

Inequality, Part II—An Analysis of Earnings

INTRODUCTION

While our first assignment's objective was **descriptive**, here we want to **analyze** some forces contributing to inequality in earnings. We've seen what the distribution of selected financial resources looks like both currently and over time. While a much larger percentage of American families are located in the top income bracket in 1990 compared to fifty years ago, there were still slightly over ¼ of American families with income of \$25K or less in 1990. Moreover, when we looked at individual earnings, over one-third (35.1%) of American workers earned below \$25,000 as of 2000! **Why?**

THEORETICAL FRAMEWORKS

This is a simplification, but some major answers to this question can be sorted into those that say, 'it's in the individual' and those that say, 'it's in the structure.' One influential set of ideas that argues 'it's in the person' is called *human capital theory*. This theory states that individuals have "capital" (e.g., their skills and abilities) much like a factory has capital (e.g., computers, tools, and machinery). If a factory owner improves and updates the equipment, all else being equal, she is likely to improve the factory's productive capacity and therefore the factory's earnings potential (profits). So, too, can individuals improve their "capital" through education, job experience, etc., and as they improve, their value increases. In other words, the market should reward their increased productive capacity by paying them more. Earnings inequalities then result from differences in the human capital that people possess.

Marx and a variety of other social scientists have argued that we need to widen our analysis. Earnings differentials may not be so much due to the individual but the *structures* in which they are located. Marx, for instance, said we have to understand the relationship between those who own the means of production (bourgeoisie) and those who simply sell their labor power (workers or proletariat). We also have to understand what's happening *to jobs* to answer our question. Another framework disputes the notion that there is one big, undifferentiated labor market in which we all compete. Rather, these theorists insist that there are different labor markets with different rates of pay, advancement opportunities etc., so that the same human capital could get rewarded differently depending on the individual's *location*. Still other critics point to barriers like discrimination, based on age, gender, race/ethnicity etc., that affect the pay that workers get.

OBJECTIVES

Unfortunately, we don't have the type of data necessary to examine all of these ideas as fully as they might warrant, but we can use census data to shed some general light on individual versus structural factors. Objectives for this exercise include addressing the following questions:

- **Is there a relationship between education and earnings as human capital theorists contend?**
- **Do the occupation or industry people work in influence at all the relationship between their education and earnings?**
- **What about ascribed statuses like race and gender? Do blacks and whites, men and women earn the same amount if they have equal levels of education?**

Other learning outcomes for this assignment include further honing analytical skills we practiced with the first assignment in terms of generating, reading, and interpreting rudimentary forms of data analysis. We increase our sophistication though by introducing a third variable into our crosstab analysis. Next, I lay out some guidelines for doing this kind of analysis.

USING CROSTABULATION AS AN ANALYTIC STRATEGY

Analyzing data, whether quantitative or qualitative in form, involves a conversation between you and the data. As researchers, we pose questions (guided by one’s conceptual understandings of the subject matter) and then use various analytic strategies to see what the data suggest in the form of answers. For example, human capital theory says that there is a relationship between education and earnings. This conceptual framework also points to *HOW* those two variables are related. Based on the theory, we can formulate a hypothesis, which, recall, is an informed expectation about their relationship. Before reading any further, answer this question: what should happen to earnings as educational level goes up? In the language that we used in the first assignment, in this case, EDUCATION is the **independent variable** (it’s influencing something) and EARNINGS is the **dependent variable** (it’s the “something” being influenced). We can enlist crosstabulation to examine whether or not our hypothesis is supported by evidence.

