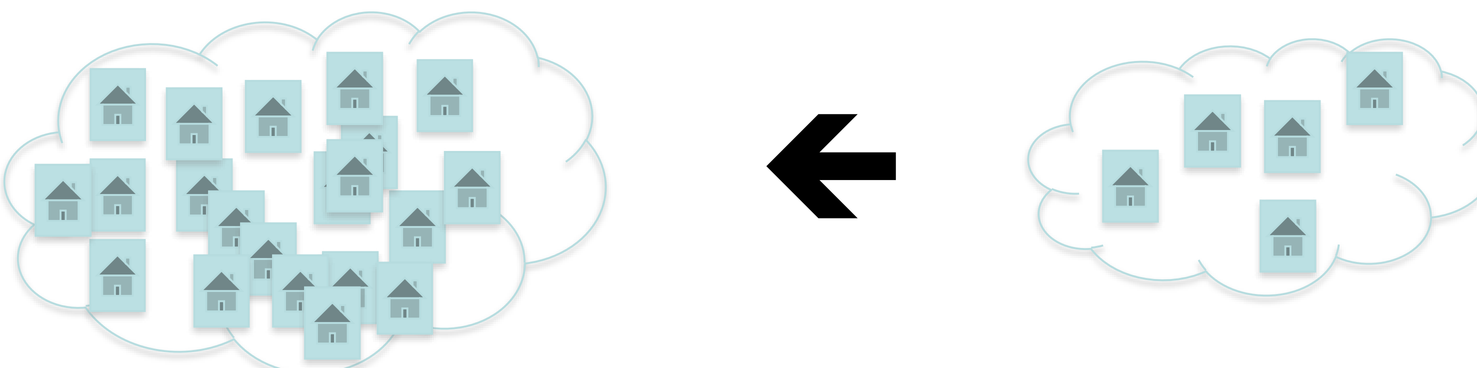
Eco 2740: Economic Statistics Fall, 2019. Sampling Distribrutions (Chapter 9)

Recall: Inferential Statistics

- Using the sample (we do see) to learn about the population (we don't see).
- We use a sample statistics like $\overline{x}$ to estimate analogous population parameter like m
- How variable is $\overline{x}$? What is its distribution?



Sample observations as random variables

- Random Sampling: Your observations $x_1, x_2, \ldots, x_n$ were drawn at random from the population.

- Before this particular group of observations were selected, the values that $x_1, x_2, \ldots, x_n$ would take was random since it depended on which observations got selected to be the sample.

- So we can think about them as random variables:

  $X_1, X_2, \ldots., X_n$

- What is the distribution of these observations?

The Distribution of a Random Sample

- Suppose $X_1, X_2, \ldots, X_n$ will be selected by random sampling from a population with population mean μ and population variance σ^2.
- Because of random sampling, the X's will be <u>independent</u> of each other and $COV(X_i,X_j) = 0$ for $i \neq j$
- Because they are all drawn at random from the same population, they will all have the same distribution. We say that they are <u>identically distributed</u>.
- And because they are drawn from the population they have population mean of $E[X_i] = \mu$ and population variance $VAR(X_i) = \sigma^2$

Distribution of Random Sample Continued

- We say that the Xi are independent and identically distributed (i.i.d.) with mean $E\left[X_i\right] = m$ and variance $VAR\left(X_i\right) = s^2$
- Usually we use short-hand notation to write this as:

$$X_1, X_2, \ldots, X_n \sim \text{i.i.d.}\left(m, s^2\right)$$

- An important special case is the one in which the Xi are also normally distributed, in which we write

$$X_1, X_2, \ldots, X_n \sim \text{i.i.d. N}\left(m, s^2\right)$$

Sample Statistics are also random variables

- Let's focus on $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ as an example
- We can think of it as a random variable generated by the following experiment:
  1. Select n observations at random from the population
  2. Take their average
- Suppose that we repeated this experiment again and drew a new random sample from the population?
- Would we draw the same x_1, x_2, x_n? Would we get the same $\bar{x}$?
- How different is it from rolling the dice twice?

