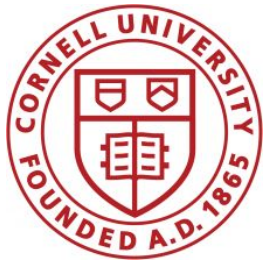


DSFA

Spring 2018

Lecture 28

Sample Size Examples

Announcements

Before break:

- Homework 6 due tonight at 9pm

After break:

- Prelim 2 on 7:30pm on Tuesday 4/10

Discussion Question

A city has 200,000 households. The annual incomes of these households have an average of $65,000 and an SD of $45,000. The distribution of the incomes [pick one and explain]:

(a) is roughly normal because the number of households is large.
(b) is not close to normal.
(c) may be close to normal, or not; we can't tell from the information given.

Questions from the Past Week

- How can we quantify natural concepts like “center” and “variability”?
- Why do many of the empirical distributions that we generate come out bell shaped?
- How is sample size related to the accuracy of an estimate?

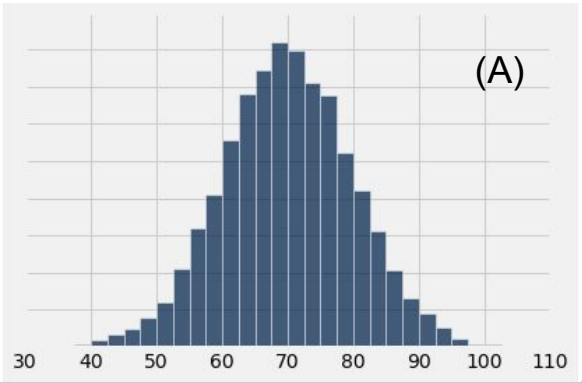
Distribution of the Sample Average

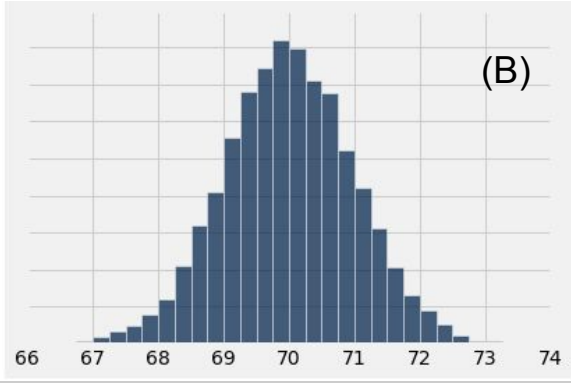
- Fix a large sample size.
- Draw all possible random samples of that size.
- Compute the average of each sample.
- You'll end up with a lot of averages.
- The distribution of those is called the *distribution of the sample average.*
- It's roughly normal, centered at the population average.
- SD = (population SD) / $\sqrt{\text{sample size}}$

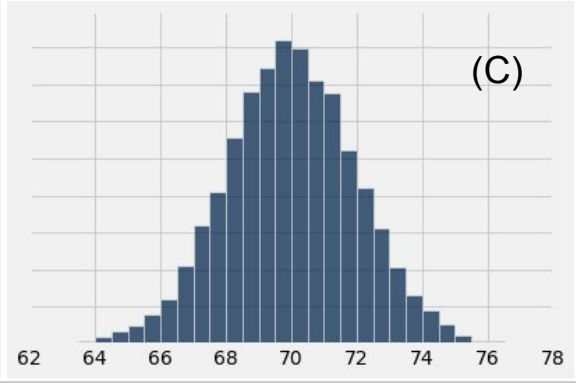
Discussion Question

Professors retire at age 70, on average; the distribution of retirement ages has SD 10.

One of the histograms below is the empirical distribution of the age at which professors retire. Which one?

