

**GSS 2014 MODULES ON
EXPLORING VIEWS ABOUT IMMIGRATION IN THE US
By Heidi Sarabia, PhD**

Purpose: These series of modules will help students think about immigration and the factors that shape views on immigration among the US population. In addition, the purpose of this assignment is to guide students to use SPSS to generate information, analysis, and a report about a large dataset (i.e. the General Social Survey).

In this assignment, you will:

- Explore variables related to views about immigration asked in the 2014 General Social Survey
- Develop a research question regarding factors that shape views about immigration, using two variables found in the General Social Survey (2014)
- Use SPSS to generate descriptive and inferential statistics
- Produce tables and synthesize the information obtained
- Write an organized, clear, and concise report on the results found

Knowledge and Skills: By completing this assignment, you will gain the following knowledge:

- A review of the skills needed to use SPSS
- A review of how to use the 2014 General Social Survey
- A review of the steps needed to conduct analysis
- The ability to analyze the results to produce a report
- The ability to create a simple report to communicate your findings

Tasks: To complete this assignment, you should:

1. Determine a topic of interest related to the factors that shape views on immigration in the US
2. Select two variables related to the topic of interest
3. Write a research question that connects the two variables
4. Analyze each variable independently
5. Conduct bivariate analysis to explore the relationship between the two variables

Summary of the Project: The exercises will be split into three parts. In the first part, the students will explore descriptive statistics that summarize the data from the 2014 General Social Survey related to views on immigration; in the second part, students will think about inferential statistics and use the 2014 General Social Survey to infer views from the US population related to perspectives on immigration; and in the final part, students will examine relationships among two variables in order to explore what factors shape views on immigration. Each part will consist of several exercises (organized as different lab sessions) for a total of 14 lab sessions or exercises. Each lab consists of a reflection question at the beginning, descriptions of the tasks, boxes with instructions to use SPSS, and practice problems at the end.

Table of Contents:

PART I: DESCRIPTIVE STATISTICS.....	3
LAB #1: INTRODUCTION TO VARIABLES	3
LAB #2: DESCRIBING THE DATA USING TABLES	6
LAB #3: DESCRIBING THE DATA USING GRAPHS	9
LAB #4: REPORTING MEASURES OF CENTRAL TENDENCY	12
LAB #5: REPORTING MEASURES OF DISPERSION	15
LAB #6: RECODING VARIABLES	17
LAB #7: CREATING DICHOTOMOUS VARIABLES.....	20
LAB #8: CREATING AN INDEX	23
PART II: INFERENTIAL STATISTICS.....	26
LAB #9: CONFIDENCE INTERVALS	26
LAB #10: ONE-SAMPLE HYPOTHESIS TESTING.....	28
LAB #11: TWO-SAMPLES HYPOTHESIS TESTING	31
LAB #12: ANOVA	33
PART III: RELATIONSHIP BETWEEN TWO VARIABLES.....	35
LAB #13: CROSSTABS	35
LAB #14: CHI-SQUARE AND CRAMER'S V	38
LAB #15: LAMBDA AND GAMMA	41
LAB #16: REGRESSION AND CORRELATION.....	44
PART IV: SUMMARY.....	47
LAB #17: FINAL PROJECT	47
INDEX: SUMMARY OF SPSS PROCEDURES	49

Lab #1: Introduction to Variables

The United States is considered to be a “nation of immigrants” due to the large share of people who have come from other places. But different waves of immigration have come from different parts of the world. Today, the largest share of immigrants come from Asia and Latin America. You can explore different waves of migration using the *New York Times*’ “Immigration Explorer”

(https://archive.nytimes.com/www.nytimes.com/interactive/2009/03/10/us/20090310-immigration-explorer.html?_r=)



Reflection Question: Do you think the current anxieties expressed about immigration are shaped by where these immigrants are coming from?

Lab #1 Learning Objectives:

- Learn how to use the GSS website to select variables
- Select a variable of interest using the GSS website and use the information on the website to determine the level of measurement of the variable, and the categories/values of the variable

Concepts you should review before you do this lab:

- Dependent and independent variables
- Levels of measurement: nominal, ordinal, interval-ratio
- Categories/Values of variables

Immigration is a topic that seems to be in the news often. From the travel ban to certain refugees and immigration from mostly Muslim countries, to Central American migrants at the US-Mexico border, to round ups by Immigration and Customs Enforcement (ICE) agents in the US, to immigration officials losing track of thousands of children at the US Mexico border. But how does the US population feel about immigrants in general? What factors shape these views?

In these series of labs, we will be using the 2014 General Social Survey (GSS), which is a random sample of the US population (for an in-depth description of the sampling technique, refer to http://gss.norc.org/documents/codebook/GSS_Codebook_AppendixA.pdf), to learn more about what factors shape people’s views on immigration.

In this first part (labs 1-7), we will be working on describing the GSS sample. Descriptive statistics refers to the summary of variables in an entire population, or an entire sample. We will focus on certain variables that are related to views among the US population about the issue of immigration. Variables are characteristics in a population that we are interested in studying, such as race, gender, income, or as in this case, views about immigration. There are different types of variables, and for the purposes of these exercises, we will only focus on three levels of measurement of variables: nominal, ordinal, and interval-ratio. In these exercises, we will focus on variables that reveal how people think about immigration in general, and immigration from different parts of the world in particular.

First, we want to explore the variables that are related to the topic of immigration. Typically, survey data has a codebook that describes each question asked in the survey, how that was coded (in terms of how the variable and the categories of the variable were coded). Since GSS asked thousands of variables, we are going to use the GSS web search engine to explore these variables.

EXPLORE THE DATA ON GSS WEBSITE

INSTRUCTIONS: Go to the General Social Survey (GSS) website (<http://gss.norc.org/>) → Go to “Explore the data” → Click on “search variable” to find questions related to immigration that were asked in the 2014 survey → Type “immigrant.”

NOTE TO STUDENT: Are you interested in other issues? Maybe opinions about family structure? Abortion rights? Gun laws? Political involvement? Police abuse? You can always explore other issues in this search engine and complete a report on other issues.

Dependent Variables and Independent Variables

For this lab, we will focus on five dependent variables related to the topic of immigration that were asked in the **2014 GSS**. Use the table below to describe them in terms of (a) the data the variable captures, (b) the categories of the variable, and (c) the level of measurement of the variable.

Dependent Variables (NAME)	What does the variable capture?	Categories of the Variable	Level of Measurement
IMMAMESCO			
IMMIDEAS			
IMMCRIME			
IMMCULT			
IMMASSIM			

What factors do you think shape how people think about these issues? Race? Ethnicity? Age? Gender?

For this lab, we will focus on five independent variables we think might influence how people think about and view immigration. Use the table below to describe these variables in terms of (a) the data the variable captures, (b) the categories of the variable, and (c) the level of measurement of the variable.

Independent Variables (NAME)	What does the variable capture?	Categories of the Variable	Level of Measurement
	Race		
	Age		
	Income		
	Education		
	Gender		

NOTE TO STUDENT: Are there other factors you think are important in shaping how people think about immigration? Maybe opinions about the government? Political affiliation? You can always explore other issues in this search engine and complete a report on other issues as well.

SUMMARY: In this lab, you used the GSS website to search for variables related to a topic of interest. Note that the website provides information about the variables that are related to different topics and the year when the GSS asked those questions. You can also go to the descriptive data on the website and see how the variable was coded.

LAB #1 PRACTICE PROBLEMS. Use the GSS website to answer these questions:

Q1. What does the variable IMMEDUC measure? __ Views about whether legal immigrants should have the same education as Americans _____

Q2. What are the categories of the variable IMMJOBS? __ (1) Agree strongly, (2) Agree, (3) neither agree nor disagree, (4) Disagree, (5) Disagree strongly _____

Q3. What is the level of measurement of the variable PATRIOT4?
__ ordinal _____

Q4. What year was the question, LETINHSP, asked? __ 2000 _____

Q5. What does the variable LETIN1 measure? __ If the person thinks the number of immigrants coming to America nowadays should be increased/maintained/reduced _____

Consider: What are other variables that measure views on immigration? What are other variables you think have an impact (or shape) people's views on immigration?

Lab #2: Describing the Data Using Tables

The United States collects a lot of data regarding the individuals it admits legally into the US (either as Legal Permanent Residents or as tourists), as well as the individuals it detains for trying to enter or living unauthorized in the US. You can view the 2016 Yearbook of Immigration Statistics online at: <https://www.dhs.gov/immigration-statistics/yearbook/2016>



Reflection Question: If you were to collect data on the individuals entering the US, what information would you collect? Why?

Lab #2 Learning Objectives:

- Learn how to open GSS data on SPSS
- Change setting options on GSS
- Run a frequency distribution table

Concepts you should review before you do this lab:

- Frequency Distribution Table
- Percentages
- Cumulative percentages

In this lab, we will explore the 2014 GSS dataset using SPSS. First, let's download the dataset and open the 2014 GSS dataset in SPSS.

OPEN 2014 GSS ON SPSS

INSTRUCTIONS: Go to the General Social Survey (GSS) website (<http://gss.norc.org/>) → Go to "Get the Data" → Click on "SPSS" → Click on "2014" → Select "Save File" → Select Folder where you will save dataset

INSTRUCTIONS: Open SPSS and then click on "GSS2014" to open.

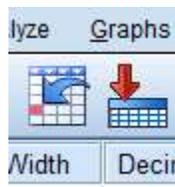
Note that there are two tabs on the bottom:

"Data View" shows all the names of the variables on top row and all the cases in the first column. GSS2014 has 2,538 cases (individual responses to the variables).

VARIABLES: CASES:

"Variable View" shows all the variables and their specifications. The top row shows all the specifications of each variable (name, type, description, categories, level of measurement, etc.), and the first column has the number of variables. GSS2014 has 896 variables.

Note that by clicking on the 7th or 8th icon at the top, you can search for a particular case or go to a particular variable.



SETTING OPTIONS

INSTRUCTIONS: The first thing we are going to do now, and every time we open a new GSS file will change the settings to make sure the program displays information the way we want it to in order to make it easier to read results.

