3 Conceptualization, Operationalization and Measurement

Earlier in the semester, we talked about the difficulty of measuring a phenomenon of interest. Sometimes it is not obvious how to best capture what we are interested in. Recall our example in which we were interested in studying the effect of exposure to violent films on the subsequent violent behavior of children. After some careful thought, we realized that it’s not obvious how to best define “violence.” By violence, do we care mostly about murder? Or does violence encompass a range of behavior including verbal threats? Does the context of the violence matter? Here, recall our comparison between Scarface and Saving Private Ryan. To what extent does the identity of the victim matter? Are children more affected by exposure to violence when the victim is a woman? Or if the victim is drawn from an especially vulnerable population such as the elderly or the physically disabled? How about racial violence? Is the effect greater if a child is exposed to violence in a film that mirrors violence he was personally witnessed?

There are many possibilities. What is important is to realize that to simply study the effect of “violence” is virtually meaningless. Without a clear idea of what constitutes violence, we might make the wrong recommendation to parents and policymakers. Likewise, if we don’t capture violence in a salient way, we might mistakenly conclude that violence doesn’t matter at all when, in fact, a particular type of violence might affect children is a very meaningful way.

Our example about violent movies jointly motivates the topics of conceptualization, operationalization and measurement which we will talk about as a single unit. In particular, it should be clear that the way we measure things matters a great deal. We can have an incredible theory, a terrific research design and great data but if we aren’t measuring our variables well, it will be all for nought.

In order to carefully map out the measurement process, we’ll begin by considering a rather vague idea of what we’d like to measure. We’ll call this vague idea a concept. We’ll then consider how to make this vague idea somewhat less vague. This is what we’ll refer to as operationalization — that is, the process of figuring out how to represent our vague idea using data. Finally, we’ll think about how precisely to measure our concept in order to form a construct called a variable.

3.0.1 Conceptualization

What do we mean by conceptualization? In short, we are referring to the process whereby we clarify a mental thought process and express it as a specific idea or concept. Basically, conceptualization is about being clear and precise. We do not want there to be any confusion about what we intend to measure or document. This is important because regardless of whether the concept we’re interested in is sensible, at a minimum, other researchers as well as practitioners will know exactly what I am doing. As an example, consider that we are interested in recidivism — the degree to which a former offender reoffends after having been
arrested or convicted. Specifically, we are interested in serious recidivism, that is reoffending that is sufficiently serious to cause a great deal of concern for society.

How to begin? Well, we might start by clarifying what we mean by serious. I think it’s fair to say that murder and rape are serious crimes. We might also consider robbery and aggravated assault to be serious. How about burglary? Is burglary a serious crime? It’s certainly worrisome — particularly if the homeowner is at home during the burglary. How about motor vehicle theft? This is a crime that is non-violent (carjacking is separate and would be classified by the FBI as a robbery). However, it is certainly costly. What would you do if your car was stolen? It certainly wouldn’t be a fun experience. How about theft of property — for example, theft of a bicycle or an iPhone? How about fraud? Or forgery? Or drug trafficking? Or drug possession? I’m pretty sure if I polled the class, each student would put together a different list of what he or she thinks of as capturing serious crime.

How can we add some precision and objectively to this? We could begin by clearly differentiating violent and nonviolent crime. In violent crimes, an offender uses force or threats of force against a victim. Nonviolent crimes either do not involve any direct contact between a victim and an offender or involve contact but no force. For example, pickpockets have direct contact with their victims but use no force. In contrast, robbery involves at least the threat to use force on victims. Using the FBI’s classification system, burglary, auto theft, shoplifting, and the theft of unattended personal property such as bicycles are examples of nonviolent crimes. Assault, rape, robbery, and murder are violent crimes.

Hence, I might conceptualize serious crime as those crimes which involve violence or the threat of violence. This might not perfectly capture everyone’s notion of serious crime but at least we are using a clear and transparent definition that captures an underlying behavior (violence or the threat of violence) that everyone is capable of understanding. If you disagree with the choice I have made, you would be free to argue that I have made an error in judgement.

