# RESEARCH METHODS 1RM -  Research Design

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 first 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 2015 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_1RM.docx) in MS Word (.docx) format is attached.

## Goals of Exercise

The goals of this exercise are to introduce the idea of a research design and to explore the basic elements of any research design – sampling, measurement, data collection, and data analysis.

### Part I—Research Questions and Research Design

All research starts with one or more research questions.  These are the questions that you want to answer in your research study.  For example, you might want to find out why some people vote Democrat and others vote Republican.  Or you might want to find out why some people don’t vote at all.  Another question you might want to try to answer is why some favor same-sex marriage and others oppose it.

There are lots of ways that we might go about trying to answer these questions.  Some might rely on what their friends or family tell them.  Others might rely on what people in authority like their religious leaders tell them.  Still others might use what is often called common sense to answer these questions.  But we’re going to use the scientific approach to try to answer these questions.  Thomas Sullivan defined science as a “method of obtaining knowledge about the world through systematic observations.”[[1]](%22%20%5Cl%20%22_ftn1%22%20%5Co%20%22)  Notice that science is empirical; it’s based on observations.  Also, notice that we’re talking about a particular type of observations – systematic observations.

A research design is your plan of action.  It lays out how you plan to go about answering your questions.  The research design includes how you plan to select the cases for analysis (sampling), how you will measure concepts, how you plan to collect your data, and how you will analyze the data.  Exercises two through five focus on the components of a research design and exercises six through thirteen deal with data analysis.

First, we have to learn how to formulate good research questions.  Let’s start by looking at some examples of poor questions.  Why are these poor questions?

* Women are more likely than men to vote Democrat in presidential elections.  This one is easy.  It’s not a question.  It’s actually a hypothesis which we will discuss in exercise [5RM](http://www.ssric.org/node/627).
* Why are women more likely than men to vote Democrat in presidential elections?  This one is a little more difficult.  We want to start with the more general question such as why some people vote Democrat and others vote Republican?  Then we would consider possible answers to this question.  One of these answers might be that gender influences voting.  Since science is empirical, we would start by looking at data to see if, in fact, gender does influence voting and we would discover that in most recent presidential elections women are more likely to vote Democrat.  This would lead us to ask why women are more likely than men to vote Democrat.  But we would start our study with the more general question.
* Why do dogs bark?  This is certainly a question and perhaps an interesting question.  But it’s not a question that social scientists would be interested in.  Social scientists focus on questions that involve behavior, attitudes, and opinions.[[2]](%22%20%5Cl%20%22_ftn2%22%20%5Co%20%22)

What are the characteristics of a good research question?

* We start by looking at general questions such as what influences voting or why do some people favor same-sex marriage and others oppose it.  As our study progresses, we move to more focused questions such as why women are more likely to vote Democrat than men.
* We focus on questions that ask about behavior, attitudes, and opinions.
* Good questions are clearly stated.  Questions such as what about voting aren’t clear and therefore aren’t useful.
* As with everything we write, we want to make sure that we use correct spelling and good grammar.  So proofread everything you write including your questions.

### Part II – Now It’s Your Turn

Write three research questions that could guide the beginning of a research study.  They can deal with any subject matter that asks about the behavior, attitudes, and opinions of people.  Be sure to follow the guidelines for writing good questions discussed in part 1.

### Part III – Sampling

Populations are the complete set of individuals that we want to study.[[3]](%22%20%5Cl%20%22_ftn3%22%20%5Co%20%22)  For example, a population might be all the individuals that live in the United States at a particular point in time.  The U.S. does a complete enumeration of all individuals living in the United States every ten years (i.e., each year ending in a zero).  We call this a census.  Another example of a population is all the students in a particular school or all college students in your state.  Populations are often large and it’s too costly and time consuming to carry out a complete enumeration.  So what we do is to select a sample from the population where a sample is a subset of the population and then use the sample data to make an inference about the population.

There are many different ways to select samples.  Probability samples are samples in which every individual in the population has a known, non-zero, chance of being in the sample (i.e., the probability of selection).  This isn’t the case for non-probability samples.  An example of a non-probability sample is an instant poll which you hear about on radio and television shows.  A show might invite you to go to a website and answer a question such as whether you favor or oppose same-sex marriage.  This is a purely volunteer sample and we have no idea of the probability of selection.

There are a number of different ways of selecting a probability sample.

* The most basic type of probability sample is the simple random sample where every individual in the population has the same chance of being in the sample.
* Samples can also be stratified.
	+ A proportional stratified random sample is one in which the sample is selected such that the sample has the same proportion on key variables as does the population.  For example, 51% of the nation is female and 49% is male.  The sample could be stratified on sex in such a way that 51% of the sample is female and 49% is male.
	+ A disproportional stratified random sample is one in which the sample is selected such that we oversample some segments and undersample other segments of the population.  For example, we might undersample whites and oversample non-whites so that our sample is 50% whites and 50% non-whites.  This would be useful if we wanted to compare whites and non-whites and wanted to have a larger sample of non-whites for comparison purposes.