First, to display abbreviated variables names alphabetically: Go to “Edit” → “Options” → Go to the “General” tab → “Variable Lists” → Click on “Display names” and “Alphabetical”

In addition, to display output tables and graphs with values and labels: Go to the “Output” tab → “Pivot Table Labeling” → Click on Variable in labels shown as “Names and Labels” and also on Variable values in item labels shown as “Values and Labels” → Click “Ok”

Frequency Distribution Tables

Now we are ready to describe our variables using frequency distribution tables and graphs.

First, let’s run a frequency distribution table for the variable that measures views on assimilation (IMMASSIM).

FREQUENCY DISTRIBUTION TABLE

INSTRUCTIONS: “Analyze” → “Descriptive Statistics” → “Frequencies” → Select variable [IMMASSIM] → Click “Ok”

Note that the first table, “Statistics,” tells us how many people answered the question (these are the “valid cases”) and the second table tells us the percent of respondents who gave each answer. Remember to always report “valid percent.”

- What percentage of people oppose acculturation by agreeing that “Immigrants should retain their culture of origin and not adopt American culture”? _____
- What percentage of people support the idea of the melting pot by agreeing that “Immigrants should retain their culture of origin and also adopt American culture”? _____
- What percentage of people support the idea of complete acculturation by agreeing that “Immigrants should give up their culture of origin and adopt American culture”? _____

SUMMARY: In this lab, you opened the 2014 GSS dataset using SPSS, changed the option settings on SPSS to make sure outputs are displayed correctly, and ran a frequency distribution table of a variable.

LAB #2 PRACTICE PROBLEMS. Use the 2014 GSS dataset to answer these questions:

Q1. Run a frequency distribution table of the variable IMMAMESCO. What percentage of the sample “agree strongly” that immigrants are good for America? ____9.3____

Q2. Run a frequency distribution table of the variable IMMCRIME. What percentage of the sample “disagree strongly” that immigrants increase crime rates? ____11.6____

Q3. Run a frequency distribution table of the variable IMMIDEAS. What percentage of the sample “agree” that immigrants make America more open? ____54.4____

Q4. Run a frequency distribution table of the variable IMMCULT. What percentage of the sample “disagree” that immigrants undermine American culture? ____50.2____

Q5. Run a frequency distribution table of the variable IMMEDUC. What percentage of the sample “neither agree nor disagree” that legal immigrants should have the same education as Americans? ____7.8____

Consider: Do you think most people in the U.S. in 2014 favored or opposed immigration? What factors do you think shape whether people favor or oppose immigration?

Lab #3: Describing the Data Using Graphs

The United States spends a considerable amount of money on policing its borders. See the US Border Patrol Budget from 1990 to 2017.

(<https://www.cbp.gov/sites/default/files/assets/documents/2017-Dec/BP%20Budget%20History%201990-2017.pdf>)



Reflection Question: What might be a better way to display this information? Further, what factors do you think shape our decision to invest more than 3 billion dollars policing the US border in 2017?

Lab #3 Learning Objectives:

- Learn how to run pie graphs, bar graphs, histograms, and line graphs to visually represent data
- Determine how level of measurement shapes type of graph use

Concepts you should review before you do this lab:

- Levels of measurement: nominal, ordinal, interval-ratio
- Bar/Pie/Histogram/line graphs

In this lab, we are going to use graphs to illustrate the 2014 GSS data.

First, remember that the level of measurement determines the appropriate chart we should use. We are only focusing on describing one variable.

Here is a summary:

	Level of Measurement		
	<i>Nominal</i>	<i>Ordinal</i>	<i>Interval-ratio</i>
Bar chart	X	X	
Pie chart	X	X	
Histogram			X
Line chart			X

Pie Chart

Let's begin by doing a pie chart of a nominal variable: race (RACE)¹.

PIE CHART

INSTRUCTIONS: “Graphs” → “Legacy Dialogs” → “Pie” → Select “Summaries for group cases” → click “Define” → Select variable RACE into “Define Slices by” → Click “Ok”

Note that you can represent slices by “N of cases” or “% of cases”

¹ Note that there are various variables that measure the concept of race/ethnicity (e.g. raceself, ethnic, hispanic, etc.), the variable RACE is only one such measure.

Note: Once you have the pie chart you can edit it by double-clicking on the pie (which will open a new window with the pie) → Then click again on the pie (which will open a new window with more options). You can change colors. Also, by right-clicking on the pie and selecting “show data labels,” you can also add the N or percent for each section of the pie (select percent).

- What percentage of the sample are White? 74.47
- What percentage of the sample are Black? 15.21

Bar Chart

Let’s now do a bar chart of an ordinal variable: about views on immigrant rights (IMMRGHTS).

BAR CHART

INSTRUCTIONS: “Graphs” → “Legacy Dialogs” → “Bar” → Select “Simple” → click “Define” → Select variable IMMRGHTS into “Category Axis” → Click “Ok”

Note: Once you have the bar chart you can edit it by double-clicking on the bars (which will open a new window with the bars) → Then click again on the bars (which will open a new window with more options). You can change colors. Also, by right-clicking on the pie and selecting “show data labels,” you can also add the N or percent for each bar (select percent).

- What percentage of the sample “agree strongly” that legal immigrants should have the same rights as Americans? 6.84
- What percentage of the sample “disagree strongly” that legal immigrants should have the same rights as Americans? 10.31

Histogram

Let’s now do a histogram of an interval-ratio variable: age (AGE).

HISTOGRAM

INSTRUCTIONS: “Graphs” → “Legacy Dialogs” → “Histogram” → Select variable AGE into “Variable” → Click “Ok”

Note: Once you have the histogram, you can edit it by double-clicking on the bars (which will open a new window with the bars) → Then click again on the bars (which will open a new window with more options). You can change colors. Also, by selecting “Binning”, you can select X Axis, Custom, and Number of Intervals (and specify the number of intervals you want 10, 20, 30, etc.).

- What is the size of the sample (N)? 2529
- What is the mean age of the sample? 49.01

Line Chart

Let's now do a line chart of another interval-ratio variable: education (EDUC).

LINE CHART

INSTRUCTIONS: “Graphs” → “Legacy Dialogs” → “Line” → Select “Simple” → Select variable EDUC into “Category Axis” → Click “Ok”

Note: Once you have the line you can edit it by double-clicking on the line (which will open a new window with the line) → Then click again on the line (which will open a new window with more options). You can change color and width of the line. By clicking on the background you can also change the background color. Also, by right-clicking on the line, you can select “show data labels” and show data points by N or % (select percent).

- What percent of the sample has zero years of education? 0.32
- What percent of the sample has 12 years of education? 27.16

SUMMARY: In this lab, you reviewed how the level of measurement of each variable determines the appropriate graphic representation of that variable. Then you learned how to do pie, bar, histogram, and line graphs.

LAB #3 PRACTICE PROBLEMS. Use the 2014 GSS dataset to answer these questions:

Q1. Run a pie graph of the variable LETIN1 and add the percentages to each section. What percentage of the sample thinks that the number of immigrants to America nowadays should be “reduced a lot”? 19.09

Q2. Run a pie graph of the variable LETIN1 and add the percentages to each section. What percentage of the sample thinks that the number of immigrants to America nowadays should be “increased a lot”? 3.73

Q3. Run a bar graph of the variable PATRIOT4 and add the percentages to each section. What percentage of the sample “agree” that Patriotic feelings lead to negative feelings toward immigrants? 36.06

Q4. Run a bar graph of the variable PATRIOT4 and add the percentages to each section. What percentage of the sample “disagree” that Patriotic feelings lead to negative feelings toward immigrants? 29.70

Q5. Run a histogram of the variable TVHOURS. What is the mean number of hours the sample spends watching TV per day? 2.98

Consider: From the reflection question at the beginning of this lab, what would be the best way to display the data reported by the US Border Patrol about its annual budget? What would such a graphic look like? What would it be the most salient characteristic?

Lab #4: Reporting Measures of Central Tendency

According to TRAC, on average, “Legal Noncitizens Receive Longest ICE Detention” (<http://trac.syr.edu/immigration/reports/321/>).



Reflection Question: Why do you think legal (as opposed to unauthorized) non-citizens spend more time in detention facilities? What factors do you think shape detention time?

Lab #4 Learning Objectives:

- Determine how level of measurement determines appropriate measures of central tendency to be reported
- Learn how to get measures of central tendency using two SPSS commands

Concepts you should review before you do this lab:

- Levels of measurement: nominal, ordinal, interval-ratio
- Measures of central tendency: mode, median, mean

In this lab, we are going to describe the sample in terms of central tendency (the average). The level of measurement of the variable will determine the appropriate measures we report on the data.

Here is a summary:

Level of Measurement of the Variable	Measures of Central Tendency			Measures of Dispersion		
	Mode	Median	Mean	Range	Variance	Standard Deviation
Nominal	X	--	--	--	--	--
Ordinal	X	X	--	--	--	--
Interval-ratio	X	X	X	X	X	X

When we think about what is average about a sample, we must also consider the level of measurement. In the case of nominal level variables, we only report the mode (the most common group); in the case of ordinal level variables, we can report mode and median (the category that splits the sample in half); and in the case of interval-ratio variables, we can report mode, median, and mean (the arithmetic average of the sample).

There are two ways to get this information on SPSS: using the frequencies procedure or using the descriptive procedure. Let's analyze the measures of central tendency for the amount of time people spend watching TV each day (TVHOURS).

FREQUENCIES PROCEDURE TO GET MEASURES OF CENTRAL TENDENCY (Interval-ratio)

INSTRUCTIONS: “Analyze” → “Descriptive Statistics” → “Frequencies” → Select variable TVHOURS

→ Click on “Statistics” → Select under “Central Tendency” the boxes for “Mean,” “Median,” and “Mode” → Click “Continue” → Click “Ok”

- What is the mode of TVHOURS? 2
- What is the median of TVHOURS? 2
- What is the mean of TVHOURS? 2.98

Note that most people in the sample (437) watched 2 hours of television per day; and half of the sample watch 2 or less hours and half watch 2 or more hours; but if we put all the hours on a weight scale, the average would be 2.98 hours.

