MATH 1700

Instructor: Mehdi Maadooliat

Chapter 7

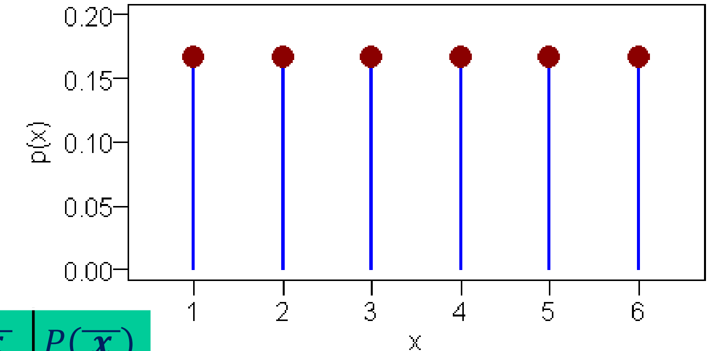
Department of Mathematical and Statistical Sciences

CHAPTER 7

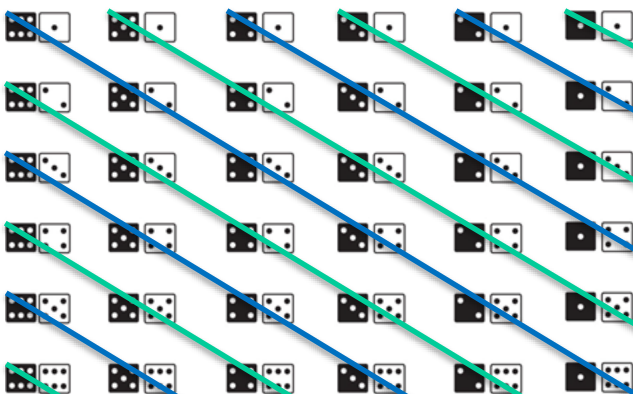
- What is the relationship between distribution of x and $\bar{x}$?
- Sampling distribution
- Random sampling
- Sampling distribution of sample means ($SDSM$)
- Central limit theorem (CLT)
- Standard error of the mean

EXAMPLE – ROLLING A PAIR OF DICE

- **Roll a die. Let** $x =$ the number we see**. The distribution of** x**:**

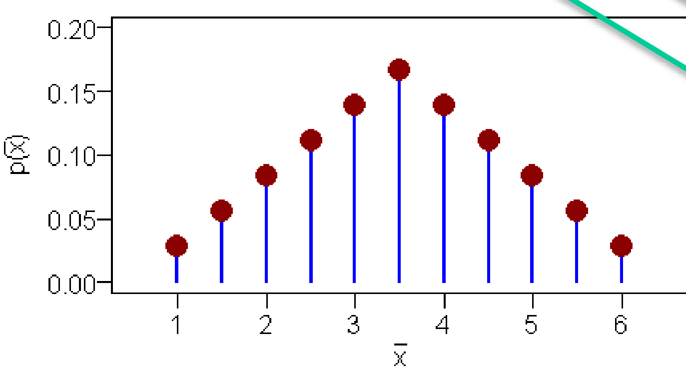


- **Roll two dice, say** x_1, x_2
  - **What is the distribution of** $\bar{x} = \frac{x_1+x_2}{2}$ **?**

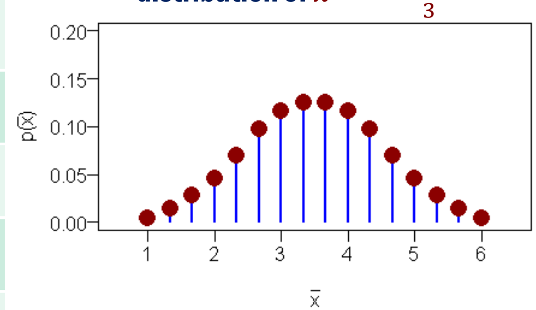


$\bar{x}$	$P(\bar{x})$
1	$^1/_{36}$
1.5	$^2/_{36}$
2	$^3/_{36}$
2.5	$^4/_{36}$
3	$^5/_{36}$
3.5	$^6/_{36}$
4	$^5/_{36}$
4.5	$^4/_{36}$
5	$^3/_{36}$
5.5	$^2/_{36}$
6	$^1/_{36}$

