First Mini-Project, 36-402, Spring 1998

Description of the problem

There is evidence that psychotheraputic interventions run by professional therapists facilitate adjustment to disease among people with cancer. However, the costs associated with formal psychotherapy groups preclude communities from adopting them. Instead, community support groups borrow some of the cognitive and behavioral principles of psychotherapy, often focusing on providing emotional support through peer discussion and informal support through education. In this study, peer discussion and education groups were examined to see if they too help cancer patients adjust to life with the disease.

The data from this study consists of records on 312 women with Stage I or Stage II breast cancer. The women were divided into groups of 8–12 women each, and these groups were randomly assigned to one of four possible interventions: education group, peer discussion group, combined education and peer discussion group, and control (no group participation). There were 7 groups for each intervention (for a total of 28 groups). The intervention lasted eight weeks.

Patients' quality of life was measured before the intervention, immediately after the intervention, and six months after the intervention. Among the quality of life measures were

- **PAF** A measure of the patient's positive affect (affect = emotional expression)
- **NAF** A measure of the patient's negative affect.

PAF and NAF are based on different questionnaires, and so even though you would expect them to be negatively correlated you wouldn't expect the correlation to be exactly -1.

An extract of the data is contained in the file `http://www.stat.cmu.edu/afs-data/402/mp1/healthed.dat`, suitable for reading into SPLUS with `read.table()`. The data looks like this:

```
group   interv  t1pa    t2pa    t1na    t2na
1       2       3.9     3.4     3.75    2.38
1       2       4       3.7     1       1.13
9       1       3.1     5       2.13    1
9       1       2.2     2.1     3.5     3.25
.
.
.
```

where the meanings of the variables are:

```
group  - which of the 28 groups was the patient assigned to

interv - which intervention was this patient's group assigned to:
         1 = peer discussion
         2 = education
         3 = combined peer and education
         4 = control
```

```
t1pa   - PAF measure, taken before the intervention

t2pa   - PAF measure, taken immediately after the intervention

t1na   - NAF measure, taken before intervention

t2na   - NAF measure, taken immediately after intervention
```

Note that there is some missing data, denoted by NA's in the data file.

Your mission

Your mission is to explore and analyze these data, and write up a report of your conclusions. Some issues to keep in mind include:

- **Coding the factors.** Part of the problem looks like a two-way ANOVA for the factors "peer discussion" and "education". But the `interv` variable isn't the right variable to think about for this kind of analysis; you will have to create new variable(s) that do what you want for this.
- **Groups.** Since "education" and "peer discussion" takes place in classes, the patients coudn't be "treated" individually, but rather had to be "treated in groups or classes of 8–12 patients each. Can the groups be ignored in the analysis? If not, how should they enter into the problem?
- **What are the study units?** Normally we would think of individual patients as the units of study (the individual observations). We can do that here too, if we can convince ourselves that the groups do not affect the outcomes. However, it seems plausible that individual responses are correlated with other responses from the same group. Does that mean the groups should be the study units, so perhaps we should use group averages as the "raw data"? If we do this, it feels like we've averaged away some potentially important information about individuals. Is there a compromise between these two?
- **What is the response variable? Or perhaps there is more than one...** You might take just one measure, such as `t2na`, as your outcome measure. The trouble is, this is a measure of both how much effect the treatment had, and how much negative affect the patient had anyway, even before the treatment; that is why `t1na` was also collected. Some possible ways to handle this include:
  - let the response be $y =$ `t2na - t1na`
  - let the response be $y =$ `t2na`, but include `t1na` as a (continuous) predictor in the model

  The second method above is called ANCOVA (Analysis of Covariance); an ANCOVA model is simply an ANOVA model that also includes one or more continuous predictors of some kind. The reason for putting the continuous predictor in the model is similar to why we include the blocks in the model in a randomized blocks experiment: we try to pull variation due to the continuous predictor out of the error sum of squares, to make F tests for factors in the model more powerful (because, presumably, we will have lowered the error sum of squares).

Which of these two works better in this problem? Why? And what do we do about the other variables, `t1pa` and `t2pa`?

The above are a few technical issues that will influence the kinds of analysis you do for your final report; there are other technical issues to think about too, like equality of variability of residuals throughout the range of data, normality assumptions, existence of outliers, and all of the other usual things that you think about with linear models and ANOVA.

Of course, when all is said and done, the main questions you are trying to answer are:

- Did any of the interventions (treatments) have an effect on PA or NA scores?
- If so, what would you recommend to new patients seeking community-based support (education, peer discussion, or both), based on the evidence of this study.

Your answers should be clearly-stated, sensible, and well justified by the data above and the methods you use.

You may find that the methods presented in class are adequate to answer these questions, and that's fine. Or you may feel you need to learn more; if you do you, may find it helpful to look at Chapters 7 and 10 of Venables and Ripley, or you may want to search the Web or consult one of the textbooks I have been using in class (Box, Hunter and Hunter, or Neter, Wasserman and Kutner), or other books in the library (E&S or Hunt).

On Writeups

A good generic outline for project reports is:

0. Executive summary (state background, problem, and conclusions in one page or less);
1. Description of problem, problem background, and data;
2. Summary of analyses performed (what, why, how, results);
3. Conclusions;
4. Technical Appendices;
5. Bibliography and Credits

Parts 1–3 should have as little technical material as you can possibly get away with: a few well-chosen graphs, the names of the techniques you applied and what parts of the data were used, and perhaps one or two parameter estimates or p-values is the most you should think about. Part 4 should contain the "story" of your technical analyses. However, it should be possible to understand what you've done from parts 1–3, without looking at part 4 at all.

If I were doing this, I would write part 4 first, as a sort of edited "diary" of my work as I learn about the problem (maybe not too different from the handouts I prepare for class). Then I would write parts 2 and 3, using the highlights of my "diary", and finally I would write parts 0 and 1, summarizing my work. I expect part 1 would be short, since you don't know anymore about the problem than I've told you.

You are welcome to discuss this problem with each other, but I would like a list of people you discussed this with (in and out of class), as well as a list of any outside sources you consulted (books, Web pages, etc.) in Part 5 of your report.

Timeline, Deadines

I would like this to be a "miniproject", not a full-blown project. So I would like to discuss with you your progress after one week, probably in class, and then I would like you to hand in your final reports one week later. So the deadlines are:

- *Tue Feb 17.* I will take questions in class on the data, the project, etc.
- *Thu Feb 19.* Bring a draft of Part 4 of your report to class for me to check, in class.
- *Thu Feb 26.* Bring your final reports to class.

I may be willing to negotiate these deadlines a little, as long as the whole class can agree on new ones.