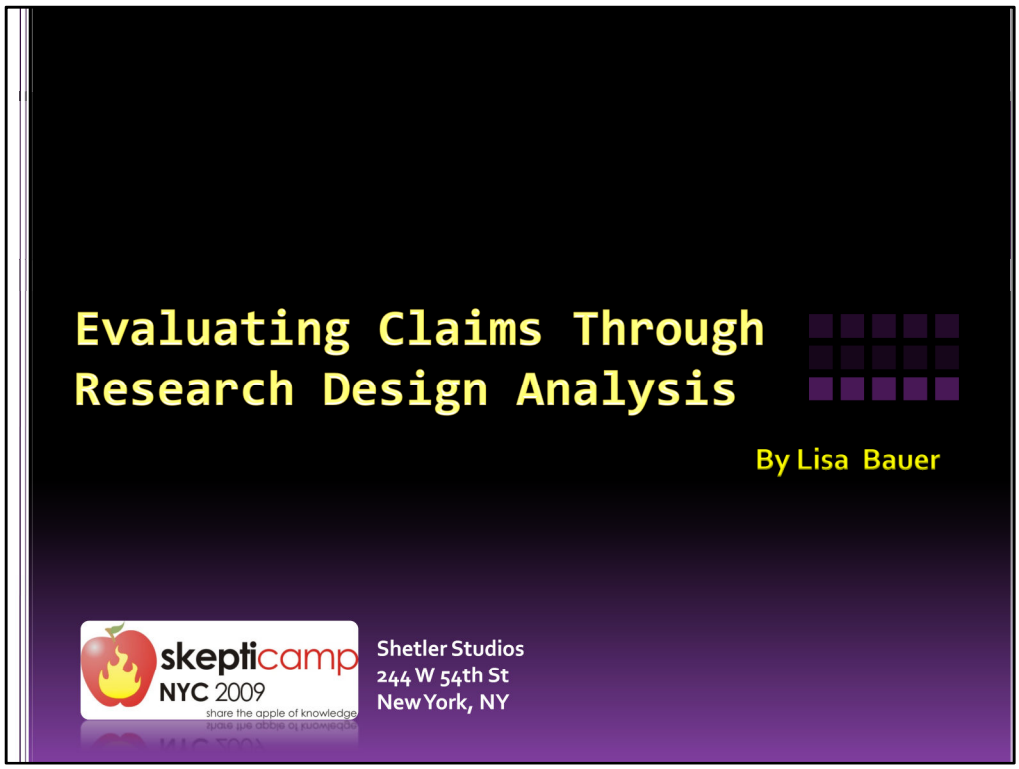
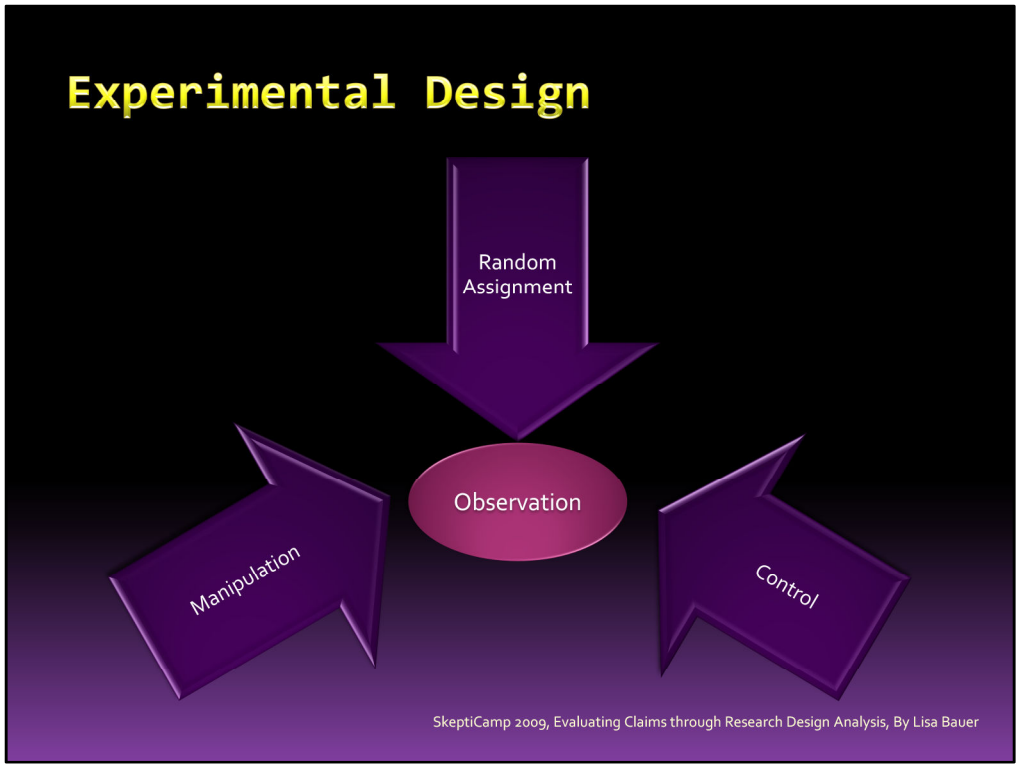


