Hypothesis Testing: z test and t-test

Data Science Immersive

May 28, 2019

FLATIRON SCHOOL

Agenda Today

- Introduction to experimental design and hypothesis testing
- Review of the standard normal distribution and z test
- Introduction to t tests
- Learn how to interpret results from a z or t tests

Students will be able to

- Formulate hypothesis in designing an experiment
- Conduct and calculate the z and t statistics
- Differentiate between the z and t tests and when they are applied in specific situations
- Interpret the results from a test

How can we know things?

?

How can we know things?

- Even a more fundamental question: **what does it mean to know something?**

School of thoughts

"I think therefore I am"

"A wise man proportions his belief
to the evidence"

Principles of Empiricism

- Knowledge is a posterior
- Proportion beliefs to evidence
- Tabula Rasa
- Ok to use knowledge from experience to synthesize new knowledge

Hypothesis Formulation

- Any scientific study begins with a question, then a hypothesis, then testing, then a conclusion is made about this hypothesis
- Following Tabula Rasa, if we cannot prove something, we assume it is untrue by default

The null hypothesis: H_0

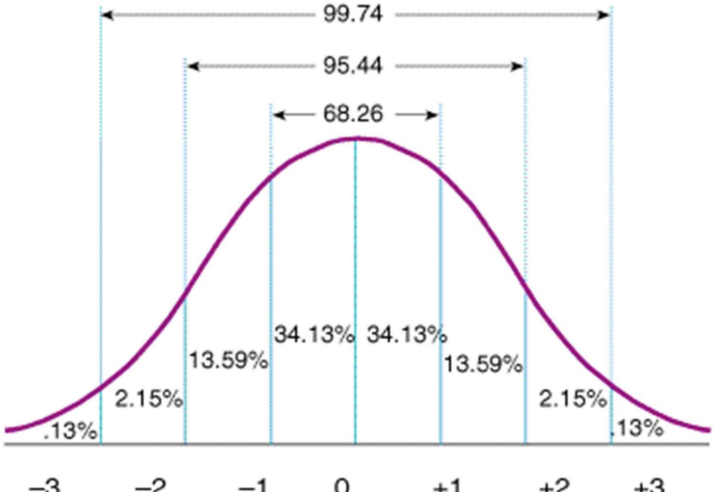
The alternative hypothesis: H_A

Examples of Formulating hypothesis

- Do we frequently engage in self-deception?
- Does smoking correlate with having cancer?
- Is having more colorful graphic design results in more purchases?
- Is allowing direct messages on instagram results in more friendships?

P values and alpha

- As we learned from last week, a p-value corresponds to the area under the curve
- But how we do actually interpret it?
- P value represents: assuming the null is true, the probability of observing the results from the experiment



Alpha

- While p value is acquired by conducting the experiments, alpha is decided by the researcher
- Traditionally, an alpha of <.05 is chosen but other conventions may apply in different fields
- The interpretation is: if the probability of achieving the result is less than .05 under the null hypothesis, then we say the result cannot be due to random chance and it must be statistically significant

Hypothesis testing with z test

- Review of z scores
  - A z score represents the number of standard deviations above or below the mean
- Review of the Z distribution
  - Also known as the standard normal distribution
  - Has a mean of 0 and a standard deviation of 1

An example of hypothesis testing with z test

- Is our class significantly different from the population in terms of egomaniac score?
- https://brainfall.com/quizzes/are-you-an-egomaniac/
- Record your answer and put it in this spreadsheet - we will conduct testing tomorrow to figure out whether our class is more or less egomaniac from the general population
- https://docs.google.com/spreadsheets/d/1de8EdVjwU2YTmYlwKw-J06ZO4A9FBHxUEHjiyajxJfU/edit#gid=0