Chapter Five: Measurements

Dr. Simona Simona

4/12/2020

This chapter will address the following topics:

- The concept of variability
- Conseptualisation and operationalisation of variables
- Variables and variable types
- Levels of measurements

The concept of variability

The concept of variability encapsulates all quantitative methods and statistics. Variability is foundational to all social research and data analysis without which there's no study which can be conducted. What is variability? It is basically the difference in conditions among individuals or objects. The opposite of variability is ***constant***, which means lack of variation or inability to change. Now imagine that the world was full of people of the same age, sex, socioeconomic status, health status, ethnicity, dwelling spaces, consumption, culture, dress code and everything else. What would be there to study? We would know pretty much everything there is to know and there will be no need for any research. Such a situation is of course unttainable. But it is important for us to conceive of it so that we may appreciate the inavitability of variability and it forms the core of all research methods.

Variability ensures comparability between and among individuals or objects and that is what instigates scientific studies. All individuals possess ***attributes*** that set them apart from others. Developments in science have shown how each one of us has their own biological makeups called the genome which are specific to us. You may be familiar with how an impression left by the friction ridges of a human finger has become an important piece of evidence in criminal investigations. This is because each person has unique fingerprints, different from anyone else who has or yet to exist in the world! These are typical examples of the importance of variability.

In the social science, variability may not be as rigid as the genome or human fingerprints. However, the differences in social attributes that exist between people in the world could raise very interesting questions that may warrant further exploration in terms of social scientific research. Variations in educational status, income levels, labour participation, age, gender, marital status, health status could raise question which make us want to compare one or two attributes. For example, we may want to compare ethnic background and educational outcomes to study whether someone's background could affect how far they go in education. We may may also want to find out whether

gender inhibits or bolsters our chances of acquaring higher income. These are possible to research on because ethnicity, educational level, gender and income are not constants but vary from one person to the other. These social attributes are called **variables**. Variables are the main ingrediants of research inquries especially in quantitative social sciences. We will, in this chapter delve into the nature, types and measurements of variables and explore how we can appropriately use them to conduct meaningful research inquries.

Variables and Variable types

Variables are attributes that can differ from one observation to the other. We have already seen what observations are in Chapter Three. They are basically entities for which we have a series of variables. They are sometimes called **cases**. In the social sciences, we typically use individual persons as observations, who would have several variables such as age, sex, employment status, ethnicity, marital status, educational status, income etc. Each variable will have **values** associated with it. Values are scores that observations will have have on a particular attribute or variable. For instance, sex as a variable may have two values, *male* and *female*. The variable 'age' is likely to have many values depending on the type of study you are conducting. You may have age values like 24, 26, 38, 54 etc. The variable 'educational status' will have values such as *primary*, *secondary* and *tertiary*.

Variables in the social sciences are often distinguished by the role they play in the research question (whether they are dependent, independent, intervening, moderating or control) and the levels at which they are measured (whether it is at nominal, ordinal, interval or ratio). In fact, before we do anything which remotely resemble data management or analysis in any software, we must have a good understanding of the types of variables we will be dealing with and how they are measured. This is because variable types have implications on the nature of analysis we are able to do. That is why we will be spending a lot of time in this chapter to cement that understanding. But before we do that let's first explore how we arrive at the variable that we are able to measure.

Conceptualisation and operationalisation of variables

Where do variables come from and why do we choose to focus on some variables are not others? Well, variables are normally a characteristic of quantitative studies and they come from the questionnaire or any data collection tool that you have used. Before that, variables are simply operational definitions and before they are operationally defined, they are concepts or conceptual definitions that you wish to consider in your study. Before they are conceptually defined, they are your research problem or research topic. This is kind of a rollercoster but I wanted to demonstrate that the link between your research topic and variables. It is actually uncommon to find your key variables mentioned in your research topic. Some quantitative research instructors will insist that this be the case. I may be interested in studying the effects of education on health outcomes. You see here that my variables of interest (education and health outcome) are in my topic. It becomes very easy for me to remain focused throughout the research process.

This means you are to think about your variables of interest right at the time you are thinking about your research problem. Therefore, all you have learnt so far about the sources of a research problem applies here. The overarching source of variables in quantitative research are *theories*.

