

QUANTITATIVE REASONING MODULE ON SOCIAL STRATIFICATION

Maria Aysa-Lastra
Winthrop University

Instructional Objective: This module contains five activities aimed at increasing quantitative reasoning abilities of students taking lower and upper level division courses in sociology. The activities introduce students to existing data sources, basic data analysis and interpretation. Activities and their corresponding assessments are described below, and can be modified for different topics/courses or time periods depending on data availability. The activities in this module can be taught independently in introductory or higher level courses in the social sciences. The module core topic as designed in this guide is social stratification with emphasis on the description of the American contemporary middle class.

Description: The back bone of a healthy society is a sizable and racially diverse middle class. The increasing diversity of the US population requires us to evaluate the participation of minority groups in the middle class, and identify the variables that explain their unequal racial and ethnic distribution and citizenship status in this important segment of our society.

Learning Objectives:

After completing the activities in this module students will be able to:

1. Explain the importance of producing, using and interpreting data
2. Explain why and how the US government collect data on the American population
3. Identify important sources of data: ACS, CPS, BRFSS, etc.
4. Gain exposure to geographical patterns based on data collected by the government
5. Design and interpret a frequency table
6. Distinguish a categorical and continuous variable
7. Identify an independent and a dependent variable
8. Design and interpret a cross tabulation
9. Create graphs and figures
10. Developing a hypothesis
11. Identify a control variable
12. Coding and creating new categories
13. Identify different analytical strategies
14. Design and interpret analysis of the difference of means
15. Interpret and critique tables designed by colleagues and other researchers.

In the following pages you will find:

Outline of Activity 1: Learning about the importance of producing quantitative data for the US

Outline of Activity 2: Univariate analysis and interpretation using Pivot tables

Outline of Activity 3: Bivariate analysis and interpretation using Pivot tables

Outline of Activity 4: Multivariate analysis and interpretation using Pivot tables

Outline of Activity 5: Interpreting and critiquing estimations

Instructional Notes (these will be modified after the modules are taught for the first time)

Resource list/guide (TBD)

Assessment (included at the end of each activity)

Bibliography (TBD)

Activity 1: Why does the government collect data? Why should I care about it?

Time: one class session (40 to 50 min.)

Level: Lower and upper level courses

Using a short, clear and FUN presentation introduce students to the importance of collecting and most importantly how to analyze and correctly interpret data. Explain the relative advantage of mastering this skill (vs. qualitative data), and list some of their potential future employers. Use NYT article on employment of data analyst.

The objective of the presentation is to introduce students to various sources of data available including the data they will be using in the activities included in the next modules. It contains a basic presentation on ACS, its importance, groups of variables, sampling, and geographical divisions. After introducing students to ACS, I will share with them the ESRI website, as an example of how it is used. A final slide will include a list of other important national level surveys: CPS and BRFFS and how are they used.

At the end of this activity students should be familiar with secondary sources as well as main variables from ACS. A short pre/post quiz will be distributed to students, as well as feedback on what do they remember from this activity at the end of the lecture and feedback on how to improve it.

This activity covers learning objectives 1 to 4 and its objective is to introduce students to sources of quantitative data and the importance of its analysis in the United States. Alternatively, sources for other countries can also be introduced if the course contents calls for it.

Assessment

Pre/Post 2 multiple choice question quiz.

Notes: This presentation is being developed based on the information provided during the workshop.

Activity 2: Analyzing and interpreting data

Time: 2 class sessions and homework

Level: Lower level courses

Before the class meets for this activity ask your students to bring their computers or schedule the class in a computer lab. This is a hands on activity designed for groups with less than 30 students.

At the end of this activity students will:

1. Identify categorical and continuous variables
2. Interpret codebooks, codes
3. Learn how to use Pivot tables in Excel.
4. Design and interpret a frequency table
5. Estimate and interpret and average
6. Create and interpret a graph

Instructions: Follow the next steps:

Download the file ACS-act2.xls from Blackboard. This files contains data from a random sample (n=10,000) of old full time year round employees (those employed at the time of the survey and who worked for more than 40 weeks during 2012) between the ages of 25 and 64 from the US PUMS 1-year file of the ACS 2013. Therefore, the unit of analysis are these individuals.

