# RESEARCH METHODS 5RM - Hypotheses and Hypothesis Testing

Author:   Ed Nelson
Department of Sociology M/S SS97
California State University, Fresno
Fresno, CA 93740
Email:  ednelson@csufresno.edu

**Note to the Instructor:** This is the fifth in a series of 13 exercises that were written for an introductory research methods class.  The first exercise focuses on the research design which is your plan of action that explains how you will try to answer your research questions.  Exercises two through four focus on sampling, measurement, and data collection.  The fifth exercise discusses hypotheses and hypothesis testing.  The last eight exercises focus on data analysis.  In these exercises we’re going to analyze data from one of the [Monitoring the Future Surveys](http://monitoringthefuture.org/) (i.e., the 2017 survey of high school seniors in the United States).  This data set is part of the collection at the Inter-university Consortium for Political and Social Research at the University of Michigan.  The data are freely available to the public and you do not have to be a member of the Consortium to use the data.  We’re going to use SDA (Survey Documentation and Analysis) to analyze the data which is an online statistical package written by the Survey Methods Program at UC Berkeley and is available without cost wherever one has an internet connection.  A weight variable is automatically applied to the data set so it better represents the population from which the sample was selected.  You have permission to use this exercise and to revise it to fit your needs.  Please send a copy of any revision to the author so I can see how people are using the exercises. Included with this exercise (as separate files) are more detailed notes to the instructors and the exercise itself.  Please contact the author for additional information.

[This page](http://ssric.org/files/Research_Methods_5RM.docx) in MS Word (.docx) format is attached.

**Goal of Exercise**

The goals of this exercise are to teach students how to write hypotheses and to provide an introduction to hypothesis testing.[[1]](%22%20%5Cl%20%22_ftn1%22%20%5Co%20%22)

**Part I—Hypotheses**

All research starts with a question.  We could ask why some high school seniors decide to go on to college and others decide not to.  Or we could ask why some people vote Republican in presidential elections and other vote Democrat.  Our goal is to answer these questions.

Theories are systematic attempts to answer these questions.  Theories are made up of the concepts that we think will be helpful in answering these questions and propositions that indicate how these concepts are interrelated.  For example, the French sociologist Emile Durkheim wanted to know why some groups had higher suicide rates than other groups.  He argued that the concepts of integration and regulation helped explain variation in suicide rates.

Concepts are abstract ideas.  There are many different concepts used in sociology including integration and regulation as well as others such as religious preference and religiosity.  Since concepts are abstract ideas, we must find ways to measure these concepts.  Measurement is the process of finding pieces of empirical data that become our measures or indicants of the concepts.  For example, we might ask people how often they attend worship services or how often they pray and use their answers to these questions as measures of their religiosity.

We test our theory by deriving hypotheses from the theory which should be true if the theory is true and then testing these hypotheses.  For example, one of the questions that Durkheim asked was why some religious groups had higher suicide rates than other groups.  The degree of integration of the group provided one possible answer to this question.  Durkheim suggested that extremes of integration lead to higher suicide rates.  Thus, groups that were very high or very low in integration should have higher suicide rates that groups that were more in the middle on integration.  Since Protestants have lower levels of integration than Catholics or Jews, he hypothesized that Protestants should have higher suicide rates.  His hypothesis specified the relationship that he expected to find between religion and suicide rates.  In other words, a hypothesis specifies the relationship that you expect to find between your measures.  These measures are often referred to as variables.  A hypothesis must be testable.  In other words, it must be capable of being shown to be false.

Let’s think about another example.  Suppose our research question is why some people vote Republican in presidential elections and other vote Democrat.  Our theory is that those with less economic power are more likely to support political parties that attempt to change the status quo, while those with more economic power are more likely to support parties that attempt to maintain the status quo.  Our concepts are economic power and support for political parties.  In the United States, family income is one possible measure or indicant of economic power and the Democratic Party is more likely to want to change the status quo than is the Republican Party.  If our theory is true, then those with less income will be more likely to vote Democrat, while those with more income will be more likely to vote Republican.

Now we need to either collect data or find existing data to test our hypothesis.  If our hypothesis is false, then there is something wrong with the theory from which it was derived.  If our hypothesis is true, then we have support for our theory.  It’s important to keep in mind that we can’t say we have proven our theory.  Instead, we say we have support for our theory.  That’s a big difference.

