Sampling Variability and the Central Limit Theorem

Part of statistics is the task of determining parameters

Parameters: A numerical summary of a population

μ: the average price of a Seattle home
μ: Average salary of a UW professor ☆
μ: Average clutch size of a penguin clutch
p: proportion of people with some genetic condition.

Take a sample of Seattle homes, UW professors, penguin clutches, people.

From the samples we calculate <u>statistics</u> that represent the parameters.

A <u>parameter</u> is a single fixed value that we want to find.

Our statistic represents the sample, but is wrong.

↳ Depends on the sample
  ↳ Sample was random
    ↳ Our statistic is <u>also</u> random

The deviation of the statistic from the parameter is the result of <u>sampling variability</u>.

Sampling variability is predictable (like probability; it can be predicted in the long run...)

Example: Where we want to measure the % of people with ↳ Americans
[illegible] for emergencies

Example: Where we want to measure the % of people with $<\$500$ savings for emergencies

→ Americans

Adults ↗ → People living in America

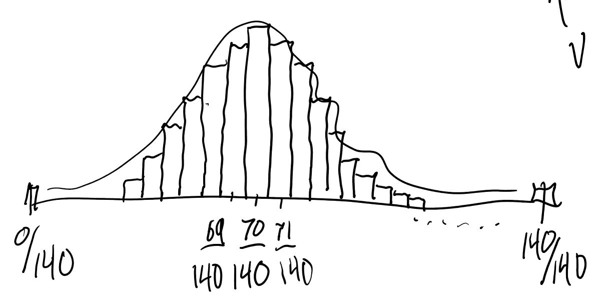
$p = 50\%$ Sample 140 people

We expect about $140 \times .5 = 70$ people with $<\$500$ savings.

Value of statistic $\hat{p}$ will vary: $\frac{70}{140}, \frac{69}{140}, \frac{71}{140}, \frac{2}{140}$

↑ very unlikely

"phat" estimate of p



What we want to do is not estimate the parameter with a single value and instead report a <u>range</u>.

To do this we use (In this context) the <u>Central Limit Theorem</u>

The Central Limit Theorem tells us how the distribution of sample statistics (Specifically means) vary due to sampling variability.

For a population, and a specific parameter, statistics estimating that parameter follow know distributions

Specifically: Estimates of a mean from a population with population mean μ, and population SD of σ, the dist of sample means $\bar{X}$

<u>Central</u> $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$

Central Limit Theorem

$$\bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

We will focus on proportions where

$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

specific special case when data are $\{0,1\}$