Categorical Data Analysis, Prelim 2

STSCI/BTRY 4110 ---- November 2, 2021

Cover Sheet

Name _____

Net ID _____

Academic integrity is expected of all students of Cornell University at all times, whether in the presence or absence of members of the faculty.

Understanding this, I declare I shall not give, use, or receive unauthorized aid in this examination.

Signatures of Students _____

Notes:
> This cover sheet should be the first page of your completed report.

> You should submit an electronic copy of your work via Canvas. The electronic copy may be checked for plagiarism.

> Your submission should be two files: 1) Your written report, 2) R code for the analysis. You do NOT need to include all the R code that you ran, just the R code that generated the analyses described in your report.

> In addition, you need to fill out and submit a Peer Group Evaluation Form. This is required for everyone. This is due 24 hours after the project has been submitted. Your comments on teammates will be kept confidential. Only Dr. Smith will see the completed forms.

> The report should be the work of you and your partners, written in your own words. Do not discuss the exam with other classmates or students, professors or other people. You may ask me questions. It should be no longer than 12 pages long.

> The deadline for submission of your report and R code is Friday November 12 at 10 pm.

Analysis of Malaria Incidence

Malaria is mosquito-borne disease caused by a parasite. It can be a debilitating disease if left untreated. It is a particularly large threat in sub-Saharan Africa.

The data in this study were collected from 749 residents of three regions of an African country. They are in the file "malaria.csv". The objective of this assignment is to analyze and determine which factors might be predictive of malaria disease onset, and how those factors increase or decrease the risk of malaria.

The following information was collected from each study participant:

Nid	Unique ID number for participant
malaria	1 = yes (subject got malaria), 0 = no (did not get malaria)
stress	measure of stress hormone levels (in nano-molar units) higher levels indicate higher levels of stress
insecticide	average concentration of insecticide from 5 indoor home surfaces
source	identifier for the source of bed nets ("free" or "paid")
behavior	evaluation scale of malaria-prevention behaviors in the household (1 to 5)
	1=few malaria-prevention behaviors
	5= frequent and correct malaria-prevention behaviors
nettype	two possible types: A or B, sprayed with different insecticides
district	three possible regional districts: "1North", "2East" or "3South"
health	indicator of relative health, on a scale of 1 to 35
work	work status of the head of household
	0 = mostly unemployed over the past two months
	1 = mostly employed over the past two months
	2 = not working (student, retired, too old or disabled)

The objective of this study is to determine what factors best describe the subject's susceptibility and prognosis with respect to malaria onset. Write a report on the factors influencing the possible onset of malaria. Health professionals in the region want to know which factors or combination of factors are most predictive of patients with the highest risk of getting malaria. They are also interested in qualities or factors that appear to be protective against malaria.

Participants in the study all received bed nets – a mesh covering to place over their beds, as a physical barrier to mosquitoes when sleeping. Some participants had to pay for their bednets, while others received them for free. Each bed net was infused with one of two types of insecticide – Type A or Type B. Each individual's home was also sprayed with insecticide. Each individual's stress hormones were measured via a saliva sample.

Your report should include the fitting of a logistic regression model and a description of the steps you took to fit the model. (Note that you should also include other types of categorical data analysis tools such as contingency analysis for preliminary analyses before fitting your model.) Your write-up should include graphical displays, as appropriate.

Use the *Take Home Project Guidelines* posted on Canvas as a guide to my expectations. The Guidelines provide a detailed list of all items that should be included in your written report.

If there are situations during the analysis when you would want to discuss a next step with a doctor or other expert for advice, mention this in your report. Make any such decisions, as best you can.

You do not have to include a description of every single thing you did or every step that you took; you should outline the steps that you took, with relevant intermediate results. Your report should have more detail than you would find in a typical published paper, but less than a step-by-step description.

If you have questions, you may email me directly at ms429@cornell.edu. If the questions are relevant to everyone, I will post them on Ed Discussions with answers. Likewise, if I have clarifications or corrections, I will post them.

Good luck and enjoy!
Melissa Smith
(ms429)