Interpreting the results. StudentChip does the number-crunching for you, but you still have to interpret what the tables are “saying.” Remember what you’re after: you want to see if the IV (education) has any association with the DV (earnings). To do that, you want to compare percentage differences **between the categories of the IV** (e.g., compare those with lower levels of education against those with higher). *[Note: to do this properly, you want to make sure each category of the IV adds up to 100%. Analytically, you’re asking what % of people with a high school education make 25K or less versus what % of people with a college education? That’s why, in building the tables in StudentChip, you’ll see below that we request that the program “percent down” through the education variable which constitutes the columns.]*

To illustrate some of what I’ve said here, below is a sample crosstab which picks up on my hypothetical example on social class and spanking introduced in our first assignment. This is actually real data (coming from the General Social Survey) analyzed by the SPSS program which I demonstrated in class. You may recall from that demo that the output looked different. I took what SPSS produced and cleaned it up a bit to make it more presentable to a lay audience. I’ll be asking you to do the same later in this assignment so look carefully at what I’ve done here.

Table 1: The Relationship Between Educational Level and Attitudes Toward Spanking for Discipline

| | Educational Level | |
|---|-----------------------------|--------------------------------|
| | High School Grad Or Less | 4-yr. College Degree or Higher |
| Favor Spanking to Discipline A Child | | |
| Agree | 76.1% (542) | 66.3% (189) |
| Disagree | 23.9% (170) | 33.7% (96) |
| | 100% | 100% |

In this example, if you compare DV categories, you see that a higher percentage of people approve of spanking than don’t. I’m not saying that’s not interesting, but it doesn’t address our conceptual question of whether or not education makes a difference in such attitudes. To answer that, we look at differences *between* educational categories. Besides, you could have already noted the differences in those who approve or disapprove spanking in a univariate analysis (frequency distribution).

There are three basic questions we can answer from this type of crosstabs output. You should frame your interpretations around these 3 questions:

- Is there a relationship between the IV and DV? (You're looking to see if there is enough of a difference in percentages between the columns to matter. Some guidelines are found in the third bullet here. Your answer at this point is a simple yes or no.)
- What is the relationship? (Here you want to put in sentence form what it is that the table is telling you. From my sample output below: "Those with a lower level of education are more likely to favor spanking compared to those with higher levels." That's the way you should phrase things. Be careful if you put it any differently because you may end up saying something that the results don't show. Also, please be aware that I've given a simple example here. Many of our variables of interest have more than two values they take. If your IV has 3 or more categories, the logic is still the same. You want to just compare across the IV categories. If the categories on each end seem to represent extreme differences, then you can look at those.)
- How strong is the relationship? (For this one, you want to look at the % differential. In the example, it's roughly 10%, which, in the social sciences, is a moderate relationship. Here's my rule of thumb for gauging strength: If it's 1-3%, then I'm hesitant to say there's much of a relationship. If it's 4-7%, then it's a slight relationship. An 8-15% difference is a moderate relationship, while anything approaching 20% is a substantial relationship. This is only meant as a rough guideline. These aren't technical rules.)

THE ELABORATION MODEL—DOING MULTIVARIATE CROSSTABULATION

Is level of education the only thing that might influence attitudes on spanking? From some of our class discussion, I think we've already seen that the world is more complex than 2 variables. We need to be able to sort out how other factors might relate to the phenomena in which we're interested. This is especially important because many of our variables are "block-booked" in the words of one sociologist (he borrowed the term from the film industry in which those showing films often had to rent them from distributors in packages of 4 or 5 rather than just getting the one big hit [e.g., Lord of the Rings]). So, too, are people "block-booked." We can describe any person in terms of a multitude of characteristics and many of those characteristics come in "blocks." For instance, someone from a lower socioeconomic status is also more likely to be working in jobs lower in occupational prestige, have lower levels of education, and be more liberal on economic issues and conservative on social ones. So, if people with less education are more likely to favor spanking, is it because of the education or is it something else associated with their social class status like the kind of jobs they have or their views on social issues? (The term "block-booked" and this example are drawn from M. Rosenberg's, *The Logic of Survey Analysis* [p. 26-27], the classic work on crosstabulation analysis of survey data.)

