journal of Exposure Analysis and Environmental Epidemiology*. 9, 56–65.

Student Voices

On Becoming a Statistics



**Jacqueline
B. Miller**

Once upon a time I wanted to become a veterinarian. Like many other people, I wound up in the corporate world instead, leaving my childhood dream long behind. After graduating from Miami University with a B.A. and a B.S., both in mathematics and statistics, I entered an actuarial consulting firm. It did not take long for me to realize that the corporate world was not for me, and that I wanted to go back to school to earn a Ph.D. in order to teach at the college level. It took a few years for me to get my ducks in a row, but I managed to find my way into the doctoral program in statistics at The Ohio State University.

Dr. Doug Wolfe can take the initial blame (or receive the initial credit!) for setting me loose on the statistics education community, as he was the first person at Ohio State to whom I talked about my desire to teach at the college level. He was the person who initially talked to me about the One-of-a-Kind degree program at Ohio State. This program allows students to combine multiple disciplines into a unique personalized program and would allow me to develop a program combining statistics with education.

I did indeed pursue a One-of-a-Kind degree in statistics education at Ohio State. By combining the core coursework of the doctoral program in statistics with carefully chosen courses from a doctoral program in education, I created a program for myself that would start me on my path to becoming a statistics educator.

Some people might wonder why I couldn't just pursue a Ph.D. in statistics and teach at the college level. I could, but I didn't want to. I knew (and still know) many faculty members with doctorates in statistics who are wonderful teachers. So, I knew it could be done. However, I felt that I needed a background in education as well as statistics in order to be the best statistics teacher I could be.

Jackie Miller is a doctoral candidate in statistics education under Dr. Emmalou Norland and Dr. Bill Notz at The Ohio State University. Jackie is currently doing a job search for a tenure-track position that will allow her to continue teaching statistics and researching the teaching and learning of statistics.

It's hard for me to describe why I felt that I needed to learn about the world of education in addition to the world of statistics. In some ways, it's so simple. I went to graduate school knowing that I had the end goal of *teaching* statistics. Since teaching was my area of interest, learning about teaching and learning seemed natural. Just as other doctoral students in statistics have areas of concentration like missing data, phylogenetic trees, and nonparametric theory, my area of concentration is the teaching and learning of statistics. When I first entered graduate school, I thought I had no interest in doing research. What I found during my graduate school years was that I do have an interest in research, but my research interests center around issues of statistics education more than around the content of statistics.

Instead of trying to describe why I wanted my degree to combine statistics with education, perhaps I can share with you the education classes that opened my eyes to many things. While immersed in a community of educators, I learned more about teaching and learning than I ever imagined I would. My coursework included:

- Six quarters of seminar with doctoral students in the Mathematics, Science, and Technology program. Topics in these seminars included:
 - How to read and share educational research articles and topics with others;
 - How to begin writing a theoretical framework, a literature review, and a dissertation proposal;
 - How to construct a data collection instrument;
 - How to do a pilot study;
 - Qualitative research methodology in mathematics education; and
 - The history of mathematics education.

■ A course devoted entirely to learning theories in the fields of mathematics and science. Here I learned about Bruner, Piaget, and Vygotsky; scaffolding (where a “more knowledgeable other” assists a learner in knowledge construction); information processing; and enough about constructivism to get me interested in researching students’ construction of knowledge.

■ A college teaching course where I learned many instructional strategies, including the jigsaw and minute papers, that I have been able to apply in my own classrooms.

■ A course in multicultural education where I learned to appreciate others’ differences and to not apologize for who I am as an individual.

■ Two courses (“Women, Technology, and Education” and “Computers in the Classroom”) with Dr. Suzanne Damarin, who not only helped me discuss the issues of the courses, but pushed me to stretch the boundaries of my world.

■ A course called “Experimentation” where I was exposed to many educational technologies including CBLs and CBRs (data collection devices for use with Texas Instruments calculators), Measurement in Motion (an interactive measurement software package that shows how measurement is used in real life situations), and ProbSim (a probability simulator).

■ Three courses in qualitative research where I went through a paradigm shift from staunch positivist researcher relying on quantitative methods only to interpretivist researcher relying on qualitative methods. I also learned to match research methods with my developing research questions.

In addition to my coursework in education, my experience with people in the College of Education taught me things that are not taught in the classroom. I met new people who are

researching many different things, but who support of the work of others and respect their diversity.

I cannot imagine *not* having had the experiences that I have had through the College of Education. It has been my experience in education, partnered with my experience in statistics that has helped me grow as a statistics educator.

The field of statistics education is not new. Research on teaching statistics has been around for decades. And, if you think about it, all statisticians are teachers to some extent, be it in the classroom or with a consulting client. Even so, as far as I know, at the time of my graduation (June 9, 2000) I will have the only existing doctorate in statistics education. I don’t see my interest in statistics education as unique, and I would like to help others pursue degrees in statistics education. In addition, I look forward to continuing research on the teaching and learning of statistics.

You know, once upon a time I wanted to be a veterinarian. Instead, I have fulfilled my love for animals as the proud mom of two dogs and two cats. And, as far as my career goes, I am very content with my choice to become a statistics educator. Teaching statistics is fulfilling for me, and doing research on the teaching and learning of statistics is a natural partner for my teaching. If you are interested in statistics education and/or if some of what I have said here rings true for you, go for it—approach your advisor about becoming a statistics educator. For details about my program and/or information about how I was able to pursue this degree, feel free to contact me at Miller.203@osu.edu.

**As far
as I
know,
at the
time**