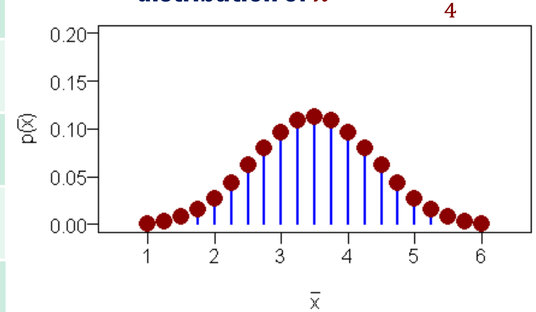
- **The distribution of** $\bar{x}$ **is:**



- **How about three dice?**
- **distribution of** $\bar{x} = \frac{x_1+x_2+x_3}{3}$



- **How about four dice?**
- **distribution of** $\bar{x} = \frac{x_1+x_2+x_3+x_4}{4}$



POPULATION SAMPLING

- **A census, a 100% survey or sampling, in the United States is done only every 10 years. It is an enormous and overwhelming job, but the information that is obtained is vital to our country's organization and structure.**
- **Issues come up and times change; information is needed and a census is impractical. This is where representative and everyday samples come in.**
- **Sampling distribution of a sample statistic**
  - **The distribution of values for a sample statistic obtained from repeated samples, all of the same size and all drawn from the same population.**

THE SAMPLING ISSUE

The fundamental goal of a survey is to come up with the same results that would have been obtained had every single member of a population been interviewed. For national Gallup polls the objective is to present the opinions of a sample of people that are exactly the same opinions that would have been obtained had it been possible to interview all adult Americans in the country.

The key to reaching this goal is a fundamental principle called *equal probability of selection*, which states that if every member of a population has an equal probability of being selected in a sample, then that sample will be representative of the population. It's that straightforward.

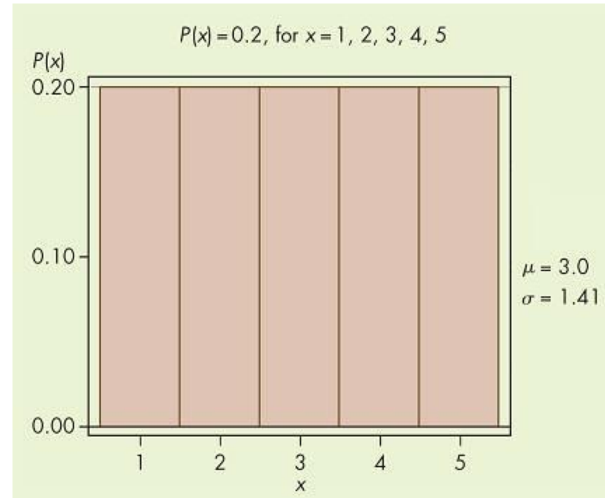
Thus, it is Gallup's goal in selecting samples to allow every adult American an equal chance of falling into the sample. How that is done, of course, is the key to the success or failure of the process.

Source: Reprinted by permission of the Gallup Organization, http://www.gallup.com/

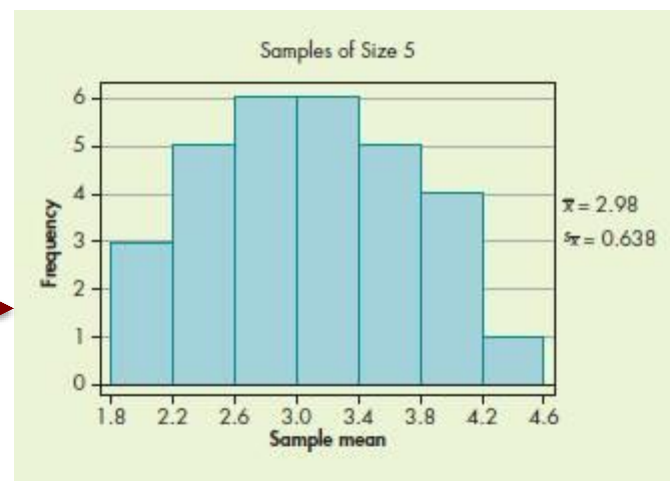
EXAMPLE 2 – CREATING A SAMPLING DISTRIBUTION OF SAMPLE MEANS

- **Let's consider a population that consists of five equally likely integers: 1, 2, 3, 4, and 5.**
- **Randomly choose 30 samples of size 5 from this population.**