Let’s now explore the mode for views about immigrant rights (IMMRGHTS) since this is an ordinal level variable:

DESCRIPTIVES PROCEDURE TO GET MEASURES OF CENTRAL TENDENCY (Ordinal)

INSTRUCTIONS: “Analyze” → “Descriptive Statistics” → “Frequencies” → Select variable IMMRGHTS

→ Click on “Statistics” → Select under “Central Tendency” the boxes for “Mode” and “Median” → Click “Continue” → Click “Ok”

Note that the mode and median are reported by SPSS as a “4.” But you must think, what does it mean to be a 4? SPSS is reporting on the category “4” as the most common response (mode) and also as the category that divides the sample in half. In this case, most people (40.3%) “disagree” (mode) with the statement that “Legal Immigrants should have the same rights as Americans,” half of the sample are at or below that category, and half of the sample are at or below that category (median).

Let’s now explore the mode for views about assimilation/acculturation (IMMASSIM) since this is a nominal level variable:

FREQUENCIES PROCEDURE TO GET MEASURES OF CENTRAL TENDENCY (Ordinal)

INSTRUCTIONS: “Analyze” → “Descriptive Statistics” → “Frequencies” → Select variable IMMASSIM

→ Click on “Statistics” → Select under “Central Tendency” the boxes for “Mode” → Click “Continue” → Click “Ok”

Note that the mode is reported by SPSS as a “2.” But you must think, what does it mean to be a 2? SPSS is reporting on the category “2” as the most common response. In this case, most people (91.5%) agreed with the statement that “Immigrants should retain their culture of origin and also adopt American culture.”

Another way to get the mean for many variables at the same time is to use the “descriptives” procedure.

DESCRIPTIVES PROCEDURE TO GET MEASURES OF CENTRAL TENDENCY (Interval-ratio)

INSTRUCTIONS: “Analyze” → “Descriptive Statistics” → “Descriptives” → Select variables TVHOURS, WWWHR, EMAILHRS
→ Click on “Options” → Select “Mean” → Click “Continue” → Click “Ok”

- What is the average number of hours people in the sample spent watching TV per day (TVHOURS)? ____ 2.98 ____
- What is the average number of hours people in the sample spent on the internet (WWWHR)? ____ 11.62 ____
- What is the average number of hours people in the sample spent on their e-mail (EMAILHR)? ____ 6.27 ____

SUMMARY: In this lab, you reviewed how the level of measurement of the variable determines the appropriate measurement of central tendency. Then, you learned how to obtain the measures of central tendency (mode, median, and mean) using two SPSS procedures: frequencies or descriptives.

LAB #4 PRACTICE PROBLEMS. Use the 2014 GSS dataset to answer these questions:

- Q1. What is the mode of the variable IMMEDUC? ____ Agree ____
- Q2. What is the mode of the variable IMMJOBS? ____ Disagree ____
- Q3. What is the median of the variable IMMAMECO? ____ Agree ____
- Q4. What is the median of the variable IMMCRIME? ____ Disagree ____
- Q5. What is the mean of the variable AGE? ____ 49.01 ____

Consider: In what ways do you think age shapes the way people think about education, or jobs for immigrants? Or whether immigrants contribute to crime in the US?

Lab #5: Reporting Measures of Dispersion

View the “Age-Sex Pyramids of US Immigrants and Native Born Populations” created by Migration Policy Institute (MPI) (<https://www.migrationpolicy.org/programs/data-hub/charts/age-sex-pyramids-immigrant-and-native-born-population-over-time>)



Reflection Question: What does the spread of the population, at each age interval, reveal about the demographic characteristics of US-born natives and immigrants in the US?

Lab #5 Learning Objectives:

- Determine how level of measurement determines appropriate measures of dispersion to be reported
- Learn how to get measures of dispersion using two SPSS commands

Concepts you should review before you do this lab:

- Levels of measurement: nominal, ordinal, interval-ratio
- Measures of dispersion: range, variance, standard deviation

In this lab, we are going to describe the sample in terms of dispersion (the spread of the data). Remember, the level of measurement of the variable will determine the appropriate measures we report on the data (refer to the table on the previous lab). We report range, inter-quartile range, variance, and standard deviation only for interval-ratio data.

There are two ways to get this information on SPSS: using the frequencies procedure or using the descriptive procedure. Let's analyze the measures of dispersion for the amount of time people spend watching TV each day (TVHOURS).

FREQUENCIES PROCEDURE TO GET MEASURES OF DISPERSION

INSTRUCTIONS: “Analyze” → “Descriptive Statistics” → “Frequencies” → Select variable TVHOURS

→ Click on “Statistics” → Select under “Dispersion” the boxes for “St. deviation,” “Variance,” and “Range” → Click “Continue” → Click “Ok”

Note that you can also get specific percentiles using this option.

- What is the range of TVHOURS? 24
- What is the variance of TVHOURS? 6.695
- What is the standard deviation of TVHOURS? 2.587

Another way to get the standard deviation for many variables at the same time is to use the “descriptives” procedure.

DESCRIPTIVES PROCEDURE TO GET MEASURES OF DISPERSION

INSTRUCTIONS: “Analyze” → “Descriptive Statistics” → “Descriptives” → Select variables TVHOURS, WWWHR, EMAILHRS
 → Click on “Options” → Select “Std. deviation,” “Variance,” and “Range” → Click “Continue” → Click “Ok”

- What is the range for the number of hours people in the sample spent on the internet (WWWHR)? ____ 122 ____
- What is the standard deviation for the number of hours people in the sample spent on their e-mail (EMAILHR)? ____ 11.335 ____

SUMMARY: In this lab, you reviewed how the level of measurement of the variable determines the appropriate measurement of dispersion. Then, you learned how to obtain the measures of dispersion (range, variance, standard deviation) using two SPSS procedures: frequencies or descriptives.

LAB #5 PRACTICE PROBLEMS. Use the 2014 GSS dataset to answer these questions:

Q1. What is the range of the variable AGE? ____ 71 ____

Q2. What is the variance of the variable AGE? ____ 303.173 ____

Q3. What is the standard deviation of the variable AGE? ____ 17.412 ____

Q4. What is the 25th percentile of the variable AGE? ____ 34 ____

Q5. What is the 75th percentile of the variable AGE? ____ 62 ____

Consider: In what ways do you think age shapes the way we think about and view immigration and immigrants in the US? Why?

Lab #6: Recoding the Data

Contrary to what the media tends to focus on, it has been widely shown that immigrants commit crimes at lesser rates than the US-born population. In fact, integration into US society is associated with increases in offenses (see a Pew Report: <http://www.pewresearch.org/fact-tank/2013/10/15/crime-rises-among-second-generation-immigrants-as-they-assimilate/>)



Reflection Question: Why do you think integration into US society is associated with higher levels of crime? Why does the first-generation (those born in another country commit less crimes) while their children, second-generation (those born in the US), look more like the native-born?

Lab #6 Learning Objectives:

- Recode one variable to reduce the number of categories in that variable
- Transform a variable from ordinal to dichotomous variable

Concepts you should review before you do this lab:

- Levels of measurement: nominal, ordinal, interval-ratio

First, run a frequency distribution table of the variable IMMAMECO, which measures opinions about whether people think that immigrants are good for America.

As you can see, the variable is an ordinal variable with 5 categories that range from “agree strongly” to “disagree strongly.” Now let’s say that we want to change the variable to (1) keep only those cases where people had an opinion, and (2) reduce the range of responses to agree strongly/agree or disagree/disagree strongly. To do this, we will recode the variable, from having five categories to having only two categories.

Table to prepare to recode the variable

Old category	Old value of the old category	RECODE	New value of the new category	New category
Agree strongly	1	→	1	Agree
Agree	2	→	1	Agree
Neither agree nor disagree	3	→	.	
Disagree	4	→	2	Disagree
Disagree strongly	5	→	2	Disagree

RECODING A VARIABLE

INSTRUCTIONS: “Transform” → “Recode Into Different Variables” → Select variable IMMAMECO

On “Output Variable” → Under “Name”: IMMAMECO_NEW (this is the name of the new variable) and under “Label”: Are immigrants good for America (this is the description of the new variable) → Click “Change”

→ Select “Old and New Values” →

- Under “Old Value”: 1; under “New Value”: 1; click “Add”
- Under “Old Value”: 2; under “New Value”: 1; click “Add”
- Under “Old Value”: 3; under “New Value”: “system missing”; click “Add”
- Under “Old Value”: 4; under “New Value”: 2; click “Add”
- Under “Old Value”: 5; under “New Value”: 2; click “Add”
- Under “Old Value”: “system-or user-missing; under “New Value”: “system missing”; click “Add”

→ Click “Continue” → Click “Ok”

Now go to the data set and go to “Variable view.” Scroll to the end of the list and change the titles of the categories by clicking on “Values”

“Value”: 1; “Label”: Agree → Click “Add”

“Value”: 2; “Label”: Disagree → Click “Add”

→ Click “Ok”

What percentage of the sample agrees that immigrants are good for America (recode variable IMMAMECO_NEW)? 74.1

- What percentage of the sample disagrees that immigrants are good for America (recode variable IMMAMECO_NEW)? 25.9

SUMMARY: In this lab, you learned how to recode a variable, from having five categories to having two categories.

LAB #6 PRACTICE PROBLEMS. Use the 2014 GSS dataset and recode the variable IMMIDEAS from having 5 categories that range from agree strongly to disagree strongly to a new variable (IMMIDEAS_NEW) having two categories (agree and disagree, dropping the cases where people felt neutral) and answer these questions:

Q1. What is the number of valid responses for the new variable IMMIDEAS_NEW?
970

Q2. What is the number of missing responses for the new variable IMMIDEAS_NEW?
1568

Q3. What percentage of the sample agree that “Immigrants make America more open”?
83.8

Q4. What percentage of the sample disagree that “Immigrants make America more open”?
_____ 16.2 _____

Q5. What is the mode of the variable IMMIDEAS_NEW? _____ Agree _____

Consider: Once we focus on people with stated options (agree/disagree), do we get a different picture of opinions? Do you think those with stated opinions differ from those with neutral opinions? How?

Lab #7: Creating Dichotomous Variables

The Pew Research Center found that political parties really shape how people think about immigration (<http://www.pewresearch.org/fact-tank/2015/09/30/on-views-of-immigrants-americans-largely-split-along-party-lines/>)



Reflection Question: How and why do you think political affiliations shape our views on immigration?

