Engineering Psychology

Prof. Naveen Kashyap

Department of Humanities and Social Sciences

Indian Institute of Technology, Guwahati

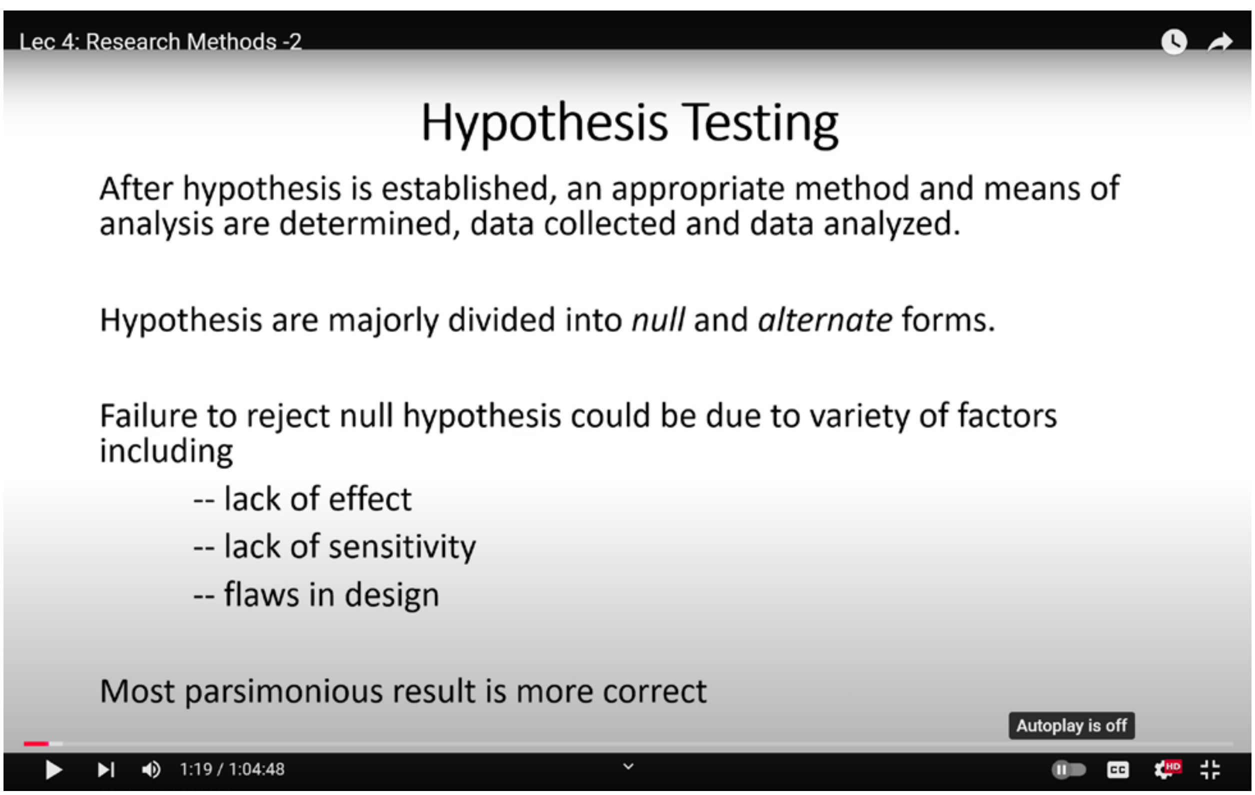
Week-02

Lecture-04

Research Methods -2

Namaskar, viewers. I welcome you to this fourth lecture in the series on engineering psychology. In the previous lecture, we focused on understanding research methods in engineering psychology. We discussed the scientific method, explained its principles, and explored the fundamentals of how research is conducted. We also examined how to identify problems for research, and once a problem is identified, how we explore the literature. Based on existing theories, we propose a hypothesis that can be tested. We looked at different types of hypotheses and discussed their characteristics.

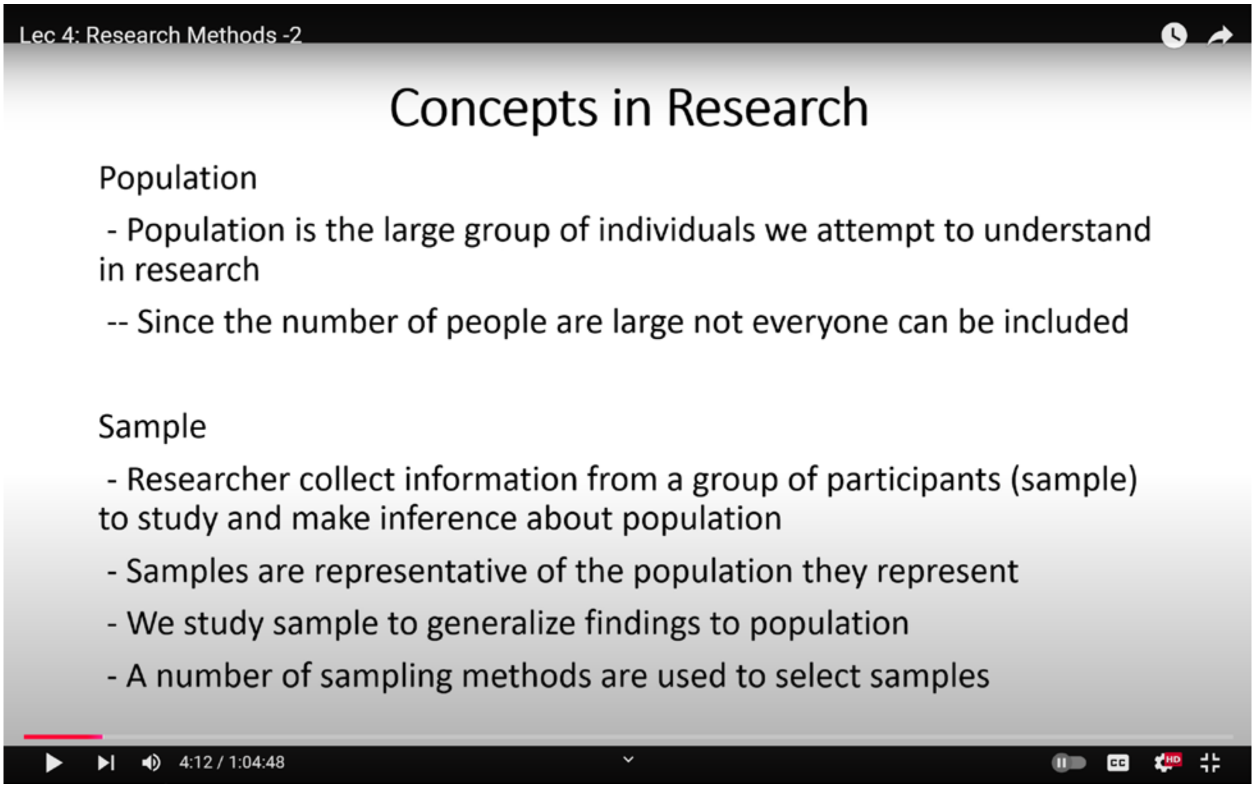
(Refer Slide Time: 01:19)



Additionally, we covered the distinction between basic and applied research and delved into concepts of research where we explored the differences between a study, an experiment, and a quasi-experiment. Today, building on the previous lecture, we will further explore the research process in greater detail.

Whenever we conduct research, the first step is to identify a problem, followed by designing a method to test this problem. I will reference the experiment discussed in the last class, where we explored whether distractions, such as using a cell phone while driving, could lead to accidents. We proposed examining whether car manufacturers could introduce solutions that allow car software to compensate for momentary lapses in attention.

(Refer Slide Time: 04:12)



That was the problem, and we came up with several intuitive solutions and hypotheses that we wanted to test. While testing these hypotheses against the problem statement, it's important to remember that this issue could affect many people. By this, I mean that when we discuss drivers and how their attention shifts, we are referring to a broad category of drivers, which includes men and women from various countries, nationalities, age groups, education levels, and training backgrounds. All individuals falling within this driving category are referred to as the population.

The population is essentially a group of individuals who possess the specific characteristics being studied. By definition, the population is the larger group of individuals that we seek to understand through research. However, one challenge we face in most studies is that it is impossible to include every individual within the population. For example, in our driving study, we cannot test every single driver on the planet or even within a specific city on how they behave when driving and interacting with a cell phone. This presents a significant challenge in population studies, populations often consist of a vast number of individuals, making it impractical to test everyone.

Populations can be both defined and undefined. Defined populations are more manageable because we know the approximate number of individuals who meet the criteria, such as the number of licensed drivers. However, an undefined population is one where the total number of individuals is unknown. For instance, a quick glance at the DMV database can tell us how many drivers have been issued licenses in a city, giving us a rough estimate of the driving population. But we are also aware that some people drive without licenses, contributing to the undefined population.

In simple terms, the population comprises all the individuals of interest to our study. Since it is not feasible to test everyone in the population, the next best approach is to select a representative sample from the population. So, what is a sample? A sample consists of a group of individuals randomly selected from the population, and we believe this group will possess the same characteristics as the overall population.

Essentially, within a population, there will be individuals from different nationalities, educational backgrounds, age groups, and genders. By using random sampling, we ensure that people from each of these categories are included in our sample. We then test this sample against the hypothesis and search for solutions. In this way, the sample becomes a subset of the population, which researchers believe represents most of the population's characteristics. These characteristics are referred to as parameters of the population.

So, how do we define a sample? Researchers collect information from a group of participants, called a sample, to study and draw conclusions about the population. This is the essence of what I was explaining. For instance, if I have a population of 10,000 drivers in a city, this population will include men, women, individuals from various age groups, nationalities, castes, and educational backgrounds. From this population of 10,000 people, I will randomly select a sample.