Most of quantitative studies (but not all) begin with theories. Hopefully you understand by now you understand why this is the case from your studies of philosophical foundations of social science research. We use theory to explain, understand and predict social phenomena. The choice of the

theory used in our research is often aligned with our world view. So even if theories may appear to pop up innocently in a research paper it is often deep-rooted in the author's own predispositions, beliefs and biases. The important thing though is that the theory that you choose determines the nature of variables you will be researching.

People who study social justice, emancipation or human rights might want to use Pierre Bourdieu's theory of capital and Amartya Sen's capability approach to ground their studies. The theory of capital considers such variables as income, property ownership, social networks, power and educational status among others, to be important concepts and variables of research. The researcher adopting the theory is inclined to utilise such concepts and variables than one who is studying the causes and consequences of rising crime rates in urban centers.

However, sometimes research is not preconceived. Which means there are no benefits of theories from which to generate variables and drive the research agenda. For instance researchers may be part of a team tasked to execute a multifaceted project commissioned by governmental on nongovernmental agencies which has a research component. In this case, researchers do not determine the nature of variables to be studied and how they are defined. They are only involved in analysing the data with variables that are arising from the main project.

Other scenarios where variables may not be predetermined by researchers is when a study is born out of curiosity and observation of regularities. Patterns of a social phenomena can be established out of curiosity due to consistencies in observed regularities. This can be from the social media, traditional media or even from one's own observations. For example you might notice that there is a lot of talks about fake posts on social media and you decide to do a study on it. You will not have already made variables but you engage in a study on the basis of hunch believing that you will arrive at useful information in due course

Whatever the sources may be, variables are often difficult to pinpoint initially because theories and other sources of variables do not bring out refined variables. Theories are notoriously abstract and general. In order for the concepts derived from them to be appropriately measured, they have to be trimmed off of ambiguities that may render them less feasible in scientific studies. In this regard, variables are first given **conceptual definitions**

Conceptual definition

Conceptual definitions tell us literal meanings of concepts and variables that are key to a particular study. Conceptual definitions are important because they provide dictionary definitions of concepts so that we have a good understanding of what the concept means before collecting any data about them. We cannot measure anything without knowing what we are measuring. It is important to note though that sometimes conceptual definitions of our research concepts may depart from their everyday use and even from dictionary definition. This is fine as long as our definition is well understood within our discipline and well described in terms of how it is being applied in our research project. It is not uncommon for specific disciplines to allocate different meanings to every day concepts and even completely manufacture new ones.

In my research, I may want to define the variable 'income' as a household's total income earnings. This gives me a good general picture of the meaning of the variable I am studying. This variable may look straightforward but in research terms which requires the highest level of specificity, things could get complicated very quickly. Total income earnings could mean a lot of things. It could mean daily household income, monthly household income, annual household income, life-time household

savings or in some societies income for the head of the household among others. However, even if the conceptual definition was to be refined to mean income earnings per household per month. It would still be problematic in the research environment. If I asked you to go and collect data on people's monthly household income, it will not take long before you come back to me and ask for more specifications. For example, do I mean the disposable (after tax) or gross (before tax) income? Is it income that remains after monthly expenditure or before? Is it only income that comes from a salary or can we also consider income other sources as well? You can see now that conceptual definitions are not adequate when it comes to actually measuring our variables. They are just the beginning of definitions that gives us a general understanding, but we need specific rules and procedures that allow us to measure income consistently from one respondent to the other. For this we need to operationalise our variables.

Operational definition

If conceptual definitions tell us the meaning of variables, an operational definition tells us how to measure them. It gives clarity in terms of what exactly we would look for when we go collect data. In the income example above, we can give the variable 'income' an operational definition of household's monthly disposable income. This is the income that remains after income tax has been accounted for. We can now use this definition to measure 'income' within households.

Whether or not disposable income is good measure of household income is a legitimate question. In fact, it is a question that represents one of the major points of disagreement within quantitative research. Operationalisation of a particular variable can take different forms and each of the forms is without problems. We used monthly disposable household income as the measure of the variable income in our example. But we might as well have chosen individual income, discretionary income, annul income etc. How do we decide to choose one over the other operational definition is the problem of **construct validity** (Argyrous, 2011). Construct validity is the degree to which a test measures what it purports to measure. For example, to what degree does disposable income measure household income? A good operational definition is one that will vary when the variable it purports to measure varies. For income, the monthly disposable income that we have chosen to represent it must vary consistently across different households if it is to be a good measure of income.