Each row corresponds to a person answering the ACS questionnaire. Each column corresponds to a variable. A variable is a characteristic being measured for all individuals in the sample. There are five variables in the file. One variable that identifies each respondent (SERIALNO) and other 2 variables measuring education, geographical location and relative income:

CATEGORICAL VARIABLES:

A categorical variable is a specific characteristic (sex) that identifies how different attributes (male, female) differ among individuals observed.

EDUC: Educational attainment

3 = Less than high school

2 = More than high school but less than college

1 = College graduates and more than college

DIVISION: Census Statistical Divisions (US Region) – Include MAP link

0=Puerto Rico

1=New England (Northeast region)

2=Middle Atlantic (Northeast region)

3=East North Central (Midwest region)

4=West North Central (Midwest region)

5=South Atlantic (South region)

6=East South Central (South region)

7=West South Central (South Region)

8=Mountain (West region)

9=Pacific (West region)

CONTINUOUS VARIABLES:

A continuous variable can take on any value within a range of units of the same length. For example, height is measured in inches, and each inch is always of the same length.

POVPIP: Income to poverty ratio (income/poverty line) – capped at a value of 501

What you just read is an example of a CODEBOOK or an instrument that links the name of a variable in a file and its coding scheme with meaning you will need to interpret the results.

Now that you know about the contents of the file and its structure you can:

- Open the file in Excel.
- Select the data, by clicking in the upper right corner of the spread sheet.
- Click the insert tab and then click on pivot tables, click OK and a new worksheet with the pivot tables prompts will appear
- Select EDUC and move it to ROW labels box in the lower left corner of your screen.
- Select DIVISION and move it to the COLUMN labels box
- Highlight/Select EDUC and drag it to the box Σ (upper-case sigma, which stands for “sum of”) Values. Click on “Count of EDUC” and then on “Value Field Settings” in the tab “Summarize Values By” you should select “Count”. Click “OK”.

COPY THE FOLLOWING TABLE 1 IN A WORD DOCUMENT and Select a DIVISION. Do the following:

FOR THE CATEGORICAL VARIABLE: Level of Education (EDUC)

- Click on “Count of EDUC” and then on “Value Field Settings” in the tab “Summarize Values By” you should select “Count”. Click “OK”. Copy the Results for your selected division
- Click again on “Count of EDUC” and then on “Value Field Settings” in the tab “Summarize Values By” you should select “Count”, in the tab “Show Values As” scroll down and select “% of Column Total”. Click “OK”. Copy the Results for your selected division
- Click once more on “Count of EDUC” and then on “Value Field Settings” in the tab “Summarize Values By” you should select “Count”, in the tab “Show Values As” scroll down and select “% of Row Total”. Click “OK”. Copy the Results for your selected division.
- Write interpretation of your results. Only interpret the RELATIVE estimates. Relative estimates refer to the percentage distributions of the data, they are relative to the size and distribution of the cases for each category of the variables EDUC and DIVISION in the population.

Table 1. Distribution of Levels of Education [your unit of analysis] in [your selected Division]

Variable	Counts	Percentage Columns	Percentage Rows
Level of Education			
Label for EDUC (1)			
Label for EDUC (2)			
Label for EDUC (3)			
TOTAL	1,000	100%	10%

Source: Class sample from ACS 2013 1-yr estimates.

FOR THE CONTINUOUS VARIABLE: Income to poverty ratio

- Remove EDUC from the box of “Row Labels”. Click on the variable and click on “Remove Field”.
- Make sure that DIVISION is in the “Column Labels” box
- Remove “Count of EDUC” from the box of “ Σ Values”. Click on the variable and click on “Remove Field”.

- Highlight/Select POVPIP and drag it to the box “ Σ Values”. Click on “Count of POVPIP” and then on “Value Field Settings” in the tab “Summarize Values By” you should select “Average”. Click “OK”.
- Select the average income to poverty ratio for the DIVISION you selected: _____ (write it here) and interpret your estimate.

GRAPHS:

Now that you are done with your table, you will create a graph using the same results. Go back to Excel and create the graph. Following the next steps.

- Drag EDUC to the box “Row Labels” and in the box “ Σ Values”.
- Click again on “Count of EDUC” and then on “Value Field Settings” in the tab “Summarize Values By” you should select “Count”, in the tab “Show Values As” scroll down and select “% of Column Total”. Click “OK”.
- Click on “PivotTable Tools” at the top of your screen. Then click on “Pivot Chart”. Select a 2 dimensional pie graph (NO 3D please, remember LESS IS MORE). Click “OK”.
- CLICK on “DIVISION” within the graph area. Take off the check mark from “(Select All)” and click ONLY on your selected Division.
- Double click on the number above the graph and write the title of your table.
- Copy and Paste this graph in your word document.

