Nursing Research

Supplemental Text

Patrick Heyman

Palm Beach Atlantic University

© 2007

Table of Contents

Chapter 1
Guide to Basic Research Terms

Research typically asks how two or more things relate. Examples:

1. What is the relationship between handwashing and nosocomial infections?
2. What is the relationship between latex exposure and latex allergy?
3. What is the relationship between injection pain and needle size?
4. What is the relationship between pneumonia and oral hygiene?
5. What is the relationship between gestational diabetes and birthweight?

One of these two things is called the response variable, and the other is called the study factor. To determine which is which ask yourself whether the relationship is:

1.	handwashing □hosocomial infections	or	nosocomial infections ●handwashing
2.	latex exposure ●latex allergy	or	latex allergy ●latex exposure
3.	injection pain ●needle size	or	needle size ●injection pain
4.	pneumonia ●oral hygiene	or	oral hygiene ●pneumonia
5.	gestational diabetes ●birthweight	or	birthweight ●gestational diabetes

Which ever you would put first in the relationship is the study factor. The second one is the response variable. The idea is that the response variable should *respond* to the factor. Another way of saying it is that research asks whether the factor affects the response variable. Sometimes the factor is called the independent variable and the response variable is called the dependent variable. In this case, the assumption is that the dependent (response) variable will change *depending* on the independent variable (which is assumed to be able to do whatever it wants or perhaps what the researcher wants).

Factors and response variables do not occur in a vacuum. They occur on or in something. In medical research, the something is usually a person. The people that the factors and response variables occur in are called your universe (population) of interest. *So the three things we need to know for any given research are study factors and response variable and the population of interest.* Let's look at each one in more detail.

Population of Interest

Experimental Unit

Before we can talk about the population of interest, we need to talk about the experimental unit. When you study something, you will have to study an object or group of objects (population of interest). The object you study is called an ***experimental unit***. In medical and nursing research, the experimental unit is usually a person. But it may be a group of people such as a family or neighborhood. Or it may be a part of a person, such as an injection site. And finally, the experimental unit may be an inanimate object such as a chart or a glove. If the experimental unit is a person, they are often referred to as subjects. And since many human subjects in medical research are also patients, you may hear them referred to as patients.

Population of Interest

Most of time, when talking about research, we don't talk in terms of experimental units. We usually talk about the group of people we are interested in, the so-called population of interest. In our first example

above, by definition, nosocomial infection means that the people we are interested in must be hospital patients, because nosocomial infections are infections that patients get in the hospital. However, we may not be interested in all hospital patients. We may only be interested in studying nosocomial infection in women who have just given birth, or patients who have HIV, or patients who are in our hospital. The particular patients we are interested in is called our population of interest.

Technically speaking, the population of interest is the collection of all experimental units. The reason this is important is because there is often a mismatch between what researchers say is their population of interest and their experimental units. They may say that they their population of interest is all patients who have just given birth, but exclude experimental units (women who have just given birth) who had twins... or caesarean deliveries... or had pre-term deliveries. It is important when critiquing or planning research to make sure that the experimental units chosen (included) and those rejected (excluded) match the stated population of interest.

Side note: Even though population of interest is the term most often used is is technically incorrect. When talking with a statistician, it is useful to know the difference. Technically speaking, the universe is the collection of experimental units, while the population is the collection of *measurements* of the response variable in the experimental units. Again, it's a technical difference, and the only time it will matter is if you are talking to a statistician, so don't worry about it (i.e., you will not be tested on the distinction or asked to apply the difference).

Sample

Most of the time it will be impractical or impossible to study all of the experimental units in the population of interest. Usually, the study is done in a small selection of experimental units called the universe sample, population sample, or study sample. The idea is that by studying the factors and response variables in a few experimental units, one may make an inference (reasoned conclusion) about the population. In order to make sure that your inference is correct, the sample must be *representative*. The methods of selecting the sample are discussed in a later chapter.

Inclusion and Exclusion Criteria

Most studies will formally define their population of interest with the use of inclusion and exclusion criteria. Inclusion criteria are any characteristics a potential experimental unit must have in order to be included in the study. For example, when studying women who have given birth recently, the inclusion criteria might be: must have given birth within the past three months. Exclusion criteria are any characteristics that would cause a potential experimental unit to be excluded from the study. For example, in our study of women who have given birth, we might exclude women who had HIV or women who gave birth at home.

It is important to remember that inclusion and exclusion criteria do not simply determine whether a patient may participate in a study. They actually define the population of interest. One of the most common errors in research, even among "top researchers" is to make inferences for a population of interest that was not actually studied. For example, a researcher may claim to have studied the relationship between smoking and depression in the elderly, but includes subjects that are 55 and up. Many people may not think that people between 55 and 60 are elderly. Alternatively, the researchers may have excluded patients in nursing homes when in fact many elderly do live in nursing homes.

When critiquing or planning research care must be taken to make sure that the stated population of interest and the inclusion/exclusion criteria match.

Response Variable (Dependent variable, outcome, outcome variable)

We have already introduced the concept of response variable, but let's define it a little more formally. The response variable is the observation of whatever phenomenon you are trying to study. What that means is that if we are studying nosocomial infections, we have to actually have some way of measuring infection. This seems fairly straightforward, but think about mothers and children for a moment. "Mom, I don't feel good. I'm sick. Can I stay home from school?" "You don't have a fever, and you haven't thrown up. You're not sick. You have to go to school."

So our response variable is not simply the concept we are studying, it is ***how we measure it***. In the case of infection, we could record the patient's temperature, we could take white blood cell counts, we could ask them how they feel, we could swab and culture their wounds, their sputum, their urine, etc. We could do a combination of all those things. Ultimately, no matter how a response variable is measure, the researcher will do one of two things: put it into categories or record a number. We will study these classifications in further detail in chapter 2.

Before going on, try to think of at least three ways each of the following concepts could be measured.

- latex allergy
- injection pain
- pneumonia
- birthweight

Some of the ways you came up with will be better than others. Some will be more practical, some will cost more or less; some will just be plain silly. The point of the exercise is to emphasize that researchers have several options at their disposal. When you critique research, ask yourself whether the researcher chose an appropriate way of measuring the response variable.

Factors

Factors are things that potentially influence the response variable. In nursing and healthcare, we often evaluate how a treatment or intervention influences our response variable. For example, giving oral hygiene to an intubated patient is an intervention that may or may not influence whether that patient gets pneumonia. In order to be a factor, there must be at least two possible levels or settings for the treatment or intervention. In the example above, the levels might be:

- Giving oral hygiene or not giving hygiene
- Three different brands of mouth wash
- Four different strengths of the same mouthwash
- The number of seconds suctioning is performed

Qualitative factors are factors where the levels are categorical. A quantitative factor is any factor that is numerical in nature. In the four examples above, which are quantitative and which are qualitative?

