

# STATISTICS



MORPHINE ACADEMY

MORPHINE ACADEMY

# The Sampling Distribution of a Sample Mean and the Central Limit Theorem

# Properties of the Sampling Distribution of $\bar{x}$

• Let  $X_1, X_2, ...., X_n$  be a random sample of size n drawn from any population with mean= $\mu$  and standard deviation =  $\sigma$ , then we have the following properties about the sampling distribution mean  $\mu$ , that is:

الفكرة المرتبيمة:

لما نوف عني (sample) مر (population) وفي المتو له نتي (لآ) ، عدد المتو حرصًا تاب مع تبلف وسعزة لعبرة.

#### : The go

21 of the september (10) so et poil plood, y - all injustice.

# orll my his org:

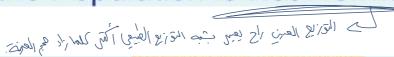
5 = 20, 1 = 100 one population 11 d

 $\frac{\sqrt{25}}{\sqrt{25}} = \frac{20}{5} = \frac{$ 

ا عدة عا لعبرة به أي مول ٥٥ ام إن معارب = ١١ س

(bej 000 min - 20 vojal)

#### Sample Mean Sampling Distribution: If the Population is **not** Normal



#### Sampling distribution properties:

Central Tendency

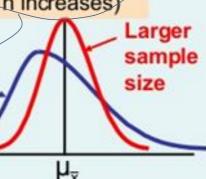
$$\mu_{\bar{x}} = \mu$$

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

Population Distribution

Sampling Distribution (becomes normal as n increases)

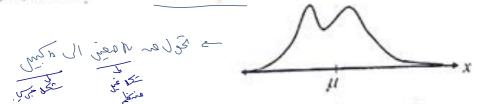
> Smaller sample size



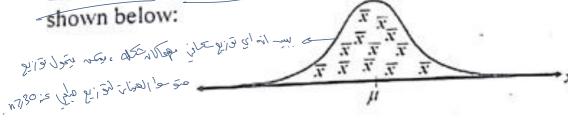
sam

( Joing Virie ) a de july july mer in solif de C

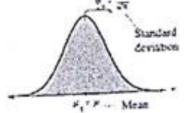
That is, if the sample size is large  $(n \ge 30)$ , and the sample is drawn from any population with mean =  $\mu$  and standard deviation =  $\sigma$ , as shown below:



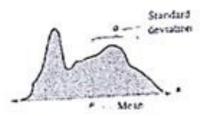
then the sampling distribution of the sample mean approximates a normal distribution. The greater the sample size, the better the approximation, as



Any Population Distribution



Distribution of Sample Mean,  $n \ge 30$ 



#### **Central Limit Theorem (CLT)**

• Example: Interpreting the CLT

Phone bill for residents of a city have a mean of \$64 and a standard deviation of \$9, random samples of 36 phone bills are drawn from this population and the mean of each sample is determined. Find the mean and standard error of the mean of the sampling distribution. Then sketch a graph of the sampling of sample mean?

$$G_{\overline{x}} = \frac{G}{NN} = \frac{9}{36} = \frac{9}{6} = \frac{15}{15}$$

#### Solution: Interpreting the Central Limit Theorem

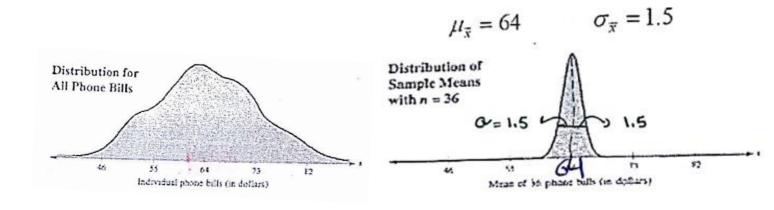
The mean of the sampling distribution is equal to the population mean:

$$\mu_{\bar{x}} = \mu = 64$$

 The standard deviation (standard error of the mean) is equal to the population standard deviation divided by the square root of n:

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{36}} = 1.5$$

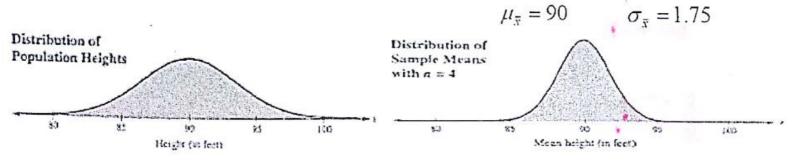
 Since the sample size is greater than 30, the sampling distribution can be approximated by a normal distribution with:



#### **Solution**

- The mean of the sampling distribution is equal to the population mean:  $\mu_{\overline{x}} = \mu$  = 90
- The standard error of the mean is equal to the population standard deviation divided by root of n:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 3.5/2 = 1.75$$



## The Probability of a Sample Mean

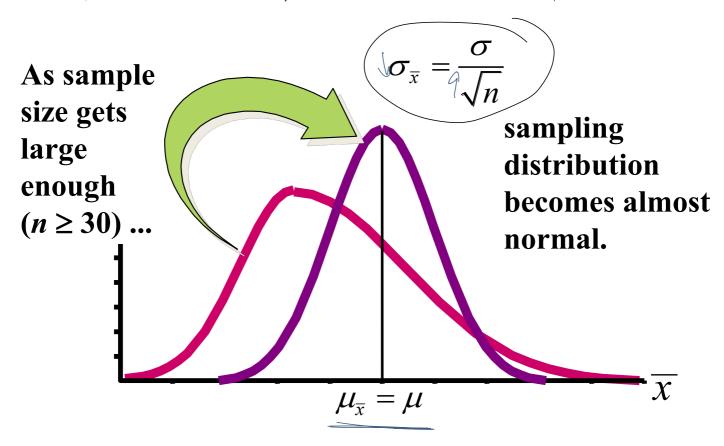
- \* How to find the probability associated with a sample mean?
- To find the probability for the sample mean, we transform X to a Z-score as follows:

Today Just

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$
 is approximately N(0, 1).

