Sta 106 - Project 1

The Senic Project

- **Background.** The data for this project came from a Study on the Efficacy of Nosocomial Infection Control (SENIC) Project conducted during the 1970s. One of the primary objectives of the project was to determine whether infection surveillance and control programs had reduced the rate of nosocomial infection in the U.S. The dataset consists of a random sample of 113 hospitals selected from the original 338 hospitals surveyed.


- **Data:** Available online at the course website.

- The project has four parts:
  
  **Part I: Reading.** Find the reference online and read the article by Haley *et al.* (pp. 472-485) that describes the study design. Write a summary of the article you read as an introduction to your project.

  **Part II:** Although the data was collected in an observational study, in the following analysis we treat it as if it were collected from an experiment (see discussion in Part IV below). The first problem of interest is whether the mean infection risk (variable 4) differs for the four geographic regions (variable 9). You may use $\alpha = 0.05$ for the analysis. You may assume that the following ANOVA model holds:

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \ldots, r, \ j = 1, \ldots, n_i,$$

where the $\epsilon_{ij}$’s are independent and distributed as $N(0, \sigma^2)$. More specifically, please give answers to the following questions:

a. What is $r$? and what are the $n_i$’s in this case?

b. What is the estimate of $\sigma^2$?

c. Use ANOVA method to address the problem of interest above.

d. What is your conclusion?

3. **Part III.** The second problem of interest is the effect of average age of patients (variable 3). You may classify the average age into four categories: Under 50.0, 50.0-54.9, 55.0-59.0, and 60.0 and over. Once again, assume that the ANOVA model above is appropriate. Repeat the parts a, b, c, d in Part II. Use $\alpha = 0.10$ for your analysis.

**Part IV: Discussion.** Discuss the possible impact of observational data on your results. Are there any (simple) tools for model diagnostics? Do so if you find a way to check the appropriateness of the ANOVA model.

- A report is due in two weeks (Feb. 8), in class.