However, this may not be difficult to achieve in some circumstances. We could record two households as having the same amount of 'income' when that may not be the case. Imagine two households responded with the same amount of disposable income to your question in your survey asking people what their disposable income was in the month prior to your survey, but one of the families receive extra monthly allowance from an adult relative staying outside the household. Clearly there is a variation between the two households which has not been captured by the measurement you have used. Imagine also that the source of income for one of the households is their own investments while that of the other is from their salary, surely although you have recorded these two households as having the same income, the reality is that they are different in very significant ways. In this regard, monthly household disposable income may not really suffice as a good measure of income.

Another example we can consider here is that of gender based violence (GBV). It is conceptually defined as harm inflicted upon individuals or groups of people on account of their gender. Imagine you were studying prevalence rates of GBV in a particular community, can you think about the operational definition you would use to measure GBV? Whom would you categorise as the victim of GBV? You might want to say anybody who has ever been assaulted by a member of the opposite sex. This definition may not be adequate. If a husband is assaulted for beating up children in the

family, does that qualify to be GBV? What about a woman who is overly scolded by the boss at work because she didn't turn for a workshop presentation. Does that qualify? The complexities of this variable is compounded by the fact that GBV is a cultural specific problem which may not have a universal applicability. What is defined as GBV in one society may not be defined as such in another. Therefore, when you are asking people about this question, they may have different interpretations depending on their own context. This renders operationalisation problematic because you wouldn't properly capture the variations between respondents that are necessary to ensure the validity and reliability of your research findings.

Variable operationalisation is further compounded by the fact that human beings, the subject matter of most research in social sciences have the capacity to conceal information.

There are a few factors that operationalisation of variables problematic and these include complexity of the concept, availability of data, and ethics.

Complexity of the concept. Although we have shown income as being very complicated. It is relatively easy compared to some variables like GBV and health status. In addition to the contextual complexities of GBV we have highlighted above, it also embodies different unequal dimensions. Violence can broadly be distinguished in terms of physical, sexual, emotional and structural. All of these types of violence can be classified as conceptual definitions in themselves. If we operationalise violence by focusing on the physical dimension of violence for example, we will have a risk of not observing important variations in this variable. But of course, even if you have selected physical violence as your area of focus, you will still face those contextual complexities as to who qualifies to be called a victim of violence. GBV has high complexity and that makes operationalisation very difficult.

Access to data. For us to pull out a good study we need to have access to the data. This is the reason why as a researcher it is important for you to consider the feasibility of your study very careful before you embark on a research project. Resource, time and ethical limitations can impede with the acquisition of information about a research question operationalised in a particular way. This is more so in small-scale research projects which are not commissioned by the government or non-governmental organisations like those conducted undergraduate and postgraduate students in institutions of higher learning. If for example, we thought the best way to measure the prevalence of heterosexual rape was to count and interview the people convicted on rape charges, we would need to have access to the prisoners and also review court, police and prison records. However, we may be prevented from doing so by the authorities on ethical grounds. Even if ethical considerations were not a factor in this case, the process of acquiring permissions to access the data may take too much time and resources. As a result, you may decide to tweak the operationalisation so that you can just collect information by interviewing prison officers which may not require stringent access mechanisms but of course, doing so will rob you from dealing with desirable data.

Ethics

I have already touched on ethical considerations above. Ethics are moral principles agreed by governments, communities and research institutions to shape the conduct of researchers. They involve judgement of right and wrong in the research process. Ethics might prevent you from pursuing answers to certain research problems either by the rules and procedures put in place by your institution. The academic world is built on the principle of collaboration and every institution of high learning has great appetite to exhibit excellent ethical credibility because it ensures trust and thus continued corporation with other institutions, policy makers, industry and the public at large. Your institution will ensure that your research abides by ethical research principles such as

minimising the risk of harm, obtaining informed consent, protecting privacy, avoiding deception and providing the right to withdraw[1]. These principles are good. They are there to protect both the researcher and the human subjects and they should be followed by everyone.

However, these principles have often times constrained the ability of researchers to carry out desired inquiries. This happens when the variable of interest is operationalised in a way that requires access to closed social spheres or secretive societies. Imagine the operational definition of your variable is to observe the number of times a member of the religious sect deviates from internal rules and regulations. You may determine that the best way to carry out such a study is through covert participant observations, but you may feel constrained by ethical implications of your study and opt instead to just interview experts in the field of sociology of religion or cultural anthropology. You will definitely not get as rich information as you would if interviewed your ideal respondents.