#### **Central Limit Theorem**

المبدأ: عندما يزداد حجم العينة (n) حتى يصبح كبيرًا بما فيه الكفاية (عادة 30 🧷 )، يصبح توزيع المتوسطات العينية ( 📆 ) تقريبًا توزيعًا طبيعيًا.



#### **Examples**

The graph shows the length of time people spend driving each day. You randomly select 50 drivers age 15 to 19. what is the probability that the mean time they spend driving each day is between 24.7 and 22.5 minutes?

Assume that  $\mu = 25$  minutes and  $\sigma = 1.5$  minutes?

**Solution** 

From the central limit theorem (sample size is greater than 30), the sampling distribution of the sample mean is approximately normal with mean and standard deviation given as follows:

$$\mu_{\bar{x}} = \mu = 25$$
  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 1.5/\sqrt{50} \approx 0.21213$ 

Then the probability that the mean time they spend driving each day is between 24.7 and 25.5 minutes can be calculated as follows:

Normal Distribution 
$$\mu_{(xbar)} = 25 \quad \sigma_{(xbar)} = 0.21213$$

Standard Normal Distribution  $\mu = 0$   $\sigma = 1$ 

$$z_{1} = \frac{\overline{x} - \mu}{\sigma / n} = \frac{24.7 - 25}{1.5 / \sqrt{50}} = -1.41$$

$$P(24.7 \le \overline{X} \le 25.5)$$

$$z_{2} = \frac{\overline{x} - \mu}{\sigma / n} = \frac{25.5 - 25}{1.5 / \sqrt{50}} = 2.36$$

$$0.9909$$

$$0.0793$$

$$\overline{x} = \frac{0.0793}{1.41}$$

$$P(24.7 \le \overline{X} \le 25.5) = P(-1.41 \le Z \le 2.36)$$
  
= N(2.36) - N(-1.41) = 0.9909 - 0.0793 = 0.9116

z +0	.50000	<b>0.01</b> .50399	<b>0.02</b> .50798	<b>0.03</b>	<b>0.04</b> .51595	<b>0.05</b> .51994	<b>0.06</b> .52392	<b>0.07</b> .52790	<b>0.08</b> .53188	<b>0.09</b>	-			<i>-</i> · .	.ک		,	// ~		12.	ب	\	
+0.1	.53983	.54380	.54776 .58706	.55172	.55567	.55966	.56360 .60257	.56749	.57142	.57535 .61409				ئوة	51	$\subseteq \emptyset$	1 Pe	7/		As 1	8	>1	
+0.3 +0.4	.61791 .65542	.62172 .65910	.62552 .66276	.62930 .66640	.63307 .67003	.63683 .67364	.64058 .67724	.64431 .68082	.64803 .68439	.65173 .68793													
+0.5 +0.6	.69146 .72575	.69497 .72907	.69847 .73237	.70194 .73565	.70540 .73891	.70884 .74215	.71226 .74537	.71566 .74857	.71904 .75175	.72240 .75490										11 -	11 -	28	
+0.7	.75804 .78814	.76115 .79103	.76424 .79389	.76730 .79673	.77035 .79955	.77337 .80234	.77637 .80511	.77935 .80785	.78230 .81057	.78524 .81327	١.									X = =	M :	25	
+0.9	.81594 .84134	.81859 .84375	.82121 .84614	.82381 .84849	.82639 .85083	.82894 .85314	.83147 .85543	.83398 .85769	.83646 .85993	.83891 .86214													
+1.1	.86433 .88493	.86650 .88686	.86864 .88877	.87076 .89065	.87286 .89251	.87493 .89435	.87698 .89617	.87900 .89796	.88100 .89973	.88298 .90147													
+1.3	.90320 .91924	.90490 .92073	.90658 .92220	.90824	.90988	.91149 .92647	.91308 .92785	.91466	.91621 .93056	.91774 .93189	Uso								5	= _	= =	1.5 = 0.212	UZ
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179 .95254	.94295 .95352	.94408	on ti grap grec									Jr		<del>آج ه</del>	
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327	scor the l												
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670													
+2.1 +2.2 +2.3	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574													
+2.4	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158													
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520					- <del></del>	- al -			Z=	24:	1-21	) ,,,,	
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736					X =	24.7		•	-	1.	,	2 - 1.41	
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861										0.21	213		
+3.1 +3.2 +3.3	.99903	.99906 .99934	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929													
+3.3 +3.4 +3.5	.99952 .99966	.99953	.99955 .99969	.99957	.99958	.99960	.99961	.99962	.99964	.99965													
+3.5 +3.6 +3.7	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983							,		, _	255	2	5 95/	-
+3.7 +3.8 +3.9	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992					χ.	<u>- 75</u> .	り _	٠. /	U =			5 ≈ 2.36	
+3.9	.99995 .99997	.99995 .99997	.99996	.99996 .99997	.99996 .99997	.99996 .99997	.99996 .99998	.99996 .99998	.99997 .99998	.99997 .99998										0.21	213		
		0.00	0.07	0.07	0.00	0.05	0.05	0.07	0.25	0.00													
-0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414													
-0.1 -0.2	.42074	.45620	.45224	.44828	.44433	.44034	.43640	.43251	.42858	.38591													
-0.3 -0.4	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827													
-0.5 -0.6	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760										: 2	Low	5 Y \	
-0.7 -0.8	.24196	.23885	.23576	.23270	.22965	.19766	.19489	.19215	.18943	.18673													
-0.9 -1	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109													
-1.1	.11507	.11314	.11123	.12924	.12714	.12507	.12302	.12100	.11900	.09853				~ (	147	, ~	20	1			11.1 -	202	-1
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226				70	₽9. <del>T</del>	$\leq \chi$	$\leq \nu$	2.5)	= P(	-1.	41 <u></u>	2 < 2.3	6 /
-1.5 -1.6 -1.7	.06681 .05480 .04457	.06552 .05370 .04363	.06426 .05262 .04272	.06301 .05155 .04182	.06178	.06057 .04947 .04006	.05938 .04846 .03920	.05821 .04746 .03836	.05705 .04648 .03754	.05592 .04551 .03673													
-1.8 -1.9	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938	native entity												
-2 -2.1	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831													
-2.2 -2.3	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101				_		\	1,		-\	11			
-2.4 -2.5	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639					à À	e) I	Ping	1 7	120	12,	Sovi	o *	
-2.6 -2.7	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357													
-2.8 -2.9	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193									1				
-3 -3.1	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100					• 1	(7	てく	2.36	$\gamma \simeq$	0.0	190	9	
-3.2 -3.3	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050								2.36					
-3.4 -3.5	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024													
-3.6 -3.7	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011					- (F	( ]	₹	-1.41	) ~	0.0	79	3	
-3.8 -3.9	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005					,								
-4	.00003	.00003	.00003	.00003	.00003	.00003	.00002	.00002	.00002	.00002													
																					_ 0		
												0.0	190	9 -	<u>()</u> .0	793	3 =	0.9	116	= 71	$y_{co}$	\	
,																	<b>N</b> 1	_					
<i>9</i> ,91	.18	رکي).	ہ ج	@ ==	و صفح	25	59	24.	7/	- لىي	ري	العي	_عا	- -	Fer	4)	Than	180		\ 	1/ 00		
· T					′ [		-		ĺ											, –	\ \ _		
			#																				
			//																				

