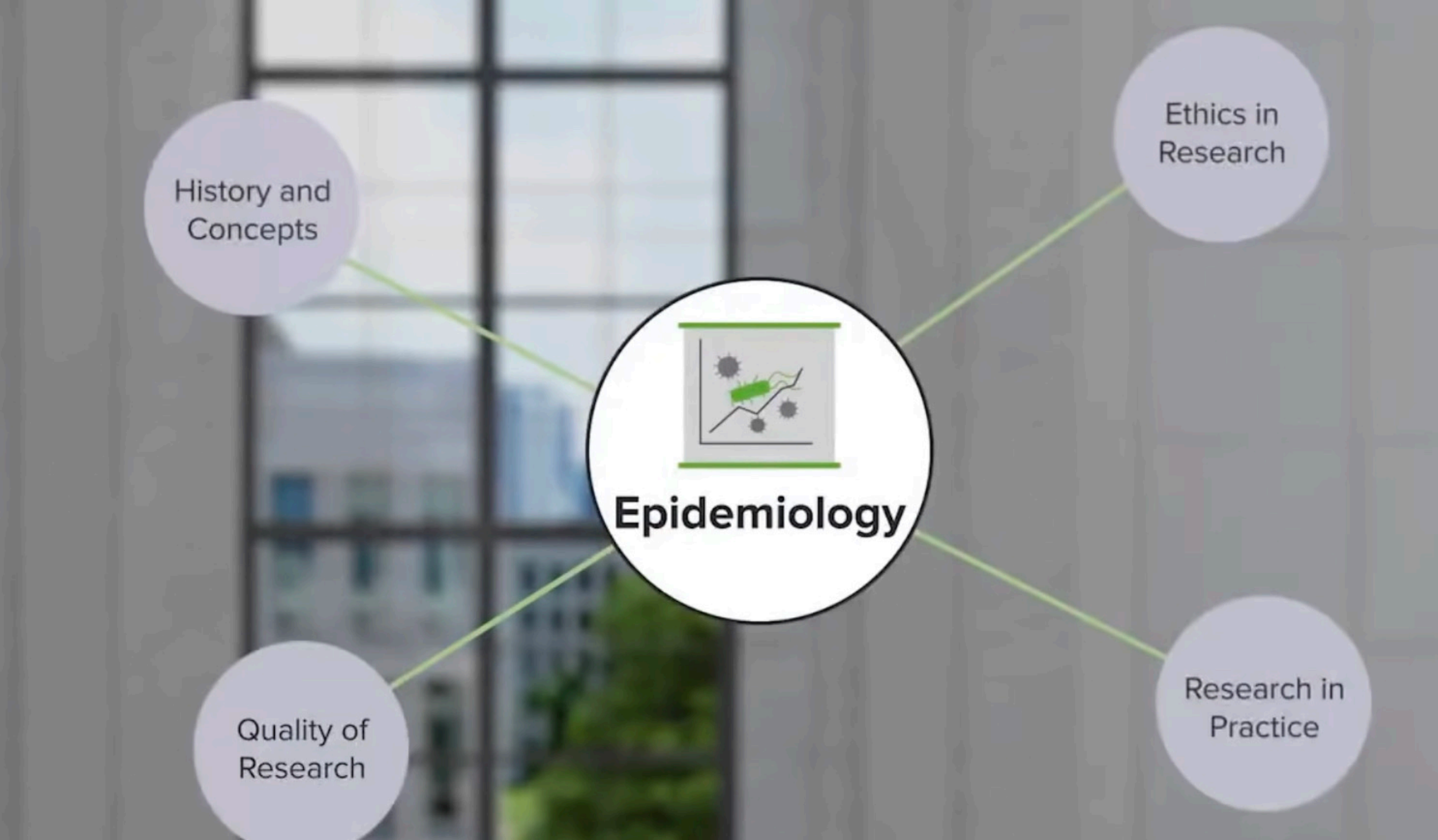
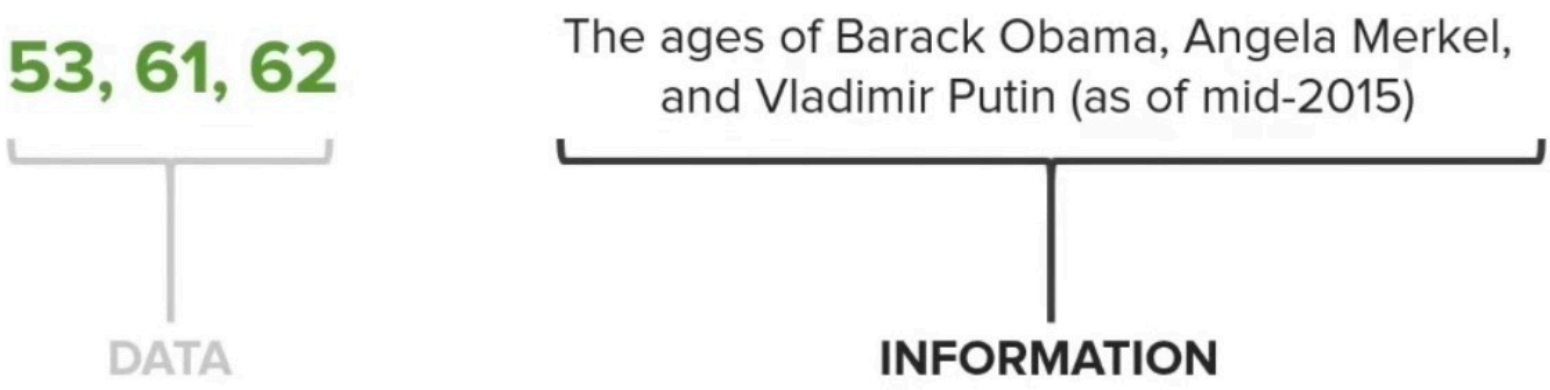

