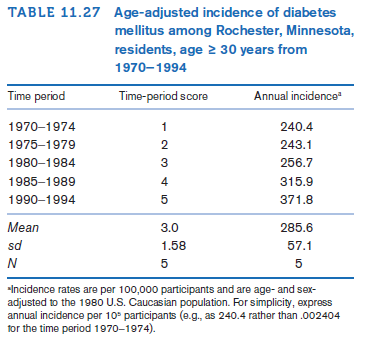
**PH1700 Homework 8**

**Instructions**.

You may work in groups, but you must submit your own write up of the homework. Create a Word or PDF document with text as if you were writing up a statistical analysis report. Embed the figures and tables in your document. You can copy tables by selecting the table, right-click, and copy as a picture – then paste into Word. Similarly, with graphs, click on the window with the graph, Ctrl-C (Command-C on Mac) to copy, then Ctrl-V (Command-V on Mac) to paste into Word. Important: it is unacceptable to just answer with a graph or a number. You have to write a sentence or two describing what you observe and your interpretation. Please include units (mm, inches, pounds, mm/L, etc.) when reporting any numbers. Be professional in your homework report. For questions involving hypothesis testing or confidence intervals, assume the 0.05 significance level, unless otherwise specified.

**For all homework problems, when performing any statistical test, please state the hypotheses, and report the test statistic, and report the p-value (or critical value, if requested) as part of performing and interpreting test.**

**Problems**

**Diabetes**

The Mayo Clinic is based in Rochester, Minnesota. Residents of Rochester and surrounding areas get almost all their medical care at the Mayo Clinic. In addition, migration in and out of Rochester is relatively low. This makes it feasible to track disease history over time and to assess whether the incidence of disease has changed over time. The data in Table 11.27 were presented concerning the incidence of diabetes over time among Rochester, Minnesota, residents ages 30 and over [19].

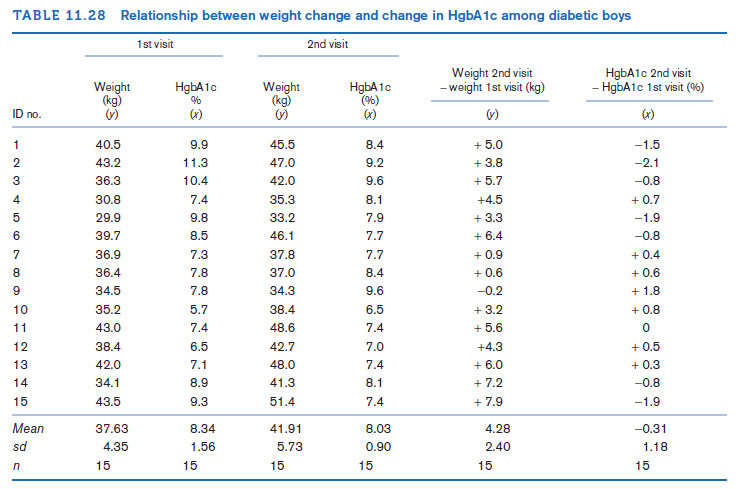
**11.89:** Clarification, fit a linear model means fit the model in STATA, and assess the assumptions, and report the equation of the line that was fitted.

**11.90:** Perform and interpret the test requested assuming all assumptions hold. If in reality, some don’t (from problem 11.89) point out that fact and name the assumptions violated.

**Diabetes**

A group of 10-year-old boys were first ascertained in a camp for diabetic boys. They had their weight measured

at baseline and again when they returned to camp 1 year later. Each time, a serum sample was obtained from which a determination of hemoglobin A1c (HgbA1c) was made. HgbA1c (also called *glycosylated hemoglobin*) is routinely used to monitor compliance with taking insulin injections. Usually, the poorer the compliance, the higher the HgbA1c level will be. The hypothesis is that the level HgbA1c is related to weight. The data in Table 11.28 were obtained.



**11.95** Compute a rank correlation between change in weight and change in HgbA1c, each over 1 year. Use this measure to directly test the hypothesis that change in weight over 1 year is related to change in HgbA1c. Report a two-tailed *p-*value, and provide a 95% confidence interval for the underlying rank correlation. (bonus points for CI)

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

**Required additional problem**

Data analysis problems:

1. Use the FEV data from Rosner to answer the research question: Is there a relationship between smoking status and FEV after adjusting for Height, Sex, and age.
   1. Categorize age by every 5 years [1-5, 6-10,….].
   2. Fit the preliminary model and assess assumptions. Discuss any violated assumptions if present, or determine they all hold.
   3. Using the gladder command, which transformation of FEV would you choose? Explain why in 2 or 3 sentences.
   4. For uniformity of answers on this homework, fit a second model using log(Fev), and assess the assumptions.
   5. Using the model from part d, answer the following questions
      1. Is the model significant? Report the test statistic and p-value used for answering the question.
      2. How much variation of log(Fev) is explained by the model?
      3. Is the coefficient for Smoking significant? (Yes or no) What is the 95% confidence interval for the coefficient, and how is it interpreted?
      4. Is the coefficient for Height significant? How would we interpret it?
      5. Are the categories of Age significant? Report the hypothesis, test statistic and p-value. How would we interpret the coefficient for the oldest category?
   6. Using the model, predict the log(FEV) for an average 17 year old male who is 60 inches tall and not a current smoker. What would the predicted FEV be?