**Part II – Now It’s Your Turn**

Our research question is why some children do better academically than other children.  Our theory suggests that media use might help explain academic performance.  Our reasoning is that the more time children spend accessing different types of media, the less time they will have available for academics, and the more likely they will be to do poorer academically.  We recognize that there are different types of media – television, radio, newspapers, magazines, and the internet.  We select a random sample of children in our state and ask their parents to complete a survey that asks several questions about the number of hours per week their children spend accessing various types of media.  The survey also asks questions about the grades their children get in school.  We decide to use grades as our measure of academic performance even though we realize it is not a perfect measure.

Write one possible hypothesis that specifies the relationship you would expect to find between the number of hours per week that children watch television and their grades in school.  Now write another hypothesis that specifies the relationship you would expect to find between the number of hours per week that children spend on the internet and their grades.

**Part III – Monitoring the Future Survey of High School Seniors**

In this exercise we’re going to consider the Monitoring the Future Survey of high school seniors in the United States that has been conducted yearly since 1975.  There is a [website](http://monitoringthefuture.org/) that will give you a lot of information about this study.  Here’s a brief description from the website’s home page.

“Monitoring the Future is an ongoing study of the behaviors, attitudes, and values of American secondary school students, college students, and young adults. Each year, a total of approximately 50,000 8th, 10th and 12th grade students are surveyed (12th graders since 1975, and 8th and 10th graders since 1991). In addition, annual follow-up questionnaires are mailed to a sample of each graduating class for a number of years after their initial participation.”

A focus of these surveys is students’ drug use.  There were three questions asked to measure marijuana and hashish use.

* “On how many occasions (if any) have you used marijuana (grass, pot) or hashish (hash, hash oil) . . . in your lifetime?
* “On how many occasions (if any) have you used marijuana (grass, pot) or hashish (hash, hash oil) . . . during the last 12 months?
* “On how many occasions (if any) have you used marijuana (grass, pot) or hashish (hash, hash oil) . . . during the last 30 days?”

Our research question is why some high school seniors are more likely to use marijuana or hashish than others.  That’s what we’re trying to explain.  What concepts do you think might help us answer this question?  Your list probably includes some of the following variables.

* Respondent’s gender – male, female
* Where respondents grew up – farm, small city or town, medium-sized city, suburb, large city
* Religiosity
	+ ​How often attended religious services – never, rarely, once or twice a month, about once a week or more
	+ How important religion is in their lives – not important, a little important, pretty important, very important
* Family factors such as parent’s education
	+ Highest grade father completed – less than high school, completed high school, some college, completed college, graduate or professional school after college
	+ Highest grade mother completed – same categories
* Political factors
	+ Political party preference – strongly Republican, mildly Republican, mildly Democrat, strongly Democrat, independent, no preference
	+ Political beliefs – very conservative, conservative, moderate, liberal, very liberal
* Number of days of school missed because respondent skipped or “cut” school
* High school grades

Choose two of these variables and for each variable write a hypothesis that specifies the relationship that you expect to find between that variable and marijuana and hashish use.  For each hypothesis explain why you expect to find that relationship.  In other words, if someone asked why you expected this hypothesis to be true, what would you say?

**Part IV – Hypothesis Testing**

Let’s review.  We start with our research question.  A theory is a systematic attempt to answer that question.  From our theory we derive hypotheses that should be true if the theory is true.  These hypotheses must be empirically testable.  We collect or find data that we can use to test these hypotheses.  So now we turn to the process of hypothesis testing.

How do we go about testing our hypotheses?  We can distinguish between quantitative and qualitative data analysis.  Our focus in these exercises is on quantitative analysis.  In no way does this suggest that quantitative analysis is better than qualitative analysis.  It also doesn’t suggest that it’s an either/or proposition.  Researchers often combine quantitative and qualitative analysis in a research study.

Quantitative analysis creates measures or indicants of the concepts in our theory.  These measures can then be related to each other.  For example, we can create a measure of religiosity based on how often respondents say they attend worship services and how important they say their religion is to them.  We can also create a measure of how often respondents say they use marijuana or hashish based on the questions that we discussed in part 3.  Based on these measures we can then determine if these two measures are related to each other.  In other words, are respondents who are more religious more or less likely to use marijuana or hashish than those who are less religious?

Statistics are tools that we use to answer these questions.  The next eight exercises will give you practice using various statistical techniques such as crosstabulation, tests of significance, and measures of association.  Statistical programs such as SDA (Survey Documentation and Analysis) carry out the computation of these different statistics so you don’t have to worry about computing them.  SDA is easy to learn and use.  You don’t have to be a computer programmer to use it. There are a few simple rules to follow which we’ll discuss in these later exercises.

[[1]](%22%20%5Cl%20%22_ftnref1%22%20%5Co%20%22) In these exercises we’re going to focus on survey research.