Lab #7 Learning Objectives:

- Transform variables from ordinal-level to nominal (dichotomous)

Concepts you should review before you do this lab:

- Levels of measurement: nominal, ordinal, interval-ratio
- Dichotomous variables

Dichotomous variables are nominal-level variables that have only two categories. For example, we can code variables to be dichotomous as follows: Gender (men and women), Race (white and non-white), Favorite color (green and non-green), etc.

In this lab, we are going to re-code 4 variables from ordinal level to dichotomous. These variables measure support for immigration and the categories will be Yes/No.

For example, IMMAMECO, which asks, “Are immigrants good for America?” right now, the options are (1) Strongly agree, (2) Agree, (3) Neutral, (4) Disagree, and (5) Strongly Disagree. It is an ordinal-level variable. We are going to re-code IMMAMECO as IMMAMECO_DICHOTOMOUS, as (1) Yes or (0) No. Yes will include (1) Strongly agree and (2) agree, and No will include (4) Disagree, and (5) Strongly Disagree.

We will re-code the following four variables into dichotomous variables: IMMAMECO, IMMEDUC, IMMIDEAS, IMMRGHTS

Table to prepare to recode the variable

Old Variable	Old values of the old category	RECODE	New Variable	New categories
IMMAMECO		→	IMMAMECO_DICHOTOMOUS	
	(1) Strongly agree	→ 1		(1)Yes
	(2) Agree			
	(3) Neither agree nor disagree	→ 1		
	(4) Disagree	→ .		.

	(5) Disagree strongly	→ 0		(0) No
		→ 0		
IMMEDUC		→	IMMEDUC_DICHOTOMOUS	
IMMIDEAS		→	IMMIDEAS_DICHOTOMOUS	
IMMRGHTS		→	IMMRIGHTS_DICHOTOMOUS	

Note that 1 expresses a support for immigrants and 0 expresses opposition to immigration.

- What percentage of the sample agrees that immigrants make American more open (recode variable IMMAMECO_DICHOTOMOUS)? 74.3
- What percentage of the sample agrees that legal immigrants should have the same education as Americans (recode variable IMMEDUC_DICHOTOMOUS)? 91.3
- What percentage of the sample agrees that immigrants make American more open (recode variable IMMIDEAS_DICHOTOMOUS)? 85.0
- What percentage of the sample agrees that legal immigrants should have the same rights as Americans (recode variable IMMEDUC_DICHOTOMOUS)? 43.4

Now we will re-code the following four variables into dichotomous variables: IMMCRIME, IMMCULT, IMMJOBS, EXCLDIMM.

We will recode them differently because we want 1 to express support for immigrants and 0 to express opposition to immigration.

Old Variable	Old values of the old category	RECODE	New Variable	New categories
IMMCRIME		→	IMMCRIME_DICHOTOMOUS	
	(1) Strongly agree	→ 0		(0)Yes
	(2) Agree			
	(3) Neither agree nor disagree	→ 0		
	(4) Disagree	→ .		.
	(5) Disagree strongly	→ 1		(1)No
		→ 1		
IMMCULT		→	IMMCULT_DICHOTOMOUS	
IMMJOBS		→	IMMJOBS_DICHOTOMOUS	
EXCLDIMM		→	EXCLDIMM_DICHOTOMOUS	

- What percentage of the sample agrees that immigrants increase crime rates (recode variable IMMCRIME_DICHOTOMOUS)? 27.2

- What percentage of the sample agrees that immigrants undermine American culture (recode variable IMMCULT_DICHOTOMOUS)? 23.7
- What percentage of the sample agrees that immigrants take jobs away (recode variable IMMJOBS_DICHOTOMOUS)? 45.2
- What percentage of the sample agrees that America should exclude undocumented immigrants (recode variable EXCLDIMM_DICHOTOMOUS)? 71.1

SUMMARY: In this lab, you learned how to recode an ordinal-level variable into a nominal-level variable with two categories, which we call dichotomous.

LAB #8 PRACTICE PROBLEMS. Use the 2014 GSS dataset and recode the variable “Should the number of immigrants nowadays be increased?” (LETIN1) as a dichotomous variable (1) yes or (0) no (LETIN1_DICHOTOMOUS), and answer the following questions:

Q1. What is the number of valid cases for the new variable LETIN1_DICHOTOMOUS?
634

Q2. What percentage of people think immigration should be increased? 24.7

Now recode the variable “Political affiliation” (PARTYID) as a dichotomous variable (1) republican or (0) democrat (PARTYID_DICHOTOMOUS),
Code as democrat: 0, 1, 2 and as republican 4,5,6, and as missing 3,7,8,9
and answer the following questions:

Q3. What is the number of valid cases for the new variable PARTYID_DICHOTOMOUS?
1931

Q4. What percentage of the sample identifies as republican? 41.8

Q5. What percentage of the sample identifies as democrat? 58.2

Consider: Again, how do you think party affiliation shapes views on immigration?

Lab #8: Creating an Index

The Migrant Integration Policy Index (MIPEX) combines 167 indicators from 38 countries to create an index of migrant-friendly countries (measured by the policies that seek to integrate migrants). According to this index, the friendliest countries are Sweden, followed by Portugal and New Zealand; and the US ranks 9th (out of 38). (<http://www.mipex.eu/key-findings>)



Reflection Question: What are some policies that you think could be adopted in the US that would make it more migrant-friendly?

Lab #7 Learning Objectives:

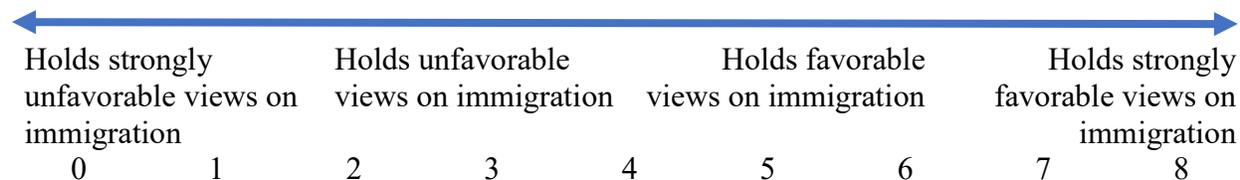
- Combine various variables to create an index (a composite measure using many variables)

Concepts you should review before you do this lab:

- Levels of measurement: nominal, ordinal, interval-ratio
- Index

We have been working with many variables that measure people's opinion about immigration. Some of the questions measure positive views on immigration (such as IMMAMECO, IMMAEDUC, IMMIDEAS, IMMRGHTS), while other variables measure negative views on immigration (IMMCRIME, IMMCULT, IMMJOBS, EXCLDIMM). In the previous lab (#7), we recoded these eight variables into dichotomous variables. In this lab, we are going to combine all those variables to create an index, that is, an indicator or measure of views on immigration that combines various variables.

When we create an index, we are counting people who answer these different questions in a particular way. For example, we are going to combine eight variables (IMMAMESCO_DICHOTOMOUS, IMMEDUC_DICHOTOMOUS, IMMIDEAS_DICHOTOMOUS, IMMRGHTS_DICHOTOMOUS, IMMCRIME_DICHOTOMOUS, IMMCULT_DICHOTOMOUS, IMMJOBS_DICHOTOMOUS, EXCLDIMM_DICHOTOMOUS) to create an index of Immigration Views. Since there are already dichotomous variables, each person will get 0 or 1 points. Those who get 8 points (the highest possible number), we know hold favorable views on immigration; while those who get 0 points (the lowest possible number) we know hold unfavorable views on immigration



To create the index, we will be counting the times people with answers of 1.

To prepare the index, let's write the expected value of the index (remember that the "expected index score" is calculated by adding all the "1" for the eight variables, and we are only counting those who answered all the eight questions):

Variables	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 11	Case 12
immameco_dichotomous	1	1	.	.	1	.	1	0
immeduc_dichotomous	1	1	.	.	1	.	1	0
immideasc_dichotomous	1	1	.	.	1	.	1	0
immrghts_dichotomous	0	0	.	.	0	.	.	0
immcrime_dichotomous	1	1	.	.	1	.	1	0
immcult_dichotomous	1	1	.	.	1	.	.	0
immjobs_dichotomous	0	1	.	.	0	.	1	0
exclimm_dichotomous	1	0	.	.	0	.	1	0
Expected Index Score	6	6	.	.	5	.	.	0
Score checked (after index creation)	6	6	.	.	5	.	.	0

Now let's create the index.

CREATING AN INDEX

INSTRUCTIONS: "Transform" → "Count Values Within Cases"

On "Target Variable," write the name of the new variable:

IMMIVIEWS_INDEX

On "Target Label," write a description of the new variable: Index on Immigration Views

On "Variables," Select the four variables we are going to use to create the index:

IMMAMECO_DICHOTOMOUS

IMMIDEAS_DICHOTOMOUS

IMMAEDUC_DICHOTOMOUS

IMMRGHTS_DICHOTOMOUS

IMMCRIME_DICHOTOMOUS

IMMCULT_DICHOTOMOUS

IMMJOBS_DICHOTOMOUS

EXCLDIMM_DICHOTOMOUS

→ Select "Define Values" →

- Under "Value": 1 → Click "Add" (this means that all the "1"s will be counted).

→ Select "If" → Click "Include if case satisfies condition:

- Type the following on textbox:

NMISS (IMMAMECO_NEW, IMMIDEAS_NEW, IMMAEDUC_NEW, IMMRGHTS_NEW, IMMCRIME_DICHOTOMOUS, IMMCULT_DICHOTOMOUS, IMMJOBS_DICHOTOMOUS, EXCLDIMM_DICHOTOMOUS)=0

→ Click "Continue" → Click "Ok"

Now check the expected scores with the actual data coded to make sure your index was coded correctly.

Run a frequency distribution table of the new variable IMMVIEWS_INDEX.

- How many people answered all eight questions for the index (N)?
_____ 351 _____
- What percentage of the sample holds strongly unfavorable views on immigration (score of 0)? _____ 1.6 _____
- What percentage of the sample holds strongly favorable views on immigration (score of 8)? _____ 19.2 _____

SUMMARY: In this lab, you learned how to create an index by combining the answers from eight variables.

