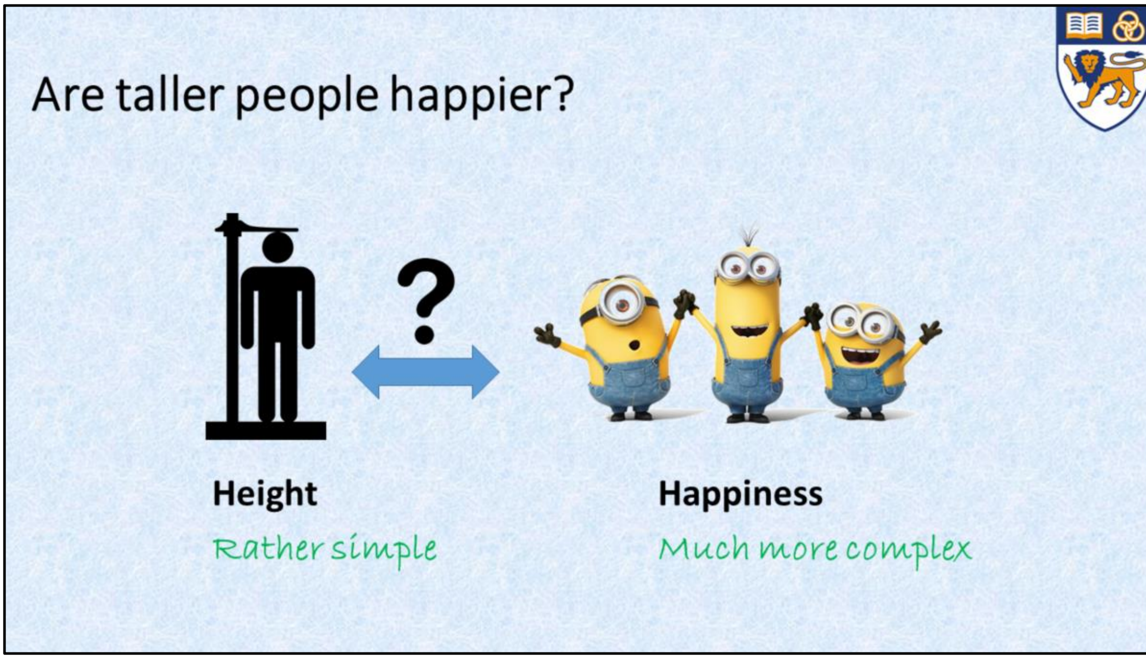
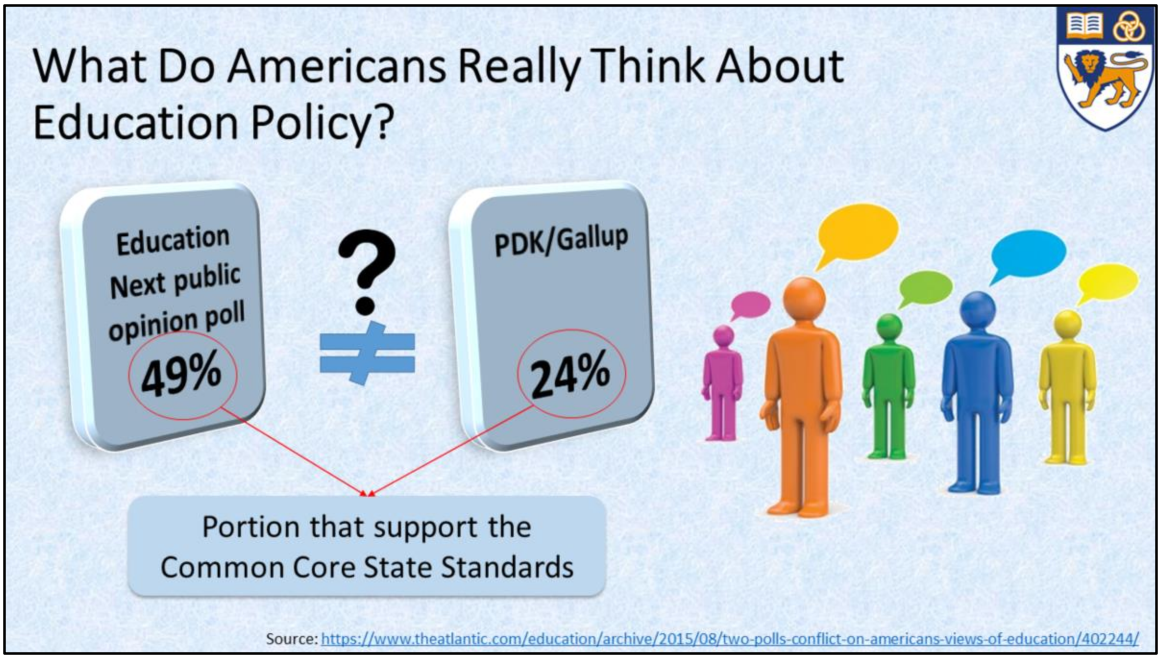
Dr. HassanAli GhaedAmini Harouni

Introduction

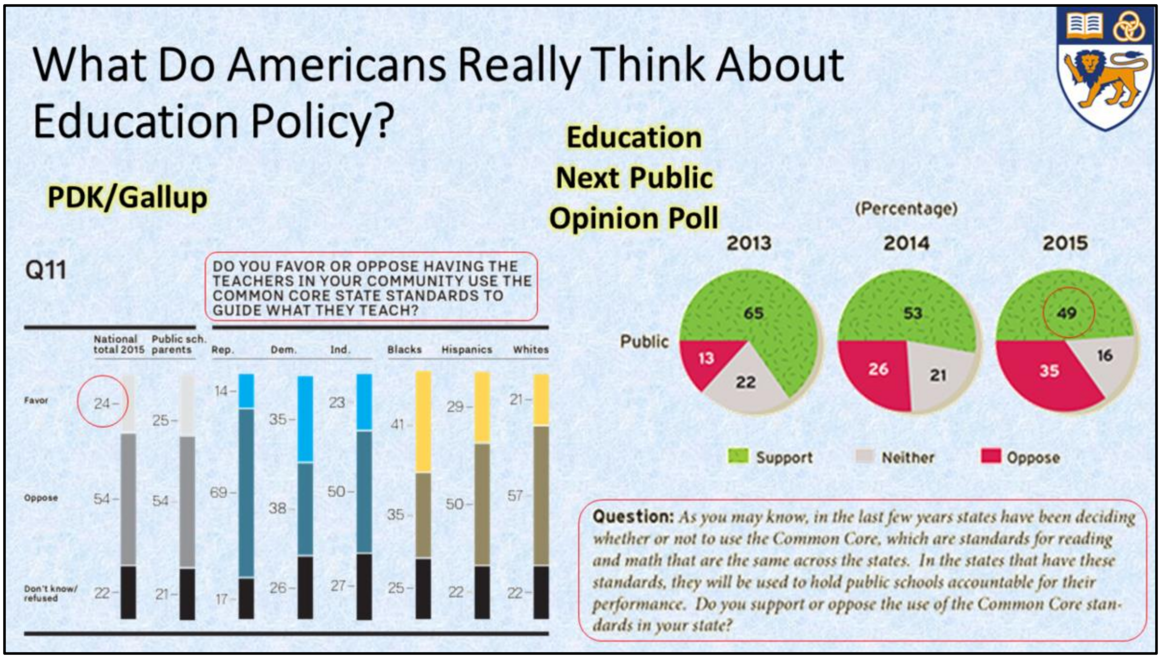
Hi there! I am Ali Amini and with my colleagues, I welcome you to the measurement chapter of quantitative reasoning.



Imagine you want to see if there is a relationship between height and happiness. Measuring height is quite simple. But what about happiness which is something much more complex? Researchers routinely attempt to measure these kind of factors and when you read the results of a study, it is important to understand in some detail how the measurement is done.



