## Sociology 3 - Critical Thinking - Spring 2020

**Note: Before you begin this paper, be sure to read chapter 2 in *Public Opinion on Social Issues* (see Canvas).**

# Instructions for Paper—Opinion on Legalization of Marijuana (40 points max)

We’re going to use the General Social Survey (GSS) for this exercise. The GSS is a national probability sample of adults in the United States conducted by the National Opinion Research Center (NORC). The GSS started in 1972 and has been an annual or biannual survey ever since. For this exercise we’re going to use the 2018 GSS. To access the GSS cumulative data file in SDA format **click** [here](https://sda.berkeley.edu/sdaweb/analysis/?dataset=gss18). SDA stands for Survey Documentation and Analysis and is the statistical program you'll be using in your paper. Read the document titled "Introduction to SDA" on Canvas. This introduction will tell you everything you need to know for running SDA.

The GSS cumulative data file contains all the data from each GSS survey conducted from 1972 through 2018. We want to use **only** the data that was collected in 2018. To select out the 2018 data, enter *year(2018)* in the SELECTION FILTER(S) box. See "Introduction to SDA" for more information on using SDA.

**Part 1. Exploring how people feel about the legalization of marijuana**

The dependent variable for this paper will be a variable named *grass* which is the respondent's answer to the following question – "Do you think the use of marijuana should be made legal or

not?" Use SDA to get the frequency distribution for *grass*.

What percent favored the legalization of marijuana?

What percent opposed legalization?

Not all respondents were asked this question. Of those who were asked the question, some said they didn't know and some refused to answer. This is referred to as missing data and these respondents were given missing value codes. This is typical of survey data. Sometimes a question is only asked of a random percent of the sample and usually some respondents fail to answer the question. So we're going to be working with a sample of a little less than 1,500 adults in the United States who answered the question. You can see how many respondents fall into this category by scrolling down to the bottom of the output and looking for the number of "cases with invalid codes on row variable." This is different from the number of "cases excluded by filter or weight."

How many cases fall into this category?

## Part 2. Age

The goal of this paper is to explore some of the reasons that people have different opinions about the legalization of marijuana. There are many variables in the GSS that might help us understand why people hold different opinions on this issue. We're going to start with *age*.

The first thing we need to do is to make sure that we understand the difference between independent and dependent variables. The dependent variable is what you're trying to explain and the independent variable is what you think will help explain some of the variation in the dependent variable. In this paper we're trying to explain why some people favor and others oppose the legalization of marijuana. The variable name for this question is *grass*. So *grass* will be our dependent variable. Age is a variable we think might help explain the variation in the dependent variable so *age* is our independent variable. We're going to put the independent variable in the COLUMN box and the dependent variable in the ROW box. That means that we want to ask for the column percents which fortunately is the default in SDA.

Age is recorded in actual years old. There's way too many categories so we will have to divide age into a smaller number of categories. This is called recoding. Rather than show you how to write recode statements, I'm going to give you them to you. All you have to do is to copy and paste them into the column box. Here's the command. This will recode age into three categories – under 35, 35 to 64, and 65 and over.

 age(r:1=18-34 "under 35"; 2=35-64 "35 to 64"; 3=65-89 "65 and older")

Be sure to read the instructions for running a crosstabulation in the "Introduction to SDA" (see Canvas) before running the table.

Each section of this paper will ask you five questions.

1. Enter your hypothesis in your paper.
2. Enter your dummy table in your paper. Do this by creating your dummy table using "Tables" in Word. We'll go over this in class.
3. Enter the crosstab in your paper. Do this using "Tables" in Word. We'll go over this in class.
4. How would you interpret this table? Describe the relationship between your independent and dependent variables. Think of the interpretation as consisting of two parts. The first part should summarize in words the pattern of the percents. The second part should use the percents to illustrate the pattern.
5. Was your hypothesis correct? How do you know?

I'm going to get you started by answering these questions for you in this part of the paper and then you're going to answer them yourself in the remaining parts of the paper.