Quantitative factors may be treated as qualitative factors during analysis, but information is lost. For example, rather than treating four strengths of the same mouth wash as numerical, one could evaluate them as though they were four different brands of mouth wash. So why would someone do that? Usually, the answer is lack of rigor, ignorance, or simply not thinking ahead.

In addition to treatments and interventions, researchers may want to know what factors influence the course of disease. Any factor thought or found to impact disease is called a risk factor.

Sidenote: You may sometimes hear about positive and negative risk factors, but usually what is meant is that the level of the risk factor is positive or negative. For example, activity level has been determined to influence coronary artery disease (CAD). Low levels of activity are associated with more CAD, while high levels of activity are associated with less CAD. Thus activity level itself is neither positive nor negative, but high and low levels of activity are positive and negative. Be careful that you understand that the popular terminology of positive and negative risk factors does not use the term factor in the correct way.

Extraneous Factors

In addition to the study factors, there are often other factors present in the experimental unit or environment that may also influence the response variable. For example, in our study of needle size and injection pain, what other things might influence pain? Don't be lazy and peek at my list yet. Come up with your own list.

When you're done with your list, turn the page.

Okay, here's my list of things that might influence injection site pain:

- Gender
- Age
- Sharpness of needle
- Smoothness of needle
- Type of injection (ID, SC, IM, IV)
- Type of liquid administered
- Amount of liquid administered
- Time of day
- Mood of subject
- Site of injection
- Diseases that predispose patients to numbness
- Diseases that predispose patients to sensitivity
- Pain drugs that patients may be on

The researcher doesn't really care about any of these things. He just wants to know about needle size. Thus all of these other factors are called **extraneous factors**. Even though researcher doesn't care about extraneous factors does not mean that they are not important. If the researcher does not adequately control the extraneous factors, he may find that he doesn't truly know whether it is the size of the needle that influences pain or whether it is gender or mood. In this case, gender and mood would be called confounding factors. The most dangerous position a researcher can be in is when he doesn't know that he has confounding factors. (In other words, **it is better to know you did poor research than it is to do poor research and think it is good**.) Good research should try to *control* extraneous factors to avoid confounding factors.

The Research Objective

The most important thing in any research project is the research objective--what the researcher is trying to do. When critiquing research, one must always try to determine what the research objective was. **Do not take what the authors of the article say at face value!!!** Authors are notorious for exaggerating their studies, or stating that they did something when in fact they did something completely different. As you read the article, sometimes you will find yourself asking, what is it they were trying to do? When this happens, it is usually a sign that the researchers didn't really know what they were doing.

Thus it is helpful to put the research objective into a standardized format. The standardized format should be expressed generically as:

- Evaluate/study/determine the relationship between **list of study factors** on **the response variables** in the **population of interest.**

Good Research Objectives

In class, we talked about the requirements to do real science. The most important requirement is intellectual honesty in not trying to make results come out a certain way. If a new drug is being tested against cancer, the appropriate research objective would be something like

- Determine the relationship between **our new experimental drug** on **cancer survival** in **patients with cancer.**

An inappropriate research objective would be something like

- Show that **our new experimental drug** makes **cancer patients live longer.**

The difference between the two is intention. It is perfectly okay to *want* the drug to make the cancer patients live longer, but that cannot be our research goal, because if we do enough studies, one or two of them will come out the way we want simply by chance. Our research objective must be free of bias. What is the difference between the two following research objectives?.

- Determine if lowering blood pressure saves lives.
- Show that lowering blood pressure saves lives.

Chapter 2
Study Types

Once we have determined our research objective, including factors and response variables, and the population of interest, the next most important decision for any researcher to make is the type of study he will perform. The reason study type is so important is that it will impact the researcher's ability to control extraneous factors and thereby reduce the chances of having a confounding factor.

There are two basic study types based on their relationship to time--prospective and retrospective. ***Prospective studies*** look forward in time; this means that the data does not yet exist. ***Retrospective studies*** look backward in time; this means that the data exists at the time the study is performed regardless of whether that data has actually been collected.

Prospective studies

There are two kinds of prospective studies, ***experimental (clinical trials)*** and ***observational (prospective cohort)***.

Clinical Trials (experimental)

The terms clinical trial and experiment are interchangeable, When experiments are done on humans, the term clinical trial is preferred, but they are really the same thing. Clinical trials are studies where the *researcher has control over the assignment of factors*. Because the researcher can control extraneous factors, it offers the researcher the best chance of reducing confounding factors. Note that the researcher cannot control whether a subject is male or female but does have control over which study factor each male and each female are assigned to.

The problem with clinical trials is that it can be difficult to find enough subjects who will let you manipulate the study factors in them. Moreover, they can create artificial situations that do not mirror real life.

A further problem with clinical trials is that sometimes just knowing that certain factor has been assigned, may influence how the subject responds. To get around this limitation, sometimes a clinical trial will be blinded--meaning that the subjects do not know what factor level they have been assigned to. For some people, this is not enough, because they think that if the researcher knows the subjects assigned factor level, that may influence the researcher to do things differently. Thus, sometimes both the researcher and the subjects do not know the assignment of factors. This is called ***double-blinded***. A clinical trial where both subject and researcher do know the factor level assignments is called ***open label.***

A final problem with clinical trials is one of ethics. It is unethical to assign patients to potentially harmful treatments in the name of science or conversely to withhold potentially life saving treatments. For example, it is very unlikely that an ethics board would allow a study that calls for the assignment of subjects who do not normally smoke to a smoking group. An ethics board may not allow a drug to be tested against placebo if other treatments are available. Thus, one will often see the study factor compared to "standard" or "conventional" treatment.

When critiquing clinical trials, the randomization process should be evaluated. Randomization is the technique used to assign study factors even though there may be no randomness involved. For many researchers the Randomized Clinical Trial is the epitome of all research studies and the only ones that should be considered. In the “evidence-based” movement, only clinical trials can achieve a grade of A.

Often for the sake of convenience, cost, or other mitigating reason, the researchers do not have a randomization process that ensures control of extraneous factors. For example, a researcher may choose one hospital unit for for one factor and another hospital unit for a second rather than assigning some of the patients from each unit to the factors. This is a non-randomized clinical trial. Nursing and the social sciences, in attempt at self-aggrandizement call these studies...***Quasi Experimental***.

Summary:

- The two criteria for a clinical trial are: 1) Data does not exist when the study starts; 2) The researcher has control over assignment of factors.
- Strengths: Most control over extraneous factors (and study factors).
- Weaknesses: Expensive, may be difficult to find enough study subjects.
- Open-label, single-blinded, and double-blinded issues
- Can run into ethics problems.
- What is the randomization method?

Example:

Research objective: What is the relationship between pneumonia and oral hygiene in nursing home patients?

