**Statistics/Forestry/Horticulture 572** **Discussion 11**


| | |
|---|---|
| **Course website** | http://www.stat.wisc.edu/courses/st572-larget/ |
| **TA** | Xu He(Mario) |
| **Email** | hexu@stat.wisc.edu |
| **Website of TA** | www.stat.wisc.edu/∼hexu |
| **Office** | 1275F, MSC |
| **Office hour** | Th 11:00-12:00, 4:00-6:00 |
| **Phone** | (608)334-9792 |


# 1 Problem 1

The data set "bp.obese" in R package "ISwR" is the random sample drawn from a small California town. The aim of this study is to see whether obesity of people affects their systolic blood pressure.

Format of this data set:

sex: a numeric vector code. 0: male,1: female.
obese: a numeric vector. Ratio of actual weight to ideal weight from New York Metropolitan Life Tables.
bp: a numeric vector. Systolic blood pressure (mm Hg).

(a) What do we need to do on this data before analysis?

(b) Regard bp as response. Use xyplot to check whether there is interaction between sex and obese.

(c) Regress bp against obese for each gender group. Simulate these two models and get the coefficient estimates for each model. Compute the difference of the intercepts and slopes from these models respectively. Then construct a linear model with "bp" as response and "sex" and "obese" as predictors, including the interaction. Comment on what you saw.

(d) Utilizing the simulation result, construct a 90% confidence interval of both the intercept and slope differences.

# 2 Problem 2

In a research of the disease malaria, a random sample of 100 children from a village in Ghana, aged 3-15 years, were followed for a period of 8 month. At the beginning of the study, values of a particular antibody were assessed. Use this data set "malaria" in R package "ISwR" and answer the following questions.

Format of this data set:

subject: subject code.
age: age in years.
ab: antibody level.
mal: a numeric vector code: Malaria, 0/1 is no/yes, respectively.

(a) What do we need to do on this data before analysis?

(b) Construct a model with age and antibody level as explanatory variables. Which model would you like to use and why?

(c) What's the risk of malaria for a 4-year child with an average antibody level, and a level 1000?

# 3 Problem 3

In the early 1990s, scientists became concerned about rapid declines in populations of amphibians worldwide. After some analysis, researchers consider global atmospheric causes, such as thinning of the ozone layer and the resulting increased exposure to damaging ultraviolet (UV-B) light to be the answer.

10 amphibians in northwestern United States were studied under a exposure to a UV-B filtering treatment having three levels: (1) a UV-B blocking filter; (2) a UV-B transmitting filter; (3) no filter-the control group. Use the subset of this data, "egg.txt", and answer following questions.

Note: In this subset, 3 species were observed in the Oregon Cascade Mountains. Each species occurs naturally in three sites chosen within this area, and each of the four sites was subject to a different UV-B treatment. In other words, we could regard the species as blocking factor, and the combination of species and UV-B treatment as a plot factor.

Format of this data set:

spe: numeric code for amphibian species studied. 1: Hyla regillar; 2: Rana cascadae; 3: Bufo boreas.
uvb: numeric code for UV-B treatment. 1: No filter; 2: UV-B transmitting; 3: UV-B blocking.
fail: percentage of eggs hatching failures.

(a) What do we need to do before analysis? Create a "plot" factor in the data set.

(b) Construct a multilevel model, and utilize the simulation process to find the 90% prediction interval for difference between Bufo boreas amphibians.

(c) Construct a 90% prediction interval for difference between UV-B blocking treatment and the control group for amphibian species Rana cascadae.