

Name: Your Name Here

Note: This worksheet is for both labs 1 & 2. This same worksheet will be submitted for both homework assignments; for lab 1 homework, only the lab 1 section should be filled out, for lab 2 we will finish the second half of the worksheet and turn it in with both sections completed. Be sure to read all of the instructions below. As you type, it may push things on subsequent questions to another page, be that you have completed ALL parts of each section (e.g., in-class vs. take-home).

Purpose

The final project will have you: Exploring a single dataset, selecting variables of interest & developing hypotheses based on the literature in that area, performing statistical tests of those hypotheses, and presenting the results in the context of an APA formatted research report. To prepare for this, labs 1 & 2 will develop the skills necessary to begin the first stages of the final project.

Lab 1 will: Familiarize you with the datasets and variables, get you thinking about how we operationalize variables – especially demographic variables, and give you practice in developing research questions.

In-Class: Complete the in-class portion of the lab 1 worksheet below by working in pairs.

Homework Assigned: More practice with a different dataset. Complete the “on your own” portion of the lab 1 worksheet below individually. Submit this worksheet to your lab section in Canvas with the lab 1 section completed. Due by 11:59 pm the night prior to your scheduled lab 2 meeting.

Lab 2 will: Familiarize you with a different dataset, give you more practice at developing questions, and have you begin formalizing those questions as research hypotheses.

In-Class: For Fall 2023, we will apply the continuity plan; there is no lab 2 meeting during week 2. Complete the homework individually.

Homework Assigned: **1)** Continue this worksheet and complete the lab 2 portion individually. Note that we do not have in-person lab for week 2; the rest of this worksheet is completed on your own. **2)** Additionally, for the second homework assignment you will need to: Select the dataset you will use for the final project, select a “core” variable, and select the other variables that you will relate to this core variable. This assignment is a *separate* worksheet specific to your chosen dataset and can be found under the lab 1 materials. Both assignments will be submitted to your lab section in Canvas under separate links. Due by 11:59 pm the night prior to your scheduled lab 3 meeting.

Looking Forward

In lab 3, we will begin the process of performing a literature review for the final project. We will dig into the literature pertaining to our core variable and find references that can be used to learn more about the core variable that you have selected. This exercise will develop skills in performing a literature search and ensure that we have a solid understanding of the core variable. Later, you will apply these skills to another lit search in which you find references to support a link between your core variable and each of the other variables of interest that you selected in lab 2.

Materials (found under the lecture section on Canvas)

- 1) Lab 1 & 2 worksheet - “new_Lab1&2_datasets_and_variables.docx”
- 2) Selecting a final project dataset and variables worksheet – *there are 3 separate .docx worksheets for this, each is specific to one of the datasets, complete & submit the *one* worksheet that corresponds to your chosen dataset.
- 3) Dataset information – each dataset has a set of files that describe that dataset’s variables, you can find these under the handouts column. For each, there is a data file (.sav), a methods file (.pdf) that tells how the data were collected, and a description (.pdf) of what the variables are.

All answers below should be in RED font!!!

IN-CLASS

1. Find a partner

Enter your partner's name in the text box to the right -> Partner's Name Here

2. Choose a dataset (highlight your choice below).

Self-descriptive Survey

Interpersonal Relationships Survey

Personal Success Survey

3. Identify types of demographic variables.

Each dataset shares the same first 6 demographic variables (listed below). For each, highlight whether it is quantitative (values reflect a count or measurement) **or** qualitative (values correspond to category membership). For each, **only if** it is quantitative, highlight whether it is continuous (can take on any value; e.g., decimal values) **or** discrete (integer or whole number values only).