By introducing other variables into our analysis, we can begin to sort out such complexities. For crosstab analysis, we can start by adding a third variable, a procedure Rosenberg labeled "elaboration." If we add more than 3, typically we need other, more sophisticated statistical techniques that are beyond our discussion here. To illustrate elaboration, I present another example, this time drawing from data I collected in a recent class of mine. The following table presents information on the connection between gender and grades on an exam

Table 2: The Relationship Between Gender and Grades in a Soc 100 Class

| Grades | Gender | |
|--------------|---------------|-------------|
| | <u>Female</u> | <u>Male</u> |
| 90 and above | 43% | 30% |
| 80-89 | 48% | 51% |
| Below 80 | 9% | 19% |
| | <u>100%</u> | <u>100%</u> |

Ok. These results are pretty clear—women did better than men on this exam. Based on my guidelines listed earlier, there is a moderate relationship between gender and grades. Are women smarter than men? Well, maybe. But aren't there other things that could be "block-booked" with gender? Does one sex study more than the other? How about where they sit? I've noticed in many of my classes that a higher proportion of women sit in front than in back (Notice I said "proportion." It's true that there are more women than men in sociology classes, but if I look at it proportionally, I'm taking that into account). Maybe seating location is really the key to women performing better since students in front are possibly more interested and pay closer attention. To analyze this, I can elaborate on the table above by adding seating location. We are, in a sense, controlling its influence by separating those who sat in front vs. those who sat towards the back. To see if seating location makes a difference, let's take those in front and compare the sexes and then do the same for those who sat in back. Here are my results:

Table 3: The Relationship Between Gender and Grades Controlling for Seating Location

| Grades | <i>First 4 Rows</i> | | <i>Last 4 Rows</i> | |
|--------------|---------------------|-------------|--------------------|-------------|
| | Gender | | Gender | |
| | <u>Female</u> | <u>Male</u> | <u>Female</u> | <u>Male</u> |
| 90 and above | 45% | 37.5% | 39% | 29% |
| 80-89 | 45% | 62.5% | 52% | 52% |
| Below 80 | 10% | --- | 9% | 19% |
| | <u>100%</u> | <u>100%</u> | <u>100%</u> | <u>100%</u> |

To interpret these results, we need to compare each of these two tables to the original. You're asking whether or not the original pattern found between gender and grades still holds when this 3rd variable is added. Women were originally 13% more likely to get a 90 or above. If we look at just those students in the first few rows, women are still outperforming the men, but the difference has decreased. Moreover, the percentage of men in those front rows who scored higher has increased compared to all men in the class (37.5% in this table vs. 30% in the original). So while women still are outperforming men, seating location does seem to play a small role by slightly moderating the original relationship. If the % difference between men and women had disappeared or totally reversed itself, that would indicate that seating location was the more important variable.

Applying all of this to our questions means that we want to first examine the connection between education and earnings. Next, we will introduce additional variables that capture some structural (e.g., occupational category and industry) and ascriptive factors (e.g., gender, race/ethnicity) that may influence how much a person earns.

Let's get on with it!!

USING STUDENTCHIP

You can access StudentChip in a couple ways. First, you can use the program on our campus terminals in Dailey. If you do it this way, you will need to download the datasets you are using onto your disk from Angel. I include the same instructions as from ur first assignment:

1. Log on to our course in Angel and click on the "Lessons" tab.
2. Under the "Lessons" tab in Angel you will find a folder called "StudentCHIP Data Sets" which contains the necessary data sets. You will need **OCEDIN2K** (which you should already have downloaded for the first assignment) and possibly **INDEDIN2KC**.

3. To download the data, click on one of the datasets you need. Angel will take you to a new page stating: "The requested file can be accessed using the link below."
4. **Right** click on the file (it's listed at the top of the page with the extension .dat). Highlight the "Save Target As..." option.
5. In the "Save In" window, select "3 1/2 Floppy (A:)" or the appropriate drive or folder where you want to save the dataset and click on the "Save" button.
6. You probably saved it on the A: drive so now you have it on your floppy. You are now able to get into StudentChip and call up the data to analyze.

The other way to access StudentChip is through a website. Use these instructions:

1. Go to the Social Science Data Analysis Network at http://www.ssdan.net/ida_resources.shtml and click on the "WebCHIP" launcher link. Click on "WebChip Launcher" again.
2. At the next screen, use the "Menu-Enabled Launcher," highlight "Custom" and click on "Select Target."
3. Scroll down through the list of data sets until you find the data set needed (indicated below). Highlight and click "Select Target." This will bring up the data set and automatically start the CHIP program. Guidelines for running the analysis are found further below.