REMINDERS:

Using the information about the attributes of this variable (1= Less than high school; 2 = High school graduates; 3 = College graduates and 4 = More than college) for your selected DIVISION. Create in a Word file a frequency table containing absolute counts and percentages (also called relative distribution).

AN ABSOLUTE COUNT is the number of observations in a given category

A PERCENTAGE indicates the relative importance of each category in the population. It is calculated by dividing the absolute count in each category by the total number of observations in the population.

Format is always important, please make sure that you include all the components of a table:

Title: What is this table about?

Labels for rows and columns: DO NOT USE THE NUMERICAL CATEGORIES IN THE PIVOT TABLE. REPLACE THOSE WITH THE LABELS FOR EACH CATEGORY (1=Less than high school; 2=High school graduates; 3=College graduates and 4=More than college).

Source:

Formatting: Use only horizontal lines as in the template provided above (LESS IS MORE!!!)

A figure (or graph) has the following elements:

Title:

Labels for the categories: as long as you describe the labels in your interpretation of the table, you do not need to include the labels for the categories in the graph.

Source:

If you have more questions about pivot tables in Excel you can go to:

<http://www.timeatlas.com/excel-pivot-tables/>

<http://www.wikihow.com/Create-Pivot-Tables-in-Excel>

<https://www.youtube.com/watch?v=9NUjHBNWe9M>

HOMEWORK: After participating in the class discussion you are ready to write the interpretation of results. Continue working on the same word file in which you have your table and graph. Below each of them write your interpretations. REQUIRED: Bring TWO copies of your work next class, do not include your name, as ID use only the first 2 and last 3 digits of your WID number.

Peer Review

Students will exchange copies of their homework during the second class session. Each student will review its colleague work according to the following rubric.

Checklist for Peer Review

- *Clear title for table
- *Clear labels for rows and columns
- *Source cited
- *Format (LESS IS MORE-only few horizontal lines)
- *Clear and correct interpretation of percentage distribution and counts.
- *Write objective and constructive comments for your colleague, including your reasons for the correct or incorrect interpretation of the distributions shown in the table and graph

Assessment

Pre/Post 2 multiple choice question quiz.

Question for learning objective 1: Types of Variables

Race is a continuous variable T/F

Income is a continuous variable T/F

Question for learning objective 2: Based on the frequency table

TBD

In the post test add the following questions:

What is the most important concept/skill that you learned after completing the activity?

Please let me know how you would change this assignment to improve your learning experience.

Instructional Notes:

Use the table and graph you have created and present to the students the correct interpretation of the data, and the differences observed by region. Emphasize that this is an example of UNIVARIATE analysis, or the analysis of one variable at a time by DIVISION.

Activity 3: Analyzing and interpreting the relation between two variables

Time: 2 class sessions

Level: Upper level

At the end of this activity students will:

1. Identify between independent and a dependent variables
2. Learn how to use Pivot tables in Excel.
3. Design and interpret cross tabulations
4. Create graphs and figures
5. Identify a control variable
6. Discuss how college graduates, those earning more than national median income and those in white collar occupations differ from all others.

Instructions: Follow the next steps:

Download the file ACS-act3.xls from Blackboard. This files contains data from a random sample (n=10,000) of old full time year round employees (those employed at the time of the survey and who worked for more than 40 weeks during 2012) between the ages of 25 and 64 from the US PUMS 1-year file of the ACS 2013. Each row corresponds to a person answering the ACS questionnaire. Each column corresponds to a variable. A variable is a characteristic being measured for all individuals in the sample. There are five variables in the file. One variable that identifies each respondent (SERIALNO) and other 4 variables measuring education, gender, geographical location and relative income:

EDUC: Level of educational attainment

3 = Less than high school

2 = More than high school but less than college

1 = College graduates and more than college

SEX: Gender

1=Males

2=Females

DIVISION: Census Statistical Divisions (US Region) – Include MAP link

0=Puerto Rico

1=New England (Northeast region)

2=Middle Atlantic (Northeast region)

3=East North Central (Midwest region)

4=West North Central (Midwest region)

5=South Atlantic (South region)

6=East South Central (South region)

7=West South Central (South Region)

8=Mountain (West region)

9=Pacific (West region)

POVPIP: Income to poverty ratio (income/poverty line) – capped at a value of 501

Follow these steps to analyze the relation between two variables, or what we called bivariate analysis.

