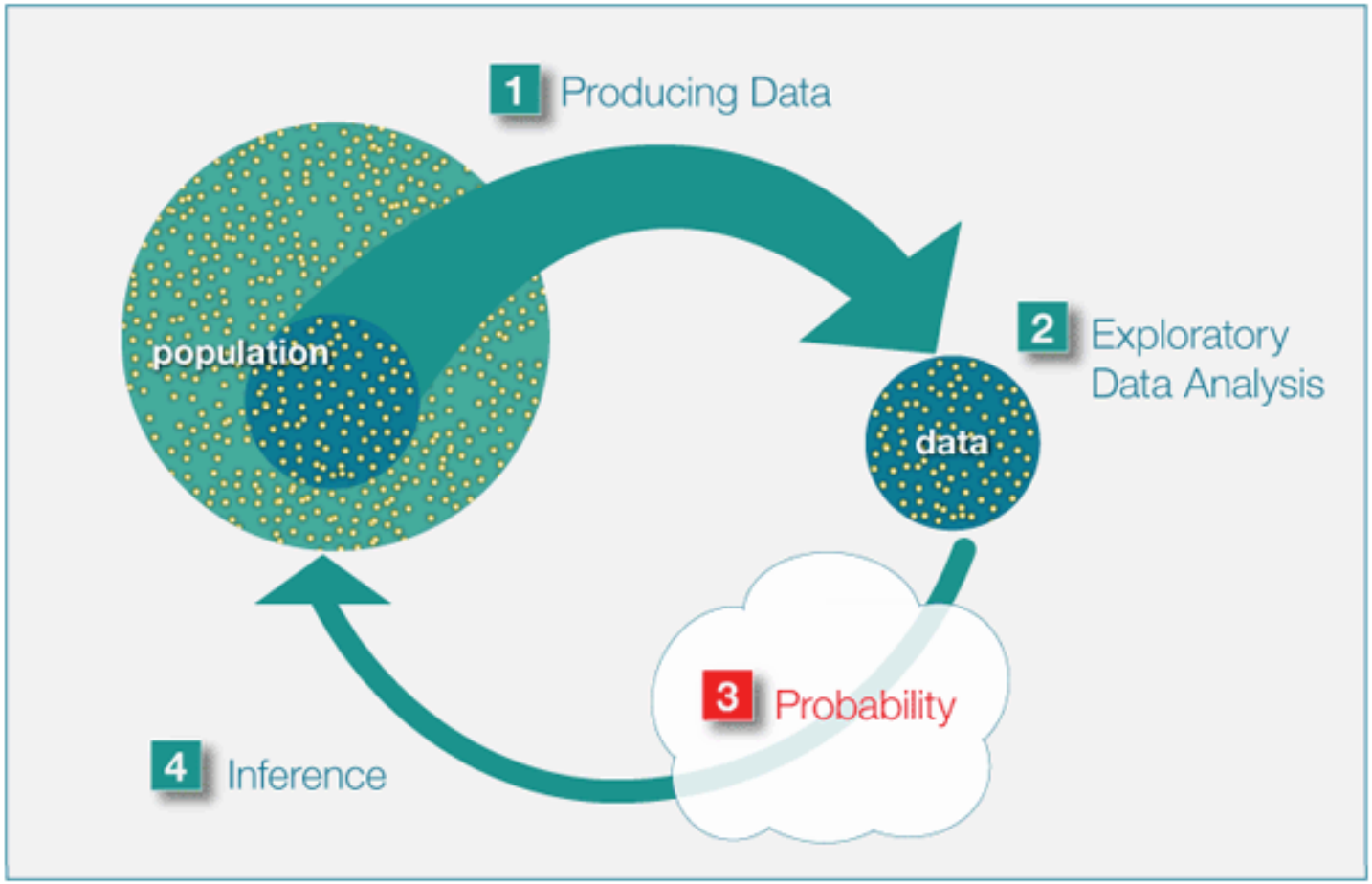
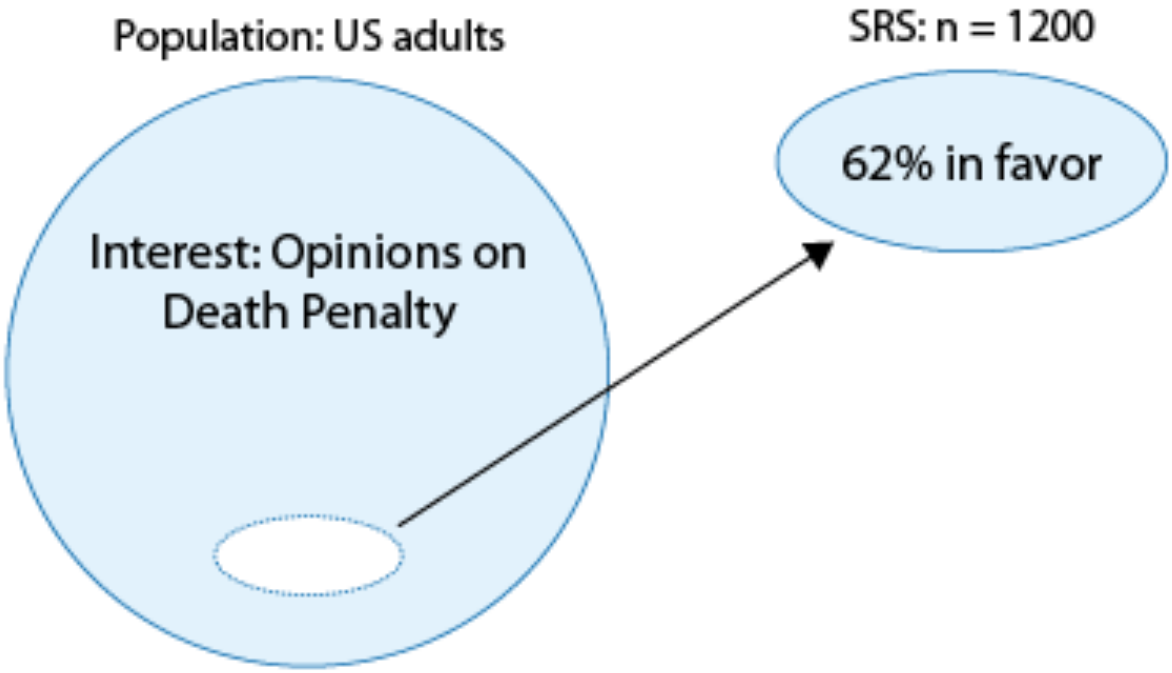
WESLEYAN

