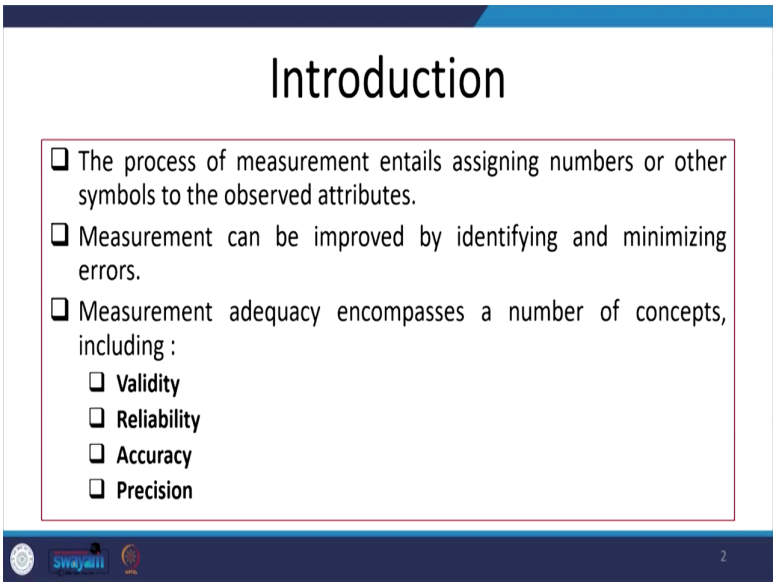
Exploring Survey Data on Health Care
Prof. Pratap C. Mohanty
Department of Humanities and Social Sciences
Indian Institute of Technology, Roorkee

Lecture - 15
Validity, Reliability, Accuracy and Precision of Sample

Welcome friends to this NPTEL MOOC module on Handling Healthcare Survey Data. We are on the 3rd week of our sessions. After explaining the very foundations of sample data and also how to go for handling with some sampling data, even we discuss missing values in the previous lectures. So, now we should explain the validity, reliability and accuracy, and precision of the sample data. The pre-testing is required before doing our final analysis.

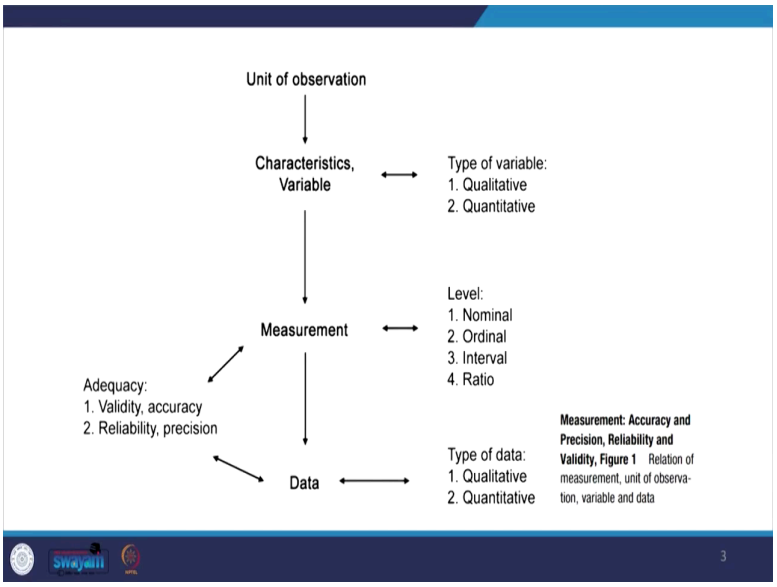
There are some aspects that I am going to emphasize. All those 4 components I will clarify in this lecture. So, coming to the validity or even reliability, accuracy, precision, this is actually the process of measurement that entails assigning numbers or other symbols to the observed attributes.

(Refer Slide Time: 01:51)



The measurement can be improved by identifying and minimizing errors. Every time I have been emphasizing the fact of how we could be able to minimize the errors in the sample data. So, we are dealing with 4 important measurement adequacy which is quite relevant in the sample data to check before final estimation.

(Refer Slide Time: 02:32)

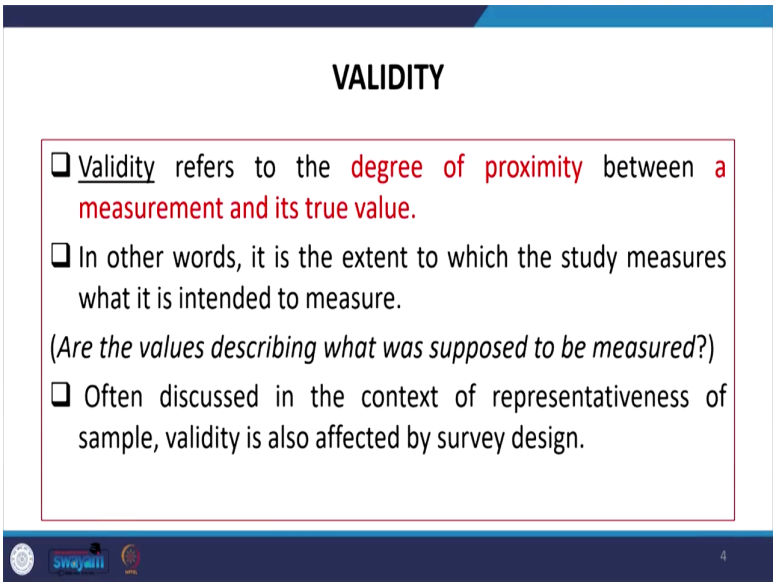


So, the first one is called validity, the second one is reliability, and the third one is accuracy and precision. We start with the unit of observation, then we will define their characteristics or their variables, those variables could be qualitative or quantitative. Once we understand the very basic aspects of the variables and characteristics then we move to the measurement.

In the measurement, we should have checked all those 4 adequacies, that is validity, accuracy or reliability, and precision. These are applied to all types of data sets may be nominal, ordinal or interval, or ratio. So, finally, after doing all those fundamental checks, we should finally get our data and those data could be again of qualitative and quantitative. So, that means, these are the adequate instruments, we do apply whether it is qualitative data or quantitative data.