1. Age	Quantitative	Qualitative	Continuous	Discrete
2. Ethnicity	Quantitative	Qualitative	Continuous	Discrete
3. Family type	Quantitative	Qualitative	Continuous	Discrete
4. Number of siblings	Quantitative	Qualitative	Continuous	Discrete
5. Number of times moved	Quantitative	Qualitative	Continuous	Discrete
6. Hometown population	Quantitative	Qualitative	Continuous	Discrete
7. Gender	Quantitative	Qualitative	Continuous	Discrete

4. Identify remaining demographic variables.

Each dataset has 8 additional *demographic* variables, some are unique to that dataset. Using your chosen dataset and the table below: Enter each other variable name in column 1, tell whether it is quantitative *or* qualitative in column 2, and **if** it is quantitative, tell whether it is continuous, discrete, or either, in column 3.

	Variable Name	Type of Variable	Continuous?
8.			
9.			
10.			
11.			
12.			
13.			
14.			
15.			

continues on next page

5. How does operationalization affect interpretation and/or generalization?

Often, there are no “natural” categories that people fit into, we can almost always “slice” the data finer, but we do need to specify some grouping in our study. For example, ethnicity might be categorical with 2 levels – white or non-white. It might also be 6 levels, as it is in the current datasets. Operationalization is the term for how we define the variable space – for example, ethnicity could be operationalized as binary (white vs. non), or as a 6-level variable. It might even be operationalized as all possible ethnicities (good for representation, but as we will see later in the course potentially bad for statistical “power” and the feasibility of collecting that data). Generalization is how well we are able to take the results we have, given that specific operationalization, and extend it to populations we didn’t specifically test or record data about. Now, consider that these datasets were collected over a decade ago from UNL students. Answer each question below.

a) For the current datasets, ethnicity is 6 levels (what are those levels?) – how might our results be conceptually different if we were to have collected just white vs. non-white; how about if we had collected more categories (e.g., what ethnic groups are “lumped” together into one level of our ethnicity variable that could have otherwise been split further)? How does this affect how ethnicities are represented in this older dataset? How would our understanding of the results change compared to how we might operationalize that variable in current times? Finally, note that each category option is given in terms of “XYZ-American” for participants to self-identify as... what assumptions are built into this, and how might that weirdly force people into the “other” category? How would that potentially skew our data?

b) Next consider the variable “Family Type” in the context of all of the considerations outlined above. What are some issues with the variable as it was operationalized in this older dataset? What might be a better way of defining that variable today... what would the levels be? What knowledge would we gain by defining it differently today?

c) Next, Gender is also a variable in all of the datasets. How is it operationalized here (i.e., what are the levels of the variable)? What are some issues with the variable as it was operationalized in this older dataset? What might be a better way of defining that variable today... what would the levels be? What knowledge would we gain by defining it differently today?

d) Finally, some variables such as age could be quantitative or qualitative – how might we do that? How might that difference in operationalization change our interpretation of the results or generalizability? Similarly, we could measure age in years, or in months – how might that change things?

6. Identifying variables within research questions.

Before we go further, let's examine some research questions. These are simplified, or informal, versions of a research hypothesis (a research hypothesis being a formal statement of specific relationships, often with a specific prediction of the result that is based on prior research or theory). For each research question below, identify the variables that we would need in order to answer the question, and tell whether they are quantitative or qualitative. For quantitative variables, also tell whether it is continuous or discrete. Finally, tell what the population is that we are studying.

Note: Depending on the phrasing of the question, some variables will need to be specifically worded – be sure to take this into account when you list the variable. For others, the variable “name” may not be the same as how it is labeled in the research question – the question may refer to only one category, or “level” of the variable. Think about what the other levels might need to be at a minimum, and what variable name might encompass all of the levels.

- Is there a relationship between weather conditions and absenteeism in Grade 9 at your local school?
- Is there a relationship between the amount of television watched and the level of physical fitness among adult women?
- Are teenage drivers who have been issued speeding tickets more likely to be men?
- Is there a difference in levels of anxiety among UNL students depending on their year in school.