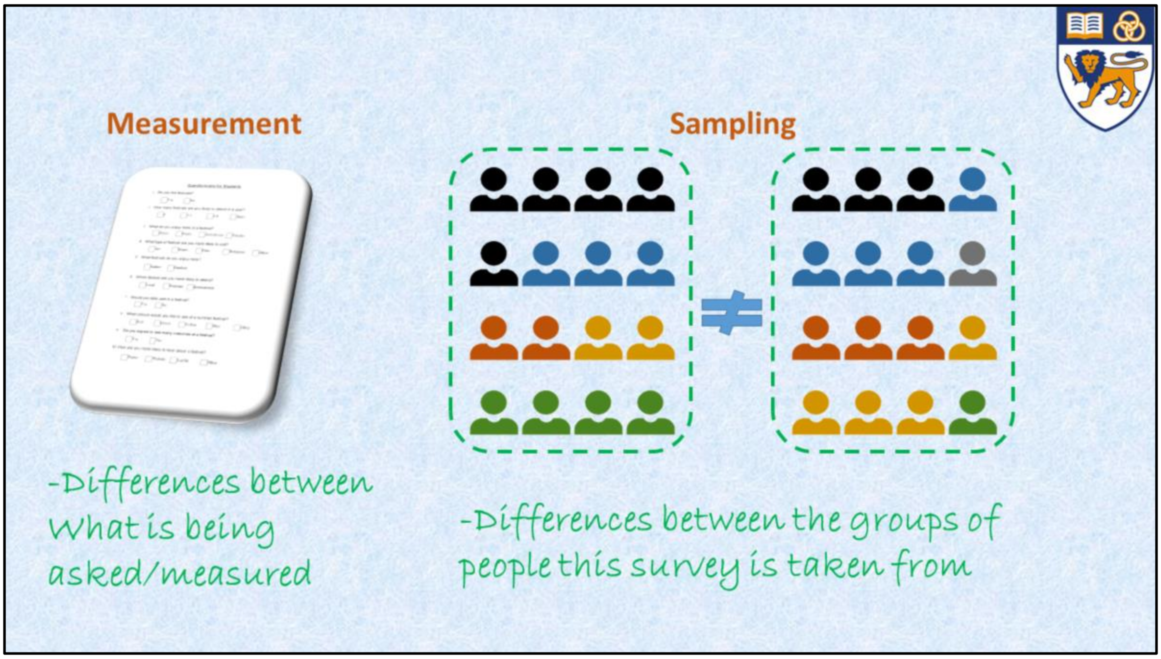
Lets look into a real life example:
When it comes to public opinion, many times we face surveys or polls that have rather conflicting results. For example, in an article published by the Atlantic asking "**What Do Americans Really Think About Education Policy?**", we see two completely different survey results.
In the *Education Next* public opinion poll, 49 percent of respondents said they support the Common Core State Standards; in the other poll, conducted by PDK/Gallup, 24 percent share that view.
What can explain the difference here?



Well, let's look into the information published by the two polls. First, I hope you can recognize the numbers reported in the previous slide. I give you some time to spot them by pausing the video.

Here they are.

Now let's look at the questions asked in the surveys. Are they different? In what aspects? It is now a good time to pause the video and study the survey questions to spot the differences between them.



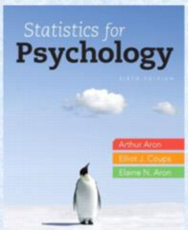
We saw that if the questions we ask about a subject are different, we may see different results and even conflicting ones out of the polls on the same subject. However, you should know that this is not the only explanation. Another reason for the differences can be the differences between the groups of individuals who were polled. The latter is the subject of the sampling chapter that we will cover later. In this chapter, we will focus on measurement, particularly the difficulties and pitfalls that can be encountered when measuring something. Please stay with us.

https://www.theatlantic.com/education/archive/2015/08/two-polls-conflict-on-americans-views-of-education/402244/

Defining a common language

Welcome back! In this unit, we will define some terms that we will use in the rest of this chapter. These definitions will help to form a common understanding.

Variables, Values

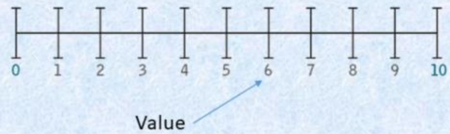


As part of a larger study (Aron, Paris, & Aron, 1995), researchers gave a questionnaire to students in an introductory statistics class during the first week of the course. One question asked was, "How stressed have you been in the last 2½ weeks, on a scale of 0 to 10, with 0 being *not at all stressed* and 10 being *as stressed as possible?*" (How would *you* answer?) In this study, the researchers used a survey to examine students' level of stress. Other methods that researchers use to study stress include creating stress with laboratory tasks (such as having to be videotaped giving a talk for humans or swimming in water for rats) and measuring stress-related hormones or brain changes.

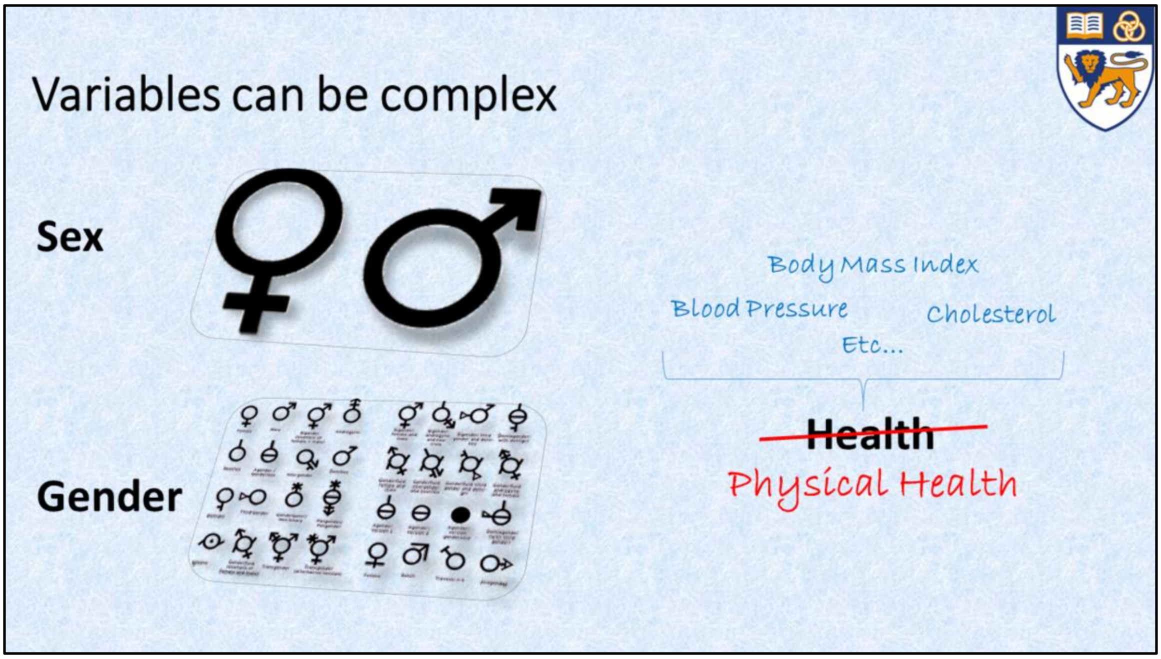
Variable: Characteristic that can have different values or categories