History and Concepts
Ethics in Research
Epidemiology
Quality of Research
Research in Practice

Learning Goals

- You will understand the meaning of p-values, confidence intervals, and the null hypothesis.
- You will know the difference between data and information.
- You will know the differences between different types of variables.

Epidemiology and Statistics

There is a close relationship between the methods of epidemiology and the methods and philosophies of statistics.


53, 61, 62
DATA
The ages of Barack Obama, Angela Merkel, and Vladimir Putin (as of mid-2015)
INFORMATION

Epidemiology and Statistics

What Is a “Variable”?

Math — A value that may change within the scope of a problem or situation (vs. a “constant”).

Research — A logical set of attributes (gender, age, etc.).

Computers — A symbolic name given to an unknown quantity.

Epidemiology and Statistics

What Is a “Variable”?

Math	x
Research	Age
Computers	A$

Epidemiology and Statistics

Relationships Between Variables

In math we write the relationship between 2 variables as a "function".

e.g. $F(x) = 210 - x$

Epidemiology and Statistics

Relationships Between Variables

In math we write the relationship between 2 variables as a "function"

e.g. $F(x) = 210 - x$

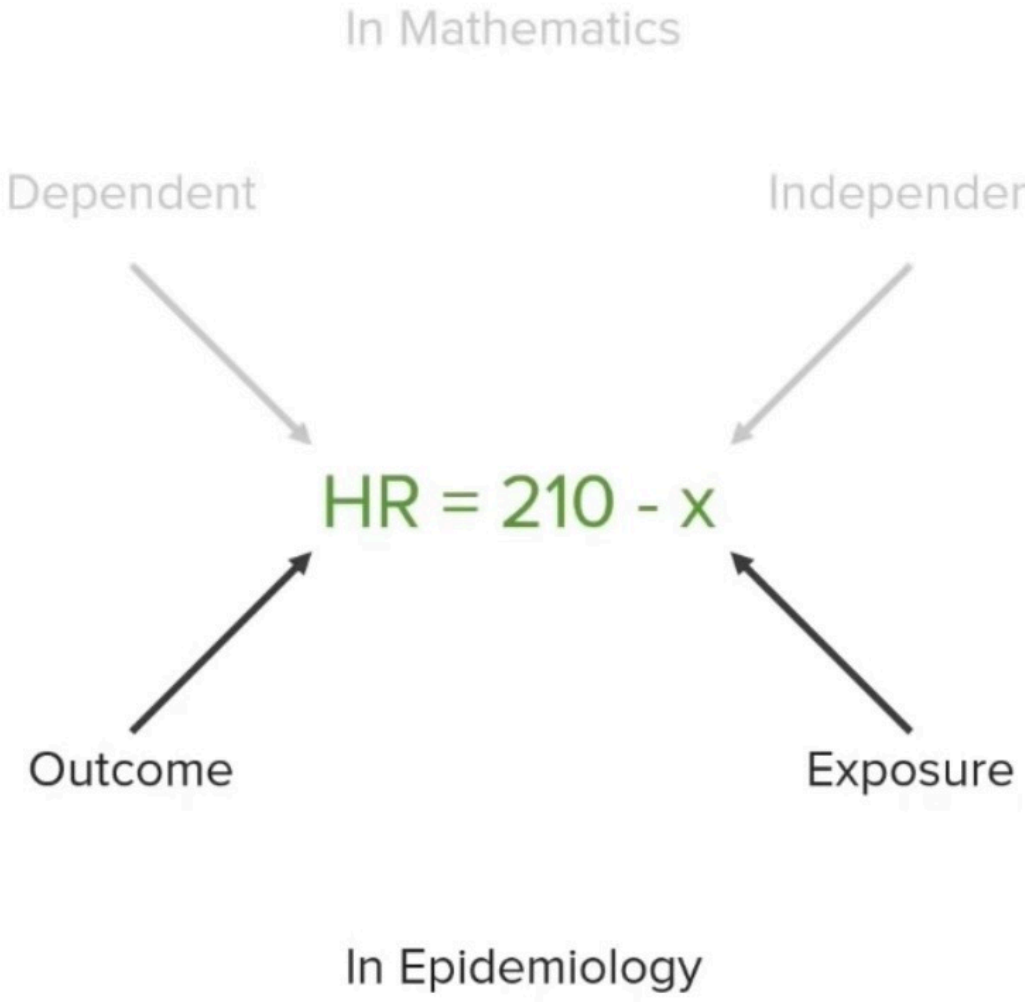
Maybe this is the relationship between age and maximum attainable heart rate.