7. Relating demographic information to other variables (e.g., surveys & scales).

Back to your datasets - Below the demographic variables in each dataset, there are 7–8 other variables listed. These reflect a participant’s “score” for a specific scale or survey that they took. Some of these variables are “single-scale score” meaning that one value reflects a participant’s overall score (e.g., anxiety level). Others have an overall score that can be further broken down into sub-scales, meaning that each of these sub-scales offers a standalone piece of information that is different from their overall scale score; each sub-scale tells us something about a unique component of the overall scale.

- a) Identify and list 3 single-scale variables below. For each, tell what it is that it seems to be measuring.
- b) Identify and list 2 variables below that have sub-scales built into them. For each, what does the overall scale seem to measure? What does each sub-scale seem to measure, and how is that piece of info different than the overall scale?
- Note: In your research projects, for any variable with subscales given, you will always be developing and testing questions about one of the sub-scales **not** the overall scale. Keep this in mind going forward as you select a dataset for the project.*

c) Finally, consider the kinds of demographic variables that you listed in the table for #4 above. Now consider the specific scales that your dataset includes. **Why** do you think the researchers chose those specific demographic variables? What kinds of research questions were they interested in (i.e., how are those variables related and why would those demographics matter)?

d) Working with your partner, list at least 6 research questions that you find interesting/important and that you could answer given the data that we currently have. Keep in mind that the **type of variable** (quantitative vs. qualitative) **determines how we phrase** those questions:

- 1 quantitative + 1 qualitative is asking questions about whether groups (qualitative variable's levels) *differ* in terms of some measurement (quantitative). These include wording such as more/less than, greater, lower. i.e., Toyota sells *more* cars than Ferrari.
- 2 quantitative is asking questions about whether there is a relationship between the variables. These include phrasing that talks about how higher/lower scores or measures on one variable will be associated with higher/lower scores on another. i.e., *higher* temperature is associated with *higher* ice cream sales (positively related; both increase together); *more* credit hours per semester is associated with *less* hours spent socializing with friends (negatively related; as one increases, the other decreases and vice-versa).
- 2 qualitative is asking about where people fall on levels of both variables in terms of some pattern. i.e. if our data is broken down into people who do/do not play sports, and people who have/don't have scholarships – we might say, people who *do* play sports also tend to *have* scholarships. We will change how we understand this type of question later, but for now you can think of it as almost like a correlation or relationship type question (like the 2 quantitative example) **except** that it is for qualitative/categorical data instead of quantitative data.

Of the 6 questions you list, make sure that you have at least 1 of each of the types above.

TAKE-HOME

1. Exploring more data and generating additional questions on your own.

Now that you have had a chance to inspect one of the datasets, let's get more practice but with a different dataset this time. On your own, select **one of the other two datasets** – highlight your choice below. Answer the rest of the take-home section below, based on this new dataset.

Self-descriptive Survey

Interpersonal Relationships Survey

Personal Success Survey

2. Identify types of demographic variables.

Each dataset shares the same first 6 demographic variables (listed below). **Even though this part is the same as the in-class, it will be easier if we repeat it here:** Highlight whether it is quantitative (values reflect a count or measurement) **or** qualitative (values correspond to category membership). For each, **only if** it is quantitative, highlight whether it is continuous (can take on any value; e.g., decimal values) **or** discrete (integer or whole number values only).

1. Age	Quantitative	Qualitative	Continuous	Discrete
2. Ethnicity	Quantitative	Qualitative	Continuous	Discrete
3. Family type	Quantitative	Qualitative	Continuous	Discrete
4. Number of siblings	Quantitative	Qualitative	Continuous	Discrete
5. Number of times moved	Quantitative	Qualitative	Continuous	Discrete
6. Hometown population	Quantitative	Qualitative	Continuous	Discrete
7. Gender	Quantitative	Qualitative	Continuous	Discrete

3. Identify remaining demographic variables.

Each dataset has 8 additional *demographic* variables, some are unique to that dataset. Using your chosen dataset and the table below: Enter each other variable name in column 1, tell whether it is quantitative *or* qualitative in column 2, and **if** it is quantitative, tell whether it is continuous, discrete, or either, in column 3.