RUNNING CROSSTABS IN STUDENTCHIP

Note: For the first assignment, you simply copied output into the Tables included in the assignment. For this one, I want you to present your results in Tables you've created. While the software saves us a lot of work, it doesn't produce very good-looking results. You want to practice putting information together in a way a general audience could understand. To help you do this, you'll need your output for reference. One way to do this is to copy (by left clicking and dragging your mouse) and paste (right click) your results into your word processor. You should be able to do this with either StudentChip or WebChip. Or, if you're using StudentChip on our computers, you can create a log of your work and save it onto disk. To do that, as soon as you start the program, click on **File→Start Log**. The program will prompt you to assign a name to the file. Give it a name that you'll understand and **make sure you specify sending it to your floppy disk in the A: drive** (for example, you'd type in **a:\human capital results**). Next, you conduct your crosstabulation. When you're all done, click on **File→End Log**. If you go to the a: drive, you should see your file. If you didn't specify the a: drive, your file will be somewhere on the mainframe. Don't ask me where. Creating a log is not available on the Web version.

Task #1: Produce and examine results for the relationship between educational level and earnings. Answer the following question: Does the observed relationship support or contradict human capital theory?

Steps:

1. Start the program. If you're in StudentChip and want to create a log, do so now (according to above instructions).
2. Open the data set called **OCEDIN2K**. This can be found on Angel under **LESSONS→STUDENTCHIP DATA SETS** or, if you're using WebChip Launcher, follow the directions from the first assignment and look under "custom."
3. Produce a crosstabulation of **EARNING** by **EDUCATION**. Click on **Command→Crosstab**. Pick **EARNING** as your first variable, then click **select**. Next, pick **EDUC** and click **select**.
4. Next we have to tell Chip how to percentage the table. Keep in mind all I said earlier—we're interested in the percentage of people from each educational level

who fall into the different earnings bracket. So, choose **Table**→**Percent down**. Your table should now appear. With WebChip, use **EARNING** as the row variable and **EDUC** as the column variable and also select **% Down** to help construct the table. After those choices, click on **CROSSTAB** and you will have your table.

Your write-up should address the following for the first part of this assignment:

1. Create a table to present the information generated. This means you must take the “official” output from StudentChip and present it in an intelligible manner similar to what I did in my examples presented above. For example, give the table a good title and make the variables and their categories clear. **WARNING: YOU ARE NOT SIMPLY COPYING THE OUTPUT HERE.**
2. Interpret the table based on the 3 questions I introduced above (and also in class) for interpreting crosstabs.
3. Based on your results, is human capital theory supported or unsupported?

Task 2: Introduce an elaborated model that considers structural factors like the occupational group or industry in which one works. Draw conclusions based on what the data suggest for the human capital and structural theories.

1. Decide which 3rd variable you are interested in—*occupation* **OR** *industry*. If you choose occupation, you need to continue with the **OCEDIN2K** data set. If you are interested in the impact of industries, use the data set called **INDEDIN2KC** from either Angel or the Web Launcher (under “custom”).
2. Using the **Marginals** command, produce frequency distributions of the structural variable you chose (if you need to refresh your memory on how to do this, please refer back to our first assignment).
3. Next, follow the instructions from Task #1 above for producing the original crosstab between EARNING and EDUCATION. But, to introduce a third variable, if you’re using StudentChip here on our computers, when you click on **Table**, first choose **Control**. In the list of variables that appears select OCCUPATION or INDUSTRY depending on what you decided. Then, click on **% Down**. A series of tables will appear portraying the education-earnings relationship for each occupational or industry category. If you’re using WebChip, proceed as above for producing the initial table. As you hit the **crosstab** button to produce your first table, it will activate the **control by** box. You will see variables listed in the window next to the button. Highlight the variable you’ve chosen for your 3rd variable. Click **control by**, and then click **crosstab**. That should produce a series of tables: your original two variables for each of the categories of your control variable.