LAB #7 PRACTICE PROBLEMS. Use the 2014 GSS dataset and create an index (ANTI_IMMVIEWS_INDEX) with the following three variables: IMMCRIME, IMMCULT, IMMJOBS. Create the index by counting the value 1 and 2 (agree and strongly agree). Then Change the “Values” under “Variable View”

- 3: Holds strongly unfavorable views on immigration
- 2: Holds unfavorable views on immigration
- 1: Holds favorable views on immigration
- 0: Holds strongly favorable views on immigration

Run a frequency distribution table of the new variable IMMVIEWS_INDEX and answer these questions:

Q1. What is the range of the index variable ANTI_IMMVIEWS_INDEX? _____ 0-3 _____

Q2. What does it mean to have a 0 in ANTI_IMMVIEWS_INDEX? _____ Person holds strongly favorable views on immigration _____

Q3. What does it mean to have a 3 in ANTI_IMMVIEWS_INDEX? _____ Person holds strongly unfavorable views on immigration _____

Q4. How many people answer all three questions for the index? _____ 1159 _____

Q5. What percentage of the sample “holds strongly unfavorable views on immigration”?
_____ 7.6 _____

Consider: Once we put the index together, do we get a different picture of views on immigration? Why do you think we get different numbers when we put an index of positive views on immigration versus an index of negative views on immigration?

Lab #9: Confidence Intervals

The Migration Policy Institute (MPI) has a “Profile of the Unauthorized Population in the United States,” which includes estimations of size, years of residence in the US, age, labor force participation, etc. (<https://www.migrationpolicy.org/data/unauthorized-immigrant-population/state/US>)



Reflection Question: Why do we only have estimates of the unauthorized population in the US? How do we get these estimates? How useful do you think these estimates are?

Lab #8 Learning Objectives:

- Created a confidence interval about the population from sample data

Concepts you should review before you do this lab:

- Sample and population
- Point estimate
- Confidence intervals
- Upper and lower bound

In the previous labs (1-7), part I, we were describing the data set, focusing on single variables, such as age or views on the impact of immigrants on jobs. In the next three labs (8-10), part II, we will now focus on how to use the information we get from the sample to draw conclusions about the population; this is called inferential statistics.

First, we have to clarify that GSS is a representative sample of the US population. In order to achieve this representative sample, particular sampling techniques are used to collect the data. For a more detailed discussion of the methodology, visit Appendix A (Sampling Design and Weighting): <https://gssdataexplorer.norc.org/documents/441/display>

The weight variable is designed to correct for things like household size and oversampling of certain populations. For 2014, the weight we are going to use for inferential statistics will be WTSSNR (for the entire list of GSS weights visit: <https://gssdataexplorer.norc.org/pages/show?page=gss%2Fweighting>).

To see the effect of weights: what is the average and standard deviation of age (AGE) of the 2014 GSS sample without weights? 49.01 (with a standard deviation of 17.41)

Now turn on the weights:

TURN ON WEIGHTS

INSTRUCTIONS: “Data” → “Weight cases”

Click on “Weight cases by,” and select variable WTSSNR

→ Click “Ok”

Now that the data is weighted: what is the average and standard deviation of age (AGE) of the 2014 GSS sample with weights? 47.42 (with a standard deviation of 17.22)

In this lab, we will be calculating confidence intervals. A point estimate is an estimate (based on sample data) that consists of a single value or point (Lane:328). For example, based on the 2014 GSS, we estimate that the average age in the US adult population is 47.42 years. However, this number does not reveal what is uncertainly associated with the estimate (Lane:337). Hence, we typically report the confidence interval, which contains the population mean, a specified proportion of the time, typically either 99% or 99% of the time (Lane: 337).

CONFIDENCE INTERVALS

INSTRUCTIONS: “Analyze” → “Descriptive Statistics” → “Explore”

On “Dependent List,” select variable AGE

Click on “Statistics,” and select box “Descriptives”, and add percent 99

→ Click “Continue” → Click “Ok”

- What is the point estimate mean of age of adults in the US? 47.42
- What is the lower bound for the estimated mean of age of adults in the US?
46.75
- What is the upper bound for the estimated mean of age of adults in the US?
48.08

Interpret this information: We are 95% confident that the actual mean age of adults in the US lies between 46.75 and 48.08 years; there is a 5% change we could be wrong and the actual mean age of adults is less than 46.75 or more than 48.08.

SUMMARY: In this lab, you learned how to weight the data and create a confidence interval of an interval-ratio variable.

LAB #8 PRACTICE PROBLEMS. Use the 2014 GSS dataset using weights, get a 99% confidence interval for the variable WWWMIN (minutes per week spent on the internet), and answer these questions:

Q1. What is the point estimate of WWWMIN? 1.36

Q2. What is the lower bound estimate of WWWMIN? .93

Q3. What is the upper bound estimate of WWWMIN? 1.78

Get a 95% confidence interval for the variable WWWMIN (minutes per week spent on the internet), and answer these questions:

Q4. What is the lower bound estimate of WWWMIN? 1.04

Q5. What is the upper bound estimate of WWWMIN? 1.68

Consider: How do you think spending time on the internet shapes views on immigration? Why?

Lab #10: One-Sample Hypothesis Testing

In 2011, Jose Antonio Vargas, an immigrant born in the Philippines, who has won a Pulitzer Prize, wrote an essay in the *New York Times* that revealed he was an undocumented immigrant in the U.S. Since, then, he has founded a non-profit organization, “Define American” to engage with the public about the issues facing undocumented immigrants in the U.S. who arrived to the U.S. at a young age (<https://defineamerican.com/>)



Reflection Question: In your opinion, what defines American? What do you think about the argument that many undocumented immigrants are Americans without papers?

Lab #9 Learning Objectives:

- Learn how to conduct a hypothesis test for proportions
- Learn how to conduct a hypothesis test for means

Concepts you should review before you do this lab:

- The steps to conduct a hypothesis test

ONE SAMPLE T-TEST FOR PROPORTIONS

In 2000, 90.6% thought that immigration to the U.S. should not be increased. In 2014, 85.9% of people agreed that immigration to the U.S. should not be increased. Is this change (decrease) statistically significant? Let’s test the hypothesis that there has been a decrease in the number of people who agree that immigration to the U.S. should not be increased at a .05 significance level.

First, let’s change the variable LETIN1 into a dichotomous variable (refer to Lab #6 on how to recode a variable). Recode the new variable LETIN1_RECODED as follows:

Old category	Old value of the old category	RECODE	New value of the new category	New category
Increased a lot	1	→	1	Yes
Increased a little	2	→	1	Yes
Remain the same as it is	3	→	0	No
Decreased a lot	4	→	0	No
Decreased a little	5	→	0	No

Run the frequency distribution table of your new variable LETIN1_RECODED to confirm that 85.9% of the sample (2) disagrees that immigration should be increased.

Now let’s set up the hypothesis test:

1. State the null hypothesis, the alternative hypothesis, and the alpha value:

$$H_0: \pi \geq .906$$

$$H_1: \pi < .906$$

$$\alpha = 0.05$$

- Run the One Sample t test on SPSS

ONE SAMPLE T TEST FOR PROPORTIONS

INSTRUCTIONS: “Analyze” → “Nonparametric Tests” → “Legacy Dialogs” → “Binomial”

- Select variable LETIN1_RECODED
 - Under “Test Proportion” write the expected percentage: .892
- Click “Continue” → Click “Ok”

Note: We are using a nonparametric test because we are testing two proportions; therefore, we are not making the assumption that the distribution of proportions follows a normal curve.

- Interpret the table: The last column of the table gives you the probably of getting a .859 by chance, which is .000
- Compare P and alpha
 - $P < \alpha$
 - $.000 < .05$
- Reach a conclusion: We reject the null hypothesis and accept the research hypothesis
- Interpret the results: We have evidence, at a 0.05 significance level, that the percentage of people who do not want immigration to increase in the U.S. has decreased from 89.2% to 85.9%, from 2000 to 2014. There is still a 5% probability that we could be observing a difference that is due to chance.

ONE SAMPLE T-TEST FOR MEANS

In 2000, people spent an average of 4.12 minutes per week on the internet (not counting e-mail), and 3.89 minutes on e-mail per week. In 2014, people reported spending an average of 1.45 minutes on the internet, a significant decrease, and 2.99 minutes on e-mail. Is this decrease in e-mail use statistically significant? Let’s test the hypothesis that there has been a decreased in the number minutes people spend on e-mail per week at .05 significance level.

Let’s set up the hypothesis test:

- State the null hypothesis, the alternative hypothesis, and the alpha value:
 - $H_0: \mu \geq 3.89$
 - $H_1: \mu < 3.89$
 - $\alpha = 0.05$
- Run the One Sample t test on SPSS

ONE SAMPLE T TEST FOR MEANS

INSTRUCTIONS: “Analyze” → “Compare means” →

- Select variable EMAILHRS
 - Under “Test Proportion” write the expected mean: 3.89
- Click “Continue” → Click “Ok”

9. Interpret the table: the third column, which reads “Sig. (2-tailed)” gives the two tail probability of getting a mean difference of .9, which in this case is .000. Note that this is a two-tailed probability, but since we are only interested in the left-hand side (one-tailed), we need to divide the probability by 2 (in this case it does not make a difference, $.000/2=.000$)
10. Compare P and alpha
 $P < \alpha$
 $.000 < .05$
11. Reach a conclusion: We reject the null hypothesis and accept the research hypothesis.
12. Interpret the results: We have evidence, at a 0.05 significance level, that the average time people spend on their e-mail per week has decreased from 3.89 minutes to 2.99 minutes, from 2000 to 2014. There is still a 5% probability that we could be observing a difference that is due to chance.

SUMMARY: In this lab you learned how to do a one-sample hypothesis test for proportions and for means.

LAB #9 PRACTICE PROBLEMS. Use the 2014 GSS dataset using weights, recode the variables EXCLDIMM (as EXCLDIMM_RECoded) into dichotomous variable (1,2 as 1(yes), and 4,5 as (no)) to answer these questions:

Q1. What percentage of people did not think that American should take stronger measures to exclude undocumented immigrants? _____ 27.3 _____

In 2004, 16.9% of people did not think that American should take stronger measures to exclude undocumented immigrants. Determine whether the increase is statistically significant at a 0.05 level.