Value: possible number or category that a variable can have

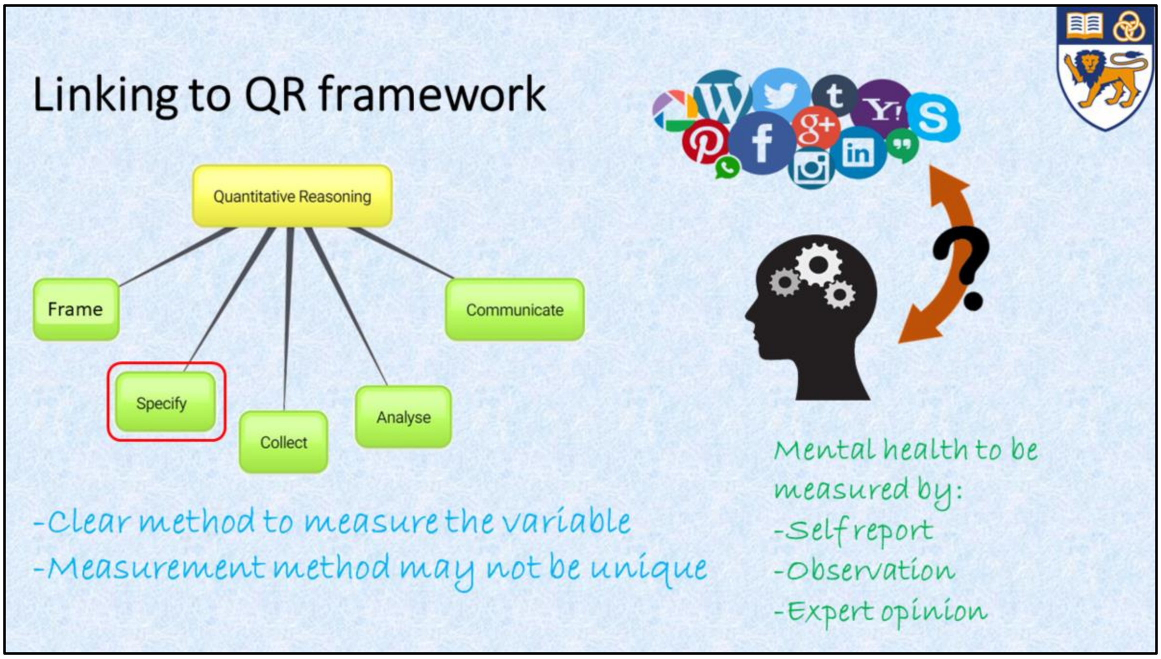
Variable: Stress level



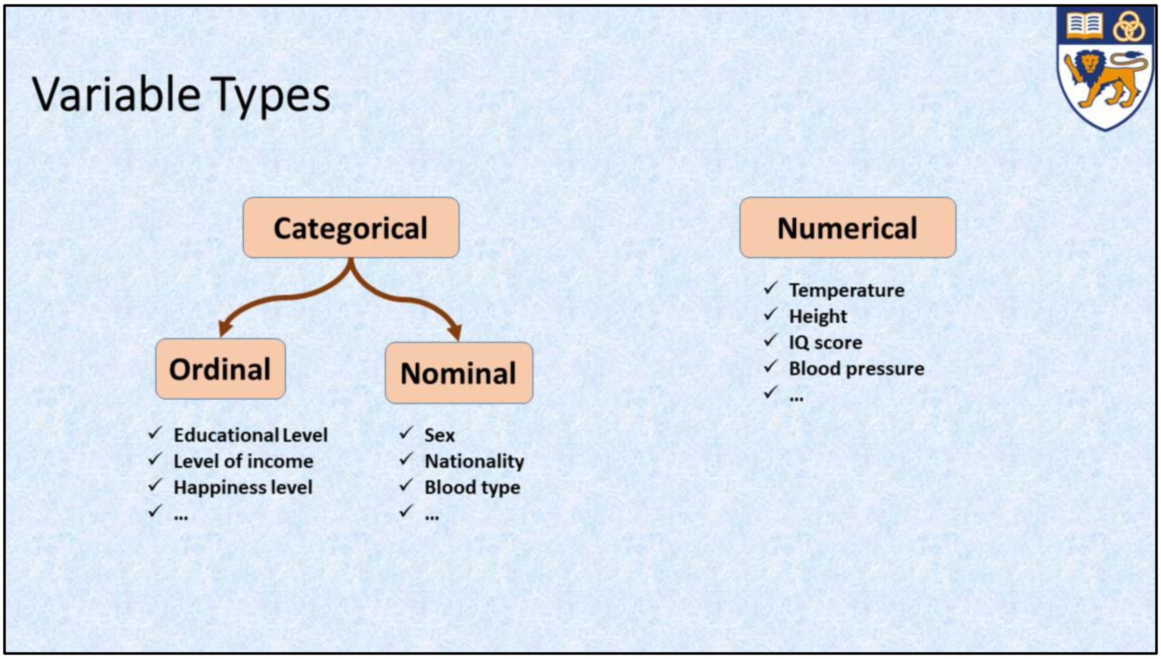
Great! Let's start with an example I have adopted from the book, Statistics for Psychology.
A question in a questionnaire asks the students: How stressed have they been in the past two and a half weeks? Students are supposed to indicate their stress level on a scale of 0 to 10.
In this study, the researchers have used a survey to measure students' level of stress but you should know that this is not the only way. Other ways to measure stress is to measure stress-related hormones or observing brain changes for example.
So, what is a variable? Well, a variable is a characteristic that can have different values or categories. Like the level of education or household income.
Values on the other hand are the possible numbers or categories that a variable can take.



Some variables are easy to define and measure. For example sex is rather an easy variable to understand and measure but gender is more complex. For another example, consider the phenomenon of health. One can measure health by looking into indicators like body mass index, blood pressure, cholesterol, etc… While one may find this logical, one should notice that in this way we are only considering physical health and not mental health. When it comes to a fuzzy phenomenon, what we actually measure may represent only a part of the phenomenon and not all of it. We should be mindful of this limitation.



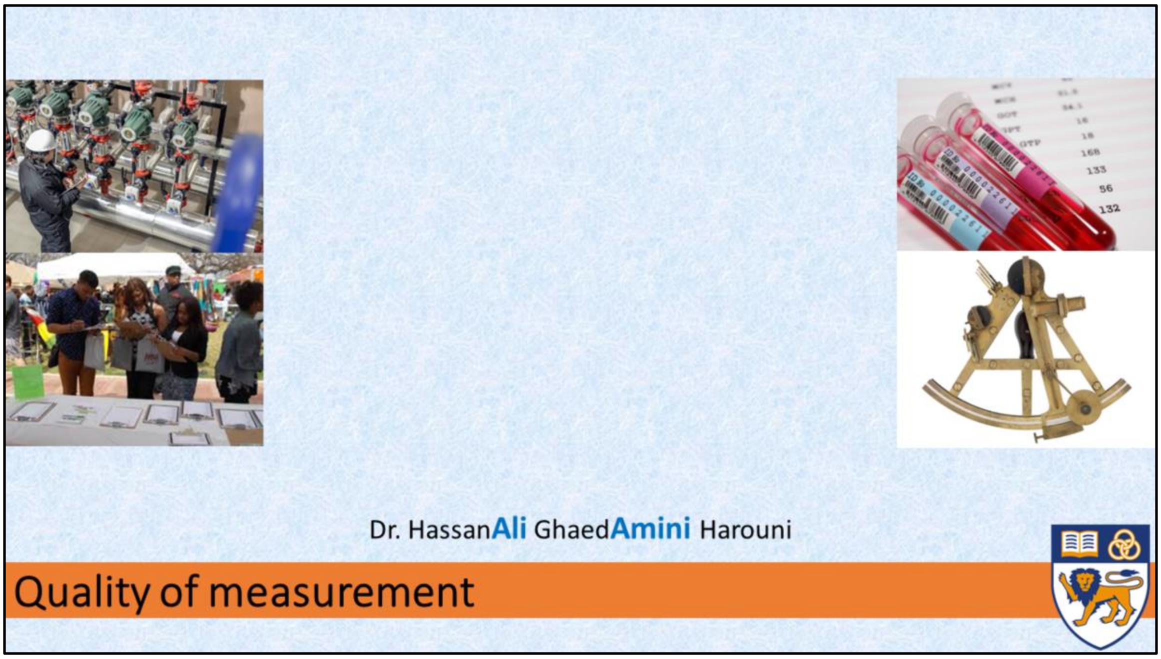
Now it is time to link back to the second step of the QR framework, Specify. Imagine we want to measure the effect of social media on mental health. Well, we can think of several ways to measure mental health. We can directly ask the participants for self-reports, or we can do observations by analyzing participants' social media posts for example, or we can ask an expert to evaluate the mental health status of the participants. As you see, when we identify our variable, we should come up with a clear method to measure it. Indeed the method used to measure something is not unique and it is very important to understand what aspect of it is being measured.