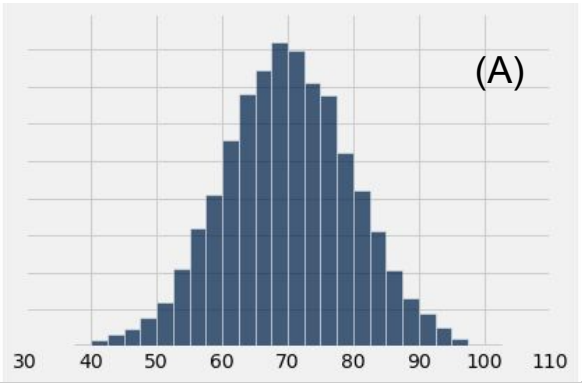


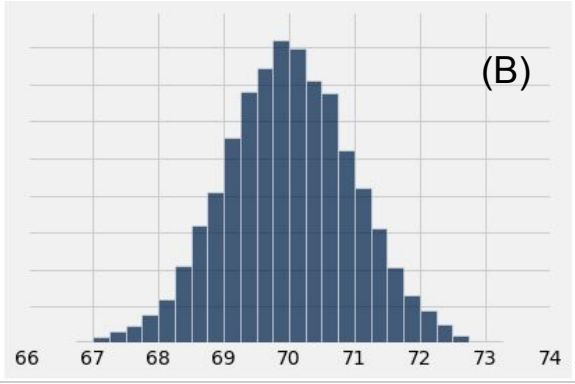


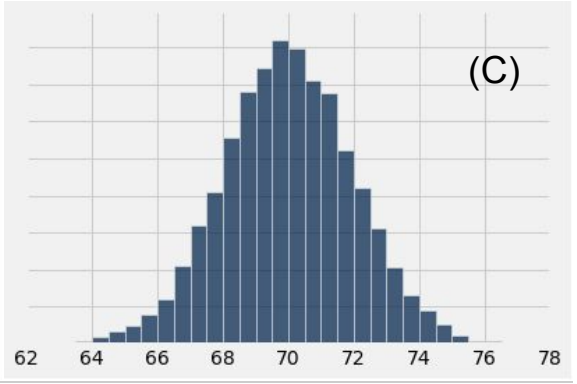


Discussion Question

These histograms show the empirical distribution of the mean retirement age for samples of N retiring professors. For which histogram is N largest?



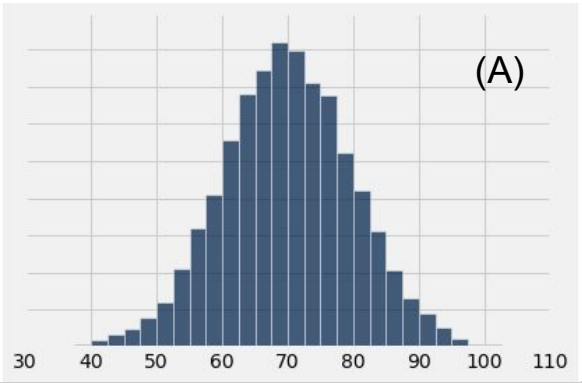


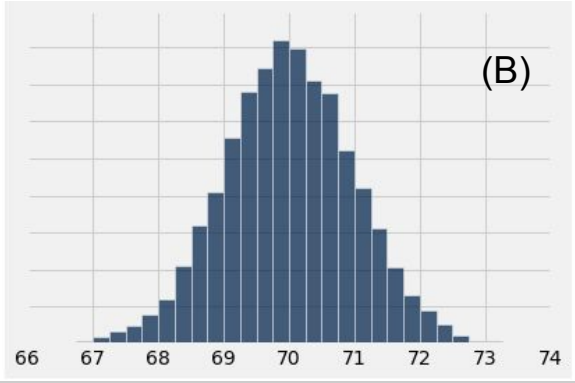


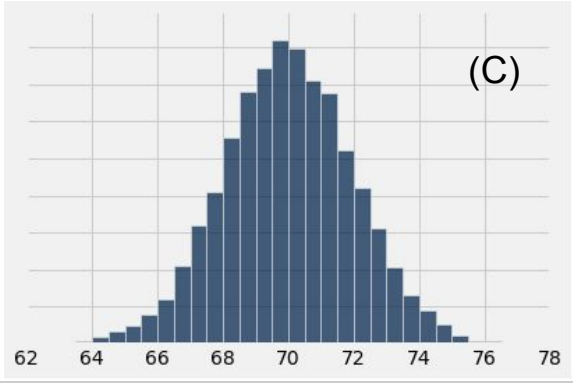
Discussion Question

Professors retire at age 70, on average; the distribution of retirement ages has SD 10.