Many areas of skepticism advocate critical thinking skills by teaching/discussing logical fallacies, showing how one can spot flaws in reasoning/arguments and demonstrating something is wrong by exposing/debunking claims. This presentation will go in a somewhat different direction; we will review the elements of experimental design (including some terminology and concepts), and how the analysis of the procedures used in arriving at a finding can be an additional, effective means of detecting 'bunk' then just evaluating the stated conclusions. It is the aim of this presentation to help you develop or advance your skills in reading primary research reports (particularly the methods and procedures).

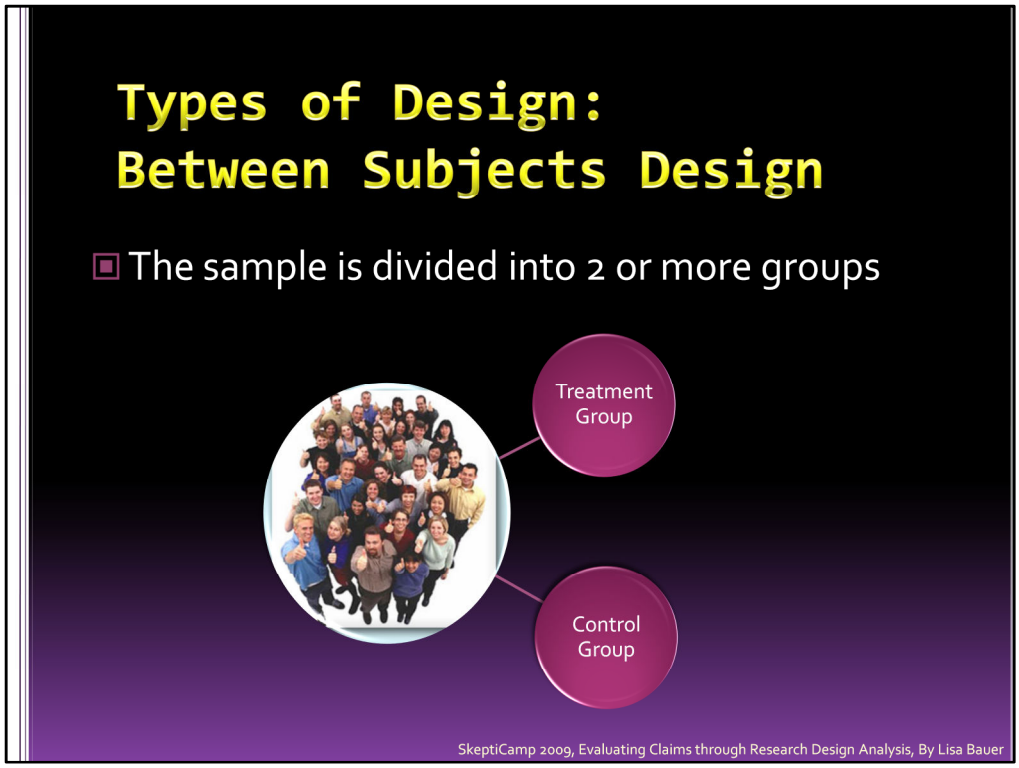


In order for a research study to be a true experiment it must have:

Manipulation: the independent variable must be introduced, meaning it cannot be a pre-existing subject variable, and there should be at least two levels of it (two experimental conditions, two levels could be presence/absence of something).

