Topic 6: Sampling Distributions

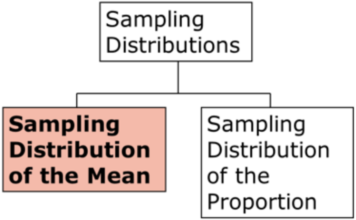
Dr. Lei Pan

School of Accounting, Economics and Finance
Curtin University

Sampling Distributions

- A **sampling distribution** is a distribution of all of the possible values of a sample statistic for a given sample size selected from a population.
- For example, suppose you sample 50 students from Curtin University regarding their mean GPA. If you obtained many different samples of size 50, you will compute a different mean for each sample. We are interested in the distribution of all potential mean GPAs we might calculate for any sample of 50 students.

Sampling Distributions of the Mean



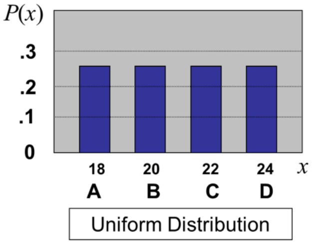
- The **sampling distribution of the mean** is the distribution of all possible sample means if you select all possible samples of a certain size.
- If the average of all possible sample means equals the population mean then the sample mean is **unbiased**.

Developing a Sampling Distribution (1 of 5)

- Assume there is a population ...
- Population size $N = 4$
- Random variable, X, is age of individuals.
- Values of X: 18, 20, 22, 24 (years)

Developing a Sampling Distribution (2 of 5)

- Summary measures for the population distribution:
  - $\mu = \frac{\sum X_i}{N} = \frac{18+20+22+24}{4} = 21$
  - $\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = 2.236$



Developing a Sampling Distribution (3 of 5)

- Now consider all possible samples of size $n = 2$.
  - 16 possible samples (sampling with replacement)

1st Obs	2nd Observation			
	18	20	22	24
18	18,18	18, 20	18, 22	18, 24
20	20,18	20, 20	20, 22	20, 24
22	22,18	22, 20	22, 22	22, 24
24	24,18	24, 20	24, 22	24,24

  - Resulting in 16 sample means

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

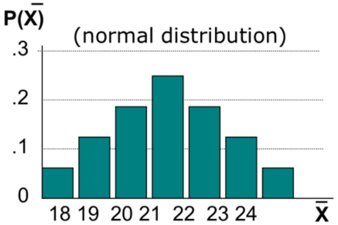
Developing a Sampling Distribution (4 of 5)

Sampling distribution of all sample means

16 Sample Means

1st Obs	2nd Observation 18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Sample Means Distribution



Developing a Sampling Distribution (5 of 5)

Summary measures of sampling distribution

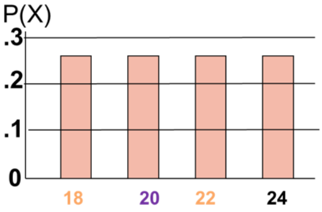
$$\mu_{\bar{x}} = \frac{18 + 19 + 19 + \dots + 24}{16} = 21$$

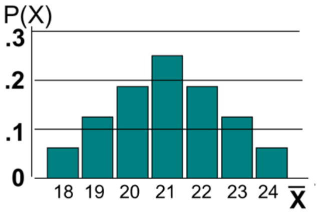
$$\sigma_{\bar{x}} = \sqrt{\frac{(18-21)^2 + (19-21)^2 + \dots + (24-21)^2}{16}} = 1.58$$

Note: Here we divide by 16 because there are 16 different samples of size 2.

Comparing the Population Distribution to the Sample Means Distribution

Population **N = 4**	**Sample Means Distribution** **n = 2**
$\mu = 21 \ \sigma = 2.236$	$\mu_{\overline{X}} = 21 \quad \sigma_{\overline{X}} = 1.58$





Standard Error of the Mean

- Different samples of the same size from the same population will yield different sample means.
- A measure of the variability in the mean from sample to sample is given by the **Standard Error of the Mean**.

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

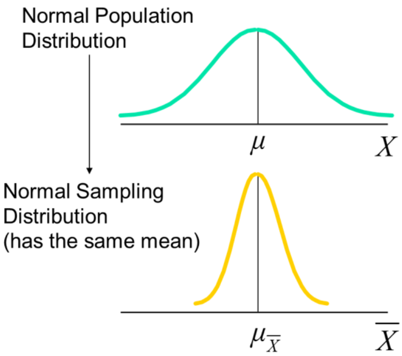
(This assumes that sampling is done with replacement or sampling is done without replacement from a large or infinite population.)

- Note that the standard error of the mean decreases as the sample size increases.

Sampling from Normally Distributed Populations

- If a population is **normal** with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ is also **normally distributed** with
  - $\mu_{\bar{X}} = \mu$
  - $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

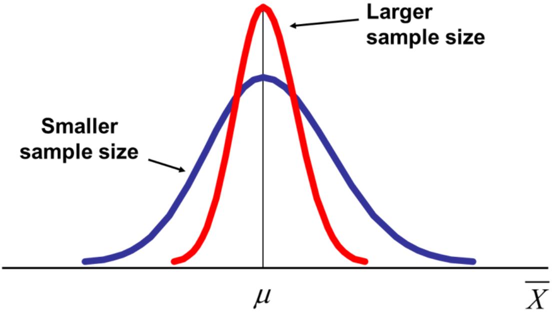
Sampling Distribution Properties (1 of 2)



$$\mu_{\bar{X}} = \mu \quad \text{(i.e } \bar{X} \text{ is unbiased)}$$

Sampling Distribution Properties (2 of 2)

As n increases, $\sigma_{\bar{X}}$ decreases.



