Null hypothesis rejection: experimental data analysis

Kalyani Ramachandran PhD

Biologist and Entrepreneur

24 Sep 2019

Experimental data analysis

- In real life, one wants to detect patterns, make comparisons, and understand differences

- Statistics – methods to do all of the above

Applied example 1

- Here is an example from my lab- liver cells
- Each group has been given a different drug; sampled values indicate quantity of gene

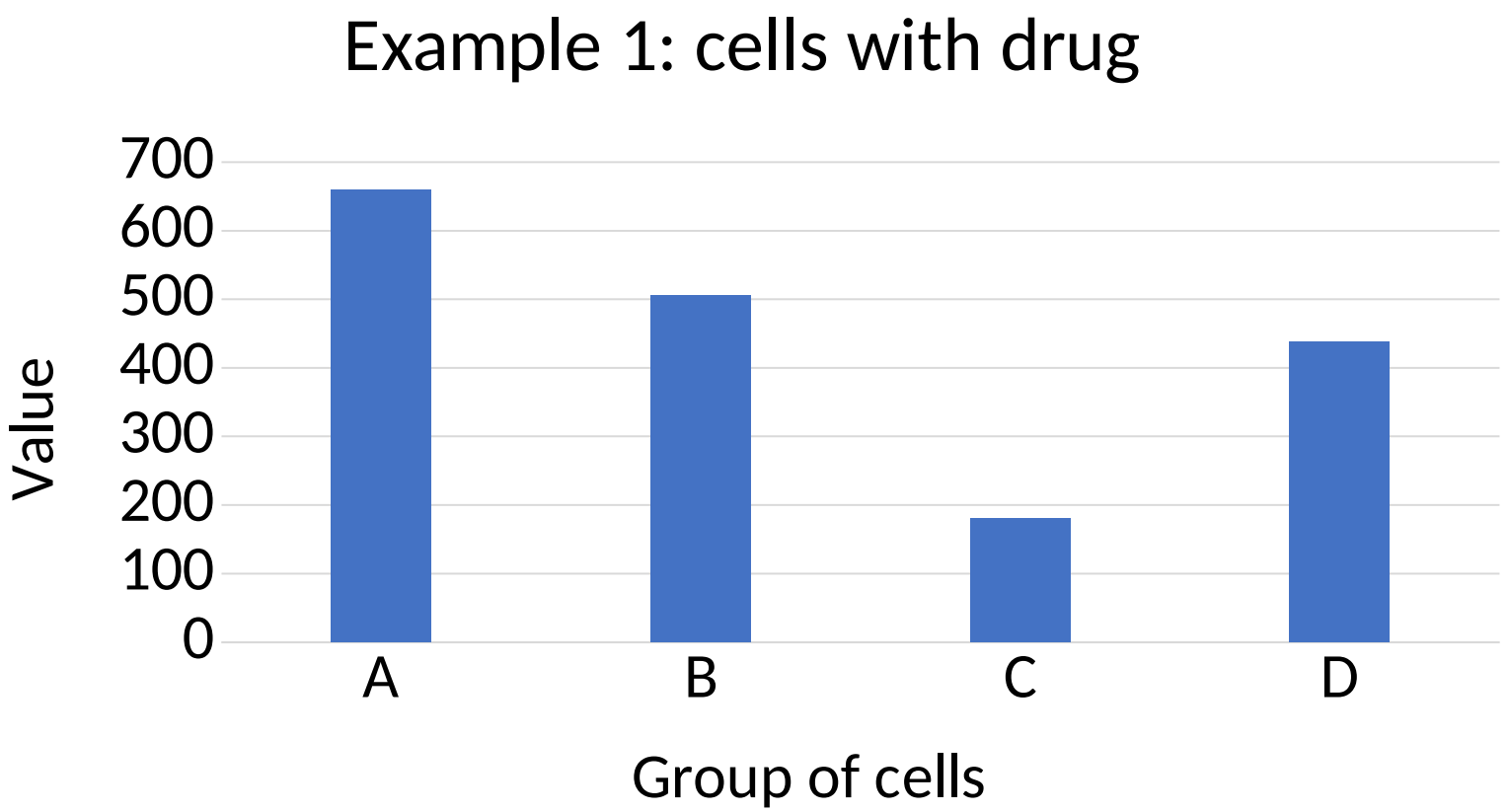
Group	Value 1	Value 2
A	606.5	713.5
B	506.5	503.5
C	133	228
D	418	459

Applied example 1

- Data from human cells – each group has been given a different drug; sampled values indicate quantity of gene

Group	Value 1	Value 2	Mean
A	606.5	713.5	660
B	506.5	503.5	505
C	133	228	180.5
D	418	459	438.5

Applied example 1:



Which group of cells are different?

Are the differences true or just by chance?