By selecting people randomly from the population, we assume that the sample will reflect the diversity of the population, including its various genders, education levels, age groups, and so on. Samples are representative of the population they are drawn from. As mentioned earlier, when a sample is taken from a population, it is assumed that the sample represents the entire population, or at the very least, that every individual in the population has an equal chance of being included in the sample.

To ensure that everyone in the population has an equal chance of being included in the sample, we use a technique known as random sampling. Random sampling involves selecting individuals from the population in a random manner, ensuring that the sample is representative of the population. We study the sample to generalize the findings to the broader population. The results obtained from the sample are generalized to the entire population, meaning we believe that the sample represents the population.

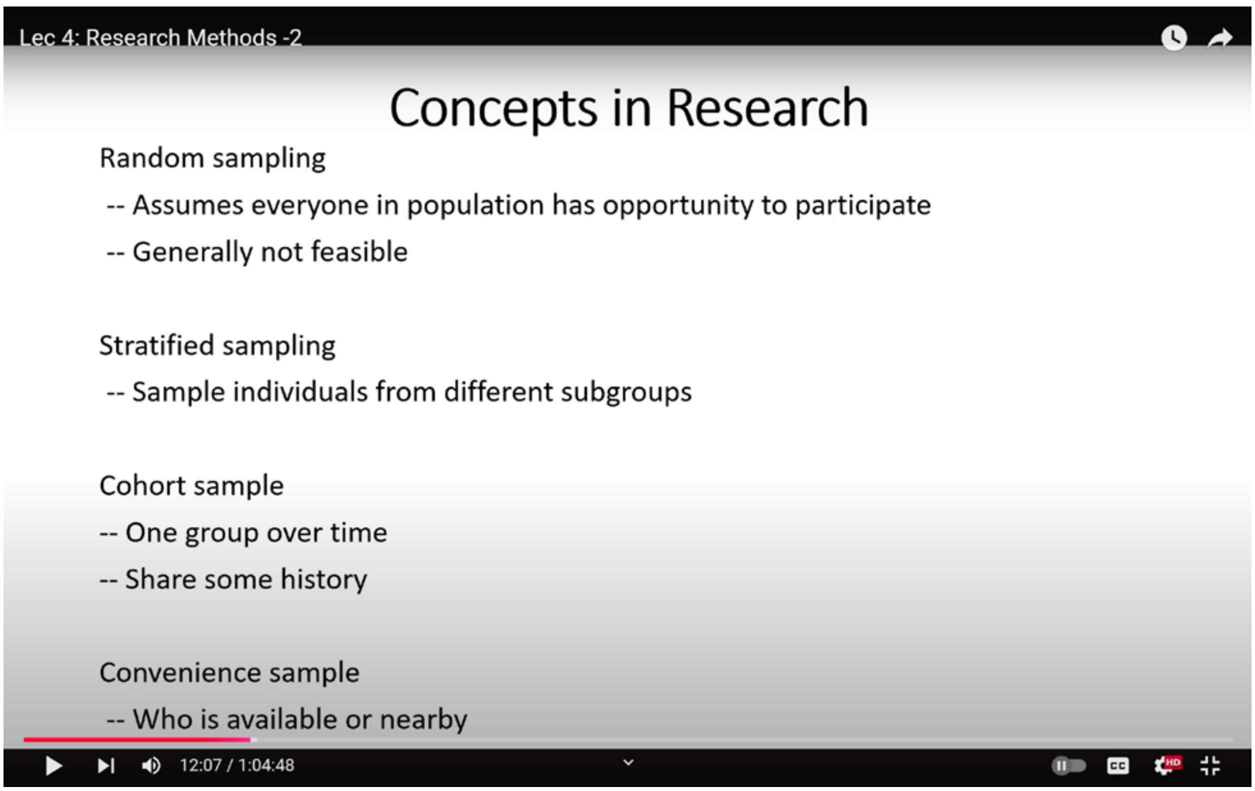
For example, if we find that providing tactile feedback through vibrations in the steering wheel helps prevent accidents in our sample, we can generalize this finding to the population, implying that this feedback mechanism will likely benefit most drivers in the city.

Now, a number of sampling methods are used to select samples. So, how do we go about sampling? One method that I have already explained is called random sampling, where we randomly select individuals from the population and use them as a sample. But are there other methods? What about situations where the distribution of different types of people within the population is unequal? For instance, we may have an unequal distribution of men and women or individuals with varying levels of education within the population. So, how do we ensure that the sample is representative, meaning that it closely approximates the population?

Let us now discuss some sampling methods that help in selecting individuals from the population in such a way that they represent the population accurately, making it easier to generalize the results from the sample to the population. The first method is random sampling. Here, the assumption that researchers make while collecting a sample is that everyone in the population has an equal opportunity to participate. Since individuals are randomly chosen from the population, each person has a probability of being selected. This means that if we take multiple samples from the population, everyone will have a chance of being represented across these samples.

For illustration, if we have a population of 100 individuals and a sample size of 10, we might randomly select individuals numbered 1, 5, 8, 20, 40, and 47 in the first sample. If we take another sample, it's possible that individuals selected in the first case may not appear in the second sample, and others might be included. By taking multiple samples, the likelihood increases that most individuals from the population of 100 will be represented in different samples. Thus, random sampling operates under the assumption of equal probability.

(Refer Slide Time: 12:07)



However, random sampling is often not feasible for several reasons. One issue could be that although a population may seem homogeneous, it is not. This means that different features of the population, such as gender, education, or training, may not be equally distributed. For instance, there might be more males than females, or the population may be undefined, making it difficult to calculate the probability of selection for each individual. In such cases, random sampling becomes challenging, leading us to consider other methods.

When the population is unequal in terms of characteristics like the number of males versus females, levels of education, or social categories, we use stratified sampling. In stratified sampling, we first create different strata by dividing the population based on the characteristic of interest. For example, if education is our focus, we could divide the population into groups such as those educated up to the 5th grade, those educated from the 5th to the 10th grade, those with high school diplomas, and those with higher degrees. We then collect individuals from these groups in proportion to their representation in the population.

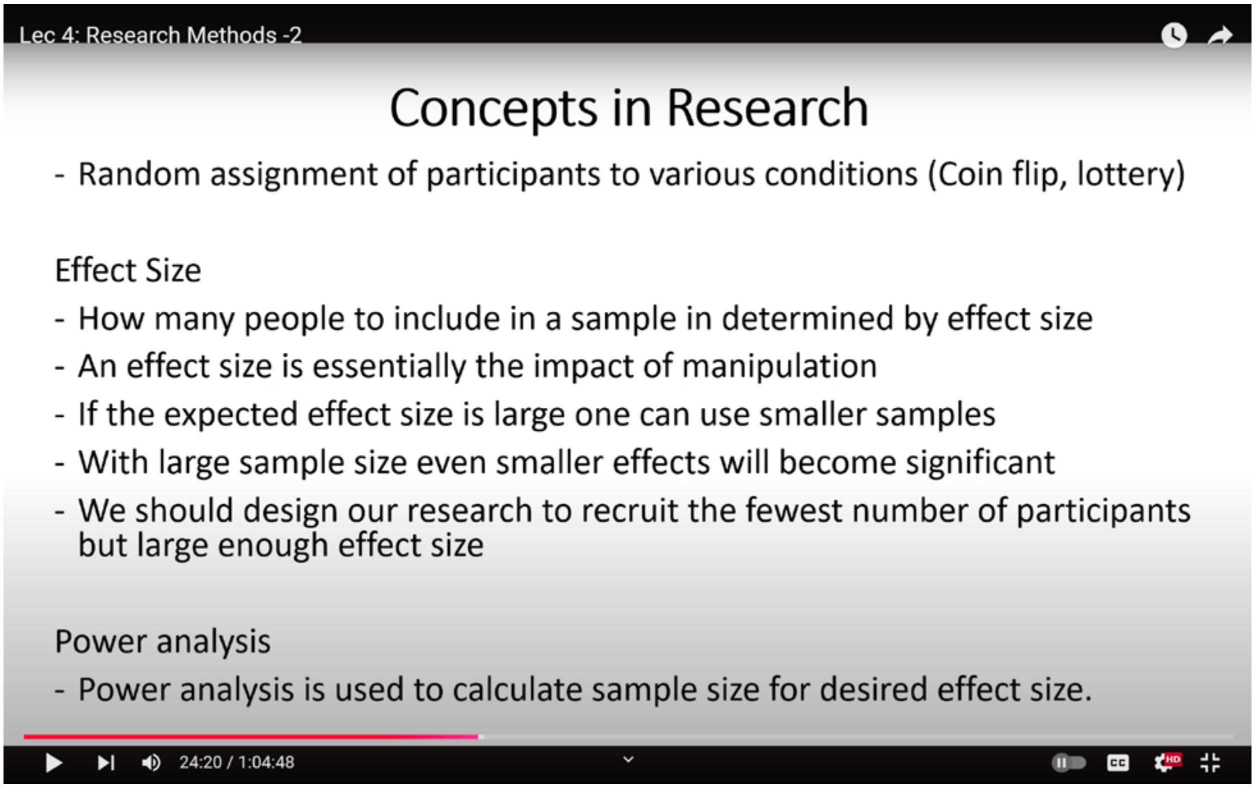
For example, let's say we have 10 people who are educated up to the 10th grade, 30 people educated up to the 5th grade, and 60 people with graduate degrees. In this case, we would select individuals from each of these categories in proportion to their numbers. So, if we want to sample from this population, we may select one person from the 10th-grade group, three people from the 5th-grade group, and six individuals from the graduate group. This way, the sample represents the population in a balanced manner.

Within each stratum, we will then randomly select individuals. For example, among the 30 people educated up to the 5th grade, we randomly pick three individuals. Similarly, from the 10 people educated up to the 10th grade, we select one individual, and from the 60 people with graduate degrees, we randomly select six. This process ensures that each subgroup is proportionally represented, and individuals within these subgroups are randomly chosen. This method is known as stratified random sampling.

