

## ST 732, DATA ANALYSIS PROJECT, SPRING 2007

*Due Thursday, April 26, 2007*

**N.B.:** This is a “closed” take-home project. Thus, you **may not** collaborate with or discuss the assignment with *any* other person (whether in the class or not), **nor** may you consult with *anyone* except the instructor for help or clarification (not even your mother). You **may** use the class notes, books, or any other reading material you like.

Because you are taking a course in longitudinal data analysis, you have been approached by a group of medical researchers to serve as a statistical consultant. The researchers have conducted a clinical trial to investigate the effects of two oral treatments for *toenail dermatophyte onychomycosis*, a common infection of the toenail. Here is some background.

Onychomycosis, known popularly as toenail fungus, is a fairly common condition that not only can disfigure and sometimes destroy the nail but that also can lead to social and self-image issues for sufferers. It has been estimated that between 2% and 18% of people world-wide and up to 5% of people in the United States are afflicted by some form of onychomycosis. The likelihood of affliction is rare among children under 18 but increases with age, and it is estimated that up to 48% of people have had at least one episode by the age of 70. The big toe and little toe are the most likely to be afflicted. Tight-fitting shoes or hosiery, the sharing of common facilities such as showers and locker rooms, and toenail polish are all thought to be implicated in the development of onychomycosis.

Onychomycosis is distinguished by discoloration of the toenail, which itself becomes thick, overgrown, and crusted with debris. This condition is often accompanied by a foul-smelling odor caused by accumulation of debris under the nails. In various versions of the disease, the nail may crumble and gradually fall off, become so thick that wearing shoes causes extreme discomfort, or develop a soft, powdery surface.

Onychomycosis can be caused by several types of fungi known as dermatophytes, as well as by nondermatophytic yeasts or molds. The study conducted by the researchers involved sufferers of the particular type caused by dermatophytes.

The standard approach to treatment of toenail dermatophyte onychomycosis involves removal of as much of the afflicted part of the nail as possible, followed by application of a topical antifungal ointment or by a course of treatment with an oral antifungal medication, for which better outcomes are generally obtained than with ointments. The study conducted by the researchers was focused on comparison of two oral medications, terbinafine (given as 250 mg/day, denoted as treatment 1 below) and itraconazole (given as 200 mg/day, denoted as treatment 2 below).

The trial was conducted as follows. 200 sufferers of advanced toenail dermatophyte onychomycosis in the big toe were recruited, and each saw a physician, who removed the afflicted nail as described above. Each subject was then randomly assigned to treatment with either terbinafine (treatment 1) or itraconazole (treatment 2). Immediately prior to beginning treatment, the length of the unafflicted part of the toenail (which was hence not removed) was recorded (in millimeters). Then at 1 month, 2 months, 3 months, 6 months, and 12 months, each subject returned, and the length of the unafflicted part of the nail was measured again. Intuitively, as the nail grows out (hopefully without signs of onychomycosis, due to the effects of treatment), this length is likely to increase if the treatments are working, with the better treatment resulting in longer unafflicted nail lengths on average.

Also recorded on each subject was gender and an indicator of the frequency with which the subject visited a gym or health club (and hence might use shared locker rooms and/or showers).

The data are available in the file `toenail.dat` on the class web page. The data are presented in the form of one data record per observation; the columns of the data set are as follows:

- 1 Subject id
- 2 Health club frequency indicator (= 0 if once a week or less, = 1 if more than once a week)
- 3 Gender indicator (= 0 if female, = 1 if male)
- 4 Month
- 5 Unafflicted nail length (the response, mm)
- 6 Treatment indicator (= 1 if terbinafine, = 2 if itraconazole)

The researchers had several questions, which they stated to you as follows:

