Chapter 1: Controlled experiments

Outline

- Controlled experiments (Ch 1):
  - acne drug
- You will learn about:
  - population, sample, study unit
  - variable, response
  - method of comparison, treatment/control groups
  - confounding
  - randomization
  - placebo
  - blind and double-blind studies

Acne drug

Background

- Drug Roaccutane given to adolescents with severe acne
- Drug used by 13 million people world-wide, very effective in reducing acne
- Reported to Medicines and Healthcare products Regulatory Agency: 1588 'adverse events', 25 deaths from suicide

Questions about background

- What are 'adverse events'?
- 1588 adverse events were reported. Can we trust this number? How many events were not reported?
- Suppose the adverse events are depression. Was the occurrence of depression higher than usual?
- First try: 1588 out of 13 million means that about 12 out of every 100,000 people were depressed. Very low rate!
- But... the "Medicines and Healthcare products Regulatory Agency" is a British organization. So probably it only received complaints from British people.
- So to compute the rates of depression, we need to compare the number of depressed people to the number of users in Great Britain.

Questions about background

- But the the number of users in Great Britain is not given in the article! We cannot compute the rate of adverse events or the suicide rate...
- Even if these rates would be higher than 'usual', could it be that depression is caused by the acne, or by being in adolescense?
- Or could it have happened by chance?

6 / 38

Experiment

- Research at University of Bath
- Researchers gave the drug to adolescent mice for a 6-week period, and monitored their behavior
- They found that mice who got the drug were less mobile. This was interpreted as a sign of depression.

7 / 38

Questions about experiment

- Why was the drug tested on mice instead of people?
- How to measure depression in mice?
- Is mobility a good measure of depression? Could there be other reasons why the mice who took the drug were less mobile? Perhaps the drug causes muscle ache? Or sleepiness?
- How did the researchers measure mobility?
- Did the researchers know which mice got the drug? If so, could it be that they saw what they wanted to see?

8 / 38

Questions about experiment

- How big is the difference in mobility between the two groups? Could it be that by chance the mice in one group were just a bit lazy?
- Did the researchers use a placebo for the control group? Perhaps just the administration of the drug caused stress in the mice and made them less mobile?
- Were the mice kept under similar conditions? Perhaps the mice who got the treatment were placed in the corner of the room, where there was less light, and that caused them to move less?
- Do we understand the chemical mechanism of the drug?

9 / 38

Conclusions

- Has it been established that the drug causes depression in mice? No!
- If the drug causes depression in mice, does that mean it causes depression in people? No!
- Many questions remain unanswered in the article.
  - Some answers we can guess (very likely the mice were kept under similar conditions)
  - Others, we really don't know (how was mobility measured? how big was the difference in mobility? how many mice were used?).
- At the end of the article, the conclusions are weakened. Titles are often exaggerated/misleading!

10 / 38

A good experiment

11 / 38

Optimal design

- Define your research question and your response variable (= what you are going to measure)
- Obtain a group of eligible study units
- Randomly assign them to a treatment and control group (randomized controlled study)
- Give a placebo to the control group, and let neither the study units nor the evaluators know who is in the treatment group (double-blind study)
- Evaluate the response in both groups

12 / 38

Many choices!

- The response variable is the variable we measure, for example mobility, hormone levels, blood pressure, etc.
- What do we want to test exactly? Define this carefully!
- Examples for acne drug:
  - Average mobility level
  - Maximum mobility level
  - Average level of a hormone related to mood
  - Average level of depression measured by a certain questionnaire
  - Number of days with diagnosed depressions per year
  - Time to first suicide attempt
  - Time to first diagnosed depression or suicide attempt

Clinically meaningful

- You want something that is both *meaningful* and *likely to show some difference*
- examples related to meaningfulness (acne drug):
  - mobility level is not very meaningful – we are primarily interested in depression, and not in how much the mice move.
  - hormone level is not very meaningful – hormone levels may have to do with depression, but again we are primarily interested in depression and not in hormone levels.
  - number of days with depression per year is meaningful – if you tell people that the drug increases the number of depressed days per year by a factor 3 or so, then that will be strong evidence against using the drug.

Likely to show difference

- You want something that is both *meaningful* and *likely to show some difference*
- examples related to likelihood of showing a difference (acne drug):
  - suicide is not likely to show a difference – it occurs so rarely that one needs a very large sample size to find any difference.
  - number of days with depression per year – more likely to show a difference.

Population, sample study, unit

Population vs sample

- Usually we want to know something about a *population*. Examples
  - All adolescents with severe acne in the world
  - All adolescents with acne in the US
- It is often infeasible to look at the entire population.
- Instead, we look at a smaller group, a *sample*.
- The sample should be *representative* of the population, i.e. similar to the population.
- Why? Otherwise the conclusions about the sample do not generalize well to the population.

18 / 38

Estimate what is going on in population

- We can never know exactly what is going on in the population, because we don't look at the entire population.
- However, we can *estimate* what is going on in the population by looking at the sample.
- A sample consists of *study units*, e.g. people, mice.

