Whitman
SCHOOL of MANAGEMENT
SYRACUSE UNIVERSITY

MBC 638
DATA ANALYSIS AND DECISION MAKING

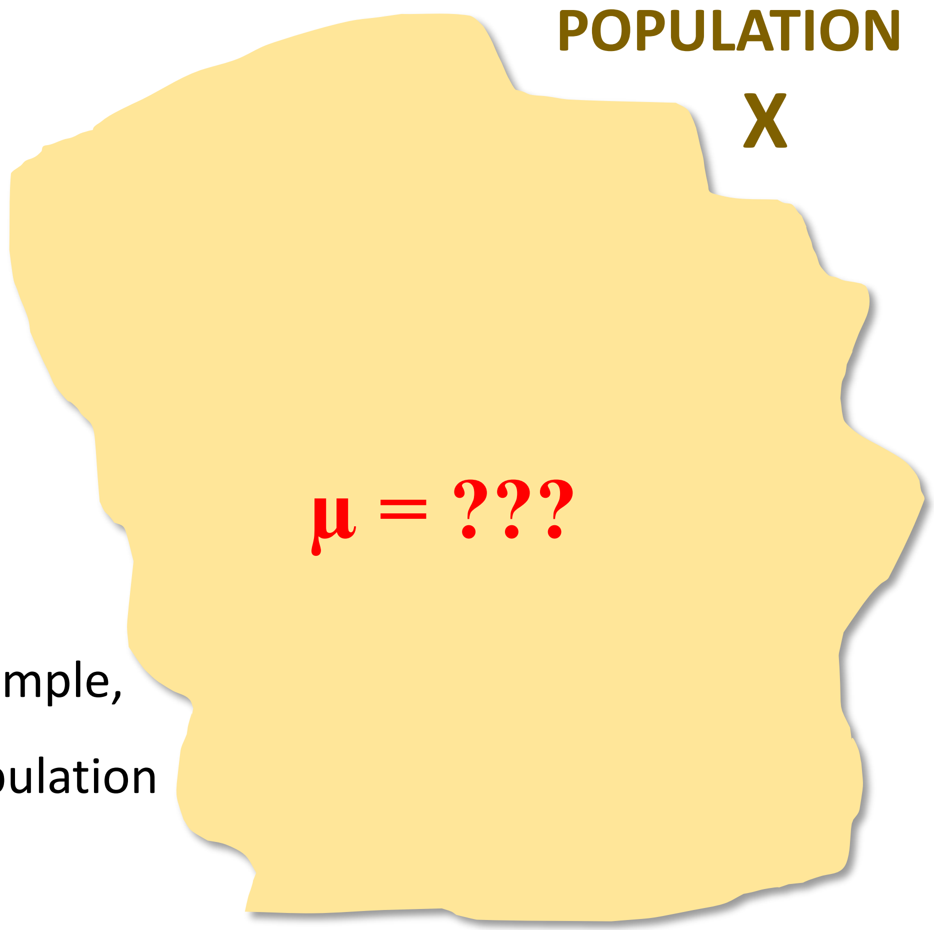
Anna Chernobai, Department of Finance

CENTRAL LIMIT THEOREM & sampling distribution of the sample mean
Chapter 7

Today…

- We are frequently interested in estimating the **population mean μ**.

- Typically, to do so, we estimate the mean of some random sample, i.e., **sample mean $\bar{x}$**, and then make inference about the population mean based on this **point estimate**.