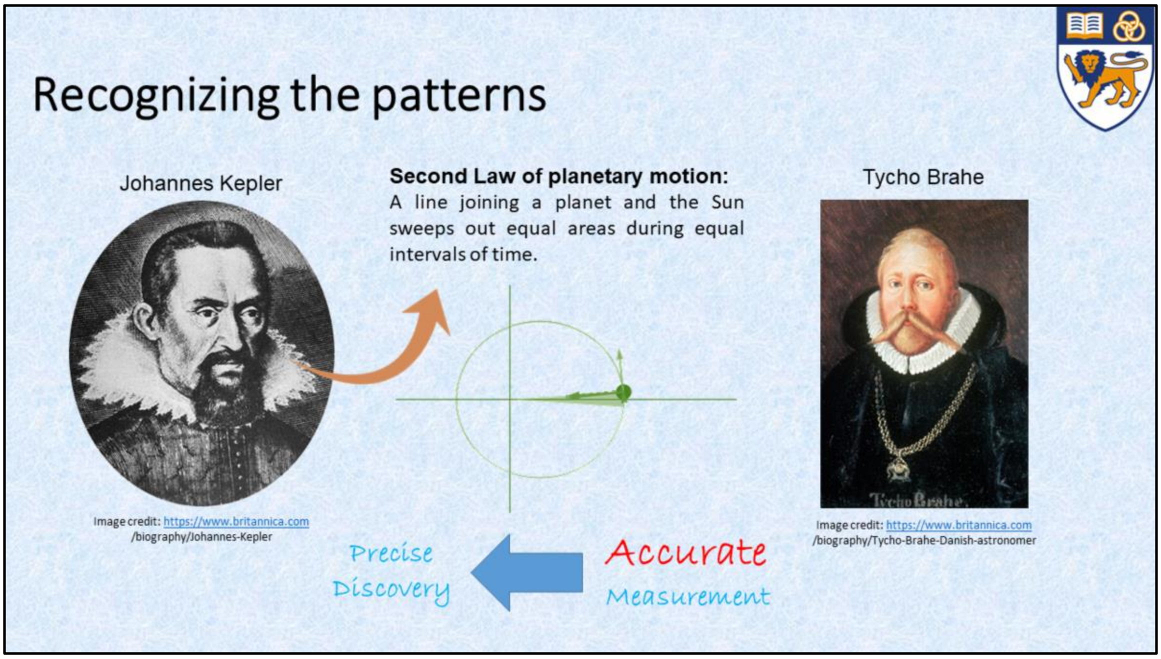
Okay! After the brief introduction into the world of variables and values, it is now a good time to talk about types of variable. Broadly, we have two types of variables. Categorical and numerical.

Categorical variables have different categories as values. For example sex is a categorical variable with two values or categories. Male and female. The same is for nationality or blood type. Sometimes, there is a natural order among the different categories. For example, educational level. In this case, we are dealing with an ordinal variable. If we cannot imagine some sort of order among the categories, then our categorical variable is nominal. For example, blood type.

Numerical variables on the other hand have numbers as values. Blood pressure, or the top speed of cars are examples of numerical variables.



Hi there!
In this unit. We will emphasize on the importance of quality of measurement and we will introduce to you the concepts of validity and reliability. At the end, we will talk about True Score Theory and how each measurement is prone to errors.



I want to start this unit with a historical example from physics. In 1609, Kepler published his first two laws about planetary motion. The second law says that: A line joining a planet and the Sun sweeps out equal areas during equal intervals of time. You may need some time to fully understand this law. Although the law is not what I want you to learn.

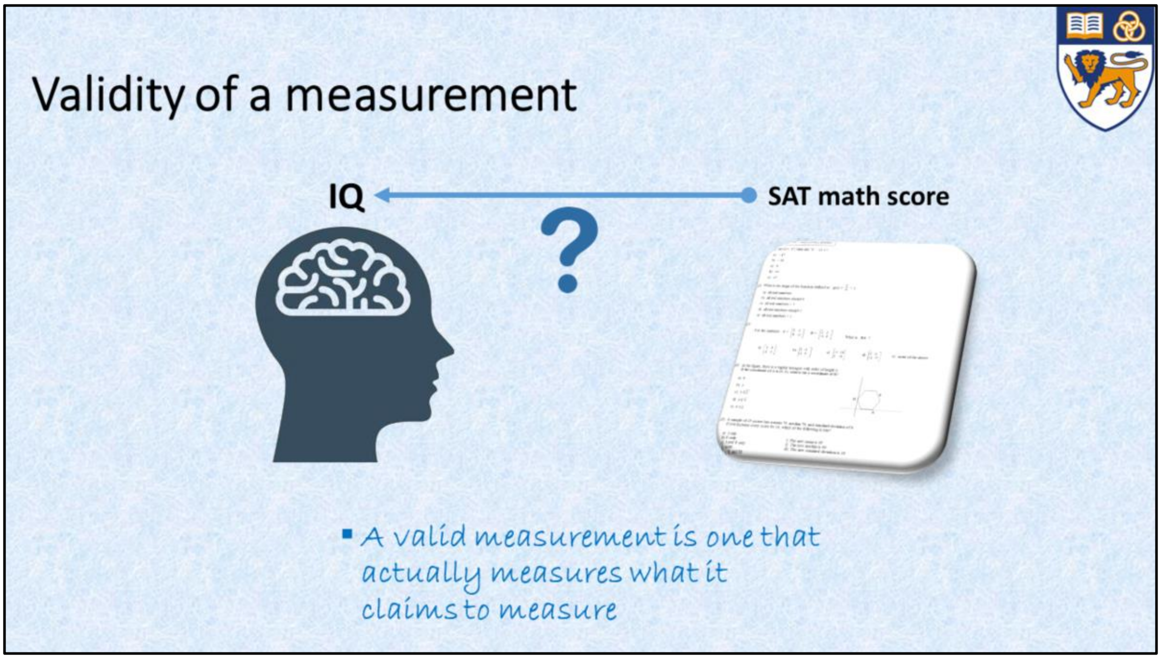
It is interesting to know that Kepler himself did not call these discoveries “laws,” as would become customary after Isaac Newton derived them from a new and quite different set of general physical principles. These were Kepler’s observations based on measurements of Danish astronomer, Tycho Brahe. If Brahe’s measurements were not accurate enough, probably Kepler could not see the pattern in planetary motion. Thus, an accurate measurement is the basis for precise discovery and that is why it is important.

https://www.britannica.com/biography/Johannes-Kepler
https://en.wikipedia.org/wiki/Kepler%27s_laws_of_planetary_motion#History
https://www.britannica.com/biography/Tycho-Brahe-Danish-astronomer

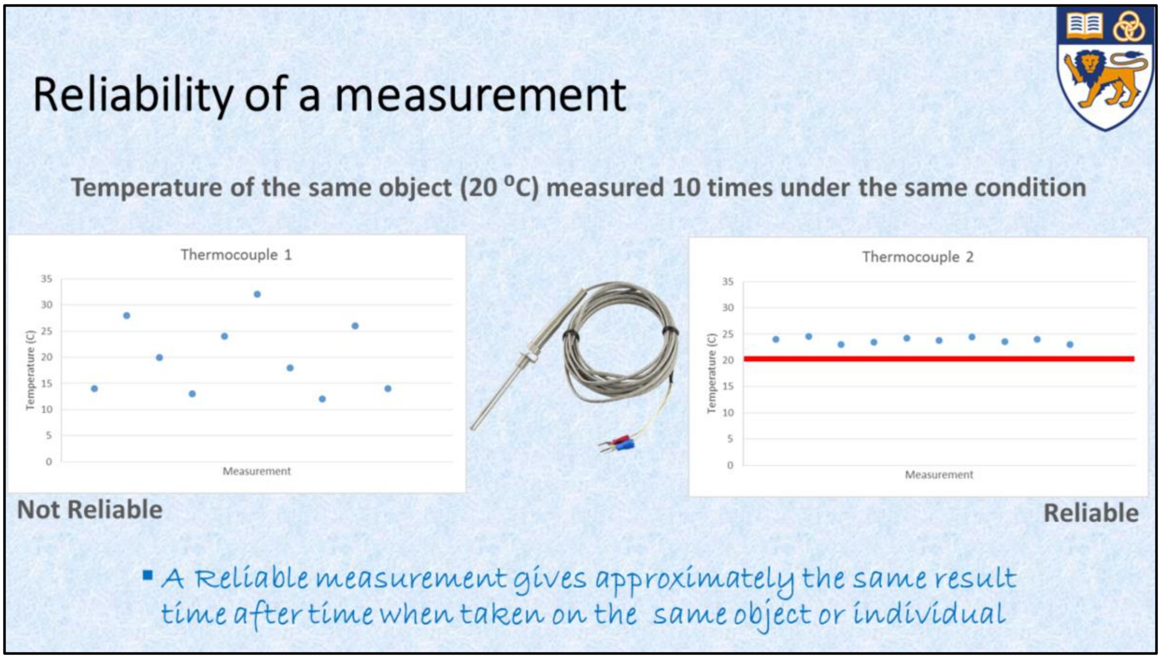
Accuracy of a measurement

- **Validity**
- **Reliability**
- **Bias**

Accuracy is a general term and it describes how close a measurement gets to the value of the variable being measured. There are three concepts that we will cover in this section:
Validity, reliability and bias. These words are commonly used in the English language, but they also have specific definitions when applied to measurement. Please be mindful of this fact.



