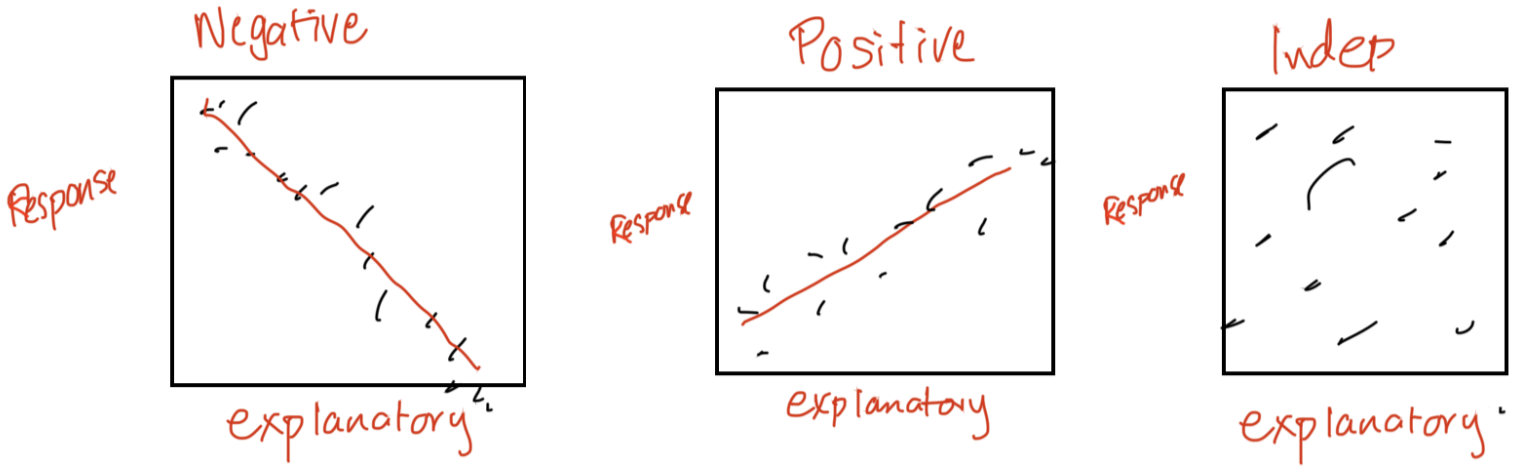
Chapter 2

2.2 Experiments

Relationships Between Variables

- Many research questions are about the relationship between two variables of interest.
- If we suspect **causality** we can use the terms **explanatory variable** and **response variable** to distinguish them.
- Typically we examine two variables using a **scatterplot**
  - *negatively associated*: downward trend
  - *positively associated*: upward trend
  - *independent*: no clear trend

Relationships Between Variables



Relationships Between Variables

Use caution with these terms. They have specific and meaningful implications!

Just because we

- *label* an explanatory/response variable, or
- *see* a trend

does not mean a relationship exists between variables.

Experiments

- In **experiments** researchers intervene and assign a **treatment** (explanatory variable) to each participant in the study.
  - *treatment group*: receives the treatment whose effects the researcher is interested in.
  - *control group*: receives no treatment, standard treatment, or placebo
- If treatment assignment is random we call this a **randomized experiment**

Experiments

A **confounder** is a variable (which may or may not have been measured) that is:

1) Associated with the explanatory variable
2) Associated with the response variable
3) Not a downstream consequence of either the explanatory or response variable

It creates a backdoor path (i.e., path of association) between the explanatory and response variables.

Experiments

Is there a confounding variable?

Do chemicals in sunscreen cause cancer?

3 minute read
Written by Cancer Center on May 28, 2020

In the summer of 2019, Dr. Neil Box toured Colorado with the Sun Bus, attending events with over 700,000 people in 46 service days. Free skin cancer screens identified 96 suspected skin cancers, including six cas also gave Dr. Box the opportunity to hear what people think about skin cancer and sun protection.

"A major bugbear for me was that people feel their sunscreen is going to harm them or damage them in s

Study shows Taylor Swift is the most listened-to artist among students with high GPAs

POPI Jr Contributor – Justine Nicholas Santos 27 July, 2023 No Comments 668

2040 33%

I guess Taylor Swift could really transport you to somewhere the culture's clever.

A study by College Rover, a student-parent information tool for college, found that American singer and songwriter Taylor Swift is the top artist for students with the highest GPAs.