Notice that simple random samples and stratified random samples assume that we have a list of the population from which to select our sample. But what if we don’t have such a list?  For example, how would we get a sample of high school seniors?  There is no list available.  But there is a list of all high schools in the United States.  So we could select a sample of high schools and then within each high school in our sample select a sample of seniors.  This is called a cluster sample because high schools are the clusters where you find seniors.

No sample is ever a perfect representation of the population from which the sample is drawn.  This is because every sample contains some amount of sampling error.  Sampling error in inevitable.  There is always some amount of sampling error present in every sample.  The question then is how can we reduce sampling error?

* One way is to increase the sample size.  The larger the sample size, the less the sampling error.  A simple random sample of 400 will have half the sampling error that a simple random sample of 100 has.  To reduce the amount of sampling error by half, you have to quadruple the sample size.
* Stratifying a sample is another way that you can reduce sampling error assuming that the variable you use to stratify the sample is related to whatever you are studying.  For example, if you are trying to explain why some people favor same-sex marriage and others oppose it, then you could stratify your sample by sex.  Assuming that sex is related to how people feel about same-sex marriage (and it is), this will reduce sampling error.

Sampling is an important component of any research design.  You need to carefully think about how you plan to select the cases for your research study.  Exercise [2RM](http://www.ssric.org/node/623) will explore sampling in more detail and give you practice in constructing a sampling design.

### Part IV – Measurement

Let’s say that we want to explain support or opposition to same-sex marriage and that we think religion might be related to how people feel about same-sex marriage.  We can distinguish between two different dimension of religion – religious preference and religiosity.  That means that we’re dealing with three different concepts.  Our concepts are:

* support or opposition to same-sex marriage,
* religious preference, and
* religiosity.

Concepts can be defined as the abstract ideas that we want to use in our study.  Another way to think about concepts is to view them as the tools we’re going to use to try to answer our research questions.  Imagine that you go to the dentist.  The dentist has a lot of tools to take care of your teeth but not all tools are appropriate.  A chain saw is a tool but you wouldn’t want to see a chain saw in your dentist’s office.

Concepts have to be defined.  There are two different ways to define concepts.

First, there is the theoretical definition.  This answers the question – what do we mean by these concepts.

* Religious preference refers to the religion with which a person identifies.  For example, some people identify themselves as Roman Catholic, others as Lutheran, others as Jewish, and still others as Muslim.
* Religiosity refers to how religious a person is.  Two individuals could identify themselves as Roman Catholic but one might be much stronger in their religion than the other.
* Opposition or support for same-sex marriage is obvious.  Do people define themselves as favoring or opposing same-sex marriage?

Second, there is the operational definition.  How do we measure these concepts?  What are the operations we go through to measure the concepts?

* Religious identification could be measured by asking people a question such as the following:  “What is your present religion, if any? Are you Protestant, Roman Catholic, Mormon, Orthodox such as Greek or Russian Orthodox, Jewish, Muslim, Buddhist, Hindu, atheist, agnostic, something else, or nothing in particular?”[[4]](%22%20%5Cl%20%22_ftn4%22%20%5Co%20%22)
* Concepts can be measured in different ways.  Religiosity could be measured by asking people how often they attend religious services, how often they pray, and how important their religion is to them.  Here are some questions that have been used in different surveys?
	+ "Aside from weddings and funerals, how often do you attend religious services... more than once a week, once a week, once or twice a month, a few times a year, seldom, or never?”[[5]](%22%20%5Cl%20%22_ftn5%22%20%5Co%20%22)
	+ “About how often do you pray?”  Categories are several times a day, once a day, several times a week, once a week, less than once a week, never.[[6]](%22%20%5Cl%20%22_ftn6%22%20%5Co%20%22)
	+ “Would you call yourself a strong [insert religious preference] or a not very strong [insert religious preference]?[[7]](%22%20%5Cl%20%22_ftn7%22%20%5Co%20%22)
* Here’s a question from the 2014 Pew Political Polarization Survey that was used to measure how people feel about same-sex marriage.  “Do you strongly favor, favor, oppose, or strongly oppose allowing gays and lesbians to marry legally?”

Your research design should identify the concepts that you want to use in your study and both your theoretical and operational definitions of these concepts.  Exercise [3RM](http://www.ssric.org/node/624) will explore measurement in more detail and give you practice in developing measures of various concepts.

### Part V – Data Collection

Science is an empirical enterprise.  That means that it is data based.  There are two ways that we collect data[[8]](%22%20%5Cl%20%22_ftn8%22%20%5Co%20%22):

* we observe people and use our observations as data, and
* we ask people questions and use what they tell us as data.

We’re going to focus on survey research in these exercises. Sometimes surveys are referred to as sample surveys because we select a sample of individuals from the population and ask them questions.  Then we use their answers to these questions as our data.  Surveys can take various forms:

* in-person interviews,
* mailed questionnaires,
* telephone surveys, and
* web-based surveys.

Error is inevitable whenever we study something.  Since we can’t eliminate all error our goal is to minimize error.  Error can enter into a survey in various ways.