$F(x)$ = max heart rate = HR

x = age

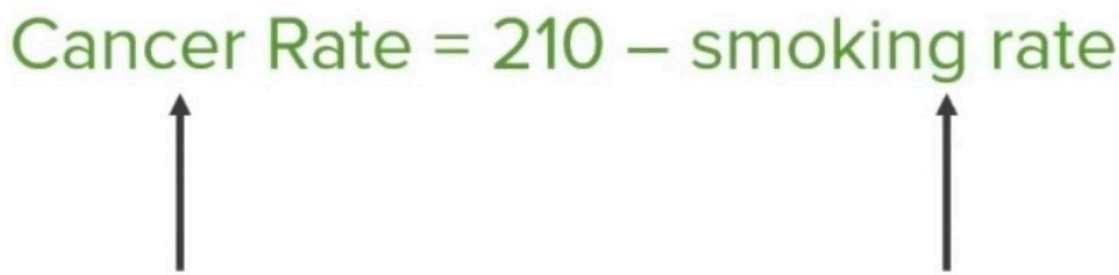
Epidemiology and Statistics

Relationships Between Variables



Epidemiology and Statistics

Relationships Between Variables



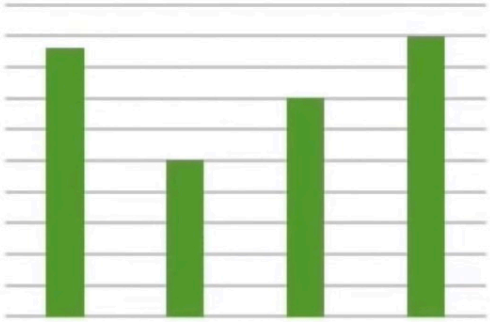
In Epidemiology

Types of Variables

There Are Two Broad Flavours of Variables

Continuous

Age, height, distance, temperature...

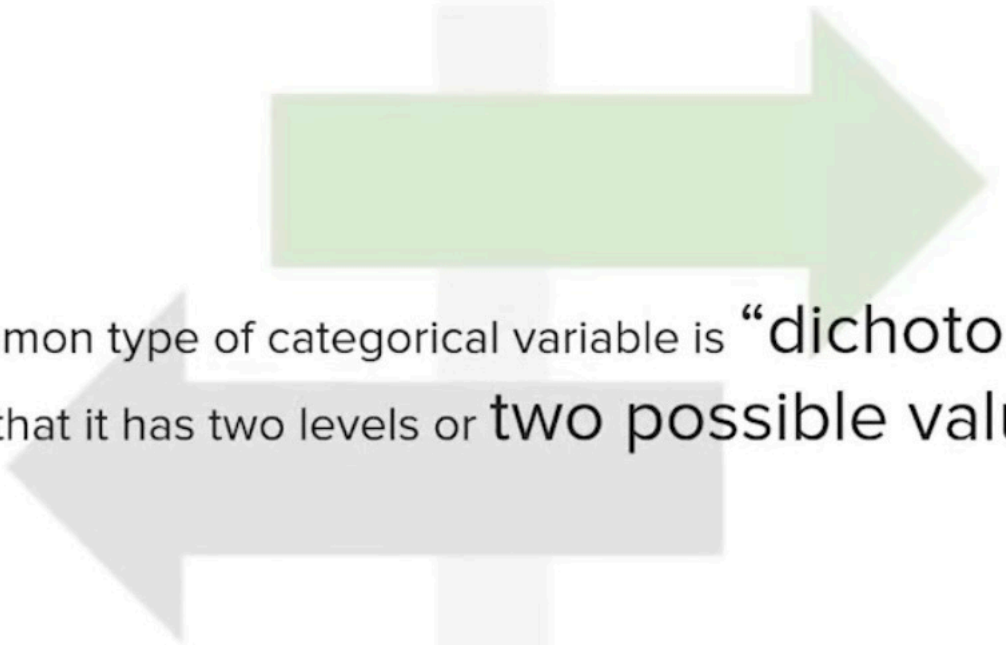


Categorical

(also called “Discrete”)

Age group, gender, number of siblings, citizenship, race...

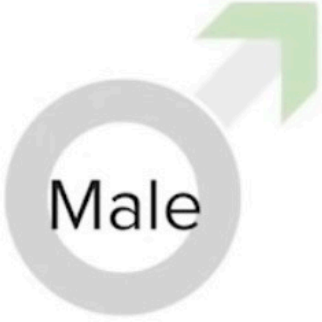
Dichotomous Variables



The most common type of categorical variable is "dichotomous", meaning that it has two levels or two possible values.

Dichotomous Variables

Gender



Male

VS.