UNIVERSITY

Probability

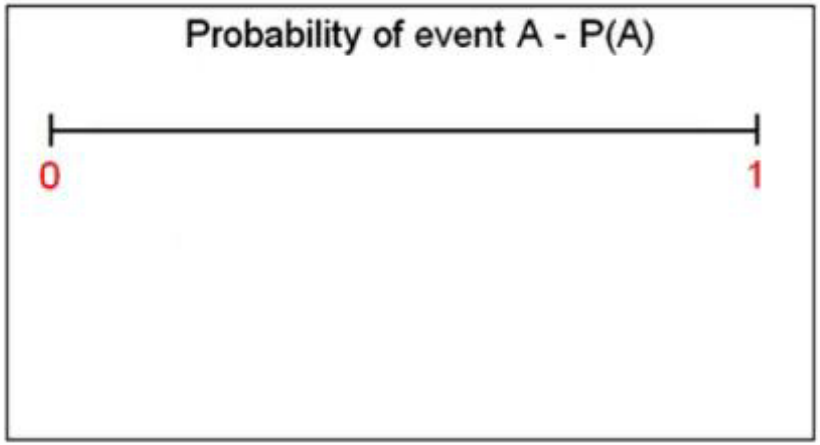


Example

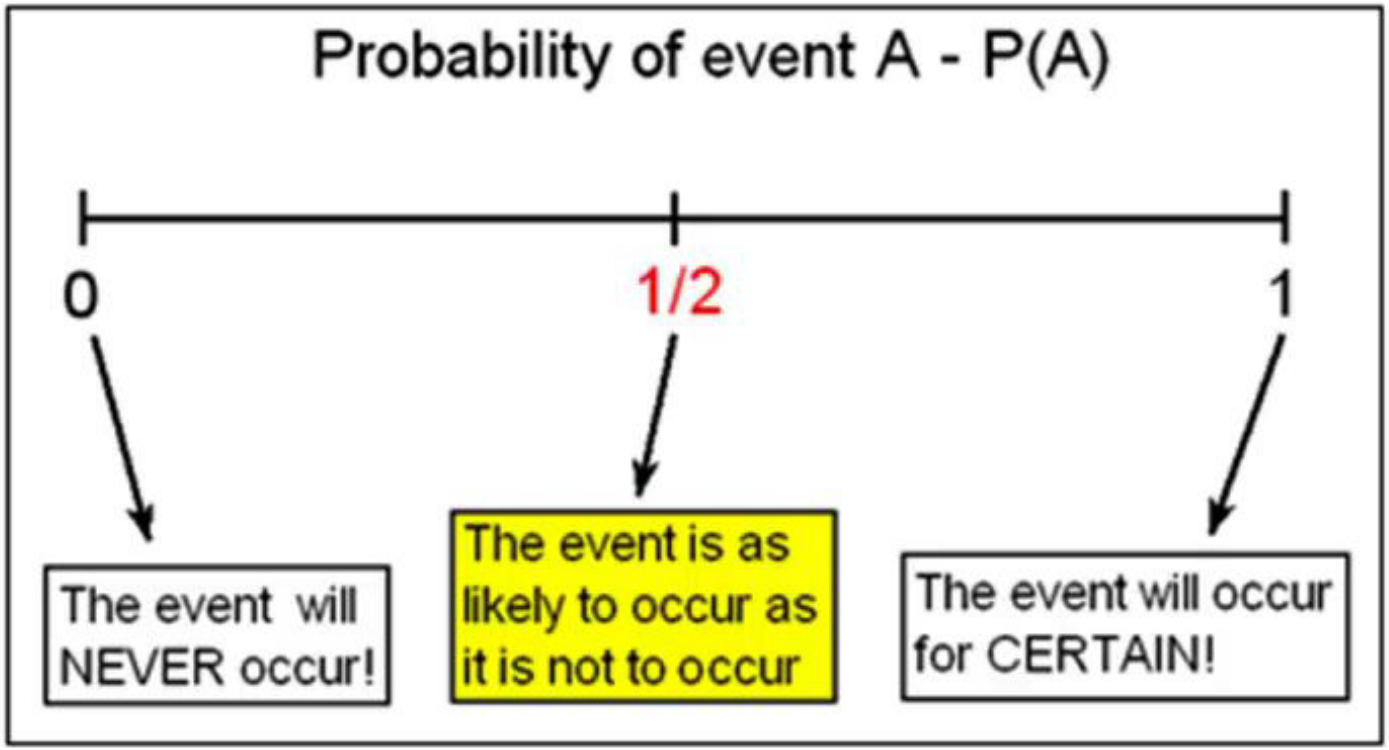


Probability is the Likelihood of Something Happening

The "probability" of an event tells us how likely it is that the event will occur.