1. Hypothesis – Younger respondents are more likely to favor the legalization of marijuana while older respondents are more likely to oppose legalization.
2. Dummy table – we'll talk about what this table means and how to construct it during class.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Age: under 35 | Age: 35-64 | Age: 65 plus |
| Favor legalization | a> | b> | c |
| Oppose legalization | d< | e< | f |

1. SDA (observed) crosstab

|  |  |  |  |
| --- | --- | --- | --- |
|  | Age: under 35 | Age: 35-64 | Age: 65 plus |
| Favor legalization | 77.4% | 65.6% | 49.0% |
| Oppose legalization | 22.6% | 34.4% | 51.0% |
| Total | 100% | 100% | 100% |

1. Interpretation – Since our percents sum down to 100%, we want to compare the percents straight across. There's a clear pattern to the percents such that younger respondents are more likely to favor legalization than those who are older. For example, approximately 77% of those under 35 favor legalization while only 49% of those 65 or older favor legalization. Those who are 35 to 64 fall between these two age groups (66%). In other words, age decreases, the percent who favor legalization increases.

We're going to use something that we'll call the 5% rule. Any percent difference less than 5 percentage points we'll consider small. We never want to make too much of small differences because of sampling error. So for this table we would conclude that as age decreases, the percent who favor the legalization of marijuana increases.

Think of your interpretation as consisting of two sentences. The first sentence should summarize in words the pattern of the percents. The second sentence should use the percents to illustrate the pattern.

1. Do the data support your hypothesis? In this example our hypothesis was correct because the table shows that younger respondents are more likely to favor the legalization of marijuana than older respondents.

## Part 3. Gender

Another variable that might help explain some of the variation in our dependent variable is gender. That variable is named *sex* in the GSS. Here are the five questions to answer in this part of the paper.

1. Enter your hypothesis in your paper.
2. Enter your dummy table in your paper. Do this by creating your dummy table using "Tables" in Word. We'll go over this in class.
3. Enter the crosstab in your paper. Do this using "Tables" in Word. We'll go over this in class.
4. How would you interpret this table? Describe the relationship between your independent and dependent variables. Think of the interpretation as consisting of two parts. The first part should summarize in words the pattern of the percents. The second part should use the percents to illustrate the pattern.
5. Was your hypothesis correct? How do you know?

## Part 4. Socioeconomic Status

Recall that we talked earlier in the semester about different variables serving as markers of a person's position in the social structure. Age and gender are important markers of a person's position in the social structure. Another marker is socioeconomic status (SES). SES is a concept that can be measured in various ways. One possible measure of SES is a person's education. The GSS includes a variable named *degree* which is the person's highest educational degree. Another way to measure SES is to ask respondents to classify themselves as lower, working, middle, or upper class. This is a subjective measure of SES which is called *class* in the GSS.

We'll need to recode each of these variables. Here are the recodes to use when running SDA.

degree (r:1=0-1"high school or less"; 2=2-4"some college or degree")

class (r:1=1-2"lower, working class";2=3-4"middle, upper class")

For both of these variables answer the same five questions.

1. Enter your hypothesis in your paper.
2. Enter your dummy table in your paper. Do this by creating your dummy table using "Tables" in Word. We'll go over this in class.
3. Enter the crosstab in your paper. Do this using "Tables" in Word. We'll go over this in class.
4. How would you interpret this table? Describe the relationship between your independent and dependent variables. Think of the interpretation as consisting of two parts. The first part should summarize in words the pattern of the percents. The second part should use the percents to illustrate the pattern.
5. Was your hypothesis correct? How do you know?

Did you get consistent results for both *degree* and *class*? What does that tell you about the relationship between SES and how people feel about the legalization of marijuana?

## Part 5. Political Party Identification

In our list of possible variables that might explain some of the variation in how people feel about the legalization of marijuana, we wouldn't want to forget political variables. One of the variables in the GSS is the political party with which respondents identify. This variable is named *partyid*. The GSS asked respondents to classify their political party preference as strong Democrat, not strong Democrat, independent leaning toward the Democrats, independent leaning neither toward Democrats or Republicans, independent leaning toward Republicans, not strong Republican, or strong Republican. They also gave people the option of choosing another party. We're going to drop that group from our analysis since there were so few people who choose this option.

There are at least two ways we could recode this variable. One would be to group the leaners with either Democrats or Republicans and leave independent by itself as a separate category. Here's the recode this would do this. Independent without leaners means that the independent category does not include the leaners.

partyid (r:1=0-2"Democrat";2=3-3"independent without leaners";3=4-6"Republican")

The other way to recode is to include the leaners with the independents. Here's the recode that would do this.

partyid (r:1=0-1"Democrat";2=2-4"independent including leaners";3=5-6"Republican")

Run the crosstab twice – once for each way of recoding *partyid*. In other words, use one of the recodes and then rerun it using the other recode.

Does the way we recode *partyid* make a difference? In other words, do you get the same results regardless of the way you recode *partyid*?

