Quantitative Research Methods

Probability and Statistics for Data Science

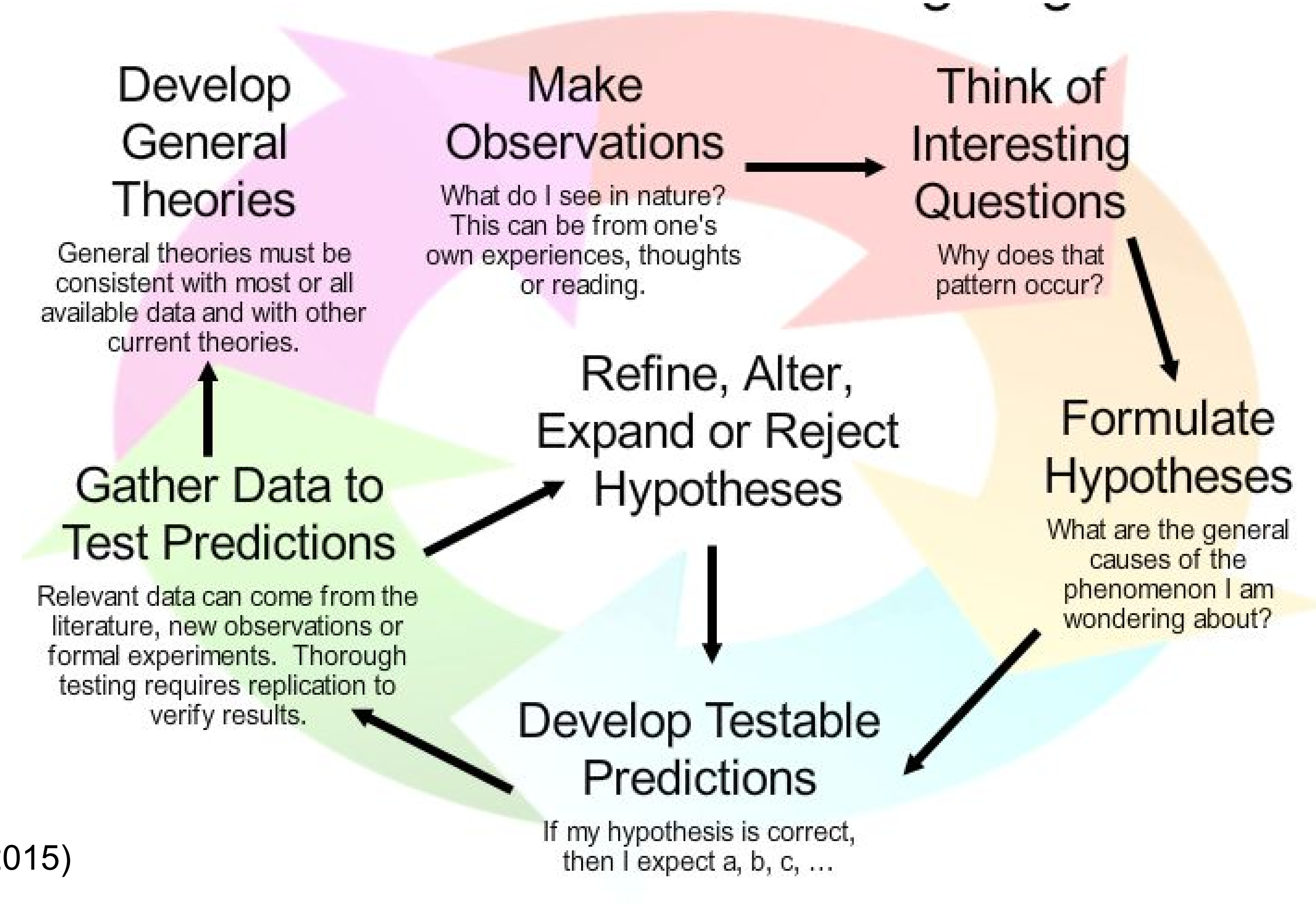
CSE594 - Spring 2016

Quantitative Research Methods

Probability and Statistics for Data **Science**

CSE594 - Spring 2016

The Scientific Method



(Garland, 2015)

Hypothesis Testing

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 Testing

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 Testing

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))

Hypothesis Testing

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. $R_{reject} \subset R$ 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))

Hypothesis Testing

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. $R_{reject} \subset R$ 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))

Hypothesis Testing

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

Assignment 1, Programming Problem “C) 9.”

Hypothesis Testing

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

Assignment 1, Programming Problem “C) 9.”

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_{x_1} + s_{x_2}} \cdot \sqrt{\frac{1}{n}}}$$

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

Hypothesis Testing

Important logical question:

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