Stat/For/Hort 572 **April 14, 2005**

Take-Home Midterm Examination

- This exam is due at the beginning of lecture **11:00 am on Thursday April 21**. For any change in due date (to be granted only under compelling circumstances), arrangements must be made in advance.
- Your solutions must be written on only **one side** of a page. You may use pen or pencil or you may use a word-processor (or use a combination); however, your solutions must be neat and legible. You should concentrate on being thorough in your thinking but concise in the write-up. It is important to provide your reasoning for the work you present. All computer output included must be integrated into the discussion; unnecessary computer output should be avoided. For each problem, you should begin with a brief summary of your findings. **The maximum number of pages for each problem is 7, including computer output.**
- **Hand in the two problems separately. For each problem please use a cover page with your name. Do not put your name on the exam otherwise.** (This cover page does not count toward the maximum number of pages described above.)
- You may use your notes and any books you wish. **However, the work is to be entirely your own. You may not collaborate with others or discuss this exam with others.** All questions concerning the exam must be addressed to the instructor — extra office hours during the exam are listed below. You may not consult with the TAs or computer assistants. There will be no discussion sections or TA office hours during the week of the exam (April 18 – 22).

Office hours for exam-related questions:	Thursday, April 14	1:00pm – 3:30pm
	Monday, April 18	2:30pm – 3:30pm
	Tuesday, April 19	2:30pm – 3:30pm
	Wednesday, April 20	2:00pm – 3:00pm

- You should strive to provide answers that are scientifically meaningful. You should attempt to analyze each problem as if it were your research problem. It is far better to provide an analysis that is simple, sensible, and easily interpreted than to demonstrate your knowledge of "fancy" statistical techniques. Use the simplest analysis that answers the question.
- The total score for the exam is 100. Each problem has equal weight.
- *Technical Note:* The following may or may not be useful to you. In R, to transform a variable in, say `y` you can type things like: `ty = log(y)` or `ty = log10(y)` or `ty = sqrt(y)` or

```
ty = y^2
```

. In SAS, suppose the data are in the variable `y`. If we use the notation `ty` to represent a transformed version of `y`, then we could say: `ty = log(y);` or `ty = log10(y);` or `ty = sqrt(y);` or `ty = y**2;` You place any of these SAS statements after the `input` and `infile` statements, but before any `proc` statements.

Problem 1: Toxic Effect of Radiation Therapy

Following removal of the tumor, women with breast cancer are often treated with radiation to shrink or destroy any remaining disease. The possible side effects of the radiation therapy, or toxicity, were evaluated in a study of 40 patients. Among the 40 patients, 20 received chemotherapy, which is a strong and often toxic medical treatment in addition to surgery and radiation, whereas 20 did not receive chemotherapy. Among the 20 patients who received chemotherapy, one half were smokers and the other half were non-smokers. Similarly among the 20 patients who did not receive chemotherapy, one half were smokers and the other half were non-smokers. Also recorded were the dose of radiation each patient received and the age of the patient.

The data appear on both the CALSHP and on the course web site in the file `therapy-05.dat` . The 8 columns correspond to the variables:

`chemo`: 0 = no chemotherapy and 1 = chemotherapy given to a patient;

`dose`: dose of radiation in Grays given to a patient;

`age`: age in years of a patient;

`smoke`: 0 = non-smoker; 1 = smoker;

`y`: toxicity, which is a side effect of radiation, rated on a scale of 0-100.

`w1, w2, w3`: alternative coding of the patients using 0-1 variables.

You are to analyze the data with an ultimate goal of understanding how the toxicity of radiation is related to the chemotherapy, dose rate, age and smoking history of a patient. You should address the following specific issues:

(a) Ignoring patient information (i.e., age and smoking history), what can you say about the relationship between toxicity and the treatment (i.e., chemotherapy and dose rate)?

(b) Ignoring the treatment (i.e., chemotherapy and dose rate), what can you say about the relationship between toxicity and a patient's age and smoking history?

(c) Based on all the data, how does toxicity relate to chemotherapy, dose rate, age, and smoking history of a patient?

Problem 2: Bark Distance of Prairie Dogs

A researcher conducted a study to assess the effect of human disturbances on the behavior of prairie dogs. A total of 12 colonies were selected, among which 4 colonies were located in urban environment, 4 in rural environment, and 4 in the boundary between the two environments (i.e., urban-rural boundary). At each colony, the researcher approached the prairie dogs and measured the distance between the prairie dog that sounded the first alarm bark and the researcher at the time of the first bark. Two starting locations were randomly selected along the edge of a colony and the experiment was repeated twice at each colony with the two starting locations. The data consist of two bark distances in m for each colony and are tabulated below.

	Colony			
Environment	1	2	3	4
Rural	44 52	42 66	53 42	55 57
Boundary	46 37	55 44	34 18	51 40
Urban	30 39	38 36	39 38	31 10

Analyze the data with the goal of understanding prairie dogs' response to human intrusion in the three different types of environment and address the following issues:

1. Are the bark distances different in the three types of environment?
2. Conduct all pairwise comparisons of the types of environment.
3. Is there a difference between the rural area and the urban/urban-rural boundary area? Is there a difference between the urban area and the rural/urban-rural boundary area?
4. Construct a 95% confidence interval for the within-colony variation.
5. Would it be useful to repeat the experiment more than twice in each colony?