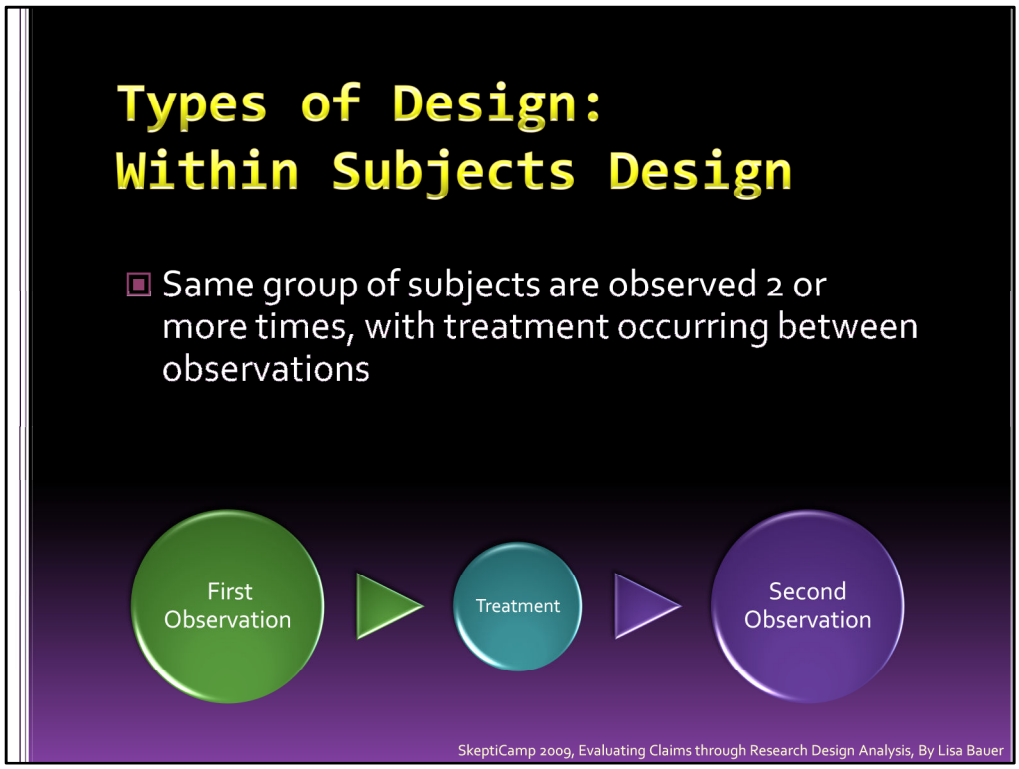
Control: variables other than the (manipulated) independent variable should be held constant. A controlled variable is not the same thing as a control group. A controlled variable is a variable that is the same across experimental conditions. For example, if I'm measuring whether or not the color a room is painted effects the productivity of persons working in the room, the color of the room would be the independent variable, their level of productivity would be the dependent variable and a control variable might be the task they are working on (they would all be performing the same task). Holding a variable constant minimizes the possibility of it becoming a confound.

Random Assignment: often, if a true random method is used to determine group assignment, variables such as age, ethnicity and sex will almost naturally be relatively similar across experimental conditions. If anything other than a truly random process determines group assignment, you do not have a true experiment.



