Experimental design

(STATS201.stat 202 10: Experimental design and descriptive statistics)

Anna Helga Jónsdóttir
Sigrún Helga Lund

December 16, 2012

Bias

Bias

Bias occurs when the applied methods give a systematically biased view of the population to be analyzed.

- Subjects are chosen in a systematically biased manner: **Sample bias**.
- A good **sampling design** minimizes sample bias.
- Interfering influences from researchers and subjects: **Experimenters bias** and **placebo effects**.
- **Blinding** minimizes researchers bias and placebo effects.

Variability

Variability

Variability occurs because the phenomena under investigation are influenced by some randomness and therefore can the outcome of the measurements change each time the experiment is conducted.

- This variability causes that our results can change every time the experiment is repeated.
- Repetitions allow us to estimate the variability of the measurements.
- The more repetitions - the better we can draw conclusions from our experiment!

Controlled experiment

The objective of many researches is to demonstrate the effect of applying certain **interventions** to our subjects. In order to do so an **controlled experiment** needs to be conducted.

Controlled experiment

In order for an investigation to be classified as an **controlled experiment** two requirements need to be fulfilled:

1. The investigator can control which subjects obtain which interventions.
2. Measurements on the subjects are made both before and after the intervention is applied.

Random sampling

Random samling

A **random sampling** is made when subjects are chosen randomly from the population and all subjects have the same probability of being selected in the sample.

A sample that is chosen with random sampling is called a **random sample**.

We will look at three types of random samples:

- **simple random sample**
- **stratified random sample**
- **paired random sample**

Simple and stratified random sample

Simple random sample

When a **simple random sample** is chosen, subjects are chosen randomly from the whole population.

Stratified random sample

When a **stratified random sample** is chosen, the population is first divided into a few groups or strata and subsequently subjects are chosen by random sampling from each stratum.
The number of subjects chosen from each stratum need to be decided beforehand, but it can vary between strata.
Good to use when the number of subject within strata varies greatly.

Paired random sample

Paired random sample

When a **paired random sample** is chosen, the subjects are paired into groups of two and two and a fixed number of pairs of subjects is randomly sampled.
Good to use when the measurements are influenced by many uncontrolled variables. Then subjects that are influenced in a similar manner paired together.

What if a random sample cannot be chosen?

Sometimes difficulties in implementation of the experiment make random sampling impossible. Then one of two actions need to be chosen:

1. To redefine the population such that random sampling will be possible.
   - Then conclusions can only by made about the "new population". Is that feasible?
2. To accept the induced bias.
   - We not the sampling bias in our discussion.
   - Discuss thoroughly which consequences it can result.
   - Can it be assumed that the bias is small compared to the phenomena to be investigated?

Volunteer samples

- Volunteer samples are only gathered when the subjects are human beings and then measurements are only made on the subjects that volunteer to participate in the research.
- This induces a sampling bias because certain subjects can be more likely to volunteer than others.
- This bias can be so large that no inference can be made on the population from the measurements gathered.

Convenience samples

- Convenience samples are gathered when measurements are only made on subjects that are (conveniently) accessible to the researchers.
- This induces a sampling bias because certain subjects are more likely to be accessible to the researchers than others.
- This bias can be so large that no inference can be made on the population from the measurements gathered.

Missing values

Missing values

Often successful measurements are not made on every subject in the sample. Then we have **missing values** for these subjects.

- One cannot simply use the successful measurements and overlook the missing values.
- Some subjects are often more likely to have missing values than others.
- These subjects will, as a consequence, be less likely to be chosen to the reduced "sample".
- That causes a samling bias!

Experimenters bias and placebo effects

Experimenters bias

Experimenters bias occurs when the anticipations of the researchers influence the measurements of the effect of the interventions on the subjects.