#### **Examples**

A bank auditor claims that credit card balances are normally distributed, with a mean of JD2870 and a standard deviation of JD900. you randomly select 25 credit card holders. What is the probability that their mean credit card balance is less than or equal to JD2500?

#### Solution

Since the population is normally distributed, the sampling distribution of the sample mean is also normally distributed. Now, we are asked to find the probability associated with a sample mean and this can be calculated as follows:

$$\mu_s = \mu = 2870$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{900}{\sqrt{25}} = 180$$

Normal Distribution

$$\mu_{\text{Obstart}} = 2870 \quad \sigma_{\text{Obstart}} = 180$$

Standard Normal Distribution 
$$\mu = 0$$
  $\sigma = 1$ 

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{2500 - 2870}{900 / \sqrt{25}} \approx -2.06$$

$$P(\bar{X} \le 2500)$$

$$y = \frac{0.0197}{2500} = -2.06$$

$$P(\overline{X} \le 2500) = P(Z \le -2.06) = N(-2.06) = 0.0197$$

Conclusion: There is only a 2% chance that the mean of a sample of size 25 will have a balance less than or equal to JD2500 (unusual event). It is possible that the sample is unusual or it is possible that the auditor's claim that the mean is JD2870 is incorrect.

	ı																								
Z	0	0.	.01	0.02	0.03	3	0.04	0.05	0.06	0.07	0.08	0.09	_					M	- M.	- = 2	187	3			
-0	.5000		601	.49202	.4880		.48405	.48006		.47210	.4681								7						
-0.1	.4601		620	.45224	.4482		.44433	.44034		.43251	.4285														
-0.2	.4207		683	.41294	.4090		.40517	.40129		.39358	.3897														
-0.3	.3820		828	.37448	.3707		.36693	.36317		.35569	.3519							_					10		
-0.4	.3445		090	.33724	.3336		.32997	.32636		.31918	.3156						(	<u>5 = 2</u>	6	= 9	00	=	(80)	1	
-0.5	.3085		503	.30153	.2980		.29460	.29116		.28434	.2809							~	TR	T	25				
-0.6	.2742		093	.26763	.2643		.26109	.25785		.25143	.2482									\	1-20				
-0.7 -0.8	.2419		885	.23576	.2327		.22965	.22663		.22065	.2177														
-0.8	.2118		897	.20611	.2032		.20045	.19766		.19215	.1894														
-0.9	.1840		141 625	.17879	.1761		.17361	.17106		.16602	.1635		٠.												
-1.1	.1356		350				.12714																		
-1.1				.13136	.1292			.12507		.12100	.1190														
-1.3	.0968		314 510	.09342	.1093		.10749	.10565		.10204	.1002														
-1.3	.0807		927	.07780	.0917		.09012	.07353	.08692	.08534	.0837			_	100	١ .		_	05	<u> </u>	1	1 2 -			~ ~
-1.4	.0668		552	.06426	.0630		.06178	.06057	.05938	.05821	.0570			$\overline{\chi}$	250		•	2	IV	500	P —	227	0~ -	-2.	06
-1.6	.0548		370	.05262	.0515		.05050	.04947		.05821	.0464									, .					
-1.7	.0445		363	.03202	.0418		.04093	.04947		.03836	.0375										80				
-1.7	.0359		515	.04272	.0336		.03288	.03216		.03074	.0300		table entry	-						· · ·				-	
-1.9	.0333		807	.02743	.0356		.03266	.02559		.02442	.0238														
-2	.0227		222	.02169	.0200		.02068	.02018		.01923	.0236														
-2.1	.0178		743	.01700	.0165		.01618	.01578		.01500	.0146														
-2.2	.0139		355	.01321	.0128		.01255	.01222		.01160	.0113														
