Biostatistics 140.621

Problem Set 2: Probability Models and their Applications in Public Health (with R)

Instructions: This problem set consists of one problem. Please submit your work as a write-up of responses to the following steps and questions. Please include your name and lecture section on the upper right-hand corner of your first page.

Due Date: BEFORE 11:59 pm ET in CoursePlus Drop Box on **Tuesday, September 16, 2025**

Problem 1. Vitamin A Supplementation to Prevent Children's Mortality in Nepal

Learning Objectives:

Students who successfully master this section will be able to:

- Use the concepts of probability to describe the effect of Vitamin A on mortality;
- Define and apply the term "effect modification" or equivalently "interaction" in a randomized trial;
- Use the binomial distribution and the Poisson approximation to the binomial to calculate probabilities of certain events.

Scientific Background:

Xerophthalmia, due to vitamin A deficiency, was long known to be a leading cause of childhood blindness in the developing world. However, in the late 1970s, past-Dean Alfred Sommer (then an assistant professor of ophthalmology) and his Indonesian colleagues observed that children with mild, non-blinding xerophthalmia (those having night blindness or Bitot's spots, which are dry patches that form on the conjunctiva) were at a 4-fold higher risk of dying than children with normal eyes. This finding, coupled with knowledge that adequate vitamin A status can improve resistance to infection, suggested that vitamin A deficiency could also be a cause of preschool child mortality. Hopkins investigators subsequently showed in large trials in Indonesia (with over 28,000 children) and Nepal (with over 25,000 children) that, in fact, dosing young children every 4-6 months with high-potency vitamin A could reduce child mortality by approximately 30%, findings that have since been replicated by other investigators in Asia and Africa. In this project, you will work with some of the original data from this community trial carried out in Nepal.

Original References (optional):

Sommer A, Tarwotjo I, Djunaedi E, West KP Jr, Loeden AA, Tilden R, Mele L and the Aceh Study Group. Impact of vitamin A supplementation on childhood mortality: A randomized controlled community trial. *Lancet* 1986;1:1169.

West KP Jr, Pokhrel RP, Katz J, LeClerq SC, Khatry SK, Shrestha SR, Pradhan EK, Tielsch JM, Pandey MR, Sommer A. Efficacy of vitamin A in reducing preschool child mortality in Nepal. *Lancet* 1991;338:67.

Methods:

i. The Nepal data set is located in the `csv` data file named `nepal621.csv`. Refer to your Class Dataset Code Book for the file format. Load the tidyverse with: `library(tidyverse)` Open the data set and name it `nepal621`. Construct a 2x2 contingency table of treatment (placebo or Vitamin A) against status (alive or dead) at sixteen months of follow-up. Calculate the rate of child mortality in Nepal for children receiving placebo; Vitamin A. Summarize the difference in mortality in a sentence as if for a journal.

```
CT = table(nepal621$trt, nepal621$status)
addmargins(CT)
prop.table(CT, margin=1)
```

ii. For a randomly chosen child from the study population, calculate the following probabilities from the 2x2 contingency table you constructed above.

Marginal Probabilities

Pr(VitA) Pr(Died)

Joint Probabilities

Pr(Died and VitA) Pr(Died and Placebo)

Conditional Probabilities

$$\Pr(\text{Died} \mid \text{VitA}) = \frac{\Pr(\text{Died and VitA})}{\Pr(\text{VitA})}$$

$\Pr(\text{Died} \mid \text{Placebo})$

By hand, use Bayes' Theorem and the 2x2 contingency table to calculate the probability that a child that died received Vitamin A. Use the observed rates for each term below to see how the calculations work.

$$\Pr(\text{VitA} \mid \text{Died}) = \frac{\Pr(\text{VitA and Died})}{\Pr(\text{Died})} = \frac{\Pr(\text{Died} \mid \text{VitA}) \cdot \Pr(\text{VitA})}{\Pr(\text{Died})}$$

$$= \frac{\Pr(\text{Died} \mid \text{VitA}) \Pr(\text{VitA})}{\Pr(\text{Died} \mid \text{VitA}) \cdot \Pr(\text{VitA}) + \Pr(\text{Died} \mid \text{Placebo}) \cdot \Pr(\text{Placebo})}$$

iii. For each treatment group, construct the 2x2 contingency table of sex (girls or boys) versus vital status (alive or dead). From these tables, calculate the overall probability of dying for boys and girls (sex defined at birth) separately by treatment group. Describe in a sentence as if for a journal the relationship between mortality and treatment. Does the effect of treatment appear to vary by sex? If so, we say: "sex is an effect modifier," or "sex modifies the effect of vitamin A on mortality" or there is an "interaction" of sex and treatment in causing mortality. Write another sentence or two describing differences in the treatment effect between boys and girls. Be quantitative and use the term "effect modification."

```
nepal.plac = filter(nepal621, trt=="Placebo")
nepal.vit = filter(nepal621, trt=="Vit A")

CT = table(nepal.plac$sex, nepal.plac$status)
addmargins(CT)
prop.table(CT, margin=1)

CT = table(nepal.vit$sex, nepal.vit$status)
addmargins(CT)
prop.table(CT, margin=1)
```

iv. Summarize in a table and/or figure the evidence (data) relevant to the null hypotheses that:
(1) Vitamin A supplementation has no effect on mortality in Nepali pre-school children; and
(2) The treatment effect is the same for both boys and girls (i.e., "not modified by sex").

v. Consider a family with 3 boys and 2 girls who received placebo. Suppose that each child's survival is independent of all the other children in the family. Calculate the probability that 0, 1, 2 or 3 boys die during the study follow-up.

Use the applet: http://homepage.stat.uiowa.edu/~mbognar/applets/bin.html

vi. Use the Poisson approximation to the binomial distribution to recalculate probabilities of 0, 1, 2 or 3 boys dying in problem v.

Use the applet: http://homepage.stat.uiowa.edu/~mbognar/applets/pois.html

vii. For each treatment group, construct the 2x2 contingency table of age (<1, 1-2, 3-4) versus vital status (alive or dead). Consider a family with triplets who are 18 months old, on treatment: J, K, and L.

a) What is the probability that J and K live and L dies?

b) Does this differ from the probability that one of the three children dies? Why or why not?