UNL Statistics PhD Qualifying Exam - January 2024

Print Your Qualifying Exam ID: ____________________

Day 2

1. In epidemiology one studies the spread of disease with mathematical and statistical models. One known model **SIR** stands for Susceptible, Infected, and Removed. In this model, individuals are assumed to be one of three types:

   - susceptible if they have not yet caught the disease,
   - infected if they currently have the disease,
   - removed if they have had the disease and have since recovered (and are now immune).

   At each time t, the infected can infect susceptibles or can recover, at which point the infected is removed. We assume none can become infected and then recover in the same time t. Let $S(t), I(t)$, and $R(t)$ denote the number of susceptible, infected and removed individuals at time t. Each infected has probability α of infecting each susceptible (we assume that each infected has equal contact with all susceptibles). At the end of each time step, after having had a chance to infect people, each infected has probability $\boldsymbol{\beta}$ of being removed. Initial conditions are

   $$S(0) = N, \quad I(0) = 1, \quad R(0) = 0.$$

   The total population size is $N + 1$ and remains fixed (assuming no death), so $S(t)+I(t)+R(t) = N+1$ for all t. Each time step t the probability that a susceptible remains uninfected is $(1-\alpha)^{I(t)}$. That is, each infected must fail to pass on the infection to the susceptible. Thus,

   $$S(t+1) \sim \text{Binomial}\left(S(t), (1-\alpha)^{I(t)}\right).$$

   As each infected has the probability $\boldsymbol{\beta}$ of being removed, we have

   $$R(t+1) \sim R(t) + \text{Binomial}(I(t), \beta).$$

   Then $I(t+1)$ is computed via

   $$I(t+1) = N + 1 - R(t+1) - S(t+1).$$

   Researchers would like to understand how the SIR model behaves for different values of α and β. For simplicity, you don't have to consider the exit condition or other complex settings.

a) Write a simulation code of the SIR process and use it to complete the following tasks. The SIR process must be coded as a function, for example,

```
SIRsim = function(alpha, beta, N, T) {
    ............
}
```

b) Simulate twenty realizations (or repeats) of the SIR epidemic using $N = 1000$, $\alpha = 0.0005$, and $\beta = 0.3$. Plot the twenty realizations in a figure of three subplots for $S(t), I(t), and R(t)$, respectively. The plots should be well labeled. Also, describe what you observe from the simulation result.

c) Next, fix $N = 1000$, $\alpha = 0.0005$, and we want to learn how β affects the SIR model. Simulate the SIR for $\beta = 0.1, 0.2, 0.3, 0.4$ (one realization for each β). Plot the simulation results $(S(t), I(t), R(t))$ for each β. Describe what you observe from the simulation study of different β values. The plots should be well labeled.

d) What happens to the SIR epidemic when $N\alpha > \beta$ and $N\alpha \leq \beta$? Use a simulation study to answer this question.

2. The global obesity epidemic is giving researchers a challenging task. To understand how different hormones can influence food intake, food scientists proposed a research project where they wanted to evaluate the weight gain of mice after hormone treatment. They randomly assigned one of two hormone treatments (Estrogen, Testosterone) and no hormone treatment to mice at the same time. Before the study, they measured the mice's weight, and one month after the hormone treatment they measured the weight again. All of the mice were given the same diet, and they kept the mice under the same environmental conditions. They collected the data on the weight before the hormone treatment and after the hormone treatment in grams. The data set can be found in data.csv.

   Let's assume you are a statistical consultant for the researchers, and you are responsible for the data analysis and interpretation.

   (a) Explain in detail what statistical model you would suggest based on the data if the researchers are interested in evaluating whether it makes a difference to apply a hormone in terms of the weight gain and they also want to compare the effect of the two hormones on the weight gain. In this part, you need to specify the design, model, and model assumptions.

   (b) Implement your model and conduct a data analysis. Make sure you include some exploratory data analysis, and model assumption checking before you fit your model. You need to summarize your findings in a report form. Don't forget to make some recommendations.

   (c) As a statistical consultant you realize that their experimental setup is not optimal. How would you improve their design? What questions would you ask from the researchers?

   (d) The researchers are also interested in a new response variable indicating whether the weight increased or not. Specify a model for this scenario. Include assumptions.

   (e) Analyze the data based on your model in part 4 and the same questions that are specified in part 1. Summarize your findings in a report form in a similar fashion as in part 2.

   (f) Based on your result, did you have enough power for your comparisons? Justify your answer. If not please justify how many experimental units you would need to reach 80% power. You need to complete this exercise for both response variables.