Null hypothesis

- Hypothesis states that there is no relationship between two data sets
- H0: $\mu 1 = \mu 2$
- Alternative hypothesis : H1: $\mu 1 \neq \mu 2$

- Statistically significant if p-value is below set value
- REJECT the null hypothesis vs fail to reject the alternative hypothesis
- p-value understanding : 0.05/0.01/0.001
- Statistical assumptions: normal/student’s t-distribution

Example 1 again

Group	Value 1	Value 2	Mean	p-value
A	606.5	713.5	660	
B	506.5	503.5	505	0.1
C	133	228	180.5	0.02
D	418	459	438.5	0.06

Significance limit: 0.05; H0: $\mu A=\mu B=\mu C=\mu D$

Which group is significantly different from group A ?

How to interpret if p-value is borderline?

t statistic calculation

$t= X_1-X_2/(S_p\sqrt{(2/n)})$

X_1, X_2 are means of the two sample populations

$S_p = \sqrt{(s^2_{x1} + s^2_{x2})/2}$

S_p is pooled standard deviation and s_{x1} , s_{x2} are variances of the two samples.

Look up corresponding p value in t table

or

In excel function

TTEST(array1,array2,tails,type)

Arrays refer to sample groups, tails is whether right/left/both, type is variance- paired, equal or unequal

or

In R

t.test(y1,y2, var.equal=TRUE)

Healthcare applications

- Drug/treatment clinical studies
- Design of study: Compare two groups of patients: control vs drug/treatment
- Hidden biases: number of samples, age and gender of participants, outcome definition, sampling, number of replicates in experiment.

It is easy to get fooled if one relies only on the p-value.
Statistical interpretation has to be done carefully.

Applied example 2: mice model

	Treatment			
Values	Pre-Control	Pre-A	Post-Control	Post-A
1	0.37315	0.86272	0.13301	0.52339
2	0.39038	1.1056	0.67946	0.55169
3	0.40688	1.03096	0.96005	0.23163
4	0.09202	1.31211	0.30404	0.63068
5	0.51352	1.04229	0.63041	0.52417
Response to injury	Yes	No	Yes	yes

Each group of mice (n=5) has been given either control or treatment A

Values indicate gene A quantity

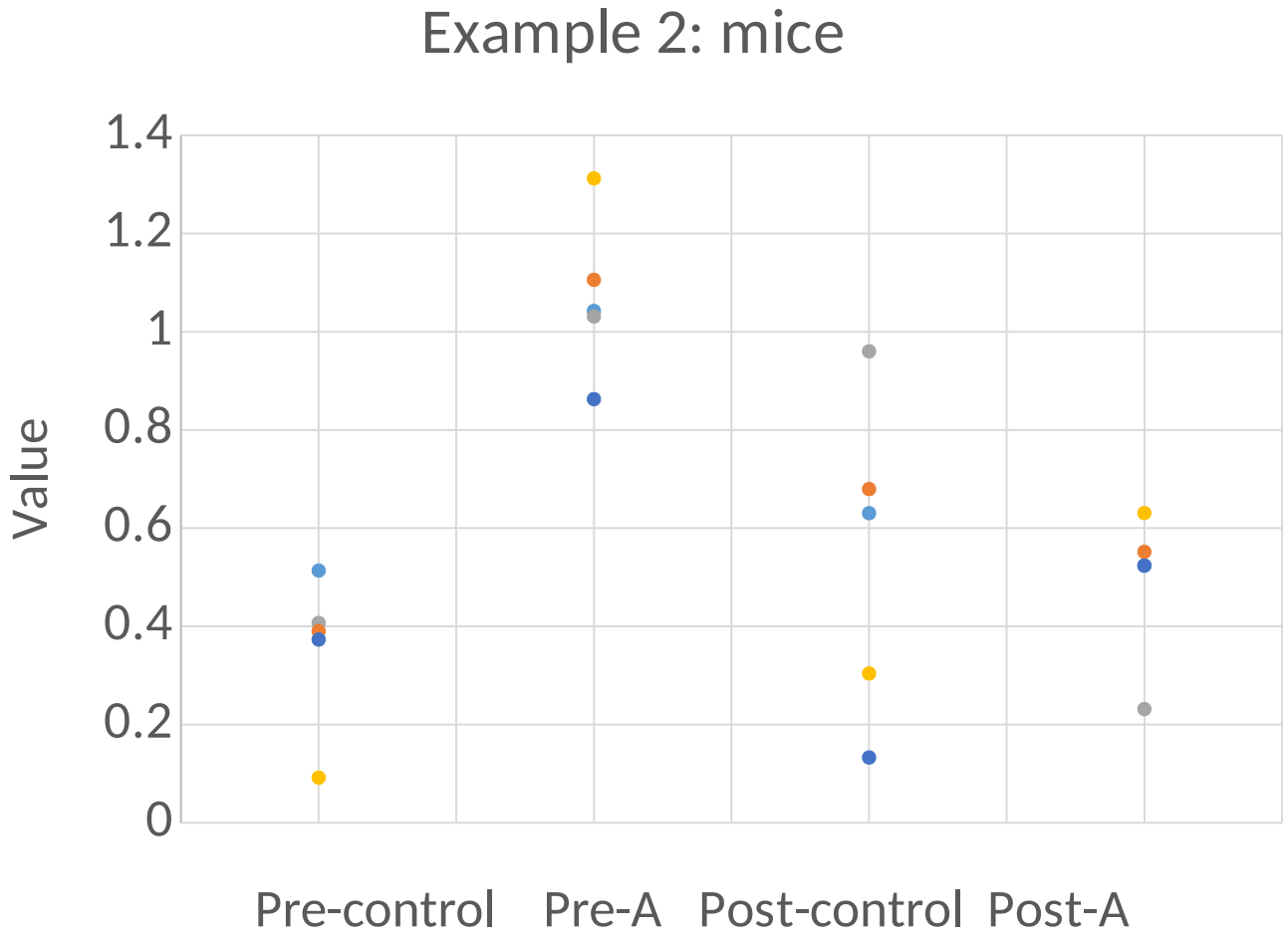
‘Pre’ and ‘post’ refer to when the treatment was given with respect to injury to mice