So, coming to the first inadequacy check is called validity, which refers to the degree of proximity between measurement and its true value.

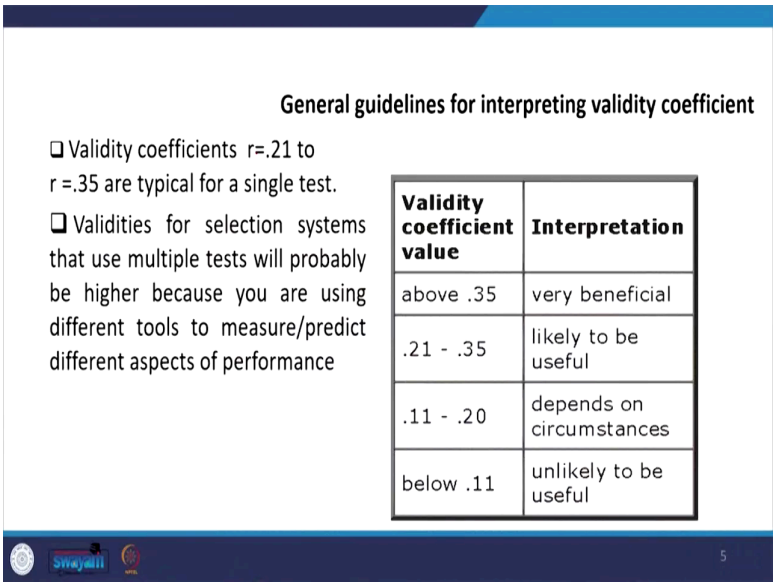
(Refer Slide Time: 03:58)



So, we do hypothesis testing. We check with the true value and its estimated value. Then we check how our sample information or sample estimates could be representing the true value.

The validity refers to the degree of proximity. This also explains the extent to which the study measures, what it is intended to measure, or the values describing what was supposed to be measured. Often discussed in the context of representativeness of the sample, validity is also reflected by survey design. If your survey design is quite apt, quite right, then there are high chances of getting validity to your true value or true representative value.

(Refer Slide Time: 05:23)



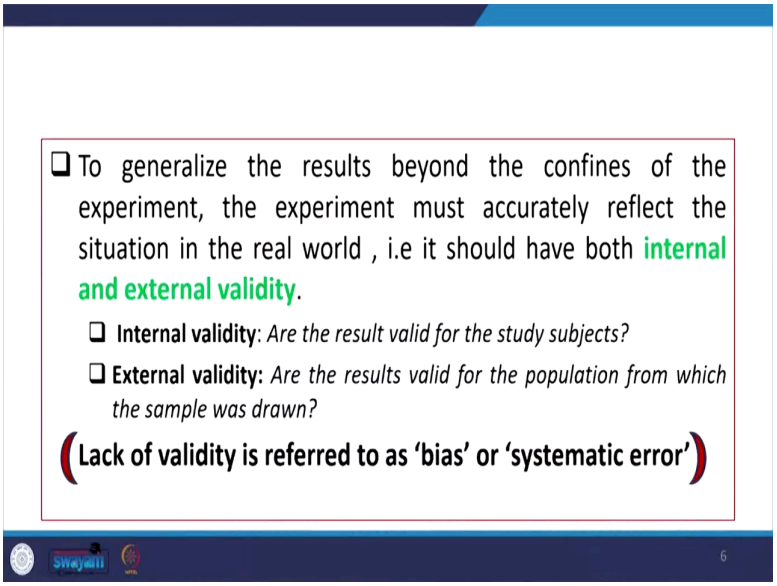
Now, there are certain guidelines for interpreting the validity coefficient. So, the validity coefficients are like r is equal to 0.21. So, as per our example data set, there are indicators we have mentioned like below 0.11, 0.11 to 0.20, then 0.21 to 0.35, and above 0.35.

If it is the value of the validity coefficient is below 0.11 that means your sample data is unlikely to be useful for estimating your population characteristics. Then, if it is above 0.35 that is very beneficial. In between, there is a certain likeliness of its uses.

So, validity for selection systems that use multiple tests will probably be higher because you are using different tools to measure or predict different aspects of performance.

To generalize the results beyond the confines of the experiment, the experiment must accurately reflect the situation in the real world. It should have both internal as well as external validity. So, then we need to understand these two concepts first what do you mean by internal validity and what do you mean by external validity.

(Refer Slide Time: 06:49)

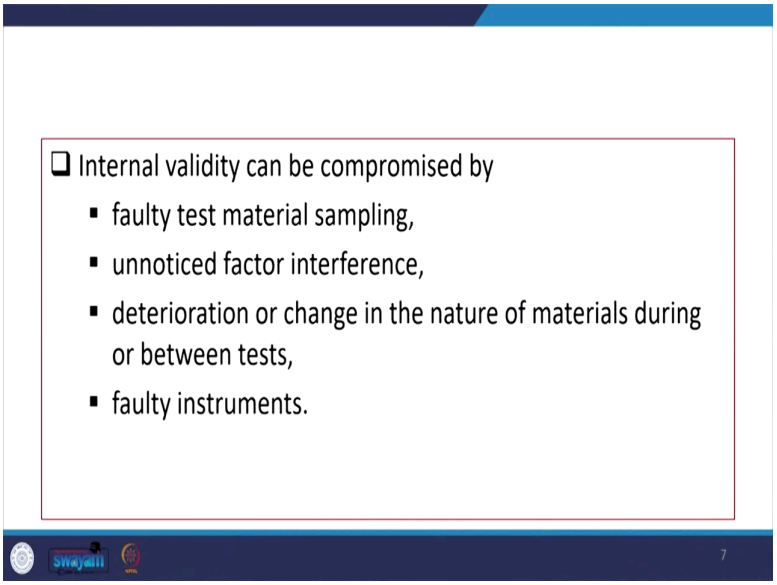


Regarding internal validity, the first question we ask is the result valid for the study itself or for study subjects. External validity: obviously, it is beyond the study subject and that is like are the results valid for the population from which the sample was drawn.