One of the histograms below is the empirical distribution of the averages of 10,000 random samples of 100 retiring professors. Which one?

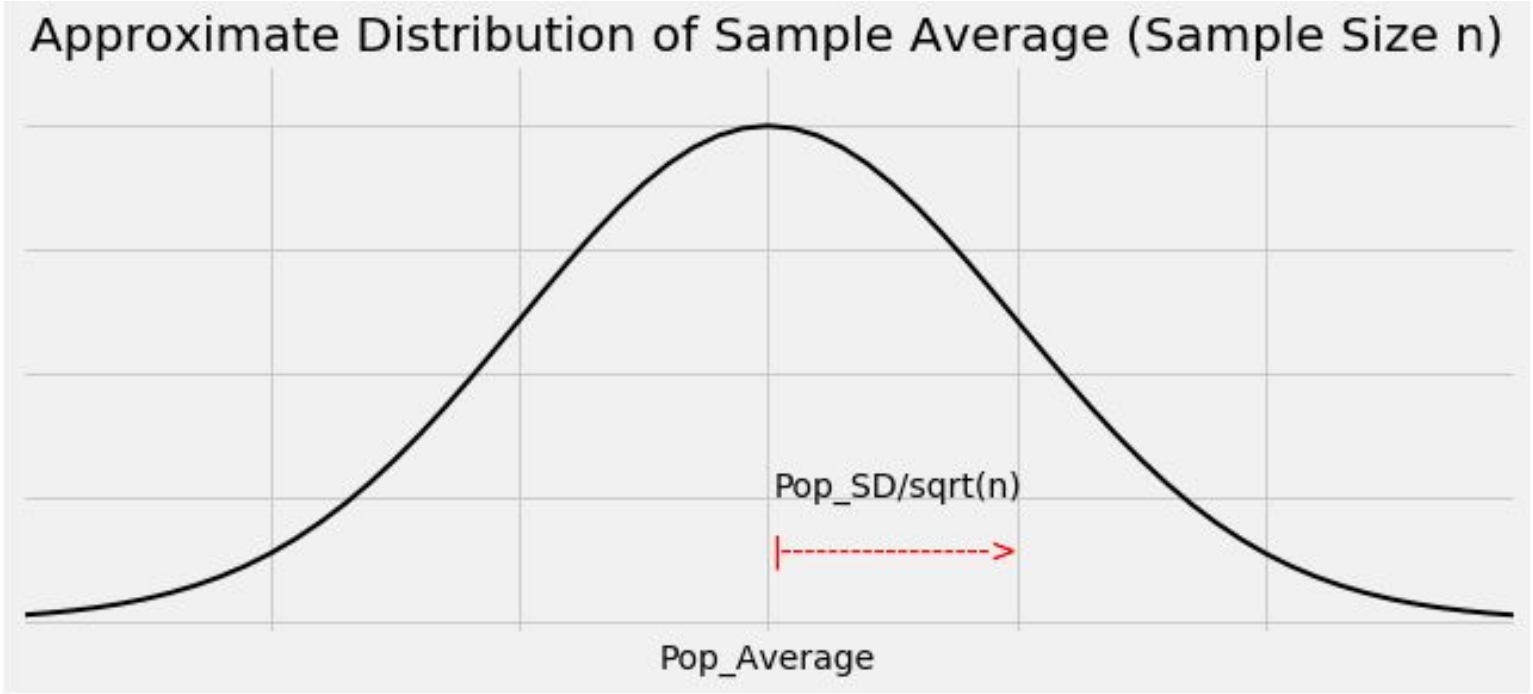




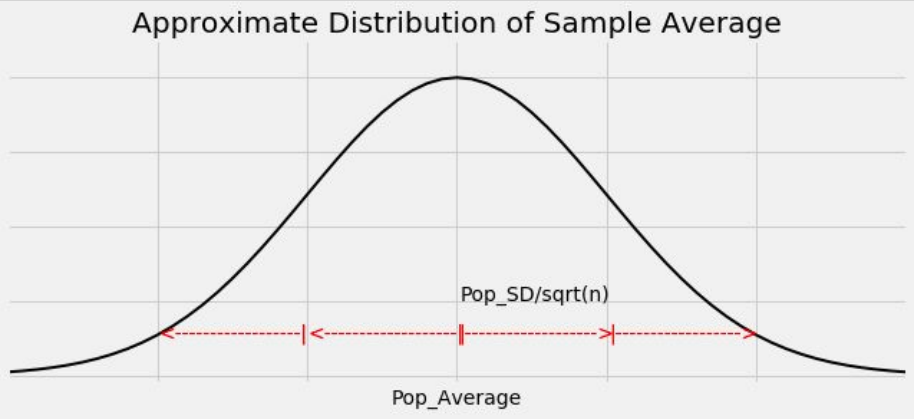


Confidence Intervals

Graph of the Distribution



The Key to 95% Confidence



- For about 95% of all samples, the sample average and population average are within **2 SD**s of each other.
- **SD** = SD of sample average

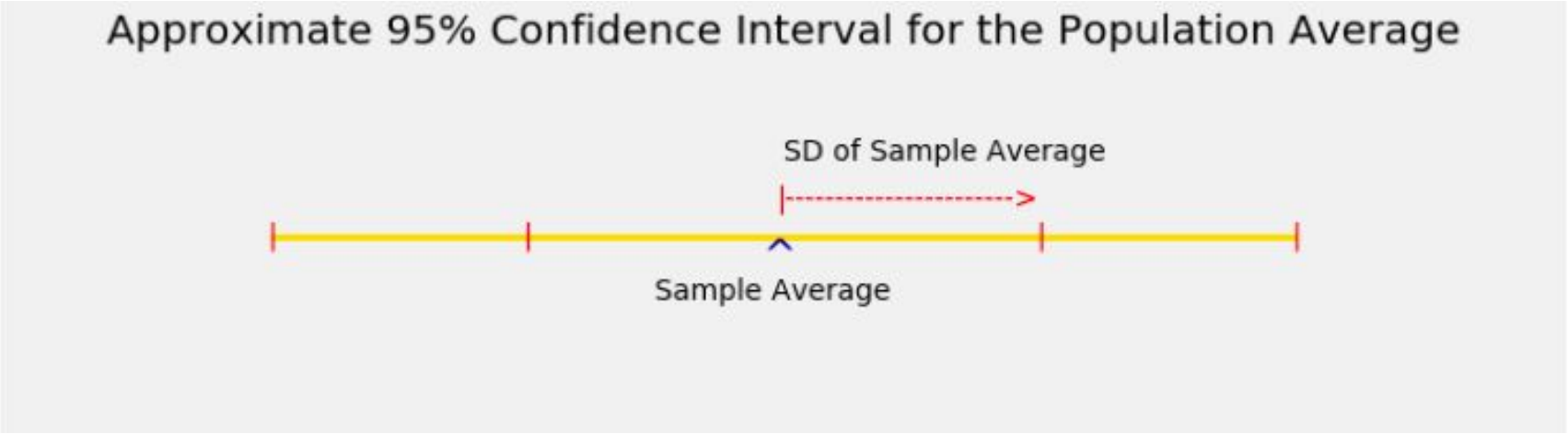
  = (population SD) / $\sqrt{\text{sample size}}$

Constructing the Interval

For 95% of all samples,

- If you stand at the population average and look two **SD**s on both sides, you will find the sample average.
- Distance is symmetric.
- So if you stand at the sample average and look two **SD**s on both sides, you will capture the population average.

The Interval



Width of the Interval

Total width of a 95% confidence interval for the population average

= 4 * SD of the sample average

= 4 * (population SD) / $\sqrt{\text{sample size}}$