Today's lecture is about research methodology in psychology.

Goals of scientific research

Description: we might all have an idea of what a psychological concept such as intelligence means, but intelligence is an abstract concept. In order to study and measure it, psychologists need to define this construct and specify the types of observable behaviors that would reflect it.

Prediction: it's about finding relationships between variables and being able to anticipate when a phenomenon will occur. It is not necessary for prediction to understand exactly what is going on. You might know that dark skies announce rain without having to understand the details of how rain happens.

Explanation: it's about identifying underlying causes of relationship.

Control: control is very useful to be able to do something with the knowledge we are getting by doing research. For example, if someone has a mental illness, we want to be able to use the knowledge obtained by doing research on mental illnesses to try to help them.

Reaching the goals of scientific research

There are various things researchers can do to reach these goals: they can do experiments, correlational studies, or descriptive studies. In this course, we will only cover correlational studies and experiments because those are the methods psychologists use most. If you want to know more about descriptive research, refer to the book.

Correlational studies

There is a correlation between two variables when a variation in one variable corresponds to some variation with another variable. For example, Weight and height tend to be correlated: taller people tend to be heavier (positive correlation). A correlation must be present in order to have possibility of a causal relationship between two variables, but a correlation is not sufficient for causality. Correlation and causation are not the same, as you have often heard. In many cases, two variable are correlated but are not necessarily causally related. Some of the correlations are meaningful, others are spurious.

Coincidence example: imagine that you just happen to think of a friend, and right then, the friend calls you; that might lead you to think that this means something. But in order to know for sure/scientifically whether or not this means something, you would have to collect the following data: all cases when you think of your friend, and your friend calls you; all cases when you don't think of your friend, and your friend calls you; all cases when you think of your friend, and your friend does not call you; and finally all cases when you don't think of your friend, and your friend does not call you. You are likely to find that the other cases are much more frequent, and that your friend calling you when you think of them might just be due to chance/randomness. Multiple observations can help us learn not to interpret coincidences as true causal relationships. We tend to make errors because we overly pay attention to cases when we turn we think of our friend and our friend calls us; coincidence is very easily noticeable.

Another way to avoid being trapped by coincidences is to use comparison group. If we were to make inferences about the Yankees and their likelihood of winning after 8 innings, we would probably need to collect data on the Yankees' wins and losses before and after 8 innings, and most importantly, compare the Yankees to other teams, and other teams' wins and losses before and after 8 innings.

We also make errors because we don't know the temporal order of two variables that are correlated. For example, we know for sure that sex and pregnancy are correlated, and we also know the temporal order: sex comes before pregnancy, not the other way around obviously. But for other variables, it's not as obvious: researchers know that there is a correlation between TV violence and aggressive behavior, but the temporal order is not clear. Is it that people who are aggressive watch violent TV, or is it that people who watch violent TV become aggressive? Why is there the relationship? Which causes which?

Even if we knew for sure that watching violent TV preceded violent behavior, we still can't establish causation. Sometimes, a third variable explains the relationship and the temporal order. In class, we talked about the following example: ice cream consumption and death rates (drowning, shark attacks) are both highest in June, July and August, so it is known that there is a correlation. But saying that ice cream consumption causes drowning and shark attacks, that would not make sense. And saying that shark attacks and drowning causes ice cream consumption would not make sense either, despite the fact that the correlation exists. A third variable causes the correlation between ice cream consumption and drowning and shark attacks: summer heat! People have more daylight hours, have more time because of summer vacation, it's hot, so they eat more ice cream and spend more time at the beach.

In class, we also talked about a magazine article marriage keeps ex-cons out of trouble (if you want to avoid a life of crime, get married). The method used for this article was the following: ex cons were followed for 7 years and some went back to jail, others did not. Finding: those who were married were less likely to end up in jail. Does this finding deserve causal statement? Can we say that getting married causes people to avoid going back to jail? It could be that marriage (values, etc) makes people stay out of trouble. Another explanation: staying out trouble makes it easier to meet people to marry. It could be that there is no causal relationship, maybe there is a third variable that explains the relationship. For instance, feeling regretful and responsible about criminal life makes people both stay out of trouble and marry. Other variables can account for this relation (personality, education, values). The point is not to say that the original claim is wrong, it's just to say that there is not enough information to make a causal claim. So in science, a lot of time is spent finding alternative causes, and isolating and contesting causes.

