**Epi 550 – Homework 1**

Due in Canvas at **8:30am ET** on 9/11/2024

1. **Practice Exercises** – Logistic Regression, Chapter 1, pg. 32 – 34 – Problems 2, 9, and 10

2.

9.

10.

1. **Practice Exercises** – Logistic Regression, Chapter 2, pg. 67 – 68 – Problems 1-4, 7, 9-14

1.

2.

3.

4.

7.

9.

10.

11.

12.

13.

14.

1. **Practice Exercises** – Logistic Regression, Chapter 5, pg. 158 – 159 – Problems 8-13

8.

9.

10.

11.

12.

13.

1. **Logistic Regression Practice**

On Canvas, you will find a SAS dataset entitled DHS\_water. To conduct these analyses in R, begin with the code below. You will also need to load the car package.

DHS\_water <- haven::read\_sas(“DHS\_water.sas7bdat")

These data come from the [Demographic and Health Surveys (DHS) Program](https://dhsprogram.com/) – a collection of more than 300 surveys from over 90 countries that are used to collect information on fertility, family planning, maternal and child health, gender, HIV/AIDS, malaria, and nutrition. This particular dataset is a subset of a DHS model dataset which has been constructed for practice, and does not represent any country’s actual data.

Access to clean water - a human right - is often used to measure progress towards eliminating poverty, morbidity, and mortality. Clean water and breastfeeding are both important for infant health globally. It is possible that the distance that someone must travel to access water may impact their ability to breastfeed. For this assignment, you are interested in learning whether there is a relationship between the distance that a person must travel to get to their water source and breastfeeding. The type of water source[[1]](#footnote-1) that someone has access to may be associated with both the distance traveled and breastfeeding – thus, you will consider this as a potential confounder. You have access to the following data:

|  |  |
| --- | --- |
| **Variable Name** | **Coding** |
| Breastfeeding | 1 – Ever or currently breastfeeding  0 – Never breastfed |
| WaterDistance | 1 – Water source is on-site  0 – Water source is off-site |
| WaterSource | 1 – Improved water source  0 – Unimproved water source |

1. What is the exposure of interest, and how are the index and reference groups defined?
2. What is the outcome of interest, and how are the index and reference groups defined?
3. Write the logit form of the model that allows you to address the research question described above using the population parameters (i.e., do not fill in the estimated regression coefficients).
4. Run this model in SAS or R to find the estimated odds ratio for the association between distance to the water source and breastfeeding, adjusted for the type of water source.
5. Use both Wald and likelihood ratio tests to determine whether there is a statistically significant association between the distance to the water source and breastfeeding, controlling for the type of water source.
   * 1. What is the null hypothesis for these tests?
     2. What are the test statistics and associated p-values?
6. Write a 1-2 sentence summary of the study findings.
7. Fit the same model as described above, but recode WaterDistance as follows: -1: off-site, 1: on-site. Estimate the odds ratio for the between the distance to the water source and breastfeeding, controlling for the type of water source using the new coding. Compare your answer to what you found in Question 3, and comment on whether your responses differ and why.

1. If you’d like to know more about improved and unimproved water sources, you can learn more [here](https://www.cdc.gov/healthywater/global/assessing.html) [↑](#footnote-ref-1)