

Data snooping

Elena Tuzhilina

Stanford University, Department of Statistics

elenatuz@stanford.edu

February 3, 2021

Definition

Data snooping is manipulating data or analysis to artificially get statistically significant results

Definition

Data snooping is manipulating data or analysis to artificially get statistically significant results

Alternative names:

- data dredging
- data torturing
- data fishing
- data butchery
- p-hacking



IF YOU TORTURE THE
DATA LONG ENOUGH, IT
WILL CONFESS.

Significance testing





Null hypothesis: there is no effect

Test result: Reject/Fail to Reject

Significance testing

Null hypothesis: there is no effect

Test result: Reject/Fail to Reject

	H_0 is true	H_0 is false
Fail to reject H_0		
Reject H_0		

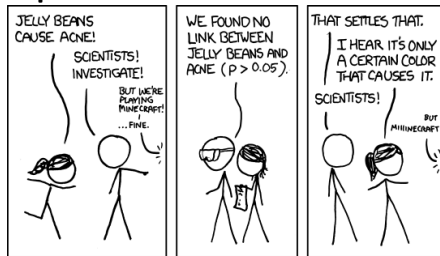
Significance testing

Null hypothesis: there is no effect

Test result: Reject/Fail to Reject

	H_0 is true	H_0 is false
Fail to reject H_0	☹️	☹️
Reject H_0	☹️	😊

Example:



<https://xkcd.com/882/>

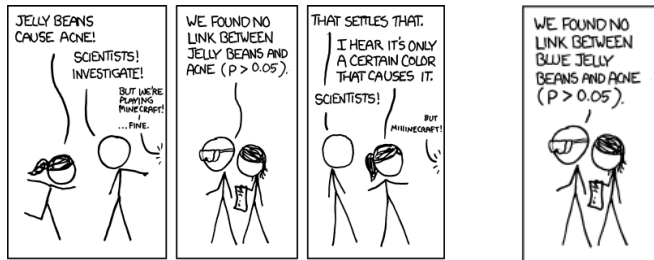
Significance testing

Null hypothesis: there is no effect

Test result: Reject/Fail to Reject

	H_0 is true	H_0 is false
Fail to reject H_0	☹️	☹️
Reject H_0	☹️	😊

Example:



<https://xkcd.com/882/>

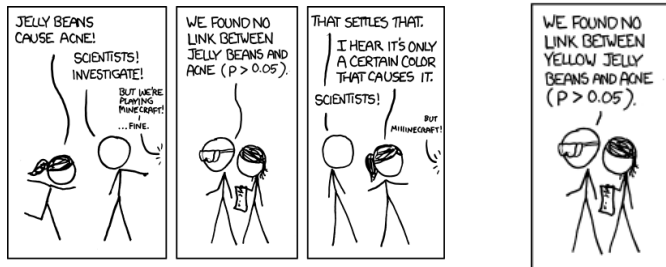
Significance testing

Null hypothesis: there is no effect

Test result: Reject/Fail to Reject

	H_0 is true	H_0 is false
Fail to reject H_0	☹️	☹️
Reject H_0	☹️	😊

Example:



<https://xkcd.com/882/>

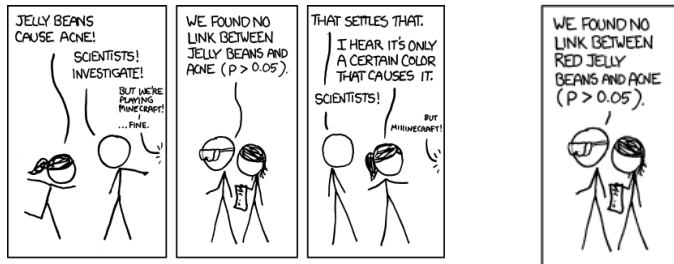
Significance testing

Null hypothesis: there is no effect

Test result: Reject/Fail to Reject

	H_0 is true	H_0 is false
Fail to reject H_0	☹️	☹️
Reject H_0	☹️	😊

Example:



