Quantitative Research Methods

Probability and Statistics for Data Science CSE594 - Spring 2016

Quantitative Research Methods

Probability and Statistics for Data Science CSE594 - Spring 2016

The Scientific Method

Make

Observations

What do I see in nature? This can be from one's

own experiences, thoughts

or reading.

Develop General Theories

General theories must be consistent with most or all available data and with other current theories.

Gather Data to Test Predictions

Relevant data can come from the literature, new observations or formal experiments. Thorough testing requires replication to verify results.

Develop Testable Predictions

Refine, Alter,

Expand or Reject

Hypotheses

If my hypothesis is correct, then I expect a, b, c, ...

(Garland, 2015)

Formulate Hypotheses

Think of

Interesting

Questions

Why does that

pattern occur?

What are the general causes of the phenomenon I am wondering about?

Hypothesis -- something one asserts to be true.

Classical Approach:

 H_0 : null hypothesis -- some "default" value (usually that one's hypothesis is false)

 H_1 : the alternative -- usually that one's "hypothesis" is true

Hypothesis -- something one asserts to be true.

Classical Approach:

 H_0 : null hypothesis -- some "default" value (usually that one's hypothesis is false)

 H_1 : the alternative -- usually that one's "hypothesis" is true

Goal: Use probability to determine if we can "reject the null"(H_0) in favor of H_1 . "There is less than a 5% chance the null is true" (i.e. 95% alternative is true).

Hypothesis -- something one asserts to be true.

Classical Approach:

 H_0 : null hypothesis -- some "default" value (usually that one's hypothesis is false)

 H_1 : the alternative -- usually that one's "hypothesis" is true

Goal: Use probability to determine if we can "reject the null"(H_0) in favor of H_1 . "There is less than a 5% chance the null is true" (i.e. 95% alternative is true).

Example: Hypothesize a coin is biased.

 H_0 : the coin is not biased (i.e. flipping n times results in a Binomial(n, 0.5))



*H*₀: *null hypothesis* -- some "default" value (usually that one's hypothesis is false)

 H_1 : the alternative -- usually that one's "hypothesis" is true

More formally: Let *X* be a random variable and let *R* be the range of X. $R_{reject} \subseteq R$ is the *rejection region.* If $X \in R_{reject}$ then we reject the null.

Example: Hypothesize a coin is biased.

 H_0 : the coin is not biased (i.e. flipping n times results in a Binomial(n, 0.5))



 H_0 : null hypothesis -- some "default" value (usually that one's hypothesis is false) H_1 : the alternative -- usually that one's "hypothesis" is true

More formally: Let *X* be a random variable and let *R* be the range of X. $R_{reject} \subseteq R$ is the *rejection region.* If $X \in R_{reject}$ then we reject the null.

in the example, if n = 1000, then then $R_{reject} = [0, 469] \cup [531, 1000]$

Example: Hypothesize a coin is biased.

 H_0 : the coin is not biased (i.e. flipping n times results in a Binomial(n, 0.5))



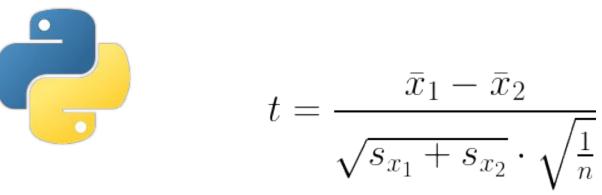
Example: Communities with higher population have different amounts of violent crimes (per capita) than those with lower population.

Assignment 1, Programming Problem "C) 9."



Example: Communities with higher population have different amounts of violent crimes (per capita) than those with lower population.

Assignment 1, Programming Problem "C) 9."



t statistic for 2 iid (independent, identically distributed) samples

Important logical question:

Does failure to reject the null mean the null is true?