INDEPENDENT VARIABLE:

It is a causal variable, or a characteristic that has influence or is related to another variable. For example: Do you think that on average levels of education differ by gender because women and men have different access to schools or study habits? In this case sex is the independent variable.

DEPENDENT VARIABLE:

It is the result variable, or the variable that has been affected by the independent variable. For example: Do you think that gender discrimination leads to lower incomes on average for women. In this case, income to poverty ration is our dependent variable, and sex is the independent variable.

In this activity we will analyze two bivariate relations. First, between sex and education. Second, sex and relative income.

RELATION BETWEEN SEX AND EDUCATION

WRITE THE NAME OF THE INDEPENDENT VARIABLE: _____

Is this variable categorical or continuous? _____

WRITE THE NAME DEPENDENT VARIABLE: _____

Is this variable categorical or continuous? _____

SELECT A GEOGRAPHICAL DIVISION: _____

EXPLAIN HOW THESE VARIABLES ARE RELATED:

Now that you know about the contents of the file, its structure, you have identified the variables you can build a cross tabulation using pivot tables:

- Open the file in Excel.
- Select the data, by clicking in the upper right corner of the spread sheet.
- Click the insert tab and then click on pivot tables, click OK and a new worksheet with the pivot tables prompts will appear
- Select DIVISION and move it to "Row Labels" box in the lower left corner of your screen.
- Select EDUC and move it to the "Row Labels" box (make sure EDUC is below DIVISION).
- Select SEX and move it to the "Column Labels" box.

In this cross tabulation our independent variable is in the column and our dependent variable is in the row, therefore, we should look at the distribution of the dependent variable for each category of the independent variable. To estimate the cross tab do the following:

- Highlight/Select EDUC and drag it to the box Σ (upper-case sigma, which stands for "sum of") Values. Click on "Count of EDUC" and then on "Value Field Settings" in the tab "Summarize Values By" you should select "Count". In the tab "Show Values As" scroll down and select "% of Parent Row Total". Click "OK". Copy the Results for your selected division.

COPY THE FOLLOWING TABLE 2A IN A WORD DOCUMENT and fill in the results for your selected DIVISION.

Table 2A. Distribution of Levels of Education by SEX [your unit of analysis] in [your selected Division]

Variable	Males	Females	Total
Level of Education			
Label for EDUC (1)			
Label for EDUC (2)			
Label for EDUC (3)			
TOTAL	100%	100%	100%

Source: Class sample from ACS 2013 1-yr estimates.

Now that you are done with your table, you will create a graph using the same results. Go back to Excel and create the graph. Following the next steps.

- Click on "PivotTable Tools" at the top of your screen. Then click on "Pivot Chart". Select a 2 dimensional stacked column graph, (NO 3D please, remember LESS IS MORE). Click "OK".
- CLICK on "DIVISION" within the graph area. Take off the check mark from "(Select All)" and click ONLY on your selected Division.

Copy and Paste this graph in your word document. Remember that in your explanation you must refer to 1 as males and 2 as females.

RELATION BETWEEN SEX AND RELATIVE INCOME

WRITE THE NAME OF THE INDEPENDENT VARIABLE: _____

Is this variable categorical or continuous? _____

WRITE THE NAME DEPENDENT VARIABLE: _____

Is this variable categorical or continuous? _____

SELECT A GEOGRAPHICAL DIVISION: _____

EXPLAIN HOW THESE VARIABLES ARE RELATED:

Now that you know about the contents of the file, its structure, you have identified the variables you can build a cross tabulation using pivot tables:

- Open the file in Excel.
- Select the data, by clicking in the upper right corner of the spread sheet.
- Click the insert tab and then click on pivot tables, click OK and a new worksheet with the pivot tables prompts will appear
- Select DIVISION and move it to "Row Labels" box in the lower left corner of your screen.
- Select SEX and move it to the "Column Labels" box.

- Highlight/Select EDUC and drag it to the box “Σ Values”. Click on “Count of POVPIP” and then on “Value Field Settings” in the tab “Summarize Values By” you should select “Average”. Click “OK”.

COPY THE FOLLOWING TABLE 2B IN A WORD DOCUMENT and fill in the results for your selected DIVISION.