Female

Dichotomous Variables

Employment status

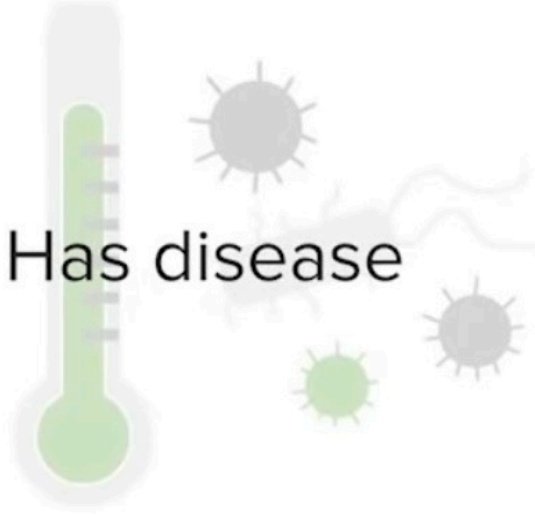
Employed

VS.

Unemployed

Dichotomous Variables

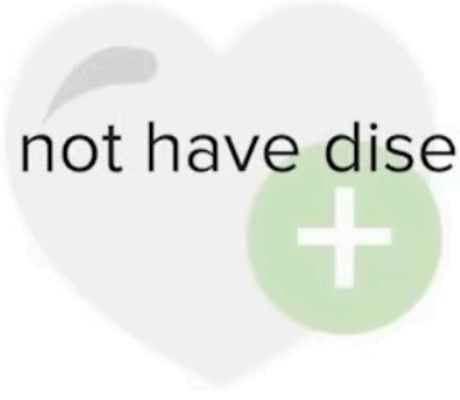
Disease status



Has disease

VS.

Does not have disease



Dichotomous Variables

Court outcome

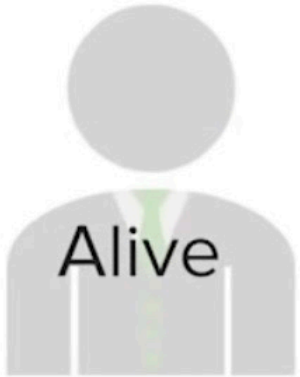
Guilty

VS.

Innocent

Dichotomous Variables

Emergency therapy outcome



Alive

VS.

Deceased

Dichotomous Variables

"Dichotomize" means to convert
a non-dichotomous variable to a dichotomous one.

Age

23, 17, 14, 35, 68, 15

Age group

< 18 vs. ≥ 18

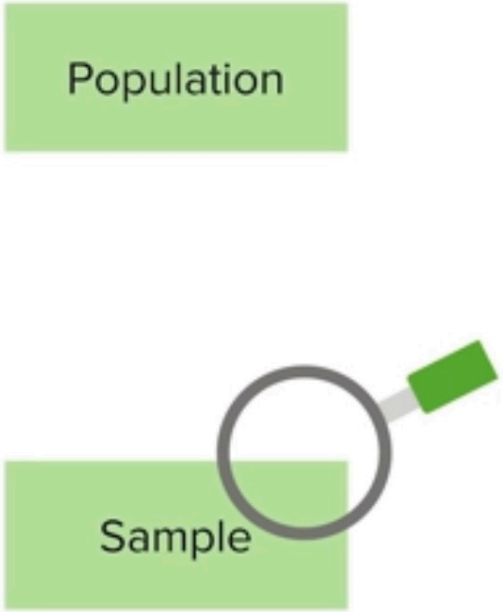
14, 15, 17 | 23, 35, 68

Categorizing Continuous Variables

We can also create categorical variables with more levels.

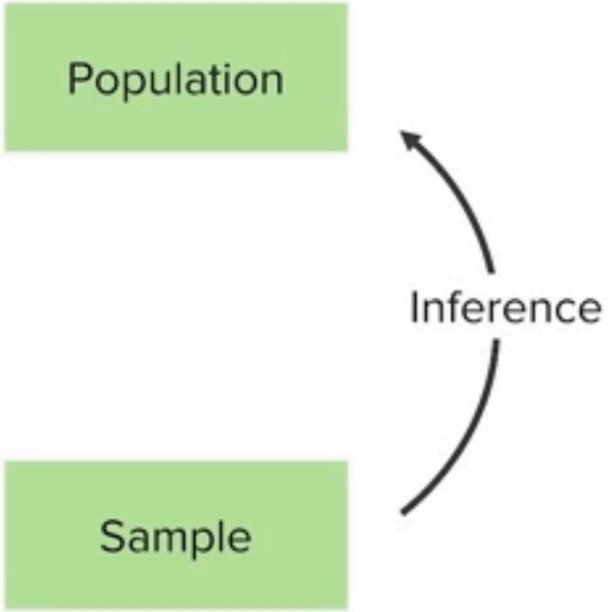
Age		Age group
23, 17, 14, 35, 68, 15	→	< 25 26-50 > 50

Sampling



When doing population research, we often need to perform analyses upon a sample of the larger population.

