Sampling Distributions

Sampling Distribution of $\overline{x}$

Sampling Distribution of $\overline{p}$

18

Sampling Distributions

- Chose a random sample of 30 managers and calculated
  - Sample Mean = x
  - Sample Proportion = p

Sample Mean	Sample SD	Proportion of Competing Training (Sample)
72372.71	3839.873894	0.666666667

- x is the point estimator of the population mean, μ
  - Point Estimate: x = $72,372.71
- p is the point estimator of the population proportion, p
  - Point Estimate: p = 0.67

What would happen if we chose a different sample of 30 managers?*

19

Sampling Distributions

If we picked 30 random managers 500 times, the results might look something like this:

Sample Number	Sample Mean ($\bar{x}$)	Sample Proportion ($\bar{p}$)
1	71,814	0.63
2	72,670	0.70
3	71,780	0.67
4	71,588	0.53
.	.	.
.	.	.
.	.	.
500	71,752	0.50

20

Sampling Distributions

Make a frequency and relative frequency distribution of the x results:

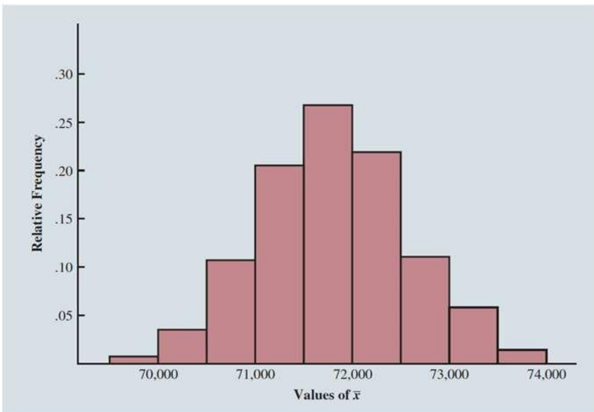
Mean Annual Salary ($)	Frequency	Relative Frequency
69,500.00-69,999.99	2	0.004
70,000.00-70,499.99	16	0.032
70,500.00-70,999.99	52	0.104
71,000.00-71,499.99	101	0.202
71,500.00-71,999.99	133	0.266
72,000.00-72,499.99	110	0.220
72,500.00-72,999.99	54	0.108
73,000.00-73,499.99	26	0.052
73,500.00-73,999.99	6	0.012
Totals:	**500**	**1.000**

21

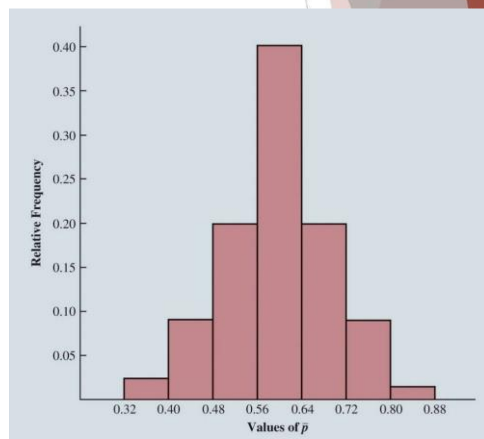
Sampling Distributions

Use the frequency distribution to make a histogram:

$\bar{x}$

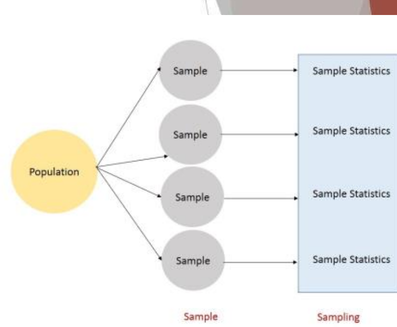


$\bar{p}$



22

Sampling Distributions



- $\bar{x}$ and $\bar{p}$ are random variables so they have:
  - A Mean, or Expected Value
  - A Standard Deviation
  - Probability Distribution or **Sampling Distribution**
    - Each Probability Distribution has a **characteristic shape** or **form**
    - Knowing the **Sampling Distribution** of $\bar{x}$, and its properties helps us gauge how close the sample mean, x is to the true Population Mean, μ

23

Sampling Distributions

- The Expected Value of the sample mean x
  - Is the mean average of all possible values of x that can be generated by the various simple random samples
- Turns out:
  - The average of all the x's we could get by sampling 30 managers over and over actually equals the Population Mean, μ