Another sampling method is cohort sampling, where we study a specific group over a period of time. Instead of testing different individuals, we select one group or individual and study them for a duration of 5 to 10 years. This is called cohort sampling, and since the same individual or group is studied over time, they share a common history, allowing for more detailed observations.

Finally, the last method commonly used is convenience sampling. This method is employed when neither random sampling nor stratified sampling is feasible, often due to the lack of knowledge about the population's composition or the inability to obtain a representative sample. In convenience sampling, we select individuals who are readily available for the study, regardless of their representation in the population. This is often a practical approach when other methods cannot be implemented.

(Refer Slide Time: 24:20)



Once we have selected individuals from the population using a sampling method, the next step is to assign them to different groups in a random manner. We can use methods like coin flips or lotteries to do this. For instance, if we have taken a sample of 1,000 individuals from a population of 10,000, we would randomly divide these 1,000 people into two groups. One group might receive tactile feedback from the car's steering wheel, while the other group receives auditory feedback. Alternatively, we could design the experiment such that one group receives tactile feedback, and the results from these groups could be compared.

In the second group, individuals do not receive any feedback at all, allowing us to examine how their driving performances differ in situations where attention may be diverted to a cell phone.

These scenarios represent emergency situations. Random assignment ensures that participants are selected randomly, as well as the conditions of the experiment.

The concept of effect size is the third element that needs to be discussed here. What is an effect size? It may occur that we have two different groups: the experimental group, which receives feedback from the car steering system, and the control group, which does not receive any feedback. We aim to test whether these groups exhibit any differences in terms of driving performance. We can evaluate these groups under varying driving conditions, such as high-load driving conditions and low-load driving conditions, or normal driving conditions.

Effect size indicates how effective we believe our experiment will be and whether the design we have created will yield successful results. The determination of how many individuals to include in a sample is based on effect size, specifically, how large the sample size should be. If we believe that the conditions under which the experiment is conducted have a minimal effect size and are not significantly different from each other, we will need to take larger samples. However, if the conditions differ significantly, we may be able to utilize a smaller sample.

Effect size essentially reflects the impact of manipulation. It addresses how much we believe that altering the type of feedback received from the car steering system affects driving performance. If we anticipate a substantial impact, smaller samples can suffice. Conversely, if we expect the impact to be less pronounced, a larger sample size is necessary. In summary, if the anticipated effect size is large, smaller samples may be appropriate, while larger samples are required for smaller expected effect sizes.

It is important to note that larger sample sizes can render even minor effects statistically significant. Therefore, in cases of very large sample sizes, even a small difference between the control and experimental conditions can yield significant results due to the sheer number of participants and the statistical methods employed. Thus, we should design our research to recruit the fewest participants necessary while ensuring a sufficiently large effect size.

When designing experiments, we must be cautious not to recruit an excessive number of participants, as this can manipulate statistical outcomes. The larger the sample size, the more likely it is that results will be deemed statistically significant. Our goal should be to determine the optimal number of participants to assess the effectiveness of our experiment accurately.

A mediocre effect size typically falls between 0.3 and 0.6, while a small effect size ranges from 0.1 to 0.3. An effect size above 0.6 to 0.8 is considered high. These thresholds can vary; for instance, we may classify 0.3 to 0.4 as a medium effect size and 0.4 to 0.6 as a high effect size. This means that 60% of the variation in the data can be explained by the changes we have implemented or the new design we have employed.

To determine the appropriate sample size, we can also utilize power analysis. In power analysis, we review existing literature to see what different experiments have achieved and how successful they have been in terms of effect size. We examine the impact of manipulation they reported and, based on this information, we can estimate the effect size and significance level, essentially, how frequently we believe an effect occurs by chance. This data can be incorporated into a formula to calculate the required sample size.

The next important concept in research is the idea of variables. Various types of variables are employed in research, including independent and dependent variables. Independent variables, sometimes referred to as predictor variables, are those that the researcher manipulates. In contrast, dependent variables are known as criterion variables, which are the outcomes measured to assess the effects of the independent variable.

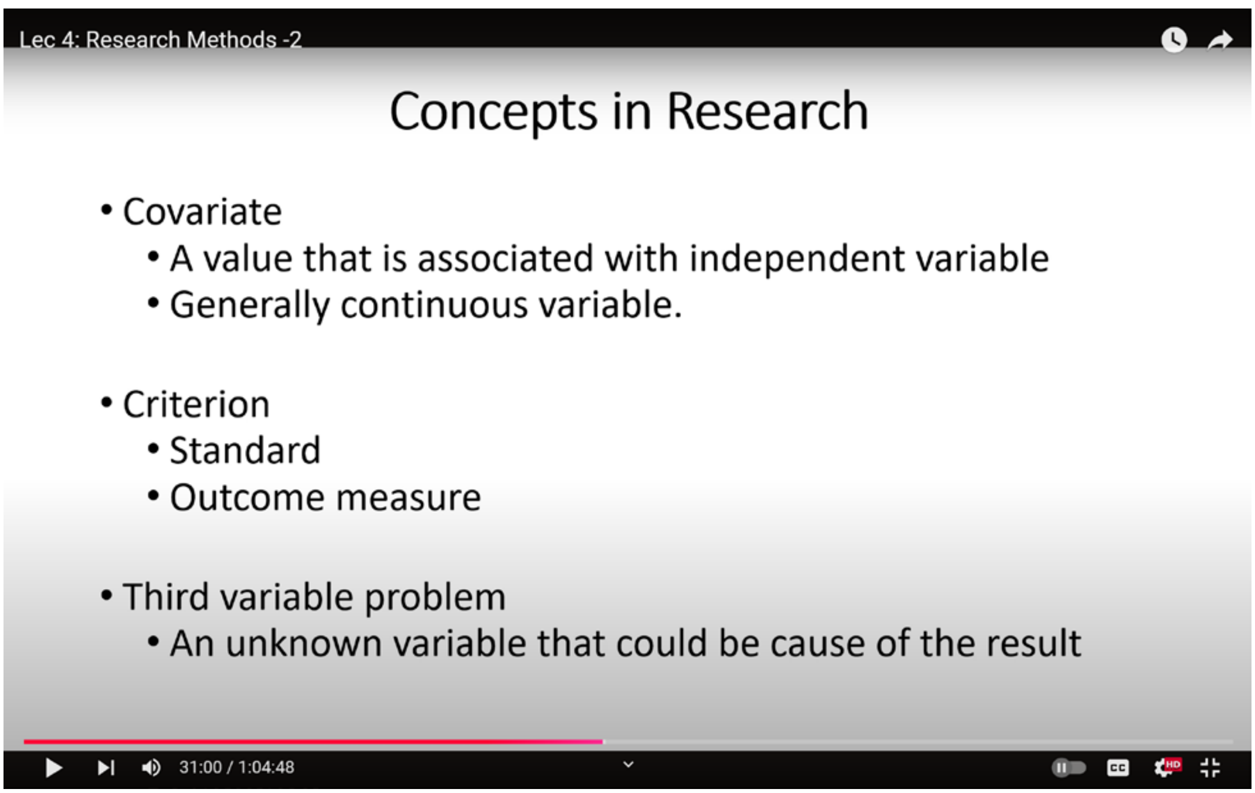
The type of feedback received from the software serves as the independent variable. This feedback can take various forms, such as tactile feedback, auditory feedback, or even no feedback at all from the car's software system or the vehicle itself. This independent variable influences the dependent variable, which is how this feedback translates into avoiding accidents or mitigating accident situations based on driving performance. The driver's performance is assessed by evaluating how accurately they can avoid an accident.

Independent variables are those that can be manipulated during an experiment. As previously mentioned, independent variables are sometimes referred to as predictive variables, and the experimenter has complete control over these variables. Examples of independent variables include environmental colors, such as blue, yellow, green, and white, as well as the type of feedback provided by the car's software. They can also include teaching methods employed by an instructor in a classroom. By altering the independent variable, we seek to measure the resulting changes in individuals' behavior. Thus, a change in the independent variable will prompt a change in behavior, which can be quantified through specific criteria. This measurable change is referred to as the dependent variable.

Consequently, changes in the independent variable lead to changes in the dependent variable. We believe that this behavior change in the subject, assessed through specific criteria, is the process of measuring the dependent variable. Dependent variables are often called outcome variables because they reflect the results of modifications made to the independent variable. Examples of dependent variables include performance, mood, and reaction time.

Defining how we measure the dependent variable is crucial. For instance, consider performance: if we change the color of the environment, we need to assess how performance is impacted. We might conduct an experiment with workers in differently colored rooms to observe variations in their performance.

(Refer Slide Time: 31:00)



However, performance itself is not inherently measurable; thus, we must define it more precisely. We can quantify performance in terms of the number of units of work completed, the accuracy of task execution, or the speed at which a job is performed. Therefore, we would examine whether different colors enhance performance in terms of productivity or efficiency. Similarly, mood can be measured qualitatively as either positive or negative.