So, you need to check whether whatever we have derived is actually correct within the sample frame or not, then we check with the population. When the population characteristics are being compared; means, we are actually looking at the external validity of the sample.

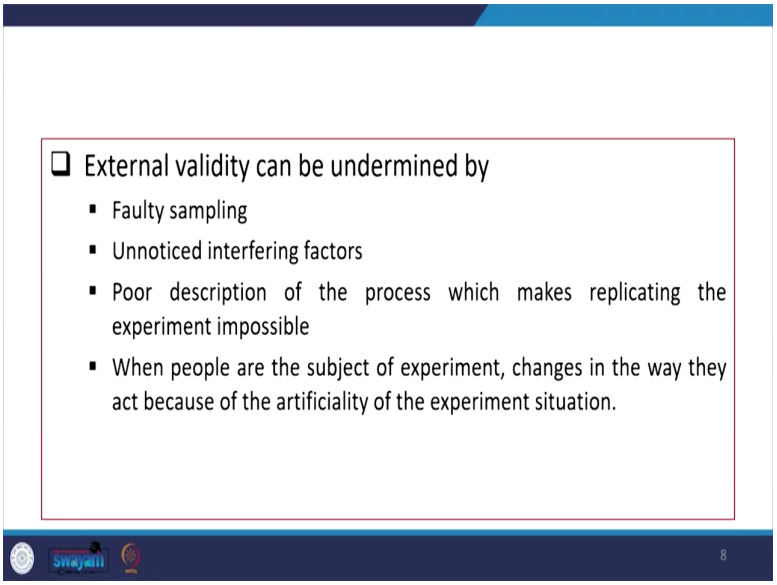
Now, lack of validity is also referred to as bias or systematic error. You might have heard about the term often used by the reviewers.

(Refer Slide Time: 08:07)



Internal validity can be compromised by faulty test material sampling, unnoticed factoring interference, deterioration or change in the nature of materials during or between tests, and faulty instruments. So, out of any of the indicators, there might be more other factors as well, but these actually refer to internal validity. Then we discussed external validity that can be understood by the faulty in its sampling frame or sampling design.

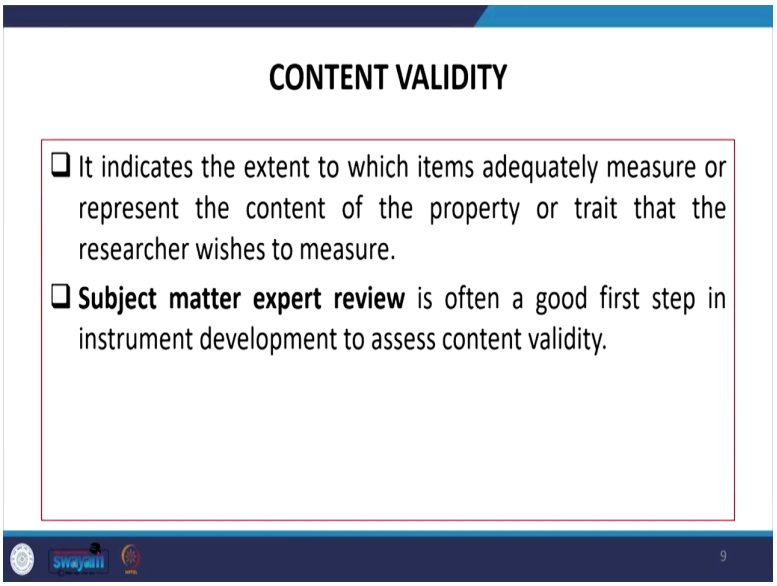
(Refer Slide Time: 08:33)



Then unnoticed interfering factors and poor description of the process make replicating the experiment impossible. When people are the subject of an experiment, changes in the way they act because of the artificiality of the experiment situation.

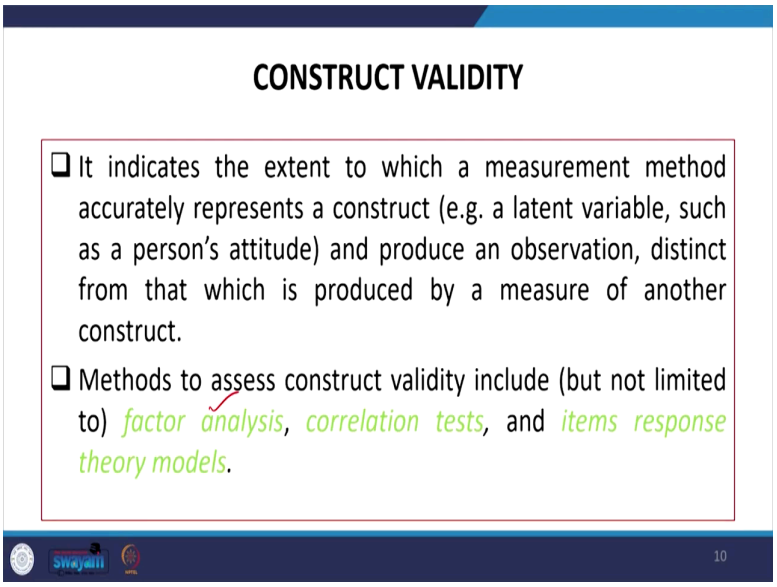
The next one is on content validity regarding the content. This indicates the extent to which items adequately measure or represent the content of the property or trait that the researcher wishes to measure.

(Refer Slide Time: 09:18)



The second important aspect of content validity is that the subject matter expert review is often a very good first step in instrument development to assess content validity.