Operational definitions that targets organised criminals like mafias, football hooligans, gangs and drug lords have similar ethical implications which forces researchers to negotiate a fine line between research interests and ethical principles. Unable to ensure their safety and credibility of their institutions, researchers have many abandoned their desired line of inquiry in favour of a more realistic approach that compromises their research agenda.

Dependent vs. independent variables

I have mentioned above that variables are often distinguished by the role they play in the research process and that often times key variables in quantitative researchers are identified right in the research problem or topic of research inquiry. The main type of variables in quantitative research are dependent and independent variable(s). These form the fulcrum of quantitative research and hardly a quantitative study exists without *both* types present. A research inquiry can have one dependent and one independent variable but that has to be a very small study or one which is done only for instructional purpose and not intended for a scientific publication. We will explore why this is he case in this section but we first need to explore what these variables are.

A **dependent variable** is an attribute that the researcher wishes to measure, explain or predict. If I want to study the impact of educational qualification on employment status. The variable I wish to explain is employment status. One of the problems with quantitative research is that one concept can be explained by different terms which sometimes confuses new comers to the field. This is one of them because sometimes the dependent variable is called 'outcome' or 'response' variables. It means the same thing.

An **independent variable** on the other hand, is an attribute that is believed to measure, explain and predict variation in the dependent variable. In the example above, I want to use educational qualifications to explain variation in employment status. Educational qualifications is my independent variable. Again, you will find several terms in literature used to describe independent variables including explanatory factors or variables, predictor variables, exposure variables or just covariates.

It is difficult to must these when you are meeting them for the first time. I struggled with them as well but I will show you the approach I use to distinguish between them. I think about in terms of cause an effect. The causes are the independent variables and the effects are the dependent variables. When I come across two variables, I ask myself which one is likely to cause the other here.

[1]There are details about these principles I have not bothered to go into here because this section is about operationalisation of variables and not research ethics. A good treatment of these as they pertain to social science research can be found in Alan Bryman's *Social Research Methods*

Immediately you ask yourself that question you will not have problems with these two variables again. In our example above I have educational qualifications and employment status and I ask myself which one is likely to cause the other here? If I perform a little mental experiment, I would come to a conclusion that having good education is likely to make someone get a better job. In the same vein, I would struggle to construct how employment status may cause educational qualification. It certainly does but not as straightforward as the reverse. Therefore educational qualification is the cause and thus independent and employment status is the effect and thus dependent.

Indeed, it is very easy to identify the dependent and independent variables in a research problem where the reverse relationship is implausible. For example if you are studying the relationship between age and health status, it is clear that this relationship only has one causal direction. Age influencing health status but health status can not influence age. It is easy to identify age as the independent and health status as the dependent variable. However, if we were looking at health status and income deprivation, distinguishing between the two types of variables become much more complex. The causal direction of these two variables can go both ways. On one hand, income deprivation can cause poor health and on the other hand, poor health can cause income deprivation. In this case, you distinguish your two variables in terms of the variable you are trying to explain and what you are using to explain like we said in the definition. If I am attempting to study population health outcomes and how they are influenced by income deprivation, then my dependent variable is health status. But if my interest is to look at the impact of poor health on income deprivation, then income deprivation is my dependent variable.

Of course, we should note that we are using cause and effect casually here just to help us identify our key variables. Otherwise, cause and effect as a goal of scientific research is a thorn issue in social science research methods. There are only a few methods that can offer cause and effect. In most cases our research in the social sciences ends at establishing associations or relationships between variables.

I mentioned earlier that we would rarely (if any at all) find a published study that only looked at two variables. What is typical is to have several explanatory variables[2] used to explain one outcome variable[3]. The reason why this is the case is because social phenomena is complex and it can hardly be adequately explained by one variable. Any social phenomena you might think of has several interacting factors associated with it. In research, it is our aim to represent as much reality as possible. Whether we are studying the emergence of fake news on social media or the rise in crime rates in urban centers, there is no one magic bullet variable responsible either factor. We have a better chance of understanding them if we explore several factors.