Q2. What is the research hypothesis? _____ $\pi > .169$ _____

Q3. What is the null hypothesis? _____ $\pi \leq .169$ _____

Q4. What is the probability of getting $p > .169$? _____ .000 _____

Q5. Comparing P and alpha, what is your conclusion? _____ We reject the null hypothesis _____

Consider: What do you have changed in the past 10 years that have made people change their mind about excluding undocumented immigrants?

Lab #11: Two-Samples Hypothesis Testing

Immigration seemed to be front and center in the presidential elections of 2016, and political party seems to be neatly aligned with views on immigration

(http://www.newamericaneconomy.org/wp-content/uploads/2014/11/PNAE_sharedvalues.pdf).



Reflection Question: Why do you think political party shapes the way people view immigration issues?

Lab #10 Learning Objectives:

- Learn how to conduct a hypothesis test for two means
- Learn how to conduct a hypothesis test for two proportions

Concepts you should review before you do this lab:

- The steps to conduct an independent samples hypothesis test

INDEPENDENT SAMPLES T-TEST FOR MEANS

We created an index in lab #8 (IMMVIEWES_INDEX) with 8 measures of views on immigration; if you get a mean of the index, the average for the sample is 5.45.

Do you think men and women have different opinions on immigration? Let's test the hypothesis that men and women have different views on immigration at a .05 significance level.

Let's set up the hypothesis test:

1. State the null hypothesis, the alternative hypothesis, and the alpha value:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

$$\alpha = 0.05$$

2. Run the Two (Independent) Samples T test on SPSS

INDEPENDENT SAMPLES T-TEST FOR MEANS

INSTRUCTIONS: "Analyze" → "Compare means" → "Independent Samples T-Test"

→ For Test Variable, select the IMMVIEWES_INDEX

○ Under "Grouping variable", select: SEX

▪ Click on "Define Groups"

• For "Group 1" type: 1

• For "Group 2" type: 2

→ Click "Continue" → Click "Ok"

Note: This is a parametric test (which means that we assume that the variable is normally distributed) and is an independent-groups t-test, which means that subjects in the first group (women) cannot also be in the second group (men).

3. Interpret the tables:
 - a. The first table, “Group Statistics” gives us the mean for each group. Note that on average, men have a mean of 5.47, while women have a mean of 5.41. Is this difference statistically significant at a 0.05 level? Let’s look at the second table.
 - b. The second table first gives us “Lavene’s Test for Equality of Variances” so we must first establish whether the variances are equal or not.
 - i. First we look at the F-statistic: .116, and the probability of getting this F-statistic (“Sig”): .734. Since P (.734) is greater than our alpha (.05), we can accept the null hypothesis of equality. Yes, variances for the two groups (men and women) are statistically equal.
 - ii. Now we look at the “t-test for Equality of Means”
 1. First we look at the t-statistic: .263, and the probability of getting this t-statistic (“Sig. (2-tailed)”): .793.
4. Compare P and alpha: Since P (.793) is greater than our alpha (.05), we can accept the null hypothesis of equality.
5. Reach a conclusion: We accept the null hypothesis and reject the research hypothesis.
6. Interpret the results: We have evidence, at a 0.05 significance level, that average index scores for men and women are not statistically different. The difference observed, of 0.062 points is due to sampling error.

SUMMARY: In this lab you learned how to do a two-sample hypothesis test for means.

LAB #10 PRACTICE PROBLEMS. Compare the means of the immigration index (IMMVIEWES_INDEX) by party affiliation (PARYID_DICHOTOMOUS):

- ➔ For Test Variable, select the IMMVIEWES_INDEX
- Under “Grouping variable”, select: PARYID_DICHOTOMOUS
 - Click on “Define Groups”
 - For “Group 1” type: 0
 - For “Group 2” type: 1

Q1. What is the average index score on immigration views for democrats? 5.69

Q2. What is the average index score on immigration views for republicans? 4.66

Q3. Are variances equal (F-statistic)? yes

Q4. What is the t-statistic? 3.755

Q5. Is the difference in averages between democrats and republicans statistically significant?
yes

Consider: How do you think political party shapes views on immigration? Why?

Lab #12: ANOVA

Many people think that the 2016 election was largely shaped by ideas about race and anxieties about demographic changes in the U.S. (<https://nyti.ms/2BwdlKO>)



Reflection Question: How and why do you think race shapes our views about immigration in the U.S.?

Lab #11 Learning Objectives:

- Learn how to conduct a one-way analysis of variance (ANOVA)

Concepts you should review before you do this lab:

- The steps to conduct a one-way analysis of variance (ANOVA)

In the previous lab (#10), we compared two groups (women and men, and democrats and republicans); now we want to see whether views on immigration, as measured by the index on views on immigration (IMMVIEWES_INDEX), is different among different racial groups (RACE). Race in GSS is measured as white, black, or other, and in this ANOVA, we are going to compare the means (on the index) among these three groups at a 0.05 significance level.

ANOVA

INSTRUCTIONS: “Analyze” → “Compare means” → “One-Way ANOVA”

- For “Dependent List”, select the variable: IMMVIEWES_INDEX
- For “Factor”, select variable: RACE

→ Click on “Post Hoc...”

- Under “Equal Variances Assumed”, select: “Tukey”

→ Click “Continue” → Click “Options” and Select “Descriptives”

→ Click “Continue” → Click “Ok”

Note: ANOVA is an omnibus test statistic; therefore, if you do find a statistically significant difference, ANOVA cannot tell you which of the different groups were different. ANOVA only tells you that at least two of groups were statistically significantly different. We run the Post Hoc Test (Tukey) to determined which differences were significant.

Let’s analyze the results:

The first table “Descriptives” tells you the mean for each of the groups: Whites had an average of 5.37, Blacks 4.91, and Other 6.23. Are these differences statistically significant?

The second table “ANOVA” gives you the ANOVA (or one-way analysis of variance), we look at the first row “Between groups” to see if there is a statistically significant difference between our group means. We get an F-statistic of: 5.355. The probability of getting this F-

statistic (“Sig.”) is: .005. Since P (0.005) is less than our alpha (0.05), then we reject the null hypothesis and accept the research hypothesis of difference. Yes, there is a statistically significant difference in the mean between the three racial groups when it comes to views about immigration. But were all the groups different? How did each group differ from each other?

The third table “Multiple Comparisons” gives us the comparisons between specific groups.

The difference between White and Black is of .456 points, and the probability (“Sig.”) of getting this difference by chance is .342, which is higher than our alpha (0.005); therefore, this difference is not statistically significant.

The difference between White and Other is of .864 points, and the probability (“Sig.”) of getting this difference by chance is .020, which is lower than our alpha (0.005); therefore, this difference is statistically significant.

The difference between Black and Other is of .425 points, and the probability (“Sig.”) of getting this difference by chance is .006, which is lower than our alpha (0.005); therefore, this difference is statistically significant.

In sum, White and Black groups have statistically equal views on immigration, but differences between White and Other groups and Black and Other groups statistically different at a .05 significance level.

SUMMARY: In this lab you learned how to conduct a one-way analysis of variance (ANOVA) and interpret the results.

LAB #11 PRACTICE PROBLEMS. Compare the means of the immigration index (IMMVIEWES_INDEX) by education. First recode education (EDUC) into 3 categories: 0-12 as High School or less, 12-14 as College or some college, and 14-20 as Graduate degree (name the new variable EDU_RECODED).

Now conduct a one-way analysis of variance (ANOVA):

- For “Dependent List”, select the variable: IMMVIEWES_INDEX
- For “Factor”, select variable: EDU_RECODED
- Click on “Post Hoc...”
 - Under “Equal Variances Assumed”, select: “Tukey”
- Click “Continue” → Click “Options” and Select “Descriptives”

Q1. What is the mean difference between those with HS or less and those with some college?
0.328

Q2. What is the mean difference between those with HS or less and those with a graduate education? 0.783

Q3. What is the mean difference between those with college and those with a graduate education? 0.454

Q4. What is the F-statistic for ANOVA? 2.471

Q5. Is the difference between groups statistically significant? No

Lab #13: Crosstabs

Discussions about immigration and walls seem to go hand-and-hand in today's political debates. While the U.S. has one of the longest and most militarized borders in the world, it is not certainly the only wall in place. See other walls:

<https://www.thisamericanlife.org/extras/walls>



Reflection Question: What do you think are the factors that shape whether the population is more or less supportive of building a longer wall between the U.S. and Mexico?

Lab #12 Learning Objectives:

- Learn how to construct a bivariate table
- Use SPSS to construct a 2x2 table

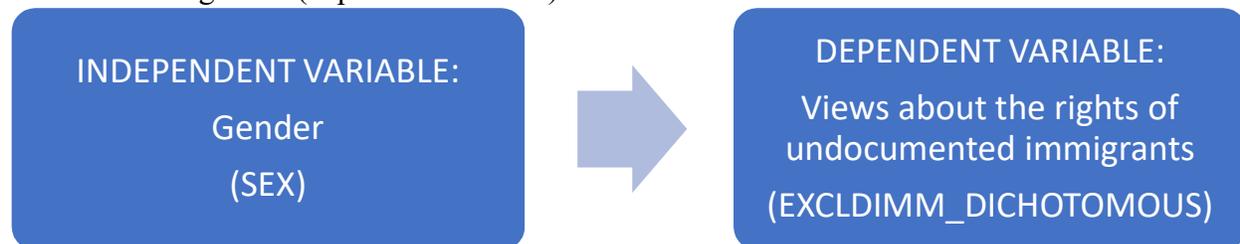
Concepts you should review before you do this lab:

- Dependent and Independent variable
- Bivariate table
- Column percentage
- Marginals

In the previous labs, we have been exploring one variable at a time—thus, we have been engaging in univariate analysis. In the next labs, we will be exploring the relationship between two variables—also known as bivariate analysis.

Let's begin by considering the relationship between two nominal (dichotomous) variables.

So first, let's consider what factor (variable) we think might shape (as an independent variable) views on immigration (dependent variable):



Hypothesis

First, let's come up with a hypothesis about the relationship between gender and views on undocumented immigrants. We might hypothesize that because women have been socialized to be more attuned to issues of inequality, they might have more supportive views on immigration. So women might be more likely to support rights for undocumented immigrants.