Sampling



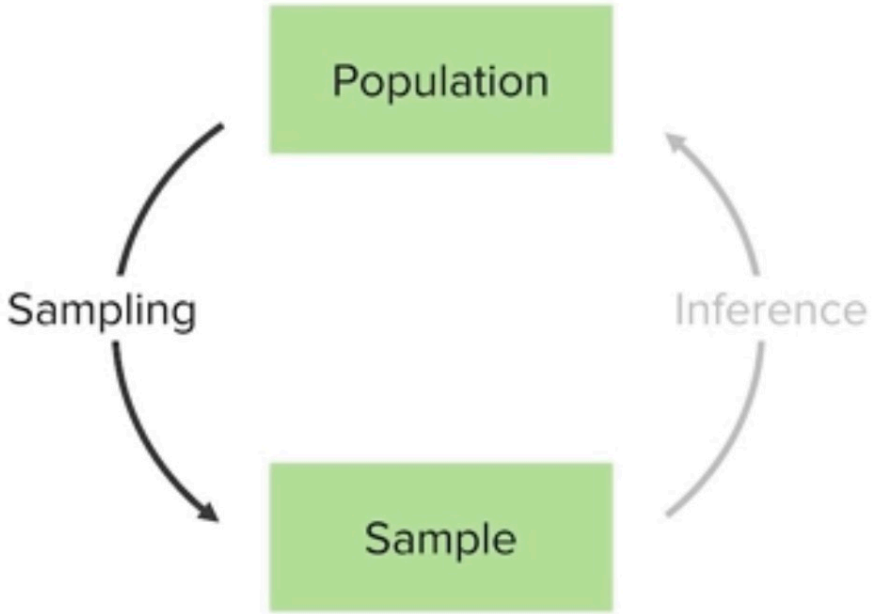
Dichotomous variables

Descriptive

Inferential

We perform statistical analyses upon a sample, from which we infer characteristics of the larger population, which is also called the “reference population”.