Population of interest: all nursing home patients

Experimental unit: a nursing home patient

Response variable: pneumonia (yes or no)

Study factors: Number of minutes oral hygiene is performed by caregivers

In our clinical trial, the researcher will assign the number of minutes that caregivers will provide oral care for each subject. This allows the researcher to control extraneous factors such as age, gender, and neurological problems by making sure that each oral hygiene level is has a similar number of patients with these extraneous factors. After assigning the factors, the researcher would then look for pneumonia in the subjects during the study period.

Observational (Prospective Cohort)

An observational study is any prospective study where the researcher does not have control over factors, but merely observes the factors and response variable.

Using the same example as above, rather than assigning the number of minutes that oral hygiene is performed, the researcher would simply record what is already being done and then look for pneumonia during the study period.

Because the subject usually already knows what factors he has, blinding is usually not an issue, although it is possible that the researcher might react to subjects differently based on their factor level. For example, if the researcher is trying to study the effect of maternal smoking on birthweight, he may be ethically obligated as a health professional to counsel the subject to stop smoking and thus confound his own research study.

Summary:

- Two criteria are: 1) Prospective 2) The researcher does not assign factors.

- Advantages: are usually easier to conduct, less expensive, may be able to study more patients than with a clinical trial
- Disadvantages: less control over extraneous factors; may find that not enough experimental units have a factor level to study

Retrospective Studies

There are three retrospective study types--***historical (retrospective cohort)***, ***cross-sectional***, and ***case control***.

Historical (Retrospective Cohort)

The historical study uses existing data to answer the study question. In our oral hygiene/pneumonia example, rather than going out and looking for patients and following them through time, a much quicker method would be to simply look at a nursing homes records and simply look in the charts to see what kind of oral hygiene the patients had received and whether or not patients had pneumonia.

Historical studies have a number of advantages. They can usually be accomplished very quickly because the data already exists. They are usually much cheaper to conduct, and sometimes, the researcher may not have to have the subjects sign consent forms. The disadvantages however can be substantial. The first is that the researcher has no control over any factors whatsoever, as the data has already been collected. A more serious problem is that oftentimes, the data collected may be incomplete or not in the correct form. In our example above, some charts may be missing oral hygiene information altogether, and the ones that do have have oral hygiene documentation probably only have whether/when it was done, not the number of minutes provided (as specified by our study).

Any study that looks at charts, birth certificates, death certificates, or incident reports is probably historical

Summary:

- Criteria: Retrospective; looks at data that has already been collected.
- Advantages: Fast, inexpensive, fewer consent problems, can study a very large number of subjects
- Disadvantages: No control over factors; no control over amount or quality of information collected.

Cross-Sectional Survey

In a Cross-sectional survey, the experimental units are chosen and the factor levels and response variables are measured at the same time. It is considered retrospective because the data exists at the time the subject is chosen for the study even though it has not yet been collected.

Cross-sectional surveys can be very convenient and allow for a relatively large number of subjects to be studies. Most mailed or telephone surveys are cross-sectional.

Case-Control Study

The case-control study offers the least amount of control, and the most risk of drawing incorrect inferences. However, it is sometimes the only feasible study to answer some questions. Case-control studies approach the process backwards: subjects are identified by their response variable. They are then matched to "similar" normal control subjects, and then the factors are assessed. Case-control trials should generally be used when there are only a very few number of people with the desired response variable. For

example, bladder cancer is very rare. In order to study 500 people with bladder cancer, over 250,000 subjects would need to be selected. This is most likely cost and resource prohibitive for the other four study types we have examined. In a case control study, we would first select 500 subjects with bladder cancer, then we would select subjects who do not have bladder cancer to compare them to (normal controls). Ideally, the normal controls should be matched to potential extraneous factors.

Summary:

- Criteria: Cases who have the response variable result already are selected first, and the normal controls are matched to them.
- Advantages: Can be used to study phenomenon that otherwise would be very expensive because they are rare.
- Disadvantages: No control over anything. High risk of drawing incorrect inferences.

Distractors

For some reason research article authors sometimes seem determined not to correctly describe their research. If you are lucky, the author will simple say that they did a clinical trial or cross-sectional survey, etc. However, even if your author identifies the type of study, you must still carefully read the article and make sure that the author is using the terms correctly.

Here are some other terms that are often used to describe research that often confuse students. Note that these terms should never be reported in a critique in place of one of the five study types above.

Longitudinal

The word longitudinal simply means that the experimental units were followed for a long period of time. Unfortunately, "a long period of time" is relative, sometimes being as short as a year. Generally speaking though, a study must follow the experimental units for at least five years in order to be considered longitudinal. As the data is collected over time, longitudinal studies are prospective, but the term longitudinal could be used to describe both clinical trials and observational studies.

Population-based

Population-based simply means that experimental units were drawn from the population at large, not patients who had come to a hospital or care provider. Any kind of study can be population based.

Epidemiological

Epidemiological studies focus on health patterns in populations, generally meaning that they are population-based. Epidemiological studies are typically observational, cross-sectional, or historical although clinical trials and case-control studies may occasionally be used. An epidemiological study will usually have one of the following objectives:

- identifying the prevalence of a disease in the population of interest (proportion of the population that has the disease)
- identifying the incidence of a disease in the population of interest (proportion of the population that acquires the disease in a given period of time)
- identifying factors related to related to occurrence of a disease
- identifying changes in disease patterns over time

Chapter 3
Choosing a Response Variable

Remember that a response variable is a measurement that reflects the concept that we are studying. Sometimes, it will be quite apparent that there is one best way to measure the concept, but there may be several different ways to measure it. If you are planning the research, you should put some time into thinking about the best way to measure the concept you are interested in. If you are critiquing research, you should ask yourself whether the researchers chose appropriate measurements. For example, if you are studying nosocomial infections, you could record:

- Infection: present/not present
- Infection: pneumonia, cellulitis, pseudomembranous colitis, etc.
- Pathogen: Pseudomonas, clostridium difficile, Klebsiella, etc.
- Severity: Mild, Moderate, Severe, Life threatening
- Patient's temperature
- Patient's white blood cell count

Systolic blood pressure could be recorded:

- Normal or abnormal
- Low, normal, high
- mm of mercury

Before going on, list at least three ways each of the following concepts could be measured:

- latex allergy
- injection pain
- oral hygiene
- birthweight

As can be seen, there are quite a few options. The choice of response variable is very important and should be guided by two concerns:

- What are most appropriate response variables that could measure the physiological phenomenon or concept and achieve the research objectives?
- Which response variables can realistically be collected

A third consideration is the response variables that other researchers have chosen in past studies. This should not be a primary concern, but you should learn from their experiences, especially if you want to compare your results to those of other researchers. The literature review is vital in this regard, but keep in mind what has been done before might not be as good, although you might find it is better than what you had planned.

