**The Condition of Inequality, Part I—Current and Historic Patterns
In the Distribution of Income**

INTRODUCTION

We've argued that societal stratification is "both a condition and a process" (Kerckhoff, 2000). The former captures what the distribution of valued resources (e.g., money, education) among other things look like in a society. The question, most simply, is 'who gets what'? In this exercise, we'll examine contemporary and historical data on financial resources. First, we'll look at individual earnings, which is defined as the money a person makes from working and includes such things as wages, salary, or a form of self-employment and expressed as an annual amount. We'll also investigate historical trends in family income. Looking at these financial resources replicates and complements some of our text's (Chapter 5 in R. Rothman's *Inequality and Stratification*, 4th ed.) discussion of economic inequality (e.g., family income, CEO pay, net worth). Our data come from various Census Bureau sources, including its decennial census and its monthly Current Population Survey.

LEARNING OBJECTIVES

Skill
Learning outcomes for this assignment include honing analytical skills associated with generating, reading, and interpreting rudimentary forms of data analysis. Our goal is to interpret the information for a lay audience (e.g., write a summary that any newspaper reader could understand).

Substance
Social science research is often categorized by two of our major objectives—**describe** and **analyze**. The first simply tries to build a portrait of the way things are (or were) while the latter is concerned with trying to explain why things are the way they are. For this assignment, our objective is descriptive. We will answer the following 2 questions:

- **What is the distribution of earnings among U.S. full-time workers, ages 25 and over, in the most recent year for which data are available? For instance, what percentage of these workers make 50-75K?**
- **How has this distribution changed in the post WWII era?**

A PRIMER ON THE LOGIC OF SOME BASIC FORMS OF DATA ANALYSIS

Social scientists are interested in whether or not there are patterns or relationships among the social phenomena on which we collect information. A major conduit for such information comes in the

form of surveys. Survey research translates our phenomena of interest into **variables.** A variable is anything that can vary, i.e., take different values, and it represents our way of measuring the concepts in which we are interested. To illustrate, for this course, social class is one of our important concepts, and it can vary, right? People can be located higher or lower in the social class structure. In order to try to measure that in the real world, I have to find a way to put the concept into operation. For instance, maybe I could construct a survey in which I ask respondents to identify their social class location and I gave them choices of "upper class," "middle class," "working class," or "poor" (in methodological terms, I call this the "**response set**" for this question) from which to select. That's one way of measuring our concept— social class. Now, to look for patterns, I might examine the variation on just this one variable by looking at the distribution of cases [respondents] across the variable's response set (we call this doing "**univariate** analysis" through generating a "**frequency distribution**"). Doing that would tell me how many of my survey respondents classified themselves as "middle class," "working class" and so on. While it doesn't take me real far, this type of analysis is still useful in its own right.

But since my hypothetical survey also asked about a lot of other stuff, I could expand my analysis to see what other kinds of things are associated with my respondent's self-reported social class status. Examining the relationship between two variables is important since social scientists often are trying to make sense of how variables are related to one another. To do so, we work with conceptual frameworks (theories) that lead us to expect certain relationships between our variables of interest. Here's an example continuing with social class. We know that social class is connected to many things including how parents might raise their children. One conceptual framework suggests that middle class parents, because they tend to work in less regimented environments might actually be less regimented in their child rearing practices.

A concrete variable that represents such child rearing practices could be whether or not someone favors physical forms of punishment. Let's say that the actual measures for these two concepts are my earlier measure on social class and the degree to which a parent believes in spanking as a form of discipline (measured with a response set of "strongly agree," "agree," etc.).

One simple, yet useful, way to investigate relationships between variables like these is through cross-tabulation. A "**cross-tab**" is a table that presents the distribution (in frequencies and/or percents) of one variable across the categories of another variable(s) (e.g., what percentage of parents from of middle class background agree with spanking). Since it let's us look at two variables, we label this a **bivariate** analysis. Typically, in crosstabs and many other statistical techniques, we conceptualize the relationship between the two variables in terms of one influencing the other. The language we use to capture such relationships is to call one variable an **independent variable (IV--it's doing the influencing)** and the second variable a **dependent variable (DV--it's the one that is being influenced).** To run a crosstab you decide on two variables that you think might be related to one another. Typically, drawing from a conceptual framework, one next states a **hypothesis** for the relationship between your chosen variables. A hypothesis is an educated guess about what you think you will find. Hypotheses should always state a specific relationship and specify the comparison. Extending my example, the conceptual framework I noted above would lead to the following expectation: respondents from a lower social class background are more likely to favor physical punishment of children compared to those from higher social class locations. In this example, "social class" is the IV and "attitude towards spanking" is the DV. Then I'd go on and examine the data to see if my expectation is borne out.

So that's a hypothetical example. We're going to be investigating some other stratification-related variables that were collected as part of the U.S. Census. As part of the sociology's participation in a special program, people at the University of Michigan have prepared some special data sets with variables I've requested. I turn now to a brief description of such data.