<https://xkcd.com/882/>

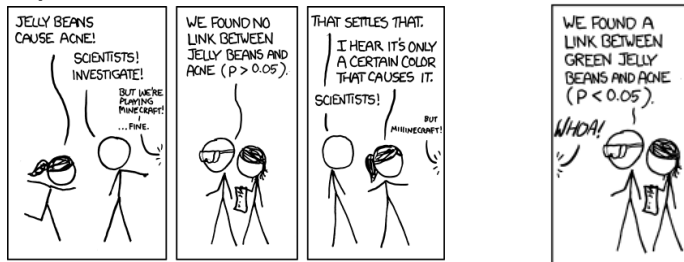
Significance testing

Null hypothesis: there is no effect

Test result: Reject/Fail to Reject

	H_0 is true	H_0 is false
Fail to reject H_0	☹️	☹️
Reject H_0	☹️	😊

Example:



<https://xkcd.com/882/>

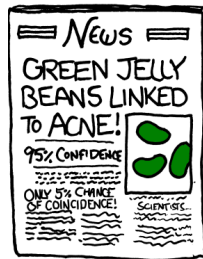
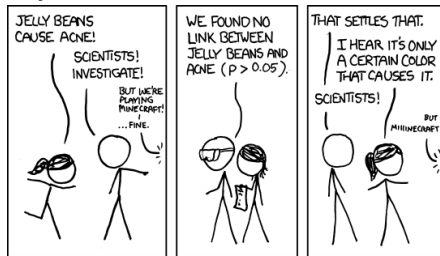
Significance testing

Null hypothesis: there is no effect

Test result: Reject/Fail to Reject

	H_0 is true	H_0 is false
Fail to reject H_0	☹️	☹️
Reject H_0	☹️	😊

Example:







<https://xkcd.com/882/>

Significance testing

Null hypothesis: there is no effect

Test result: Reject/Fail to Reject

	H_0 is true	H_0 is false
Fail to reject H_0		
Reject H_0		

Example:

If jelly beans are not linked to acne, then each test has 5% chance to be significant.

Significance testing

Null hypothesis: there is no effect

Test result: Reject/Fail to Reject

	H_0 is true	H_0 is false
Fail to reject H_0	☹️	☹️
Reject H_0	☹️	😊

Example:

If jelly beans are not linked to acne, then each test has 5% chance to be significant. Among 20 tests there is $1 - 0.95^{20} \approx 64\%$ chance to observe *one or more* significant test.

Significance testing

Null hypothesis: there is no effect

Test result: Reject/Fail to Reject

	H_0 is true	H_0 is false
Fail to reject H_0	☹️	☹️
Reject H_0	☹️	😊

Example:

If jelly beans are not linked to acne, then each test has 5% chance to be significant. Among 20 tests there is $1 - 0.95^{20} \approx 64\%$ chance to observe *one or more* significant test. **Just by chance!**

Cornell's Food and Brand Lab experiment

Experiment: Charge half people full price, gave other half a 50% discount.

Data:

- amount of food
- food items
- eating alone/in a group
- ...

Hypothesis: There is an effect of price on the amount that people eat.

Result: No statistical difference

Cornell's Food and Brand Lab experiment

Experiment: Charge half people full price, gave other half a 50% discount.

Data:

- amount of food
- food items
- eating alone/in a group
- ...

Hypothesis: There is an effect of price on the amount that people eat.

Result: No statistical difference \implies do some digging!

Experiments:

- check smaller groups
- test other variables

Examples of significant observations: food tastes worse when buffet prices are low, men eat more when dining with women in order to impress them.

Cornell's Food and Brand Lab experiment

Experiment: Charge half people full price, gave other half a 50% discount.

Data:

- amount of food
- food items
- eating alone/in a group
- ...

Hypothesis: There is an effect of price on the amount that people eat.

Result: No statistical difference \implies do some digging!

