S 01

Data Analytics for Data Scientists

Design of Experiments (DoE)

Suggested solutions for Exercise 01: Introduction

February 2025

Prof. Dr. Jürg Schwarz
MSc Adrian Bieri

Suggested solution 01 ____ **2**
A famous study ____ 2
Suggested solution 02 ____ **3**
Spurious Correlation ____ 3

Suggested solution 01

A famous study

Sir Austin Bradford Hill, called Tony Hill, (* July 8, 1897 in London; † April 18, 1991) was a British epidemiologist, statistician and pioneer of clinical trials. Together with Richard Doll, he was the first to establish a relation between smoking and lung cancer.

In 1948, Hill published the results of a study[1] he had conducted on the treatment of tuberculosis with the antibiotic *streptomycin*.

Have a look at Table II from the publication and discuss a **possible research design** that Hill may have chosen and the results of the study.

TABLE II.—*Assessment of Radiological Appearance at Six Months as Compared with Appearance on Admission*

Radiological Assessment	Streptomycin Group		Control Group	
Considerable improvement ..	28	*51%*	4	*8%*
Moderate or slight improvement	10	*18%*	13	*25%*
No material change	2	*4%*	3	*6%*
Moderate or slight deterioration	5	*9%*	12	*23%*
Considerable deterioration ..	6	*11%*	6	*11%*
Deaths	4	*7%*	14	*27%*
Total	55	*100%*	52	*100%*

Suggested answers to the questions

Individuals were randomly assigned to a treatment group (*streptomycin* group) and a control group – see *The Control Scheme* in the study on page 770.

The persons in the treatment group were given the antibiotic *streptomycin*.
This is an RCT study (RCT = *Randomized Controlled Trial*)

Both the persons in the two groups and the medical staff did not knew whether a person had been assigned to the treatment group or the control group.
This is a double-blinded RCT study (*Double Blinded Randomized Controlled Trial*)

The treatment with the antibiotic *streptomycin* was effective:
51% of the treated persons experienced a significant improvement (8% in the control group)
7% of the persons with treatment died (27% in the control group)

[1] Medical Research Council (1948): Streptomycin Treatment of Pulmonary Tuberculosis. In: BMJ 2 (4582), p. 769. DOI: 10.1136/bmj.2.4582.769.

Suggested solution 02

Spurious Correlation

From the lecture notes (script)

*Confounding means that a factor (**confounder**) that is not directly investigated is associated with both the independent variable and the dependent variable and accordingly causes the relationship between the two variables (Spurious Correlation).*

Given the definition above, interpret the following three statements:

- *The more firefighters fight a fire, the greater is the damage.*

- *The fewer storks nest in an area, the lower the human birth rate is in that area.*

www.istockphoto.com

- *The more books a family owns, the better the children are able to read.*

  A few comments on the last statement:

  The *Statistical Analysis Report* of February 2000 by the *National Education Longitudinal Study*, which the U.S. Department of Education carries out for children, states:

  A strong indicator of the literacy environment in the home may be the number of child-oriented resources (e.g., books, tapes and CDs).

  https://nces.ed.gov/pubs2000/2000070.pdf

  Based on these findings, the governor of the State of Illinois in 2004 wanted to send a book a month to all families with newborns until the children enter kindergarten. The project would have cost USD 26 million a year, but government rejected the project.

  The PISA study on student assessments asked a similar question:

  How many books do you have at home?

  with this comment:

  One meter of bookshelf can hold about 40 books.
  Don't include magazines, newspapers, and your school books.

  and these answers to choose from:

  0-10 books / 11-25 books / 26-100 books / 101-200 books /
  201-500 books / more than 500 …

Suggested answers to the questions

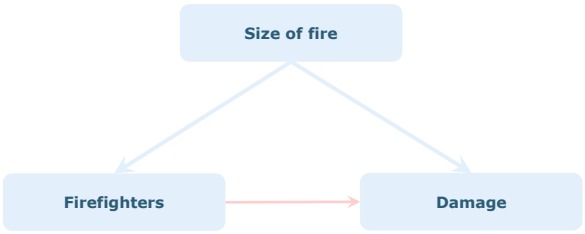
Firefighting

In this case, there is a third, causal variable: The **size of the fire**.

In addition, a certain direct, causal effect can be assumed, since a larger number of firefighters tend to cause more damage due to a lack of coordination.

Another approach would be that the **wealth of the residential area** acts as a confounder.

In wealthy neighborhoods, there is more public safety and emergency services (police services, fire services, ...) Therefore, in the event of a fire, more firefighters are present and the damage is greater compared to a residential area with lower wealth.

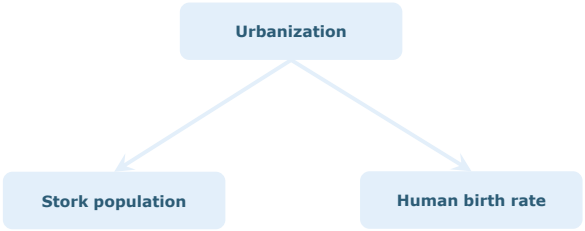


Storks

A direct relationship is most unlikely.

The association can be explained with the degree of **urbanization**. Urbanization is the third, causal variable (*confounder*) for the human birth rate and stork population.

Urbanization leads to a decline in the stork population because more land is being developed. Urbanization also leads to a lower birth rate because women tend to be more educated.



Reading

Here a third, causal variable in the **parental education level** at home can be assumed.

In addition, a greater direct, causal effect can be assumed, since reading books can increase reading skills, especially if children start at an early stage.