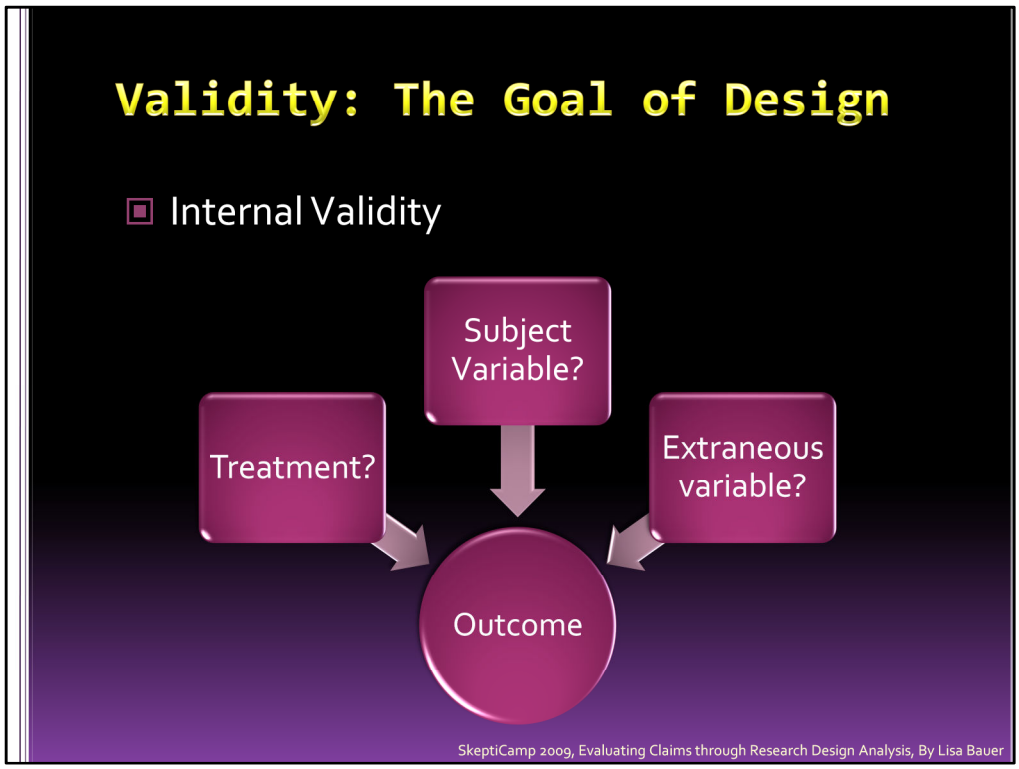
Advantages of between subjects design include a minimization of practice effects. Practice effects are when subjects improve on several observations over time due to having had practice at the test, and the improvement can not be unambiguously attributed to the treatment condition. There are other problems discussed later.

Disadvantages of a between subjects design is that individual variation may play a part in any between group differences that you might observe. With a within subjects design, individual variation is controlled for in each observation as the same subjects are producing the same scores.



Disadvantages of within subjects design are practice effects, or attrition (when subjects participate in the first observation but not the second).

An advantage of within subjects design is that subjects act as their own control group, establishing their own baseline measure and variation.