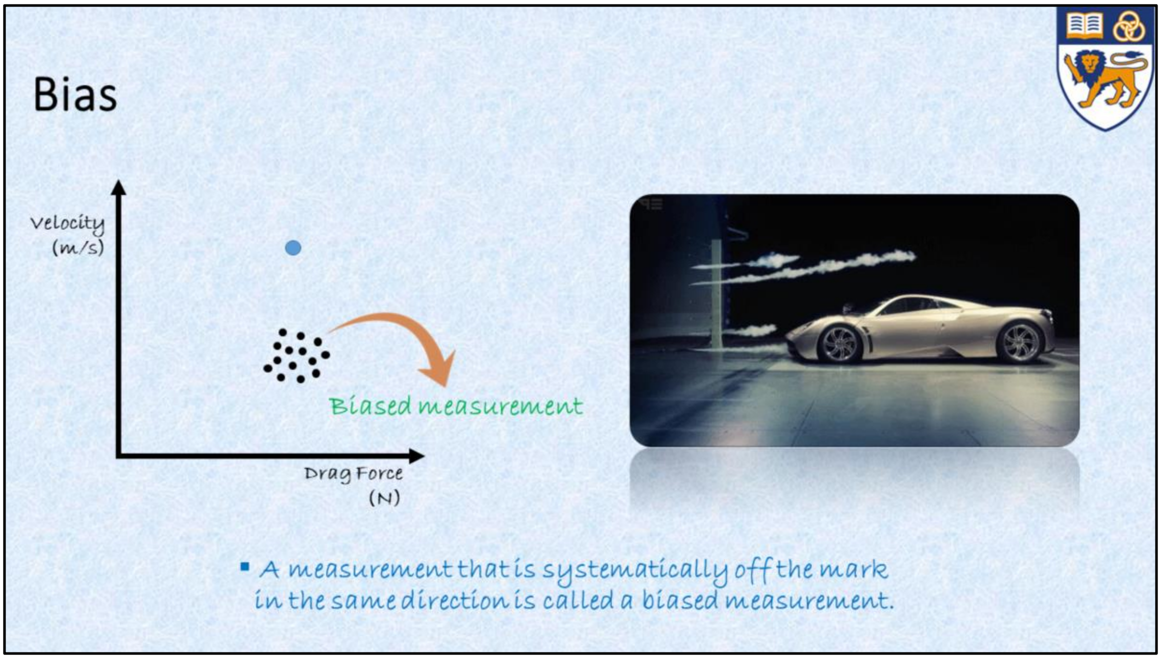
Imagine you want to measure IQ of a person. If the test you are using is actually measuring person's math ability, for example SAT math score, then this measurement is not valid. A valid measurement is one that actually measures what it claims to measure. Be mindful of measurement validity specially when facing researches in social and behavioral sciences.



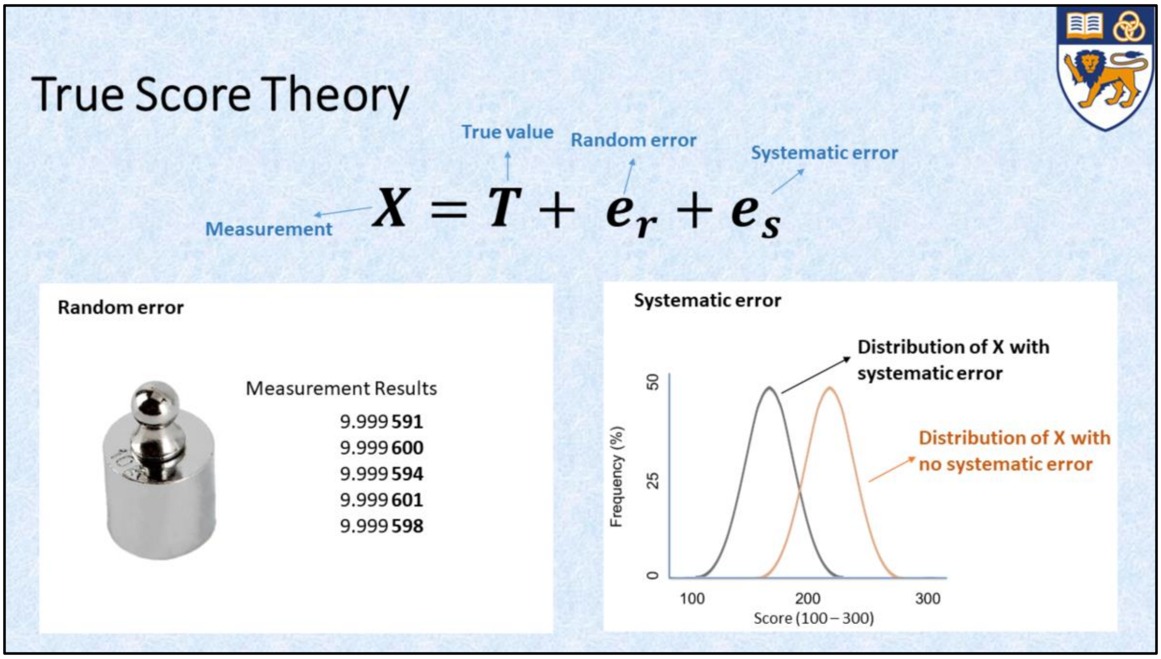
Imagine there is an object that we know is at 20 C. Using two thermocouples to measure its temperature for 10 times, the first thermocouple gives readings that are scattered and all over the place. On the other hand, the second thermocouple gives readings that are similar. The first thermocouple is hence not reliable.

Here, you may notice that the second thermocouple readings are overstated and on average different from the actual temperature. Is second thermocouple reliable? The answer is yes. This exactly is the difference between reliability in measurement terminology and its common meaning in the English language. In the common terminology used to define the accuracy of measurement, reliability means that the same result is being obtained after repeating the measurement.

Reliability is also a useful concept in psychological and aptitude testing. An IQ test is not of that much use if it shows the IQ score of a person to be 80 one time and 120 the next time.



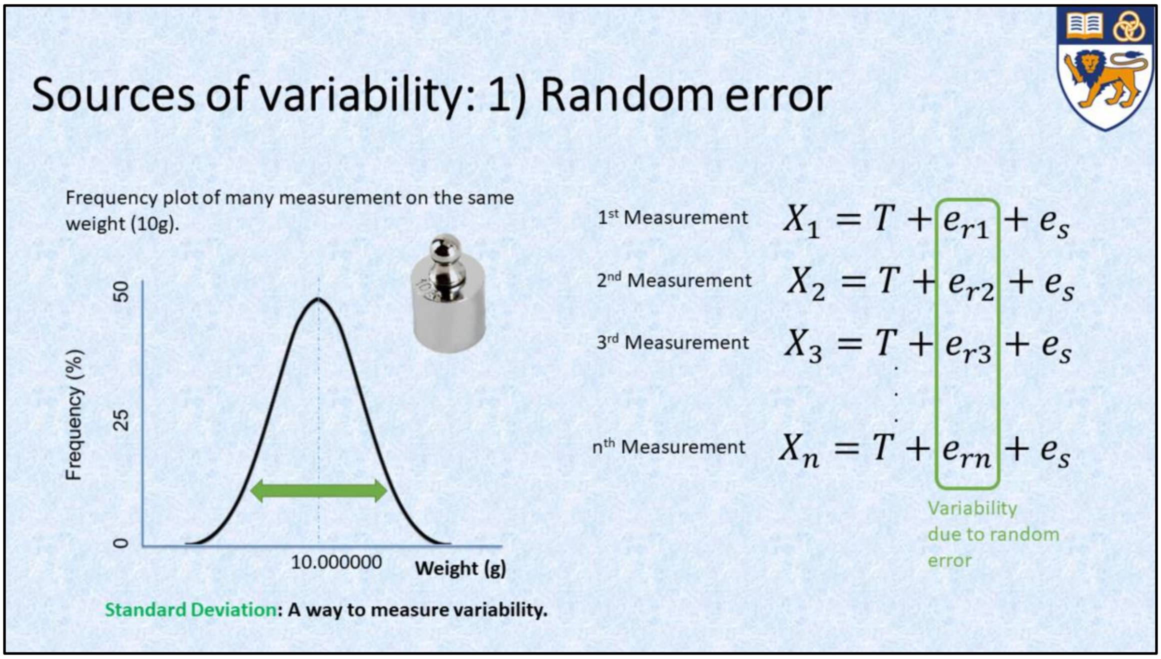
To explain the concept of Bias, let's again use an example from the engineering world. Imagine you want to measure the amount of drag force exerted on a car when it moves at a certain velocity. Assume that the true values of the velocity and drag force are indicated by the blue dot shown. After several rounds of measurement, we have a cloud of measured points like these. One explanation for this observation can be that the velocity sensor is not properly aligned, and is understating the true velocity. This measurement is biased.