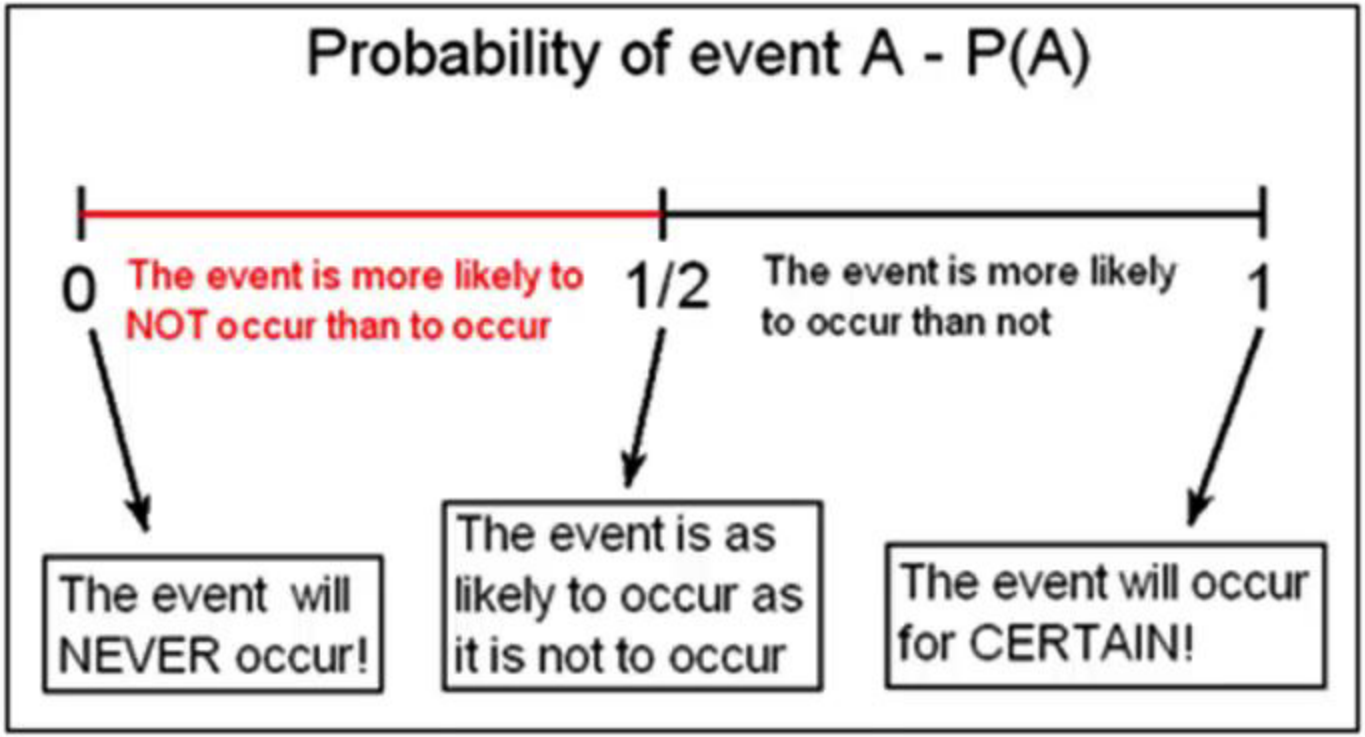


What is Probability?

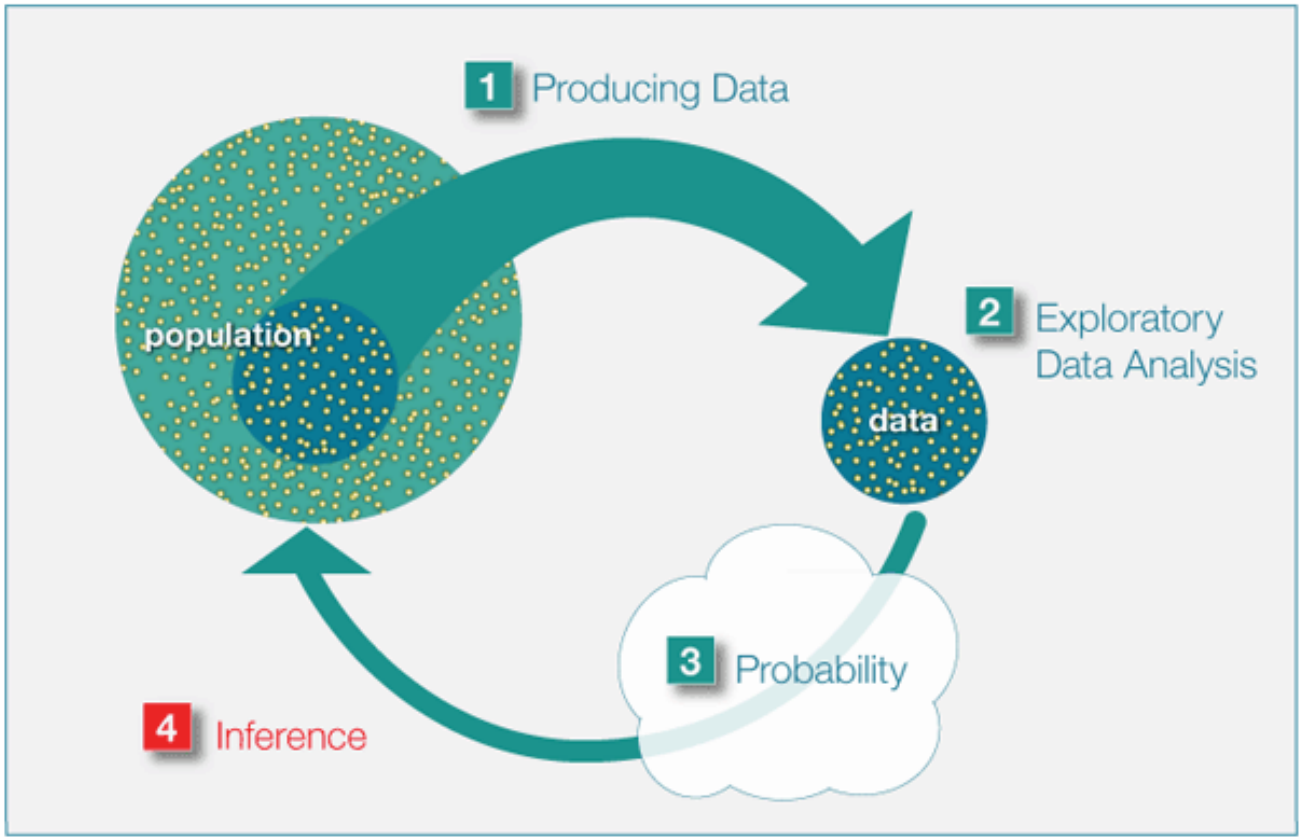


What is Probability?