-2.3	.0107		044	.01017	.0099		.00964	.00939		.00889	.0086														
-2.4	.0082		798	.00776	.0075		.00734	.00714		.00676	.0065											_ 31			
-2.5	.0062		604	.00587	.0057		.00554	.00539		.00508	.0049									,	1/	Jon.	1		
-2.6	.0046		453	.00440	.0042		.00415	.00402		.00379	.0036										Ů				
-2.7	.0034		336	.00326	.0031		.00307	.00298		.00280	.0027														
-2.8	.0025		248	.00240	.0023		.00226	.00219		.00205	.0019								_						
-2.9	.0018		181	.00175	.0016		.00164	.00159		.00149	.0014							0	(=	1-	_	١ _	1-		00
-3	.0013		131	.00126	.0012		.00118	.00114		.00107	.0010							P	(T	$\subseteq \mathcal{U}$	100	ツニト	(z≤	-6.	06
-3.1	.0009		094	.00090	.0008		.00084	.00082		.00076	.0007														
-3.2	.0006		066	.00064	.0006		.00060	.00058		.00054	.0005														
-3.3	.0004		047	.00045	.0004		.00042	.00040		.00038	.0003										-	-		-	
-3.4	.0003	.00	032	.00031	.0003		.00029	.00028		.00026	.0002														
-3.5	.0002		022	.00022	.0002		.00020	.00019		.00018	.0001										<u> </u>				
-3.6	.0001	6 .00	015	.00015	.0001		.00014	.00013	.00013	.00012	.0001													$\setminus \top$	
-3.7	.0001	.00	010	.00010	.0001	10	.00009	.00009	.00008	.00008	.0000	8 .00008										_	1 \ .	/\	ا ا
-3.8	.0000	7 .00	007	.00007	.0000	06	.00006	.00006	.00006	.00005	.0000	5 .00005	-										100	1.9	
-3.9	.0000	.00	005	.00004	.0000	04	.00004	.00004	.00004	.00004	.0000	3 .00003													
-4	.0000	.00	003	.00003	.0000	03	.00003	.00003	.00002	.00002	.0000	2 .00002	_												
			1	1		1					1							/-	_ ~		١				
																	F	[ [ ]	2	. 06	∫ %	0.0	(97		
- 0	,																						$\sim$		
2%	0 _ 0	101	ار آ	ועכ	1.0	) J	1 6	, 9/	1.97	\$	2.60	ر کی رہی	en le	\e_{i}	= . ~ .	mo	ا ا	م صح	Ter	1. 6	Ken				
" " "	- 70	2,	7 NO.	2.2.1	\ \	_ 1	, Am	* (	<del>- \'   \                                </del>		Ly	- Jac	1 7	10,	Just	1 E.		y -	U_V	W. A.	(m) ()	1 2			
	.1																								-
	#		v Li	S	~. \_	ابربد	1 8/0	of he la	کارک لا	1	i eei	اعتم	فایہ ا												
	7,	-95	h h	~ 427	12 A	ال کا	11 30	12 416	امت رد	100	لر مت	mare	1 45	-								1			
		12											1	1	1			1	1	1	1	1	1		

### **Examples**

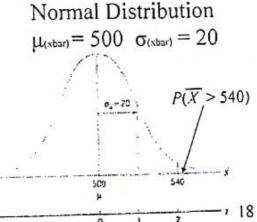
Suppose that the TOFEL exam scores for pharmacy students graduated from the Hashemite University are normally distributed with  $\mu = 500$  and  $\sigma = 100$ .

In a random sample of size n = 25 students, what is the probability that the sample mean would be greater than 540?

#### Solution

Since the population is normally distributed, the sampling distribution of the sample mean is also normally distributed. Now, we are asked to find the probability associated with a sample mean and this can be calculated as follows:

 $P(\overline{X} > 540) = 1 - P(\overline{X} \le 540)$  $= 1 - P(Z \le (540 - 500)/20)$  $= 1 - P(Z \le 2)$ = 1 - N(2)= 1 - 0.9772= 0.0228



#### **Thinking Challenge**

You're an operations analyst for AT&T. Long-distance telephone calls are normally distributed with  $\mu = 8$  min. and  $\sigma$ = 2 min. If you select random samples of 25 calls, what percentage of the sample means would be between 7.8 & 8.2 minutes?