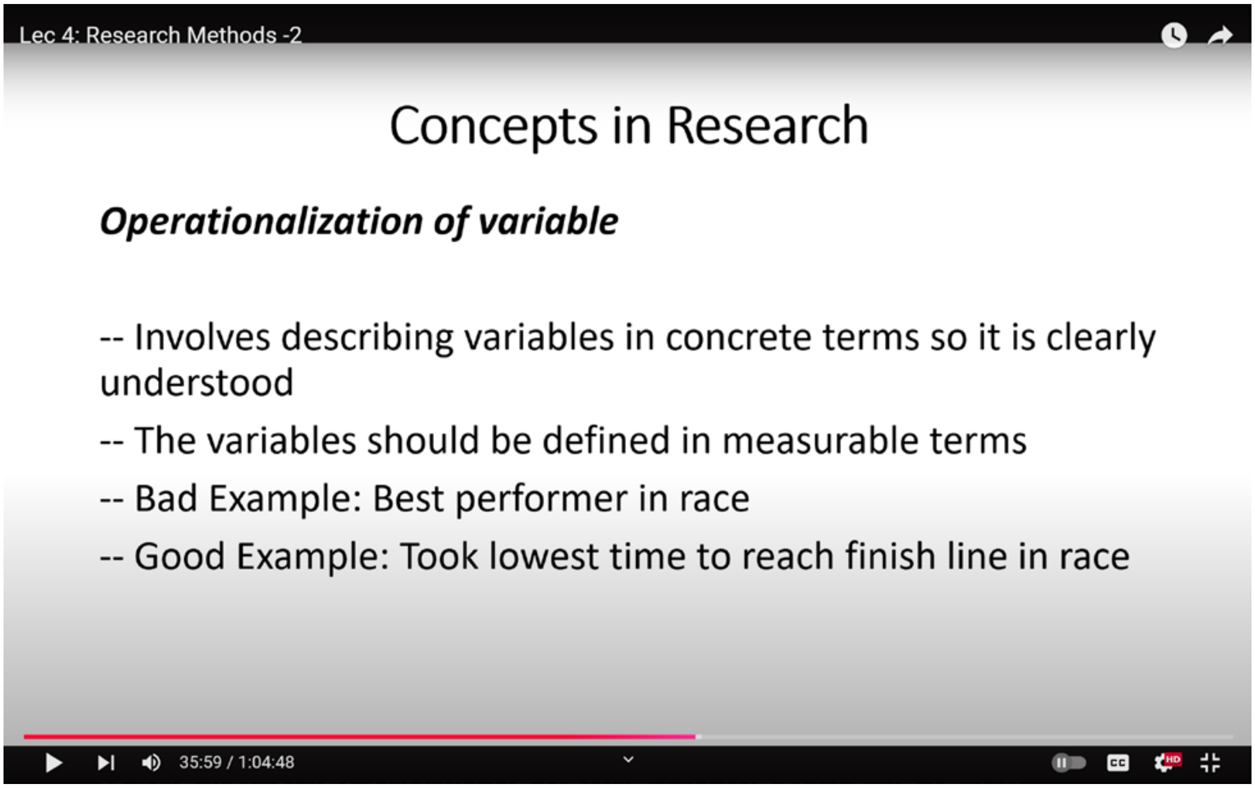
This leads us to investigate whether the color of the environment influences people's moods, as these mood changes can, in turn, affect performance. This illustrates the definitions of independent and dependent variables. Additionally, we may encounter a covariate in our research. Sometimes, independent variables do not exert a direct effect on dependent variables but instead influence them through an intermediary variable known as a covariate.

A covariate is a value associated with the independent variable and is continuous in nature. For example, while colors may influence performance, this relationship may not be direct. Instead, colors could affect mood, which subsequently impacts performance. Alternatively, colors might influence personality traits, leading to variations in performance. For instance, extroverted individuals might perform better due to color changes, while introverted individuals may not experience any performance improvement from the same color adjustments. Here, personality acts as a covariate. Although it may not be immediately apparent, colors affect performance differently across various groups of people, influenced by this hidden factor, or intermediary factor, known as personality.

This is the essence of a covariate, and it should be continuous. As previously noted, dependent variables are measured using specific criteria. What constitutes a criterion? A criterion serves as the standard for performance, typically measured in terms of how many units of work have been completed or the accuracy of the tasks performed. These metrics are considered outcome variables known as criteria.

The concept of the third variable problem is also important to consider. This refers to an unknown variable that could potentially be the cause of the observed results. In some instances, a hidden variable might be responsible for the effects seen in an experiment, often without the researchers' awareness of this variable. This is referred to as the third variable problem. For example, if we fail to account for participants' intelligence, an attribute that relates to motor skills, it may be that individuals with higher intelligence are better able to avoid accidents, even without feedback from the car's steering system. Thus, intelligence could represent a hidden variable that we did not consider in our experiment, illustrating a potential third variable problem.

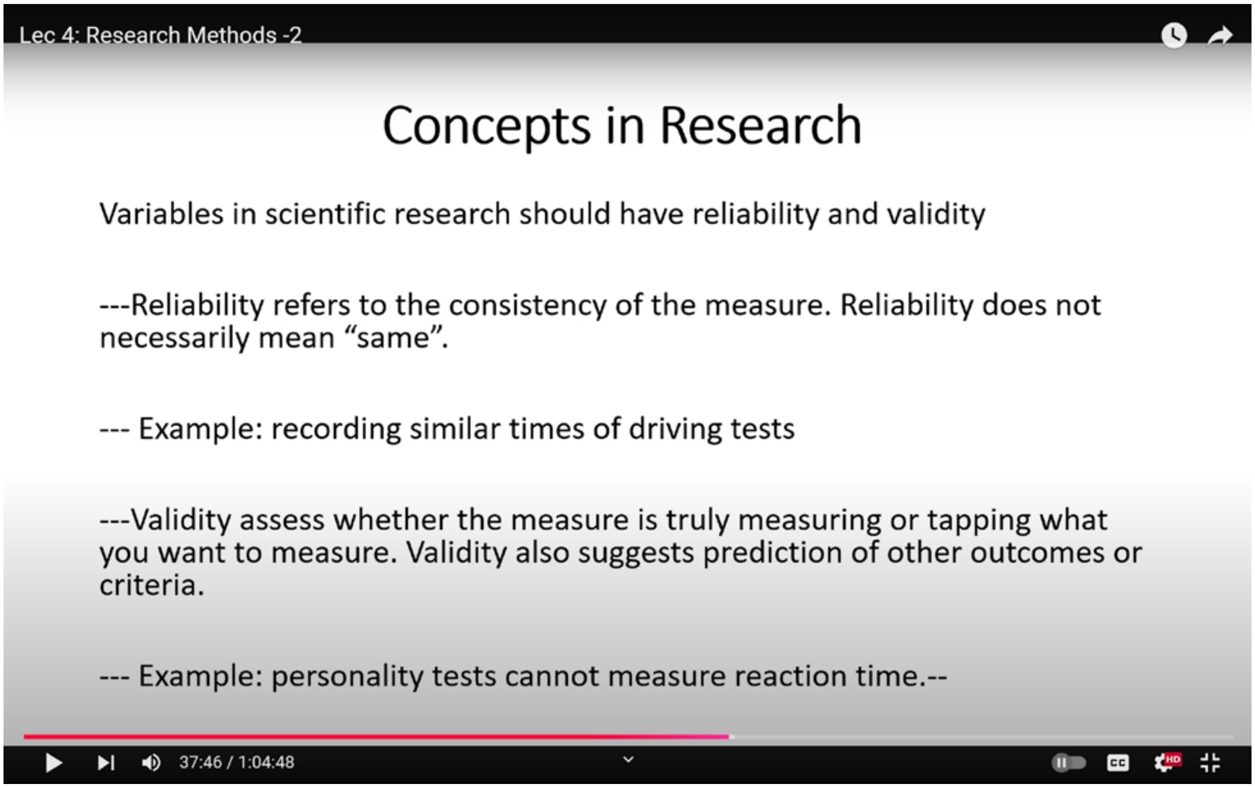
(Refer Slide Time: 35:59)



Another significant concept is the operationalization of the variable. In the field of behavioral sciences, variables can possess multiple interpretations. For instance, when defining happiness, how might one proceed? Happiness could manifest as elation, a sense of well-being derived from receiving a reward, or it could be characterized through various indicators, such as changes in facial expressions, physiological responses, or alterations in behavioral patterns.

When defining happiness, it is essential to specify which aspect of behavior is being targeted as a measure of happiness. Are we focusing on physiological changes, facial expressions, personality alterations, or another factor? This process is known as operationalization. Operationalization involves establishing a specific definition of happiness for the purposes of the experiment, and the results will only be applicable within this defined framework.

(Refer Slide Time: 37:46)