(Refer Slide Time: 09:57)

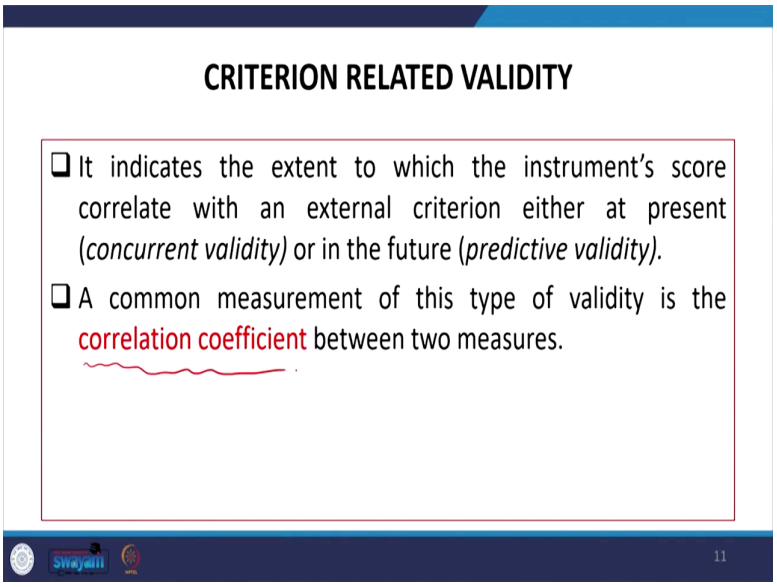


The third one is called construct validity. Construct validity here refers to the extent to which a measurement method accurately represents a construct. The construct would be your latent variable, such as a person's attitude, and produce an observation, distinct from that which is produced by a measure of another construct.

I know construct could be your latent variable. Latent variables are usually referred to as such variables consider being the proxy or sometimes refer to as the qualitative variable as well, but most often it is referred to as a proxy. So, likewise, we have cited here an example of a person's attitude. So, it is very difficult to define or quantify, so it has qualitative characteristics.

So, then methods to assess construct validity include factor analysis, correlation test, and items response theory models. So, these 3 are mostly used, so the factor analysis we will also discuss sometimes within the you know for view of the SPSS. We may not do it in detail, but the overview of this will be discussed because you know this one I have covered in my previous module. So, you can follow it.

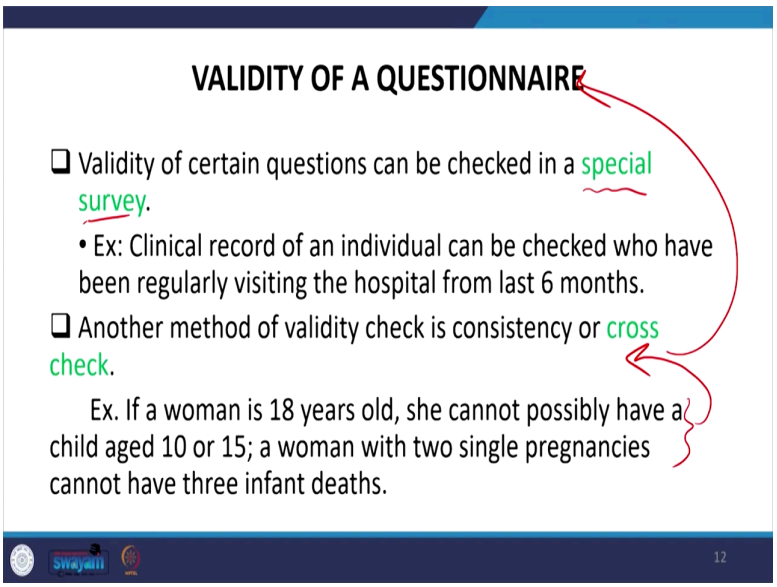
(Refer Slide Time: 11:31)



The next aspect of discussion is criterion related validity. What are the criteria we usually discuss? This indicates the extent to which the instrument's score correlate with an external criterion either at present or in the future. In the present, this is called concurrent validity, and, in the future, it is called predictive validity.

A common measurement of this type of validity is the correlation coefficient between two measures. So, you can take note of it that this is the common measurement related to validity.

(Refer Slide Time: 12:16)

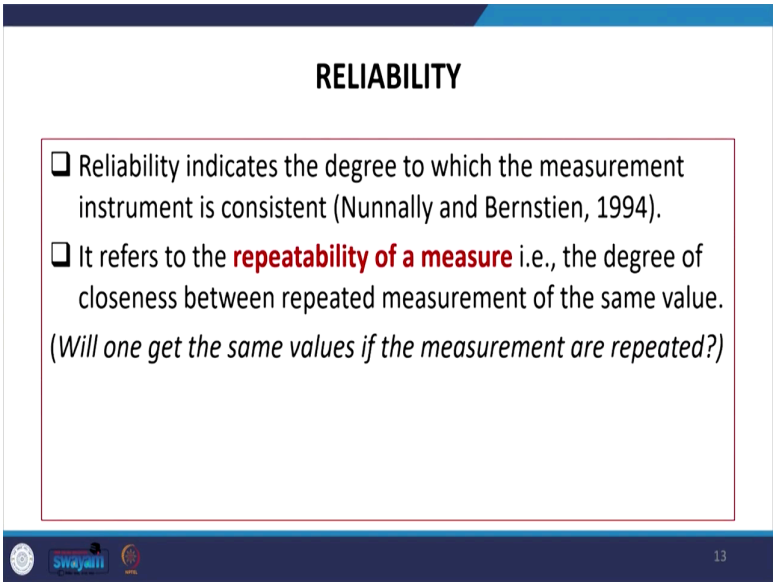


So, next comes the validity of a questionnaire. The validity of certain questions can be checked in a special survey like a clinical record of an individual can be checked who has been regularly visiting the hospital for the last 6 months or a special survey.

Then another method of validity check is consistent consistency or cross-check. Like if a woman is 18 years old she cannot possibly have a child as a 10 or 15; A woman with two single pregnancies cannot have 3 infant deaths. So, these inconsistencies should also be checked initially. Those kinds of the check are called cross-check and corrects your faults with the questionnaire.

Then, the first aspect we said related to a questionnaire is that you need to have a special check or special survey. A special survey that has to be directly heating to the requirement of the study like a clinical record of an individual can be checked who has been regularly visiting the hospital for the last 6 months.