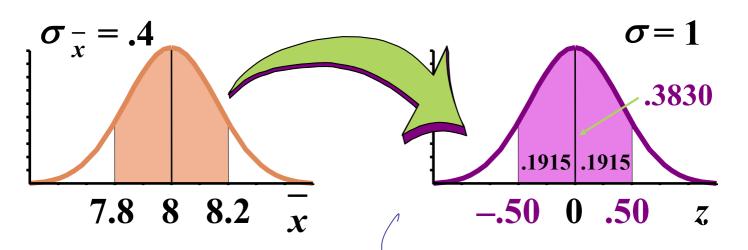
## **Sampling Distribution Solution\***

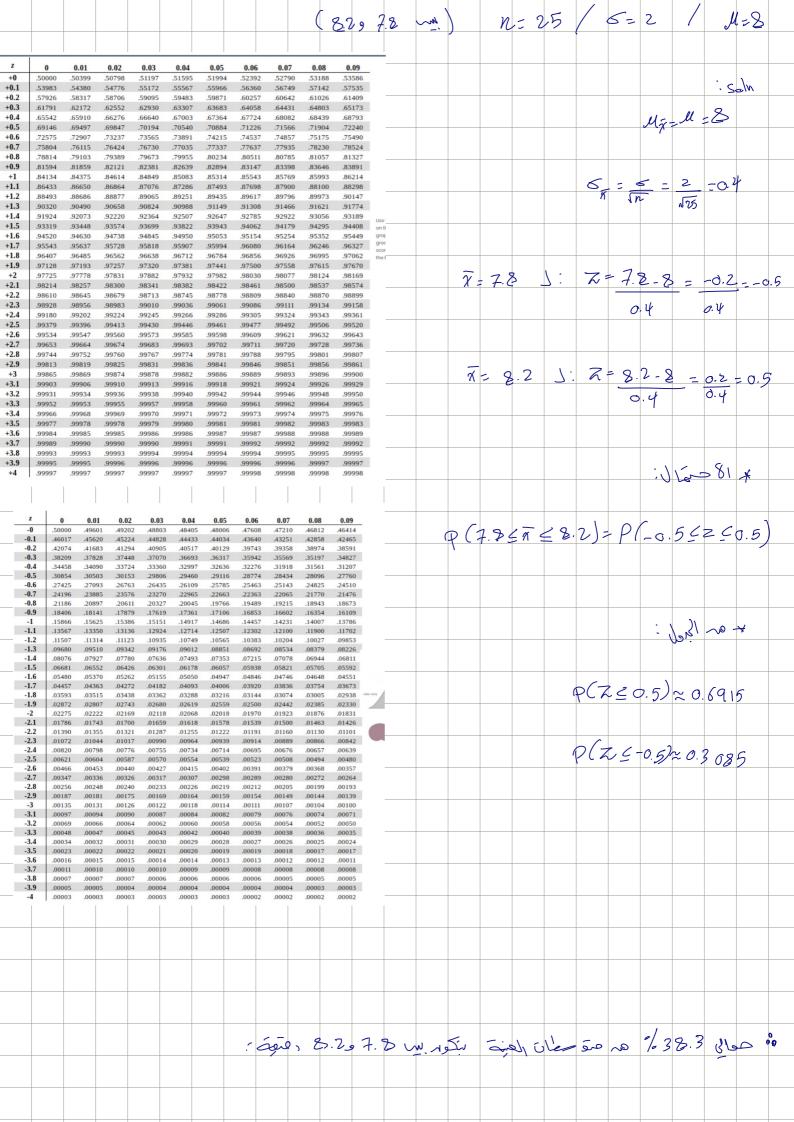
$$z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} = \frac{7.8 - 8}{2 / \sqrt{25}} = -.50$$

**Sampling Distribution** 

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{8.2 - 8}{2 / \sqrt{25}} = .50$$

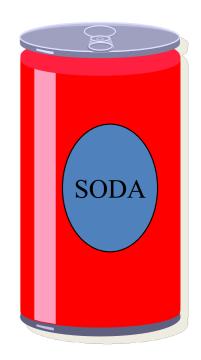
Standardized Normal Distribution



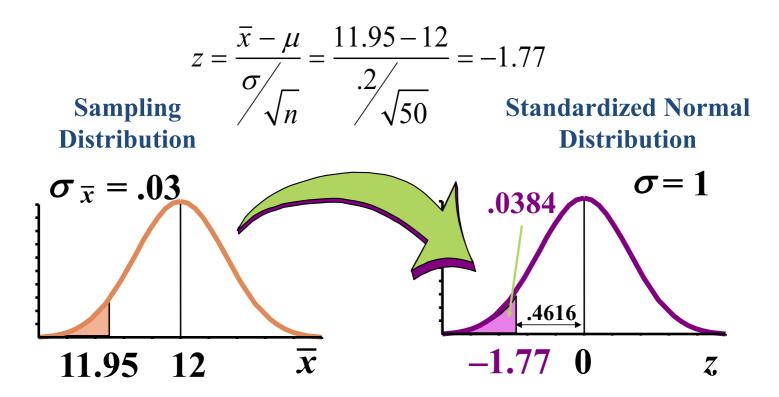


#### **Central Limit Theorem Example**

The amount of soda in cans of a particular brand has a mean of 12 oz and a standard deviation of .2 oz. If you select random samples of 50 cans, what percentage of the sample means would be less than 11.95 oz?



#### **Central Limit Theorem Solution\***



Shaded area exaggerated

## Sampling Distributions

## **Common Statistics & Parameters**

	Sample Statistic	Population Parameter
Mean	$\overline{x}$	μ
Standard Deviation	S	$\sigma$
Variance	$s^2$	$\sigma^{2}$
Binomial Proportion	<b>^</b>	p

## **Sampling Distribution**

The **sampling distribution** of a sample statistic calculated from a <u>sample of n measurements</u> is the probability distribution of the statistic in all possible samples of size n taken from the same population.