We may still be interested in using one factor to explain the other. Indeed, this perfectly fine. For example, we might want to study the effect of income deprivation on crime rates. But the fact that there are many other variables that can be used to explain crime rates other than income deprivation causes problems to our study. We can not be too sure how income deprivation compares with other unknown factors in explaining variations in crime rates. Sometimes income deprivation may not have a direct impact on crime, it may be helped by other factors in doing that such as age, level of education and marital status. This problem is solved by including a reasonable number of possible factors that is believed to influence crime rate in our study. We will learn the practicalities

[2]There is no rule of the thumb for the number of explanatory variables you need in your study. It all dependents on the nature of the research inquiry you are pursuing

[3]Notice that I have changed the names of our key variables. I will be using the names interchangeably in this book because that is what you would expect in the literature. It is important to familiarise ourselves with these terms now so that they will not confuse us later

of doing this in later chapters. These extra variables that we use to explain our dependent variable are generally dependent variables but they have their own special names and that is what we explore below.

Other types of variables

Extraneous variables This is a general term given to all unwanted and mostly unintentional variables that may affect the results of your study. These are typically weak and may not affect your study in a significant way but it is important to be aware of them so that you can minimise their impact even less. Extraneous variables include **experimenter effect**, **participant effect**, **situational variables** and **demand characteristics**.

Experimenter effect occurs when the researcher unintentionally (and even intentionally) influences the research outcomes. This can happen at stage in the research process from design to data analysis. The research process is largely controlled by the researcher. They choose the topic of study, tools to be used in data collection and methods of data analysis. At any point in this processes, their biases, prejudices, beliefs and expectations can creep into their research agenda. These may also be projected unintentionally onto the research respondents as cues about how they should behave, act or respond to questions. A more extreme form of experimenter effect is when the researcher doesn't like what is coming out of the study and is tempted to doctor the results according to their expectations. Consequently, the research outcomes may be more of the researcher's viewpoints rather than that of the respondents.

To mitigate against experimenter effect, researchers should constantly be **reflexive**[4] about their role in the research process. They need to keep checking if they have began to be motivated by factors other than the interest of science for its own sake. Other factors can be pre-registration of studies so that the procedures are stipulated before the study is conducted and could not be changed by circumstances arising from the study environment. Furthermore, triangulation of researchers could provide checks and balances – making sure that the research doesn't depart significantly away from the standard disciplinary procedures. Moreover, rigorous training of interviewers and research assistants can also to minimise experimenter effect.

Participant effect occurs when bias in the research process is induced by the research participants. Research participants are not the same. They vary from one another on several fronts including age, gender, socioeconomic backgrounds, intelligence, level of concentration, anxieties and motivation among others. Variations in these and many other characteristics could affect the research outcome either positively or negatively. In sociology a typical example of participant effect is what is call the **Hawthorne Effect**, a situation where an individual or group of people know they are being observed and so change their behaviour accordingly. Some respondents might come in the research environment with prior knowledge of the purpose of the study and would be motivated to react in a certain way to either advance their interests or suit a preconceived perception. Some motivations such as socially desirability and compensations are also examples of participant effects and can affect the research outcome.

The best counter-measures against participant effect are **random sampling**, **double-blinded designs**, **computerised data collection** and **covert observation**. Randomisation is important because it helps to evenly distribute the confounding introduced by participant effect across the selected sample(s). Double-blinded design is the type whereby researchers and participants don't

[4]The concept and practice of reflexivity in very popular in qualitative research. It means being able to your own norms, beliefs, politics, feelings and desires and how these influence your behaviour and action in a particular situation

know each. Computerised data collection has an element of double-blindness to it. Additionally, data can also be collected retrospectively i.e after the research participants have already produced the data which are the subject of research. Covert observations are basically deception i.e the research participants are not aware that they are being observed for research purposes.

Situational variables are aspects of the environment that may affect the participant's responses or bahaviour during data collection. Some of these may be easy to identify and thus controllable. They can be things like noise, high or low temperature, hunger, etc. Others which are consequences of the social and cultural environment may be more subtle and difficult to determine. For example, the presence of husbands and older people during the interview in some societies can impede truthfulness among women and youthful respondents.

It is important to standardise procedures in research to ensure that conditions are the same for all respondents. Situation and cultural awareness are equally important important to minimise the negative effects of situational variables.