3.0.2 Dimensions and Indicators

Now that we have conceptualized serious crime as that which involves violence or the threat of violence, we next turn to a related concept — just how serious is each crime we’re considering? In other words, within our notion of serious crime, we want to have some measure of each crime’s relative seriousness. With regard to violent crimes, we could consider the extent of the victim’s injuries. Obviously the most serious injuries would be those which led to the victim’s death. Short of this, the extent of injuries could range from the very serious (e.g., paralysis) to something more minor (e.g., an abrasion to an arm). We could also consider the psychological injuries stemming from the crime — that is the victim’s level of trauma. For example, a sexual assault which involves a very serious personal violation and is carried out over a period of time likely involves a great deal more emotional trauma than a robbery in which the victim is approached from behind and knocked down while their property is stolen, even if the robbery victim has more serious physical injuries. The extent to which there is trauma might depend on the length of an attack, the seriousness of
the personal violation and the degree to which the threat of force was used by the perpetrator.

All this is to say that within the category of “serious” crimes, we need to have some way to conceptualize seriousness. The FBI classifies the four violent “index” crimes of murder, rape, robbery and aggravated assault in that same order of seriousness. This is what they believe to be true, in general. Obviously, for a given crime it will depend on the details of that crime.

Whenever we are trying to conceptualize a concept, it is important to provide a definition of that concept that is as specific and precise as possible. It is also important to conceive of a way to consider the dimensionality of the concept (that is, how will we measure seriousness).

3.0.3 From Concept to Operationalization

Once a researcher conceptualizes what he or she would like to measure, the next step is to determine how this conceptualization will be operationalized to produce a variable. Let me re-state this because it is important. Operationalization is the process of taking a concept and translating it into something that can be measured.

\[
\text{Concept} \rightarrow \text{Operationalization} \rightarrow \text{Measurement}
\]

To continue our “serious crimes” example, once we conceive of a serious crime as one which involves either violence or the threat of violence we want to be able to generate a variable that we can use directly in our research. This may sound simple and sometimes it is but often it is not. One of the trickiest issues here is that the data we would like to have are not always available.

For example, let’s consider a scenario in which we are interested in the recidivism of individuals who have been convicted of a violent crime (which we have defined to include murder, rape, robbery or aggravated assault). First, we will need to consider the follow-up period. Do we want to consider recidivism as of one year of release from prison? Two years? Three years? (In practice, usually, CJ researchers look at one- and/or three-year recidivism rates.)

Next, we will need to think about what type of recidivism we will count. We’ll certainly count an individual as having recidivated if we observe that he commits a new serious crime. Let’s say we’ll also count as recidivism a new crime that is less serious (for example, burglary, larceny or motor vehicle theft). However, a large percentage (depending on the state, possibly as high as one third!) of individuals returned to custody are returned not because they committed a new crime but rather for a technical violation. Typically these are parole violations (a positive drug test, leaving the state, failing to show up for a probation/parole appointment). Do we want to count such individuals as having recidivated? On the one hand, they’re back in prison and the taxpayers have to pay for this. On the other hand, there is no new crime and therefore no new crime victim. The choice of whether or not to count this type of violation as recidivism is up to the researcher. The important thing is that the researcher must document all of the choices made in the course of research.

This brings me to my penultimate point about operationalization. In an ideal world, we would be able to observe whether every individual in our sample has committed a new crime.
But in the real world, we don’t know this! After all, to know if someone has committed a crime we’d literally need to follow that person around twenty four hours, seven days per week. Instead, what we have are administrative data from the criminal justice system. We usually can observe whether the individual has been arrested, whether he has been charged with a crime, whether he has been convicted of a crime (including a guilty plea) and whether he has been returned to custody (incarceration). Hence, when we operationalize our notion of recidivism, we will be limited to what we can actually measure. We might conceptualize recidivism as being about whether or not an ex-inmate continues his or her bad behavior upon being released from prison. However, we will operationalize this concept using data that indicates either re-arrest or a return to custody.

3.0.4 Variables

We have talked about how a researcher begins by conceiving of a concept and next thinks about how to operationalize that concept. Now let’s think about types of measures that we can construct based on that operationalization.

A variable is the end product of the measurement phase of a research project. In other words, we begin with a rough concept of what we’re interested in measuring. Next, we think of a way to operationalize that concept. Next, we consider a way to measure what we’re interested in. What we ultimately must come up with is called a variable. Accordingly, we can add to the schema introduced above:

Concept → Operationalization → Measurement → Variable

A variable must satisfy certain properties:

1. The values of a variable must be mutually exclusive. Researchers must be able to classify every observation in terms of one and only one value.