EXPECTED VALUE OF $\bar{x}$

$$E(\bar{x}) = \mu \qquad \textbf{(6.1)}$$

where

$$E(\bar{x}) = \text{the expected value of } \bar{x}$$
$$\mu = \text{the population mean}$$

What is it called when the expected value of a point estimator equals the population parameter?*

24

Sampling Distributions

Sample Number	Sample Mean ($\bar{x}$)	Sample Proportion ($\bar{p}$)
1	71,814	0.63
2	72,670	0.70
3	71,780	0.67
4	71,588	0.53
.	.	.
.	.	.
.	.	.
500	71,752	0.50

$$\frac{71{,}814 + 72,670 + 71{,}780 + 71{,}588 + \ldots + 71{,}752}{500} = \$71{,}800 = \mu = Population\ Mean$$

25

Sampling Distributions

The formula for the standard deviation of x depends on whether the population is finite or infinite.

Using the following notation:
σ_x = the standard deviation of x, or the standard error of the mean.
σ = the standard deviation of the population.
n = the sample size.
N = the population size.

STANDARD DEVIATION OF $\bar{x}$

Finite Population	*Infinite Population*	
$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{\sigma}{\sqrt{n}}\right)$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	**(6.2)**

26

Sampling Distributions

- **Finite population correction factor:**

$$\sqrt{\frac{N - n}{N - 1}}$$

Finite Population

$$s_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{s}{\sqrt{n}}\right)$$

- In many practical sampling situations,
  - The finite population correction factor is close to 1
  - So, the difference between the finite and infinite standard deviations is negligible.

In general, you can use *Infinite Population* $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ when, $\frac{n}{N} < 0.05$

27

Sampling Distributions

ESTIMATED STANDARD DEVIATION OF $\bar{x}$

Finite Population	Infinite Population	
$s_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{s}{\sqrt{n}}\right)$	$s_{\bar{x}} = \left(\frac{s}{\sqrt{n}}\right)$	**(6.3)**

Estimated standard error: $s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{3,348}{\sqrt{30}} = 611.3.$

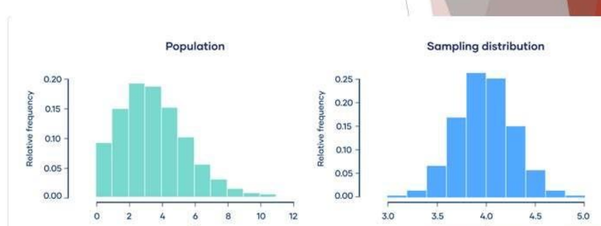
True standard error: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4,000}{\sqrt{30}} = 730.3.$

The difference between $s_{\bar{x}}$ and $\sigma_{\bar{x}}$ is due to sampling error.

28

Sampling Distributions

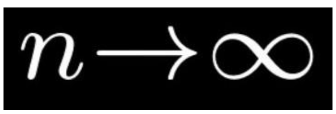
- When the population has a normal distribution,
  - The sampling distribution of x is normally distributed for any sample size
- When the population does not have a normal distribution
  - The central limit theorem is helpful in identifying the shape of the sampling distribution of x

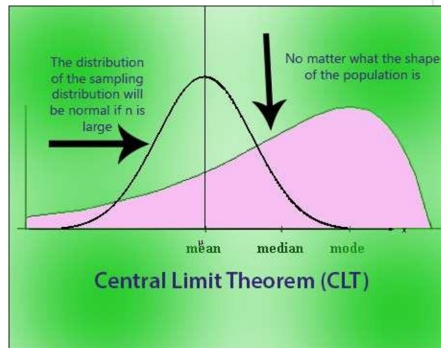


29

Sampling Distributions

- **Central Limit Theorem**
  - Distribution of the sample mean x can be approximated by a normal distribution as the sample size becomes large.