Demand characteristics are clues that give away the purpose of the research to the respondents. These can compel the participants to react in a manner that conforms with expectations. Research respondents can be influenced by the the researcher's characteristics, nonverbal communications and their own perceptions. The researcher's gender, dress code, accent and language can motivate respondents to alter their responses and behaviour. Keeping the research environment natural, researcher triangulation and adopting standard procedures and regulations can help reduce demand characteristics

Confounding variables

When you fail to deal with extraneous variables and they creep into your study providing alternative explanations to the relationship between independent and dependent variables, they become confounding variables. Confounding variables can introduce ***spurious relationship*** (showing correlation which does not exist in the real world data) between variables. But there are too many extraneous variables in any one study and it is not always the case that whenever they are introduced in the study, they become confounding. Confounding variables are only created when extraneous variables cause confirmed threat[5] to internal validity. This will be discussed further below when you look at validity and reliability in social research.

Intervening variable Intervening variables are abstract or hypothetical constructs that can provide the link between the independent and dependent variables. Intervening variables are not normally measured in research projects but they are usually common sense logical connection between the key variables being studied. For instance if you are interested in the relationship between education and income, you may find that people who have higher educational qualifications tend to have higher income. But this relationship has no causal connection if people are not in employment. Employment in this case acts as the link between education and income.

Another example would be socioeconomic status and life expectancy. We know that poor people or those who live in poor environments have shorter life span. However, there isn't any real connection between poverty unless linked by poor healthcare systems, poor transport to health centers and lack of qualified medical personnel among others. We can say that these are variables are intervening in the relationship between socioeconomic status and life expectancy. Sometimes researchers call these variables ***mediator variables***.

[5]One that has has been shown to be so through empirical studies

Since intervening variables are not measured in empirical studies, researchers are not usually able to quantify the influence of intervening variables in research outcomes. They can only use them to provide theoretical explanations for research results.

Moderator variable Moderator or moderating variables are variables that have a conditional effect on the relationship between the independent and dependent variables. The value of the moderating variable can alter the strength of the relationship the independent and dependent variable. In the example of the relationship between education and income, we said employment is the mediating factor. However, the availability and quality of employment is moderated by government policy on job creation. If the government policy is not favourable for employment, then the impact of education on income will be less compared to what would occur in reverse circumstances. Government policy is the moderator variable in this case.

Control variables. These are variables that you know (maybe from the theory or literature) can influence your dependent variable but you wish to eliminate that influence in your study so that you can determine the relationship between your main independent variable and dependent variable without bias. Control variables are included in empirical studies to reduce the risk of attributing explanatory power to independent variables that in fact may not be responsible for the variation in the dependent variables. You do this by *controlling* the effects of these variables in your research inquiry. Again going back to our education-income example, we know that experience does determine not only to your ability to acquire employment but also the amount of money you will be paid in that employment. Therefore for use to properly measure the effect of educational qualification on the level of income, we need to control for years of experience. Yes! experience is the control variable.

Discrete vs. continuous variable

Variables are also classified with respect to their unit of measurements. In this regard, variables can either be ***discrete*** or ***continuous***. Discrete variables are those whose numerical values can only be measured in whole numbers or positive integers. This means their unit of measurements cannot be broken down into smaller or finer pieces to take on float or decimal numbers. Number of hospital admissions, number of friends on social media and number of children per household are all discrete variables because people or human beings are their unit of measurement. Human beings can only be counted in terms of whole numbers and as such these variables will never take on decimal values. You cannot say 8.5 people were admitted in the hospital or that she has 20.8 friends on Facebook. Other discrete variables include number of political parties in a country, number of vehicles on the road, number of countries in a continent etc.

Some researchers describe discrete variables as those that take on distinct and countable values. We can count the number of social media friends that anyone has, we can count the number of people admitted in a hospital and we can indeed count the number of vehicles on the road. Some may take us quite some time to count but the point is that discrete take on finite amounts. Compare with a variable like time, which can take on an infinite amount of values and good luck on trying to count all the values can time can take. Time is an example of a ***continuous variable***

Continuous variables are those whose values can be broken down infinitely into smaller pieces. Values of continuous variables consist of real numbers[6]. Any distance between two values can take on infinite more real numbers. As such, values in continuous variables cannot be counted. I have

[6]These are all numbers that can be represented on a number line including integers, natural numbers, fractions, rational numbers and irrational numbers

already given an example of time which can take on infinite number of smaller values. Time can be dived into days, hours, minutes, seconds, milliseconds, nanoseconds etc. The distance between any two values of a unit is infinite. For example, the distance between 1 and 2 minutes may begin from 1.00000001, 1.00000002, 1.00000003 and so on. Even then I have broken the number of zeroes but they can continue up to infinite and we would never reach 2. Other types of continuous variables include age, height, weight, temperature, income and so on.