2. The values of a variable must be collectively exhaustive. As Maxfield and Babbie note, if the variable is to have any utility in research, researchers must be able to classify every observation in terms of one of the attributes composing the variable.

These concepts are potentially confusing so let’s consider a few examples:

<table>
<thead>
<tr>
<th>Candidate Variable</th>
<th>Values of the Variable</th>
<th>Mutually exclusive</th>
<th>Collectively exhaustive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0-120</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Gender</td>
<td>male, female, other?</td>
<td>Probably</td>
<td>Yes</td>
</tr>
<tr>
<td>Race</td>
<td>black, white, asian</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

In the table above, we’ve determined that race apparently does not meet the criteria outlined above for a variable because the categories are not mutually exclusive. Why not? Because an individual might consider himself to be both white and black (e.g., Barack Obama) or both white and Asian (e.g., Keanu Reeves). Using a simple conception of race means that we cannot construct a proper variable. Does this mean we cannot do research on race? No!
The solution is to re-define race so that it does meet the criteria of a variable. In particular, since the issue is that while many individuals identify as only black, only white or only Asian, a growing number of individuals identify with multiple racial groups, instead of classifying individuals as either black, white or Asian, we could classify individuals in the following way:

1. black only
2. white only
3. Asian only
4. white and black only
5. white and Asian only
6. black and Asian only
7. white, black and Asian
8. other race
9. choose not to identify

Now, race can be thought of as a variable because the categories are mutually exclusive. There is no one who can be in multiple categories, given how we have written down the choices. At the same time, we have added so many categories that it might become unwieldy to analyze the data in a parsimonious way. One solution that researchers have developed is to recode a variable — that is, combine the categories in such a way as to be parsimonious but also mutually exclusive. For example, we could recode the above categories according to the following scheme:

1. black only: [1]
2. white only: [2]
3. Asian only: [3]
5. other/choose not to identify: [8]+[9]

This new variable still satisfies our demand of mutual exclusivity but it is considerably more parsimonious than our first attempt. We simply decided to recode the variable such that all multi-racial individuals are lumped together in one category. We also combined the “other” and “choose not to identify” categories. These categories are always an issue in survey research — often individuals do not answer our questions at all. Other times, individuals provide an answer but not necessarily one we have anticipated or offered them. For example, in response to being asked about their race in the U.S. Census, millions of Americans do not choose one of the boxes offered and instead simply write in “American.” In other cases, answers may be completely unintelligible either because the respondent does not know how to
answer a question or because he is trying deliberately to be difficult. For example, in response to a question about a respondent’s income (always a sensitive question), a respondent might simply write “jello.” Obviously it is not clear what to do with this information!

Aside: This will obviously not be on the exam but the above discussion reminds me of an incredible (and true) story from World War II. Anthony Clement “Nuts” McAuliffe was a United States Army general who was the acting division commander of the 101st Airborne Division troops defending Batsogne, Belgium during World War II’s “Battle of the Bulge.” Here is an account of how he earned the nickname “Nuts”:

In December 1944, when the German army launched the surprise Battle of the Bulge, Major General Maxwell D. Taylor, commander of the 101st Airborne Division, was away, attending a staff conference in the United States. In Taylor’s absence, acting command of the 101st and its attached troops fell to McAuliffe. At Bastogne, the 101st was besieged by a far larger force of Germans under the command of General Heinrich Freiherr von Lttwitz. On December 22, 1944, through a party consisting of a major, a lieutenant, and two enlisted men under a flag of truce that entered the American lines southeast of Bastogne (occupied by Company F, 2nd Battalion, 327th Glider Infantry), General von Lttwitz sent the following ultimatum to Gen. McAuliffe:

To the U.S.A. Commander of the encircled town of Bastogne: The fortune of war is changing. This time the U.S.A. forces in and near Bastogne have been encircled by strong German armored units. More German armored units have crossed the river Our near Ortheuville, have taken Marche and reached St. Hubert by passing through Hompre-Sibret-Tillet. Libramont is in German hands. There is only one possibility to save the encircled U.S.A. troops from total annihilation: that is the honorable surrender of the encircled town. In order to think it over a term of two hours will be granted beginning with the presentation of this note. If this proposal should be rejected one German Artillery Corps and six heavy A. A. Battalions are ready to annihilate the U.S.A. troops in and near Bastogne. The order for firing will be given immediately after this two hours term. All the serious civilian losses caused by this artillery fire would not correspond with the well-known American humanity.