Table 2A. Distribution of Levels of Education by SEX [your unit of analysis] in [your selected Division]

Variable	Males	Females	Total
Relative Income to Poverty Ratio Average			

Source: Class sample from ACS 2013 1-yr estimates.

Now that you are done with your table, you will create a graph using the same results. Go back to Excel and create the graph. Following the next steps.

- Click on “PivotTable Tools” at the top of your screen. Then click on “Pivot Chart”. Select a 2 dimensional pie column, (NO 3D please, remember LESS IS MORE). Click “OK”.
- CLICK on “DIVISION” within the graph area. Take off the check mark from “(Select All)” and click ONLY on your selected Division.
- Copy and Paste this graph in your word document. Remember that in your explanation you must refer to 1 as males and 2 as females.

REMINDERS:

Use the labels in the codebook provided in this assignment to correctly label the categories of your variables. Include your cross tab, the stacked column graph, the table for the difference of averages and the bar chart in a word file. Include your interpretations for all tables and graphs.

Format is always important, please make sure that you include all the components of a table:

Title: What is this table about?

Labels for rows and columns: DO NOT USE THE NUMERICAL CATEGORIES IN THE PIVOT TABLE. REPLACE THOSE WITH THE LABELS FOR EACH CATEGORY (See codebook included in this assignment).

Source:

Formatting: Use only horizontal lines as in the template provided above (LESS IS MORE!!!)

A figure (or graph) has the following elements:

Title:

Labels for the categories: as long as you describe the labels in your interpretation of the table, you do not need to include the labels for the categories in the graph.

Source:

If you have more questions about pivot tables in Excel you can go to:

<http://www.timeatlas.com/excel-pivot-tables/>

<http://www.wikihow.com/Create-Pivot-Tables-in-Excel>

<https://www.youtube.com/watch?v=9NUjHBNWe9M>

HOMEWORK: After participating in the class discussion you are ready to write the interpretation of results. Continue working on the same word file in which you have your tables and graphs. Below each of them write your interpretations. REQUIRED: Bring TWO copies of your work next class, do not include your name, as ID use only the first 2 and last 3 digits of your WID number.

Peer Review

Students will exchange copies of their homework during the second class session. Each student will review its colleague work according to the following rubric.

Checklist for Peer Review

- *Clear title for table
- *Clear labels for rows and columns
- *Source cited
- *Format (LESS IS MORE-only few horizontal lines)
- *Clear and correct interpretation of percentage distribution and counts.
- *Write objective and constructive comments for your colleague, including your reasons for the correct or incorrect interpretation of the distributions shown in the table and graph

Assessment

Pre/Post multiple choice question quiz.

Question for each learning objective

In the post test add the following questions:

What is the most important concept/skill that you learned after completing the activity?

Please let me know how you would change this assignment to improve your learning experience.

Instructional Notes:

Use the tables and graphs you have created and present to the students the correct interpretation of the data, and the differences observed by division. Emphasize that this is an example of BIVARIATE analysis, or the analysis of the relation between two variables at a time by DIVISION.

Activity 4: Analyzing and interpreting the relation between two variables, controlling for a third variable and recoding.

Time: 2 class sessions

Level: Upper level

At the end of this activity students will:

1. Identify independent, dependent and control variables
2. Distinguish between categorical and continuous variables
3. Coding, recoding and creating new variables
4. Develop a hypothesis
5. Learn how to use Pivot tables in Excel.
6. Design and interpret tables and figures

Instructions: Follow the next steps:

Download the file ACS-act4.xls from Blackboard. This file contains data from a random sample (n=10,000) of old full time year round employees (those employed at the time of the survey and who worked for more than 40 weeks during 2012) between the ages of 25 and 64 from the US PUMS 1-year file of the ACS 2013. Each row corresponds to a person answering the ACS questionnaire. Each column corresponds to a variable. A variable is a characteristic being measured for all individuals in the sample. There are seven variables in the file. One variable that identifies each respondent (SERIALNO) and other 6 variables measuring education, gender, geographical division and relative income, level of relative income and a variable combining race/ethnicity, nativity and citizenship.

SEX: Gender

1=Males

2=Females

DIVISION: Census Statistical Divisions (US Region) – Include MAP link

0=Puerto Rico

1=New England (Northeast region)

2=Middle Atlantic (Northeast region)

3=East North Central (Midwest region)

4=West North Central (Midwest region)

5=South Atlantic (South region)

6=East South Central (South region)

7=West South Central (South Region)

8=Mountain (West region)

9=Pacific (West region)

POVPIP: Income to poverty ratio (income/poverty line) – capped at a value of 501

The variables listed above were obtained directly from the ACS 2013 1-year sample. However, data analysts usually need to recode variables to convert them from continuous to categorical, combine categories or combine different variables to create new variables.

