

Prepared by:

**Jana L. Jasinski, University of Central Florida**

Note: The purpose of this module is to introduce students in an online Criminology course to the idea of data analysis using an online analysis program and the General Social Survey. Prior to this module, students will have read several documents describing the scientific method including terminology such as statistical significance, independent and dependent variables, and operationalization. Once students have read this document, they then complete a guided online analysis and turn in their answers by completing an online quiz. Listed Below is the actual module and quiz questions. The module begins with an analysis that the students can follow along with. After they become acquainted with how the program works, students can then complete a data analysis exercise on their own.

Data Analysis Module for Criminology  
Department of Sociology & Anthropology

### **FEAR AND CRIME: DATA ANALYSIS MODULE**

#### **Learning Objectives:**

##### *Skill*

- Using software to access and analyze census data
- Identifying independent and dependent variables
- Forming testable hypotheses using quantitative data
- Learning how to construct, read, and interpret bivariate tables displaying frequencies and percentages
- Using real world data to enhance and support key course concepts

##### *Substance*

- Examining the correlation between crime statistics and demographic data

While you are thinking about the relationship between media consumption and crime you may have developed a hypothesis about the relationship between media consumption and crime. To introduce you to one of the data analysis activities you will be using in this course we can actually test a hypothesis about these two variables. We will do this using data from the General Social Survey (GSS). The GSS is a national survey designed with a standard core of demographic and attitudinal variables, and special interest topic modules that are rotated into the survey. The purpose of this biannual survey is to gather data on contemporary American society and to provide easy access to high quality data.

This data set is commonly used by social scientists and we have access to an interactive online analysis program.

The GSS online can be located at the following address:

<http://www.norc.unc.edu/GSS+Website/Data+Analysis/>

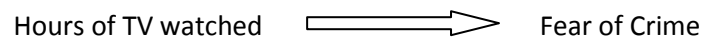
Once you arrive at this screen you should select the data file that you will be using (The GSS Cumulative Data File) as well as options for analysis.

For this exercise we will be conducting both univariate (single variable) and bivariate (two variables) analyses, specifically frequency distributions and comparison of means (also known as contingency tables).

After thinking about the possible relationship between how fearful someone might be and their level of media consumption we might develop the following hypothesis:

**Hypothesis:** People who watch more hours of television will be more likely to report being fearful of crime.

We could also use a causal diagram to illustrate the hypothesis:



Now to test this hypothesis, we need to have a way to measure each of these variables (Fear and television consumption). It just so happens that both of these are questions asked in the GSS.

To measure FEAR, the GSS asks respondents "Is there any where around here - that is within a mile- where you would be afraid to walk alone at night?" - Respondents answer either yes or no. In the language of science, this question is our **operationalization** (the way we are actually going to measure) of the variable FEAR. The name for this variable in the data file is FEAR.

To measure Media Consumption the GSS includes a question asking "On the average day, about how many hours do you personally watch television?" The name for this variable in the data file is TVHOURS.

The program wants to know what variables you are interested in analyzing. Use the left hand tool bar to select your variables. For this exercise we will use FEAR as our dependent variable (the outcome) and TVHOURS as our independent variable (that variable that influences the outcome).

If you are interested in viewing a list of the variables in this data file use the mnemonic link, which will give you an alphabetical subject listing of variables included in the GSS. Clicking on any of the letters listed at the top of the screen will produce a list of variables beginning with that letter. The names you see are the variable names you should use to run your analyses. For example, the variable with the name RACE represents the racial group to which survey respondents indicated that they belonged.

First let's find out how many people in the sample answered yes to the following question: "Is there any area right around here--that is, within a mile--where you would be afraid to walk alone?" Remember this is the operational definition of the concept FEAR.

Find the variable FEAR and add it to a row. This will produce a frequency table for the variable **FEAR**. You can replace the FEAR variable with any other variable and run the same analysis.

1. How many people said Yes to the question? .
2. What percent of the sample said they were fearful?

Now lets look at the other variable **TVHOURS**. Follow the same steps as you did earlier but in addition to find the variable.. Look at the frequency analysis of the variable **TVHOURS**. You should see a long list of categories and the number and percent of people in each. For variables like this (we call them continuous) it may be more interesting to look at the average or mean number of tv hours watched. You can find this statistic at the bottom of the page. In this case the average number of hours of television watched by the respondents was 2.97 hours.

You can also run a comparison of means to see if the mean number of tv hours watched is different for those who said they were fearful compared to those who did not report they were fearful.

To do this go back to the tabulation option. Add the variable name TVHOURS to the columns and add the variable FEAR in in the rows and then run the table. You should see a new table where you can see how many hours a day people that people who reporting being fearful watched tv compared to people who were not fearful.

3. Is this consistent with our hypothesis?

*Yes - our hypothesis was that people who watch more hours of television will be more likely to report being fearful of crime.*

We also want to know if the difference we are seeing is statistically significant. To do this we need to look at the table with the heading analysis > correlation. In this table we can find the P-value under the column with the heading "P".

4. Is this number smaller than .05? If yes, we can say that the relationship is statistically significant.

So what have we found out? After our data analysis we can say that people who reported being fearful watched significantly more hours of television than people who did not report being fearful.

You can also run a crosstabulations (also known as a contingency table) in this online analysis program. Below are the instructions for a crosstabulation looking at the relationship between the independent variable (SEX) and the dependent variable (FEAR). To run this analysis with other independent variables, simply replace the name for SEX with the variable name you are interested in looking at.

To find out if there are gender differences in fear, we can conduct a contingency table (also known as crosstabulation). Add the FEAR variable to the Rows and the variable name SEX to the

columns. The default for the percentaging option should be column. If it is not, use the drop down menu to select column. Click on the Run Table button.

If you want some practice why not look at the average number of hours of television watched by men compared to women? Do you think that this has any relationship to their levels of fear?

### **DATA ANALYSIS EXERCISE**

Now that you have practiced using the online data analysis program you are ready to conduct your own data analysis.

(Note: These questions would appear to the students in the format of an online quiz)

1. For this quiz, you will need to use the online GSS program as described previously. What would you hypothesize is the relationship between gender (SEX) and victimization (GUN)? What would you hypothesize is the relationship between race (RACE) and victimization (GUN)?
2. What is the dependent variable in these analyses?
3. What is the operationalization of the concept of victimization?
4. What percentage of people said they had been shot at or threatened by a gun?
5. What percent of men were victims? What percent of women were victims?
6. Who was more likely to be a victim, men or women? Were the differences significant? How do you know this?
7. Which racial group was more likely to report being a victim? Were the differences significant? How do you know this?
8. Did your data analysis support your hypotheses?
9. What other factors do you think would be related to victimization?