(Refer Slide Time: 13:42)



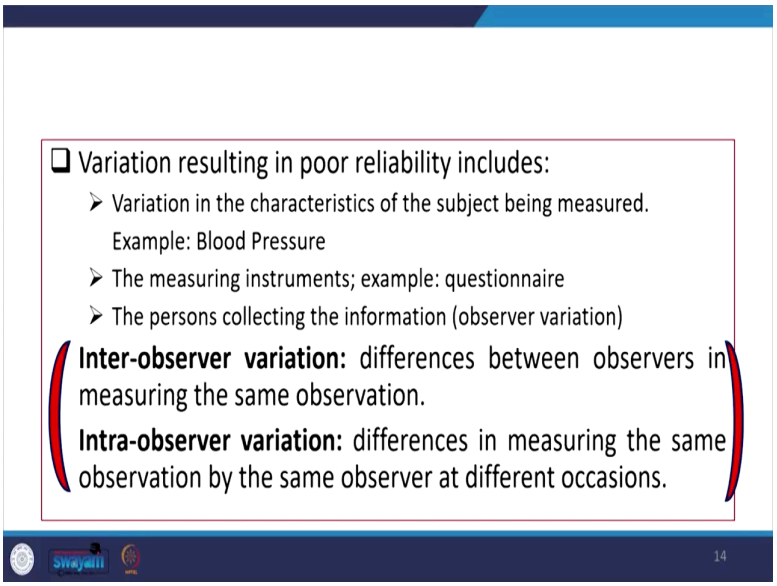
Now, we are discussing reliability after the discussion of validity. Reliability is something where we are related to the understanding, relate this understanding to the consistency of a result. So, reliability indicates the degree to which the measurement instrument is consistent. This also refers to the repeatability of the measure.

You can take some subsamples and then estimate from the survey data to cross-check whether the result is still valid or not. Reliability is also useful in the discussion of robustness

as well because this is repeatedly estimated and can give consistency to your result. This explains the degree of closeness between repeated measurements of the same value.

Will one get the same values if the measurements are repeated? It is not necessarily that someone is going to get the same value. If the values are not very different, not so diverse then your result is in fact very consistent. If it is having huge dispersion then you need to cross-check your questionnaire or sample once again.

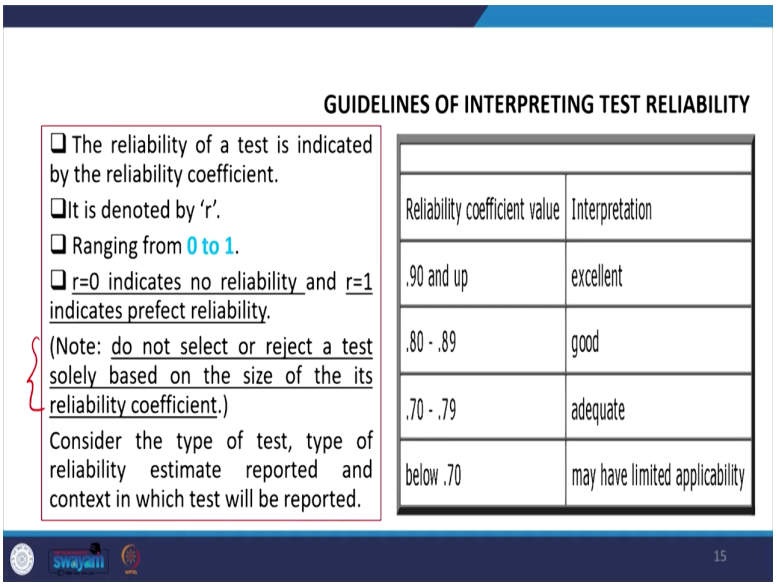
(Refer Slide Time: 15:37)



Variation results in poor reliability this includes variation in the characteristics of the subject being measured like blood pressure. So, there are variations in the characteristics of the subjects. So, blood pressure is one such example. The second one is called a measuring instrument like a questionnaire. Then, the third one is called persons collecting the information or the object observer variation. So, observer variation is also important in understanding the poor reliability of your sample.

There are two aspects we need to discuss; one is called inter-observer variation, and another is called intra-observable variation. Inter-observer variations differentiate between observers in measuring the same observation. Intra-observer variation differences in measuring the same observation by the same observer on different occasions.

(Refer Slide Time: 16:41)



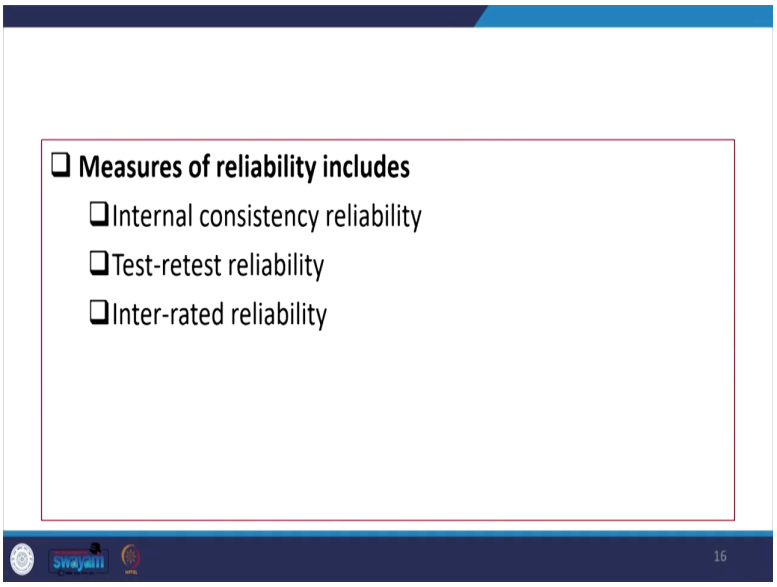
Here we are defining some other aspects of reliability. We find the reliability coefficient values and how to interpret them once you have got your reliability values.

If it is 0.9 and up that means your sample is excellent. If it is 0.8 to 0.89 that is good and if it is below 0.7 then you may have limited applicability. So, the reliability of a test is indicated by the reliability of the coefficient and denoted by r. This ranges from 0 to 1, 0 stands for no reliability, and 1 stands for the perfect reliability.