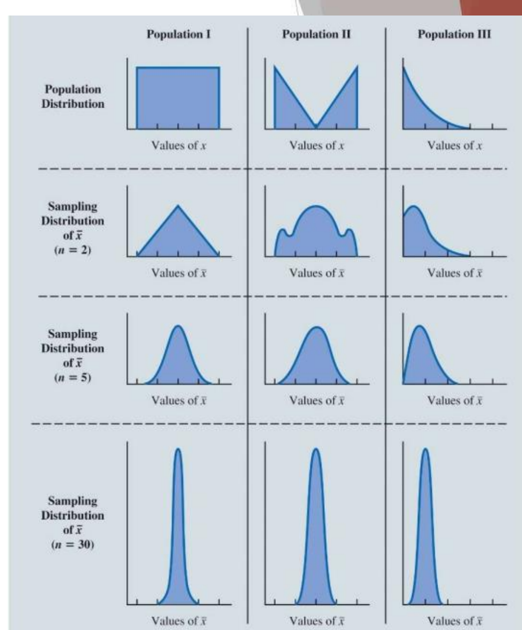


30

Sampling Distributions

Central Limit Theorem for Three Populations

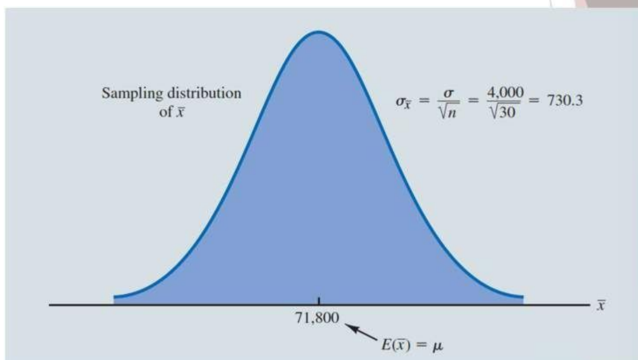
- Top panel shows that none of the populations are normally distributed.
- Bottom three panels show the shape of the sampling distribution for samples n = 2, n = 5, and n = 30.
- For sample size of 30 or more, it looks closer to normal
- We can assume then, for a sample of 30 or more, the sampling distribution can be approximated by normal distribution.



31

Sampling Distributions

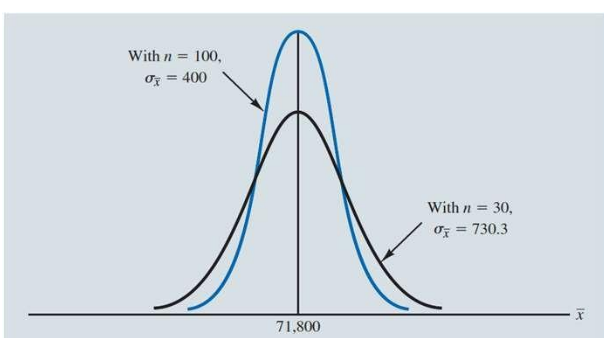
Sampling Distribution of $\bar{x}$ for the Mean Annual Salary of a Simple Random Sample of 30 EAI Employees



32

Sampling Distributions

A Comparison of the Sampling Distributions of $\bar{x}$ for Simple Random Samples of:
$n = 30$ and $n = 100$ EAI Employees.



33

Sampling Distributions

Sampling Distribution of $\bar{p}$:

The sample proportion $\bar{p}$ is the point estimator of the population proportion p.

The formula for computing the sample proportion is:

$$\bar{p} = \frac{x}{n}$$

where

x = the number of elements in the sample that possess the characteristic of interest.

n = sample size.

34

Sampling Distributions

Sampling distribution of $\bar{p}$: The sampling distribution of $\bar{p}$ is the probability distribution of all possible values of the sample proportion $\bar{p}$.

EXPECTED VALUE OF $\bar{p}$

$$E(\bar{p}) = p \tag{6.4}$$

where

$E(\bar{p})$ = the expected value of $\bar{p}$
p = the population proportion

STANDARD DEVIATION OF $\bar{p}$

Finite Population | *Infinite Population*

$$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}} \qquad \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \tag{6.5}$$

35

Sampling Distributions

ESTIMATED STANDARD DEVIATION OF $\bar{p}$

Finite Population	Infinite Population	
$s_{\bar{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$	$s_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$	**(6.6)**

The sampling distribution of $\bar{p}$ can be approximated by a normal distribution whenever $np \geq 5$ and $n(1 - p) \geq 5$.