Parameters and Statistics

A **parameter** is a single number that summarizes the entire *population of interest*. A **statistic** is a single number that describes the *study sample*. For example, if a researcher wanted to know the prevalence of teen pregnancy in the United States, he might sample teens from across the country. The prevalence of teen pregnancy in the researcher's sample is a statistic. Using that statistic, the researcher will try to estimate the prevalence of teen pregnancy for the entire country. The true prevalence of teen pregnancy in the United States would be a parameter. Thus, we can see that a sample statistic is used to estimate the population of

interest's parameter. And that is why it is so important that a sample be representative of the population of interest. (Parameters and statistics have the same name but different symbols.)

Classifications of Response Variable

Ultimately, there are only two ways to measure a response variable. You can put it into categories, or you can record a number (just like factors). However, unlike factors, there are two subclassifications for each one. We will examine the categorical classifications first. In our example of nosocomial infections, we could have a number of classifications. The simplest is simply to measure whether the patient had an infection or not. Alternatively, one could record the type of infection, or one could record the type of pathogen, or one could record the severity of the infection.

Nominal Response Variables

Any time a the categories of a response variable cannot be put into any particular order, it is a *nominal* response variable. Thus, in our bulleted examples above, the first three are nominal variables. There is no way to rank the various categories, therefore, they are nominal. The simplest (and most common) nominal response variables have only two values: present/not present; yes/no; dead/alive. However, sometimes, a nominal response variable has more than one category; for instance, the Florida Department of Vital Statistics recognizes seventeen causes of death. The only restriction on the number of possible categories is that there must be enough data in each category to make analysis meaningful.

The parameter/statistic for a nominal response variable is the proportion of experimental units that fall into each category, and is often reported as a percentage. Examples:

Response Variable	**Categories**	**Parameter/Statistic**
Mortality	dead/alive	proportion of subjects who died
Morbidity	hospitalized/not hospitalized	proportion of subjects hospitalized
Prevalence	have disease/don't have disease	proportion of people with disease
Incidence	contracted the disease this time period/did not contract the disease this time period	proportion of those who contracted the disease this time period
Satisfaction	satisfied/not satisfied	proportion of those who are satisfied
Competence	competent/not competent	proportion of those who are competent
Birth control method	nothing/pill/diaphragm/condom/rhythm/etc.	proportion of those who use nothing/pill/diaphragm/condom/rhythm/etc.

Please note that for binary (two category) response variables, only one proportion is needed because the other can be calculated by subtracting from 100%. (If the death rate is 12%, then the survival rate is 100% - 12% = 88%.)

Ordinal Response Variables

If the categories can be put into an order, then the response variable is called *ordinal*. The two keys to an ordinal response variable are 1) the categories must be comprehensive; and 2) the order must be universally agreed upon. Examples:

Response Variable	**Categories**
Satisfaction	Very Satisfied > Satisfied > Neutral > Unsatisfied > Very Unsatisfied
Blood Pressure	Low > Normal > Prehypertension > Hypertension 1 > Hypertension 2

Because the categories can be ordered, sometimes ordinal response variables are referred to as *ranked data*. Sometimes, numbers are assigned to the categories to help keep track of the rank. The important thing to remember in this case is the *order; the numbers themselves are arbitrary and meaningless*. Example:

Response Variable	**Categories**				
Satisfaction	Very Satisfied > Satisfied > Neutral > Unsatisfied > Very Unsatisfied				
	1	2	3	4	5
	0	2	300	300.1	1,000,000,002

In the table above, the only thing that matters is which category comes before the others (the order). When reading research, a common mistake is to pretend that the numbers have true meaning and analyze the data as though it were a true numerical response variable.

It must also be recognized that many times, the categories themselves may be arbitrary. In such a case it becomes necessary to ask whether additional information is truly being gained by the addition of categories. Which of the following do you think is most likely to accurately assess satisfaction?

Response Variable	**Categories**
Satisfaction	Very Satisfied > Satisfied > Neutral > Unsatisfied > Very Unsatisfied
	Satisfied > Neutral > Unsatisfied

Adding additional categories does not necessarily increase the ability to measure the phenomenon. How often have you heard a patient asked to rank their pain on a scale of 0 to 10? Would it have made a difference if they had ranked it from 0 to 5, or 0 to 100? Adding numbers to an arbitrary scale *does not make the scale more precise*.

Two types of ordinal response variables often used in clinical research are the Likert Scale and the Visual Analog Scale.

Likert Scale

Response Variable	**Categories**				
Satisfaction	Very Satisfied > Satisfied > Neutral > Unsatisfied > Very Unsatisfied				
	1	2	3	4	5

Visual Analog Scale:

Response Variable	Categories
Satisfaction	0----------X----------5
	0----------X----------10
	0----------X----------100

The subject is then asked to mark where on the scale their satisfaction level lies. A ruler is then used to determine the corresponding number. Note that as stated above, adding a larger number to the scale does not make it more accurate. Moreover, you may be saying to yourself, "If someone asked me to mark my satisfaction on a scale like that, I'd have no idea what to put." This brings us to a very important problem with ordinal response variables. How do we know that the ordinal response variable is any good at actually measuring what we want it to?

There are two terms that are used to describe how well an ordinal response variable measures what we are trying to measure: reliability and validity. Both are usually reported as either a decimal or a percentage. **Reliability** indicates how good the response variable is at getting the same response each time. If a researcher was researching pain in children, and asked them to mark their pain several times on a visual analog scale, and each time the child marked a different spot, then the response variable would have a low reliability.

Low reliability

Response Variable	Categories
Pain	0----------X----------5
	0----------X-----5
	0----------X----------5
	0----------X----------5

If the researcher then tried the same scale on an adult, and the adult consistently chose similar spots, the response variable would have a high reliability

High Reliability

Response Variable	Categories
Pain	0----------X----------5
	0----------X----------5
	0----------X----------5
	0----------X----------5

It is important to keep in mind that a response variable that is highly reliable for one population (e.g., adults who speak English) may be very unreliable in a different population (e.g. children who cannot read). One way of measuring a questionnaire's reliability is to see if the answers to various questions correlate with one another. A patient who says he is comfortable but has a pain rating of 10/10 would have low correlation, while a patient who is comfortable and has a low pain rating would have a higher correlation.

This concept is called internal consistency and is often measured by Cronbach's alpha. A Cronbach's alpha score of 0.8 is usually desirable. (A Cronbach's alpha of 1 would not be desirable, because it simply indicates a high degree of redundancy.)

Just because a response variable is reliable (produces consistent results) does not mean that the response variable is actually measuring what we want it to measure. For example, say a researcher was trying to study pain using an ordinal response variable, but patients rated their hunger or boredom rather than pain, the researcher would come to an incorrect conclusion. In order to be meaningful, the response variable must actually measure the concept being studied. **Validity** indicates how well a concept is measured by an ordinal response variable or questonnaire. Generally speaking reliability and validity must be established for each new population (age, ethnicity, education/reading level).