The Expectation and Variance of the Sample Mean

- Using the laws of expectations and variances we can obtain:

$$E\left[\overline{X}\right] = m$$

$$VAR\left(\overline{X}\right) = \frac{s^2}{n}$$

<u>**ON BLACKBOARD:**</u> PROVIDE STEP BY STEP DERIVATIONS FOR TWO FORMULAS ABOVE

(if you missed class then please ask a friend for notes on this important topic)

Distribution of Sample Mean in Special Case of Normally Distributed Random Sample

- Since we have a normally distributed random sample:

$$X_1, X_2, \ldots, X_n \sim \text{i.i.d. } N(\mu, \sigma^2)$$

- And $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is just a <u>linear</u> <u>combination</u> of the X_i
- Thus the sample mean is also a <u>normal random variable</u>
- And we know it has mean μ and variance σ^2/n so:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

The Central Limit Theorem

- What if we have no idea what the distribution of X_i,
- But, we do have a random sample

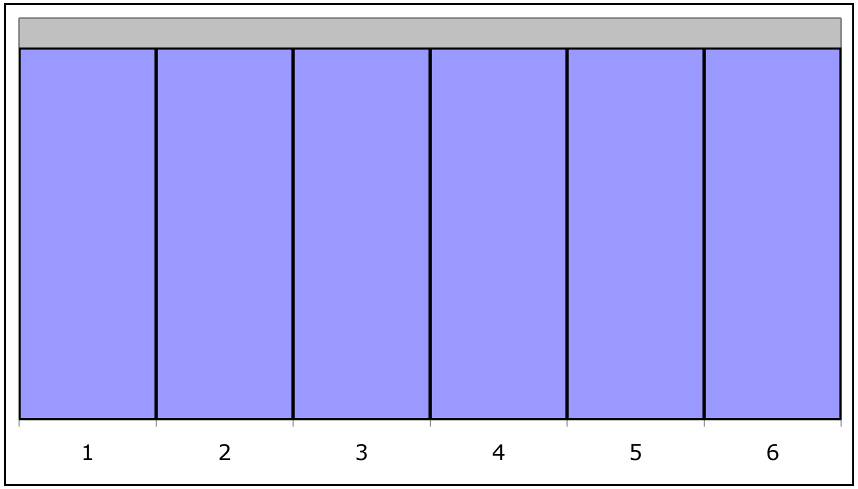
$$X_1, X_2, \ldots, X_n \sim \text{i.i.d.}(\mu, \sigma^2)$$

- Then what is the distribution of $\overline{X}$?
- In small samples – we have no idea!
- In large samples the Central Limit Theorem (CLT) tells us that $\overline{X}$ is approximately normal.
- Since it still has the same mean and variance:

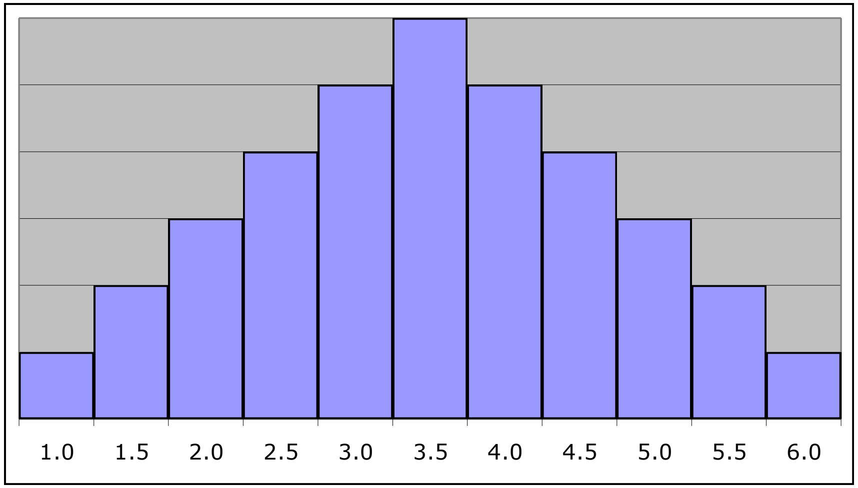
$$\overline{X} \sim \text{Aprox N}\left(\mu, \sigma^2 / n\right) \text{ for large n}$$