I believe now is a good time to introduce to you one of the most commonly used theories of measurement. True Score Theory. Experience has shown True Score Theory to be very useful for making good measurements in a wide variety of disciplines. True score theory maintains that every measurement is an additive composite of three components. True value which we don't know it unless why we are doing measurement, random error and systematic error. True score theory is a simple yet powerful model which reminds us that most, if not all, measurements have an error component.

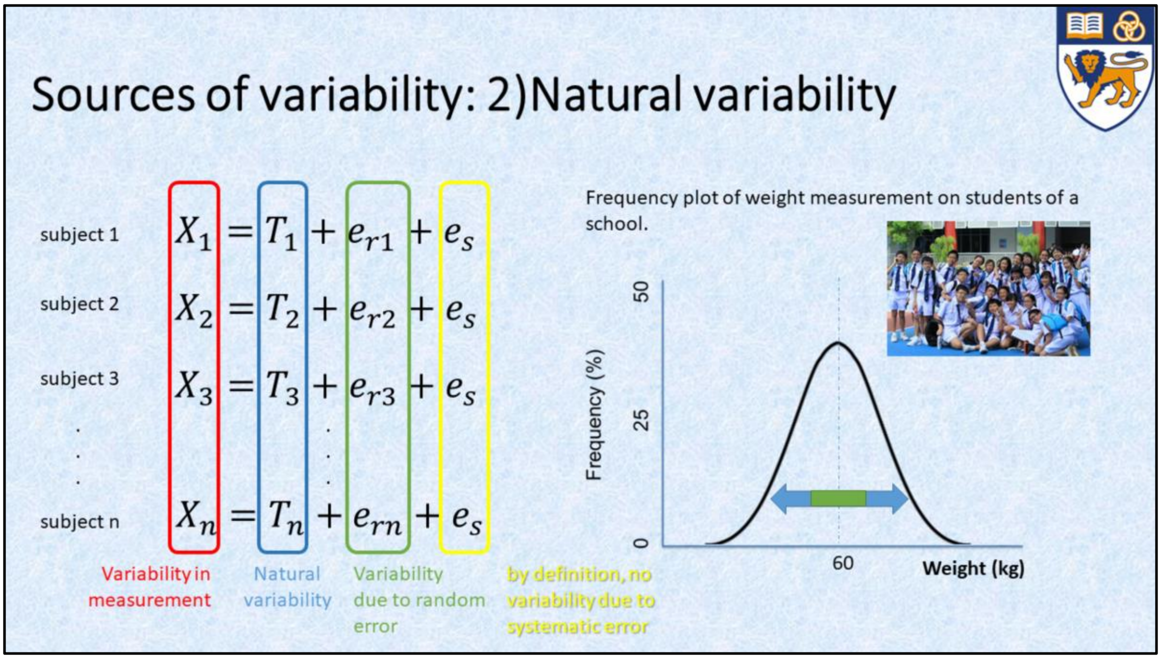
Let's go back to the theory and the equation. I want to use a simple example to explain the theory. Let's start with systematic error. Imagine I want to measure math ability of a group of students using a test. After the test, I use a frequency plot to see what percentage of students scored a certain value. Most likely, the plot would have a normal distribution. Something that all of you are already familiar with it, the bell curve. Now imagine if there is a loud traffic going by just outside of the classroom, the noise is liable to affect all of the students' score. In this case systematically lowering them. This exactly is the concept of bias we talked about earlier.

Random error is caused by any factor that randomly affects measurement. And it is likely to be cancelled out by averaging. For example, we have provided here five weightings that are done on one standard weight that is supposed to be 10 grams. Each weighting is done under the same condition, with the same procedure and by the same technicians. As you can see, the last digits are shaky, they change from measurement to measurement. This is random error at work. What is the source of it? In many cases we don't know!



Okay. Now is a good time to have a better look at the concept of variability. For the first example, let's revisit the measurement on the weight that I just mentioned in the previous slide.
Remember from the true score theory that each measurement is the sum of three components: True value, Random error and Systematic error.
Now, If we repeat the measurement on the same object, the first measurement is the true value plus the random error component and then systematic error component. Note that measurement after measurement, the true value and systematic error does not change. Now if we present the data with a histogram, it might be like the picture. The variability that we are observing here is due to measurement random error.

One way to measure the variability is to calculate the standard deviation. This concept was introduced in previous chapters. Larger standard deviation means higher variation and larger standard deviation means that the instrument is less reliable.



Now, consider we are measuring the weight of students of the same grade in a school. And hence measurement is done for a group of subjects. For subject one we have: $X_1 = T_1 + e_{r1} + e_s$
Similarly for subject 2, 3 and up to subject n.

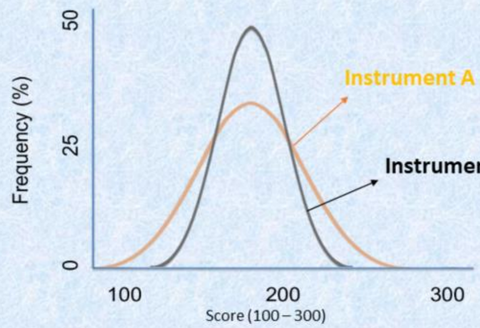
Again, if we create the histogram, it may have a shape as shown. Although some of the variability is due to our measuring instrument, random error, most of it is simply due to the fact that not everyone has the same weight. This is called natural variability.

The natural variability we talked about is the variability of the true scores which are not known to us. Random errors are also variable between measurement to measurement. Here notice that by definition, systematic error does not have any variability.

Natural variability can also happen if the variable that you are measuring depends on time. For example if we are measuring the pulse rate, it is not likely to remain constant throughout the day.

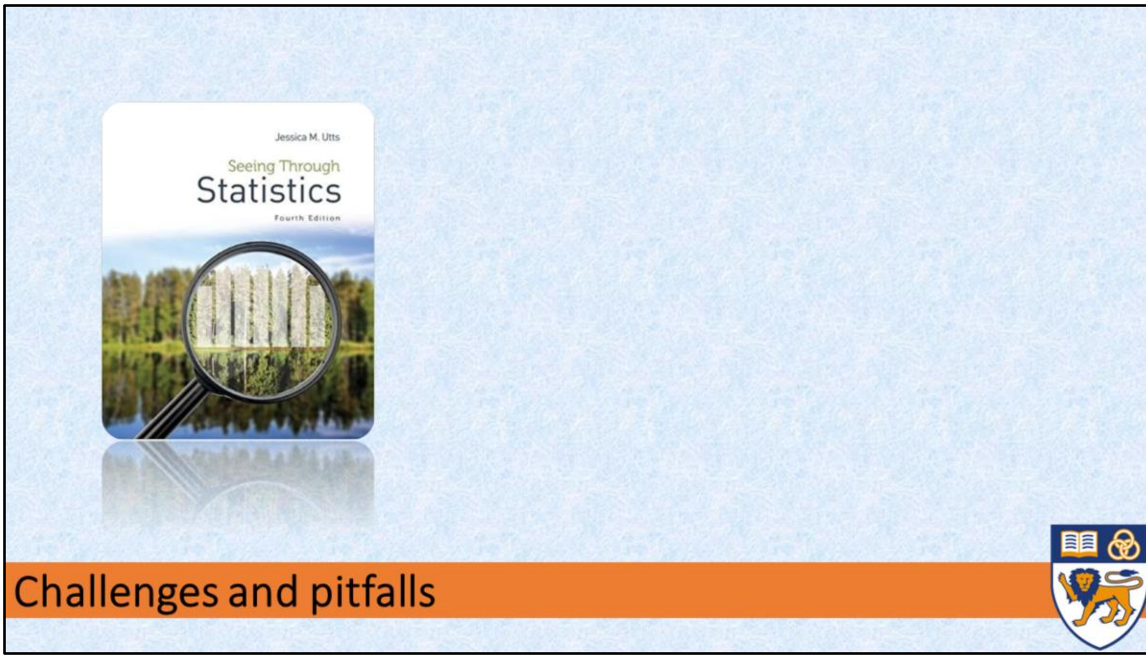
Comparing reliability of instruments

Variability in measurement = ***Natural variability*** + ***variability due to random error***



Instrument B is more reliable.