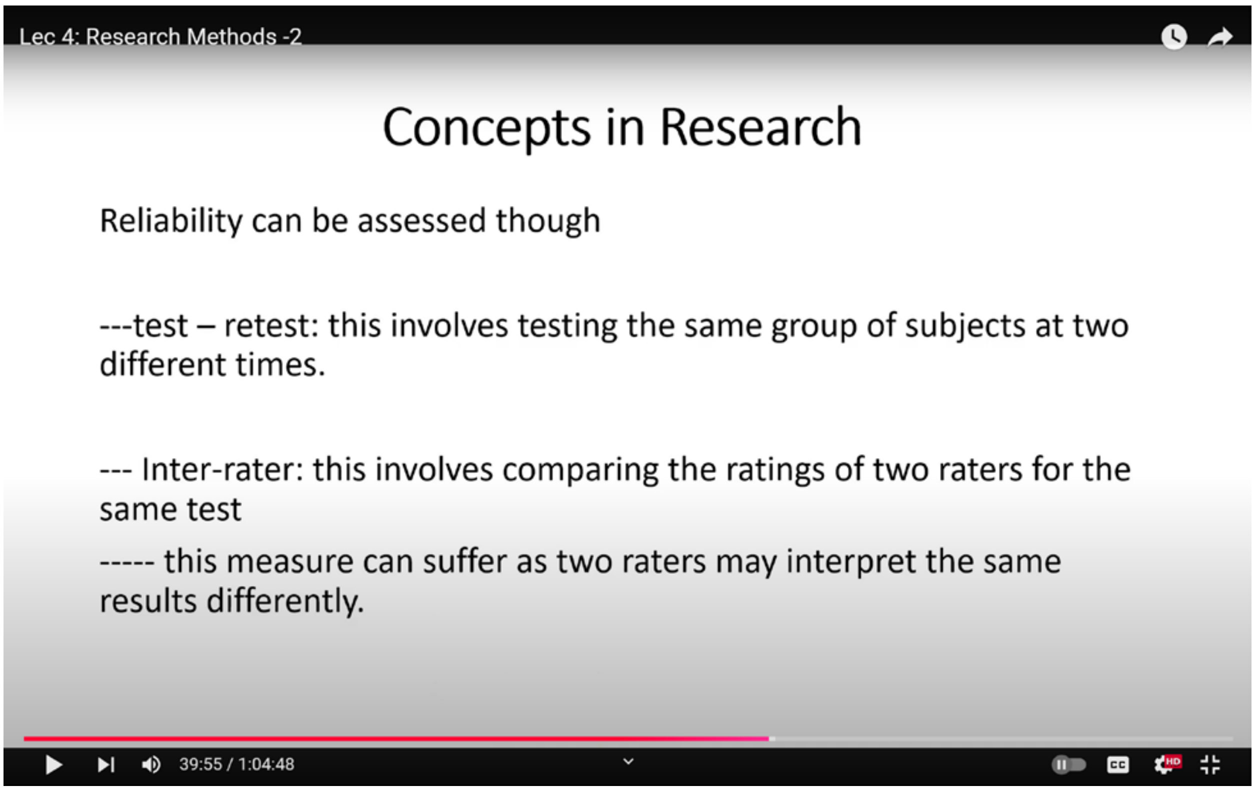


If happiness is defined or measured in alternative ways, the outcomes of an experiment aimed at exploring happiness may not be valid. This concept is termed operationalization, which requires describing variables in concrete, measurable terms to ensure clarity. For instance, happiness might be assessed through physiological indicators, personality assessments, or other relevant factors. When discussing physiology, we could measure it through Galvanic Skin Response (GSR) or arousal levels, which clearly outlines what is being measured.

A poor example of operationalization is defining a variable as the "best performance in arrays." This definition lacks specificity. What criteria determine this "best performance"? Is it based on timing, the number of errors made, or some other factor? A more effective example of operationalization would be defining performance in terms of the shortest time taken to complete a race. Here, the variable is defined by a specific metric, time, demonstrating effective operationalization.

In scientific research, it is also crucial for variables to possess reliability and validity. Reliability pertains to consistency. It does not necessarily imply that the same results will always be obtained, but rather that there is a stable level of performance. For example, if I take a driving test and my performance is assessed based on the number of errors, and I am subsequently retested without having taken any driving lessons, the test can be considered reliable if my scores remain consistent across both tests. Thus, reliability is indicated by similar scores obtained across multiple assessments over time, such as recording consistent times in driving tests.

(Refer Slide Time: 39:55)



To illustrate, suppose I take a driving test multiple times without any additional driving instruction and achieve comparable scores in each instance; this indicates that the test is reliable, as I am consistently achieving similar results. In contrast, validity evaluates whether a measure truly assesses what it is intended to measure. Validity also involves the capacity to predict other outcomes or criteria. Essentially, validity ensures that the test measures what it claims to measure.

For instance, consider a driving test that includes questions unrelated to driving skills, such as personal feelings or household members' driving abilities. Such questions do not assess driving competency and therefore lack validity. A valid driving test should ask questions that pertain directly to driving skills, such as how one avoids dangerous situations or understands vehicle controls. Valid questions would assess driving-related knowledge and skills, while irrelevant questions would fail to measure the intended driving ability.

Furthermore, a valid example would be that a personality test cannot accurately measure reaction time. While personality assessments can provide insights into individual traits, they do not assess speed or responsiveness.

To measure reliability, two primary methods are commonly used, among others. The first is known as the test-retest method, where the same group of subjects is tested at two different points in time using the same test. If the results remain similar, the measure is deemed reliable. For instance, I might conduct multiple tests over various time intervals for the same group, and if the results are consistent, this indicates reliability through the test-retest method.

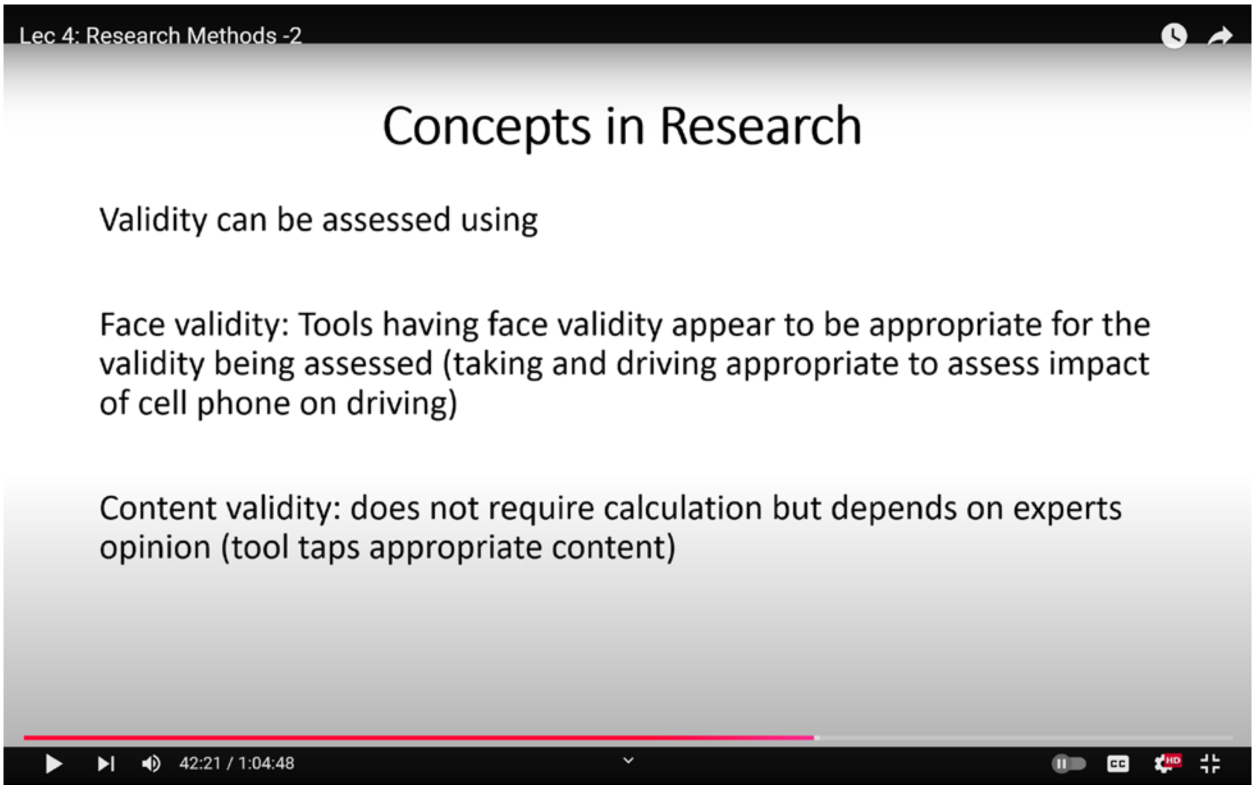
The second method is inter-rater reliability, which involves multiple experts rating the performance of a sample group. If the average ratings from different raters show agreement, this indicates inter-rater reliability. An analogy for this might be a beauty contest where several judges evaluate the participants. If multiple judges provide similar scores for a particular contestant, this indicates inter-rater reliability. However, a challenge with this approach is that different raters may interpret the same situation differently. People often have varied perceptions of beauty, leading to discrepancies in how they rate the same individual based on specific criteria.

Validity can be categorized into four types, which I will briefly outline. The first type is face validity. This form of validity refers to tests that intuitively appear to measure what they claim to measure. For example, using a driving simulator to assess the impact of cell phone use on driving performance provides a seemingly appropriate measure for this effect. Tools demonstrating face validity seem relevant and suitable for a given experiment or measurement purpose.

The second type of validity is content validity, which assesses whether the content of a measure accurately reflects the construct it is intended to assess.

This process does not involve any calculations but rather relies on expert opinion. When I assert that a particular device or tool possesses content validity, it signifies that the tool effectively assesses the content being measured. For example, if I am comparing introverts and extroverts, I would develop a questionnaire that addresses the characteristics of extroverted individuals. Initially, I would document the traits associated with introverts. Then, experts would evaluate the tool to determine whether the questions align with the established characteristics of extroverts and introverts. This evaluation process is known as content validity.

(Refer Slide Time: 42:21)



To clarify, how do extroverts and introverts differ? Such distinctions can only be articulated or understood by a subject matter expert. If an expert reviews the tool and compares it to the expected features of extroverted and introverted personalities, and if the tool's content aligns with these features, it is deemed to have content validity.

The third type of validity is referred to as criteria-related validity. This form of validity aims to assess how well the measurement tool correlates with some predetermined criteria. For instance, let us consider a scenario where I investigate how different colors affect mood. In this case, I would present various colors and observe the variations in people's moods. However, a challenge arises in defining how we measure mood. A positive mood is typically associated with happiness, while a negative mood correlates with sadness. Thus, when individuals experience a good mood, they are happy, and when they experience a bad mood, they are sad.

