

Lab 1 Instructions

BSTA 513/613

Nicky Wakim

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

Ready to be worked on! (4/2/26)

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

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](#)
- [NIH's page: Food Insecurity, Neighborhood Food Environment, and Nutrition Health Disparities: State of the Science](#)
- [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, set up your folders for the class and project, including making an `.Rproj` file. Make sure you are working with the project by using the `here()` function to display your working directory.

- You can reference [Lesson 2 from BSTA 512/612](#) for help on folder organization and the `here` package

! Task

Display your working directory using the `here` package and `here()` function.
Insert an image of your folder organization with the `.Rproj` file in the correct place.

Access and download the data

1. Go to the [Health and Medical Care Archive](#) page for the 2023 Well-Being and Basic Needs Survey

At A Glance	Data & Documentation	Variables	Data-related Publications	Export Metadata
Name		Size	Preview	
DS0 Study-Level Files		737 KB		
DS1 Public-Use Data		28 MB		
DS2 Restricted-Use Data		3 MB		

2. Go to the Data & Documentation tab

At A Glance	Data & Documentation	Variables	Data-related Publications	Export Metadata
Name		Size	Preview	
DS0 Study-Level Files		737 KB		
DS1 Public-Use Data		28 MB		
DS2 Restricted-Use Data		3 MB		

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”



For the best experience, please use the email address associated with your institution, work, or school.



Sign in with email



Access through your institution



Sign in with Google



Sign in with ORCID



Sign in with LinkedIn

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.)

9. Read the User Guide! There is some important information about the data and how to use it.

! Task

No task to report back on. Just make sure you have the data and read the User Guide!

Decide on your project type

In Lab 4, you will need to build a model for your outcome. That's really far away from this lab, BUT we will have two options for model building in Lab 4 that affects the work you do in this lab.

For this project, you may either:

1. Use LASSO regression to build a prediction model with no interactions
2. Use Purposeful model selection to build an association model with at least one interaction

We will discuss prediction modeling in Lesson 14 of this class. We discussed association modeling in BSTA 512/612, also Lesson 14.

I highly suggest picking the model selection strategy based on **your desired learning objectives**:

1. **LASSO regression** will help stretch your R coding and machine learning skills. You will not answer a research question about the relationship between a variable and food insecurity. You will ask: What variables are most important for predicting food insecurity?
2. **Purposeful model selection** will allow you to cement many concepts that we learned within 512/612 and 513/613. You will answer a research question about the relationship between a variable and food insecurity.

You can change your mind about which model selection strategy you want to use later on, but your dataset will change a little.

Just note that labs (including this one) will have different activities for each model building strategy.

! Task

Pick between LASSO regression and purposeful model selection. You can just write your choice.

Decide on list of variables to focus on

For purposeful model selection and LASSO regression

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 variable, 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

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,

nor Survey Questions. 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 the type of variable (numeric and continuous, counts, binary, or multi-level categorical)

For LASSO regression only

I wanted you to think critically about which variables you want to include in your model, but for LASSO regression, you can include as many predictors as you want! You can even include all of the variables in the dataset if you want. Even though you just selected 10 variables to focus on, I want to keep all variables from the dataset.

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

For purposeful model selection only

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 _____.

For LASSO regression only

For our project, we will build a prediction model for food insecurity. We will not have a set research question on association. Instead, I want you to remember your main question: What variables are the most predictive of food insecurity?

From the above readings, survey information, and your list of predictors in Section , which predictors do you think will be the most important to predict food insecurity?

! Task

Which predictors do you think will be the most important to predict food insecurity?
You can list 5-10 predictors.

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

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)}))
```

For purposeful model selection only: 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.

We will once again do a complete case analysis (aka drop observations with any missing data). You can use the `drop_na()` function from the `tidyr` package to do this.

! Task

- Remove the parentheses with values that are in front of the category names.
- **Purposeful model selection only:** Select the variables that will be used in your analysis and make a new dataset. Include the code that you used.
- Drop observations with any missing data.

Explore the outcome and predictors

The codebook online gives some plots of each variable. Please take a look at the codebook online to see the spread of each variable. You can also make your own plots - the online codebook is not perfect. 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 at 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

For purposeful model selection only: Make a Table 1

Make a Table 1 that describes the distribution of your outcome and predictors. You can use the `tbl_summary()` function from the `gtsummary` package to do this. Make sure to include the code you used to make the table.

[Check out the Table 1 I made](#) for the gapminder data in BSTA 512/612.

! Task

For purposeful model selection only: Make a Table 1

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 using 3-5 bullets.