Your write-up should address the following for the second part of this assignment:

Present and interpret (give a one or two sentence summary) the frequency distribution of your third variable (OCCUPATION or INDUSTRY). (*Note: StudentChip output will use abbreviations for the “occupation” and “industry” categories. Some of the categories are self-evident whereas others may not be. Please find on the last page of this assignment a brief description and some examples of these classifications which may aid your understanding and interpretation of the output.*)

1. Create Tables (for a lay audience) to present the information generated from your crosstab output (again, use my earlier example as a guide).
2. Interpret the table based on the guidelines for using the elaboration model, presented in class and summarized earlier in this handout.
3. Based on your results, is there any evidence that structural factors play a role in earnings inequality? Some possible ways to think about this is to answer questions such as: If you have less than a high school education, what industry do you want to be in? If you are a

college grad, what industry do you want to be in? (Ask yourself similar questions if you used the occupation variable.)

Task #3: *Introduce a second elaborated model that considers ascriptive factors like the respondent's gender or racial status. Draw conclusions based on what the data suggest for the human capital and structural theories.*

1. Use the **OCEDIN2K** data set. Decide what 3rd variable you are interested in—*race* **OR** *gender*.
2. Using the **Marginals** command, produce a frequency distribution of the ascriptive variable you chose (if you need to refresh your memory on how to do this, please refer back to our first assignment).
3. Next, follow the instructions from Task #2 above for producing a crosstab with a third variable. A series of tables will appear portraying the education-earnings relationship for each gender or racial/ethnic category.

Your write-up should address the following for the third part of this assignment:

1. Present and interpret the frequency distribution of your third variable (either race or gender).
2. Create the tables to present the information generated from your crosstab output.
3. Interpret the tables based on the guidelines already presented (in class and above).
4. Based on your results, is there any evidence that ascriptive factors play a role in earnings inequality? Some possible ways to think about this is to answer questions such as: Do both men and women get equal returns on their educational investment? How about for people from different racial/ethnic backgrounds?

So, to sum up what your report should contain:

- Include an introductory paragraph stating what this assignment is about.
- The necessary statistical output (frequency distributions and crosstabs in tables you've created). You must construct the tables so a lay audience could understand. Do your crosstabs like I've done the examples in this handout.
- Your interpretation of the crosstabs results.
- Be sure to answer any questions found in the items located at the end of each task.
- A concluding paragraph that ties up what your results suggest for the human capital and structural views. Organize this paragraph around this question: Are earnings totally the result of something "in the individual" or something "in the structure?" Or is it a little bit of both?

(Below is a summary list of the Occupation and Industry categories. For a more complete breakdown, go to www.bls.census.gov/cps. At that site, click on the “Methodology and Documentation” box at the top. From the next list, click on “Basic Monthly Survey,” and from the resulting list click on either “occupation classification” or “industry.”)

Occupation Classification

| | | |
|---------|----|--|
| TopWC | -- | Top White Collar (e.g., managers, professionals) |
| OtrWC | -- | Other White Collar (e.g., technicians, sales, clerical support) |
| Service | -- | Service (e.g., protective [fire, police], private household, Food prep, personal [barber]) |
| Farm | -- | (e.g., farming, forestry, fishing) |
| TopBC | -- | Top Blue Collar (e.g., precision production, craft, repair, construction) |
| OtrBC | -- | Other Blue Collar (e.g., operators, fabricators, laborers, transportation) |

Industry Classification

| | | |
|--------|----|---|
| AgForM | -- | Agriculture, Forestry & Fisheries, Mining |
| ConsTr | -- | Construction |
| Manufa | -- | Durable and NonDurable Goods |
| Trade | -- | Wholesale and Retail Trade |
| FinIns | -- | Finance, Insurance, & Real Estate |
| BusRep | -- | Business and Repair Services |
| Health | -- | Health Services |
| EducGo | -- | Education, public administration, govt. |