— the German Commander

General McAuliffe’s reply:

To the German Commander: NUTS!

— The American Commander
Obviously, the German commander was pretty confused!

Postscript: The 101st was able to hold off the Germans until the 4th Armored Division arrived on December 26 to provide reinforcement. For his actions at Bastogne, McAuliffe was awarded the Distinguished Service Cross by General Patton on December 30, 1944, followed later by the Distinguished Service Medal.

In any case, the lesson from this lengthy aside is that often researchers have to exercise judgement in constructing variables. We will return to the issue of weird and hard to interpret survey responses when we study survey research later in the semester.

### 3.0.5 Types of Variables

Now that we have defined what a variable is and what criteria it must satisfy, let’s consider the types of variables that we might employ in a research project. Broadly speaking there are three types of variables: 1) nominal measures, 2) ordinal measures and 3) continuous variables. We’ll define these in turn:

1. **Nominal variables**: These are variables which have only the properties of mutual exclusiveness and collective exhaustiveness — nothing more and nothing less. That is, a nominal variable assigns each unit in the dataset to a particular category but there is no natural order to the categories. Examples include:
   
   a) Gender (Individuals can be either male or female, one is not higher or better than the other)
   b) Marital status (Individuals can be either single, married, divorced or separated)
   c) State of Residence (Individuals might live in any one of the 50 states in the U.S., again one is not better than another)

2. **Ordinal variables**: These are variables which in addition to having the properties of mutual exclusiveness and collective exhaustiveness, can be logically rank ordered. These variables continue to assign individuals to categories but now one category can be characterized as either higher or lower than another in a meaningful and objective way. Examples include:
   
   a) Highest education completed (Individuals can have less than high school, a high school diploma, a college degree, a graduate degree. Each level indicates a higher level of education than the one before it)
   b) Rank within the police department (Individuals can be officers, sargent, lieutenants, captains, etc. Each rank is associated with a higher level of authority than the one preceding it)

3. **Continuous variables**: These are variables which do not group individuals into categories and instead provide a simple numerical measure of some attribute. Examples include:
   
   a) Age: Police officers are generally between the ages of 18 and 65. We could even measure each individuals age more precisely (e.g., 18.54 years, 47.26 years, etc.)
b) Income: Individuals’ incomes might vary between $10,000 and $150,000. There are no categories. The variable is simply a number representing each individual’s income.

c) SAT score: Though SAT scores are measured in 10-point intervals (i.e., there are no scores like 1076 or 1203), this is still a continuous variable since each 10-point interval represents a unique score.

Sometimes, variables are classified as being either continuous or categorical. In this case, the term “categorical” refers to variables that are either nominal or ordinal.

In general, it is fairly straightforward to tell if a variable is nominal, ordinal or continuous. However, sometimes these determinations can be a little tricky. Let’s consider a couple of additional examples to make sure we’re all on the same page:

Example #1: A variable measures each individual’s income in $ for the 2012 calendar year.
This variable is continuous. The variable does not classify individuals into groups based on their income. It is a numerical measure of how many dollars an individual earned last year.

Example #2: A variable classifies an individual’s economic status based on that individual’s income in $ for the 2012 calendar year. In particular, individuals are divided into those living below the poverty line, individuals whose incomes make them “working class,” “middle class” and wealthy.
This variable is ordinal. Even though the underlying measure (income) is continuous, once we use income to group individuals into categories, the variable is no longer continuous. The variable is not nominal because there is still a logical ordering (i.e., “middle class” refers to a higher income than “working class.”

Example #3: A variable documents each individual’s occupation — e.g., teacher, police officer, therapist, lawyer, fast food worker, etc.
This variable is nominal. Even though each of us might have some sense of which jobs we think are better than others (e.g., most of us would probably prefer to be a lawyer than a fast food worker), there is no logical ordering that we could all agree upon. Some of us might rank “teacher” above “lawyer.” Others might disagree. Each category simply describes a particular occupation. This is not to say that we could not conceive of a different variable that seeks to measure an individual’s occupational status (say on a scale from 1 to 10). But this would be a different variable.