Okay. Now with the foundation laid just now, we can discuss more complex cases. Do you remember the math ability example I used during the introduction of true score theory? If I have two instruments, A and B, to measure this variable for a large number of students and measurements done by instrument B has a smaller standard deviation, then instrument B is more reliable.



Let's start the last unit of this chapter with an example adopted from the book Seeing Through Statistics by Jessica M. Utts.

Is Marijuana Easy to Buy but Hard to Get?

In a survey one of the questions asked **teens** about the relative ease of getting cigarettes, beer, and marijuana. About half of the teens were asked about “**buying**” these items and the other half about “**obtaining**” them.

“*Which is easiest for someone of your age to* ***buy****: cigarettes, beer or marijuana?*”

“*Which is easiest for someone of your age to* ***obtain****: cigarettes, beer or marijuana?*”

Response	Version with “buy”	Version with “obtain”
Cigarettes	35%	39%
Beer	18%	27%
Marijuana	34%	19%
The same	4%	5%
Don’t know/no response	9%	10%

In this example, we see how much answers to a question can change based on simple changes in wording.
In a survey, one of the questions asked **teens** about the relative ease of getting cigarettes, beer, and marijuana. About half of the teens were asked about “**buying**” these items and the other half about “**obtaining**” them. The results are interesting. It is not a bad idea to pause the video and try to make sense of the table provided. Let me help you with that. For example, 18% here means that 18% of the respondents selected beer to be the one that is easier to buy.

Comparing marijuana and beer, it is easier to buy marijuana but harder to obtain it. This observation is understandable as we know that regulations make it difficult for teenagers to buy alcohol and not to obtain it in other ways.
In what comes in the next slides, we will talk about the pitfalls that can be encountered when asking questions in a survey or experiment.

Pitfalls when asking questions …

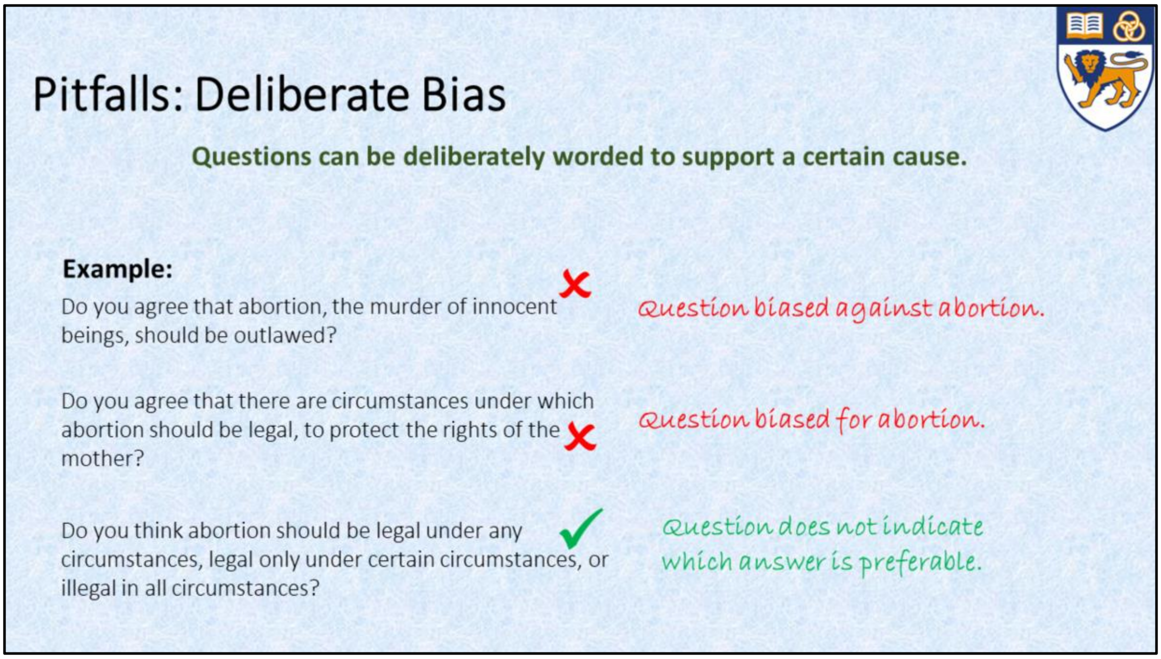
1. Deliberate bias
2. Unintentional bias
3. Desire to please
4. Asking the uninformed
5. Unnecessary complexity
6. Ordering of questions
7. Confidentiality and anonymity

These includes:

1. Deliberate bias
2. Unintentional bias
3. Desire to please
4. Asking the uninformed
5. Unnecessary complexity
6. Ordering of questions

and

7. Confidentiality and anonymity



Sometimes, if a survey is being conducted to support a certain cause, questions are deliberately worded in a biased manner. For example, both of the questions asked here about abortion are biased. I give you some time to pause the video to figure it out yourself.

Appropriate wording should not indicate a desired answer. For instance, a better way to ask the opinion about abortion is presented here. Note that this question does not indicate which answer is preferable.

Pitfalls: Unintentional Bias

Questions are worded such that the meaning is misinterpreted by many

Example:

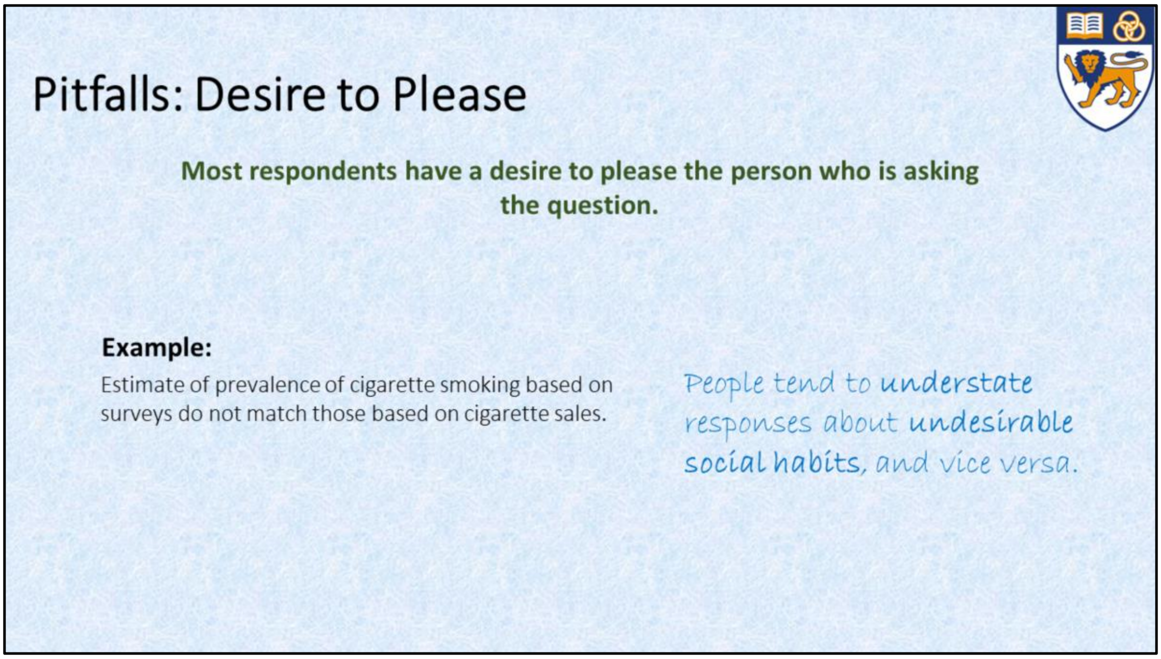
Do you use drugs? ✗

Need to specify if you mean prescription drugs, illegal drugs, etc...