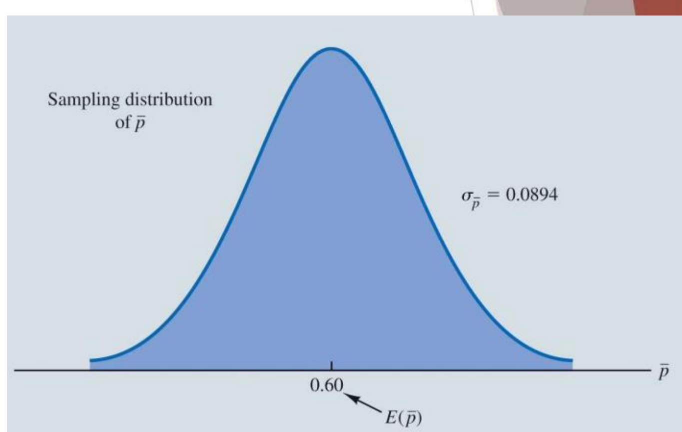
For our example: $p = 0.63$, n = 30

$$np = 30(0.63) = 18.9 > 5 \; and \; n(1-p) = 30(0.37) = 11.1 > 5$$

36

Sampling Distributions

Sampling Distribution of p for the Proportion of EAI Employees Who Participated in the Management Training Program



37

Interval Estimation

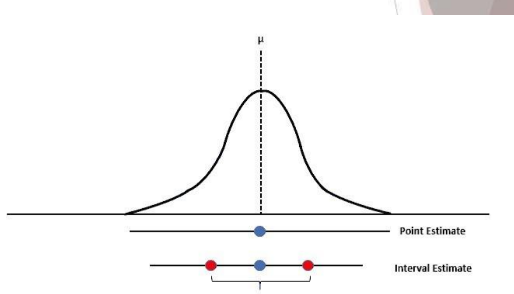
Interval Estimation of the Population Mean

Interval Estimation of the Population Proportion

38

Interval Estimation

- Point estimators are not perfect!
  - They do not provide the exact value of the population parameter
- An **interval estimate**
  - computed by adding and subtracting a value, called the **margin of error**, to the point estimate.
- The general form of an interval estimate is:



Point estimate ± Margin of error

39

Interval Estimation

Interval Estimation of the Population Mean:

- An interval estimate provides information about how close the point estimate is to the value of the population parameter.
- General form of an interval estimate of a population mean is:

$$\bar{x} \pm \text{Margin of error}$$

- General form of an interval estimate of a population proportion is:

$$\bar{p} \pm \text{Margin of error}$$

40

Interval Estimation

Interval Estimation of the Population Mean (cont.):

For any normally distributed random variable:

- 90% of the values lie within 1.645 standard deviations of the mean.
- 95% of the values lie within 1.960 standard deviations of the mean.
- 99% of the values lie within 2.576 standard deviations of the mean.

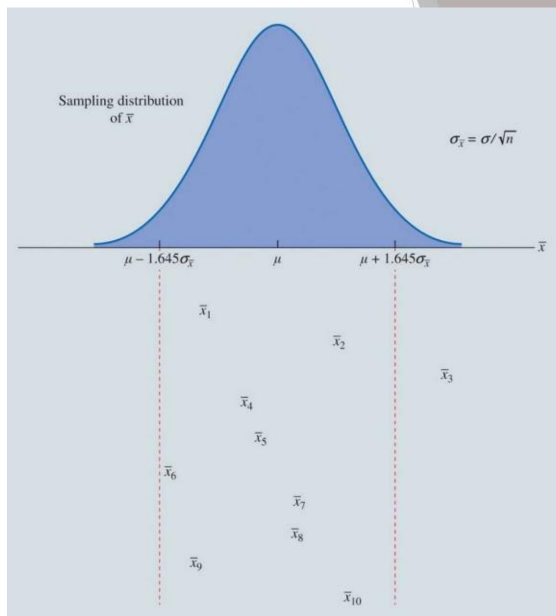
41

Interval Estimation

Sampling Distribution of the Sample Mean

How many of the x's are in between the lines? What does this mean?

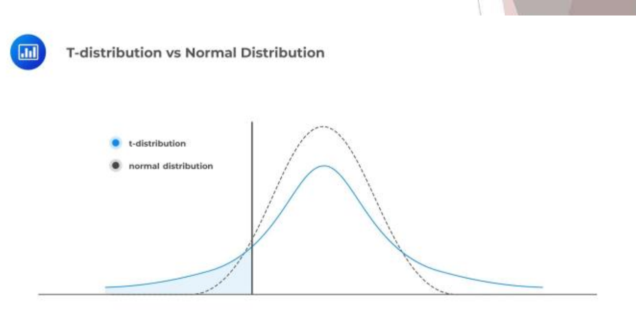
- Remember, we do not generally know the population standard deviation, σ
  - We have to use the sample data to estimate:
    - σ and μ
  - This introduces more uncertainty about the distribution values of x.