Summed scale Response variables

Sometimes a response variable is made up a number of smaller ordinal response variables which are then summed. For example, the Apgar score used to indicate a newborns wellbeing and vigor is made up of five ordinal responses that are then summed to obtain a single Apgar Score. Similarly, the Glasgow Coma Score is used to indicate the level of consciousness in patients who are have suffered head injuries, although now it is also used to assess patients in recovery from surgery.

Apgar Score

Response Variable (Score)	2	1	0
Heart Rate (pulse)	Normal (above 100 beats per minute)	Below 100 beats per minute	Absent (no pulse)
Breathing (rate and effort)	Normal rate and effort	Slow or irregular breathing	Absent (no breathing)
Grimace (Responsiveness or "reflex irritability")	Pulls away, sneezes, or coughs with stimulation	Facial movement only (grimace) with stimulation	Absent (no response to stimulation)
Activity (muscle tone)	Active, spontaneous movement	Arms and legs flexed with little movement	No movement, "floppy" tone
Appearance (skin coloration)	Normal color all over (hands and feet are pink)	Normal color (but hands and feet are bluish)	Bluish-gray or pale all over

The lowest score possible is 0, and the highest score possible is 10. Scores above seven are considered normal.

Glasgow coma Score

Response Variable	6	5	4	3	2	1
Eyes open			Spontaneous	To speech	To Pain	None
Verbal response		Oriented	Confused	Inappropriate words	Incomprehension	None
Motor response	Obeys commands	Localizes pain	Withdraws from pain	Flexion to pain	Extension to pain	None

The lowest score possible is 3, and the highest score possible is 15.

Even though a number is generated from such a summed scale, care should be used when analyzing the data. The summed scale must still be evaluated for reliability and validity. Generally speaking, the summed scale number should not be used as a numerical response variable in analysis unless the number of questions is "sufficiently high." In scales where there are a large number of questions asked such as the Hamilton Depression Scale (Ham-D), the number generated may be used as a numerical response variable. See sidebar for an excerpted explanation of the Ham-D development and rationale.

The parameter/statistic that describes an ordinal response variable is the **median**--that is the middle number where half the responses fall above the median, and half the responses fall below. Note that a common mistake is to use means (averages) to describe ordinal response variables. This almost always results in incorrect inferences.

Numerical Response Variables

We have already examined response variables where the possible responses are categories. The other two kinds of response variables are numerical, where the possible responses are a number. As with the categorical response variables, there are two subcategories of numerical response variables.

Discrete Response Variables

Discrete Response variables are a special kind of numerical response values where the number must absolutely be an integer and is usually a small number. Moreover, the number must not change over time. As a result, there are few true discrete response variables. Examples might include the number of broken bones a patient has had at the time of of admission, or the number of children a menopausal woman has had. Generally speaking, however, a researcher need not agonize over whether a response variable is discrete, because most discrete variables are treated as continuous when they are analyzed anyway.

Continuous Response Variable

A continuous response variable is any variable that is measured with a number where the phenomenon being studied can take on any number within the range of possibilities. Most numerical response variables are continuous. Almost all physiological measures such as height, weight, and age are continuous.

Theoretically, no two experimental units should have the exact same continuous measurements, but we are limited by our ability to make precise measurements. For example, if

Development of HAM-D Depression Inventory

"The first consideration was that it should be applicable to all the subgroups of depressive illness, i.e., to cover all types of symptoms, though only common ones. It takes much time to inquire after rare symptoms and the information gained is meagre. The scale had to have a length of 12 to 20 items, because too short a scale is insufficiently reliable and when too long it is burdensome to fill in. Above all, it had to be clearly relevant and easy to use by clinicians working in their usual setting.

"The items selected covered the major symptoms of the depressions. The number of grades of severity chosen, for those present, was four: trivial, mild, moderate, and severe. Too many grades makes judgement very difficult and too few loses sensitivity. Four grades also ensured the elimination of the common bias to choose a mid-point. Preliminary tests showed that the patients could not provide sufficient information to make these fine distinctions for some of the symptoms, so they were reduced to two: doubtful or trivial, and clearly present. Experts are very dubious about two ranges of grading, but clinicians find them appropriate. This is the ultimate basis for the popularity of the scale. It is simple and easy to use in the routine of clinical practice, and it is meaningful and relevant. It is highly valid against clinical judgement and its reliability is equally high (far higher than many biochemical tests). All these account for its acceptance as a standard all over the world and its translation into many languages. Despite its deficiencies, it has lasted over 20 years and continues to flourish,[12] although doubtless it will be replaced in time."

you record a subjects age, you probably record it in years, but one could be more accurate and record, the months, days, hours, seconds, milliseconds.... And this brings us to an important data collection tip.

Tip: The recording of a continuous variable should always be as precise as possible. If such precision is not needed, it can be ignored during analysis, but the researcher can never go back and record the data more precisely. For example, if records the date of birth and the study date, he may later calculate age to within a day's precision or simply calculate the number of years. But if the researcher records only the number of years, he will never be able to calculate age more accurately. Now age itself is not so cumbersome, because a person's date of birth does not change, so as long as the researcher can track the subject down, he can still record the subjects date of birth. But if the researcher is recording some physiological measurement prone to change, such as blood pressure or heart rate, the researcher cannot go back in time and re-record the measurement more precisely. Bottom line: measurements should always be as precise as the instrument will allow.

Two parameters/statistics are used to describe continuous response variables. The first is the **mean** or average, and the second is the **standard deviation**. The mean is the sum of all responses divided by the number of responses. The standard deviation is the square root of the sum of differences of each response and the mean. The mean represents the a center balance point of the measurements, while the standard deviation represents how far from the mean the data lies. For example, two thirds of all responses will be accounted for within two standard deviations from the mean, and 96% of all responses will be accounted for within three standard deviations from the mean.

Sometimes, continuous variables are further broken down into interval and ratio. Interval variables are numerical but have no natural zero. For example, temperature as measured in either Celsius or Farenheit have a zero, but it is arbitrary. It does not make sense to say that a temperature of 10 is twice as hot as 5. Ratio variables on the other hand, do have a natural zero, so it makes sense to say that 10 kilograms is twice as heavy as 5 kilograms. Hence, ratios make sense in ratio response variables. However, for most purposes, the distinction is not necessary.

Choosing a Response Variable

When choosing a response variable for study, the researcher should ask himself the following questions:

1. What will best measure the concept I am trying to study?
2. What can I practically measure (time, money, technology, manpower)
3. What have other people done, and what advantages/disadvantages are there to their approach?

When reading or critiquing research, one should ask the same questions. Did the researcher choose an appropriate response variable that reflects the concept he is studying? Did the researcher run into problems because he chose a response variable that was impractical? Did the researcher take into account the experiences of other researchers?

The most appropriate response variable is not always self-evident.

