STATISTICS Ph.D. QUALIFYING EXAM: TAKE HOME
Due 2PM on Fri. January 16, 2004. 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 four 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 several formats (an EXCEL spreadsheet, a MINITAB worksheet, and a flat text (ascii) file in comma-separated-value form) and can be downloaded from the WWW site: http://www.math.unm.edu/~bedrick. Look for the link titled: Comprehensive Exam: Data for Problem 1.

The rise in abundance of algae in coastal waters is thought to be due to increases in nutrients such as nitrate and other forms of nitrogen. It is theorized that the excessive amounts of nitrate are due to human influences. Researchers gathered data to gauge the evidence that nitrates in the discharges of rivers around the world are associated with human population density. Human populations can affect nitrogen inputs to rivers through industrial and automobile emissions to the atmosphere (causing the nitrogen to enter the river through rainfall), through fertilizer runoff, through sewage discharge, and through watershed disturbance. The data set for this problem records the following variables (from left to right in the dataset) on 42 different rivers throughout the world.

1. RIVER: Name of river
2. COUNTRY: Location of river
3. DISCHARG: Estimated annual average discharge of the river into an ocean (in m^3/sec)
4. RUNOFF: Estimated annual average runoff from the watershed (in liters/(sec x km^2))
5. AREA: Area of watershed (in km^2)
6. DENSITY: Population density (in people/km^2)
7. NO3: Nitrate (NO_3) concentration (in μM/l)
8. EXPORT: Nitrate export, which is the product of runoff times nitrate concentration
9. DEP: Deposition, which is the product of precipitation times nitrate concentration
10. NPREC: Nitrate precipitation, the concentration of nitrate in wet precipitation at sites located near the watersheds (in μmol NO_3/(sec x km^2))
11. PREC: Precipitation (in cm/year)

The response variables are nitrate concentration and nitrate export. It is desired to determine whether these (separately) are associated with deposition or nitrate precipitation,

after accounting for discharge, runoff, precipitation, and area of watershed. It is also of interest to estimate the effect that human population density has, over and above its influence on the previous variables. Thus the analysis attempts to determine the extent to which the human effect is from pollutants discharged into the river directly as opposed to those discharged indirectly through atmospheric pollution.

Carefully analyze these data addressing the aims of the analysis above. You need to prepare a report with the following sections: A brief introduction to this problem, a methods section describing the statistical methods you used to analyze the data, a results section describing your findings in technical terms, and a conclusions section summarizing your findings in layman's terms.

Movie preference An experiment was designed to determine which of four movies, *Return of the King* (A), *The Station Agent* (B), *Elf* (C), or *Brother Bear* (D), had the greatest appeal to moviegoers. It is conjectured that both the time during the day and the day of the week may affect a movie's appeal. The movies were randomly assigned to the time slots and days in such a way that each movie would be shown once a day but in a different time slot from one day to the next. Each movie was shown to 50 adult (ages 18-25) volunteers who were asked individually upon exiting the theater if they would recommend the movie to a friend. The response is the number out of the 50 moviegoers who *would* recommend the movie to a friend. Note: At each showing a *different* group of 50 people watched the movie. The data are

	Monday	Tuesday	Wednesday	Thursday
10:30 AM	*C* 32	*D* 23	*B* 36	*A* 40
1:00 PM	*B* 33	*A* 36	*C* 31	*D* 22
4:30 PM	*D* 17	*C* 37	*A* 34	*B* 41
8:00 PM	*A* 35	*B* 37	*D* 18	*C* 31

(a) Analyze the data. Be complete. There may be more that one satisfactory approach to modeling these data.
(b) Discuss strengths and weaknesses of the experimental design. Can you suggest any modifications to the design that would improve the experiment?