CLT in Pictures…

- Roll a single six-sided die once:
- X_1 and X_2 is outcome of rolls 1 and 2
- $\overline{X} = \frac{1}{2}(X_1 + X_2)$ is sample mean



Distribution of X_1



Distribution of $\overline{X}$

Sampling Distribution: The Concept

- What we just derived is called the **sampling distribution** of $\overline{X}$
- It is really just the distribution of $\overline{X}$
- But because $\overline{X}$ is a sample average we add "sampling"
- It reflects the idea that $\overline{X}$ is random because the sample we happen to select is itself random.
- As a thought experiment, we could imagine repeating the random sampling exercise over and over a billion times to generate an entire population of $\overline{X}$ values each based on a different random sample.
- The particular $\overline{X}$ we obtained is just one of these possible $\overline{X}$ values based on just one of the samples we could draw.

Example 9.1(a)…

The foreman of a bottling plant has observed that the amount of soda in each "32-ounce" bottle is actually a normally distributed random variable, with a mean of 32.2 ounces and a standard deviation of .3 ounce.

If a customer buys one bottle, what is the probability that the bottle will contain more than 32 ounces?

$$X \sim N\left(m = 32.2, s^{2} = (0.3)^{2}\right)$$

$$P(X > 32) = P\left(\frac{X - \mu}{\sigma} > \frac{32 - 32.2}{.3}\right) = P(Z > -.67) = 1 - .2514 = .7486$$

Example 9.1(b)…

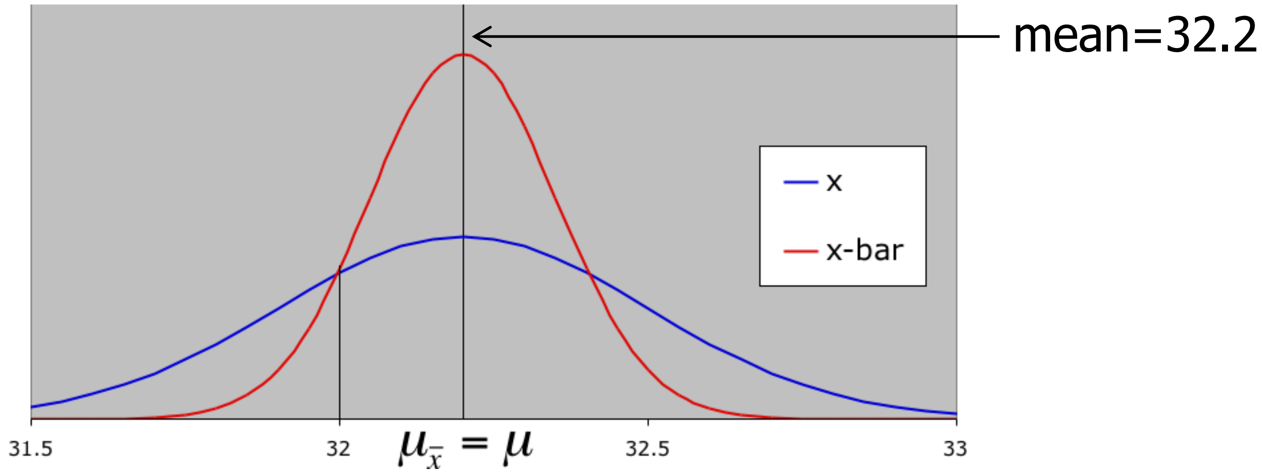
If a customer buys a carton of **four** bottles, what is the probability that the ***mean amount of the four bottles*** will be greater than 32 ounces?

$$X \sim N\left(\mu = 32.2, \sigma^2 = (0.3)^2\right)$$

$$\bar{X} \sim N\left(\mu, \sigma_{\bar{X}}^2\right) \quad \text{for} \quad \sigma_{\bar{X}}^2 = \sigma^2 / n = (0.3)^2 / 4 = \left(0.3/2\right)^2 = 0.15^2$$