My advice is to go home and try to think of a few potential variables and how you might classify them. I’m sure to ask you at least one question about this on your midterm exam.
3.0.6 Scales

Another type of variable to consider are scale variables. A scale is really special type of ordinal variable but it is sufficiently distinct from other ordinal variables that it merits its own discussion. Scales are ordinal variables that arise from surveys of people’s attitudes towards a particular issue. Specifically, it is common in survey research to ask people whether they agree or disagree with a particular statement. Examples of statements that one might ask survey respondents to evaluate in these terms include:

1. The level of immigration to the United States is currently too high.
2. Police in my community are doing a good job.
3. The United States should effect military intervention in Syria.

Respondents are often asked to provide one of the following responses to such a statement:

Strongly disagree (1) → Disagree (2) Somewhat disagree (3) → Neutral (4) → Somewhat agree (5) → Agree (6) → Strongly agree (7)

When the respondent is offered the seven choices listed above, the scale is known as a Likert scale. Why seven choices as opposed to ten or even twenty? After all, more choices would offer greater precision. As it turns out, psychologists believe that there is a tradeoff between precision and accuracy when it comes to attitudinal variables. Human beings are generally able to accurately describe their feelings on a seven-point scale but once we move beyond seven choices, people have a tough time figuring out how to translate their feelings into an answer choice.

Often scale variables are combined into constructs called factors. The idea is that if you want to measure an individual’s overall attitude with respect to an item of interest (e.g., public safety), more questions is better than a single question. For example, we might ask individuals the following questions:

1. Police in my community are doing a good job.
2. It is safe to walk around my neighborhood at night.
3. I would allow my children to walk home alone from school.
4. I do not worry about crime in my neighborhood.
5. My neighborhood is safer than other neighborhoods.

As you can see from the list of questions above, each question asks individuals to evaluate how safe they feel in their neighborhood. But each question is subtly different. Instead of analyzing each question separately, we could combine the questions into a single factor that we might call “feelings of safety.” There are various techniques for combining variables. The simplest is to simply average them or add them up. Other techniques are extremely complex. The choice of technique really doesn’t matter. The core idea is that a factor is a combination of topically-related scales.
3.0.7 Assessing the Quality of a Variable

At this point, we have considered lots of different variables that we might be interested in studying. Each variable came about by operationalizing a particular concept and seeking to measure it. At the end of the day, what we want to have are high quality variables — that is, variables that are good measures of what we are interested in. But how do we know if a measure is “good” or not? Ideally, we would like to have a precise and scientific way to discern between “good” and “bad” variables. As it turns out, social scientists have developed various schema and rules or thumb in order to judge the quality of variables. In general, there are four criteria for assessing the quality of a variable:

1. Precision: This refers to the specificity of a variable. For example, it is more specific and therefore more precise to say that a man is 45 years of age rather than simply saying that the man is in his “mid 40s.” Likewise, it is more precise to say that the man is in his mid 40s rather than simply saying that the man is in his 40s. In almost all circumstances, the more precise our measure is, the better. However, there will be some cases in which precision matters only to a point. For example, if we are interested in assessing whether or not an individual passed an exam to be admitted to the police academy on his or her first attempt, then we really don’t need to know that individual’s exact score — we would just have to know if it were above or below a certain threshold.

2. Accuracy: This refers to the extent to which the variable contains errors. As social scientists, we would love to believe that all of our variables are correct all of the time but, in reality, there are often errors in variables. For example, I have seen datasets in which individuals who are male are erroneously classified as female. Likewise, it is often the case, especially in survey research, where individuals lie to the researcher. People often lie about attributes such as education and income which are often perceived to be measures of one’s status. But people also often lie about their behaviors as well. For example, individuals do not always admit to drug use, criminal activity or participation in risky sexual behaviors.