Unlike discrete variables, continuous variables can take on decimals as we have seen in the above example. In practice however, researchers usually work with continuous variables rounded to smaller numbers of decimal places. Also, it is not uncommon for researchers to convert continuous variables into discrete variables that can easily be counted. For example, we may choose to consider age in terms of number of years a person has lived or height in terms of the number of meters or centimeters. In this regard, you may find that the values of both variables are represented by similar configurations of numbers and yet they are different. Yet the distinction between the two is important and forms an important criteria upon which different statistical and graphical procedures are chosen in quantitative social science.

Levels of measurements

Another important distinction of variables is on the basis of their levels of measurements. Levels of measurements are a critical in quantitative social science as they determine the kind of statistical methods we apply to our data as well as the interpretation of our research findings. There are procedures can be applied use with variables measured at one level that we may not apply on variables measured at a different level. Variables at the highest level of measurement have meaningful numerical values and a broad range of statistical procedures can be used to analyse them. While those that are measured at the lowest level may have meaningless values that may just serve as labels only and not a great deal of statistics can be applied to them. I would say that levels of measurements are the fundamental building blocks of quantitative methods and statistics and it would be wise for you to develop a good understanding of how they operate.

There are basically three levels of measurements: nominal, ordinal, interval/ratio. The interval/ratio level is sometimes separated in some texts to make four levels but are going to simplify things here by considering them as three levels because the difference between interval and ratio is too minimal. Before we consider each of the levels however, you may want to note that the level of measurements to use on a variable is usually determined by the presence or absence of four characteristics: distinctiveness, order of magnitude, equal intervals and absolute zero. These characteristics are sometimes confusing but should be able to have a good grasp after we explain them in details. Don't worry if you are still rusty after your first go at them. It sometimes takes time for you to be grounded. Like I said in the beginning, this is a new language you are leaning and like the spoken language, it takes a bit of sometime to master the rules but after you have done so you will sour through like a bird in the sky.

Nominal level

This is the lowest level of measurement. *Nominal* is Latin for name only. This level of measurement only has the characteristic of distinctiveness, which means that the variable here only measures category or identity and nothing else. You can use numbers, letters or alpha-numeric symbols to refer to the values of attributes of the variable but these do not mean anything more than just for categorising your values. For example, religion is a variable that is measured at the nominal level.

You may assign numbers to categories of world religions like 1 = Catholic, 2 = Protestant, 3 = Islam, 4 = Hinduism, 5 = Judaism, 6 = No religion. We have used numbers from 1 to 6 here but these numbers do not in any way signify some kind of order between the categories. A person in the Judaism category is nominally different from someone in the Catholic but even if Judaism has the value of 5 assigned to it, it doesn't mean people who belong to that category are more religious than those who belong to the Catholic category. Other variables that are measured at the nominal level include gender (male or female), political affiliation (Labour, Conservative, Lib-dems, Greens etc), marital status (single, married, separated, widowed) etc.

Nominal level variables sometimes have only two categories (yes and no/true and false/ male and female). When this is the case, the variable is sometimes called binary or dichotomous. When these kinds of variables are assigned values of 0 and 1, they are called dummy variables. You don't have to master these for now. I am just giving you a heads-ups because you are likely to encounter them in the broader literature.

We will deal with more with what we can do with nominal level data in terms of statistical analysis in later chapters but suffice to say that at the nominal level of measurements, we can only only compare the relative sizes of categories. We can say that there are more Protestants than Catholics in this country or that there are more people who voted for the labour party than those who voted for the Greens in the previous election. There are no other mathematical operations we can apply to this data. We cannot order it with respect to magnitude, we cannot add or divide the values associated with the categories.

However, there are certain rules and procedures that need to follow when categorising variables to ensure correct measurements. These can apply at any level but we will discuss them here because they make more sense the nominal and ordinal level: 1. Categories should be mutually exclusive of each other without any overlaps. 2. Categories must be exhaustive, which means that there should be no value(s) that is uncategorised. Each category must have only one value or score associated with it. 3. Categories should be comparable or homogeneous. You can mix oranges with apples if you are categorising fruits but you may not add ketchup or gender to that list. Researchers must select categories with respect to their research objects. There are no straightforward guidelines as every research inquiry is unique on it's own.