	Variable Name	Type of Variable	Continuous?
8.			
9.			
10.			
11.			
12.			
13.			
14.			
15.			

continues on next page

4. Relating demographic information to other variables (e.g., surveys & scales).

Below the demographic variables in each dataset, there are 7–8 other variables listed. These reflect a participant's "score" for a specific scale or survey that they took. Some of these variables are "single-scale score" meaning that one value reflects a participant's overall score (e.g., anxiety level). Others have an overall score that can be further broken down into sub-scales, meaning that each of these sub-scales offers a standalone piece of information that is different from their overall scale score; each sub-scale tells us something about a unique component of the overall scale.

*Note: In your research projects later, for any variable with subscales given, you will always be developing and testing questions about **one** of the sub-scales, **not** the overall scale. Keep this in mind going forward as you select a dataset for the project. Also keep in mind that for the project, you will be selecting **one** (non-demographic) "core" variable from the list, and 4–6 other variables you're interested in. For your project, you will be creating **separate** questions about how that **one core variable** is related to each of the other variables you chose.*

a) Identify and list **all** of the single-scale variables below. For **each**: Tell what it is measuring in your own words, then list **one** of the demographic variables, and give a research question that links them. (refer to in-class #7d for the description of types of questions we can ask.)

b) Identify and list below, **all** of the variables that have sub-scales built into them. For each, what does the overall scale seem to measure? What does each sub-scale seem to measure, and how is that piece of info different than the overall scale?

c) For **each** of the variables with a sub-scale, pick just **one** of the sub-scales along with **one other** (non-demographic) variable from the list of scale variables to pair it with (this can be any of the single-scale or sub-scale variables) and give a research question that links them.

LAB 2 TAKE-HOME

Note: This begins the worksheet section for LAB 2. This is the individual work for week 2 and is to be submitted by 11:59 pm the day prior to your lab 3 meeting.

1. Exploring *even more* data and generating additional questions on your own.

Now that you have had a chance to inspect two of the datasets, let's get more practice but with the dataset that you have not yet worked with. On your own, **select the remaining dataset that you have not used in either of the activities above** – highlight your choice below. Answer the take-home section below, based on this last dataset.

Self-descriptive Survey

Interpersonal Relationships Survey

Personal Success Survey

2. Identify types of demographic variables.

Each dataset shares the same first 6 demographic variables (listed below). **Even though this part is the same as before, it will be easier if we repeat it here:** Highlight whether it is quantitative (values reflect a count or measurement) **or** qualitative (values correspond to category membership). For each, **only if** it is quantitative, highlight whether it is continuous (can take on any value; e.g., decimal values) **or** discrete (integer or whole number values only).

1. Age	Quantitative	Qualitative	Continuous	Discrete
2. Ethnicity	Quantitative	Qualitative	Continuous	Discrete
3. Family type	Quantitative	Qualitative	Continuous	Discrete
4. Number of siblings	Quantitative	Qualitative	Continuous	Discrete
5. Number of times moved	Quantitative	Qualitative	Continuous	Discrete
6. Hometown population	Quantitative	Qualitative	Continuous	Discrete
7. Gender	Quantitative	Qualitative	Continuous	Discrete

3. Identify remaining demographic variables.

Each dataset has 8 additional *demographic* variables, some are unique to that dataset. Using your chosen dataset and the table below: Enter each other variable name in column 1, tell whether it is quantitative *or* qualitative in column 2, and **if** it is quantitative, tell whether it is continuous, discrete, or either, in column 3.