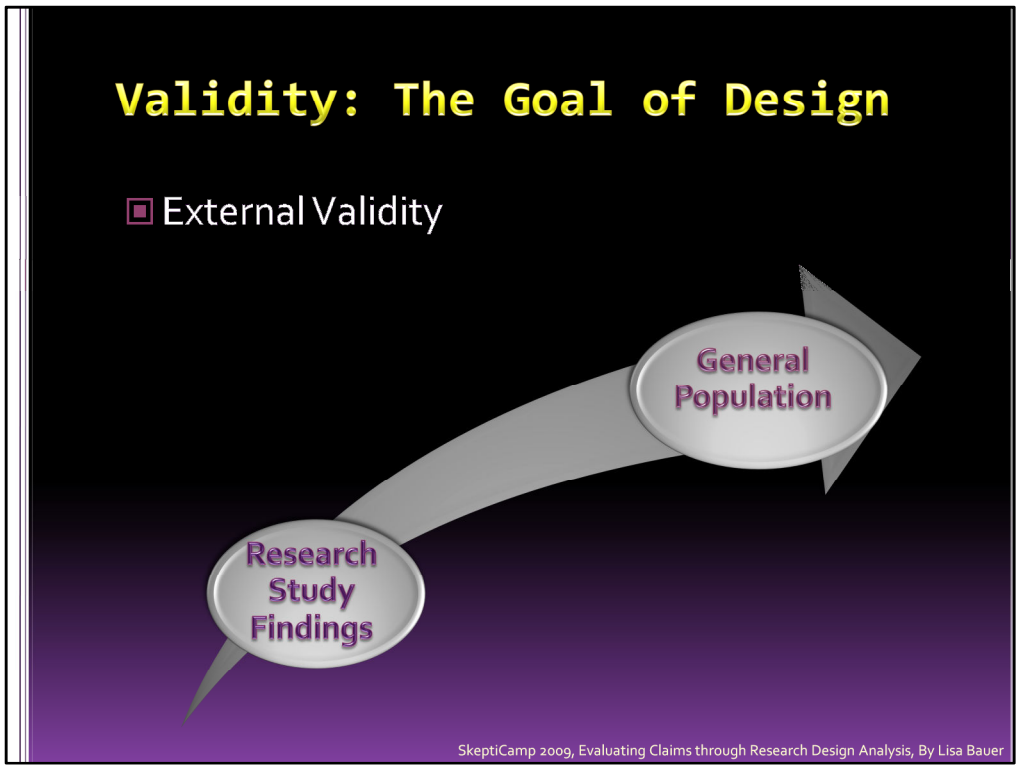


Validity refers to the degree to which a testing procedures actually measures the construct it is attempting to measure. To put it another way, how well the researcher has operationalized the construct being investigated. It is typical to perform validity tests for a measure before it is adopted. However, there is a difference between a measure proving to be consistent, and a measure that is valid. For example, if I'm measuring intelligence, and I do so be measuring the diameter of a person's head, I would probably produce consistent measurements over time. This, however, does not allow me to conclude that this is actually measuring 'intelligence'.

Internal validity refers more specifically to the procedures of a research design, and whether or not they prevent alternative explanations of the observed outcome from being plausible. For a procedure to have high internal validity, there must be a single, unambiguous explanation for the observed outcome.

There are factors that can threaten internal validity:

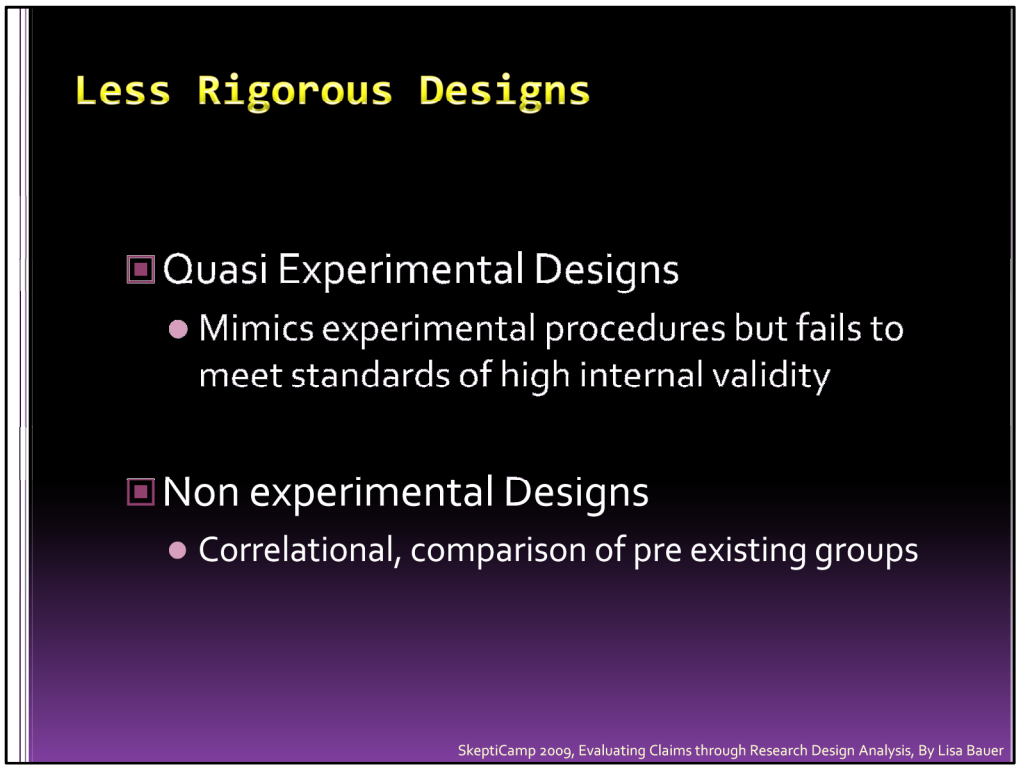
Environmental variables: if conditions are different within or between the observations
Assignment Bias: If participants in one condition differ systemically in their demographics
History: extraneous events outside of the parameters of the study that occur between observations
Maturation: participants 'mature' or change over time naturally, between the observations
Instrumentation: degradation to the instrument over time can lead to flaws in the measurements
Testing Effects: the experience of the observation effects behavior on future observations
Regression towards the mean: extreme scores have a tendency to fall closer to the mean on repeat trials