WESLEYAN UNIVERSITY

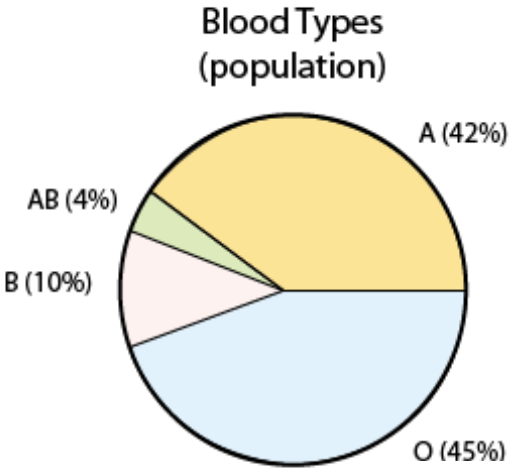


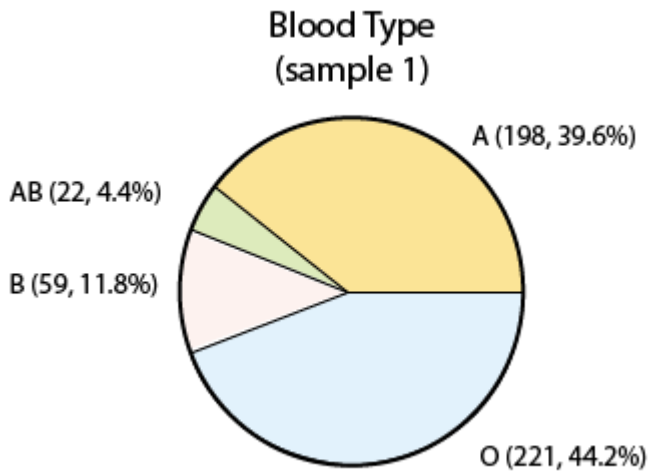
Inference

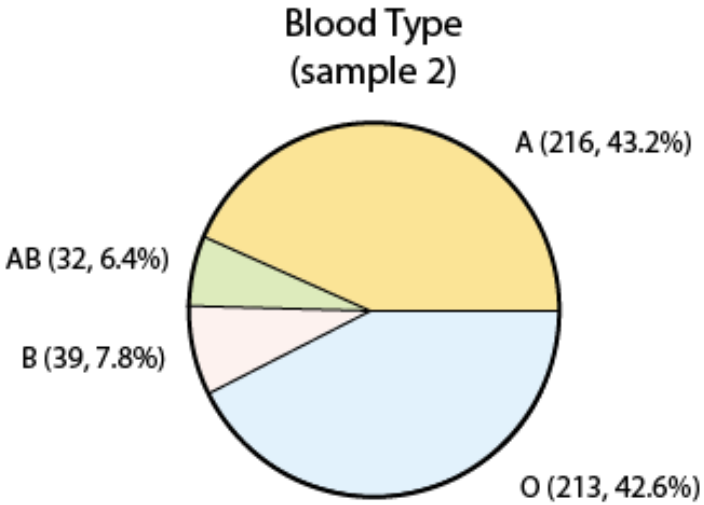


Sample Variability Example

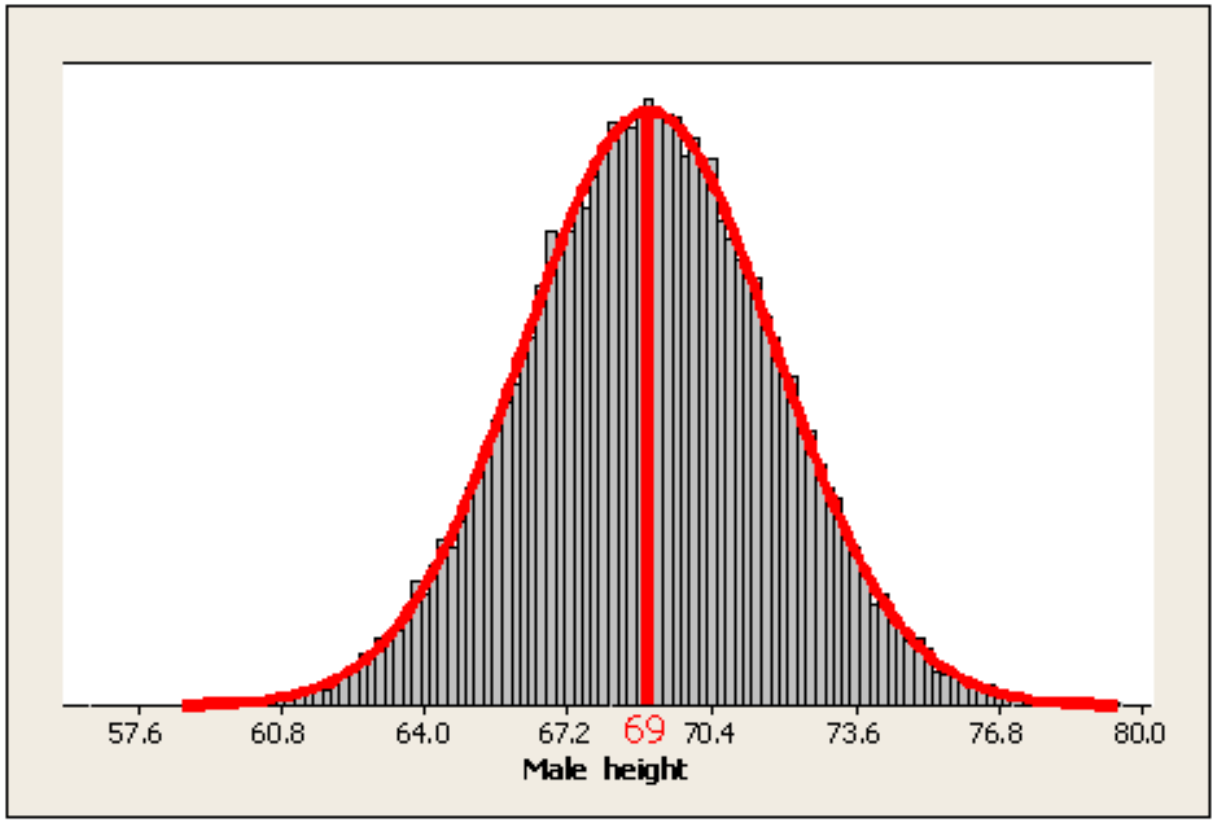
WESLEYAN UNIVERSITY







Male Height: US population

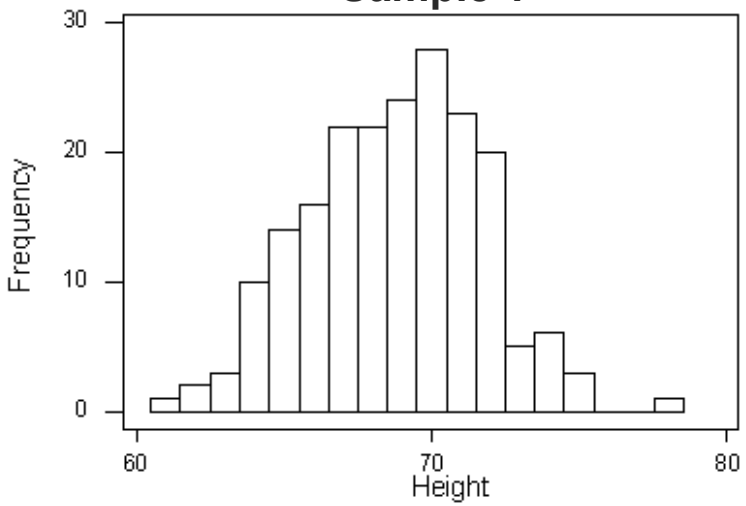


$\mu = 69 \qquad \sigma = 2.8$ inches.