- Is there evidence to suggest that these two oral medications differ in their effects on toenail dermatophyte onychomycosis, as reflected by the lengths of the unafflicted part of the nail?
- Is there a difference in the pattern of change of lengths of the unafflicted part of the nail between subjects receiving terbinafine and itraconazole over a 12 month period? Does one treatment show results more quickly?
- What is the difference in lengths of the unafflicted part of the nail at the end of the study? Is there evidence to suggest that one of the treatments produces longer lengths on average than the other after 12 months?
- Recent literature reports have suggested that males tend to have more serious cases of onychomycosis than females. This would be reflected by an association between lengths of the unafflicted part of the nail and gender in this population at the beginning of the study (before treatment), with lengths being shorter on average for males than for females (meaning that the disease afflicts more of the nail in males). Is there evidence to suggest that this is the case?
- As noted above, the time spent in shared locker rooms and showers is thought to be associated with developing onychomycosis. A recent report has further hypothesized that the severity of the disease is also associated with this practice. Is there evidence in these data that subjects in this population who spend more than one day per week at a gym or health club have nail lengths different from those who spend less time at the beginning of the study (before treatment)? (One might expect that shorter lengths would be seen in those who spend more time if this were true.)
- Is there an association between the pattern of change of nail lengths and gender and/or health club frequency in subjects taking terbinafine? This might indicate that this drug brings about relief more swiftly in some kinds of subject versus others.
- Is there an association between the pattern of change of nail lengths and gender and/or health club frequency in subjects taking itraconazole? This might indicate that this drug brings about relief more swiftly in some kinds of subject versus others.

As the statistical consultant, you have been called in to carry out an appropriate analysis of the data from the study in order to address these questions. Your job is to conduct a thorough analysis and write a report for the researchers describing what you did, why you did it, and what inferences may be drawn regarding the questions of interest.

*YOUR REPORT SHOULD BE WRITTEN FOR THE RESEARCHERS (NOT FOR ME).* You should assume that, although the researchers know about some basic statistical models and

methods, such as linear regression, they know virtually *nothing* about longitudinal data models or methods. Thus, your report should “tell the story” of what you did, why you did it, and what your conclusions are at a level that the researchers can understand. Your grade will be based on how well you communicate and justify the statistical modeling and analysis principles you use and how well you explain and interpret the results in a way that researchers not familiar with longitudinal data models and methods can understand. In particular, a good report will explain clearly in a non-technical way:

- Why specialized statistical models and methods are required for analysis of longitudinal data (so why methods familiar to the researchers are not appropriate).
- The statistical model you have chosen, including its interpretation and how the researchers’ questions may be cast formally in terms of the model.
- Your rationale for choosing the model, any assumptions you have made, and why these assumptions are reasonable for the researchers’ situation.
- The method used to fit the model.
- Each step of your analysis in “layman’s” terms (*not* “technical” terms).
- The results of the analysis, described in terms of the subject-matter.

The researchers will not be interested in seeing gory matrix expressions and lots of equations and formulæ, because they *will not understand them*. Thus, you will need to communicate the above information to them mainly *in words*, with very few equations and symbols (being sure to define clearly any symbols you do use).

It is not a good idea to provide collaborators with results from several different analyses, because it can lead to confusion. A good report should report a single analysis based on a particular statistical model and methods and justify why the model and methods were chosen. Thus, your report should explain clearly to the researchers the considerations that led you to adopt the particular model and analysis whose results you provide and interpret. The researchers should not have to encounter terminology or concepts that they are not likely to know or that are not explained, nor should they have to search through SAS programs and output, which they are *sure* to not understand, to find results.

Your report should adhere to the following requirements:

- Your report should be *typed*.
- A good statistical data analysis report always provides background on the situation, gives a general statement of the problem (even if both you and your collaborators know what it is), summarizes the data (this is often most effective when done graphically), and states clearly the scientific objectives and why they are of interest. It also always summarizes what was done and gives clear a statement of the conclusions regarding questions of interest from a subject-matter perspective. Thus, your report should be organized into a sequence of sections that presents all of this in a logical way.
- Any code that produces results cited in your report and the associated output should be included as an appendix to the report. *However*, this is for my information only; your report *should not* refer to the appendix. (I.e., the report should not ask the researchers to go to pages of code or output; any results that the researchers need to see should be cited in the body of your report. Keep in mind that the researchers will know nothing about SAS procedures you might use nor how to interpret their output.)