

**STATISTICS Ph.D. QUALIFYING EXAM: TAKE HOME**  
**Due 2PM on Mon. August 18, 2003. Return to Dept Math and Stat Office**

*Directions:* The exam has two questions of equal value. Your answer to each problem should be word-processed, double spaced, and should be no longer than three pages. An appendix is allowed for each problem but will be examined only at the discretion of the graders. The better constructed your appendix, the more likely it is to get examined.

1. The data for this problem are stored in an EXCEL spreadsheet and can be downloaded from the WWW site: <http://www.math.unm.edu/~bedrick>. Look for link titled: Comprehensive Exam: EXCEL Spreadsheet for Problem 1.

In the early 1960's, several studies were conducted to assess whether pollution kills people. The data described below are from one study of 60 Standard Metropolitan Statistical Areas (SMSA) in the United States, obtained for the years 1959-1961. From left to right, the columns in the EXCEL spreadsheet are

1. SMSA - Standard metropolitan statistical area
2. All cause mortality (age adjusted, deaths per 100,000)

Columns 3-6: Weather Variables (1960 data)

3. Mean annual precipitation (in inches)
4. Humidity (annual average percent)
5. Mean January Temperature (degrees F)
6. Mean July Temperature (degrees F)

Columns 7-14: Demographics (1960 data)

7. Percent of population over 65 years old
8. Mean number per household
9. Median number of years of education completed for persons age 25 or older
10. Percent of housing that is sound with all facilities
11. Population density (population per square mile of urbanized area)
12. Percent that is non-white
13. Percentage of employment in white collar professions
14. Percentage poor (household income below \$3000)

Columns 15-17: Relative Pollution Potentials

15. Hydrocarbons
16. Oxides of Nitrogen
17. Sulfur Dioxides

The question for you to examine is whether there is evidence that mortality is associated with one or more of the pollution variables, after accounting for the effect of the climate and

demographic variables. Analyze the data, and write a report of your findings, including any important limitations of the study.

Remarks on Data:

A. Two cities, Lancaster and York, are heavily populated by members of the Amish religion, who prefer to teach their children at home. The lower years of education for these two cities do not indicate a social climate similar to other cities with similar years of education.

B. “Relative pollution potential” is the product of the tons emitted per day per square kilometer and a factor correcting for SMSA dimension and exposure.

2. An agricultural experiment was designed to compare the effect of five row spacings (18, 24, 30, 36, and 42 inches) on the yield of two soybean varieties (OM=Ottawa Mandarin; B=Blackhawk). A field was partitioned into 12 long blocks, each of which was subdivided into five plots. Each variety of soybean was randomly assigned to six blocks. Within each block, the five plots were randomly assigned a spacing to be used between rows of the planted seeds. The number of plants per plot was kept fixed, as was the plot size. The data below are the yield (in bushels per acre) for each of the plots.

		Block per Variety					
Variety	Spacing	1	2	3	4	5	6
OM	18	33.6	37.1	34.1	34.6	35.4	36.1
	24	31.1	34.5	30.5	32.7	30.7	30.3
	30	33.0	29.5	29.2	30.7	30.7	27.9
	36	28.4	29.9	31.6	32.3	28.1	26.9
	42	31.4	28.3	28.9	28.6	18.5	33.4
B	18	28.0	25.5	28.3	29.4	27.3	28.3
	24	23.7	26.2	27.0	25.8	26.8	23.8
	30	23.5	26.8	24.9	23.3	21.4	22.0
	36	25.0	25.3	25.6	26.4	24.6	24.5
	42	25.7	23.2	23.4	25.6	24.5	22.9

A primary interest was determining the optimal spacing between rows (i.e. the spacing that maximized mean yield), and whether the differences in mean yield for different spacings depend on the variety. Provide a careful analysis of these data, making sure to address the questions of interest.