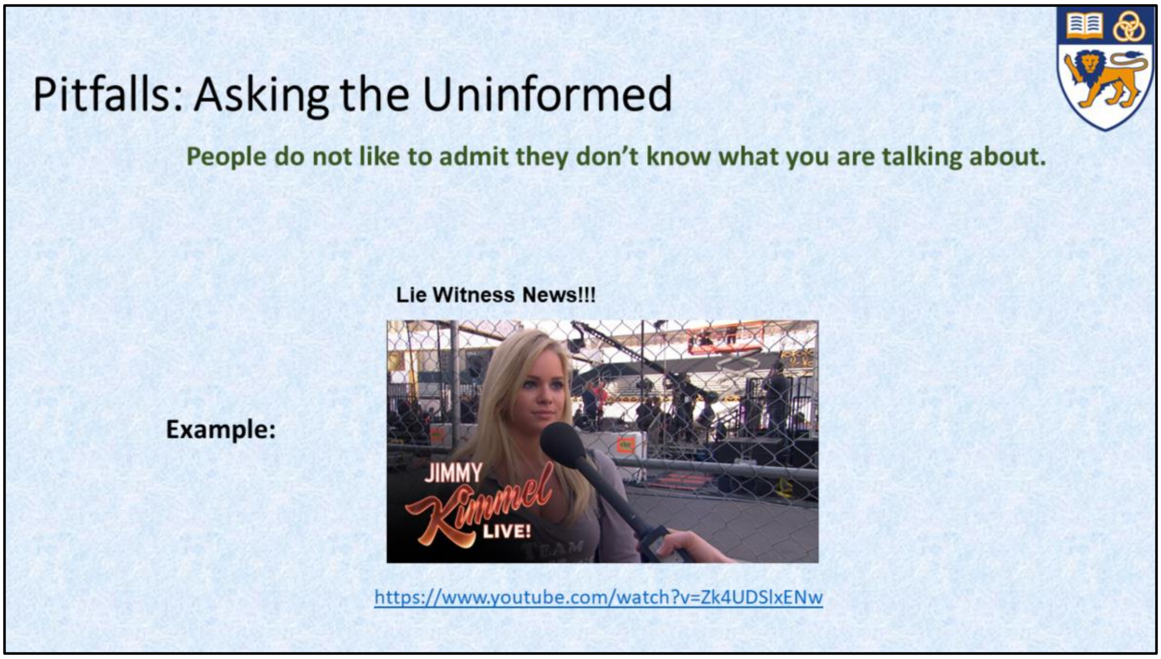
What is the most important date in your life?? ✗

Need to specify if you mean calendar date or social engagement.

Sometimes questions are worded in such a way that the meaning is misinterpreted by a large percentage of the respondents. For example, if you were to ask people whether they use drugs, you would need to specify if you mean prescription drugs, illegal drugs, over-the-counter drugs, or common substances such as caffeine.



Survey respondents tend to understate their responses about undesirable social habits and opinions and vice versa. For example, estimate of prevalence of cigarette smoking based on surveys do not match those based on cigarette sales.



People do not like to admit that they don't know what you are talking about when they are being asked a question.

The comedy host, Jimmy Kimmel, has an item in his show with the name "Lie Witness News". Almost all the time we see people who lie about an event that has not even happened. I suggest you watch some of them.

Pitfalls: Unnecessary Complexity

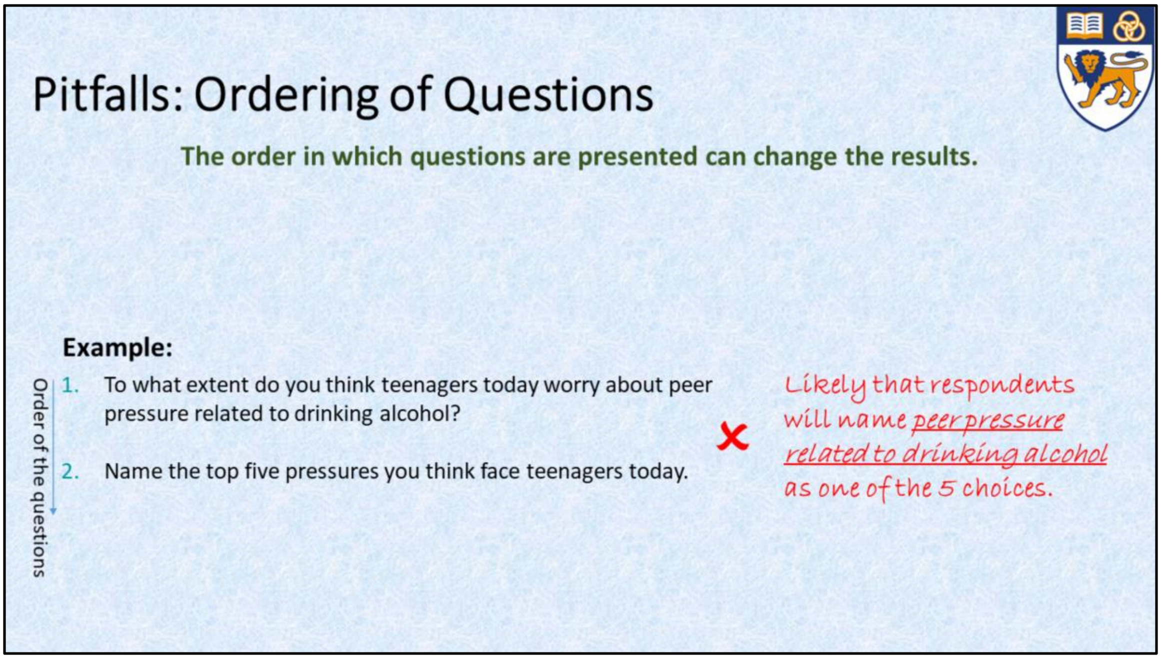
If questions are to be understood, they must be kept simple.

Example:

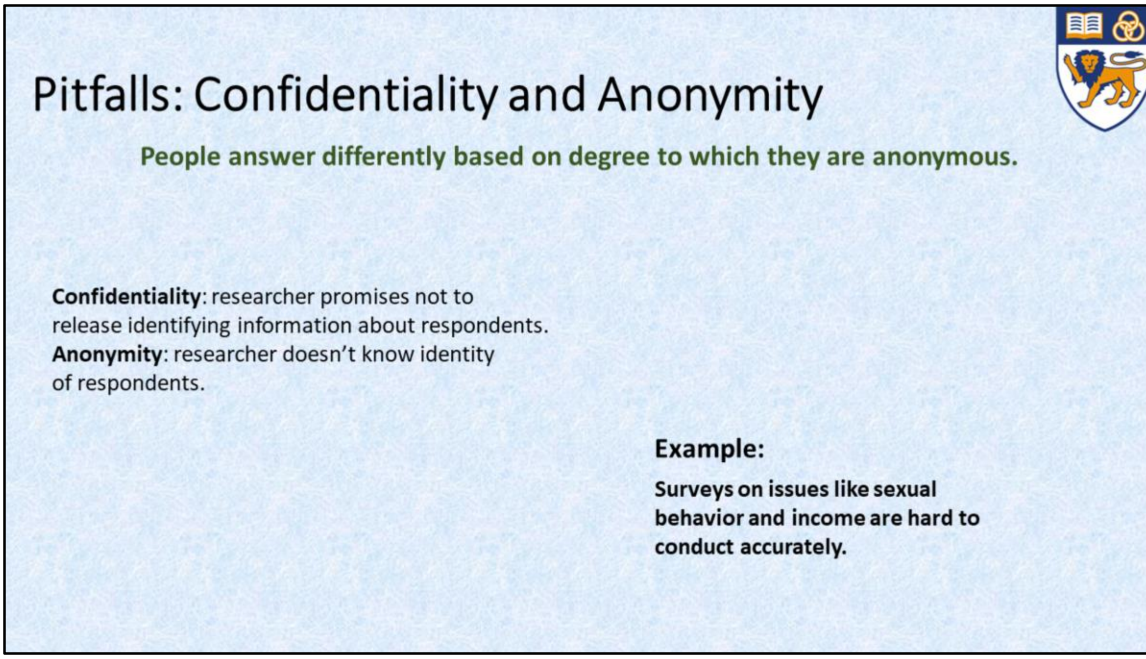
"Shouldn't former drug dealers not be allowed to work in hospitals after they are released from prison?" ✗ Too confusing

"Do you support the president's health care plan because it would ensure that all Americans receive health coverage?" ✗ Asking more than one question at once

When asking a question in a survey, it is important to avoid complex language or to ask more than one question at once. If questions are to be understood, they must be kept simple.



If a question requires respondents to think about something that they may not have otherwise considered, the order in which questions are presented can change the results. For example, if in a survey you first mention about peer pressure related to drinking alcohol and later you ask about top five pressures teenagers face, it is very likely that respondents will chose peer pressure related to drinking as one.



Answers to questions may be different based on the degree people believe they are anonymous. That is why surveys on issues like sexual behavior and income level are hard to conduct accurately.

End