Injury response is counted by looking at liver of mice after injury plus treatment.

Injury to liver mimicking alcohol consumption/metabolic syndrome.

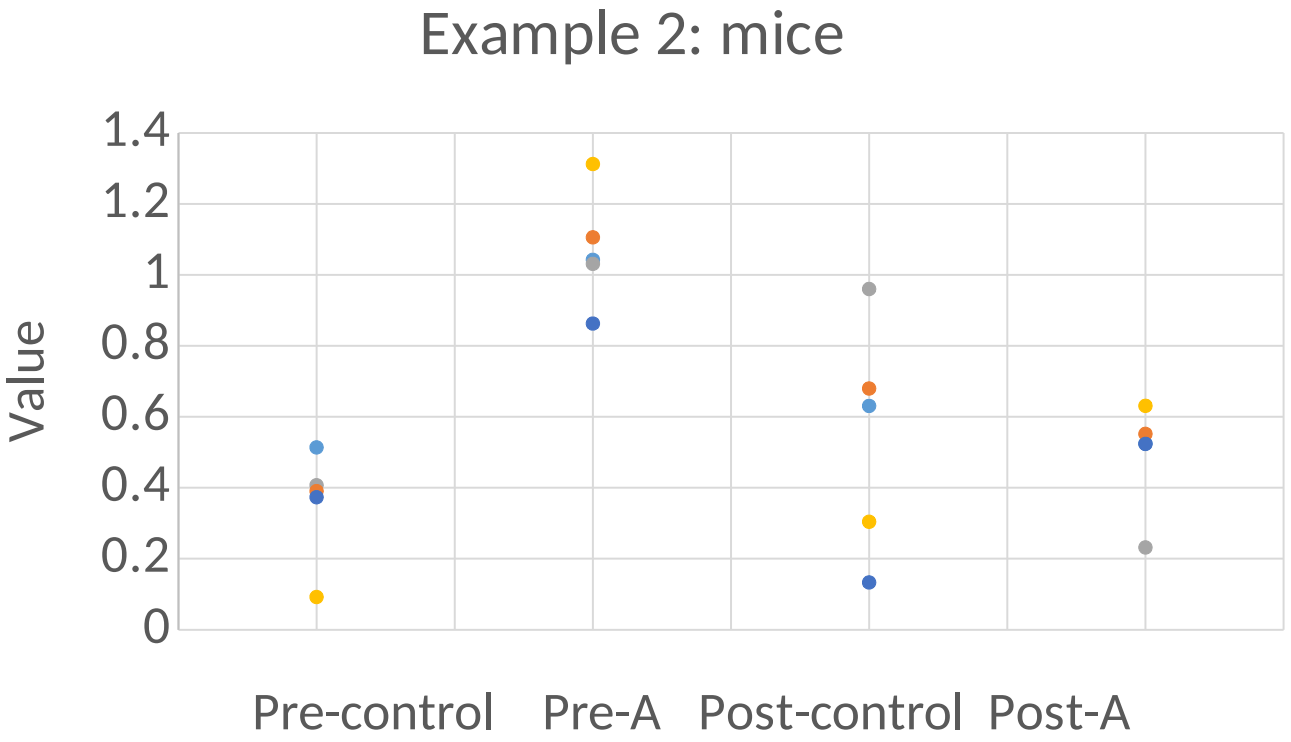
Applied example 2

- Pairwise comparisons of all groups
- Null hypothesis and significance level set
- Pre-A was significantly different from all other groups (p-values of less than 0.01)



Applied example 2

	Treatment			
Values	Pre-Control	Pre-A	Post-Control	Post-A
1	0.37315	0.86272	0.13301	0.52339
2	0.39038	1.1056	0.67946	0.55169
3	0.40688	1.03096	0.96005	0.23163
4	0.09202	1.31211	0.30404	0.63068
5	0.51352	1.04229	0.63041	0.52417
Response to injury	Yes	No	Yes	yes



The data suggests that pre-treatment with A gives protection to injury, correlating with gene value significantly higher in protected group.

Vaccine but for lifestyle diseases!

Summary

- Rejection of null hypothesis is commonly used in analysing experimental data
- The p-value at a significance level above the set threshold gives us confidence to reject the null hypothesis
- One should check the assumptions –both statistical and experimental ones to interpret correctly.