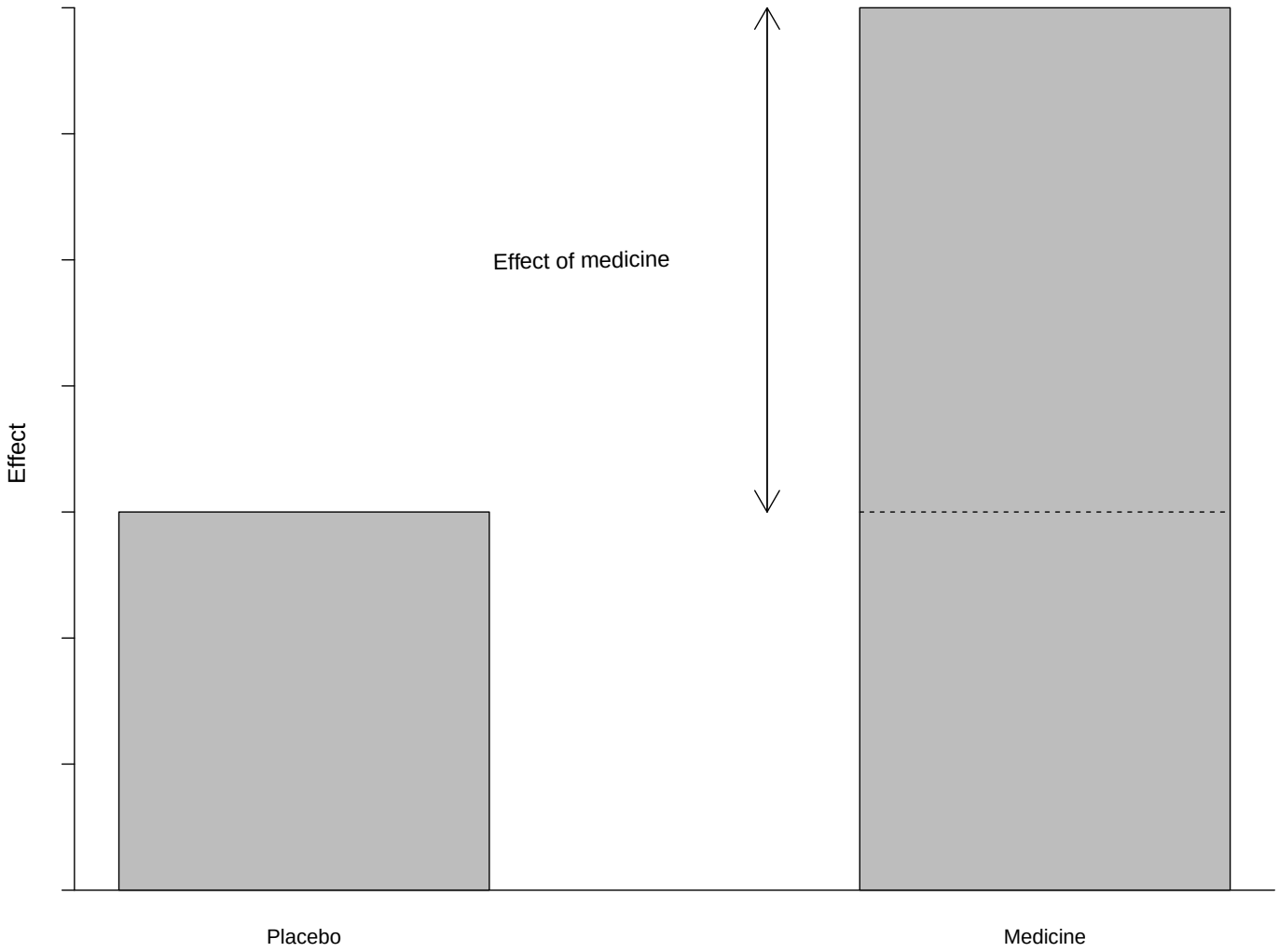
Placebo effect

Placebo is every intervention that the subject wrongly assumes that is the intervention under investigation.

Placebo effect

The difference in the measurements of the subjects before and after the placebo intervention is called a **placebo effect**.

Placebo effects



Single-blinded and double-blinded experiments.

Double blinded experiments

When an experiment is **double-blind** neither the investigator nor the subjects know which intervention they will receive. notice that an intervention can be a placebo intervention.

Single blind experiments

When an experiment is **single-blind** either the subjects don't know which intervention they received but the investigator does or vice versa.

Repetitions

- We assume that there is always some variability in our measurements - they are influenced by some randomness.
- The results can change every time a new sample is chosen and the investigation repeated.
- As soon as we have measured more than one subject, that is we have **repetitions**, we have some idea of the variability in the measurements.
- The more measurements we have, the better idea we have of this variability.

Drawing conclusions

- Statistical inference is drawing conclusions about a population based on investigations of a sample from that population.
- The more repetitions of subjects, the more likely it is that we can draw conclusions about the population.
- The main rule is: "The more repetitions, the better".

Causation

Causation

Causation between two variables is when the outcome of one variable influences the outcome of the other. Causation can only be shown with controlled experiments.

MYND Causal relationship /pictures/Orsakasamband

Good experimental design

Which requirements should a controlled experiment fulfill?

Every investigator should seek that his experiment fulfills the following conditions:

1. Sampling design.
   The subjects are chosen by random sampling and/or divided into groups by random sampling.
2. Blinding.
   The experiment is by all means double-blind but at least single-blind if that is impossible.
3. Repetitions.
   The intervention is applied to a repeated number of subjects.

If a controlled experiment fulfills these conditions, claims of causal relationship can be made.