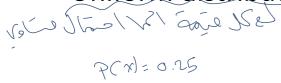
The **sampling distribution** are used to calculate the <u>probability that sample statistics</u> could have occurred by chance and thus to decide whether something that is true of sample statistics is also likely to be true of a population parameter.

(2) 2x visiliste alli de 18 min) 5 lenis ligher y 12/2.

## **Developing Sampling Distributions**

#### Suppose There's a Population ...

- Population size, N = 4
- Random variable, x
- Values of x: 1, 2, 3, 4
- Uniform distribution



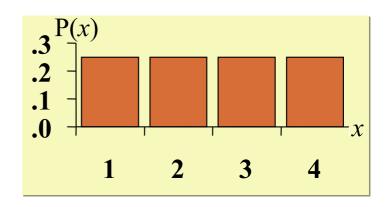


## **Population Characteristics**

#### **Summary Measure**

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = 2.5 = \frac{1 + 2 + 2 + 4}{4}$$

#### **Population Distribution**



## All Possible Samples of Size n = 2

#### 16 Samples

1st	2nd Observation			
Obs	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,4
3	3,1	3,2	3,3	3,4
4	4,1	4,2	4,3	4,4

#### 16 Sample Means

1st	2nd Observation			
Obs	1	2	3	4
1	1.0	1.5	2.0	2.5
2	1.5	2.0	2.5	3.0
3	2.0	2.5	3.0	3.5
4	2.5	3.0	3.5	4.0

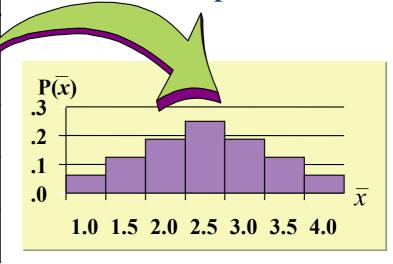
Sample with replacement

# Sampling Distribution of All Sample Means

#### 16 Sample Means

1st	2nd Observation			
Obs	1	2	3	4
1	1.0	1.5	2.0	2.5
2	1.5	2.0	2.5	3.0
3	2.0	2.5	3.0	3.5
4	2.5	3.0	3.5	4.0

**Sampling Distribution** of the Sample Mean



## **Summary Measure of All Sample Means**

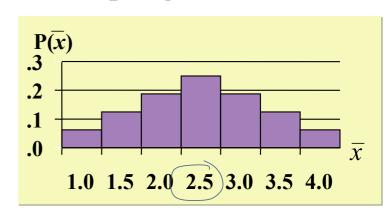
$$\mu_{\bar{X}} = \frac{\sum_{i=1}^{N} \bar{x}_{i}}{N} = \frac{1.0 + 1.5 + ... + 4.0}{16} = 2.5 = 2.5$$

## **Comparison**

#### **Population**

# 

## **Sampling Distribution**



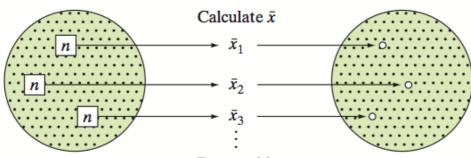
$$\mu = 2.5$$

$$\mu_{\bar{x}} = 2.5$$

## **Key Ideas**

## Generating the Sampling Distribution of $\bar{x}$

Select sample size *n* (large) from target population



Population:

Mean =  $\mu$ Std. Dev. =  $\sigma$ 

Unknown shape

Repeat this process an infinite number of times

Sampling distribution of  $\bar{x}$  (i.e., theoretical population of  $\bar{x}$ 's)

 $Mean = \mu_{\overline{x}} = \mu$ 

Std. Dev. =  $\sigma_{\overline{x}} = \sigma/\sqrt{n}$ 

Normal distribution (Central Limit Theorem)

# The Sampling Distribution of the Sample Proportion $p^{\wedge}$

#### The Sampling Distribution of the Sample Proportion $ho^{\wedge}$

- We hope that the sample proportion is close to the population proportion.
- How close can we expect it to be?
- Would it be worth it to collect a larger sample?
  - —If the sample were larger, would we expect the sample proportion to be closer to the population proportion?

How much closer? , Range Just and & Mills

## **Example**

The Jordanian Ministry of Health did a survey of in a random sample of 10904 Jordanian peoples and smoking habits. The researchers defined:

Frequent smoking as having 5 or more cigarettes in a row three or more times in the past two hours. According to this definition, 2486 persons were classified as frequent smoking. Based on these data, estimate the proportion p of all Jordanian peoples who admits to frequent smoking?

#### Solution

The true proportion (p) of Jordanian peoples who are frequent smoking can be estimated by using the sample proportion  $(\hat{p})$  which can be calculated as follows:

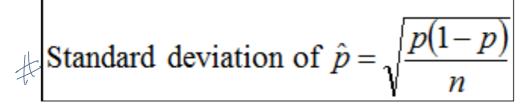
$$\hat{p} = \frac{Number\ of\ Jordanian\ persons\ who\ are\ classified\ as\ frequent\ smokin}{Number\ of\ Jordanian\ persons\ in\ the\ sample}$$

$$\hat{p} = \frac{X}{n} = \frac{2486}{10904} \neq 0.228$$

## The Sampling Distribution of the Sample Proportion $\rho^{\wedge}$

 Choose an SRS of size n from a large population that contains population proportion p of successes. Let  $p^{\wedge}$ be the sample proportion of successes,

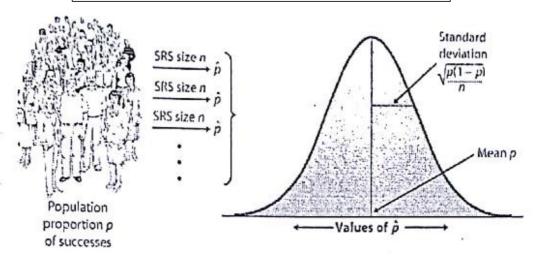
- > As the sample size increases, the sampling distribution of p becomes approximately normal.
- The mean of the sampling distribution is **p**,
- > The standard deviation of the sampling distribution is



## The Sampling Distribution of $p^{*}$

• It turns out that the sampling distribution of  $p^{n}$  is approximately normal with the following parameters.