Now, we need to note one thing, do not select or reject a test solely based on the size of its reliability coefficient. Just based on the size of the coefficient, we should not take a decision in either selecting or rejecting a test.

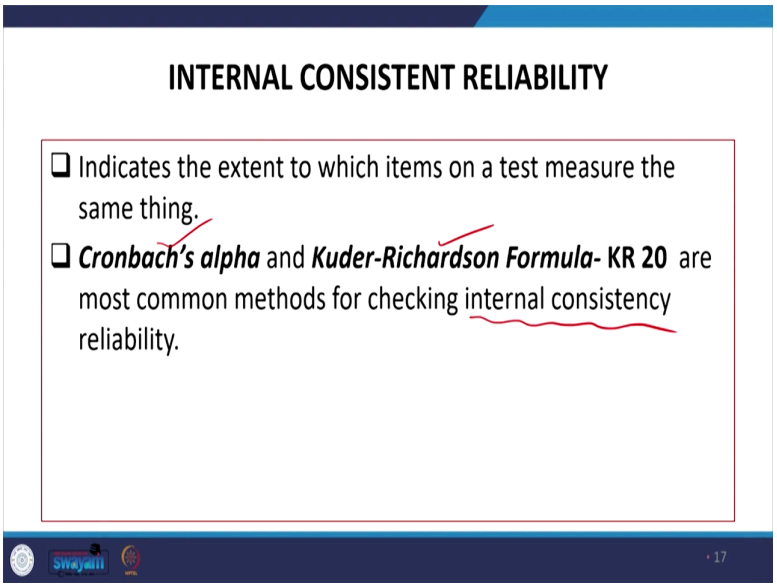
Then, consider the type of test, type of reliability estimate reported, and the context in which test will be reported. We would also take care of it.

(Refer Slide Time: 18:34)



The next set of discussions are on measures of reliability, this includes internal consistency reliability, test-retest reliability, and inter-rated reliability. These 3 measures are essential. Internal consistent reliability indicates the extent to which items on a test measure the same thing.

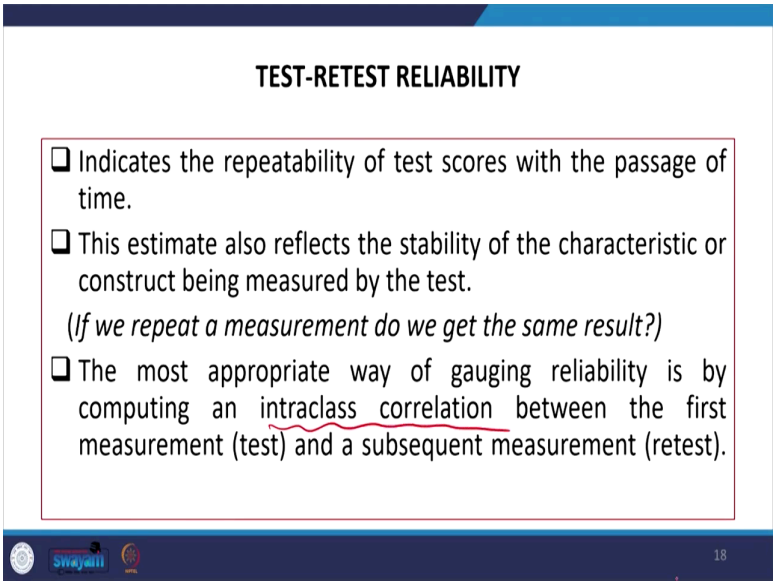
(Refer Slide Time: 18:50)



In the case of internal consistency reliability, we use Cronbach's alpha and Kuder-Richardson's formula. So, Cronbach's alpha and Kuder-Richardson Formula-KR 20

are the most common methods for checking internal consistency reliability. So, we usually apply these two for internal consistency reliability.

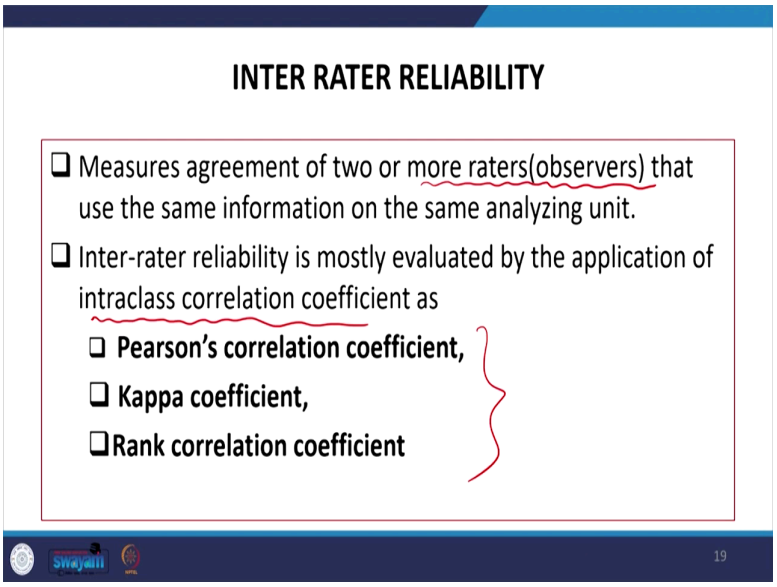
(Refer Slide Time: 19:45)



Then, the second one is called test-retest reliability. This indicates the repeatability of test scores with the passage of time. With the time dimension, we check the test again. This estimate also reflects the stability of the characteristic or construct being measured by the test.

If we repeat a measurement, do we get the same result again? we need to cross-check. The most appropriate way of gauging reliability is by computing in the intra-class correlation between the first measurement and a subsequent measurement or retest. So, intra-class correlation is in fact, essential in this case.

