First look at the Data

Population	entire group that you want to draw conclusions about.
Sample	he specific group that you will collect data from. The size of the sample is always less than the total size of the population
Mean	average (μ mean of population; x̄ mean of sample)
Median	separates the sample (Mittelpunkt)
Mode	highest score
Variance	measures dispersion around the mean
Standart Deviation (SD)	estimates the SD of the sampling distribution
	FORMULA
Standard Error	Square root of the variance (σ SD of population; s SD of sample)
	s/√n
Confidence Intervalls (CI)	This is the range of values you expect your estimate to fall between if you redo your test, within a certain level of confidence. Confidence, in statistics, is another way to describe probability
Quantitative data	s expressed in numbers and graphs and is analyzed through statistical methods.
Qualitative data	is expressed in words and analyzed through interpretations and categorizations.
Hypothesis Testing	

First look at the Data (cont)

H0	the null hypothesis of a test always predicts no effect or no relationship between variables
H1	alternative hypothesis states your research prediction of an effect or relationship
Randomisation	
completely randomized design	every subject is assigned to a treatment group at random.
	Ex. Subjects are all randomly assigned a level of phone use using a random number generator.
randomized block design	subjects are first grouped according to a characteristic they share, and then randomly assigned to treatments within those groups
	Ex. Subjects are first grouped by age, and then phone use treatments are randomly assigned within these groups.
Between-subjects vs. within-subjects	
between-subjects design	AKA independent measures design or classic ANOVA design
	individuals receive only one of the possible levels of an experimental treatment.
	EX. Subjects are randomly assigned a level of phone use (none, low, or high) and follow that level of phone use throughout the experiment.
within-subjects design	AKA repeated measures design

First look at the Data (cont)

every individual receives each of the experimental treatments consecutively, and their responses to each treatment are measured.

EX. Subjects are assigned consecutively to zero, low, and high levels of phone use throughout the experiment, and the order in which they follow these treatments is randomized.

Different Scales of Measurement

Nominal Categories	do not correspond to numerical value
	Ex. British Team, German Team, ...
Ordinal Measurement or Ranks	scores can be ordered from smallest to largest, only a rank order is implied
	Ex. 1st, 2nd, 3rd, ...
Interval Measurement	size of the difference between scores is an indication of magnitude
	Ex. Bill was 5 seconds behind the winner, ... (equal interval scale of measurement - interval of 1 second)
Ratio Measurement	like Interval Measurement, but allows ratios to be meaningfully calculated between scores
	Ex. Tom took 50 seconds and Bill took 100 seconds -> Tom is twice as fast as Bill

Types of Variables

Dependent Variable	Variables that represent the outcome of the experiment.
	Ex. Any measurement of plant health and growth: in this case, plant height and wilting.
Independent Variable	Variables you manipulate in order to affect the outcome of an experiment
	Ex. The amount of salt added to each plant’s water.
Controlled Variable	Variables that are held constant throughout the experiment.
	Ex. The temperature and light in the room the plants are kept in, and the volume of water given to each plant.
Confounding Variable	A variable that hides the true effect of another variable in your experiment. This can happen when another variable is closely related to a variable you are interested in, but you haven’t controlled it in your experiment.

Types of Variables (cont)

	Ex. Pot size and soil type might affect plant survival as much or more than salt additions. In an experiment you would control these potential confounders by holding them constant.
Latent variables	A variable that can’t be directly measured, but that you represent via a proxy.
	Ex. Salt tolerance in plants cannot be measured directly, but can be inferred from measurements of plant health in our salt-addition experiment.
Composite variables	A variable that is made by combining multiple variables in an experiment. These variables are created when you analyze data, not when you measure it.
	Ex. The three plant health variables could be combined into a single plant-health score to make it easier to present your findings.
Quantitative Variables	
Discrete/ integer variables	Counts of individual items or values.

Types of Variables (cont)

	Ex. Number of students in a class; Number of different tree species in a forest
Continuous variables (aka ratio variables)	Measurements of continuous or non-finite values.
	Ex. Distance, Volume, Age
Categorial Variables	
Binary/dichotomous variables	Yes/no outcomes
Nominal variables	Groups with no rank or order between them.
	Ex. Species, Names, Colors, Brands
Ordinal variables	Groups that are ranked in a specific order.
	Ex. Finishing place in a race, Rating scale responses in a survey

Sampling

Probability sampling methods

Probability sampling means that every member of the population has a chance of being selected. It is mainly used in quantitative research. If you want to produce results that are representative of the whole population, probability sampling techniques are the most valid choice.

Simple random sampling	every member of the population has an equal chance of being selected. Your sampling frame should include the whole population.

Sampling (cont)

Systematic sampling	is similar to simple random sampling, but it is usually slightly easier to conduct. Every member of the population is listed with a number, but instead of randomly generating numbers, individuals are chosen at regular intervals.
Stratified sampling	involves dividing the population into subpopulations that may differ in important ways. It allows you draw more precise conclusions by ensuring that every subgroup is properly represented in the sample. To use this sampling method, you divide the population into subgroups (called strata) based on the relevant characteristic (e.g. gender, age range, income bracket, job role).
Cluster sampling	also involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole sample. Instead of sampling individuals from each subgroup, you randomly select entire subgroups

Non-probability sampling methods

Sampling (cont)

In a non-probability sample, individuals are selected based on non-random criteria, and not every individual has a chance of being included. This type of sample is easier and cheaper to access, but it has a higher risk of sampling bias. That means the inferences you can make about the population are weaker than with probability samples, and your conclusions may be more limited. If you use a non-probability sample, you should still aim to make it as representative of the population as possible. Non-probability sampling techniques are often used in exploratory and qualitative research. In these types of research, the aim is not to test a hypothesis about a broad population, but to develop an initial understanding of a small or under-researched population.

Convenience sampling	A convenience sample simply includes the individuals who happen to be most accessible to the researcher. This is an easy and inexpensive way to gather initial data, but there is no way to tell if the sample is representative of the population, so it can’t produce generalizable results.

Sampling (cont)

Voluntary response sampling	Similar to a convenience sample, a voluntary response sample is mainly based on ease of access. Instead of the researcher choosing participants and directly contacting them, people volunteer themselves (e.g. by responding to a public online survey). Voluntary response samples are always at least somewhat biased, as some people will inherently be more likely to volunteer than others.
Purposive sampling	This type of sampling, also known as judgement sampling, involves the researcher using their expertise to select a sample that is most useful to the purposes of the research. It is often used in qualitative research, where the researcher wants to gain detailed knowledge about a specific phenomenon rather than make statistical inferences, or where the population is very small and specific. An effective purposive sample must have clear criteria and rationale for inclusion.

Sampling (cont)

Snowball sampling	If the population is hard to access, snowball sampling can be used to recruit participants via other participants. The number of people you have access to “snowballs” as you get in contact with more people.

Data Cleansing

Data cleansing involves spotting and resolving potential data inconsistencies or errors to improve your data quality.

Type I vs Type II error

Type I error (false positive)

Type II error (false negative)