External validity refers to the degree to which results from a study can generalize to the general population.

Threats to external validity include:
Subject variables: characteristics of the sample are unique to the sample population (ex: a lot of research is conducted using college freshman)
Selection bias: sampling procedure favors one population more than others
Volunteer bias: participants who volunteer may already have dissimilar characteristics (subject variables) from the general population
Experimenter characteristics: differences in the experimenters collecting the data (specifically how they collect/interpret it) can affect the outcomes of the study
Novelty effect: the experiment environment differs from the natural environment, which may create difficulty generalizing to naturalistic conditions

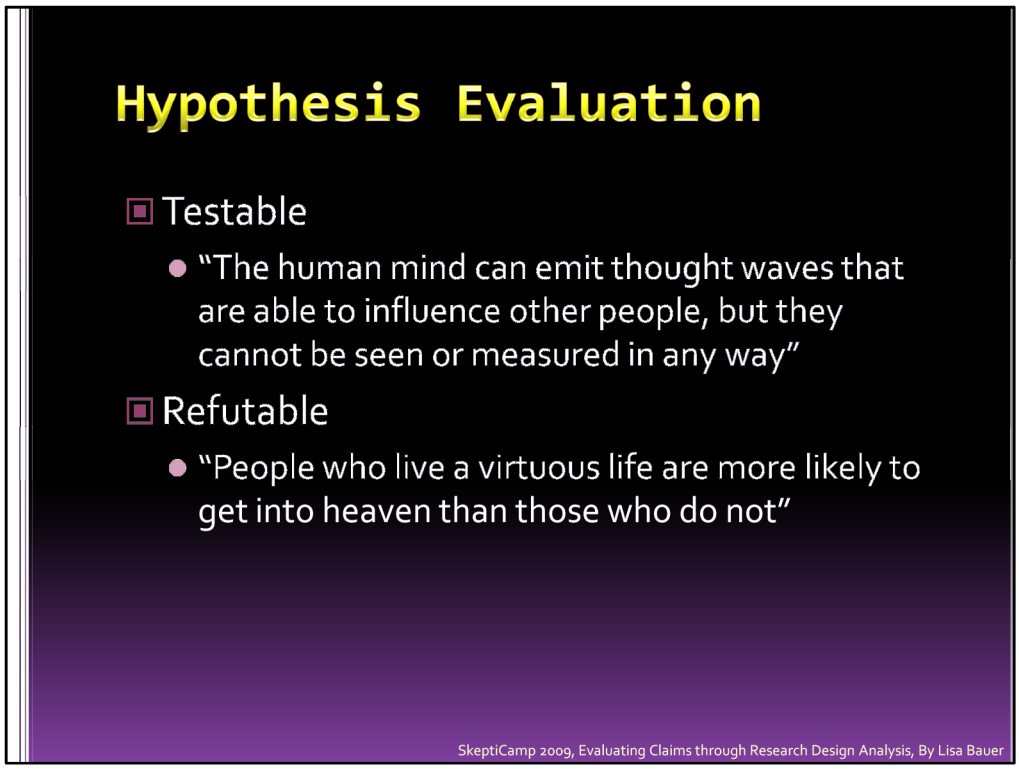
Some threats affect both internal and external validity:
Experimenter Bias: the experimenter believes the outcome will occur a certain way, and subtly influences the results towards achieving that end
Demand Characteristics: the subjects may suspect the hypothesis being investigated, and behave accordingly