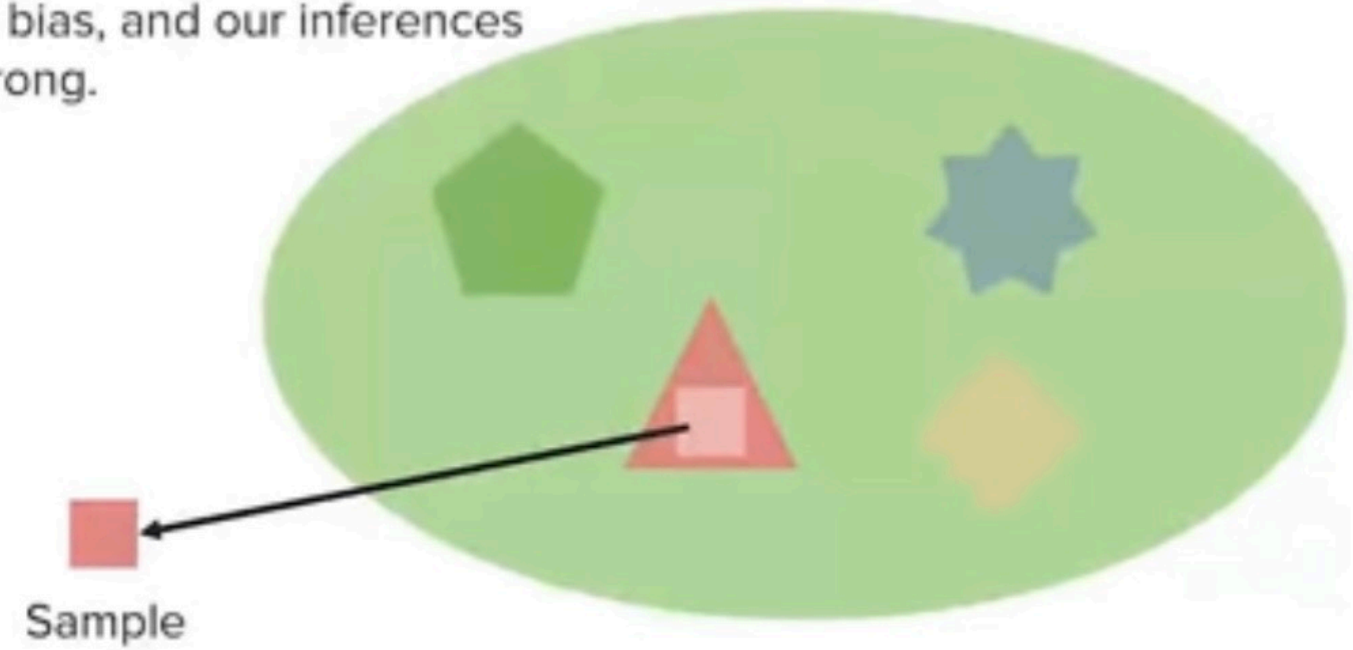
Sampling



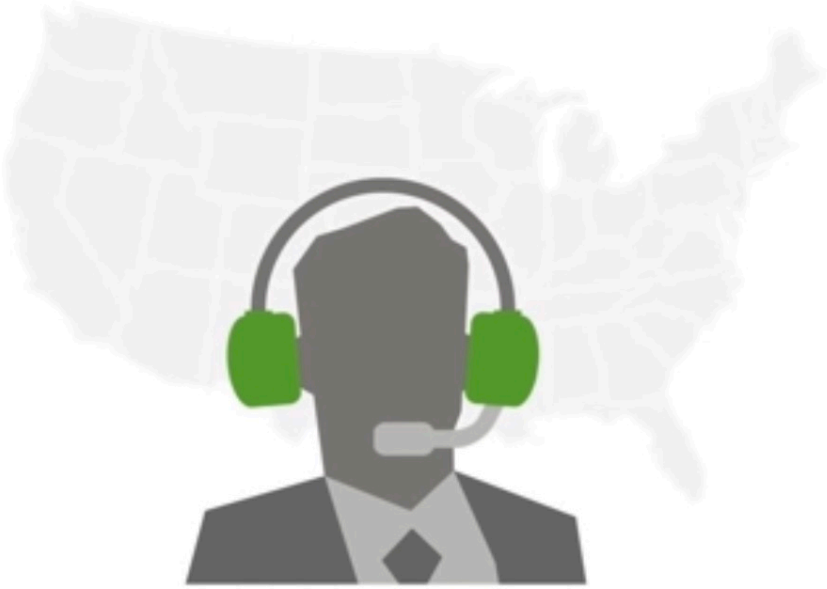
The statistics and epidemiologic approaches we use are affected by the assumptions of the sampling strategies used.

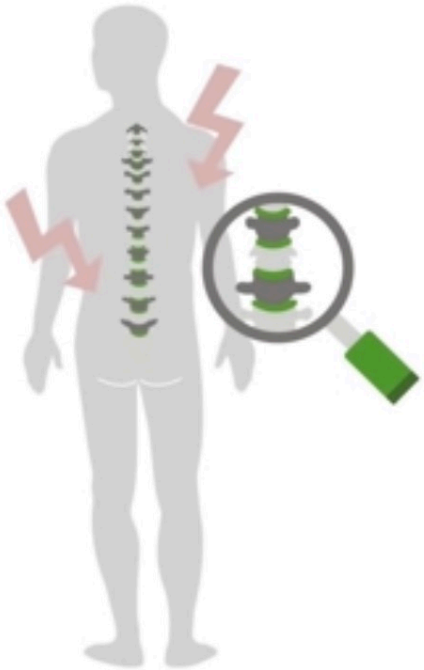
Sampling

If the sample is **not representative** of the population, then we have sampling bias, and our inferences will be wrong.



Sampling
Example





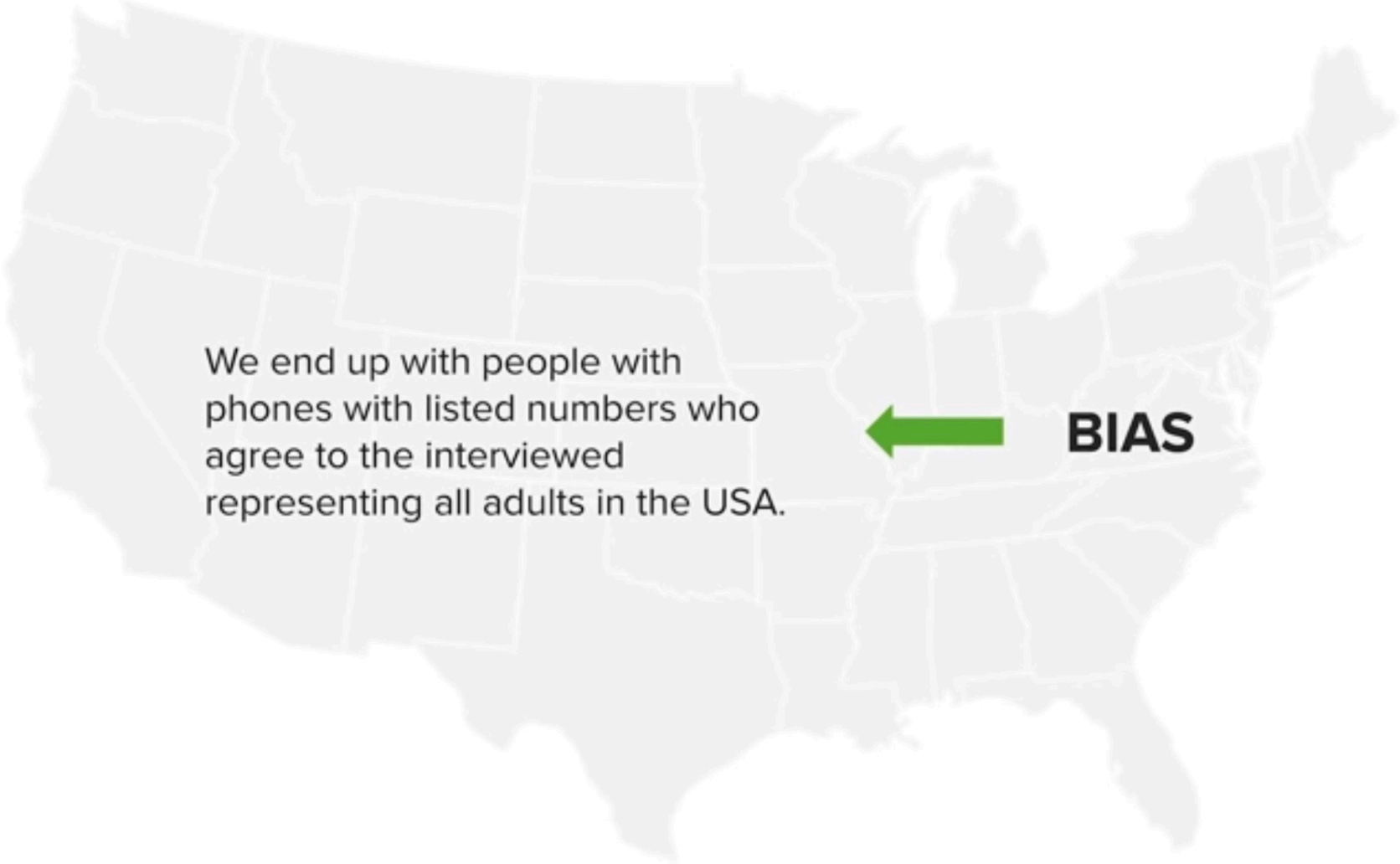
Let's say you want to run a telephone survey in the USA to measure the prevalence rate of perceived back pain.