(Refer Slide Time: 20:52)

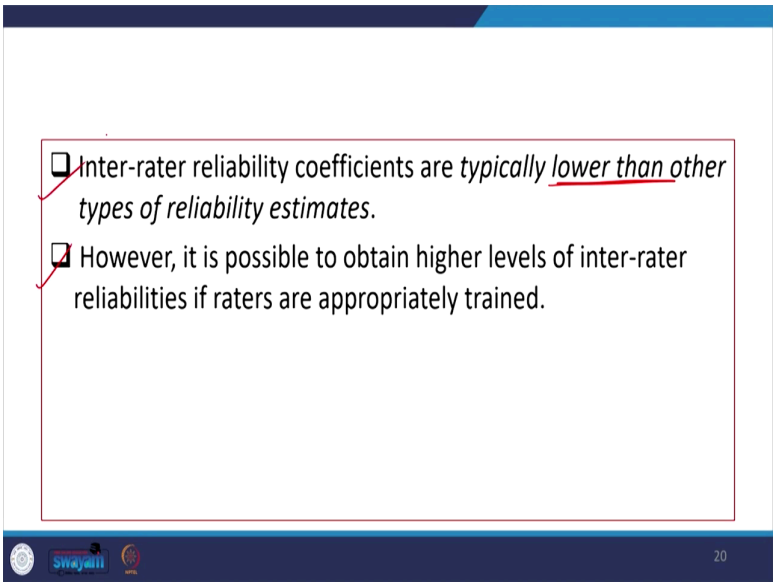


The next one is called inter-rater reliability,this measures agreement of two or more raters that use the same information on the same analyzing unit. So, these raters will actually give information repeatedly. So, they are called inter-rater reliability.

Inter-rater reliability is mostly evaluated by the applications of an intra-class correlation coefficient. So, the correlation coefficient within the class is in fact useful. These co-coefficients are the most important one is called the Pearson correlation coefficient, the second one is called the Kappa coefficient, and the third one is called the Rank correlation coefficient.

And these are quite useful in the case of inter-rater reliability or intra-class correlation coefficients. So, these 3 are important I have already mentioned. I suggest that you should prepare accordingly.

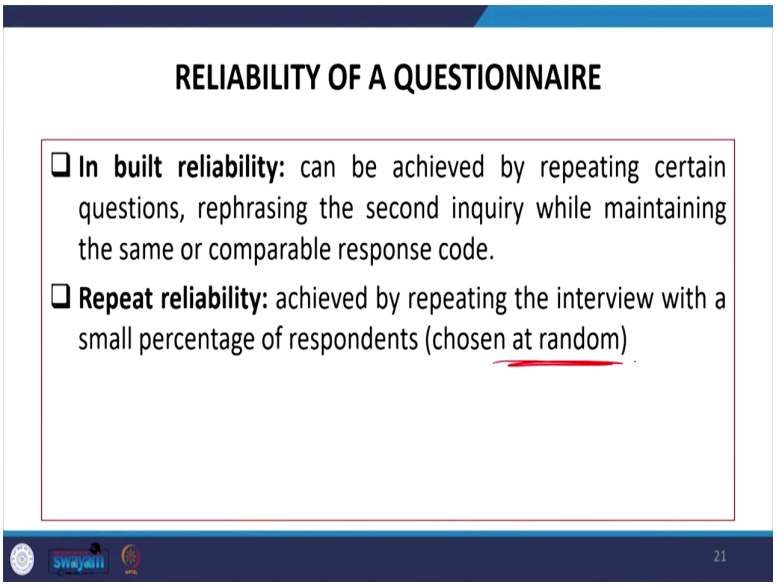
(Refer Slide Time: 22:13)



The inter-rater reliability coefficients are typically lower than other types of liability estimates. So, this is in fact to be noted again, these reliability coefficients are usually lower than the other types of reliability estimates. However, it is possible to obtain higher levels of inter-rater reliability if raters are appropriately trained. If raters are well trained, then there is no problem and they may get higher levels of reliability as well.

So, the next is the reliability of a questionnaire. First, we tested the validity through the questionnaire, and now we are discussing the reliability of a questionnaire. One is called inbuilt reliability; the second one is called repeat reliability.

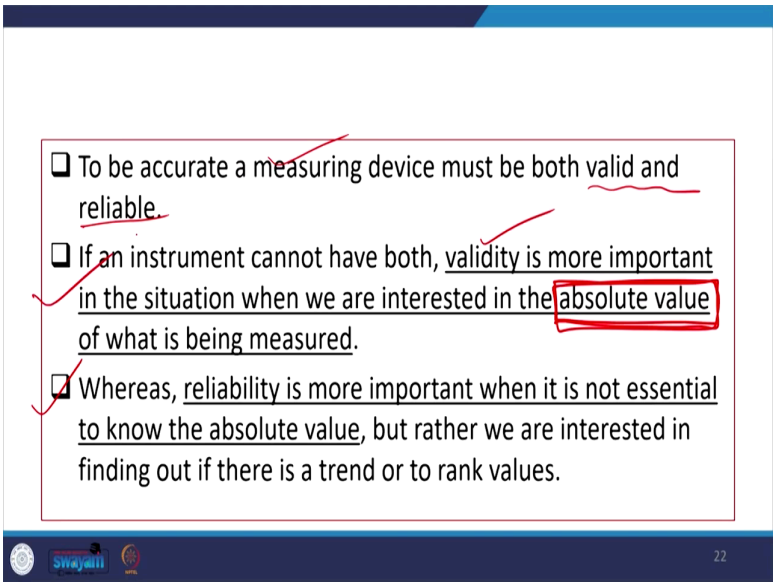
(Refer Slide Time: 23:29)



Inbuilt reliability can be achieved by repeating specific questions and rephrasing the second inquiry while maintaining the same or comparable response code. Whereas the response codes are the same, we have rephrased the second enquiry to test whether it is giving the same result or not.

Then, in the repeat reliability, what is that all about in the questionnaire? Its repeated reliability was achieved by repeating the interview with a small percentage of respondents. So, those are in fact, chosen at random. So, some sample percentages of respondents are chosen and based on that the reliability test is derived. So, these two are in fact, very useful and this has to be followed by at random method.