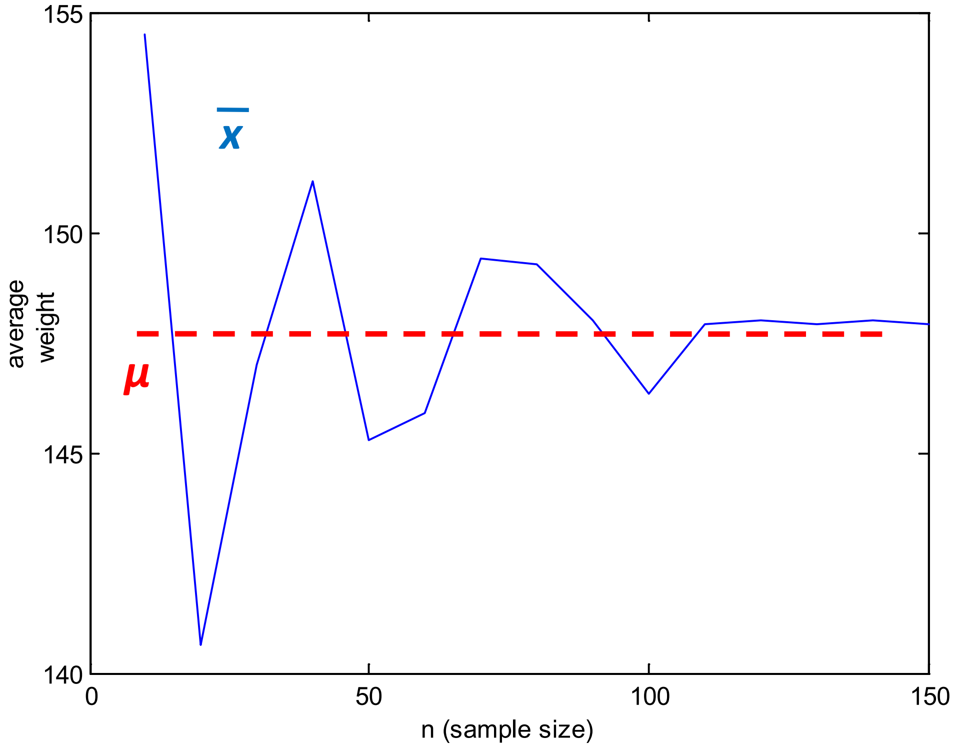
- How accurate is our estimation of the **population mean** based on the **sample mean**? Is the sample mean a good estimate of the population mean? What is the **estimation error** (also called **sampling error**)?

Today: CLT and sampling distributions

1. Law of Large Numbers
2. Central Limit Theorem
3. Sampling distribution: examples

Law of Large Numbers (LLN)

As sample size $\boldsymbol{n}$ increases, $\bar{\boldsymbol{x}}$ approaches $\boldsymbol{\mu}$.



What we observe:

- $\bar{x}$ approaches μ as n increases (LLN)
- $\bar{x}$ has a distribution
- $\bar{x}$ is more volatile for small n

Central Limit Theorem

- We just saw that **sample mean** is a random variable.

  (It has its own mean and its own standard deviation.)

  What can we tell about this **distribution** of the sample mean?

- **Central Limit Theorem:**

  When sample size *n* is large, then:

  Sample mean ($\bar{X}$) is approximately Normally distributed.

- Notice that I did *not* say whether the distribution of the population X from which the sample was taken was Normal. In fact, it does not matter!

If a **random** sample of size *n* is drawn from **any distribution** X that has mean μ and standard deviation σ, then, **if *n* is large**:

sample mean $\bar{x}$ is **approximately Normally distributed** with

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

How large is *large*?

n is considered large when $n \geq 30$

Examples

Example

Application: Insurance

The mean annual cost of automobile insurance is \$939 (*CNBC*, Feb. 23, 2006). The standard deviation is \$245 per policy.

a) A random sample of 50 policies was selected. What is the probability that a sample of 50 policies will have mean cost within \$25 of the population mean?

b) A certain sample of 50 policies showed that the mean cost is \$990. Would you say it's a **biased** sample?

c) Why do you think insurance companies benefit from insuring a very large pool of people as opposed to a very small pool?

Hint 1: Think about large payments.

Hint 2: Think about the standard deviation of a sample mean.

Example

Application: Insurance

The standard deviation of annual cost of automobile insurance is $245 per policy.

a) A random sample of 50 policies was selected. What is the probability that a sample of 50 policies will have mean cost within $25 of the population mean?

Example

Application: Insurance

The monthly premium that you pay for your health insurance reflects reflect the *average cost* that the insurer will incur in case you get sick and require medical attention.

Insurance companies often determine the premiums of its policyholders as some high percentile (e.g., 99%) of the **average cost distribution** of the insureds' pool. By doing so, they make sure that there is only 1% chance that their average payout would exceed your average medical costs.

Explain why, the more there are people in the insureds' pool, the lower your **monthly premium** will be. In other words, explain why *you* would benefit from purchasing insurance from a large insurance company.

Example

Suppose that a person's weight has a mean of 170 pounds and a standard deviation of 55 pounds.

An elevator is designed to carry no more than 4,000 pounds. Manufacturers must design elevators in such a way that they can handle a necessary weight. Help the manufacturers figure out the following.

Can they specify the maximum number of persons allowed in this elevator as 35? If not, then what would be a reasonable number of persons that they should specify?

Today: CLT and sampling distributions

1. Law of Large Numbers
2. Central Limit Theorem
3. Sampling distribution: examples

What's next? Confidence intervals.