While we will always know the degree to which a variable is precise, in practice, it is often difficult to know whether a variable is accurate or not. There are sometimes statistical procedures to assess a variable’s accuracy but these are unavailable in most instances. The best way to know if a variable is accurate or not is to have some background information about how that variable came about. For example, if you are looking at rap sheets for criminal offenders, it would be helpful to know something about how police departments and courts generate these rap sheets. Who fills out the paperwork? Who enters the data into a computer? Which fields in the paperwork are often left blank or misclassified. This is why it is very, very important for researchers to talk to cops! By the way, do you know the origin of the term, “rap sheet?” It is actually an acronym for “record of arrest and prosecution.”

3. Validity: We have already talked about various types of validity in Chapter 2. In this case, validity refers to the extent to which a particular variable captures the underlying concept that we are interested in measuring — essentially this is what we meant
when we discussed *construct validity* in Chapter 2. Suppose, for example, that we are interested in measuring a high school student’s readiness to take college classes. This is a difficult concept to measure because it involves many different attributes (academic ability, maturity, the student’s ability to concentrate, the student’s family situation, whether a student chooses an appropriate major given his or her skills set, etc.) Unable to measure all of the various components of college readiness, researchers often, as a proxy, use a student’s SAT score. Is a student’s SAT score a valid measure of college readiness? For most students, probably not but it is debateable. Some researchers view the SAT score as offering high validity; others argue that it is a completely invalid measure of college readiness.

In fact, there are several different ways to think about validity:

a) Face validity: This refers to whether a variable meets the criteria of plain old common sense. For example, suppose we are interested in measuring whether individuals are satisfied with the quality of policing in a given community. We might consider measuring satisfaction using the number of citizen complaints against police as a proxy. Surely this is imperfect — after all, lots of citizens may be dissatisfied but will never file a complaint. But the variable satisfies face validity in the sense that at least complaints should have something to do with perceptions of police quality. If instead I sought to measure perceptions of police quality by asking individuals in a community if taxes are too high or too low, this simply doesn’t have any face validity. How on earth would citizen views on tax rates have anything to do with perceived police quality? Basically, face validity can be assessed using the “man on the street” test. If I stopped a random dude walking down the street and explained my concept and my measure, it should be clear to this random guy who presumably knows nothing about research how the two are related.

b) Criterion-related validity: This refers to the degree to which a candidate variable predicts something that is known to be valid from prior research and is also thought to be highly related to the concept you are interested in measuring. For example, if it is true that SAT scores predict a student’s GPA in his or her first year of college, to the extent that first year college grades are what we mean when we refer to college readiness, we might conclude that SAT scores satisfy criterion-related validity.

Maxfield and Babbie use a nice example from a real research study of an attempt to establish criterion-related validity:

Timothy Heeren and associates (Heeren, Smith, Morelock, and Hingson 1985) offer a good example of criterion-related validity in their efforts to validate a measure of alcohol-related auto fatalities. Of course, conducting a blood alcohol laboratory test on everyone killed in auto accidents would be a valid measure. Not all states regularly do this, however, so Heeren and colleagues tested the validity of an alternative measure:
single-vehicle fatal accidents involving male drivers occurring between 8:00 p.m. and 3:00 a.m. The validity of this measure was shown by comparing it with the blood alcohol test results for all drivers killed in states that reliably conducted such tests in fatal accidents. Because the two measures agreed closely, Heeren and associates claimed that the proxy, or surrogate, measure would be valid in other states.

c) Construct validity: This refers to the logical relationships among variables we are interested in using. For example, suppose we are interested in measuring an individual’s fear of crime. In order to gather data, we conduct a survey of individuals in a particular community and ask each survey respondent whether they typically leave their home after 11:00pm. The degree to which this is a valid measure of fear of crime depends on whether individuals who fear crime more are less likely to leave their home after 11:00pm. If there is a strong association with fear and remaining home late at night, then our measure would meet the requirement of construct validity. If instead, leaving one’s home after 11:00pm depended mostly on whether an individual works the night shift, then our variable would probably not meet requirement so construct validity.

4. Reliability: Reliability refers to whether a particular measurement technique applied repeatedly to the same thing will repeatedly yield the same result. Reliability is really about the consistency of a variable in terms of measuring the underlying concept you are interested in.