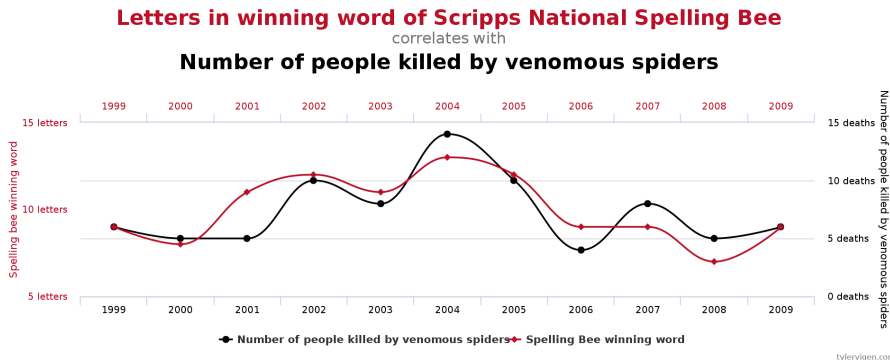
Experiments:

- check smaller groups
- test other variables

Examples of significant observations: food tastes worse when buffet prices are low, men eat more when dining with women in order to impress them.

The significant results were reported without the context of all non-significant ones

Scripps National Spelling Bee's example

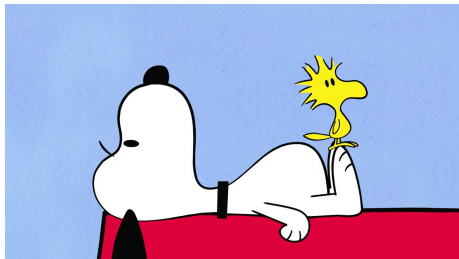


There is a correlation between the number of letters in Scripps National Spelling Bee's winning word and the number of people in the United States killed by venomous spiders!

<https://en.wikipedia.org/wiki/Data-dredging>

Data snooping summary

Data snooping is probing the data in unplanned ways, finding and reporting an “attractive” result, without accurately conveying the course of analysis

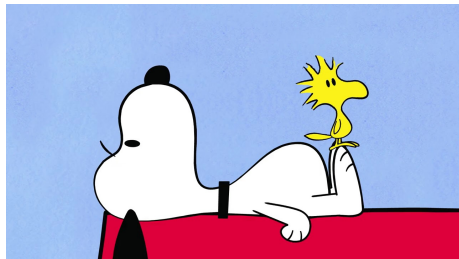


Data snooping summary

Data snooping is probing the data in unplanned ways, finding and reporting an “attractive” result, without accurately conveying the course of analysis

Reasons:

- Lack of statistical knowledge
- Belief in a specific theory
- Desire to make the result publishable

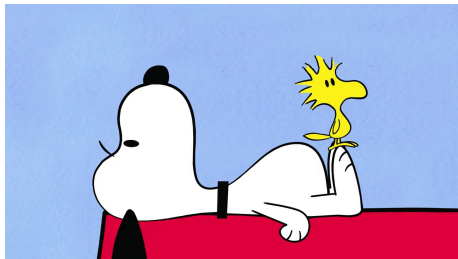


Data snooping summary

Data snooping is probing the data in unplanned ways, finding and reporting an “attractive” result, without accurately conveying the course of analysis

Reasons:

- Lack of statistical knowledge
- Belief in a specific theory
- Desire to make the result publishable



Suggested remedies:

- Use randomized out-of-sample tests (cross validation)
- Use p-value correction for multiple testing
- Choose the analysis and assumptions before you see the data

Questions

- ① Have you ever encountered data snooping in your field? Please provide an example.
- ② Is it possible to track data snooping and who should take the responsibility for preventing data snooping (publisher, institution, etc.)?
- ③ How can the data science community help prevent data snooping?

References:

- ① Data-dredging bias. Catalog of biases
- ② Data Snooping. Common mistakes in using statistics
- ③ Wikipedia page on data dredging
- ④ Video on p-hacking. Crash Course Statistics
- ⑤ Video on p-hacking. Statquest
- ⑥ Cornell Food Researcher's Downfall Raises Larger Questions For Science
- ⑦ Here's How Cornell Scientist Brian Wansink Turned Shoddy Data Into Viral Studies About How We Eat
- ⑧ A Credibility Crisis in Food Science