Conclusion for correlational research: it's always possible for a 3rd variable not considered by the researcher to account for the relationship between two other variables. Researcher needs to identify as many alternative explanations as possible and see which are viable and which are not. This is an ongoing process.

Experiments

As discussed above, correlation can never for sure establish causation because it's always possible that there is a 3rd variable that explains the correlation between two variables. Experiments, when it's possible to conduct them, are a much better method for identifying causation. Unfortunately, it's not always possible to do them.

For every experiment, there is one or more independent variable (IV). An IV is what the researcher expects to be the cause and what the researcher takes control of and manipulates. In correlational studies, the researcher just measures variables, nothing gets manipulated. There is also one or more dependent variable (DV), which is what the IV would affect. If we believed marriage to cause staying out of trouble, we might want to conduct an experiment in which marriage is IV, and staying out of trouble DV. The problem is that the researcher cannot decide whether or not people are allowed to marry, it's unethical. But if we were to suppose that this was possible, the researcher would manipulate the IV by manipulating who is married and who is not. This process is called random assignment; it is how researchers control IV. In this study, the researcher would randomly choose 50 people who are going to marry, and 50 others who are not. Choosing them at random allows for the 2 groups not to be different before the IV is introduced. Otherwise the outcome might be different, not because of marriage, but because of other factors. Every person has enough chance to be in either married or not married case (if there is a large enough sample) so that the IV that the researcher introduced is the only factor that can explain the difference in the groups.

Again, a researcher cannot make people marry or not marry. In the following cases, we can't conduct experiments:

- It's not possible to randomly assign people to being male or female

- We can't assign people to smoking for 40 years, it's unethical. In these cases, we do correlation research.
- Random assignment is not practical (eating oreos in childhood causes happiness, it doesn't make sense to get people to eat a bunch of oreos all of their lives)
- Essential realism: it's just not going to happen in real life, for example we can't make people fall in love in a research lab, and choose who does and who does not.

Example of an experiment: does anxiety affect test performance? If we were to conduct a correlational study, we would measure anxiety before the test, then we would observe performance. We would not be able to establish causality however: maybe anxious people are anxious because they did not prepare for the test. Not preparing for the test would both explain anxiety and poor performance. If we were to conduct an experiment: all people in a class would flip a coin. All people who get heads (50%) will take a surprise test that will count as 50% of their grade (this will create anxiety!), and people who get tails (50%) will just take the test for fun, it will not count. Then we measure test performance. We have manipulated what we think to be anxiety but some other researcher might not agree with how we have defined anxiety.

Error and bias

Error: random variability in effect. If for example, the experimenter has to estimate height by eye balling participants (maybe because researcher does not want Participants to know that their height is being measured), it is likely that the experimenter will not always be perfectly accurate, even with training, this would count as an error. An error means that the variable is not being measured exactly.

Bias: systematic, non random error that is systematically orienting the data in a certain direction. The researcher is not measuring what he or she thinks is being measured.

Imagine we were interested in the relationship between SAT and intelligence. If there was no error, the SAT and the score on some intelligence test would always match perfectly. But because there is likely to be error, sometimes, it will be off a little. If however we were to predict intelligence by measuring shoe size, we might be totally accurate in measuring shoe size, but we are completely not on target, because shoe size does not have anything to do with intelligence. The worse to have is bias because usually, random error is taken care of by averaging. In research, we are always averaging because we are not interested in one particular person but the average person in a group. Using shoe size to measure intelligence, regardless of how accurate measure is, is wrong.

Avoiding bias and error

There are few things researchers can do to avoid bias:

Reliability of scale: this can be improved by using a scale vs. using eyeballing

Validity : are we doing what we think we are doing? There are two types of validity:

- **internal validity** : does the design do what it's supposed to do? Can I draw the conclusions I want to draw from this design?
- **External validity:** does the finding that I have with 10 UVA students generalize to other UVA students? Other people in Charlottesville? Other sorts of designs? Does it extend beyond just the particular context of this particular study?