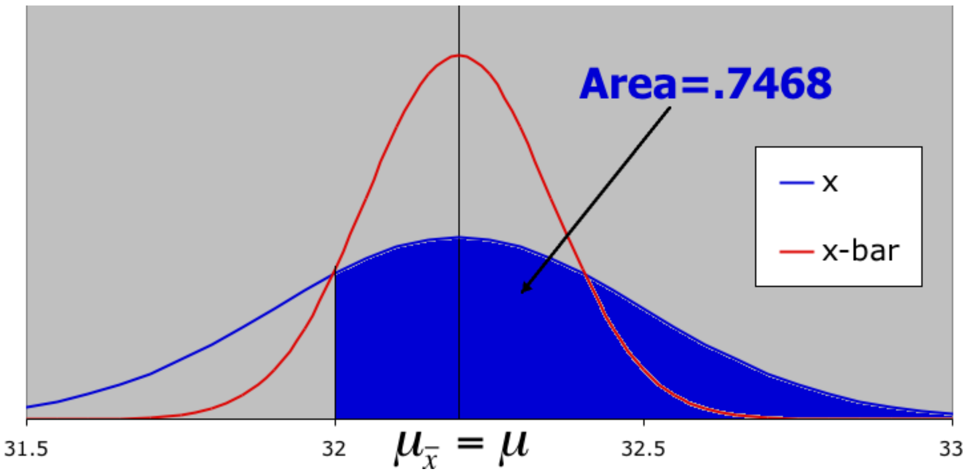
$$P(\bar{X} > 32) = P\left(\frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{32 - 32.2}{.15}\right) = P(Z > -1.33) = .9082$$

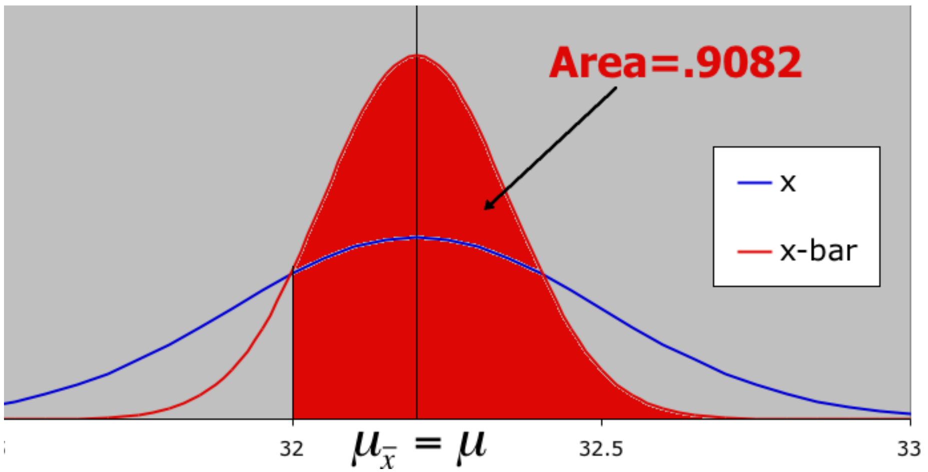
Graphically Speaking…



what is the probability that one bottle will contain more than 32 ounces?

what is the probability that the mean of four bottles will exceed 32 oz?





Chapter-Opening Example-Graduate Salaries

A B-School dean claims that the average salary of the school's graduates one year after graduation is $800 per week with a standard deviation of $100.

A skeptical second-year student double checks the claim by surveying 25 of last year's graduates and recording their weekly salary. He obtains a sample mean to be $750.

What is the probability that a sample of 25 graduates would have a mean of $750 or less if the Dean's claim is correct?

Chapter-Opening Example

Salaries of a Business School's Graduates

He does a survey of 25 people who graduated one year ago and determines their weekly salary.

He discovers the sample mean to be $750.

To interpret his finding he calculates the probability that a sample of 25 graduates would have a mean of $750 or less when the population mean is $800 and the standard deviation is $100.

Chapter-Opening Example

We seek $P(\overline{X} < 750)$ If the dean is correct then:

$\overline{X}$ is approx $N(\mu_{\overline{x}} = \mu = 800$, $\sigma^2/n = 100^2/25 = (100/5)^2 = 20^2)$

$$P(\bar{X} < 750)$$

$$= P\left(\frac{\bar{X} - m_{\bar{x}}}{s_{\bar{x}}} < \frac{750 - 800}{20} \right)$$

$$= P(Z < -2.5)$$

$$= .0062$$

Pretty small chance – this does not seem to support the Dean's claim.