Sampling

Example

To whom do you wish to generalize?	All adults in the USA.	**Reference population**
What population can you access?	Those with telephones.	**Study/Accessible population**
How can you access them?	Purchase a block of listed numbers from a phone company.	**Sampling frame**
Who is in your study?	Those who answer the phone and agree to participate.	**Sample**

Sampling
Example



We end up with people with phones with listed numbers who agree to the interviewed representing all adults in the USA.

BIAS

The Null Hypothesis

What is it?

It is a statement that there is no relationship between the variables we are testing.

The Null Hypothesis

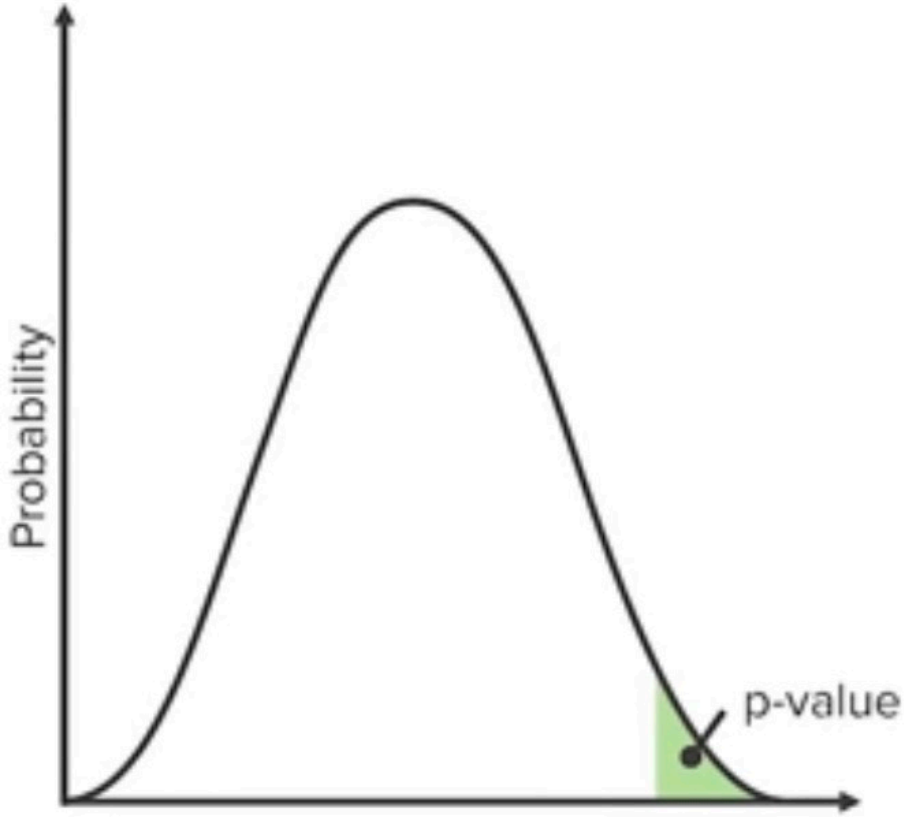
Why do we care?

Statistical tests allow us to either “reject” or “fail to reject” the null hypothesis.

The p-Value

A "p-value" is computed from a statistical test.