Mean of 
$$\hat{p} = p$$
  
Variance of  $\hat{p} = \frac{p(1-p)}{n}$   
Standard deviation of  $\hat{p} = \sqrt{\frac{p(1-p)}{n}}$ 



## The Sample Proportion

- Let p be the population proportion.
- Then p is a fixed value (for a given population).
- Let  $p^{\wedge}$  ("p-hat") be the sample proportion.
- Then  $p^{\wedge}$  is a random variable; it takes on a new value every time a sample is collected.
- The sampling distribution of  $p^{\wedge}$  is the probability distribution of all the possible values of  $p^{\wedge}$ .

The Central Limit Theorem for the Sample Proportion (
$$\hat{p}$$
)

For any population proportion ( $p$ ), the sampling distribution of the sample proportion ( $\hat{p}$ ) is approximately normal if the sample size ( $n$ ) is sufficiently large. As a general guideline, the normal distribution approximation is justified when  $np \ge 5$  and  $n(1-p) \ge 5$ , that is:

Sampling Distribution of the Sample Proportion

 $\hat{p}$  Approximately  $N(p, \frac{p(1-p)}{n})$  if  $n$  is sufficiently large.

 The approximation to the normal distribution is excellent if:

$$np \ge 5$$
 and  $n(1-p) \ge 5$ .

## **Example**

Suppose that the population of interest is the Hashemite University pharmacy students. Assume that the proportion of students in the population who wear eyeglasses is p=0.25. if a random sample (SRS) of 50 students is to be selected, then define the characteristics of the sampling distribution of the sample proportion  $p^{\prime}$ , where  $p^{\prime}$  is the proportion of HU pharmacy students in the random sample who wear eyeglasses?

## **Example**

#### Solution

We know that the mean of the sampling distribution of  $\hat{p}$  is:

and we know that the variance and standard deviation of the sampling distribution of 
$$\hat{p}$$
 are:

$$\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n} = \frac{(0.25)(0.75)}{50} = 0.00375$$
;  $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{0.00375} = 0.0612$   
Finally, for sample size of  $n = 50$ , we have  $np = (50)(0.25) = (12.5)$  and  $n(1-p) = (50)(0.75) = (37.5)$  thus the large sample conditions are satisfied and the sampling distribution of  $\hat{p}$  can be approximated by a normal distribution, that is:  $\hat{p}$  Approximately  $N(p, \frac{p(1-p)}{n}) = N(0.25, 0.00375)$ .

How to Find the Probability associated with a Sample Proportion  $(\hat{p})$ ? Since the sampling distribution of  $\hat{p}$  can be approximated by a normal distribution, we can use the areas under the curve of the standard normal distribution to answer the probability question. With the normal distribution, we need to compute a Z-score value, where

$$Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p} - P}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

Notation: The above Z-score is identical to the Z value we have always used, with the exception of the notation  $\hat{p}$  and  $\sigma_{\hat{p}}$ , which is used to remind us we are dealing with the sampling distribution of the sample proportion  $(\hat{p})$ .

## **Example**

Thiazide diuretics are often the first, but not the only choice in high blood pressure medications. Suppose that 20% of all doctors in Jordan favour Thiazide diuretics. If the manger of a pharmaceutical industries company in Jordan take a random sample (SRS) of size n = 600 doctors, then what is the probability that the sample proportion ( $p^{\wedge}$ ) of doctors who favour Thiazide diuretics will be:

- 1. Between 0.18 and 0.22?  $P(0.18 \le p^* \le 0.22)$
- 2. Less than or equal to 0.18? P ( $p^{4} \le 0.18$ )
- 3. More than 0.22? P ( $p^{\wedge} \ge 0.22$ )