To improving external variability is not always easy. An ideal sample of humans would have somebody of each corner of the planet, but it's virtually impossible to have such a sample. Another solution can be to study only a specific small sample, like people with whom you share a dorm. The problem is that the findings only apply to a small group of people. Researchers can also try to select a **representative sample:** if I want to say something about US population, I can find out statistics in the US and try to get a sample with the same properties. If 12% of the US population were African American, I could have a sample with 12% of my sample being African American.

Internal validity is much more important than external validity: if your study has low internal validity, it does not matter whom you can generalize to. If on the other hand the design works, even if you can't generalize to the whole population, you have at least learned something about some group.Sometimes, nobody cares about generalization: in a study interested on the effects of smoking on rats, nobody cares whether or not we can generalize to the rat population, or whether we have a sample of rats from all parts of the world since ultimately, we want to learn about humans.

The **representativeness** is important, but what's a representative enough sample depends on the question the researcher is interested in: For example in our class, 90% of the students are right handed , this is the same as the US population, about 60% of us prefer dogs to cats, this is also representative of the US population, put a large part of the class is from Virginia and has a roommate, this is not at all representative of the US population (only 6.7% of people are from Virginia in the US). If the researcher is interested in right-handedness, UVA is representative, but for a researcher studying housing arrangements, UVA would not be a representative sample.

Class exercise: the purpose of this exercise is to use what we have learned today to identify problems with the following study : 60 heterosexual men aged 25-33, who lived in urban areas in 4 cities were interviewed about their views of marriage. Finding: men delaying marriage because of cohabitation, makes it easy to get sex, so there is no point in rushing to marry. There are several issues with this study: All the men in the sample are from the city, not married, from 4 cities only, are heterosexual are between 25-33 only. This creates a biased sample. Additionally, the US average marriage age for men is 27, so the men in the study are likely to be older than the average unmarried man; the sample might be of men who are already self selected to avoid marriage. The sample is small and there is no comparison group. This claims to be a study on gender and marriage, women have not even been studied. The evidence is not internally or externally valid.

Probabilistic causation

The relation between cause and effect in scientific research is never perfect, it's usually probabilistic. In order to say that something is a cause, we do statistical testing. Having **significance** in statistics means that there is less than 5% chance that something happens by chance. After researchers reach significance, they can say that there is a pretty good chance that a hypothesized relationship is true, in fact, there is a 95% chance of it being true, and a 5% chance of the finding being due to chance. The **effect size** is another statistical measure that informs us of how strong a relationship is. Significance and effect size are not the same. For example: smoking causes cancer. This may be a true causal relation (significant, below 5% chance of this relationship being due to chance) but the effect size is very small (a lot of people who smoke don't get cancer).

Improving correlational studies: cross lagged designs

The purpose of a cross lagged design is to increase temporal order confidence. Let's go back to the study on the relationship between TV violence and aggressive behavior and construct a cross lagged design to study this question. The researcher would first measure aggressive a sample of kids age 8. Then the researcher would rate how much violent TV the kids watch. When those kids are 30, the researcher would measure their aggressive behavior again. The following findings can be obtained: 1)aggressive behavior and TV violence are related/correlated) 2) there is consistency so that people who were aggressive at 8 were aggressive at 30 3) violent TV at age 8 predicted aggression at 30 better than aggressive behavior at age 8. Again, this is probabilistic, but informative with respect to the temporal order between those two variables. It's still not an experiment. The following experiment has been conducted to study the relationship between violent TV and aggressive behavior: a sample of Kids were randomly assigned to watch 2 movie clips, one violent (Karate kid) and one not. After the kids watched the clips, they were sent to another room where they watched a video of 2 kids engaging in hostile behavior towards one another. The kids thought that the actors in the video were really fighting in another room. Then the experimenter went outside, and told the subject to call him/her when something goes wrong in the video. The DV was the time it took the subjects to call the experimenter and report the violence on the video that was allegedly happening right at the moment. Finding: kids who had watched Karate Kid took twice as long to call the experimenter than the kids who did not. The experiment showed that watching violence on TV desensitized kids to real violence.

<u>Summary of today's lecture</u>

- correlation is necessary but not sufficient to identify causation
- experiments are better than correlation but we can't do them all the time
- causality in science is probabilistic not definitive