Z-value for Sampling Distribution of the Mean

- Z-value for the sampling distribution of $\bar{X}$:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{(\bar{X} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

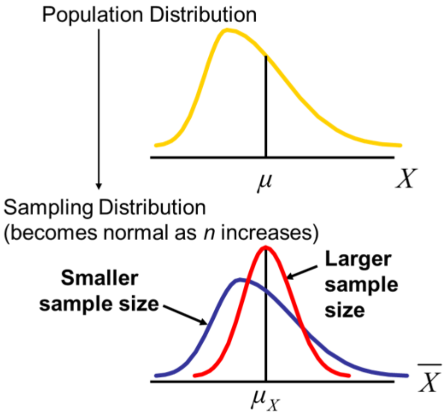
where

$$\begin{aligned} \bar{X} &= \text{sample mean} \\ \mu &= \text{population mean} \\ \sigma &= \text{population standard deviation} \\ n &= \text{sample size} \end{aligned}$$

- If a population is not normal, we can apply the **Central Limit Theorem**.
  - even if the population is not normal
  - as long as the sample size is large enough
  - sample means from the population will be approximately normal with

$$\mu_{\bar{X}} = \mu \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem



How Large Is Large Enough?

- For most distributions, $n \geq 30$ will will give a sampling distribution that is nearly normal.
- For fairly symmetric distributions, $n > 15$.
- For a normal population distribution, the sampling distribution of the mean is always normally distributed.

Example

- Suppose a population has mean $\mu = 8$ and standard deviation $\sigma = 3$. Suppose a random sample of size $n = 36$ is selected.
- What is the probability that the sample mean is between 7.8 and 8.2?

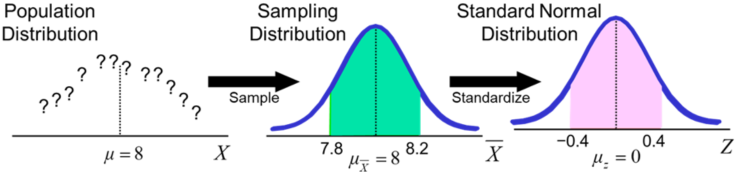
Solution (1 of 2)

- Even if the population is not normally distributed, the central limit theorem can be used ($n > 30$).
- ... so the sampling distribution of $\bar{X}$ is approximately normal
- ... with mean $\mu_{\bar{X}} = 8$
- ... and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{36} = 0.5$

Solution (2 of 2)

$$P(7.8 < \bar{X} < 8.2) = P(\frac{7.8-8}{\frac{3}{\sqrt{36}}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{8.2-8}{\frac{3}{\sqrt{36}}})$$

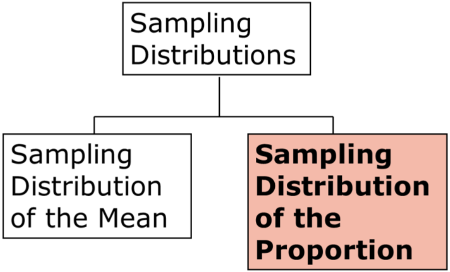
$$= P(-0.4 < Z < 0.4) = 0.6554 - 0.3446 = \boxed{0.3108}$$



You Do It

The mean and standard deviation of the tax value of all vehicles registered in a certain state are $\mu = \$13,525$ and $\sigma = \$4,180$. Suppose random samples of size 100 are drawn from the population of vehicles. What are the mean $\mu_{\bar{X}}$ and standard deviation $\sigma_{\bar{X}}$ of the sample mean $\bar{X}$?

Sampling Distribution of the Proportion

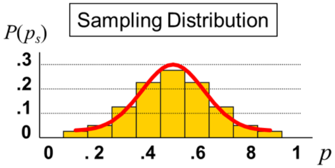


Population Proportions

- $\pi =$ the proportion of items in the **population** with a characteristic of interest
- p is the **sample proportion** and provides an estimate of π
  - $p = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$
  - $0 \leq p \leq 1$
  - p is approximately distributed as a normal distribution when n is large (assuming sampling with replacement from a finite population or without replacement from an infinite population)

Sampling Distribution of p

- Selecting all possible samples of a certain size, the distribution of all possible sample proportions is the sampling distribution of the proportion.
- The underlying distribution of the sample proportion is binomial.
- Approximately by a normal distribution if:
  - $n\pi \geq 5$ and $n(1-\pi) \geq 5$
    where $\mu_p = \pi$ and $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$



Z-value for Proportions

- Standardize p to a Z value with the formula:

$$Z = \frac{p - \pi}{\sigma_p} = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Example

- If the true proportion of voters who support Proposition A is $\pi = 0.4$, what is the probability that a sample of size 200 yields a sample proportion between 0.40 and 0.45?
- i.e. if $\pi = 0.4$ and $n = 200$, what is

$$P(0.40 \leq p \leq 0.45)$$

Solution

- Step 1: Find σ_p: $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.4(1-0.4)}{200}} = 0.03464$
- Step 2: Convert to standardized normal

$$P(0.40 \leq p \leq 0.45) = P(\frac{0.40 - 0.40}{0.03464} \leq Z \leq \frac{0.45 - 0.40}{0.03464})$$
$$= P(0 \leq Z \leq 1.44) = 0.9251 - 0.5000 = \boxed{0.4251}$$