Male Height: 2 samples

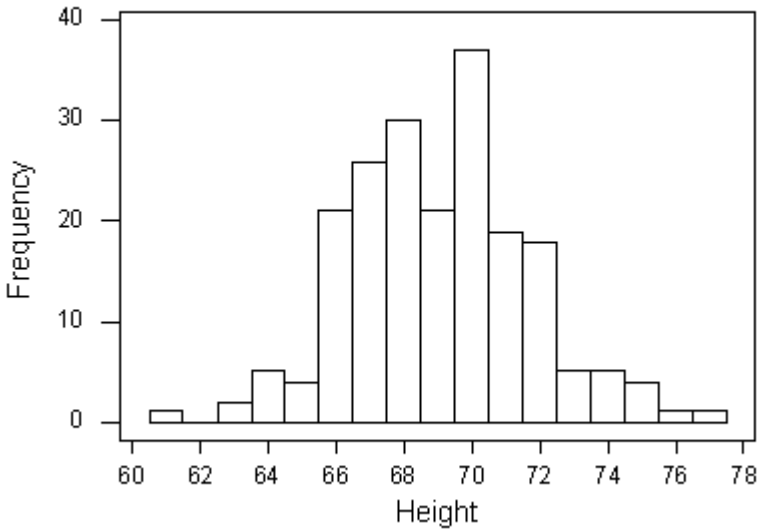
WESLEYAN UNIVERSITY

Sample 1



$\bar{x} = 68.7$ s = 2.95 inches

Sample 2



$\bar{x} = 69.065$ s = 2.659 inches

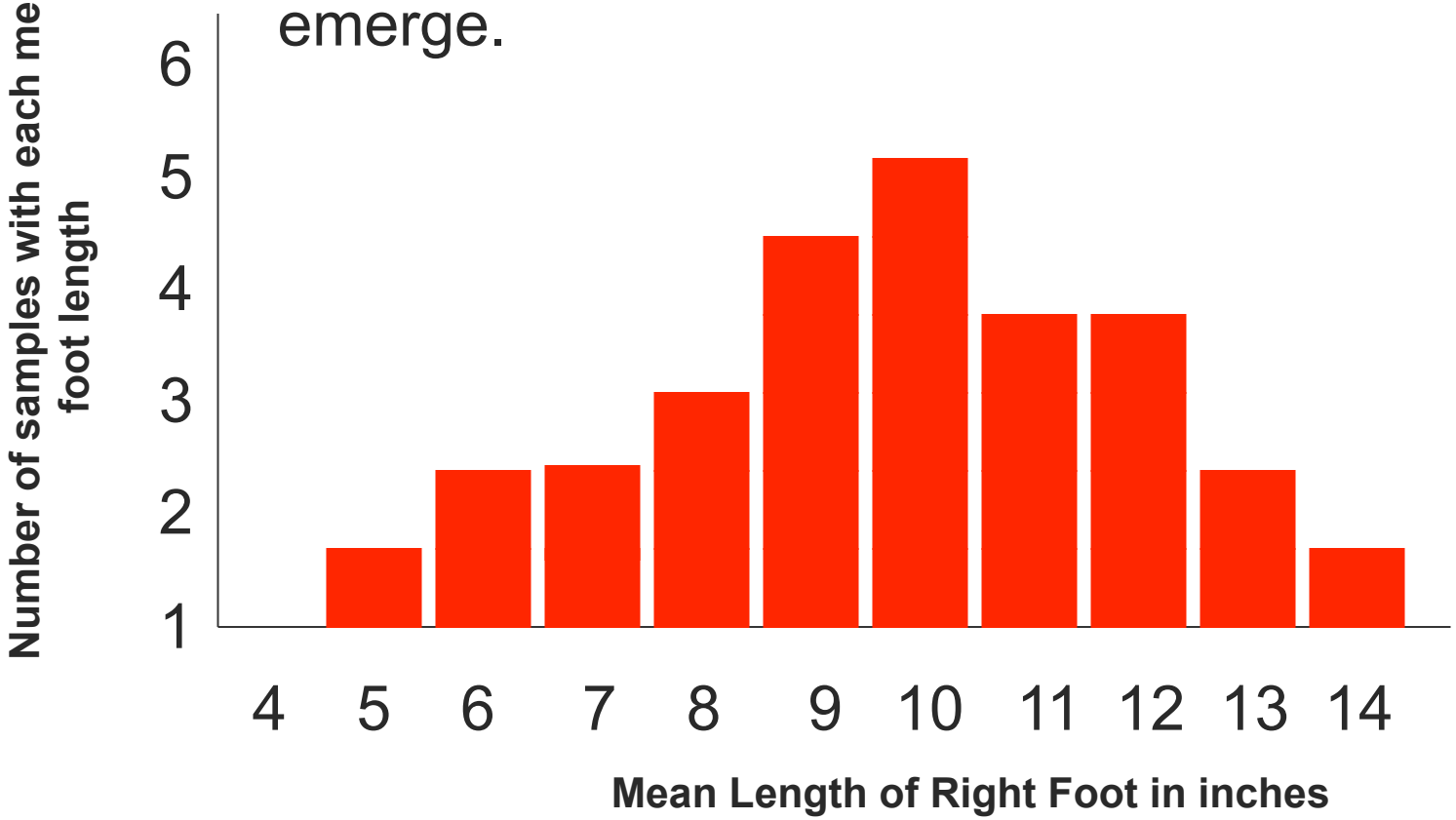
$\mu = 69 \quad \sigma = 2.8$ inches.

