

# SAMPLING

MPA 630: Data Science for Public Management

November 1, 2018

*Fill out your reading report  
on Learning Suite*

# PLAN FOR TODAY

---

## Exam 2

---

**Sampling vocabulary**

**Sampling in real life**

**Sampling with computers**

# EXAM 2

# SAMPLING VOCABULARY

# DEFINING THE POPULATION

---

## **Population**

A collection of things  
in the world

## **Population parameter**

Something we want to  
know about the population

# COUNTING THE POPULATION

---

## Census

Count every single thing in the whole population

## Sampling

Select parts of the population and count those

# MEASURE THE SAMPLE

---

## Sample statistic or point estimate

The population parameter,  
but for the sample

Uses the hat sign; p-hat

$p$        $\hat{p}$

# IS THE SAMPLE GOOD?

---

## Representativeness

Does the sample look like the population?

## Bias and randomness

Does every part of the population have the same chance of being sampled?

## Generalizability

Is  $\hat{p}$  a good guess of  $p$ ?



# WHY EVEN DO THIS?

---

Censuses are expensive  
and often impossible

**If a sample is taken at random...**

**...it will be unbiased and representative...**

**...and the sample estimates can  
generalize to the whole population**  
(within a confidence interval)

# WHAT IF \*YOU\* AREN'T COUNTED?

---

**Sampling gets us accurate estimates  
of population parameter—even if  
samples seem small!**

Statistical power

# SAMPLING IN REAL LIFE

# M & M SAMPLING

---

**Define the population**

What thing are we counting?

What parameter are we measuring?

**Count the population**

Census or sample?

**Measure the sample**

What is our  $\hat{p}$ ?

**Is the sample good?**

Is the sample representative?

Is the sample biased?

Is  $\hat{p}$  a good guess?







# THE TRUE *p*

---

Plant	City	Blue	Brown	Green	Orange	Red	Yellow
CLV	Cleveland, OH	<b>20.7%</b>	12.4%	<b>19.8%</b>	<b>20.5%</b>	13.1%	13.5%
HKP	Hackettstown, NJ	<b>25.0%</b>	12.5%	<b>12.5%</b>	<b>25.0%</b>	12.5%	12.5%

“Our color blends were selected by conducting consumer preference tests, which indicate the assortment of colors that pleased the greatest number of people and created the most attractive overall effect.”

“Each large production batch is blended to those ratios and mixed thoroughly. However, since the individual packages are filled by weight on high-speed equipment, and not by count, it is possible to have an unusual color distribution”

# IMPROVING $\hat{p}$

---

**What can we do to get a better estimate of the whole population of M&Ms?**

More samples?

Bigger samples?

Bigger sample size = better sampling

# SAMPLING WITH COMPUTERS