This is a take home final assignment (in place of an exam). You can use any inanimate objects (e.g., the book and your notes) but do not discuss it with anyone except the professor. It is due on Friday, December 12th by 5 p.m. in my office, Huntsman Hall 464. If I am not there, put it in my mailbox in the front of the department.

1. (10 points) Stanley and Walton (1961, *J. Ment. Sci.*) ran a controlled clinical trial to investigate the effect of the drug stelazine on chronic schizophrenics. The trials were conducted on chronic schizophrenics in two closed wards. In each of the wards, the patients were divided into two groups matched for age and length of time in the hospital. One member of each pair was randomly assigned to receive stelazine and the other was randomly assigned to receive a placebo. Only the hospital pharmacist knew which member of each pair received the actual drug. The data set stelazine.JMP gives the staff psychiatrists’ behavioral rating scores for the patients at the beginning of the trial and after 3 months. The variable stelazine patient improvement is equal to stelazine patient after minus stelazine patient before (and placebo patient improvement is defined similarly). Assume that the ideal normal model holds for these data.

(a) Conduct a test of whether stelazine is associated with improvement in the patients’ scores for each ward. Give the p-values (one for each ward) and state your conclusion. [Note: To do the test for Ward A, exclude the observations from Ward B and likewise to do the test for Ward B, exclude the observations from Ward A.]

(b) Test whether there is any difference in improvement between the wards. Give the appropriate p-value(s) and state your conclusion.

(c) Suppose that some of the staff psychiatrists are friendly with the hospital pharmacist. What potential problem does this pose for making inferences about the causal effect of stelazine on behavior? How could this problem have been avoided?

2. (10 points) A credit card lender distributed cards to university students using five methods of distribution and collected data on the income (interest and fees) from each card during the year. The method used to distribute the card to a given student was randomly assigned. The data is in credit.JMP.

(a) Check whether the assumptions of the ideal normal model for the one-way layout are satisfied. Briefly report your conclusions.

For the rest of the problem, assume that the assumptions of the ideal normal model for the one-way layout are satisfied

(b) Test the null hypothesis that the mean incomes of the five methods of distribution are all the same and give the appropriate p-value.

(c) Write a brief report for management (which wants to maximize income) about what you consider are appropriate conclusions as to which distribution methods ought to be used or avoided.

3. (15 points) It has been theorized that developing countries cut down their forests to pay off foreign debt. Two researchers examined this belief using data from 11 Latin American nations (Data from R.T. Gullison and E.C. Losos, “The Role of Foreign Debt in Deforestation in Latin America,” *Conservation Biology* 7(1) (1992): 140-7). The data on debt, deforestation and population are in debtdeforest.JMP.
(a) Consider the simple linear regression in which the response is deforestation and the explanatory variable is debt. If forced to choose between the models

\[ \mu\{\text{deforestation}|\text{debt}\} = \beta_0 + \beta_1 \text{debt} \quad (1) \]

and

\[ \mu\{\log \text{deforestation}|\log(\text{debt})\} = \beta_0 + \beta_1 \log \text{debt}, \quad (2) \]

which would you choose and why?

(b) Using model (2), test whether an increase in debt is associated with a change in mean deforestation. Give the p-value and state your conclusion.

(c) Consider the multiple regression model

\[ \mu\{\log(\text{deforestation})|\log(\text{debt}), \log(\text{population})\} = \beta_0 + \beta_1 \log(\text{debt}) + \beta_2 \log(\text{population}) \quad (3) \]

Test whether an increase in debt is associated with a change in mean deforestation when population is held constant. Give the p-value and state your conclusion.

(d) Which hypothesis test (b) or (c) do you think is more relevant to answering the question of do developing countries cut down their forests to pay off foreign debt? Explain briefly. (Note: neither test can be used to decisively show causation because this is an observational study but answer which test provides more relevant evidence for causation).

4. (15 points) Natal dispersion distances are the distances that juvenile animals travel from their birthplace to their adult home. An understanding of dispersal distances helps to identify which species in a community are vulnerable to the loss of connectedness of habitat. To further the understanding of determinants of dispersal distances, researchers gathered data on body weight, diet type, and maximum natal dispersal distance for various animals. The data is in nataldispersal.JMP.

(a) Consider the following two parallel regression lines models, one which uses the original variables and the other which uses the log of the variables:

\[ \mu\{\text{nataldispersaldistance}|\text{bodyweight}\} = \beta_0 + \beta_1 I_{\text{Carnivore}} + \beta_2 I_{\text{Omnivore}} + \beta_3 \text{bodyweight} \quad (4) \]

\[ \mu\{\log(\text{nataldispersaldistance})|\log(\text{bodyweight}), \text{DIETTYPE}\} = \beta_0 + \beta_1 I_{\text{Carnivore}} + \beta_2 I_{\text{Omnivore}} + \beta_3 \log(\text{bodyweight}) \quad (5) \]

Draw two coded residual plots to show that model (5) is better and explain why they show this.

(b) Based on the model (5), test whether an increase in body weight is associated with an increase in natal dispersal distance for omnivores. Give the p-value and state your conclusion.

(c) Based on the model (5), test whether for a fixed body weight, carnivores have a higher mean natal dispersal distance than herbivores.

(d) Based on the model (5), find a 95% confidence interval for the ratio of median dispersal distance for omnivores to the median dispersal distance for herbivores of the same body size.

(e) Someone objects to the use of model (5), saying that the lines relating the mean of log (natal dispersal distance) to log (body weight) will not be parallel for carnivores, omnivores and herbivores. Test the null hypothesis that the three lines are parallel. Give the p-value and state your conclusion.