42

Interval Estimation

- To address this additional source of uncertainty
  - Use a probability distribution known as the t **distribution:**
    - A family of similar probability distributions.
    - The shape of each depends on a the **degrees of freedom**.
    - Similar in shape to the **standard normal distribution**, but wider.

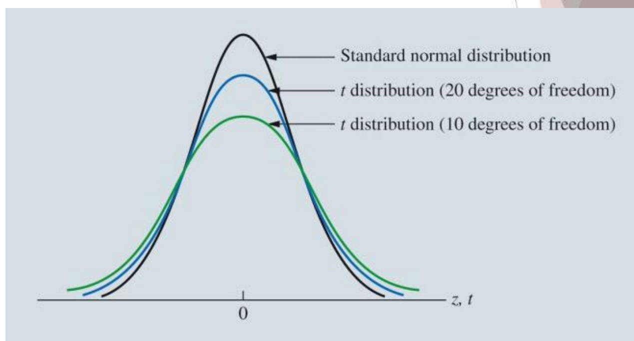


43

Interval Estimation

Comparison of the Standard Normal Distribution with t Distributions with 10 and 20 Degrees of Freedom

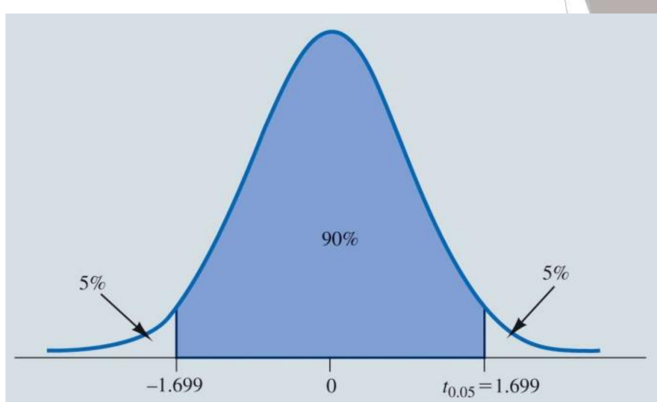
As the degrees of freedom increase, the t distribution narrows, its peak becomes higher, and it becomes more similar to the standard normal distribution.



44

Interval Estimation

t Distribution with 29 Degrees of Freedom



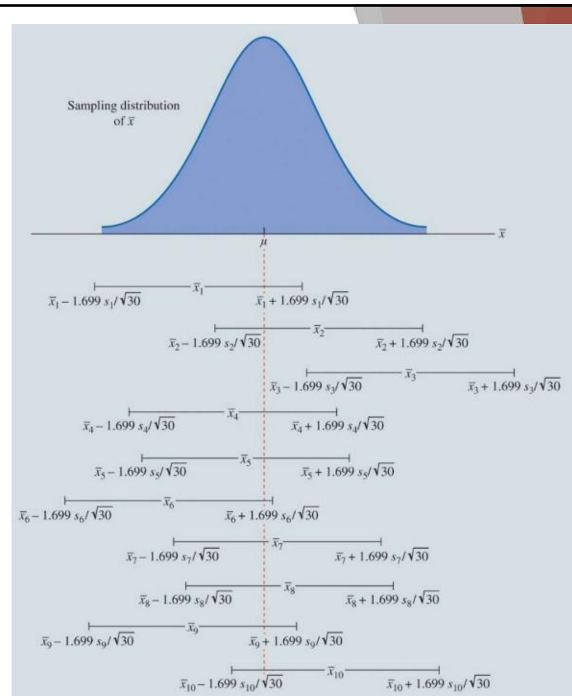
Use Excel's T.INV.2T function to find the value from a t distribution such that 95% of the distribution is included in the interval $\pm t$ for 29 degrees of freedom.*

45

Interval Estimation

Intervals Formed Around Sample Means
from 10 Independent Random Samples

- Approximately 90% of all the intervals constructed will contain the population mean
- We are approximately 90% confident that the interval will include the population mean:
  - The value of 0.90 is referred to as the **confidence coefficient.**
  - The interval is called the 90% **confidence interval.**



46

Interval Estimation

- The **level of significance**
  - is the probability that the interval estimation procedure will generate an interval that does not contain the population mean:

$$\alpha = \text{level of significance} = 1 - \text{confidence coefficient}$$

47