

Cheat Sheet:

Statistical Studies

Essential Concepts

- Experimental design refers to the structure of an experiment (a specific type of research method). The primary goal of an experiment is to provide evidence for a cause and effect relationship between two variables.
- In an observational study, a researcher will observe an outcome without changing who is and who is not exposed to some sort of treatment. Observational studies cannot be used as evidence for a cause and effect relationship.
- A confounding variable is a variable that was not accounted for in a study and may actually influence other variables in a study.
- In an experimental study, the factor of interest (also known as the explanatory variable or independent variable) is the factor that the researcher purposely changes or manipulates to see if it impacts a specific outcome.
- The treatments are the different levels of the factor of interest (or explanatory variable) you are changing. The group that receives the treatment is commonly called the experimental group. The group that does not receive the treatment, or receives a placebo as their treatment, is usually called the control group.
- The response variable (also known as the dependent variable or response factor) is an objective measure of the research question that is measured at the end of an experiment and compared across the different levels of the factor of interest (or explanatory variable).
- The nuisance factors are factors that are kept the same across all levels of the factor or are explicitly controlled in the experimental design. These factors are not of interest in the study but may effect a change in the response variable.
- To ensure that there are no unexpected differences between the experimental and control groups, good experimental design uses three key components:

- Randomization (or random assignment) to determine which participants are in each group.
 - Replication to ensure that the results of an experiment are truly caused by the change in the factor of interest and not by other hidden factors or natural variation in data.
 - Control group to determine whether any observed changes are due to the treatment, rather than external factors
- The single object or individual to be measured in the experiment is called an experimental or observational unit.
 - A placebo is a harmless version of the treatment that does not contain any active ingredients (e.g., a sugar pill). The placebo effect is a positive response that people who believe they are receiving treatment for a condition have, even if what they are actually receiving is a placebo.
 - Blinding refers to nondisclosure of the treatment an experimental unit is receiving. When neither the subject nor those having contact with the subject know the treatment assignment, the study is called double-blind.
 - A completely randomized block design occurs when experimental units are divided into homogeneous (same characteristic) groups called blocks, and then treatments are randomly assigned within each block. This design is particularly useful when there are known sources of variation that we want to control for.
 - When analyzing statistical visualizations it's important to apply critical thinking. Most data visualization will include a legend, or key, that explains the colors and symbols utilized in the visual.
 - Heat maps are particularly useful for showing relationships between categorical variables or displaying patterns in large datasets. They use color intensity to represent data values, with darker or more intense colors typically representing higher values.
 - Misleading graphs can unintentionally, or intentionally, misrepresent the data it is meant to display. A good graph will always include clear axes, data lines, bars or shapes, and a legend.
 - The criteria for effective visualizations include design, precision, efficiency, and use of time/ink/space. Some visualizations will represent multiple variables (multivariate).
-

Glossary

blinding

a randomized experiment preserves the power of suggestion

block

a group of subjects that are similar

blocking

the grouping together of homogeneous (similar) experimental units followed by the random assignment of the experimental units within each group to a treatment

completely randomized block design

when the experimental units are divided into homogeneous groups called blocks

confounding variable

a variable that was not accounted for in a study and may actually influence other variables in a study

control group

the group that does not receive the treatment of interest or the placebo

double-blind experiment

both the subjects and the researchers involved with the subjects are blinded

experiment

provide evidence for a cause-and-effect relationship between two variables

experimental design

the structure of an experiment

experimental group

one group that receives the treatment of interest

experimental unit

the single object or individual to be measured in the experiment

factor of interest, explanatory variable

the factor that the researcher purposely changes or manipulates to see if it impacts a specific outcome

heat map

a representation of data in the form of a map or diagram where different values are grouped into different colors

legend

the description, explanation, or table of symbols provided with a map or visualization that helps readers interpret the map

nuisance factors

factors that are kept the same across all levels of the factor or are explicitly controlled in the experimental design

observational study

to learn about characteristics of a population or to compare groups within a population with respect to some characteristic

placebo

a harmless version of the treatment that does not contain any active ingredients that appears identical to the treatment

placebo effect

a positive response that people who believe they are receiving treatment for a condition have, even if what they are actually receiving is a placebo

population

a group of individuals or entities that the research question pertains to

random assignment

random assignment of which participants are in each group in a study

replication

including large numbers of participants, or repeating the entire experiment with new groups of participants

response factor, response variable, dependent variable

an objective measure of the research question that is measured at the end of an experiment and compared across the different levels of the factor of interest

sample

a group of individuals or entities on which we collect data

treatments

different levels of the factor of interest