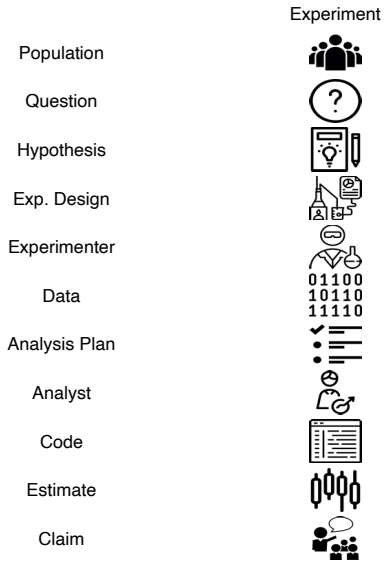
Statistical Inference II - PhD ECOSTAT - University of Milano-Bicocca

Homework 0

To submit via e-mail by 20/04/2020 h 14:00.

A scientific study consists of document(s) specifying a population, question, hypothesis, experimental design, experimenter, data, analysis plan, analyst, code, parameter estimates, and claims about the parameter estimates.
Reference: Patil, P., Peng, R.D. and Leek, J. (2016) A statistical definition for reproducibility and replicability. BioRxiv.



This homework is about two experiments:

- Darwin's experiment
- Galileo's experiment

1 Darwin's experiment

Charles Darwin collected data over a period of years on the heights of Zea mays plants.

Population: Zea mays plants

Darwin's hypothesis: Height of a plant depends on the type of fertilization.

Experimental Design: The plants were descended from the same parents and planted at the same time. Half of the plants were *self-fertilized*, and half were *cross-fertilized*, and the purpose of the experiment was to compare their *heights*. To this end Darwin planted them in pairs in different pots.
The focus of interest is the relation between the height of a plant and something that can be controlled by the experimenter, namely whether it is self or cross-fertilized. This means that you can regard the height as random with a distribution that depends on the type of fertilization, which is fixed for each plant.
Note that in order to minimize differences in humidity, growing conditions, lighting, etc. Darwin had decided to plant the seeds in pairs in the same pots. The height of a plant would therefore also depend on these factors, which are not of interest, not only on the type of fertilization.

Data:

	Pot	Cross	Self
1	I	23.500	17.375
2	I	12.000	20.375
3	I	21.000	20.000
4	II	22.000	20.000
5	II	19.125	18.375
6	II	21.500	18.625
7	III	22.125	18.625
8	III	20.375	15.250
9	III	18.250	16.500
10	III	21.625	18.000
11	III	23.250	16.250
12	IV	21.000	18.000
13	IV	22.125	12.750
14	IV	23.000	15.500
15	IV	12.000	18.000

Height is measured in eighths of an inch.

Questions:

1. Does the height depend on the type of fertilization?
2. Could you estimate the height difference, and assess the uncertainty of the estimate?

Task: Answer 1. and 2. by specifying

1. **Analysis Plan:** Define a statistical model, and be very specific about the required assumptions. The less the assumptions the better. You can use different models to answer 1 and 2.
2. **Code:** Write the R code to get the results. You can add plots, statistical summaries, etc. Your code must be replicable.
3. **Claim:** Comment your results and write your conclusions about 1 and 2.

Reference: A.C. Davison (2003). Statistical Models. Cambridge University Press. Examples 1.1, 3.11, 3.16, 7.24, 7.28, etc.

2 Galileo's inclined plane experiment (1604)

- Question of interest: If a ball rolls down a ramp, what is the relationship between time and distance?

- Aristotle hypothesis: Constant velocity (zero acceleration): distance $\propto$ time
- Galileo hypothesis: Increasing velocity (constant acceleration): distance $\propto$ time ^{2}
- Experimental data:

time	distance
1	33
2	130
3	298
4	526
5	824
6	1192
7	1620
8	2104

Time (equal intervals), Distance (in points)

Task: Reject Aristotle hypothesis in favour of Galileo's hypothesis

1. **Analysis Plan** Choose the statistical model and specify all required assumptions
2. **Code** Write the R code to get the results. Your code must be replicable.
3. **Claim** Comment on the results.

Reference: Galileo's Great Discovery: How Things Fall

`https://www.springer.com/cda/content/document/cda_downloaddocument/9781461454434-c1.pdf`