	Variable Name	Type of Variable	Continuous?
8.			
9.			
10.			
11.			
12.			
13.			
14.			
15.			

continues on next page

4. Relating demographic information to other variables (e.g., surveys & scales).

Below the demographic variables in each dataset, there are 7–8 other variables listed. These reflect a participant's "score" for a specific scale or survey that they took. Some of these variables are "single-scale score" meaning that one value reflects a participant's overall score (e.g., anxiety level). Others have an overall score that can be further broken down into sub-scales, meaning that each of these sub-scales offers a standalone piece of information that is different from their overall scale score; each sub-scale tells us something about a unique component of the overall scale.

*Note: In your research projects later, for any variable with subscales given, you will always be developing and testing questions about **one** of the sub-scales, **not** the overall scale. Keep this in mind going forward as you select a dataset for the project. Also keep in mind that for the project, you will be selecting **one** (non-demographic) "core" variable from the list, and 4–6 other variables you're interested in. For your project, you will be creating **separate** questions about how that **one core variable** is related to each of the other variables you chose.*

a) Identify and list **all** of the single-scale variables below. For **each**: Tell what it is measuring in your own words, then list **one** of the demographic variables, and give a research question that links them. (refer to in-class #7d for the description of types of questions we can ask.)

b) Identify and list below, **all** of the variables that have sub-scales built into them. For each, what does the overall scale seem to measure? What does each sub-scale seem to measure, and how is that piece of info different than the overall scale?

c) For **each** of the variables with a sub-scale, pick just **one** of the sub-scales along with **one other** (non-demographic) variable from the list of scale variables to pair it with (this can be any of the single-scale or sub-scale variables) and give a research question that links them.

5. Getting ready to pick a dataset to use for the final project.

Now you have had a chance to explore each of the available datasets, and have had plenty of practice generating potential research questions. Let's think ahead to the final project – next week, lab 3, we will learn how to do a literature search and find information about the variables you will choose for the project. It's fun to come up with research questions as we did thus far, but consider that a big part of research is having **support** for these questions. In this sense, support doesn't just mean a hunch... it means that someone else has already established an experimental or theoretical link between the pairs of variables you will choose. In this context, support means in the form of published, peer-reviewed, research articles.

a) Think about which dataset was most interesting for you to work with; which one was most natural for you to come up with research questions for? This is likely the dataset you should pick for your project as it will be easier if you find that you are personally interested in the variables you are working with. Below, list the one dataset that you plan to use for your final project.

b) For your final project, you will then select **one** “core” variable from the list of scale (non-demographic) variables in your dataset. Further, for any variable with a sub-scale, we will **not** be using the overall single scale score... instead our “core” variable will be **one** of the specific sub-scales (another way to think about it is that these variables really have 2–4 separate variables built into them and if we choose from these, we pick one of the sub-scales in it... not the overall name of the variable that the sub-scale is a part of). Keep in mind that whichever of these we choose, it will be included in **each** of the research questions we generate and then later test. Which variable seems most interesting to you? Choose one to be your core variable and list it below.

c) Finally, you will need to choose 4–6 **other** variables. Each of these will be related to your core variable in its own research question that you will test. You will be constrained in what you can choose though; 2–3 must be qualitative, and 2–3 must be quantitative! Take a look through the variables and select **3 quantitative and 3 qualitative** variables (other than the core variable, though they can be other sub-scales from the same overall scale that your core variable belongs to, i.e., if you use the self-descriptive dataset and your core variable is the Positive Affect sub-scale of the Affective Balance Scale, you may choose as one of your other variables, the Negative Affect sub-scale of the Affective Balance Scale). For each of the **6 other** variables you chose here, list them below and give a research question you might ask; the question should relate that variable to the core variable.

That's it! Now you are well prepared for lab 3 in which you will perform a literature search to find some research articles that help you learn about your core variable. In future steps of the final project, you will do another search; however, this time, you will be looking for articles that somehow link each of your variables to the core variable – since our research questions will then include a prediction of the outcome of the test, this is the **support** you will use to defend your reasoning for that prediction.