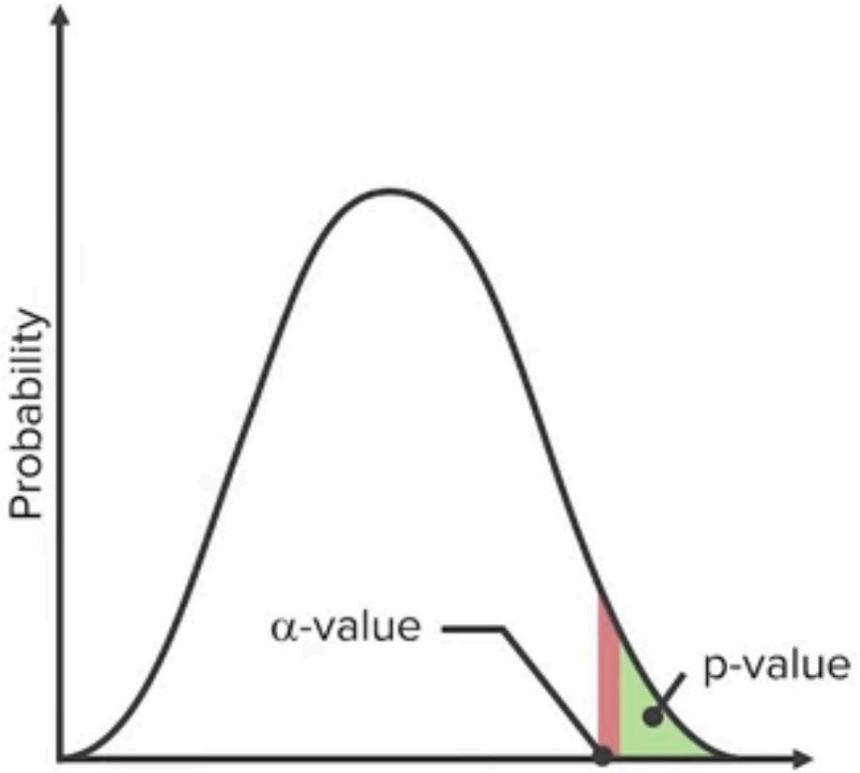
It tells us whether we should reject the null hypothesis.



The p-Value

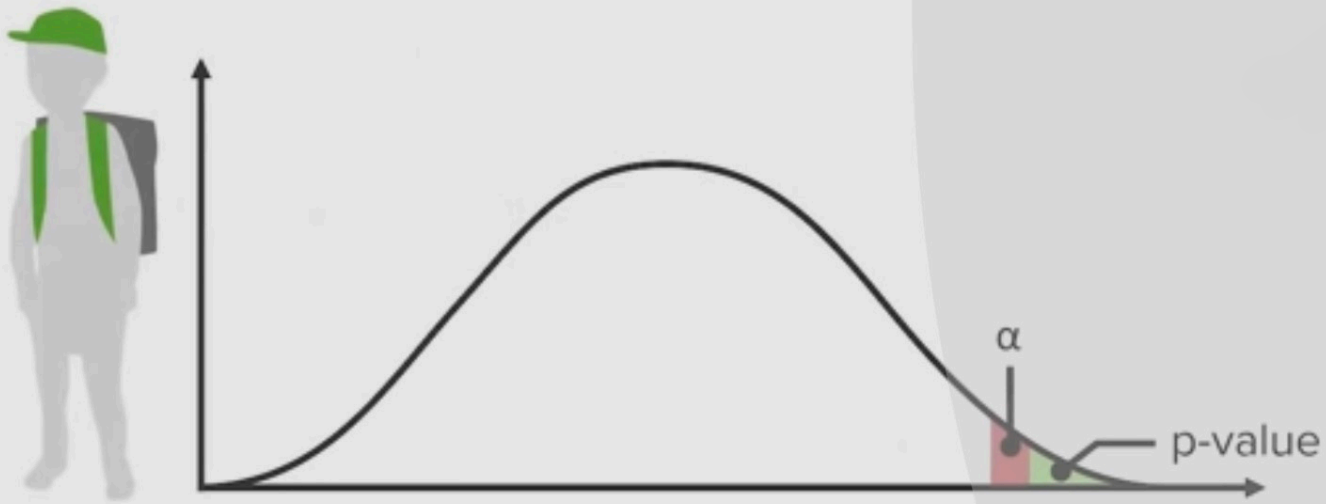
Whether or not we reject the null is determined by whether the p-value is below a certain cut-off, which we call the alpha value.

Traditionally, we tend to set alpha at either 0.05 or 0.01.



The p-Value

For example, if we are testing whether the average heights of two groups of children are different, and perform a t-test to produce a p-value of 0.02, setting $\alpha = 0.05$, we can conclude that null hypothesis is rejected and that the two groups do indeed have different average heights.



The p-Value

A convenient, though inaccurate, interpretation is that the p-value is the probability that your result was due to chance.

More accurately, the p-value is the probability of your test incorrectly rejecting the null, when indeed the null hypothesis is true.

The p-Value

A useful memory aid:

"If the p is low, the null (hypothesis) must go."

Confidence Intervals

A confidence interval is another way to express a statistical result along with its significance level, without having to use a p-value.

Example: The mean age of university students is

21 years (18, 21.5).

The actual parameter "point estimate".

Confidence Intervals

A confidence interval is another way to express a statistical result along with its significance level, without having to use a p-value.

Example: The mean age of university students is 21 years **(18, 21.5).**

The "confidence interval" of the parameter estimate.

Commonly Used Statistical Tests

T-test	Chi-square	ANOVA	Correlation	Regression
For comparing means of two groups, or mean of one group to an external standard.	For determining whether two categorical variables are associated with one another.	For comparing means of three or more groups, and/or across multiple domains.	For determining whether two continuous variables are related.	For establishing the influence of one variable on an outcome variable.

Learning Outcomes

- ✓ You understand the meaning of p-values, confidence intervals, and the null hypothesis.
- ✓ You know the difference between data and information.
- ✓ You know the differences between different types of variables.