Parameter vs. Statistic

WESLEYAN UNIVERSITY

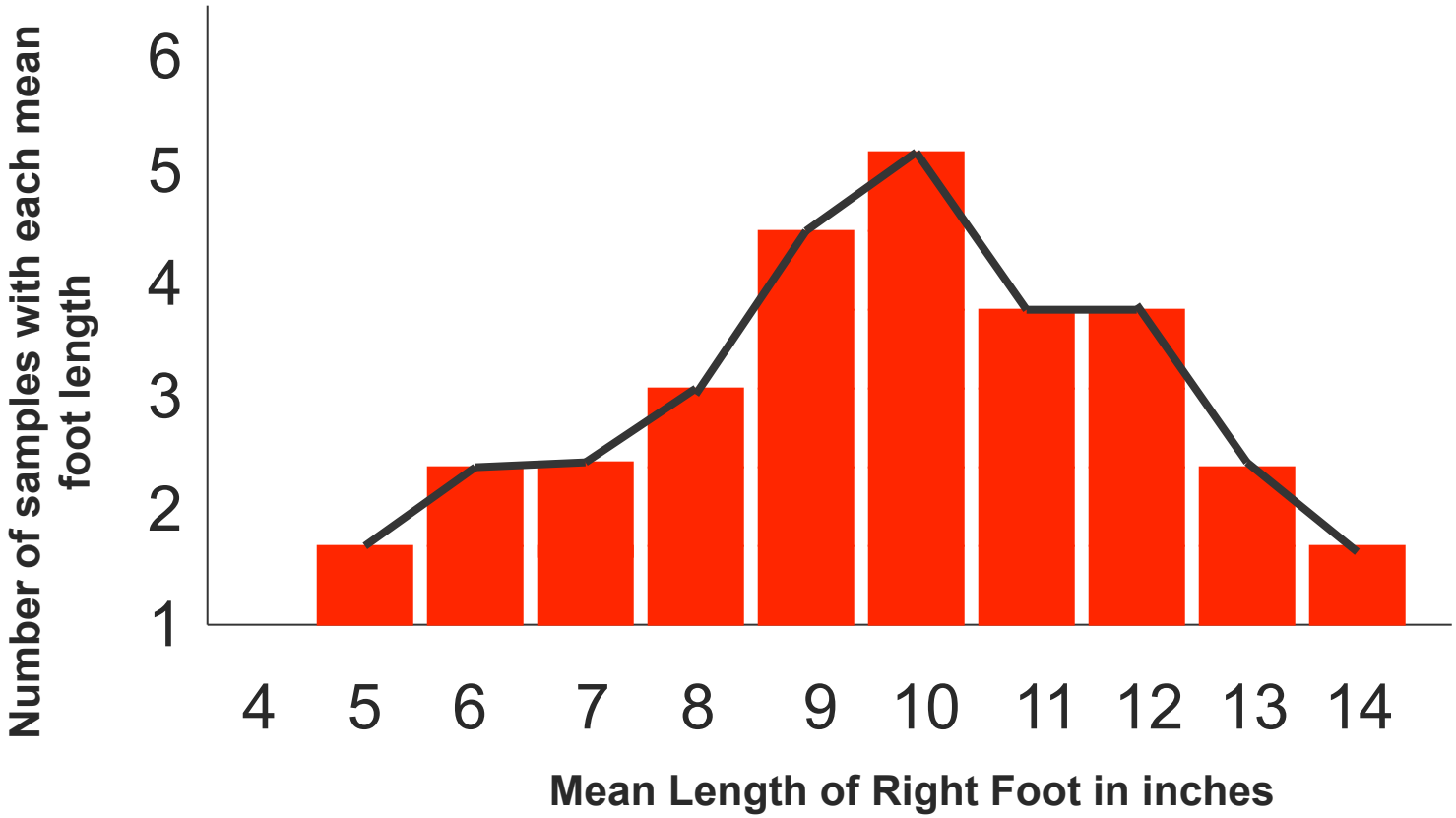
	(Population) Parameter	(Sample) Statistic
Proportion	p	$\hat{p}$
Mean	μ	$\bar{x}$
Standard Deviation	σ	s

As we continued to plot the mean foot length of each random sample, a pattern would begin to emerge.