Pick one of the recodes and answer the same five questions.

1. Enter your hypothesis in your paper.
2. Enter your dummy table in your paper. Do this by creating your dummy table using "Tables" in Word. We'll go over this in class.
3. Enter the crosstab in your paper. Do this using "Tables" in Word. We'll go over this in class.
4. How would you interpret this table? Describe the relationship between your independent and dependent variables. Think of the interpretation as consisting of two parts. The first part should summarize in words the pattern of the percents. The second part should use the percents to illustrate the pattern.
5. Was your hypothesis correct? How do you know?

**Part 6. Political Views**

Another political variable that might account for some of the variation in how people feel about the legalization of marijuana is political views. Here's the question that the GSS asked respondents. "We hear a lot of talk these days about liberals and conservatives. I'm going to show you a seven-point scale on which the political views that people might hold are arranged from extremely liberal - point 1 - to extremely conservative - point 7. Where would you place yourself on this scale?" This variable is named *polviews* in the GSS.

Here's the recode to use for *polviews*.

polviews (r:1=1-3"liberal";2=4-4"moderate";3=5-7"conservative")

Run the crosstab in SDA and answer the same five questions.

1. Enter your hypothesis in your paper.
2. Enter your dummy table in your paper. Do this by creating your dummy table using "Tables" in Word. We'll go over this in class.
3. Enter the crosstab in your paper. Do this using "Tables" in Word. We'll go over this in class.
4. How would you interpret this table? Describe the relationship between your independent and dependent variables. Think of the interpretation as consisting of two parts. The first part should summarize in words the pattern of the percents. The second part should use the percents to illustrate the pattern.
5. Was your hypothesis correct? How do you know?

Now think about why you stated your hypothesis as you did. In other words, if you said that conservatives were more likely that liberals to oppose the legalization of marijuana, why do you expect this to be true. That requires you to develop an argument for which the hypothesis is the conclusion to your argument and the premises explain why you think your hypothesis will be true.

* Write out your argument so it is clearly stated using correct grammar and spelling.
* Use inference indicators (both premise and conclusion indicators) in your argument. Circle the inference indicators.
* Place the conclusion (i.e., the hypothesis) at the end of your argument and underline it.

**Part 7. Religiosity**

The last independent variable we will consider is religiosity which refers to how religious the person is. We're going to use two different measures of religiosity – attendance at religious services (name of variable is *attend*) and subjective strength of religion (name of variable is *reliten*).

The GSS measured attendance at religious services (*attend*) by asking respondents, " How often do you attend religious services?" The response categories were never, less than once a year, once a year, several times a year, once a month, two to three times a month, nearly every week, every week, more than once a week. Use the following recode for *attend*.

attend (r:1=0-3"seldom or none";2=4-5"sometimes"; 3=6-8"often")

The GSS measured subjective strength of religion (*reliten*) by asking respondents, " Would you call yourself a strong [religious preference] or a not very strong [religious preference]?" Some respondents volunteered "somewhat" and that was added as another response category. Respondents who said they had no religion were still another category. So the response categories were strong, somewhat strong, not very strong, and no religion. Use the following recode for reliten.

reliten (r:1=1-1"strong";2=2-4"not strong")

Run the two crosstabs – one for *attend* and another for *reliten*. Did you find the same relationship between each of your measures of religiosity and how people felt about the legalization of marijuana? In other words, did you get the same results regardless of which measure you used?

Pick one of the two measures of religiosity and answer the same five questions.

1. Enter your hypothesis in your paper.
2. Enter your dummy table in your paper. Do this by creating your dummy table using "Tables" in Word. We'll go over this in class.
3. Enter the crosstab in your paper. Do this using "Tables" in Word. We'll go over this in class.
4. How would you interpret this table? Describe the relationship between your independent and dependent variables. Think of the interpretation as consisting of two parts. The first part should summarize in words the pattern of the percents. The second part should use the percents to illustrate the pattern.
5. Was your hypothesis correct? How do you know?

Now think about why you stated your hypothesis as you did. In other words, if you said that those who were more religious were more likely to oppose the legalization of marijuana than those who were less religious, why do you expect this to be true? This requires you to develop an argument for which the hypothesis is the conclusion to your argument and the premises explain why you think your hypothesis will be true.

* Write out your argument so it is clearly stated using correct grammar and spelling.
* Use inference indicators (both premise and conclusion indicators) in your argument. Circle the inference indicators.
* Place the conclusion (i.e., the hypothesis) at the end of your argument and underline it.