For this exercise I created two new variables: levels of education and a variable that combines race, nativity and citizenship.

Converting a continuous variable to a categorical variable.

I wrote this program to create the new categorical variable (EDUC – levels of education) from the original variable SCHL (educational attainment):

Here is the ACS 2013 codebook for the original variable

```
SCHL
Educational attainment
bb .N/A (less than 3 years old)
01 .No schooling completed
02 .Nursery school, preschool
03 .Kindergarten
04 .Grade 1
05 .Grade 2
06 .Grade 3
07 .Grade 4
08 .Grade 5
09 .Grade 6
10 .Grade 7
11 .Grade 8
12 .Grade 9
13 .Grade 10
14 .Grade 11
15 .12th grade - no diploma
16 .Regular high school diploma
17 .GED or alternative credential
18 .Some college, but less than 1 year
19 .1 or more years of college credit, no degree
20 .Associate's degree
21 .Bachelor's degree
22 .Master's degree
23 .Professional degree beyond a bachelor's degree
24 .Doctorate degree
```

I wrote the following program to create the new variable

```
educ=0;
```

In this line I create a new variable (EDUC), with values of zero for all cases.

```
if schl<16 then educ=3;
```

In this line I am telling the program to take as the base variable SCHL (original variable) and code my newly created variable EDUC with the value of three, if the respondent's educational attainment was below a regular high school diploma (16).

```
if 15<schl<21 then educ=2;
```

Here I am asking the program to code EDUC with the value of two, if the respondent's educational attainment was regular high school diploma and any other similar or higher level credentials such as Associate's degree, but less than Bachelor's degree.

```
if 20<schl then educ=1;
```

Here I am asking the program to code EDUC with the value of one, if the respondent's educational attainment is college graduate or any degree that required graduate education.

Combining categorical variables to create a new variable (RACE).

I combined the variables rac1p (race), hisp (ethnicity) and cit (nativity and citizenship)

Codes for the original variables

RAC1P

Recoded detailed race code

- 1 .White alone
- 2 .Black or African American alone
- 3 .American Indian alone
- 4 .Alaska Native alone
- 5 .American Indian and Alaska Native tribes specified; or American Indian or Alaska Native, not specified and no other races
- 6 .Asian alone
- 7 .Native Hawaiian and Other Pacific Islander alone
- 8 .Some Other Race alone
- 9 .Two or More Races

HISP

Recoded detailed Hispanic origin

- | | |
|---------------------------------|---------------------------------------|
| 01 .Not Spanish/Hispanic/Latino | 13 .Argentinean |
| 02 .Mexican | 14 .Bolivian |
| 03 .Puerto Rican | 15 .Chilean |
| 04 .Cuban | 16 .Colombian |
| 05 .Dominican | 17 .Ecuadorian |
| 06 .Costa Rican | 18 .Paraguayan |
| 07 .Guatemalan | 19 .Peruvian |
| 08 .Honduran | 20 .Uruguayan |
| 09 .Nicaraguan | 21 .Venezuelan |
| 10 .Panamanian | 22 .Other South American |
| 11 .Salvadoran | 23 .Spaniard |
| 12 .Other Central American | 24 .All Other Spanish/Hispanic/Latino |

CIT

Citizenship status

- 1 .Born in the U.S.
- 2 .Born in Puerto Rico, Guam, the U.S. Virgin Islands, or the Northern Marianas
- 3 .Born abroad of American parent(s)
- 4 .U.S. citizen by naturalization
- 5 .Not a citizen of the U.S.

I wrote this program to create the new categorical variable (RACE):

```
race=0;
if hisp=1 and rac1p=1 then race=1; /* Non-Hispanic White */
if hisp=1 and rac1p=2 then race=2; /* Non-Hispanic Black */
if hisp=1 and rac1p>2 then race=3; /* Non-Hispanic other minority */
if hisp>1 and cit in (1,2,3) then race=4; /* Hispanic born in the US */
if hisp>1 and cit=4 then race=5; /* Hispanic naturalized citizen */
if hisp>1 and cit=5 then race=6; /* Hispanic non-citizen */
```

Read in detail the codes of the original variable and the program. Write your comments about any problems you find with this program (or the assumptions behind it) and the new variable RACE:

As you have noticed, the file contains both categorical and continuous variables. The analytical strategies researchers pursue depend on the types of variables in the analysis. You will learn more about this in SOCL 316 and SOCL 516. Here is a general guideline:

2 Categorical variables: contingency tables, also called cross-tabs. The categories of one variable are placed in the rows and the categories of the other variable in the columns.