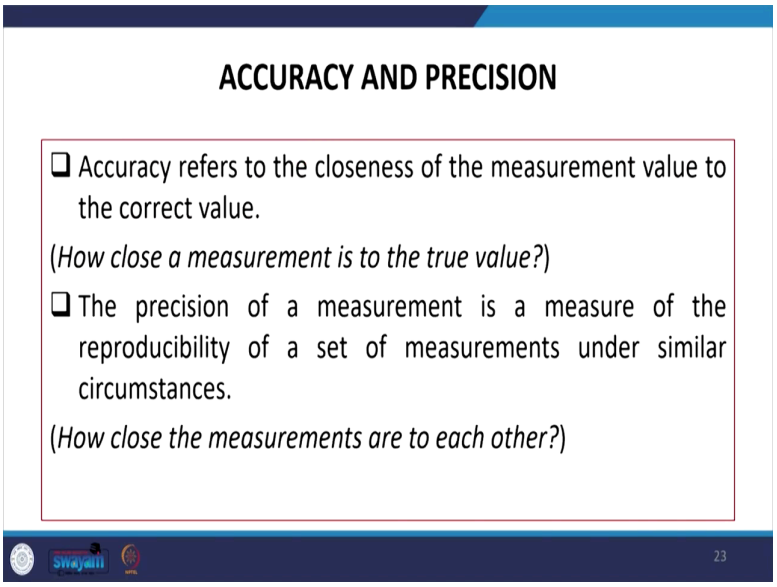
(Refer Slide Time: 24:25)



Now, to be accurate a measuring device must be both valid and reliable. So, this must be valid as well as reliable. If an instrument cannot have both, validity is more important in the situation when we are interested in the absolute value of what is being measured. This is also quite important. When we are comparing the absolute value then I think the validity method is more important.

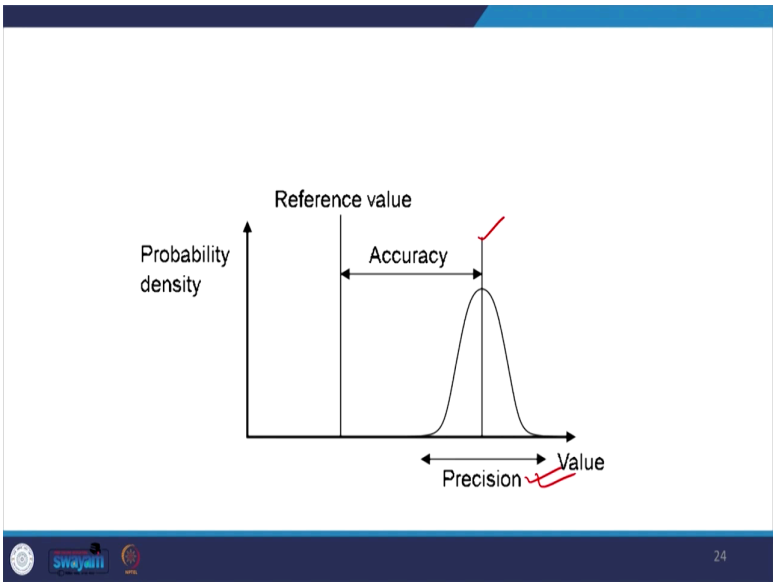
Whereas reliability is more important when it is not essential to know the absolute value, but rather we are interested in finding out if there is a trend or to rank values. If that is the case, then in that case reliability is most appropriate. So, these two points are also very important.

(Refer Slide Time: 25:42)



Regarding accuracy and precision, accuracy refers to the closeness of the measurement values to the correct value. How close a measurement is to the true value is usually asked as the question. The precision of a measurement is a measure of the reproducibility of a set of measurements under similar circumstances. So, now, the question you generally ask yourself that how close the measurement is to each other.

(Refer Slide Time: 26:18)

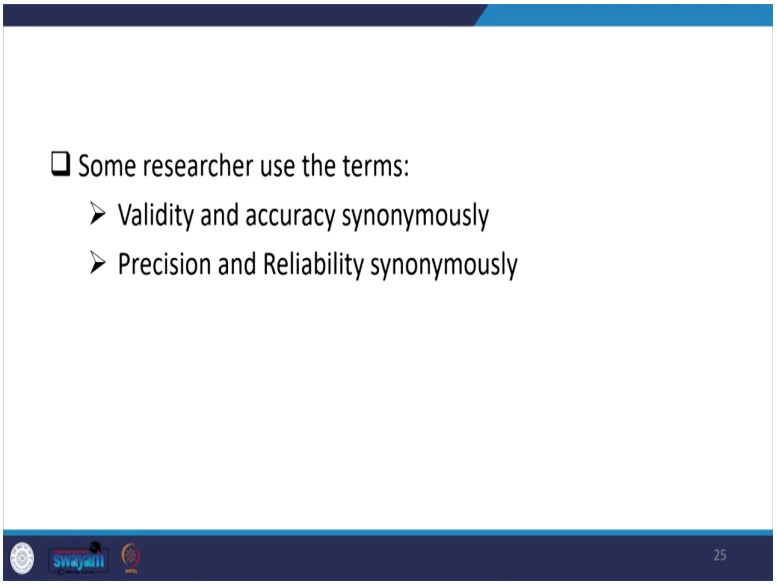


In the diagram, you can just have a check this is guidance for you on the vertical axis. We have probability density whereas the values are given on the horizontal axis. Now, how

precise your result depending upon the distribution of your data. If it is quite skewed, then there is no question of precision. Again, if it is platykurtic in the data then the alpha region is going to be very high. So, the precision is expected to be lesser.

But in the case of less sparse from the central tendency then the alpha region is expected to be lesser, in that case, you are expecting a better precision of your result. So, these give the mean value of the distribution if we are reaching around the mean value, which means we are actually defined with the accuracy of the data.

(Refer Slide Time: 27:28)



Some researchers use validity and accuracy synonymously, and precision and reliability also are interchangeably used. So, these are all the details for this class.

We have one lecture dedicated to SPSS. We will be guiding about to run spaces on your screen and some basic steps will guide you.

These are all for today and we will continue the lecture the next week. So, let me stop here.

Thank you.