

Basics of Data Analysis , Research Design and R :Lecture 1

Zahid Asghar

11/3/22

**Learn to Explore Data alongwith
fundamentals of quantitative
reasoning and logic**

Simply focussing statistics and techniques: Miss the forest for the trees

Introduction to Data and exploring it with R

thinking clearly with data is very important ,once thinking is focused, techniques will follow

Learning R for making sense of data: real data based case studies

Exploring WHO, Gapminder, cia_factbook/UN-voting data sets

Most politicians facing scandals win reelection

Correlation doesn't imply causation

Correlation : description, forecasting

Counterfactuals , Causal Inference

Learning R and Quarto

R: not only a statistical language but all documentation under one R

Beautiful reports/slides/pptx, pdf documents , CV, research paper... in R

Exploratory Data Analysis to make sense of data

Shallow men believe in luck, believe in circumstances. Strong men believe in cause and effect.

Ralph Waldo Emerson, The Conduct of Life

Basics of data visualization under ggplot2

- Exploratory Data Analysis : nycflights, bikes, cia, tax...
- Probability Theory , Lab based on Kobe's golden hand
- Probability Distributions, Sampling Distributions, CLT
- Random and non-random sampling procedures
- p-value, p-hacking, p-screening, publication bias
- Statistical modeling : Linear Regression modeling (Simple and Multiple)

Manual table

The randomized controlled
double-blind experiment

Size	Rate	
Treatment	200,000	28
Control	200,000	71
No consent	350,000	46]

The NFIP study

Size	Rate	
Grade 2 (vaccine)	225,000	25
Grade 1 and 3 (control)	725,000	54
Grade 2 (no consent)	125,000	44]

.footnote[Source: Tohmas
Francis, Jr. "An evaluation of the
1954 poliomyelitis vaccine trials-
summary report," American
Journal of Public Health vol.45
(1945) pp1-63]

atomcamp

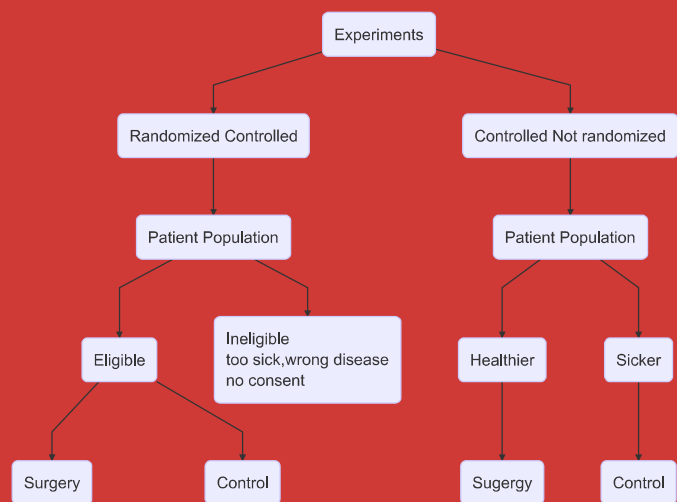


Table 3. Randomized controlled experiments vs. controlled experiments that not randomized: three-year survival rates in studies of the porta-caval shunt. (Percentage are rounded)

	Randomized	Not randomized
Surgery	60%	60%
Controls	60%	45%

Table 4. A Study of studies. Four therapies evaluated by RCTs and trials using historical controls. Conclusions of the study are summarized as positive about value of the therapy (+), or negative (-).

Therapy	Randomized controlled	Hisotrically controlled
	-	+
Coronary bypass surgery	1	7
5-FU	0	5
BCG	2	2
DES	0	3

Table 5: Randomized controlled experiments vs. studies with historical controls: three-year survival rates for surgery patients and controls in trials of coronary by surgery. Randomized trials differ from historical controls.

	Randomized	Historical
Surgery	87.6%	90.9%
Controls	83.2%	71.1%

Observational Studies

Controlled studies not the same as observational

- **observational studies:** its the subjects who assign themselves to different groups: the investigator watch just what happens

[Controlled : investigator decides who to be in control group and who to will be in treatment group]

Control has two senses

- a control is a subject who did not get the treatment:
- a controlled experiment is a study where the investigators decides who will be in the treatment group and who will not.

Smoking and lungs cancer

Main issue: was the control group really similar to the treatment group- apart from the exposure of interest?

Statisticians talk about controlling for confounding factors in an observational studies. This is a third word of the use control.

Examples

The Clofibrate trial

Pellagara

Cervical cancer and circumcision

Ultrasound and low birthweight

The Samaritan and suicide

Data Tables

```
1 library("DT"); library("dplyr")
2 iris %>%
3   DT::datatable(class = "compact", rownames = FALSE, extensions = "Buttons",
4                 options = list(dom = 'tBp', buttons = c("csv", "excel"),
5                 pageLength = 8)) %>%
6   DT::formatRound(1:4, digits = 1)
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa

CSVExcel

Previous12345...19Next



ABL : Always Be Learning.