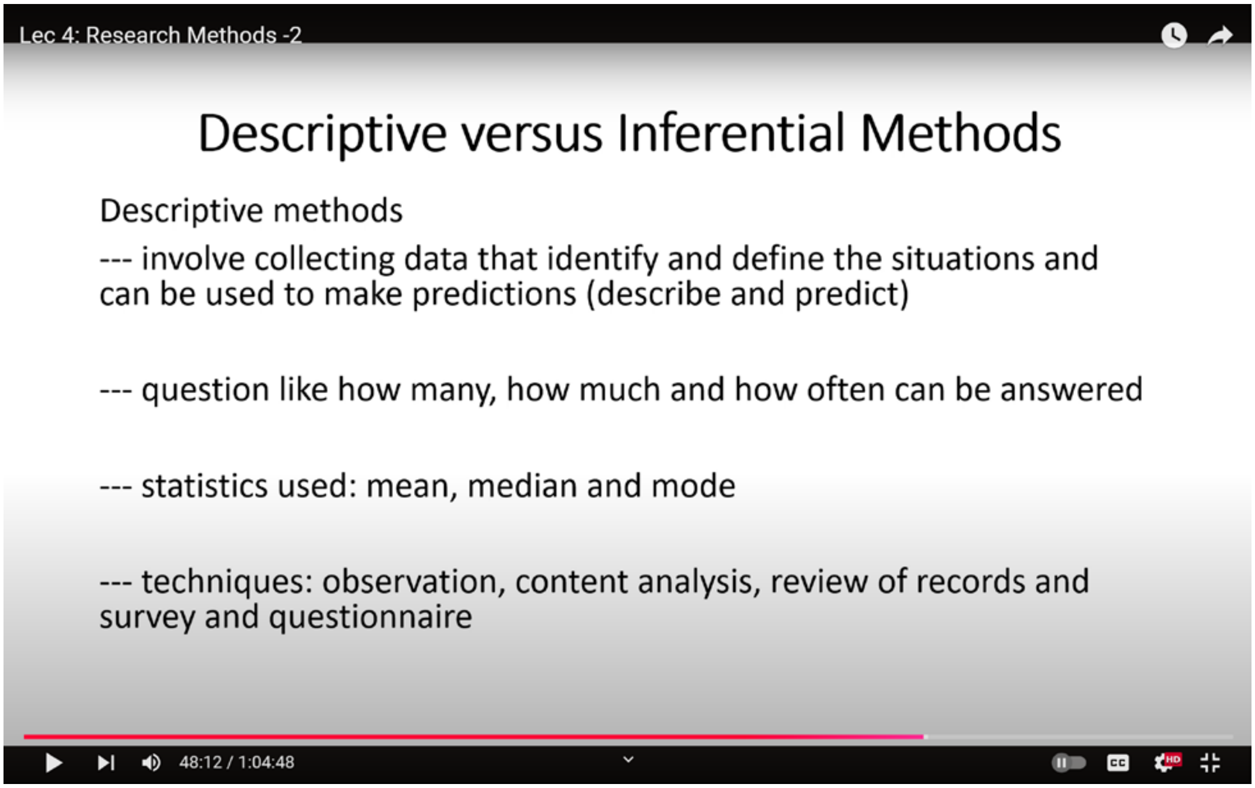
The scale for measuring mood is defined such that "good" corresponds to happiness and "bad" corresponds to sadness. If this scale is utilized to represent mood changes induced by color, the tool is said to possess criteria validity. For example, if individuals become happier when exposed to specific colors, we can interpret this as an indication of a good mood. Conversely, if exposure to a particular color results in sadness, we can conclude that the individual is in a bad mood. This approach exemplifies how we translate observations into conclusions regarding mood changes. Additionally, this form of validity is often termed predictive validity, as it helps forecast potential changes that may arise from utilizing a specific tool.

The final type of validity we consider is construct validity, which pertains to whether the tool is related to an established construct, such as personality. To clarify, what constitutes a construct? A construct is a hypothetical framework used to explain specific behaviors or outcomes, and these constructs are frequently developed within the behavioral sciences. For instance, I might determine that an individual's openness to color preferences is related to their personality. As previously discussed, introverts are generally less inclined to embrace color changes, while extroverts tend to be more receptive to them.

To measure an individual's preference for color, I could assess their personality and relate it back to their openness to color selection. It is understood that individuals who exhibit a greater openness to color are often extroverted, while those who are less open tend to be introverted. Thus, the personality measure can be effectively operationalized. We could administer a personality scale designed to differentiate between extroverts and introverts to participants involved in the color preference experiment. By assessing their personalities, we can predict that if someone exhibits extroverted traits, they are more likely to be open to color.

In construct validity, we examine how the construct is related to the concept being measured and subsequently evaluate the tool against that particular construct. Assuming the tools we have employed are both valid and reliable, we have gathered a sample from the population, formulated a hypothesis, and clearly defined a problem. The next phase involves statistical analysis, which can be conducted using both descriptive and inferential methods.

(Refer Slide Time: 48:12)



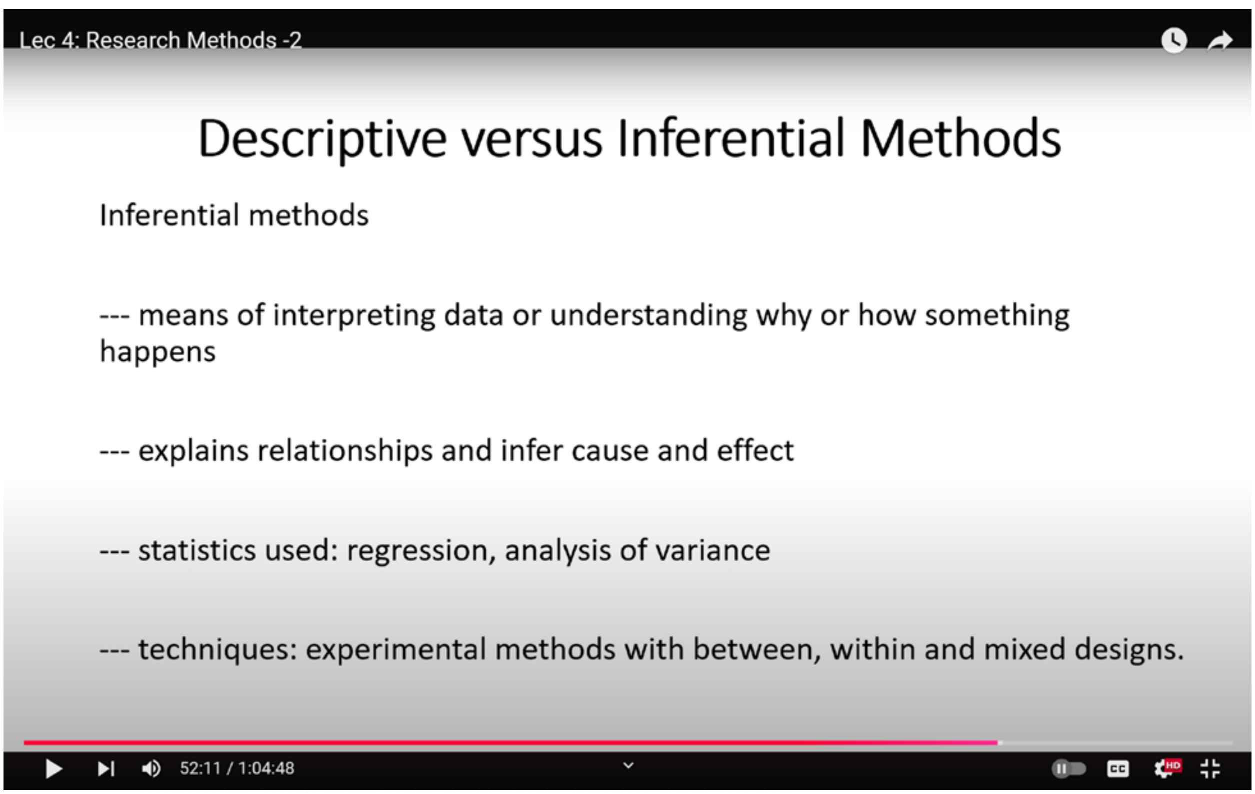
What, then, is the descriptive method of analyzing data? In the descriptive method, we collect data and aim to describe it. For instance, through descriptive analysis, we can represent the data and compare two groups based on this representation. Suppose we gathered data regarding performance on a driving test, which involved two groups: the experimental group that received feedback from a vehicle support system and the control group that did not.

We would also measure performance based on how effectively individuals navigated or managed an accident scenario. Following data collection from a sample population, we would create a visual representation comparing the performance of the feedback and non-feedback groups. If we analyze the data based solely on the number of participants who successfully completed the test within the control and experimental groups, we are employing descriptive research methods. This approach involves gathering data to identify and define specific situations, which can then be used for predictive purposes.

Descriptive analysis does not involve in-depth statistical evaluation or extraction of specific results from the data; it merely focuses on providing a description of the collected data. Questions such as "How many?" "How much?" and "How often?" can be effectively addressed through this method. In essence, descriptive statistics relate to answering inquiries about the frequency of occurrences, the number of individuals who have had particular experiences, and similar types of questions.

The primary types of statistics utilized in the descriptive method are the mean, median, and mode. The mean represents the central value of the data. The median is the precise midpoint, while the mode refers to the most frequently occurring value within a dataset. These statistical measures are integral to the descriptive method. The techniques employed in descriptive methods include observation, where we observe a specific group of individuals performing in various situations, and content analysis, which involves first documenting individuals' explanations and subsequently extracting relevant themes.

(Refer Slide Time: 52:11)



For instance, participants can be asked to describe their experiences in the experimental group versus the control group. They would write sentences, from which we would identify and analyze themes. Additionally, we can review existing records as part of secondary data analysis, as well as utilize surveys and questionnaires for descriptive research.

In contrast, the inferential method focuses on interpreting data and understanding the underlying reasons for observed phenomena. While the descriptive method merely describes data, inferential methods involve drawing conclusions, extracting results, and hypothesizing or predicting potential explanations for the observed outcomes. Inferential methods establish relationships in terms of cause and effect. For example, if tactile feedback is provided, participants may perform better; conversely, without tactile feedback, their performance may decline. Similarly, if auditory feedback is given, performance may worsen, whereas in its absence, performance may return to normal levels.