CENSUS DATA

       (This description draws from W. Frey's *Investigating Change in American Society.*) The U.S. government can probably be accused of a lot of things, but when it comes to data-gathering efforts, they're not too shabby. The reason, of course, is that they have much more money to spend on it than do most researchers. As you're probably aware, one of the government's biggest efforts is the decennial census (mandated by our Constitution). The Census's original purpose was for apportioning our congressional representatives every ten years. The first one was done in 1790; the most recent, of course, was the 2000 census. Although originally meant to count every single person, it does a lot more than that since it collects a lot of other information. Recently, for instance, the census bureau sent out two different surveys, a "short" and a "long" form. The "short" form asked just for basic sociodemographic information (e.g., age, gender, race/ethnicity etc.), while the "long" form added many more items on a variety of social characteristics (e.g., occupation, education, language proficiency). These characteristics represent some variables. The information on these characteristics is of immense importance for both planners at all governmental levels (federal, state, and local) and social science researchers. For this class, we can examine stratification-related phenomena like people's earnings, their occupational category, educational level, and much more. Our assignments will require us to do both the univariate and bivariate types of analyses I introduced above for these types of variables.

       Before we get to the actual assignment, one thing to remember is that no research effort is perfect, including the census's. They are never able to count everyone (e.g., all the homeless), and some people may be more likely to answer the different forms than others (we refer to this potential problem as **response bias**). The sample sizes are so large; however, that it still represents a pretty accurate picture of American people. Now, we turn our attention briefly to our software program to see how we will access our census data.

USING WebCHIP

You can access WebCHIP through the SSDAN website. Use these instructions:

1. http://www.ssdan.net/datacounts/data/
2. From there, click "Browse" on the left sidebar. Find **"custom"** in the drop-down box and select it.
3. Scroll down through the list of data sets until you find "**OCEDIN2K.DAT".** Highlight and click "submit." This will bring up the data set in the WebCHIP program and it is ready for analysis. Guidelines for running the analysis are found further below.

Now, on to our assignment! Please follow the instructions and complete the necessary information for the three tasks found below. You may put your name on the top of the next page. Separate and turn in these next 3 pages for your assignment.

Soc 325   Name: _____

*TASK #1: Univariate Analysis—Produce a descriptive portrait of **CURRENT** income inequality using the most up-to-date information available.*

Directions:

1. Following the directions above, open the dataset **OCEDIN2K**. You can also click here to launch the dataset in WebCHIP.
2. For this data set, the Info output will say "2000 Personal Earnings by Occupation …" It will also give you the total number of individuals these data are drawn from (N=98,436,674).
3. Next, create a marginals table. This procedure results in a listing of each **variable** in this file with its **frequency distribution.** Find the "Earnings" variable and fill in below the % of people falling in the various categories.

< $25K	$25-35K	$35-50K	$50-70K	$70-100K	$100K+

Question: Summarize in words what the current income distribution looks like.

*TASK #2: Produce a descriptive portrait of income inequality **OVER TIME**.*

Directions:

1. Following the directions above, open the dataset **OCIN5090**. You can also click here to launch the dataset in WebCHIP.
2. One of the variables in this data set is YEAR which draws from information from each decennial census since 1950. Your job is to produce a crosstabulation of earnings distribution by year (Note: In this case, however, we're not really thinking of the variables so much in terms of independent and dependent. We're not insinuating that the year in which the data was collected actually influenced the income gained. We're simply tracking over time how the distribution changed).
3. Pick **Fmincome** as your row **variable** and **Year** as your column **variable.** We're interested in the percentage of people from a given year who fall into the various earnings bracket, so create a percent down crosstab
4. Note: The income variable is a bit different here. We're looking at income for the whole family. Also, the categories are different from the first part of the assignment since we're using historical data going back 50 years ago. I realize that in this day-and-age, making over $50K doesn't exactly qualify someone for Forbes's "richest in America" list, but it still gives us a picture of how things have changed over time.
5. Fill in the matrix below based on your data.

Year	1950	1960	1970	1980	1990	All
Family Income						
Below 15K						
15-25K						
25-35K						
35-50K						
Above 50K						

Question: Describe the trends in family income distribution since WWII. Has the % who are making less than 15K gone up, down, or some of both? How about for those who make above 50K? What would you conclude about the distribution of income among U.S families since WWII? Are American families becoming more affluent? You want to aim for a complete yet concise summary of the results.

TASK #3: Produce a descriptive portrait of other potentially relevant trends over this same time period and speculate on connections between these and the income trends witnessed above.

<u>*Directions:*</u>

Finally, we'll produce one more table. YEAR will be the column variable. Pick either OCCUPATION or EDUCATION as the row variable. Fill in the matrix below with the results (put in your row variable of choice and its respective categories).

Year	1950	1960	1970	1980	1990	All

(Note: If you've chosen "Occupation" as your other variable to look at, some of the categories are self-evident whereas others may not be. Please find on the following page their description and some examples.)

Question: In a few sentences, first summarize what the table reveals. Then, speculate on what you think the connection might be between results from the earlier table that looked at trends in *Family Income* since 1950 and this one that examines similar trends in *Occupation* or *Education* in the same time period. Draw from relevant material in ch. 3 and possibly 5 from our text to help support your reasoning.

<u>Occupation Classification</u>

Below is a summary list of the occupational categories.

TopWC	--	Top White Collar (e.g., managers, professionals)
OtrWC	--	Other White Collar (e.g., technicians, sales, clerical support)
Servic	--	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)