Ordinal level

Variables measured at the ordinal level still retain distinctiveness or category but the difference from nominal level measurements is that these categories are also ordered. Therefore ordinal level of measurement has two characteristics: distinctiveness and order. This means categories of ordinal level variables are not only different from each other but also that higher values used for categories represent more of the phenomena being measured than lower values. Individual categories are comparable in terms of their order of magnitude. For instance, imagine we wanted to measure the concept of educational status and we assign values to categories as 4 = tertiary education, 3 = Secondary education, 2 = Primary education, and 1 = No education. Here the values assigned to categories are meaningful as the high indicate more education compared to lower values. Thus, we can say that those who have attained tertiary education are more educated than attained secondary education and so on.

Other variables that are often measured at the ordinal level include social class (upper class, middle class, working class, lower class) and personal opinions either represented by a scale of numbers (e.g from 1 to 7) or measured using the likert scale (strongly agree, agree, neutral, disagree, strongly

disagree). It is also very common to find higher level measurements grouped together to create ordinal variables. For example, age may be grouped to make ordinal levels of measurements (15-24, 25- 4, 35-44, 45-54, 55-64, 65+). We will learn more about collapsing higher level measurements in the next sections.

Ordinal level measurements have a limitation of lack of precise distance between categories. The only thing we know is that one values representing categories are ranked. Because of this limitation, we are unable to carry out sophisticated mathematical computations on ordinal level variables. Just like it is the case with nominal level measurements, we can not compute mathematical operations like subtraction, division or addition on ordinal level measurements. Most mathematical and statistical operations we use everyday are only possible because of the assumption of equal distance between numbers. For example, in a set of numbers from 1 to 8 {1,2,3,4,5,6,7,8}, the distance between each of the numbers is 1. Thus, we are able to calculate operations like the mean by adding all the values and dividing by the total number of items. If the distance between 1 and 2 and that between 4 and 5 was not equal, adding the values of the set to calculate the mean would be inappropriate. That's why calculating the mean or other sophisticated computations on the nominal and ordinal level measurements is meaningless.

Interval/ratio level

In addition to distinctiveness and order of magnitude, interval/ratio level measurements adds equal distance between intervals. In other words, with the equal distance characteristics of the interval/ratio level, we are able to determine exactly how much the different values differ from each other. This characteristics is missing in the ordinal level measurement. For example we can tell that strongly agree means more than agree but we cannot tell by how much. On the other hand, if you take an example of age which is measured at the interval/ratio level, someone who is 19 years old is older than someone is 15 years by 4 years. Again, the reason we are able to do this is because interval level measurements allow for equal distance between values.

Apart from age, there are several other variables which are measured at the interval level and they may include income, age, weight, height, life expectancy, number of houses in a village, population size etc. Because of the characteristic of equal distance between intervals, we are able to mathematically manipulate variables measured at the interval level. We can add, subtract and divide values associated with variables measured at the interval level.

Interval level variable have one limitation though. The absence of an ***absolute zero***. What this means is that in an interval level, zero does not always represent the category that possesses no value of the variable. Take temperature for example, zero degrees Celsius is a reading and it doesn't represent a point where there is no heat. This is where the ratio level differs from an interval level because it has an absolute zero or ***true zero point*** characteristics. With ration level zero represents the complete absence of the phenomena being measured. If you take money as an example, zero means the completely absence of money. Money is therefore measured at the ratio level.

However, like I indicated earlier the difference between the interval and ratio levels of measurements is so minimal that we are justified to put them together as interval/ratio because all the computations that we are able to do with variables measured at the ratio level, we can do with interval level variables as well. Note that sometimes the interval/ratio level of measurement is referred simply to as continuous variables.

I indicated earlier that we are able to collapse higher level measurements to lower level ones. In

this regard, we can collapse interval/ratio variables into an ordinal level measurement. We gave an example of age above we can be collapsed into categories. It is important to note though that lower level measurement cannot be upgraded to provide higher measurements. We cannot change from ordinal to interval/ratio measurements

The differences between levels of measurements is predicated on the amount of information we are able to gather from each level. Nominal level of measurements provide the least amount of information as it only provides categories of variables. Ordinal level variables provide a little more information because it can give rank of order to a variable. Interval/ratio provide the most information because they permit us to measure the difference between values. See the table below for a summary of information that can be gathered from each level and the mathematical operations we can perform with the data collected at each level of measurements.