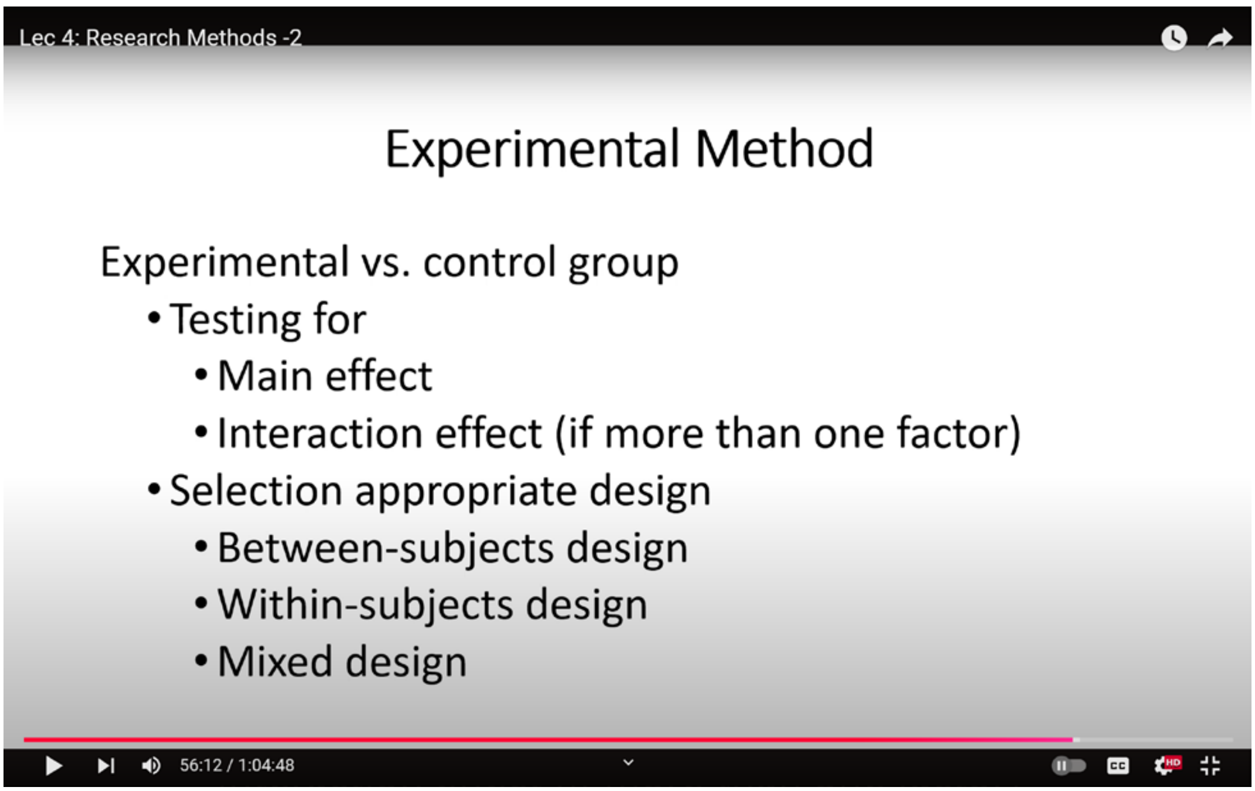
By comparing the three groups, we can conclude that tactile feedback yields the best results, while auditory feedback results in the poorest performance, and no feedback still produces acceptable outcomes. The statistics employed in the inferential method include regression analysis and analysis of variance (ANOVA).

What does regression analysis entail? In regression, we aim to fit a best-fit line that explains deviations within the data. When collecting data, an independent variable is plotted on the x-axis, while a dependent variable is plotted on the y-axis. When the data is represented as a scatter diagram, it may appear dispersed, indicating that individuals exhibit varying responses to changes in the independent variable or predictor variable. The regression line attempts to account for most deviations or data points, thereby explaining a significant portion of the variance within the dataset. Thus, regression lines serve as best-fit lines that elucidate individuals' behavior concerning the predictor and outcome variables.

Analysis of variance (ANOVA) is employed to assess variances or performance changes using either a t-test or an F-test. The techniques utilized in inferential methods often involve experimental designs. In the experimental method, we establish an experimental group and a control group. The experimental group undergoes variable manipulation, while the control group experiences no such manipulation. Various design types may be employed, including within-subject designs, between-subject designs, and mixed designs.

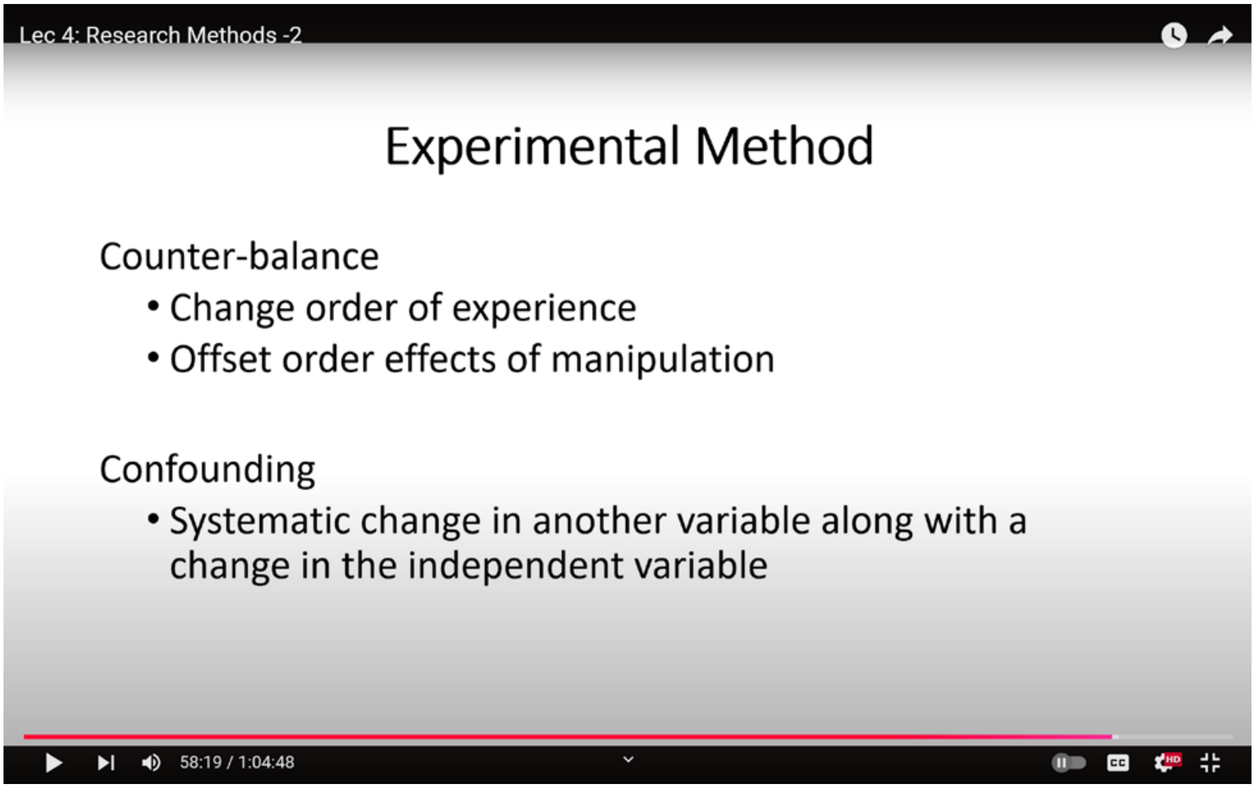
In a between-subject design, we utilize two different samples of participants for the experimental and control groups. Conversely, within-subject designs involve using the same participants under both experimental and control conditions. Mixed designs incorporate both distinct groups of individuals and the same group under different conditions. For instance, when assessing reaction times related to driving performance, we might have two different groups for one aspect of the study, but if examining errors, we can use the same group of participants under two different conditions.

(Refer Slide Time: 56:12)



Thus, mixed designs essentially combine elements of both within-subject and between-subject designs. Experimental control groups are tested for a variable referred to as the main effect, which is the primary effect of interest. For example, in a driving scenario, determining whether tactile feedback is superior to no feedback would constitute the main effect. The interaction effect may arise under specific conditions, such as low driving visibility benefiting from tactile feedback, whereas in high driving conditions, tactile feedback may perform as poorly as no feedback at all. These are classified as interaction effects when multiple factors are involved.

(Refer Slide Time: 58:19)