19 / 38

Example

- We want to know the average height of people in the US.
  - Population: all people in the US
  - Variable of interest: height
- We don't have the time and money to measure all people.
- Therefore we take a *random sample* of 100 people (=study units).
- We measure their heights and average those. That gives an *estimate* for the average height of people in the US.
- Typically, this estimate does not equal the true average height. If we take a different sample, we get a different estimate. The bigger our sample, the more precise our estimate is.

20 / 38

Treatment/control groups

- Divide sample in treatment and control groups
  - Treatment group: gets treatment
  - Control group: often gets placebo, sometimes nothing
- Groups must be similar. Just like in physics/biology experiments, you only want to vary one thing at a time: the treatment.

22 / 38

How to get similar groups?

- We could try to make similar groups by hand. Acne example:
  - same number of male/female mice in each group
  - age distribution of mice similar in each group
  - etc.
- This is quite hard, and often does not work well because we do not know all the factors that are important. Example: weight of the mice might be an important factor for mobility of the mice, but we forgot to take this into account.
- The best way to get similar groups, is to assign study units to the groups *at random*, for example by flipping a coin. In that way, any differences are likely to even out.

23 / 38

There is one problem...

- It is not always possible to do a controlled experiment.
  - Example: the effect of smoking on lung cancer.
  - People do not want to start or stop smoking for a period of say 10 years because an investigator tells them to do so.

24 / 38

Groups must be comparable

Confounding

- If the treatment and control group differ by some other factor than the treatment, then we cannot separate the effect of the treatment and this other factor. This is called *confounding* (=mix-up).
- Example:
  - acne drug: suppose mice who got the drug were kept in a dark corner. Then is the difference in mobility due to the drug or due to the difference in light? Confounding factor: amount of light.

Non adherence

- Consider the following example:
  - We want to test whether a new cancer drug increases life expectancy
  - We do a double-blind randomized controlled trial - the optimal design
  - Suppose the drug does no good, but has bad side effects
  - As a result, the somewhat sicker people in the treatment group can't stand the drug. They stop taking it and die within months (just because they were sicker to begin with).
  - The people in the treatment group that stick to the drug are healthier. They live on average 1.5 years.
  - The people in the control group live on average 1 year.
  - By comparing the numbers 1.5 and 1, it looks as if the drug works. But it only has bad side effects!

Non adherence

- To make a fair comparison between life expectancy in the treatment and control group, we also need to drop the sicker people in the control group. Then the groups are similar again. But there is no way of knowing who these people are.
- Hence, we need to compare the entire treatment group to the entire control group, even though some people in the treatment group did not take the treatment. This is called *'intent-to-treat analysis'*.
- Terminology: people who don't stick to the treatment are called *non-adherent*.
- A similar problem exists when people are *lost to follow-up*, i.e. completely disappear from the study. In this case one cannot measure the response anymore, so they cannot be analyzed at all.Try to keep people involved in the study!

Blinding

29 / 38

Placebo

- We want to measure the effect of the treatment, not the effect of the *idea* of getting a treatment.
- These two effects are confounded (mixed-up), unless we use a placebo.
- A placebo is something that looks similar to the treatment, but has no effect.
- It is not always possible to use a placebo.
  - Example: testing a new intensive form of psychotherapy for alcohol addicts.

30 / 38

Blind and double-blind

- A *blind study*: the study units don't know whether they get the treatment or not.
- A *double-blind study*: both the study units and the evaluators don't know who gets the treatment.
- Why use blinding?
  - Blinding the study units (e.g. by using a placebo) ensures that we only measure the effect of the treatment, and not the effect of the idea of getting the treatment.
  - Blinding the investigators makes sure that they don't see what they want to see.

31 / 38

Experiments for acne drug

32 / 38

Proposal 1

- Define research question and response
- Double-blind randomized controlled trial:
  - Take a sample of eligible adolescents with acne
  - Random assignment to treatment/control group
  - People in control group get a placebo
  - Neither the study participants nor the investigators know who gets the treatment and who gets the placebo
- Question: would this be ethical?
  - If we think the drug is potentially dangerous, can we give it to people? Probably yes.
  - We know the drug reduces acne. Is it ethical to give a placebo to people with severe acne? Probably yes, if they consent to enroll in the trial.

33 / 38

Problems

- The blinding will be broken quickly, since only people in the treatment group get less acne.
- A difference in depression levels between the two groups may be caused by a difference in acne levels. Acne becomes a confounding factor.

34 / 38

Proposal 2

- Define research question and response
- Take a sample of adolescents *without acne*
- Double-blind randomized controlled trial (see previous slide)
- Now blinding will not be broken
- Question: would this be ethical?
  - Is it OK to give a potentially dangerous drug to healthy people? I'm not sure.

35 / 38

Problems

- The effect of the drug in adolescents without acne may differ from its effect in adolescent with severe acne. Our sample is not representative of the population of interest.
- But it seems likely that if the drug causes depression, it would do so both in people with and without acne. So this may not be a big problem.

36 / 38

What's next?

37 / 38

Observational studies

- Quiz section on Thursday:
  - Salk vaccine trial - polio
- Lecture on Friday: Observational studies (Ch 2)
  - These are not ideal, but sometimes it is not possible to do controlled experiments (example: effect of smoking).
  - Finding weak points in such studies is something of an art, and often depends on outside knowledge (this is also the case in homework problems).
  - We will look at several examples.

38 / 38