**Part 8. Spuriousness**

**NOTE: Be sure to read chapter 3 in *Public Opinion of Social Issues* and the discussion of “spuriousness” which is on Canvas before you do Part 8.**

In the previous sections of this paper we found that some of our independent variables were related to how people felt about the legalization of marijuana while others weren't. In Part 7 we discovered that people who said they were more religious were more likely to oppose the legalization of marijuana. We already know from Part 2 that older respondents were more likely to oppose legalization. Although we didn't run the table, it's also true that older respondents are more likely to be more religious. We can diagram this in the following causal model.

Ch

This raises the possibility that the relationship between reliten and grass might be spurious. In other words, more religious individuals might be more opposed to the legalization of marijuana **because** they are older and less religious individuals might be more in favor of the legalization of marijuana **because** they are younger. The relationship between *reliten* and *grass* might be due to this third variable (*age*).

Spuriousness means that there is a statistical relationship between the independent variable and the dependent variable, but it's not a causal relationship. The relationship between the independent variable and the dependent variable is a result of this third variable which is called the control variable.

For a relationship to be spurious we need to identify a third variable that might explain away the original relationship. That third variable has to be related to both the independent and the dependent variables. We have already seen that *age* is such a variable. It's related both to the independent variable (*reliten*) and the dependent variable (*grass*).

How do we test to see whether a relationship is spurious or not? To do that we need to run a three-variable table which controls for *age*. Use *reliten* as your independent (or column) variable, *grass* as your dependent (or row) variable, and *age* as your control variable. Be sure to use the following recoded versions of both *reliten* and *age*.

 age(r:1=18-34 "under 35"; 2=35-64 "35 to 64"; 3=65-89 "65 and older")

 reliten (r:1=1-1"strong";2=2-4"not strong")

Your output will show four tables stacked on top of each other. The top three tables show the relationship between *reliten* and *grass* for each category of the control variable (*age*). The top table shows the relationship for those under the age of 35; the second table shows the relationship for those 35 to 64; the third table shows it for those 65 and over. The bottom table shows the relationship for all ages. This is the two-variable table that you ran in Part 7.[[1]](#footnote-1)

What you just did was to control for age. Control means to hold that variable constant. *Age* has three categories – those under 35, those 35 to 64, and those 65 and over. That means you'll have three partial tables. Each partial table contains part of the data broken down by age. That’s what it means to control for a third variable or to hold that third variable constant.

Now we want to compare the top three tables with the bottom table. Notice that the bottom table contains all the respondents (i.e., all ages).

Note that I'm not asking you in Part 8 to write a hypothesis, create a dummy table, or create a table showing the percents from the SDA table. What I am asking you to do is to carefully answer the questions in the following two paragraphs. Use the percents from the SDA crosstabulation in your answer. Don't just write a sentence or two. Rather carefully think about the questions and write your answers.

Write a paragraph explaining what you found when you controlled for *age* (i.e., when you compared the top three tables with the bottom table). Use the percents to help you make these comparisons. Did the relationship in each of the top three tables stay basically the same as in the bottom table or did the relationship partially go away when you controlled for *age*?

Compare this to the example in the handout on “spuriousness” on Canvas. What does spuriousness mean? Are the tables that you just ran an example of spuriousness? How do you know?

**Part 9. Conclusions**

What did you learn about how people felt about the legalization of marijuana? Which variables help explain how people feel about marijuana? Which don't? Are some of these variables more strongly related to how people feel about marijuana? Which ones? How did you decide?

**Notes on Preparing your Paper:**

Double space the paper. Use one-inch margins and 12-point type. Number the pages. Organize the paper by parts. Use the part number as the heading for each section of the paper.

Be sure to check to make sure that you answered **all** parts of each section of the paper. Use the spell and grammar checker. Proofread your paper at least twice.

The papers will be read and graded by the instructor using several criteria. These criteria include the extent to which the papers show an understanding of how to construct and interpret tables and how to develop and write hypotheses and arguments. Other important criteria are the extent to which the paper is logically organized and the clarity and quality of your writing.

Remember that it is permissible for students to work together to get their tables from SDA but you **must** write your own answers to all parts of the paper. You **cannot** hand in a group answer. Answers that appear to be identical or nearly identical will **not** be accepted.

1. If the numbers in the table don’t match what you got in Part 7, don’t worry about it. We’ll talk about why this happened in class. [↑](#footnote-ref-1)