Bivariate Table

Let's build a bivariate table to see the relationship between these two variables:

BIVARIATE TABLE

INSTRUCTIONS: “Analyze” → “Descriptive Statistics” → “Crosstabs”

- For “Row”, select the dependent variable:
EXCLIMM_DICHOTOMOUS
- For “Column”, select the independent variable: SEX

→ Click on “Cells”

- Under “Percentages”, select: “Column”

→ Click “Continue” → Click “Ok”

- What is the marginal percentage (total percentage) for people who think undocumented immigrants should be excluded in the U.S.? 71.1
- What is the percentage (column percentage) for women who think undocumented immigrants should be excluded in the U.S.? 69.3
- What is the percentage (column percentage) for women who think undocumented immigrants should be excluded in the U.S.? 73.2

Analysis

Notice that if there was no relationship between gender and views on immigration, 71.1% of men and 71.1% of women would think undocumented immigrants should be excluded in the U.S. Instead, we see that women are less likely to support that statement, and men are slightly more likely to support that statement. The percent difference between men and women supporting the statement that undocumented immigrants should be excluded in the U.S. is 3.9%. So we have some evidence that there is a relationship between gender and views on immigration. But is this difference significant? In the next lab we will test whether this difference is statistically significant.

SUMMARY: In this lab you learned how to construct a bivariate table, and add the column percentages to conduct a preliminary analysis of the relationship between two variables.

LAB #12 PRACTICE PROBLEMS. Recode the variable RACE as a dichotomous variable (white/nonwhite) and examine if race (RACE_DICHOTOMOUS) shapes support of rights of legal immigrants (IMMRGHTS_DICHOTOMOUS), and answer these questions:

Q1. What is the total percentage (marginal percentage) of people who think legal immigrants should NOT have the same rights as Americans (IMMRGHTS_DICHOTOMOUS)?
56.6

Q2. What is the percentage of whites (column percentage) who think legal immigrants should NOT have the same rights as Americans (IMMRGHTS_DICHOTOMOUS)?
60.4

Q3. What is the percentage of non-whites (column percentage) who think legal immigrants should NOT have the same rights as Americans (IMMRGHTS_DICHOTOMOUS)?
_____ 45.4 _____

Q4. Do you see preliminary evidence that whites and non-whites have different views about the rights of legal immigrants? _____yes _____

Q5. What is the percent difference between whites and non-whites in their views about the rights of legal immigrants? _____ 15.0 _____

Consider: Do you think this difference is statistically significant? Why do you think whites and non-whites have different views about the rights of legal immigrants?

Lab #14: Chi-Square and Cramer's V

The Pew Research Center reported that when it comes to refugees, factors such as political affiliation, race, age, and education, shaped views about U.S.'s responsibility to accept refugees (http://www.pewresearch.org/fact-tank/2018/05/24/republicans-turn-more-negative-toward-refugees-as-number-admitted-to-u-s-plummets/ft_18-05-23_refugeeviews_demographic/)



Reflection Question: What do you think are the factors that shape whether people think the U.S. has a responsibility to accept refugees also shapes how people think about legal residents and undocumented immigrants? Or considerations for refugees are unique and different?

Lab #13 Learning Objectives:

- Learn how to obtain and interpret a chi-square and cramer's v from a bivariate table

Concepts you should review before you do this lab:

- Dependent and Independent variable
- Bivariate table
- Cross-tabulation
- Hypothesis testing (null and research hypothesis)

CHI-SQUARE

From the previous lab, we observed that women are less likely to support that statement that undocumented immigrants should be excluded from U.S.-society, and men are slightly more likely to support that statement. But we could not conclude whether this difference of 3.9% was statistically significant at a 0.05 level or was just the result of sampling error.

So now, we are going to conduct a hypothesis test, using the chi-square statistic, to determine if the percent difference observed is statistically significant.

Null Hypothesis (statement of equality): There is no relationship between the dependent (IMMRGHTS_DICHOTOMOUS) and independent variable (SEX)—they are independent

Research hypothesis (statement of difference): There is a relationship between the dependent (IMMRGHTS_DICHOTOMOUS) and independent variable (SEX)—they are dependent

Distribution: What distribution should you use for this hypothesis test? Chi-square Why? Because we are testing for independence.

CHI-SQUARE AND CRAMER'S V

INSTRUCTIONS: “Analyze” → “Descriptive Statistics” → “Crosstabs”

- For “Row”, select the dependent variable:
EXCLIMM_DICHOTOMOUS
- For “Column”, select the independent variable: SEX

→ Click on “Cells”

- Under “Percentages”, select: “Column”

→ Click on “Statistics”

- Select “Chi-square,”
- Under “Nominal” and select: “Phi and Cramer’s V”

→ Click “Continue” → Click “Ok”

Chi-Square (X^2) value: ___ 18.571 _____

P-value: ___ .000 _____

Compare P and alpha: ___ $P < \alpha$ _____

Reach a conclusion: Reject the null hypothesis, accept the research hypothesis.

Interpret decision: We have evidence at a .05 significance level that the variables gender and views on the rights of legal immigrants are dependent; the observed differences is not due to sampling error.

Cramer’s V: ___ .132 _____

Strength of this association: Very weak

Interpret strength of association: Gender shapes views on rights of legal immigrants but weakly.

Table of Levels of Association

Value	Strength of Association
0.00	No relationship
.00 to .15	Very weak
.16 to .20	Weak
.21 to .25	Moderate
.26 to .30	Moderately strong
.31 to .35	Strong
.36 to .40	Very strong
.40 to .50	Worrisomely strong (variables might be measuring the same thing)
.50 to .90	Redundant (variables might be measuring the same thing)
1.00	Perfect Relationship

SUMMARY: In this lab you learned how to obtain and interpret a chi-square and v from a bivariate table, and (1) using chi-square, conduct a hypothesis test to determine if the differences observed are statistically significant, and (2) using Cramer’s V determine the strength of association between the independent and dependent variable.

LAB #13 PRACTICE PROBLEMS. Use the recode the variable RACE as a dichotomous variable (white/nonwhite) and examine if the relationship between race (RACE_DICHOTOMOUS) and support of rights of legal immigrants (IMMRGHTS_DICHOTOMOUS) is statistically significant (X^2) and the strength of this association (Cramer's V), and answer these questions:

Q1. What is X^2 ? _____ 30.536 _____

Q2. What is the probability (sig.) of getting this X^2 ? _____ .000 _____

Q3. What conclusion did you reach? _____ reject the null hypothesis _____

Q4. What is Cramer's V value? _____ .174 _____

Q5. What is the strength of association between the two variables? _____ Weak _____

Consider: Why do you think whites and non-whites have different views about the rights of legal immigrants?

Lab #15: Lambda and Gamma

Even though much of the recent discussion has been about building a wall (or continuing the construction of the wall) along the U.S.-Mexico border, most U.S.-Americans (57%) actually oppose expanding the construction of such wall

(<https://news.gallup.com/poll/235775/americans-oppose-border-walls-favor-dealing-daca.aspx>)



Reflection Question: Why do you think there is a disconnect between the focus of the national government (insisting on building the wall) and popular option (as shown by the polls)?

Lab #14 Learning Objectives:

- Learn how to obtain and interpret lambda and gamma values depending on the level of measurement of the variables.

Concepts you should review before you do this lab:

- Dependent and Independent variable
- Cross-tabulation

Lambda (λ)

Lambda (λ) is a measure of association for nominal variables (like Cramer's V); this value is based on the Proportional Reduction in Error (PRE). If we multiply lambda by 100, it gives us the percentage by which, by using the independent variable as a predictor, how much we are able to predict the dependent variables without any error. Also, note that this an asymmetrical measure of association, which means that the value depends on which value you consider independent and which value you consider a dependent.

Let's see how strongly gender (SEX) shapes views on immigration (IMMAMECO_DICHOTOMOUS).

LAMBDA

INSTRUCTIONS: "Analyze" → "Descriptive Statistics" → "Crosstabs"

- For "Row", select the dependent variable:
IMMAMECO_DICHOTOMOUS
- For "Column", select the independent variable: SEX

→ Click on "Cells"

- Under "Percentages", select: "Column" → click "continue"

→ Click on "Statistics"

- Under "Nominal" and select: "Lambda"

→ Click "Continue" → Click "Ok"

The value of Lambda is: 0.038

Table of Levels of Association

Value	Strength of Association
0.00	No relationship
.00 to .15	Very weak
.16 to .20	Weak
.21 to .25	Moderate
.26 to .30	Moderately strong
.31 to .35	Strong
.36 to .40	Very strong
.40 to .50	Worrisomely strong (variables might be measuring the same thing)
.50 to .90	Redundant (variables might be measuring the same thing)
1.00	Perfect Relationship

Using the Table of Levels of Association, we see that this is a very weak prediction. We can also interpret lambda by multiplying the value by 100 and getting a percentage, which means that by using the respondent's gender identification (SEX) to predict views on immigration (IMMAMECO_DICHOTOMOUS), we have reduced our error of prediction by 3.8% (a weak predictor).

Gamma (Y)

Gamma (Y) is a measure of association for ordinal variables (or also dichotomous variables). Let's see how strongly education (EDUC_RECODED) shapes views on immigration (IMMCULT).

First, recode EDUC into three categories (High School or less: 0-12 years of education, College or some college: 13-16 years of education, and Graduate education: 17+ years of education).

Now let's run the analysis.

GAMMA

INSTRUCTIONS: "Analyze" → "Descriptive Statistics" → "Crosstabs"

- For "Row", select the dependent variable: IMMCULT
- For "Column", select the independent variable: EDUC_RECODED

→ Click on "Cells"

- Under "Percentages", select: "Column" → click "continue"

→ Click on "Statistics"

- Under "Ordinal" and select: "Gamma"

→ Click "Continue" → Click "Ok"

The value of gamma is: .245. This is a positive number, which means that the association is positive (as one value increases the other value increases as well). That is, as education increases, the view that immigrants undermine American culture increases as well, but note that the way the variable is coded means people tend to agree less with that statement. This means that as education increases, people tend to have more positive views about immigration. In addition, if we look at the Table of Strength of Association, this is a moderate association.

SUMMARY: In this lab you learned how to obtain and interpret lambda and gamma values depending on the level of measurement of the variables.

LAB #15 PRACTICE PROBLEMS. Use the 2014 GSS dataset and answer these questions:

Determine if views of the bible (BIBLE) shapes views on immigration (IMMAMECO_DICHOTOMOUS).