Related Stories

Smash Mouth's former frontman

Experiments

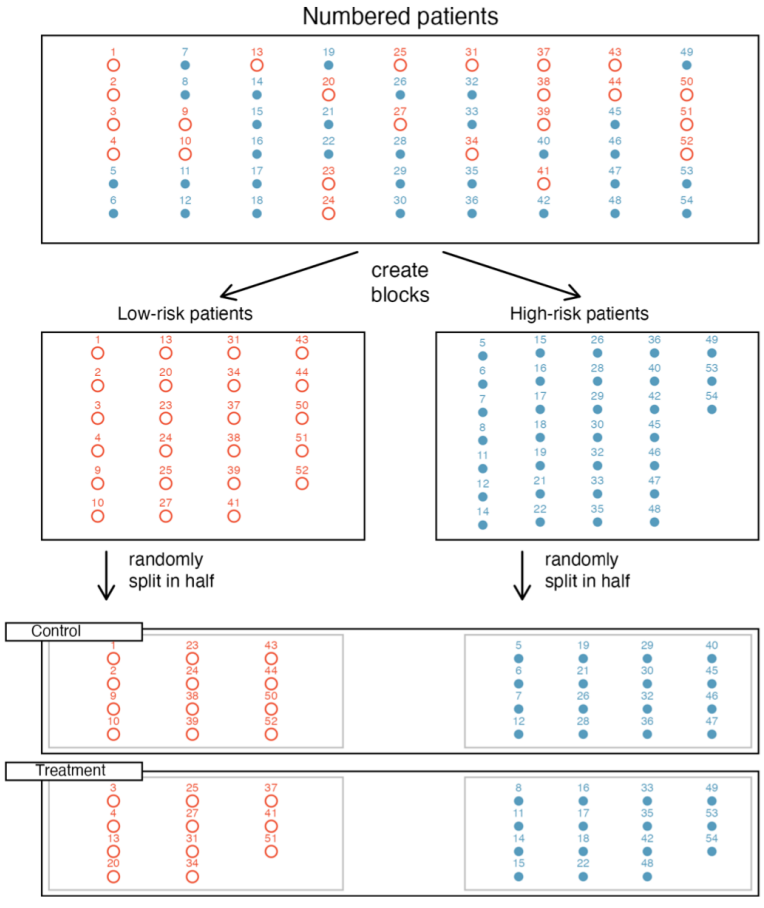
The **replication crisis** refers to influx of scientific studies that have not yielded the same results when repeated.

This has been caused by several issues.

Principles of Experimental Design

1) **Control**: isolate the relationship being studied.
2) **Randomization**: randomize patients into treatment groups to control for confounding variables.
3) **Replication**: the more cases observed, the more likely the true effects of the data are captured.
4) **Blocking**: ensure treatment and control groups are balanced with respect to a possible confounding variable.
   - Group individuals into blocks then randomize cases into treatment groups within each block.
   - Not always required, may need more advanced methods to analyze.

Blocking Example: Use blocking to study the effect of a drug on heart attacks



Reduce Bias in Human Experiments

- **Blind Study**: when researchers keep patients uninformed to which group they are in
- **Double Blind Study**: researchers and patients are unaware on which patients are received treatment
- Ethical concerns on not giving treatments.

2.3 Observational Studies

Observational Studies

- **Observational Study**: researchers observe both explanatory and response variable without interfering with how the data arise
- Typically cannot make causal conclusions on observational studies.
- A **prospective study** identifies individuals and collect info as events unfold, and a **retrospective study** collects data after events take place.

Big Picture

		Assignment of Explanatory Variable		
		Random allocation of explanatory variable	Individual decides explanatory variable (non-random)	
Selection of Observational Units from the Population	Random sample	The observational units are randomly selected from the population; then the explanatory variable (treatment) is randomly assigned.	The observational units are randomly selected from the population, but the value of the explanatory variable is not randomly assigned by the researcher.	Conclusions generalize directly to the population.
	Other sampling method (non-random)	The observational units are observed (somehow!) and then randomly allocated to the levels of the explanatory variable.	The observational units are observed (somehow!) and the value of the explanatory variable is not randomly assigned by the researcher.	Conclusions might not be generalizable because of volunteer bias.
		↓	↓	
		Significant conclusions are considered to be cause and effect.	Significant conclusions must be framed with possible confounding variables.	

- *Top left box*: Ideal, but very few datasets come from the top left box because usually ethics require that random assignment of treatments can only be given to volunteers.
- *Bottom right box*: Common for survey data.

Big Picture

Both experiments (random assignment of treatments) and representative (ideally random) sampling are important for how statistical conclusions can be made on populations.

- Was the explanatory variable randomly allocated?
- Dose the data represent a random sample from our target population?

This in turn determines *what* sorts of conclusions we can draw and to *whom* we can generalize those results.