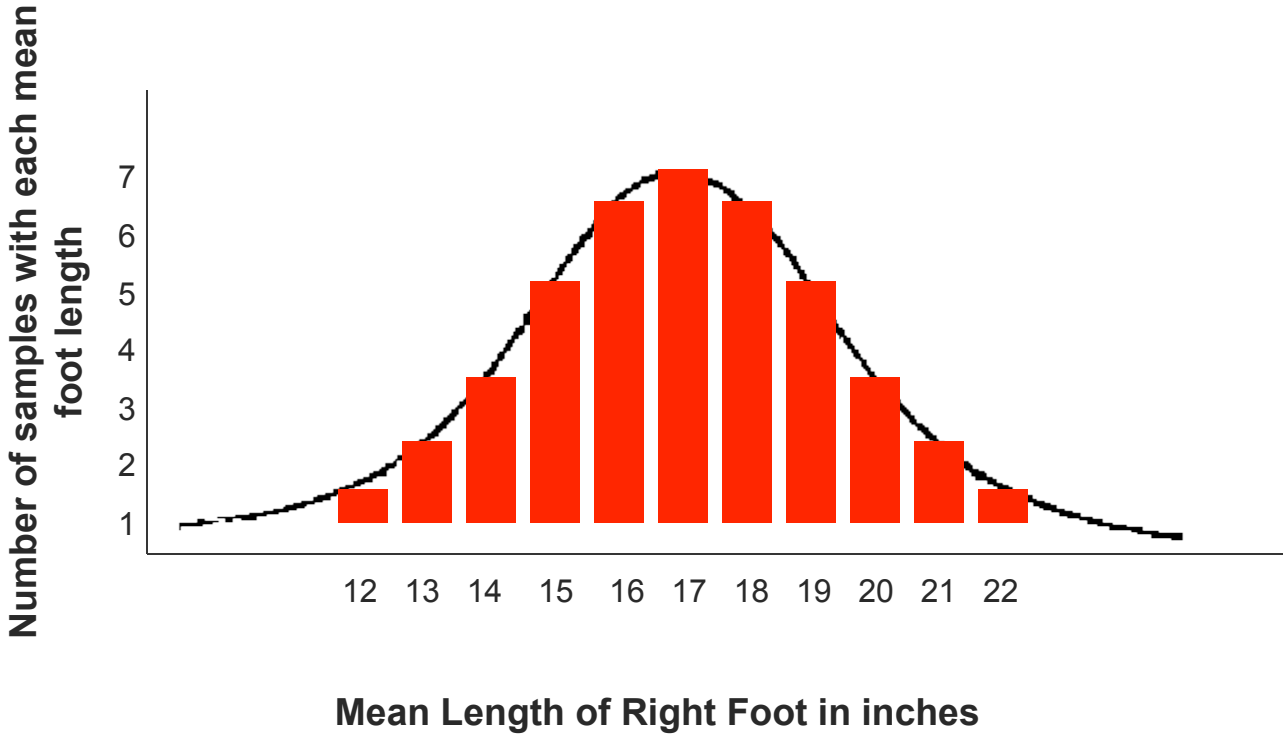
Sampling Distribution

WESLEYAN UNIVERSITY



Sampling Distribution

WESLEYAN UNIVERSITY



Steps in Hypothesis Testing

WESLEYAN UNIVERSITY

Specifying the null (H_o) and alternate (H_a) hypothesis

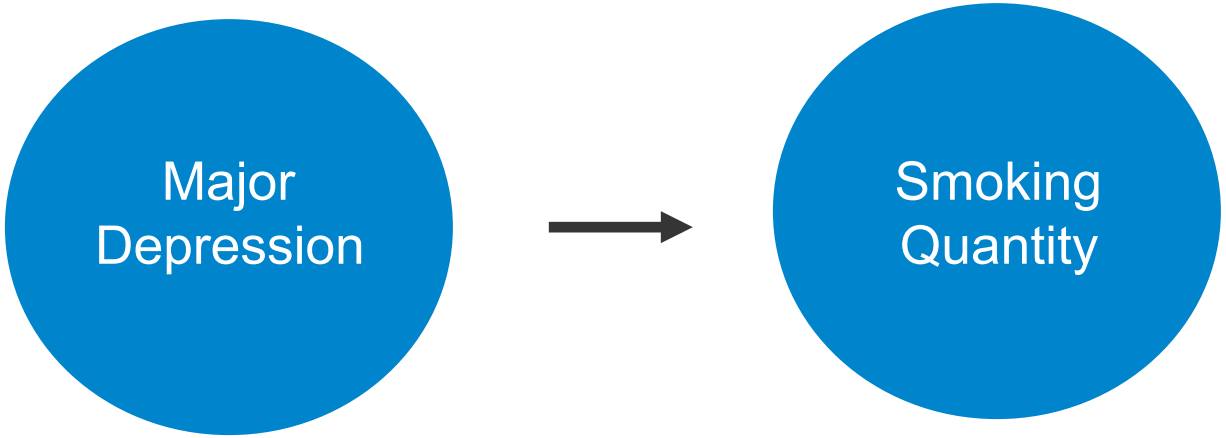
Choosing a sample

Assessing the evidence

Drawing conclusions

Specifying the Null and Alternate Hypothesis

WESLEYAN UNIVERSITY



H_o: There is no difference in smoking quantity between smokers with and without depression.

H_a: There is a difference in smoking quantity between smokers with and without depression.

Choosing a Sample

WESLEYAN UNIVERSITY

The NESARC, a representative sample of 43,093 adults in the U.S.

1) current daily smokers

2) age 18 to 25.

n=1320

Young adult, daily smokers with depression smoked an average of **13.9** cigarettes per day (s.d. 9.2).

Young adult daily smokers without depression smoked on average **13.2** cigarettes per day (s.d 8.5)

Assessing the Evidence

WESLEYAN UNIVERSITY

Is there strong enough evidence against the null hypothesis?

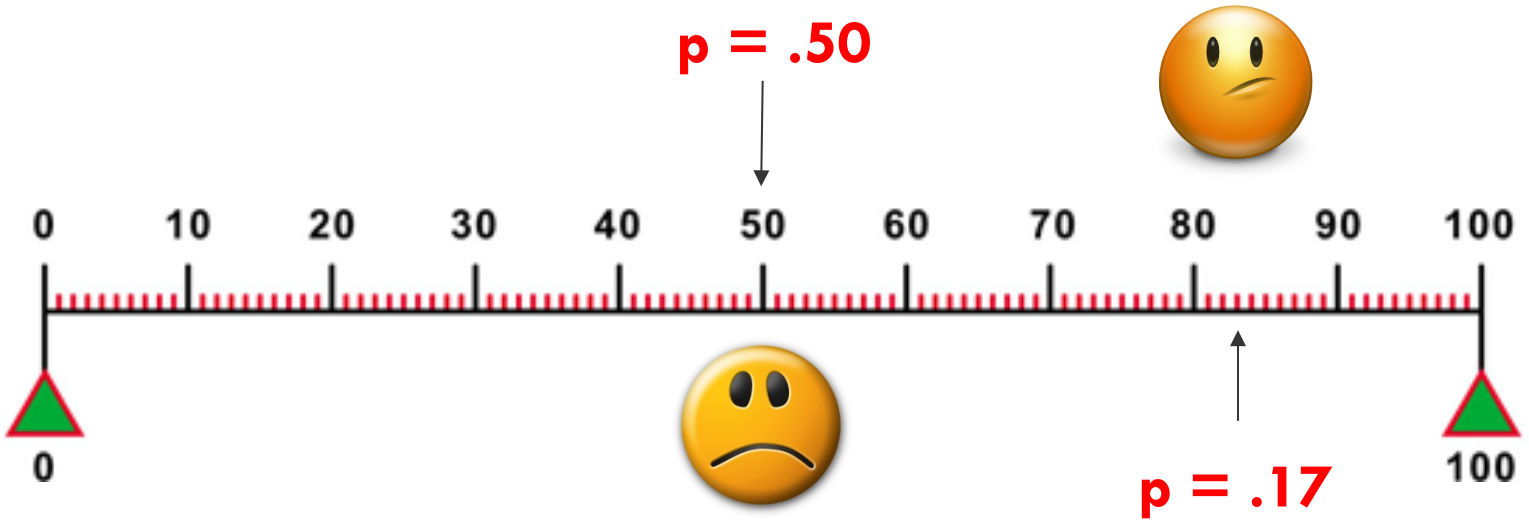
Smokers with depression = **13.9** cigarettes per day (s.d. 9.2).