Example 1: Blood Pressure

If a researcher wants to measure blood pressure, he finds out that he may measure blood pressure a number of ways:

- Cuff systolic pressure
- Cuff diastolic pressure
- Cuff mean pressure
- Cuff pulse pressure
- Intra-arterial measurements of the same four concepts above

As the researcher does his literature review he will find that intra-arterial measurements taken in the radial or brachial artery are very different from carotid or aortic measurements. Which is the best way to measure blood pressure? In the end, the researcher will probably record a cuff systolic and diastolic pressure for the simple, expedient reason that everyone else is doing the same thing (and it's cheap). Moreover, mean and pulse pressure can be calculated from systolic/diastolic.

The researcher will also find that the recommended technique is to take two blood pressure measurements at least two minutes apart, which should be averaged. As noted in the tip in the previous section, the researcher should record both readings separately and not simply record the average of the two readings.

Now the researcher may not truly care about the actual blood pressure; he only wants to know if the blood pressure is normal or high. He should resist the urge to only record the classification of normal or high, because later on, he may find that he wants to analyze the data according to the JNC VII guidelines for high blood pressure (normal < preHTN < HTN-1 < HTN-2). Moreover, if the researcher has done any kind of literature review, he will know that the JNC blood pressure definitions have changed over time. By recording the actual blood pressure numbers rather than only the category, the researcher ensures that he will be able to re-interpret the data if the classifications change again.

Bottom Line: When a number is available, always record the number. Resist the urge to categorize numbers during data collection. The numbers can always be categorized during analysis, but if the number was never recorded, we can never go back the other direction.

Example 2: Coronary Artery Disease

A researcher who wants to study heart disease may find that there are a number of ways to measure coronary artery disease (CAD):

- Degree of atherosclerosis based on intravascular ultrasound
- Percentage of narrowed coronary arteries based on angiography
- Positive stress test yes/no
- Ischemic ECG changes yes/no
- Myocardial infarction (MI) yes/no
- Presence of Angina Pectoris yes/no

The researcher could treat each of the above as a separate response variable, or he could combine them together and say that heart disease will be measured by the presence of any of the above. Such a combination response variable is called a **composite response variable.**

Let us assume for a moment that the researcher has chosen myocardial infarction yes/no to measure the concept of CAD. The researcher must still decide what constitutes a myocardial infarction. Is it merely presence of chest pain? Chest pain plus positive troponin enzymes? If troponin levels, then what level will be considered normal and what level will be considered? What will the researcher do with "grey zone" troponin levels? If patients from multiple hospitals are participating in the study, what do we do if the individual hospitals use different reference values for troponins? So we can see that even something relatively straightforward such as MI can be very tricky once the researcher starts trying to operationalize the response variable. As in the tip above, no matter what the researcher decides, it is best to record the actual troponin values rather than simply record "MI" or "not MI".

Summary of Response Variables

- The response variable is the actual measurement/observation of the concept being studied.
- There are two major classifications of response variable: categorical or numerical
    - There are two subclassifications of categorical: nominal and ordinal.
    - There are two subclassifications of numerical: discrete and continuous
- If a response variable can be measured numerically, ordinally, or nominally, then the numerical response variable should be recorded, and the numbers can later be put into categories during analysis.
- If it seems as though there is a lot more information on ordinal response variables than the others, that is because 1) ordinal response variables are more complex than the others, and 2) ordinal response variables are very popular in nursing research and are often misused.

Chapter 4
Obtaining a Sample

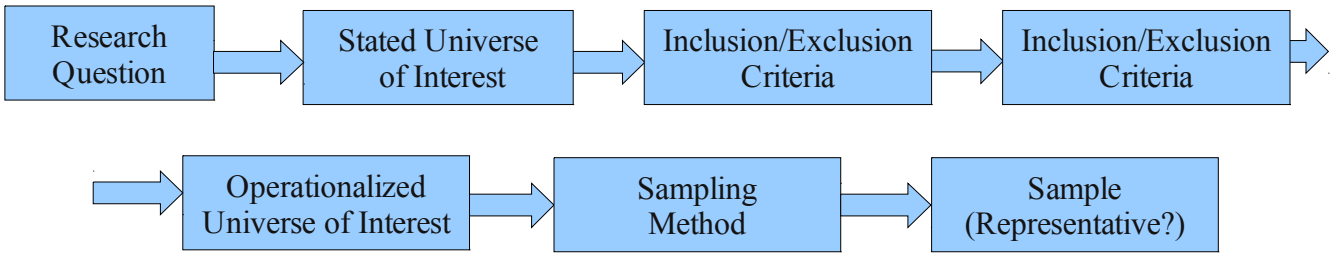
Once the research objective has been operationalized through inclusion/exclusion criteria, enumeration of study and extraneous factors, and selection of response variables, it becomes necessary to choose the actual experimental units that will be included in the study. Obtaining a sample actually is actually a two step process: 1) determining the appropriate sample size (number of subjects to be studied; and 2) choosing a method to obtain the sample. Determining the appropriate sample size is a rather complex subject and will be covered in a later chapter. In this chapter, we will examine the actual process of selecting experimental units.

In order for a researcher to make valid and true inferences about the population of interest, the researcher must obtain a representative sample. That is, the sample should be representative of the population of interest. If the researcher wants to study how a drug affects all adults, but only studies men, then the sample is not representative of the population of interest. Many people want to skip directly from the research question to the sample, however, there are important steps that must be taken to ensure a representative sample.

Before selecting experimental units, the researcher should have gone through this process:

- Decide upon a research objective, including the population of interest
- Refine the population of interest through inclusion/exclusion criteria
    - This will yield an operationalized (actual) population of interest.
- Decide upon extraneous factors

If the above process has not been carefully thought through, no sampling technique will produce a representative sample except by pure ~~luck~~ chance. Another way to express the relationship is:



When planning research and when critiquing research, one is often tempted to skip directly from the population of interest to the sample, but errors and distortions may be introduced at each step in the process. The researcher should make sure that each step is as consistent as possible, and critics should examine whether the researcher has followed the process appropriately.

Researchers have been trained through the years to desire a "random" sample. The drive has been so inbred that many researchers report using a random sample when in fact they did no such thing. In fact, as we will see shortly, a random sample is one of the less reliable methods of obtaining a sample. Before we discuss individual sampling methods, we must discuss bias.

Types of Bias

Selection Bias

Selection bias occurs any time the levels of an extraneous factor do not match the population of interest. For example, if the population of interest is all U.S. adults, and the sample has a larger proportion of Hispanic subjects than the U.S adult population, that sample is said to be biased with respect to ethnicity. In other words, our sampling technique has *selected* more Hispanics than would be expected if the sample were truly representative. If the population of interest is all traditional PBA students, and the sample has a larger proportion of females than the population of interest, then the sample would be biased with respect to gender.