1 Categorical and 1 Continuous variable: differences of means. You will calculate the mean of the continuous variable for each of groups defined by the categorical variable.

2 Continuous variables: correlation coefficient. It is a summary measure that indicates if the relation between the variables is positive (when one increases, the other increases) or negative (when one increases, the other decreases).

Now that you know about all variables in the analysis and the appropriate analytic strategies, follow the next steps to analyze the relation between a dependent, and independent and a control variable. Remember to also select one geographical division.

DIVISION SELECTED _____

For the analysis you know that:

DEPENDENT VARIABLE: income to poverty ratio (POVPI)

Type of Variable: _____

INDEPENDENT VARIABLE: Level of Education (EDUC)

Type of Variable: _____

Appropriate analytical strategy: _____

CONTROL VARIABLE: Racial/Ethnic and Citizenship status (RACE)

Type of Variable: _____

A control variable is a characteristic that influences the relation between the dependent and the independent variables. If it is held constant the researcher can observe the relation between the dependent and the independent variable without the effect of the control variable. For example, the relation between education and income is different across racial and ethnic groups, therefore, to observe the relation between education and relative income net of the differentials by race, we need to control by race.

The hypothesis then is: education is positively related to relative income after controlling for race.

Now that you know about the contents of the file, its structure, the appropriate analytic strategy, and the hypothesis you will design a table to analyze the relation between education, relative income and race using pivot tables:

- Open the file in Excel.
- Select the data, by clicking in the upper right corner of the spread sheet.

- Click the insert tab and then click on pivot tables, click OK and a new worksheet with the pivot tables prompts will appear
- Select DIVISION and move it to “Row Labels” box in the lower left corner of your screen.
- Select RACE and move it to the “Row Labels” box (make sure RACE is below DIVISION).
- Select EDUC and move it to the “Column Labels” box (make sure RACE is below DIVISION).

To estimate the means of POVPIP for each EDUC*RACE group for your selected DIVISION do the following:

- Highlight/Select POVPIP and drag it to the box Σ (upper-case sigma, which stands for “sum of”) Values. Click on “Count of POVPIP” and then on “Value Field Settings” in the tab “Summarize Values By” you should select “Average”. Click “OK”. Copy the results for your selected division in the following table

COPY THE TABLE 3 IN A WORD DOCUMENT and interpret your table.

Table 3. Distribution of Levels of Education by RACE [your unit of analysis] in [your selected Division]

Variable	EDUC (1)	EDUC (2)	EDUC (3)	ALL
Race/Ethnicity & Citizenship				
Label for RACE (1)				
Label for EDUC (2)				
Label for EDUC (3)				
Label for EDUC (4)				
Label for EDUC (5)				
Label for EDUC (6)				
TOTAL				

Source: Class sample from ACS 2013 1-yr estimates.

Now that you are done with your table, you will create a graph using the same results. Go back to Excel and create the graph. Following the next steps.

- Click on “PivotTable Tools” at the top of your screen. Then click on “Pivot Chart”. Select a 2 dimensional column graph, (NO 3D please, remember LESS IS MORE). Click “OK”.
- CLICK on “DIVISION” within the graph area. Take off the check mark from “(Select All)” and click ONLY on your selected Division.

Copy and Paste this graph in your word document. Remember that in your explanation you must refer to 1-3 according to the education levels (color coding of the bars) and 1-6 to racial/ethnic groups (groups on the X-axis).

REMINDERS:

Use the labels in the codebook provided in this assignment to correctly label the categories of your variables. Include your cross tab, the stacked column graph, the table for the difference of averages and the bar chart in a word file. Include your interpretations for all tables and graphs.

Format is always important, please make sure that you include all the components of a table:

Title: What is this table about?

Labels for rows and columns: DO NOT USE THE NUMERICAL CATEGORIES IN THE PIVOT TABLE. REPLACE THOSE WITH THE LABELS FOR EACH CATEGORY (See codebook included in this assignment).

