Lab 1 Instructions

BSTA 513/613

Nicky Wakim

> **⏶ Caution**
>
> This project includes analysis on food insecurity. Please let Nicky know if this is a triggering topic for you. I can find a different dataset for you to work on!

> **⏶ Caution**
>
> Nicky needs to edit

Directions

You can download the `.qmd` file for this lab here.

The above link will take you to your **editing** file. **Please do not remove anything from this editing file!!** You will only add your code and work to this file.

Lab activities

> **ℹ Note**
>
> I have left it up to you to load the needed packages for this lab.

Reading and listening activities

I will not check that you have read or listened to any of these, but it is a good starting point for understanding the context of our data: food insecurity in the United States. I haven't fully read them all yet, but I will be reading and sharing throughout the quarter.

Here are some articles:

- NPR: Millions of American families struggle to get food on the table, report finds
  - With option to listen to article

- Identification of factors related to food insecurity and the implications for social determinants of health screenings

- NIMHD's page on Food Accessibility, Insecurity and Health Outcomes

- Food Insecurity among American Indians and Alaska Natives: A National Profile using the Current Population Survey–Food Security Supplement

- A Framework for Evaluating Social Determinants of Health Screening and Referrals for Assistance

Familiarize yourself with the Well-Being and Basic Needs Survey

Please read the Urban Institute's page on the Well-Being and Basic Needs Survey (WBNS) and more of their published information of the survey (at least the first 4 pages). You can also read about the overarching From Safety Net to Solid Ground Initiative that started the survey. Answer the following questions:

- What is the motivation for this study?
- How could an analysis that looks at associations with food insecurity help facilitate change in policy?
- Why is it important to study food insecurity?

> **❗ Task**
>
> Answer the following questions using information on WBNS:
>
> - What is the motivation for this study?
> - How could an analysis that looks at associations with food insecurity help facilitate change in policy?
> - Why is it important to study food insecurity?

File organization

Before downloading the data, go back to Lesson 2 and follow the file setup for our project. This includes making an `.Rproj` file within the main folder. Make sure you are working with the project by using the `here()` function to display your working directory.

> **❗ Task**
>
> Display your working directory using the `here` package and `here()` function.

Access and download the data

1. Go to the Health and Medical Care Archive page for the Well-Being and Basic Needs Survey

2. Go to the Data & Documentation tab

3. Download the R version of the Public use data

4. Read and agree to the Terms of Use. After this, you will be redirected to a new page.
5. Log into ICPSR by clicking "Access through your institution". You should be taken to a new page where you need to select "Oregon Health & Science University"

6. Login using standard OHSU login. Then the download should begin!
7. Make sure to move this into your project folder under the Data folder.
8. Take a look at the folders/files you just downloaded. Make sure to locate and understand the difference between the Codebook, Questionnaire, and User Guide. (Note that the website also contains the codebook if you go to the variable tab. I think the online one has an easier user interface than the pdf.)

> ❗ Task
>
> No task to report back on. Just make sure you have the data!

Decide on list of variables to focus on

From the codebook, I want you to explore the variables and create a list of 10 predictors that you would like to focus on. Our outcome is `FOOD_INSEC` so we cannot use this as a predictor. Feel free to take a look at the Urban Institute's list of publications to get ideas of variables and relationships.

There are a few requirements for your predictors:

- 1 variable must be a numeric (i.e. `PPAGE`)

- 1 variable must be binary
- 1 variable must be multi-level categorical (categorical with more than 2 groups)
- You must choose at least 10 predictors (does not include the outcome)

There is a good online version of the codebook with information about the variables. I have linked you to the ID varaible, but you can take a look at all the other variables using the left hand side navigator:

ADMINISTRATIVE VARIABLES ⌄

SURVEY QUESTIONS ⌄

DEMOGRAPHIC VARIABLES ⌄

FAMILY INCOME ⌄

INSURANCE STATUS ⌄

MATERIAL HARDSHIP ⌄

SCHOOL ENROLLMENT AND CHILD CARE ⌄

You can look under survey questions to get a better sense of how questions were asked, but please stick to variables under Demographic Variables, Family Income, Insurance Status, and Material Hardship. Do not choose variables from the Administrative levels, Survey Questions, nor School Enrollment or Child Care variables. May leave in the `ID` variable for easier tracking on individuals, but it does not count towards the 10 predictors.

> **!** Task
>
> List the 10 predictors that you plan to use in your analysis. Note which variables are numeric, binary, or multi-level categorical.

Get a sense of how you would like to analyze the data

For our project, we will examine the association between the food insecurity and one other variable (our main explanatory variable). From the above readings, survey information, and your list of predictors in Section , which association are you most interested in analyzing?

Please write this in the form of a research question statement. Feel free to copy this sentence and insert your chosen predictor: We will investigate the association between food insecurity and _____.

> **!** Task
>
> Complete the following statement to identify your research question:
> We will investigate the association between food insecurity and _____.

Save data for processing with `.Rda`

Within this document, or in a separate document, use R to save a copy of the dataset so that you can process it without changing the raw data. Recall the file organization that we discussed to set up proper folders. Include a screenshot showing the new `.Rda` file within your Data folder.

> **!** Task
>
> Include a screenshot showing the new `.Rda` file within your Data folder.

Getting data in working format

You can start by selecting only the variables you will use in your analysis. Again, you can keep `ID` in addition to your outcome and predictors for easy tracking.

Use the following code (with your dataset's name) to remove the parentheses with values that are in front of the category names. Make sure to change `old_df` and `new_df`.

```
new_df = data.frame(lapply(old_df, function(x) {gsub(".*) ", "", x)}))
```

> **❗ Task**
>
> - Select the variables that will be used in your analysis and make a new dataset. Include the code that you used.
> - Remove the parentheses with values that are in front of the category names.

Explore the outcome and predictors

The codebook online gives some nice plots of each variable. Please take a look at the codebook online to see the spread of each variable. Make note of any categorical variables that have less than 100 observations in a group. This may cause issues in our analysis later.

> **❗ Task**
>
> - To check that you have looked that the variables, please report the percent of respondents that were food insecure in the past 12 months.
> - List any categorical variables that have less than 100 observations in a group

Compile above work into an introduction

Please check out this source for what a research article introduction includes and how to organize it. The only thing I would add is mentioning the Well-Being and Basic Needs Survey.

> **❗ Task**
>
> Write an introduction to the analysis.