MA-120 Probability and Statistics

Nazar Khan
PUCIT
Lecture 2: Design of Experiments

Design is more important than the experiment itself

Is Polio vaccine effective?

- Some one makes a vaccine for Polio.
- You need to find if it is effective or not.
- How will you go about finding an answer to this question?
  - Follow the scientific method

Is Polio vaccine effective?

- Some one makes a vaccine for Polio.
- You need to find if it is effective or not.
- How will you go about finding an answer to this question?
  - Follow the scientific method

Is Polio vaccine effective?

- Some one makes a vaccine for Polio.
- You need to find if it is effective or not.
- How will you go about finding an answer to this question?
  - Follow the scientific method

Is Polio vaccine effective?

- Some one makes a vaccine for Polio.
- You need to find if it is effective or not.
- How will you go about finding an answer to this question?
  - Follow the scientific method

How to lie with statistics

- What if the treatment and control groups are different?
  - Treatment group is from people that are more immune to Polio – poor children
  - Control group is from people that are less immune to Polio – rich children
  - Statistics will falsely show that the vaccine is effective.

	Treatment	Control	Conclusion
Outcome	Less Polio	More Polio	Effective

How not to lie with statistics

- The treatment and control groups **must be similar**.
- How to ensure that?
  - Pick randomly
- Statistical studies can mix up hidden factors.
  - Polio is a disease of hygiene. This factor must be accounted for in the treatment and control groups.

Questions that statistics can answer

- Is Homeopathy effective?
  - Bad study: ask those that use Homeopathy.
  - Good study: perform a **controlled** experiment with **randomly selected** treatment and control groups.
- How important is Misbah-ul-Haq to the team?
  - Bad study: look at his averages.
  - Good study: look at the circumstances of his innings.

Statistics is also an ‘art’

- Blind statistics: Plugging numbers into formulae without thinking.
- Proper statistics: Controlling for the hidden factors.
- Blind application of statistics can be disastrous.
  - Terming a medical treatment effective when, in fact, it is harmful.
  - Terming a medical treatment ineffective when, in fact, it is effective.

Confounding Factor

- The treatment and control groups should differ from each other **only in terms of the treatment**.
- If they differ with respect to some other factor, then this is a **confounding factor**.
  - Are the results due to treatment or due to the confounding factor?

The Randomized Controlled Experiment

- **Random** selection of treatment and control groups
- Elimination of **confounding factors**
  - **Placebo effect** – some people are cured by the idea of treatment. So give both groups the impression that they are being treated.
  - **Double blinding** – neither the patient nor the doctors know which group the patient is in. So the doctors don’t give different treatment.

The Observational Study

- Is smoking harmful?
  - Can we do a randomized controlled experiment to answer this?
- We can **observe** smokers and non-smokers over time.

Is Smoking Harmful?

- If smokers are less healthy compared to non-smokers, then “yes, smoking is harmful”.
- **Wrong!!**
- Association is not causation!
- What about confounding factors?
  - For example, a gene that causes lung cancer and also causes people to smoke.
- Careful studies have concluded that there are no confounding factors. **Smoking really is harmful.**

Association is not Causation

- Many people die in hospitals.
- Do hospitals cause death?
- Hospitals and deaths are **associated** with each other.
- But hospitals don’t **cause** deaths (in general).
- Blind statistics will tell you that hospitals lead to deaths.

Randomized Controlled Experiment vs. Observational Study

- Observational studies prove association but not causation.
- Confounding factors can be at work.
- Randomized controlled experiments try to minimize the effects of confounding factors.
- So, wherever possible, a randomized controlled experiment should be performed to understand the real world.

Next Lecture

- Descriptive statistics
  - Histograms
  - Mean, Standard Deviation
  - The Normal Curve
- Read Ch. 3-5.