Selection bias is especially prone to certain kinds of study design and recruitment. If a researcher must advertise in order to obtain subjects, the kind of advertising may affect the experimental units selected. If print advertising is used, then the subjects who volunteer may be biased toward reading level or the language the ad is printed in. Where the print advertising is distributed may bias the sample: newspaper, fliers posted on trees, billboards, etc. If radio advertising is used, the kind of music or talk on the station may bias the sample.

If payment is going to be made in compensation for participation, it may bias the sample toward those without jobs or college students. The time demands of the research may bias the sample toward housewives and students (persons with free time during the day).

Finally, the time frame or nature of the experimental unit can influence bias. If infection is being studied in a limited time frame, epidemics or disasters may bias the sample. If only a small pool of experimental units are available, that may also bias the sample.

Geographical Bias

Geographical bias occurs when there are differences in the response variable based on location. A researcher studying breathing might get very different results if he studied subjects in Denver instead of subjects from Atlanta. Additionally, a large proportion of research is done at universities, which may introduce bias in the form of age (students) or education level. (In Gainesville, there are so many college students that your local minimum-wage Blockbuster clerk is likely to have a bachelor's degree.)

Berkson's Bias

Berkson was a researcher who proposed that care must be taken when studying illnesses only in the clinical setting, because results may be very different than if studied in the community. For example, if a researcher wanted to know the percentage of person's who have both diabetes and hypertension together instead of separately would get very different answers depending on whether he conducted the research at a hospital vs. out in the community.

Berkson's Bias also explains why very rare illnesses are routinely seen in the hospital. Persons who are not ill do not typically go to the hospital, while people who are sick go to the hospital to seek care. Thus, although a disease may be rare in the general population, it may be quite common in a hospital.

Sampling Techniques

Probability Sampling

The best way to ensure that a sample is representative of the population of interest is to use probability

sampling. Probability sampling is any sampling method where *each experimental unit has a known and fair chance of being selected.* That is a *fair* chance, not an equal chance. Any time a sampling technique does not allow every experimental unit a fair chance, it is called a **convenience sample**.

There are four types of probability samples. Three of them require a list of all possible experimental units

Simple Random Sampling

Simple Random Sampling requires a list of all possible experimental units. Once a list is obtained, each experimental unit selected is selected randomly from the list. That is the first experimental unit is randomly selected, then the second is randomly selected, and so on.

Simple Random Sampling is the crudest method of selecting a probability sample. It does not guarantee a representative sample, and is deceptively difficult to do properly. The proverbial "pull numbers out of a hat" does not produce a true random sample. When simple random sampling is done incorrectly, it results in a convenience sample. The problem is that the researcher probably does not realize that his sampling technique is flawed. Even when it is done properly, there is still a chance that the sample is not representative of the population of interest. Despite these limitations, it seems that "random sampling" is very common in the research literature. Many articles claim a "random" sample, but offer no details as to how the sampling was actually done. In such a case, the critic should ask themselves whether it is possible to compile a list of all possible experimental units. If such as list is not possible, then quite simply, the researcher is lying and actually used a convenience sample.

Stratified Random Sampling

One of the problems with simple random sampling is that a representative sample is not guaranteed. Stratified random sampling tries to ensure that the sample is more likely to be representative. If the list of all experimental units already has some extraneous factors within the list such as age, gender, or ethnicity, then the list may be grouped by the extraneous factor. For example, if the population of interest were all students at PBA, and the researcher wanted to make sure that the study sample is representative regarding gender. It would be quite easy to obtain a list of all students at PBA that includes their gender. From this list, the researchers would find that males make up 35% of all PBA students; 65% female. The list would then be sorted by gender. The researcher would then take a simple random sample of males and a simple random sample of females, such that the final sample contains the same proportion of males and females as the population of interest.

It is important to note that stratified random sampling must be done two or more times, once for each group. So while it helps to ensure that the sample is more likely to be representative, it actually complicates the procedure and provides more opportunity for error.

Systematic Sampling

Systematic sampling is the simplest probability sampling technique both to understand and to perform. Once the list is obtained, the number of experimental units in the list is divided by the sample size needed. For example, if the list of all PBA students has 2000 students, and the researcher needs to study 100 of them, the researcher will divide 2000 by 100 to come up with 20. The researcher will then choose every 20^{th} student on the list to participate in the study. In order to ensure that each student has a fair chance of being chosen, the researcher must randomly choose the first student.

Because only one random number needs to be chosen, systematic sampling is far less prone to error in

technique than either simple or stratified random sampling. Moreover, if the list is ordered by known extraneous variables, systematic sampling will ensure a representative sample. In our example, we might sort the list by gender, then by major, then by GPA, and then by age. Our systematic sample will then have the same proportions of gender, majors, GPA's, and ages as the population of interest.

Cluster Sampling

The major drawback to the other three probability methods is that they require a list of all possible experimental units. Such a list might be difficult to come by or change before the study is completed. Additionally, if the population of interest is spread over a large area, it might be very time consuming or expensive to go to every place that has experimental units. In such cases, cluster sampling is the sampling technique of choice. In cluster sampling, first a list of all locations or sites that have experimental units is compiled. Then a random or systematic sample is chosen of *sites*. Then a list of all experimental units at the chosen sites is compiled, and another random or systematic sample is chosen from the second list.

For example, say a researchers population of interest was all patients in Palm Beach County hospitals. There are 20 hospitals in Palm Beach County and he needs to study 100 patients total to achieve the research objective. If the researcher used simple or stratified sampling or systematic sampling, the researcher would study on average 4 patients at each hospital. It would probably be too time consuming to go to each hospital for only a few subjects. In such a case, cluster sampling allows the researcher to study more subjects in a shorter time period as well as reducing travel costs. The first step would be to compile a list of all Palm Beach County hospitals. He would then choose five hospitals using simple random sampling, stratified random sampling, or systematic sampling. Then a list of all patients at those five hospitals is compiled and subjects are chosen from that list. The researcher will now study, on average, 20 patients from five hospitals—which is much more manageable. In fact, if the researcher saves enough time, he may find that he can study 30 or 40 patients from each hospital helping to ensure enough subjects to meet the researcher objective.

Notice that the researcher still needed a list of all patients at each of the sampled hospitals, but did not require a list of patients at the other fifteen hospitals. Cluster sampling is best suited for when subjects are clustered by location or when a list of all subjects is not available.

Convenience Samples

Although probability sampling is the best way to achieve a representative sample, most medical research is conducted with convenience sampling. Many convenience samples live up to the name. For example, a researcher studying left-handedness might simply go to the hospital cafeteria and watch for people eating with their left hand and then ask them participate in her study. (I was once recruited for a study in this way.) Or a researcher studying breast cancer might ask local oncologists to recruit all new patients until the researcher has enough subjects to study. Alternatively, a researcher studying sleep might take out an ad on the radio advertising sleep studies to persons with insomnia.