( 1/2 lot siz ) h.P=600 x 0.2 = 1207,5 1/ ng = 600 + 0. 8 = 420 7/5 Mô= P=0.2  $\frac{\sqrt{500} - \sqrt{500}}{\sqrt{500}} = \frac{0.16}{\sqrt{500}} = \frac{0.0002667}{\sqrt{500}} \approx 0.01633 \qquad N(0.2, 0.01633)$ Z = p - Mp (P) P(0.18 = P = 0.22) 7 = 0.18-0.2 ~ -0.02 2-122 : p=0.18  $Z_{n} = 0.22 - 0.2 \times 0.02 \times 0.02 \text{ if } = 0.22$ P(-1.22 52 51.22)=P(25122)-P(Z 5-1.22) 1988 8.0x (2 51.22) x0.888 ME Pigl =18 = P(S & 1,22) 20.8888 JEnogy - 0.8 882 - 0.1112 = 0.7776 + 0.09 .53586 .57535 .47210 .46812 .43251 .42858 .46414 .42465 .50000 .50399 .53983 .54380 .50798 .51595 .55567 .52790 .53188 .56749 .57142 +0.2 .57926 .58317 .61791 .62172 .58706 .62552 .59095 .62930 .59483 .59871 .63683 .60642 .61026 .64431 .64803 .61409 .65173 +0.4 .65910 .69497 .66276 .69847 .66640 .67003 .70540 .67364 .67724 .70884 .71226 .68082 .68439 .71566 .71904 .69146 .72240 .24510 .71566 .71904 .74857 .75175 .77935 .78230 .80785 .81057 .83398 .83646 .75490 .78524 .81327 .83891 .86214 .88298 .19766 .17106 .14686 .12507 .10565 .08851 .07353 .06057 .04947 .04006 .03216 .02559 .02018 .18673 +1.2 .89065 .90824 .89435 .91149 .89973 .91621 .90147 .91774 .90320 .91308 .91466 .91466 .92922 .94179 .95254 .96164 .96926 .97558 +1.4 .91924 .93319 .93056 .94295 .93448 .93574 .94630 .95637 .96485 .97193 -2 .02275
-2.1 .01786
-2.2 .01390
-2.3 .01072
-2.4 .00820
-2.5 .00621 .97831 .98300 .97882 .98341 .98077 .98124 .98500 .98537 .98169 .98574 +2.1 .97725 .98214 +2.2 .98610 .98645 .98928 .98956 .98679 .98713 .98745 .98983 .99010 .99036 .98778 .98809 .99061 .99086 .98840 .98870 .99111 .99134 .98899 .99158 .99010 .99036 .99245 .99266 .99430 .99446 .99573 .99585 .99683 .99693 .99767 .99774 .99831 .99836 .99878 .99836 .99111 .99134 .99324 .99343 .99492 .99506 .99621 .99632 .99720 .99728 .99795 .99801 .99851 .99856 .99224 .99413 .99560 .99674 .99361 .99520 .99643 .99736 +2.4 .99180 .99379 .99202 .99396 .99286 .99305 .99461 .99477 -2.6 .00466 -2.7 .00347 -2.8 .00256 -2.9 .00187 .00415 .00402 .00440 .00233 -3 -3.1 -3.2 .00069 -3.3 .00048 .99900 -3.3 .00048 -3.4 .00034 -3.5 .00023 -3.6 .00016 -3.7 .00011 -3.8 .00007 -3.9 .00005 -4 .00003 +3.2 .99931 .99934 .99952 .99953 .99946 .99948 .99962 .99964

#### Solution

We know that the mean of the sampling distribution of  $\hat{p}$  is:

$$\mu_{\mathfrak{P}} = p = 20\% = 0.2$$

and we know that the variance and standard deviation of the sampling

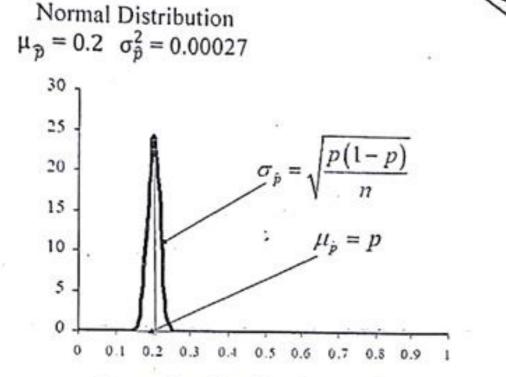
distribution of 
$$\hat{p}$$
 are:  

$$\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n} = \frac{(0.2(0.8)}{600} = 0.00027 \quad ; \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{0.00027} = 0.0164$$

Finally, for sample size of n = 600, we have np = (600)(0.2) = 120 and n(1-p) = (600)(0.8) = 480; thus the large sample conditions are satisfied and the sampling distribution of  $\hat{p}$  can be approximated by a normal distribution, that is:  $\hat{p}$  Approximately  $N(p, \frac{p(1-p)}{n}) = N(0.2, 0.00027)$ .

Normal Distribution

Distribution of the Sample Proportion



Samplingdistribution of  $\hat{p}$ 

Now, the required probabilities can be found as follows:

(i) Probability that the sample proportion (
$$\hat{p}$$
) of doctors who favor Thiazide diuretics will be between 0.18 and 0.22 can be calculated as follows:

$$P(0.18 \le \hat{p} \le 0.22) = P(\frac{0.18 - 0.2}{0.0164} \le \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \le \frac{0.22 - 0.2}{0.0164})$$

$$= P(-1.22 \le Z \le 1.22)$$

$$= N(1.22) - N(-1.22) = 0.8888 - 0.1112$$

= 0.7776

(ii) Probability that the sample proportion 
$$(\hat{p})$$
 of doctors who favor Thiazide

diuretics will be less than or equ 18 can be calculated as follows:  $P(\hat{p} \le 0.18) = P(\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \le \frac{0.18-}{0.0164}$ 

$$= P(Z \le -1.22) = N(-1.22) = 0.1112$$

(iii) Probability that the sample proportion (p) of doctors who favor Thiazide diuretics will be more than 0.22 can be calculated as follows:

diuretics will be more than 0.22 can be calculated as follows:  

$$P(\hat{p} > 0.22) = 1 - P(\hat{p} \le 0.22) = 1 - P(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \le \frac{0.22 - 0.2}{0.0164})$$

 $= 1 - P(Z \le 1.22) = 1 - N(1.22)$ = 1 - 0.8888 = 0.1112