A true experiment is essentially testing for a cause/effect interaction. As such, the design would need to prevent there from being an alternative explanation for the findings. A quasi experimental design differs from a true experiment in this way; it holds as much constant as would be allowed by the testing procedures, but is unable to produce a clear, cause/effect relationship. For example, say I wanted to conduct a study to determine if persons' who suffered childhood trauma, but repressed the memory, would have a therapeutic benefit in experiencing recovery of that memory. I recruit subjects who are experiencing depression, that I believe is attributed to repressed memories of trauma, and I begin undertaking strategies to bring such memories to a conscious level. Now , say that 60% of the subjects state that they now remember traumatic events, and report decreases in their depression that significantly differs from the 40% that did not recover memories of abuse.

Now, there are several things wrong with this study, but they all follow the same theme: the procedures allow for too many alternative explanations as to the attribute of the results.

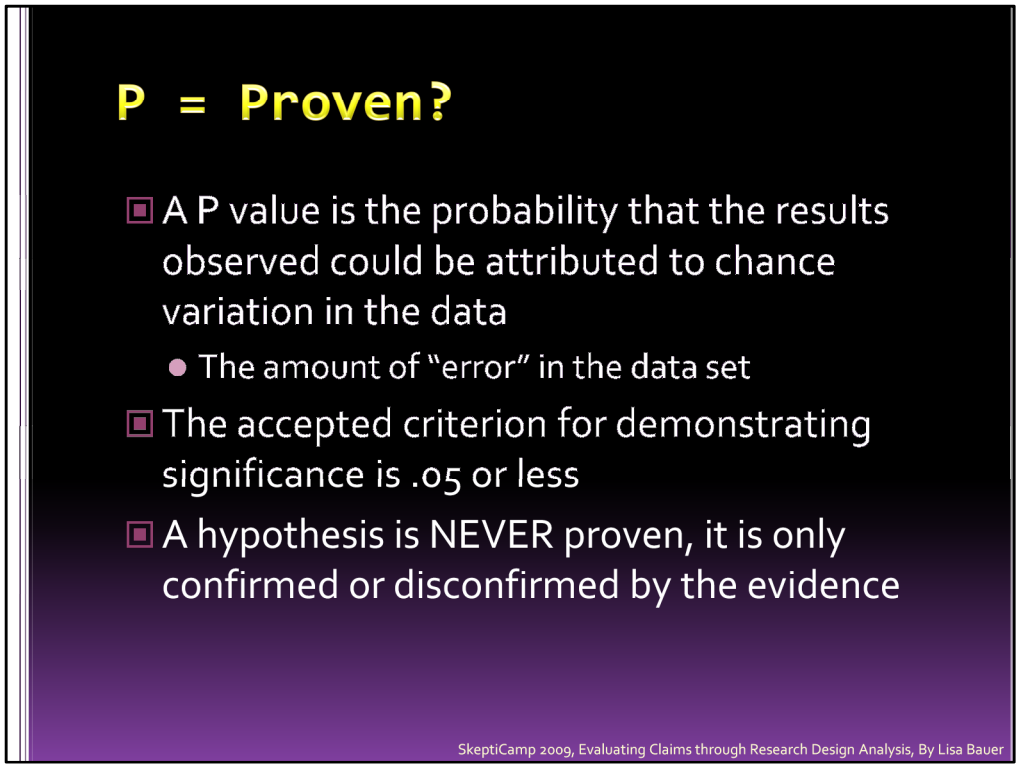
Non experimental designs (descriptive studies, correlational studies) fail to meet the high standards of true experimentation as well. These procedures examine differences in pre existing groups, and in some design structures there is no manipulation of the variable under investigation. The adage "Correlation does not equal causality" is the most recognizable problem with any non-experimental classification. These studies can describe differences, and relationships, but cannot explain them.