Once a potential experimental unit has been identified, they are screened for inclusion/exclusion criteria and asked to participate in the study. This process of advertising, screening, and asking a patient to participate in research is called **recruitment.**

The defining characteristic of a convenience sample is that not every experimental unit has a fair chance of being selected. Thus, no matter how sophisticated the sampling technique if there are experimental units who will be left out of the recruitment process for any reason (cost, language, distance, ability to read, etc.) the sample should be considered a convenience sample. There is no shame in using a convenience sample, but the

researcher must be honest and accurate in reporting how the sample was obtained. For example, if a researcher studying obesity in children recruits his sample from an obesity clinic, the researcher should be honest and report it so that persons reading the research can realize that the conclusions of the research are not necessarily valid for obese children who are not in treatment. Alternatively, if a researcher must advertise, the researcher should report the advertising and comment on the possible bias it introduces. One of the most common problems with medical research is potential geographical bias—even in large multinational, multi-center trials.

Damage control

Once the sample is obtained and data is collected, one method of establishing whether a sample is representative or not is to compare levels of extraneous factors to the population of interest. For example, in our sample of 100 PBA students, we may want to compare the distribution of gender, age, major, and ethnicity to the distribution of those factors for all PBA students. If the distributions are similar, then it is likely that the sample is representative. If the distributions are different, then the sample is likely biased and generalizations to the population of interest may be invalid.

The best time to deal with bias is in the planning phase. Probability sampling should be used whenever possible. When it is not possible, the researchers should try to make the sampling technique as broad as possible. However, even the best sampling technique is still subject to bias if the operationalized population of interest (as defined by inclusion/exclusion criteria) is not representative of the stated population of interest. The honest researcher should comment on the limitations of the sample obtained as well as the appropriateness of the inclusion/exclusion criteria.

Sidebar: Please not that Case-Control studies present a special challenge as there are effectively two samples. Sampling technique should be described for the cases and separate sampling technique should be described for the controls. The goal of the control sampling technique should be to ensure that the cases and controls are similar for matched extraneous variables.

Chapter 5
Types of Research Questions

Most research ultimately asks a question about reality. These questions are answered using *inference statistics*. Ultimately, there are only two kinds of inference questions or objectives: estimation and hypothesis testing. Estimation asks, "what is the true parameter?" Hypothesis testing asks, "is there a difference among two or more parameters?" In addition to setting up the questions properly, the researcher must also ensure that he has enough experimental units to make a good statistical inference. Remember that research is ultimately a form of inductive logic, and in order to avoid faulty inductive leaps, we have to study enough instances to be reasonably certain that our conclusions are correct. Estimation and hypothesis testing have different processes and implications for sample size. We will examine each separately as well as common elements that affect sample size determination for both.

Estimation

Estimation asks the question, what is the true parameter?

- What is the average weight of PBA students?
- What is the percentage of males at PBA?
- What is the average age of Junior nursing students?
- What is the average GPA of Junior nursing students?
- What is the percentage of PBA graduates who pass the NCLEX on their first attempt?

In addition to these kinds of straightforward questions, estimation may also compare two parameters:

- What is the average weight change of PBA students over Christmas break?
- What is the average change in cholesterol while on drug X.

Estimation of changes may be analyzed by absolute changes (arithmetic difference), percent change, or co-variance.

Estimation involves a four step process. We will examine each step in detail below.

1. Identify the response variable and the parameter to be estimated.
2. Determine the meaningful difference
3. Determine the confidence interval
4. Collect the data and calculate the statistics and estimate the parameter.

Identify the response variable and the parameter to be estimated

The response variable influences the number of experimental units needed (sample size). Generally speaking, it takes larger sample sizes to make inferences about nominal response variables than continuous response variables, but each type of response variable has its own considerations. The key consideration for determining sample size for nominal response variables is the approximate frequency of each response.

For example, despite its certainty, death is relatively rare in the United States—less than one percent per year. If a researcher wants to study mortality (R.V. Died: yes/no), he will need to study enough people to make sure he has enough dead people to meaningfully estimate. Otherwise, the researcher will come to the

incorrect conclusion that the true death rate is zero. Armed with this knowledge, a researcher has several techniques at his disposal to increase the number of deaths in his sample: 1) He can extend the study period, hoping that more people die, 2) Study higher risk people who are more likely to die, but this will change the study question; 3) Increase the sample size.

I know what you are thinking: Why do the study if the researcher has to know the proportion ahead of time in order to know how many people to study in order to estimate the proportion? The answer is, that the researcher does not need to know the true proportion, only have an idea of how many to expect. He can obtain this info from three sources: 1) Literature review; 2) Estimate from other populations; 3) Pilot studies.

Estimating a continuous response variable generally takes fewer experimental units than nominal response variables. The key consideration for continuous response variables is the variance. The more the phenomenon being studied varies, the more experimental units need to be studied. Just as with nominal response variables, the researcher needs to have an idea of the variance. Variance can be learned from either literature review or pilot studies. If neither are available, the variance can be estimated from approximate range of likely values. See Sidebar for details

Sidebar: Recall from statistics that 99% of all responses will fall within three standard deviations from the mean (±3 St.Dev.). Therefore, the expected standard deviation can be estimated by taking the range of possible responses and dividing by six. For example, if a researcher is studying systolic blood pressure in patients with high blood pressure, the range of likely values is 140 – 200. The range is 200 – 140 = 60. The approximated standard deviation is 60 ÷ 6 = 10.

General rules

- Nominal response variables require larger sample sizes than continuous response variables
- Rare nominal responses require larger sample sizes
- Continuous response variables with large variances require larger sample sizes

Determine the meaningful difference

Imagine for a moment that you want to lose weight. How much weight would you have to lose before you considered it meaningful? One pound? Two pounds? Three pounds? Ten pounds? When estimating any parameter the researcher must ask the same kind of question, "how precise do I want to be?" This is called determining the *bound on error*, and is usually measured as ± y. So if above, you decided that five pounds was meaningful, your bound on error would be 5 ÷ 2 = ±2.5 lbs. The more precise the bound on error, the larger the sample size needed.

Determine the confidence level

The bound on error has nothing to do with how likely the research is to produce the correct estimation, only how precise the estimation produced will be. The confidence level determines how likely the research is to produce the correct parameter. Confidence level is expressed in percentage and refers to chance that any given research using the same sample size will produce the correct parameter within the bound on error. Common confidence levels are 90%, 95%, and 99%. A 95% confidence level is most common in medical research.

Calculate the statistics and estimate the parameter

Hypothesis testing

Chapter 6
Additional Statistical Analysis

In addition to simple parameter and mean estimation and hypothesis testing, researchers use a variety of methods to answer the research questions. This chapter will introduce several of these methods and point out techniques for interpreting and abusing them.

Relative risk ratio (Odds ratio)

Cox regression and Hazard Ratio

Kaplan-Meier Curves

Correlation

Regression

Simple Linear

Multiple Regression

Logistic Regression