Smokers without depression = **13.2** cigarettes per day (s.d 8.5)

A difference of 0.7 cigarettes per day

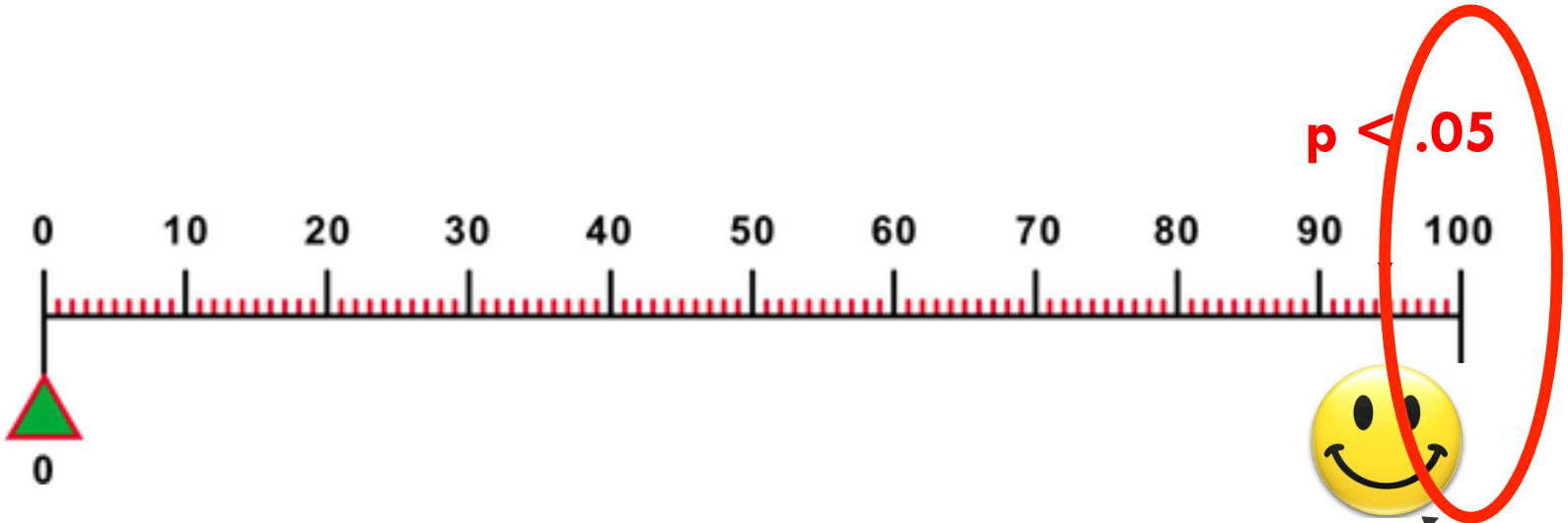
The probability of getting a difference of this size is roughly **0.17**

Drawing Conclusions



Translation: If we took 100 random samples from our population we would be wrong 17 out of 100 times if we rejected the null hypothesis and said that the difference in smoking quantity was difference for smokers with depression compared to those without.

p-value



statistically significant

reject H_o and accept H_a

p =.05 is the Type I error rate

If I change my example…

WESLEYAN UNIVERSITY

The NESARC, a representative sample of 43,093 adults in the U.S.

1) current daily smokers 1) smoked in past 12 months

2) age 18 to 25.

n=1320 n=1706

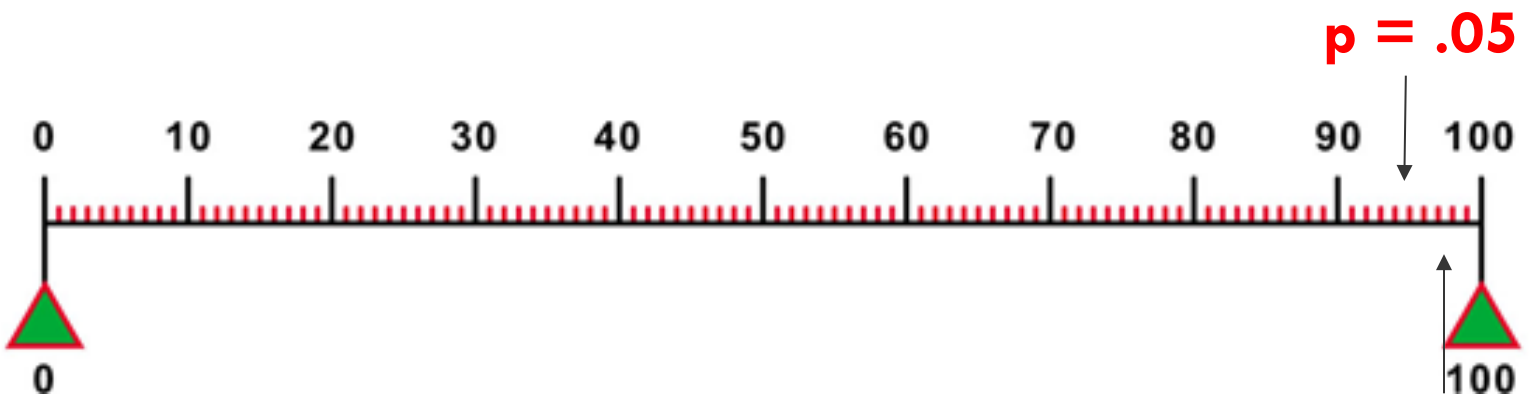
Young adults with depression smoked an average of **351.7** cigarettes per month (s.d. 300.0).

Young adult without depression smoked on average **313.5** cigarettes per month (s.d 268.2)

A difference of 38.2 cigarettes per month (almost 2 packs) p=.0285

p-value < .05

WESLEYAN UNIVERSITY



p = .0285

Translation: If we reject the null hypothesis and say that there is a difference between the average number of cigarettes smoked per month among young adults with and without depression, we would be wrong fewer than 3 out of 100 times. We would be correct more than 97% of the time.

Variable types and inferential tools

WESLEYAN UNIVERSITY

		Response	
		Categorical	Quantitative
Explanatory	Categorical	C → C	C → Q
	Quatitative	Q → C	Q → Q