To continue our example from above, let’s assume, despite its questionable validity, that we would like to use the SAT exam as a measure of college readiness. We would need to know whether a student’s SAT score is a reliable measure of his or her underlying ability with respect to the SAT. In other words, if I know that a student received a score of 1500 on the SAT exam, does this mean that the student would score 1500 again if I asked him to retake the exam? What if I asked the student to retake the exam a third time? Or a fourth time? Of course, if a student took the SAT exam 100 times, he or she would probably receive many different scores. The student’s scores might range from 1400 to 1600. The degree to which there is variability in these scores reflects how reliable the SAT score is. If the scores ranged from 800 to 2200, we certainly wouldn’t think a student’s SAT score, measured from only one test taking experience, was a reliable measure of anything. However, if the scores ranged from 1470 to 1530, we might conclude that the exam is, in fact, a reliable measure, whether or not it is valid.

As you think through issues of validity and reliability, it may be useful to go home and think through a couple of examples. Maxfield and Babbie have a nice visual depiction which may be helpful:
The above picture shows targets which you might imagine were used in target practice during a pistol course. In the first panel, the individual has consistently missed the target. However, all of his shots missed in the same area. Thus we can say that we can reliably predict where the shots will end up — the problem is that they do not end up where we want them to end up. As such, this shooter is reliable but not valid. In the second panel, the shooter is, on average, around the target but his misses are all over the place. Sometimes he fires too high; other times he fires too low. Likewise, sometimes he misses to the left; other times to the right. This shooter is valid but not reliable. In the final panel, the guy is a crack shot. He is consistently hitting the target. Thus he is both valid and reliable.

3.0.8 Data

Since we’ll be referencing “data” a great deal throughout the semester, I’d like to pause here and say more about what data are and what data look like. Before defining data, let me bring up a strange fact about data – the word is plural. Hence, we will make statements such as “the data are well-measured” rather than statement “the data is ready to be analyzed.” Do you find this confusing? If so, you are certainly not alone. I am a native speaker of English having been raised in a household in which only English was spoken and, somewhat embarrassingly, I did not figure out that data are plural until I was 23 years old and had a master’s degree! To this day, sometimes I slip up and treat the word as though it were singular.

So what are data? Simply put, data consist of a collection of variables. These variables, in turn, describe the attributes or characteristics of each unit of analysis in our dataset. Let’s consider an example in which the unit of analysis is police officers working for the Cincinnati PD and the variables measure characteristics of these officers. Such characteristics might include the officer’s age \((A)\), the officer’s gender \((G)\), the number of years of experience the officer has as a police officer \((EXP)\), the number of years of experience the officer has with Cincinnati PD \((SEXP)\) and the number of arrests the officer made in the previous year \((Y)\).

What do data look like? It will always depend on how these variables are stored and how the researcher chooses to arrange them. But here is a reasonable depiction of what the data on Cincinnati PD officers referenced above might look like (note: the data are not real):
<table>
<thead>
<tr>
<th>Last Name</th>
<th>First Name</th>
<th>A</th>
<th>G</th>
<th>EXP</th>
<th>SEXP</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones</td>
<td>Charles</td>
<td>39</td>
<td>M</td>
<td>16</td>
<td>12</td>
<td>62</td>
</tr>
<tr>
<td>Johnson</td>
<td>Rachel</td>
<td>29</td>
<td>F</td>
<td>4</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>Mason</td>
<td>Joseph</td>
<td>58</td>
<td>M</td>
<td>30</td>
<td>28</td>
<td>42</td>
</tr>
<tr>
<td>Smith</td>
<td>David</td>
<td>30</td>
<td>M</td>
<td>6</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>Thomas</td>
<td>Richard</td>
<td>32</td>
<td>M</td>
<td>18</td>
<td>14</td>
<td>45</td>
</tr>
</tbody>
</table>

Let’s answer a couple of quick questions about the data:

1. Which officer has the most total years of experience?

2. Which officer has the most years of experience with Cincinnati PD?

3. Which officer has spent the smallest percentage of his/her career with Cincinnati PD?

4. Can you spot a likely error in the data?

Bonus? Can you suggest a procedure for spotting such errors?

Answer: Let \( CHECK = A - EXP \). If \( CHECK \leq 18 \), then flag as an error. What might we call our new variable, \( CHECK \)? Well, check is actually the age at which the officer began working as a police officer. If the age is less than 18, this is likely an error since police departments do not, as a rule, employ minors.