* Sampling error occurs when we select a sample from a population.  No sample is a perfect representation of a population.
* Coverage error occurs when the list of the population from which we select our sample does not perfectly match the population.  For example, about 98% of all households in the United States have a telephone (either landline or cell or both).  So when we do a telephone survey we fail to cover about 2% of all households.
* Nonresponse error occurs when we fail to reach the entire sample.  This type of error can occur in two ways – refusals and the failure to contact some individuals in the sample.
* Measurement error occurs when our measures of some concept fall short in some way.  For example, the way we word our survey questions can introduce error.  It turns out that it matters a great deal whether we refer to global warming or climate change when we ask people questions.

Our survey design should clearly describe how we plan to collect our data.  We should consider the different ways that error might enter into the data and how we will try to minimize that error.  Exercise [4RM](http://www.ssric.org/node/626) will explore survey design in more detail and give you practice in constructing questions.

### Part VI – Data Analysis

Once we have our data, then we want to analyze the data in such a way that we can begin to answer our research questions.  Exercises 6RM through 13RM will explore data analysis and give you practice in analyzing survey data.  We’ll have much more to say about data analysis in these exercises.

Typically we start by looking at variables one at a time (i.e., univariate analysis).  We can use various statistical tools such as frequency distributions, measures of central tendency, and measures of dispersion to help us describe variables.

Then we look at relationships between pairs of variables (i.e., bivariate analysis).  We’re going to use crosstabulation and Chi Square to help us explore these relationships.  From here, we look at sets of variables (i.e., multivariate analysis) to see what they can tell us.

One very important point to consider is the question of causality.  Survey design can never give us a complete picture of the causal patterns in our data but it can help us begin to tease out what these causal patterns might look like.

We’ll come back to data analysis in exercises 6RM through 13RM.

### Part VII – Research Study We’ll Be Using

The research study that we’ll be using in these exercises is 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 major focus of these surveys is students’ drug use.  But the surveys include a lot more information than just drug use.  The website describes the range of questions asked.

“Questions include drug use and views about drugs, delinquency and victimization, changing roles for women, confidence in social institutions, concerns about energy and ecology, and social and ethical attitudes.”

These are only a few of the areas that students are asked about.  Other areas include, for example, their educational goals, religion, politics, the military, race, health, and background information including their family.

Questions about drug use include a variety of questions including several questions about alcohol use.  These include questions about:

* whether the respondents had ever consumed alcohol,
* how often they drank over their lifetime, in the last twelve months, in the last 30 days,
* how often they had consumed alcohol to the point of feeling “pretty high,” and
* how often during the last two weeks they had consumed “five or more drinks in a row” which is a common definition of binge drinking.

### Part VIII – Now It’s Your Turn Again

In part 2 you were asked to write three research questions that could guide the beginning of a research study.  This time write three questions that relate specifically to drinking alcoholic beverages.  Think about what you want to find out about drinking by high school students.  For example, we might ask whether males or females are more likely to binge drink.  (Don’t use this example as one of your three questions.)

[[1]](%22%20%5Cl%20%22_ftnref1%22%20%5Co%20%22) Thomas J. Sullivan.  1992. Applied Sociology. New York: Macmillan Publishing Company (p. 10).

[[2]](%22%20%5Cl%20%22_ftnref2%22%20%5Co%20%22) These are not the only things that social scientists are interested in.  For example, we might also study businesses or nations.  We’re going to focus in these exercises on individuals and their behavior, attitudes, and opinions.  In other words, our unit of analysis will be the individual.

[[3]](%22%20%5Cl%20%22_ftnref3%22%20%5Co%20%22) Just a reminder that we need not limit ourselves to studying individuals.  We could also study objects like businesses or nations.  So it might be better to define a population as the complete set of objects that we want to study.  But in these exercises our focus is on individuals so we’ll define a population as the complete set of individuals we want to study.

[[4]](%22%20%5Cl%20%22_ftnref4%22%20%5Co%20%22) This is the question used in the Pew Research Center’s Political Polarization Survey conducted in 2014.  More information on this survey can be found on their [website](http://www.pewresearch.org/fact-tank/2014/06/12/7-things-to-know-about-polarization-in-america/).

[[5]](%22%20%5Cl%20%22_ftnref5%22%20%5Co%20%22) This is also from the 2014 Pew Political Polarization Survey referred to in footnote 4.

[[6]](%22%20%5Cl%20%22_ftnref6%22%20%5Co%20%22) This question is from the 2014 General Social Survey conducted by the National Opinion for Social Research.  More information about this survey can be found on their [website](https://gssdataexplorer.norc.org/).

[[7]](%22%20%5Cl%20%22_ftnref7%22%20%5Co%20%22) This question is also from the 2014 General Social Survey referred to in footnote 6.

[[8]](%22%20%5Cl%20%22_ftnref8%22%20%5Co%20%22) Matilda White Riley. Sociological Research I: A Case Approach. 1963. New York: Harcourt, Brace & World (pp. 184-190).