Source:

Formatting: Use only horizontal lines as in the template provided above (LESS IS MORE!!!)

A figure (or graph) has the following elements:

Title:

Labels for the categories: as long as you describe the labels in your interpretation of the table, you do not need to include the labels for the categories in the graph.

Source:

If you have more questions about pivot tables in Excel you can go to:

<http://www.timeatlas.com/excel-pivot-tables/>

<http://www.wikihow.com/Create-Pivot-Tables-in-Excel>

<https://www.youtube.com/watch?v=9NUjHBNWe9M>

HOMEWORK: After participating in the class discussion you are ready to write the interpretation of results. Continue working on the same word file in which you have your tables and graphs. Below each of them write your interpretations. REQUIRED: Bring TWO copies of your work next class, do not include your name, as ID use only the first 2 and last 3 digits of your WID number.

Peer Review

Students will exchange copies of their homework during the second class session. Each student will review its colleague work according to the following rubric.

Checklist for Peer Review

*Clear title for table

*Clear labels for rows and columns

*Source cited

*Format (LESS IS MORE-only few horizontal lines)

*Clear and correct interpretation of percentage distribution and counts.

*Write objective and constructive comments for your colleague, including your reasons for the correct or incorrect interpretation of the distributions shown in the table and graph

Assessment

Pre/Post multiple choice question quiz.

Question for each learning objective

In the post test add the following questions:

What is the most important concept/skill that you learned after completing the activity?

Please let me know how you would change this assignment to improve your learning experience.

Instructional Notes:

Use the tables and graphs you have created and present to the students the correct interpretation of the data, and the differences observed by division. Emphasize that this is an example of BIVARIATE analysis, or the analysis of the relation between two variables at a time by DIVISION.

Table 1. Relative Distribution of Middle Class Indicators by Race, Nativity and Citizenship, United States, 2013
(Percentage of each group^a by race/ethnicity, sex and nativity)

Indicator	Non-Hispanic White	Non-Hispanic Black	Non-Hispanic Other	US born Hispanic	Hispanic Naturalized Citizen	Hispanic Non-Citizen
Education						
College Graduates						
Males	38.86 ^b	23.64	52.96	20.74	18.83	6.66
Females	42.13	29.31	51.61	26.84	23.53	9.55
English Fluency						
Speaks English well or very well						
Males	99.57	99.14	91.16	97.38	82.72	50.08
Females	99.62	99.18	90.34	97.77	82.21	44.60
Occupation						
Males						
Top White Collar	40.96	25.62	49.79	26.81	21.51	7.65
Other White Collar	17.00	16.92	16.01	18.79	13.55	7.51
Females						
Top White Collar	48.71	37.20	49.11	37.70	28.94	10.59
Other White Collar	31.46	28.65	24.96	36.34	27.39	17.01
Income						
More than three times the poverty line						
Males	70.55	47.55	67.37	52.85	47.92	21.68
Females	73.72	56.30	68.74	56.87	46.82	21.69
Homeownership						
Owned						
Males	75.59	52.21	62.40	58.44	65.01	33.91
Females	75.55	49.85	65.60	57.22	63.11	38.23
All indicators						
Males	27.89	13.97	29.04	14.04	11.51	2.00
Females	24.61	10.12	26.95	10.31	8.47	1.18

a. Currently employed civilian population between ages 25 and 65, who worked at least 40 weeks in the last 12 months.

b. This estimation reads as: 38.86 percent of the non-institutionalized civilian Non-Hispanic White male population between the ages of 25 and 65, who worked at least 40 weeks in 2012, and were employed at the time of the survey had at least a bachelor's degree.

Source: American Community Survey 2013, PUMS file, own calculations

Teaching/facilitation notes (later)

Resource guide – collection of docs

NOTES: Check with ICPSR page for working with data online (not recommended for lower division courses, but useful for upper division).

Assessments (pre, post the activity) – Included in each activity

Outline activities (pre, post the activity) – Included in this file

References (TBD)

http://www2.census.gov/acs2013_1yr/pums/

https://www.census.gov/acs/www/Downloads/data_documentation/pums/DataDict/PUMSDataDict13.pdf

Elements (check for final draft)

- ✓ Name
- ✓ College
- ✓ Module title
- ✓ Brief Description
- ✓ Learning Objectives
- ✓ Teaching/facilitation notes
- ✓ Resource list/guide
- ✓ Outline of lectures/activities
- ✓ Assessment
- ✓ Bibliography