Q1. What is the value of lambda (λ)? 0.012

Q2. By using the respondent's view of the bible (BIBLE) to predict views on immigration (IMMAMECO_DICHOTOMOUS), we have reduced our error of prediction by: 1.2%

Q3. How strong is the prediction? Weak

Now determine if education (EDUC_RECODED) shapes views on immigration (IMMCRIME).

Q4. What is the value of gamma (γ)? 0.213

Q3. How strong is the association? Moderate

Consider: In what ways would views of the bible and years of education could shape views on immigration?

Lab #16: Regression and Correlation

Most U.S.-Americans (65%) support creating a path for undocumented immigrants in the U.S. to gain U.S. citizenship over time (<https://news.gallup.com/poll/184577/favor-path-citizenship-illegal-immigrants.aspx>)



Reflection Question: Given this strong support to create a path for undocumented immigrants to legalize their status in the U.S. and eventually become U.S. citizens, why do you think unauthorized migration remains a very contentious and divisive topic in politics?

Lab #15 Learning Objectives:

- Learn how to graph two interval-ratio variables
- Learn how to obtain and interpret r and r^2
- Learn how to draw a regression line

Concepts you should review before you do this lab:

- Dependent and Independent variable
- Level of measurement: nominal, ordinal, interval-ratio
- Regression line
- Correlation (r)
- Coefficient of determination (r^2)

Correlation (r)

Do you think that education shapes our views on immigration? How much?

Correlation is the measure of association between two interval-level variables. But before we do this, we need to make sure our data does not have any outliers. Running a histogram of the variables in question will show the spread of the data (both graphically and numerically).

Let's run a histogram on education (EDUC) and the index we created about views on immigration (IMMVIEWES_INDEX).

The spread of EDUC should show the data from 0 to 20 (years of education), with a mean of 13.69 and a standard deviation of 3.08.

The spread of IMMVIEWES_INDEX should show the data from 0 to 8, with a mean of 5.38 and a standard deviation of 2.21.

Then you can obtain r:

PEARSON's R:

INSTRUCTIONS: "Analyze" → "Correlate" → "Bivariate"

- For "Variables", select the two variables: EDUC and IMMVIEWS_INDEX
- Under "Correlation Coefficients", select "Pearson"

→Click "Ok"

The output will be a matrix. Notice that it will correlate EDUC and EDUC, which is a perfect 1, and IMMVIEWS_INDEX and IMMVIEWS_INDEX which is also a perfect 1.

But the correlation of EDUC and IMMVIEWS_INDEX is: ____0.097____

Using the table of strength of association: we can conclude that this is a very weak association, and not significant at a 0.05 level (since $P(0.07) > \alpha(0.05)$).

Coefficient of Determination (r^2)

We can also draw a scatter plot to obtain the coefficient of determination (r^2):

COEFFICIENT OF DETERMINATION:

INSTRUCTIONS: "Graph" → "Legacy Dialogs" → "Scatter/Dot"

→ Select "Simple Scatter," Click "Define"

- For "Y-axis", select the dependent variables: IMMVIEWS_INDEX
- For "X-axis", select the independent variables: EDUC

→Click "Ok"

If you double click on the graph, you should get a pop-up window with just the graph, then select "Elements" → "Fit Line at Total"

This will add a line and also give you the coefficient of determination (r^2) ____0.009____, which means that 0.9% of the variation in the dependent variable (views on immigration, IMMVIEWS_INDEX) can be accounted by the variation in the independent variable (education, EDUC).

In addition, note that the graph has also given us a regression equation: $Y = A + B(X)$, where A is the intercept, and B is the slope.

$Y = 4.5 + 0.06(X)$; this means that when a person has no education (0), the predicted value of the index is 4.5 (out of 8 possible points), and for every unit of education increase (that is for every year of education increase), there is a .06 units in the immigration view index increase.

We can also use this equation to predict the index value for someone with 12 or 16 years of education.

SUMMARY: In this lab, you learned how to graph two interval-ratio variables, how to obtain and interpret r and r^2 , and how to draw a regression line

LAB #16 PRACTICE PROBLEMS. Use the 2014 GSS and answer these questions:

Correlate age (AGE) and views on immigration (IMMVIEWES_INDEX):

Q1. What is the value of r ? _____ - .155 _____

Q2. Is the correlation positive or negative? _____ negative _____

Q3. How strong is the correlation? _____ very weak _____

Now graph the data:

Q4. What percentage of the variation in age explains variation in views on immigration (r^2)?
_____ 2.4% _____

Q5. Are older people more or less likely to have favorable views on immigration? _____ less favorable views _____

Consider: Why do you think age shapes how people see immigration? How come?

Lab #17: Final Project

INSTRUCTIONS: Open GSS 2016 in SPSS, then *answer the following 10 questions, and please use a different text color (from black) when typing your answers.*

1. As a group, determine a topic of interest (example: race, immigration, gender, gun control, trust of government, etc.).

TOPIC OF INTEREST: _____

Write a research question that links two variables, and that is related to the group's topic of interests.

RESEARCH QUESTION: _____

2. List the dependent and independent variables (make sure you can find the independent and dependent variables on the same GSS panel. You might want to use a previous GSS version).

GSS version/year: _____

INDEPENDENT VARIABLE NAME: _____

What does it measure? _____

Level of measurement: _____

DEPENDENT VARIABLE NAME: _____

What does it measure? _____

Level of measurement: _____

3. Run a frequency distribution table for both variables and place the tables here:

FREQUENCY DISTRIBUTION TABLE FOR INDEPENDENT VARIABLE:

FREQUENCY DISTRIBUTION TABLE FOR DEPENDENT VARIABLE:

4. Write two short paragraphs describing the variable (use the appropriate measures of center and dispersion to describe each variable—mode, median, mean, standard deviation).

DESCRIPTIVE SUMMARY OF INDEPENDENT VARIABLE:

DESCRIPTIVE SUMMARY OF DEPENDENT VARIABLE:

5. Write a hypothesis linking one of the dependent and one of the independent variables (how do you think they are connected?)

HYPOTHESIS ABOUT THE RELATIONSHIP BETWEEN DEPENDENT AND INDEPENDENT VARIABLES:

6. Run a 2x2 crosstab of dependent and independent variables, with column percentages (you might have to recode variables).

BIVARIATE TABLE:

7. Do you see an association between the dependent and independent variable? Explain (using specific percent differences).

EXPLAIN ASSOCIATION:

8. Run a chi-square test and test the hypothesis that your independent and dependent variables are associated at a .01 significant level.

CHI-SQUARE TABLE:

- i. H_0 : _____
 - ii. H_1 : _____
 - iii. What distribution should you use for this hypothesis test?
 - iv. Why did you choose this distribution?
 - v. χ^2 =
 - vi. p-value=
 - vii. Compare P and alpha
 - viii. Reach your decision
9. Explain in your own words the findings from the chi-square hypothesis test.

EXPLAIN ASSOCIATION:

10. Report Cramer's V's: _____
 Strength of this association: _____
 Interpret strength of association:

11. (Extra-credit) Produce a scatterplot with a regression line for two interval-ratio level variables (one of your dependent and one of your independent variables related to the topic of interest). Copy and paste the graph here:
- a. Produce a correlation matrix with Pearson's r of the two variables. Exclude cases pairwise. Copy and paste the matrix here:
 - b. Describe the strength of the relationship between the two variables by noting what the value of Pearson's r is, and explaining whether the relationship is weak, moderate, or strong:
 - c. Calculate the coefficient of determination (r^2) and interpret:

SUBMIT YOUR FINAL REPORT ON CANVAS.

INDEX: SUMMARY OF SPSS PROCEDURES
--

LAB 1: INTRODUCTION TO VARIABLES.....	3
EXPLORE THE DATA ON GSS WEBSITE.....	4
LAB 2: DESCRIBING THE DATA USING TABLES	6
OPEN 2014 GSS ON SPSS.....	6
SETTING OPTIONS	7
FREQUENCY DISTRIBUTION TABLE.....	7
LAB 3: DESCRIBING THE DATA USING GRAPHS.....	9
PIE CHART	9
BAR CHART	10
HISTOGRAM.....	10
LINE CHART	11
LAB 4: REPORTING MEASURES OF CENTRAL TENDENCY	12
FREQUENCY PROCEDURE TO GET MEASURES OF CENTRAL TENDENCY (IR).....	13
DESCRIPTIVES PROCEDURE TO GET MEASURES OF CENTRAL TENDENCY (ORDINAL).....	13
FREQUENCY PROCEDURE TO GET MEASURES OF CENTRAL TENDENCY (ORDINAL).....	14
DESCRIPTIVES PROCEDURE TO GET MEASURES OF CENTRAL TENDENCY (IR).....	14
LAB 5: REPORTING MEASURES OF DISPERSION.....	15
FREQUENCY PROCEDURE TO GET MEASURES OF DISPERSION.....	15
DESCRIPTIVES PROCEDURE TO GET MEASURES OF DISPERSION	16
LAB 6: RECODING VARIABLES.....	17
RECODING A VARIABLE.....	18
LAB 7: CREATING DICHOTOMOUS VARIABLES.....	20
RECODING A VARIABLE.....	18
LAB 8: CREATING AN INDEX.....	23
CREATING AN INDEX	24
LAB 9: CONFIDENCE INTERVALS.....	26
TURN ON WEIGHTS.....	26
CONFIDENCE INTERVALS	27
LAB 10: ONE-SAMPLE HYPOTHESIS TESTING	28
ONE SAMPLE T TEST FOR PROPORTIONS	29
ONE SAMPLE T TEST FOR MEANS	29
LAB 11: TWO-SAMPLES HYPOTHESIS TESTING	31
INDEPENDENT SAMPLES T-TEST FOR MEANS	31

LAB 12: ANOVA	33
ANOVA	33
LAB 13: CROSSTABS	35
BIVARIATE TABLE.....	36
LAB 14: CHI-SQUARE AND CRAMER'S V	38
CHI-SQUARE AND CRAMER'S V	39
LAB 15: LAMBDA AND GAMMA	41
LAMBDA	41
GAMMA	42
LAB 16: REGRESSION AND CORRELATION	44
PEARSON'S R	45
COEFFICIENT OF DETERMINATION.....	45