A hypothesis must be relevant to the design and procedures of a study, capable of being measured and refutable. For example, if your hypothesis is that a vitamin will improve your memory, you cannot test this hypothesis by administering vitamins and then asking people to name as many U.S. presidents as they can. The procedures do not fit the hypothesis. More so, the construct under investigation must be observable and testable. Now, not everything studied is directly observable. Constructs like "motivation", "happiness", "fear" are abstract concepts and are not external and obvious like height, weight, speed, etc. This is why we develop operational definitions and translate these abstract concepts into behaviors and measures we can observe.

A hypothesis must also be refutable, meaning that for it to be capable of being confirmed through testing procedures, there must also be a way to disconfirm it. When evaluating a research hypothesis, the actual statistical procedures measure the degree to which the observations made deviate from the null hypothesis. The null hypothesis is the absence of an effect, difference, etc. For example, if my research hypothesis (alluded to above) is that vitamins would improve memory (and let's assume I've developed better procedures to test this), the null hypothesis would be that there would be no benefit in memory performance from taking vitamins. If the observations made deviate from this assumption to a significant degree, the null hypothesis is rejected.

We never prove a hypothesis, we are only able to demonstrate that the absence of a treatment effect is not supported by the data observed. As such, for a hypothesis to be truly testable, there must first be a way one could demonstrate that the it could be refuted.



Whenever you are measuring something, there is always the possibility of error. In this respect, error does not refer to mistakes made in the collection/analysis of the data, but to the random fluctuations that might occur in the data due to individual/group differences that may not be attributable to the treatment conditions.

For example, say I hypothesize that skeptics have more critical thinking skills than non-skeptics. I have a standard and valid measure of critical thinking skills, I administer it to everyone attending skeptic camp, and then another group (of equal size, matched for education, demographics, etc), and measure if there is a significant difference between the average score of the groups. To do this, I could not simply examine the means, I would need to examine the within and between group variation. For instance, if the skeptics had a mean of 94 out of 100, and the non skeptics had a mean of 86, could I conclude that skeptics have more critical thinking skills than non skeptics? No, I could not, at least not with that information alone. If the variation amongst the skeptics were equal to 17 and the variations amongst the non skeptics were 14, we would have overlapping estimated population parameters. It could be the case that moving 2 or 3 scores from one group to the other would result in no difference in means, or the inverted difference.

To establish significance, we must be able to say that the difference observed is large and consistent enough that to say that it is due to chance (that the difference can be explained by subject variables and individual variation) is refuted The .05 criterion means that a less than a 5% probability is sought in establishing significance.

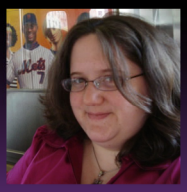
Additional Information

Glossary

- **Construct**: the abstract object/idea under investigation (i.e. "motivation", "poverty", etc.). This is not directly observable
- **Variable**: an element of a study (i.e. "response time", "score")
- **Operational definition**: a description of the construct that identifies how it will be measured, and what variables will be observed
- It is a goal of the procedures of a study to 'operationalize' the construct
- **Independent variable**: a manipulated variable (manipulated by nature or the researcher)
- **Dependent variable**: the outcome measure, the variable you expect will differ in some consistent way along with the independent variable
- **Confound**: when an extraneous variable has a systemic effect on the dependent measure, but is not the object under investigation
- **Validity**: The degree to which your instrument is actually measuring what it claims to, allowing you to draw meaningful conclusions.

Resources

- Free online text book: Trochim, William M. The Research Methods Knowledge Base, 2nd Edition. Internet WWW page, at URL: <http://www.socialresearchmethods.net/kb/> (version current as of October 20, 2006).
- Great book: Meltzoff, Julian (1997) Critical Thinking About Research http://books.apa.org/books.cfm?id=4318640
- Great book regarding statistics: Field, Andy (2009) Discovering Statistics using SPSS 3rd Edition http://www.statisticshell.com/dsus.html



Lisa Bauer is a statistics specialist and instructor in research procedures and experimental psychology . She has been a member of NYCS since 2007, and writes for The Gotham Skeptic, the official blog of NYCS http://www.nyskeptics.org/blog/

Lisambauer@gmail.com