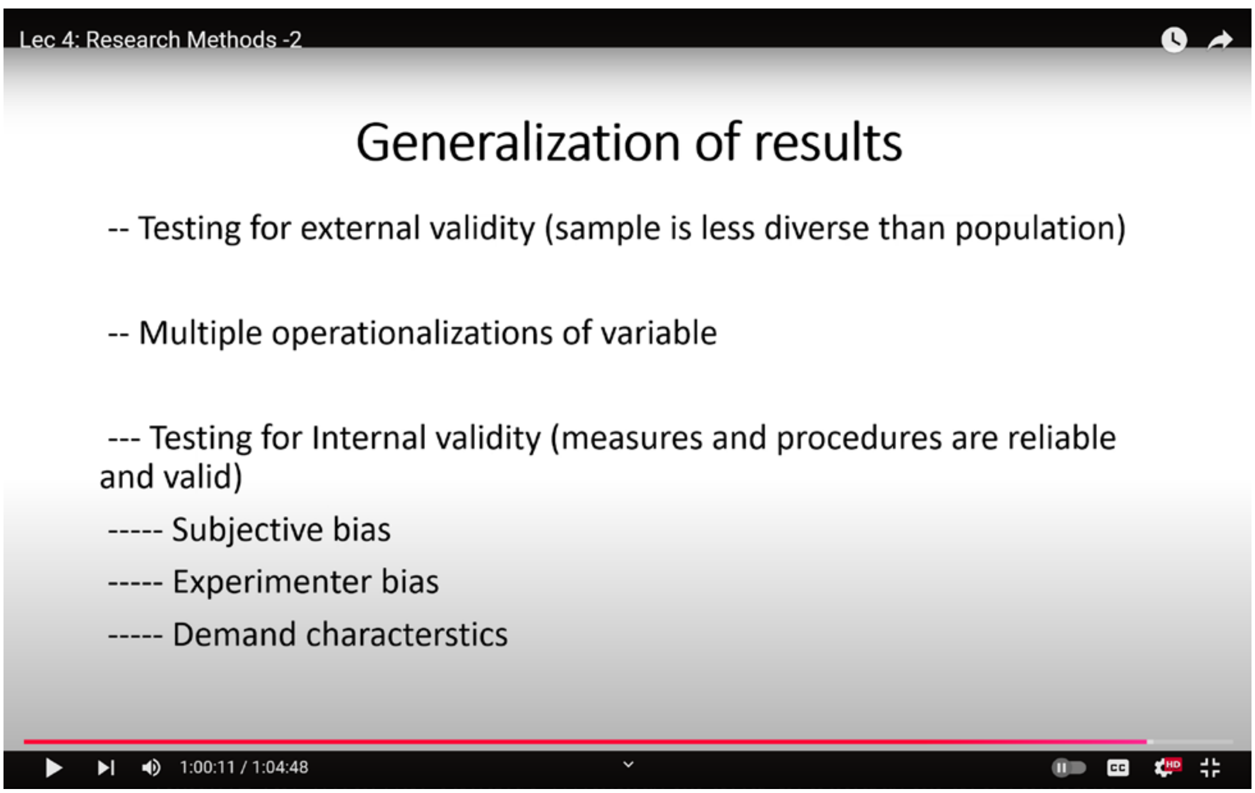


The design types previously discussed include between-subject designs using two different groups, within-subject designs employing the same group of individuals, and mixed designs combining both approaches. In experimental methods, we may implement a strategy known as counterbalancing. In counterbalancing, certain individuals in the experimental group may first experience the control condition, followed by the experimental condition. Conversely, other participants would start with the experimental condition and subsequently transition to the control condition.

For example, consider two variables: different types of accident scenarios categorized as low, medium, and high severity, with the experimental group employing a new design and the control group lacking any design or feedback. Counterbalancing would involve some participants starting in the low-accident condition of the experimental group, then proceeding through medium to high-accident conditions. In contrast, other participants might begin with the medium accident condition, proceed to high, and finally experience the low condition. This strategy is intended to mitigate the potential for participants to become habitual, as they might consistently transition from low to medium to high conditions. By employing counterbalancing, we can reduce the risk of participants developing predictable patterns.

To further address this issue, we utilize randomized methods to randomize the various conditions. Confounding can occur when systematic changes in another variable coincide with alterations in the independent variable. This means that a third variable of interest may also be influenced, potentially serving as the reason for the observed results. As previously discussed in relation to intelligence, this situation exemplifies what constitutes confounding.

(Refer Slide Time: 1:00:11)



The next step we want to address is the results we have obtained. We previously established a hypothesis, designed the experimental methods, collected data, and now we arrive at the stage of generalization. But what exactly is generalization? Generalization, particularly in the context of testing for external validity, involves taking the results from a sample and applying them to a larger population. This means making statements about the population based on the results derived from the sample. However, generalization can be problematic due to issues related to external validity, especially when the sample is less diverse than the population.

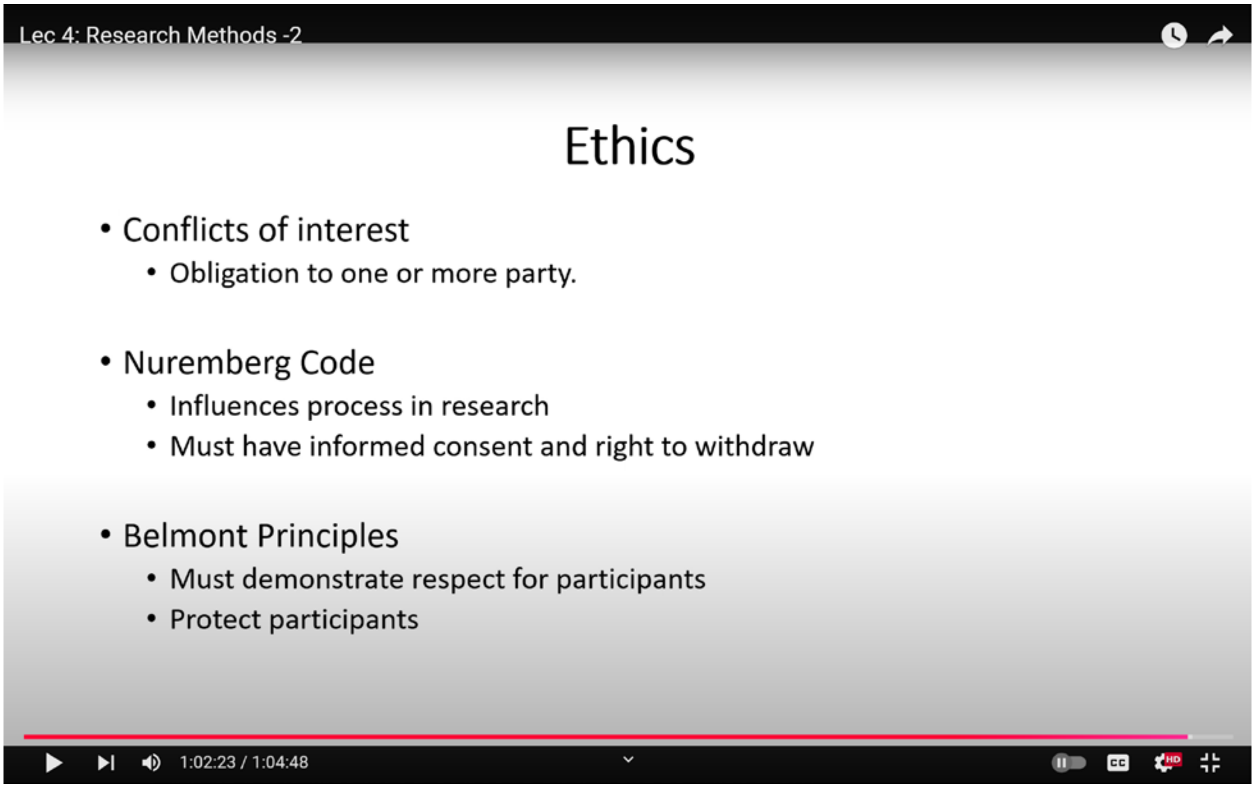
For example, if my driving experiment only involves a population of young individuals, and I find that changes in feedback lead to improved performance, this finding applies solely to younger people. If I then extrapolate this result to assert that all drivers would benefit from this type of feedback, it may not hold true. This limitation arises from the issue of external validity, as the results are only applicable to young drivers and may not translate to other age groups or types of drivers.

To address this challenge, we need to employ a strategy known as multiple operationalization of variables, which entails defining our variables in various ways. Generalization also requires testing for internal validity, which means ensuring that the measures and procedures used are both reliable and valid. It is crucial to avoid subjective biases that could influence the results. We must ensure that the participants do not exhibit bias in their responses, and the experimenter’s biases, stemming from their own beliefs and expectations, do not distort the interpretation of the results.

Thus, while I may obtain certain results from the sample, the experimenter's perspective should not lead to overinterpretation or assumptions about broader applicability. It is essential that any findings derived from the sample are conveyed clearly and concisely when applied to the population, thereby minimizing experimental bias and the influence of the experimenter's beliefs. Demand characteristics can also affect certain types of tests; for instance, if we are conducting a memory test, demand characteristics might suggest that participants should anticipate a recall task because they were not provided with a list to memorize.

These demand characteristics, essentially the implicit cues from the experiment, must be minimized. The final section of our discussion pertains to ethics. When conducting experiments, several ethical considerations must be upheld. Firstly, there should be no conflict of interest between the experimenter and the participants; it is important to ensure that there are no obligations that may bias the research.

(Refer Slide Time: 1:02:23)



The Nuremberg Code should be adhered to, as it establishes ethical guidelines for conducting research involving human participants. This code mandates that informed consent is obtained and that participants have the right to withdraw from the study at any time. The Nuremberg Code emerged from the trials that followed World War II, leading to the formulation of guidelines that protect the rights of individuals involved in research. Participants must be informed of their rights, the nature of the study, and the processes involved prior to their involvement.

Additionally, the Belmont Principles should also be respected, emphasizing the importance of demonstrating respect for participants and protecting their rights. One historical violation of these principles occurred during a study involving a group of African American men with syphilis, who were not informed of their condition and were denied treatment with penicillin. Such ethical breaches are prohibited by the Belmont Principles, which affirm that all individuals possess rights that must be safeguarded.

Before initiating any research, it is crucial to obtain informed consent, ensuring that participants are aware of their rights, the nature of the study, and the measurements being employed. While some level of deception may be permissible in research, it must be accompanied by informed consent, and participants should voluntarily agree to take part in the study.

Moreover, the risk-benefit ratio should be carefully considered, with all potential risks associated with the experiment clearly explained to the participants, alongside the benefits they may derive from it. The design of the experiment should aim to minimize risks as much as possible. Deception, when necessary, should be followed by debriefing; this means that after the experiment, participants should be informed about what data was collected, the true nature of the experiment, and how they performed.

Lastly, institutional review boards should be established to oversee research protocols and ensure that experiments are conducted in accordance with ethical guidelines. In summary, the discussions from this lecture and the previous one have outlined the fundamental principles of conducting research in behavioral sciences, specifically within the context of engineering psychology relevant to this course. In our next meeting, we will explore additional engaging topics within psychology. Thank you, and Namaskar from the MOOCs studio for today.