No.	Sample	$\bar{x}$	No.	Sample	$\bar{x}$
1	4,5,1,4,5	3.8	16	4,5,5,3,5	4.4
2	1,1,3,5,1	2.2	17	3,3,1,2,1	2.0
3	2,5,1,5,1	2.8	18	2,1,3,2,2	2.0
4	4,3,3,1,1	2.4	19	4,3,4,2,1	2.8
5	1,2,5,2,4	2.8	20	5,3,1,4,2	3.0
6	4,2,2,5,4	3.4	21	4,4,2,2,5	3.4
7	1,4,5,5,2	3.4	22	3,3,5,3,5	3.8
8	4,5,3,1,2	3.0	23	3,4,4,2,2	3.0
9	5,3,3,3,5	3.8	24	3,3,4,5,3	3.6
10	5,2,1,1,2	2.2	25	5,1,5,2,3	3.2
11	2,1,4,1,3	2.2	26	3,3,3,5,2	3.2
12	5,4,3,1,1	2.8	27	3,4,4,4,4	3.8
13	1,3,1,5,5	3.0	28	2,3,2,4,1	2.4
14	3,4,5,1,1	2.8	29	2,1,1,2,4	2.0
15	3,1,5,3,1	2.6	30	5,3,3,2,5	3.6

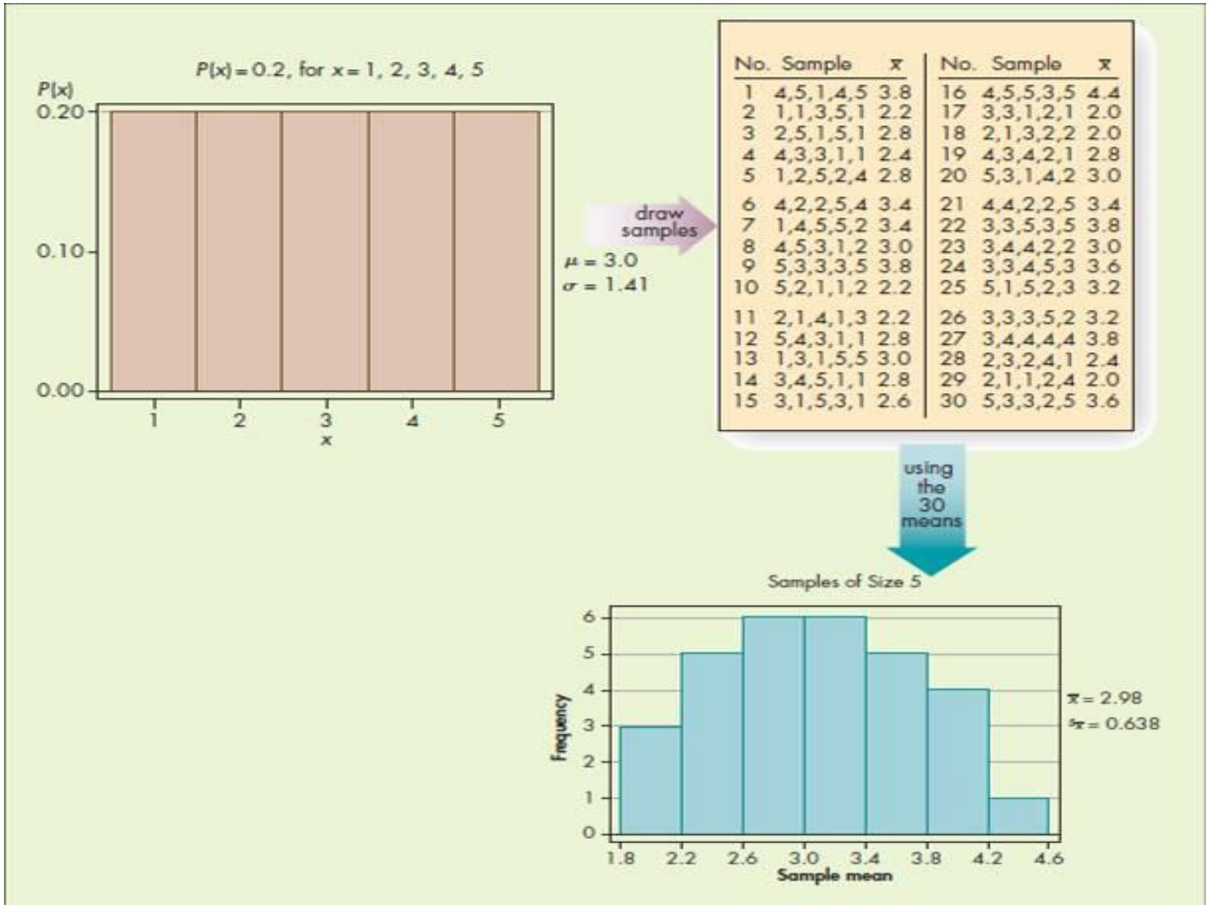


The Population: Theoretical Probability Distribution



Frequency Distribution of Sample Means

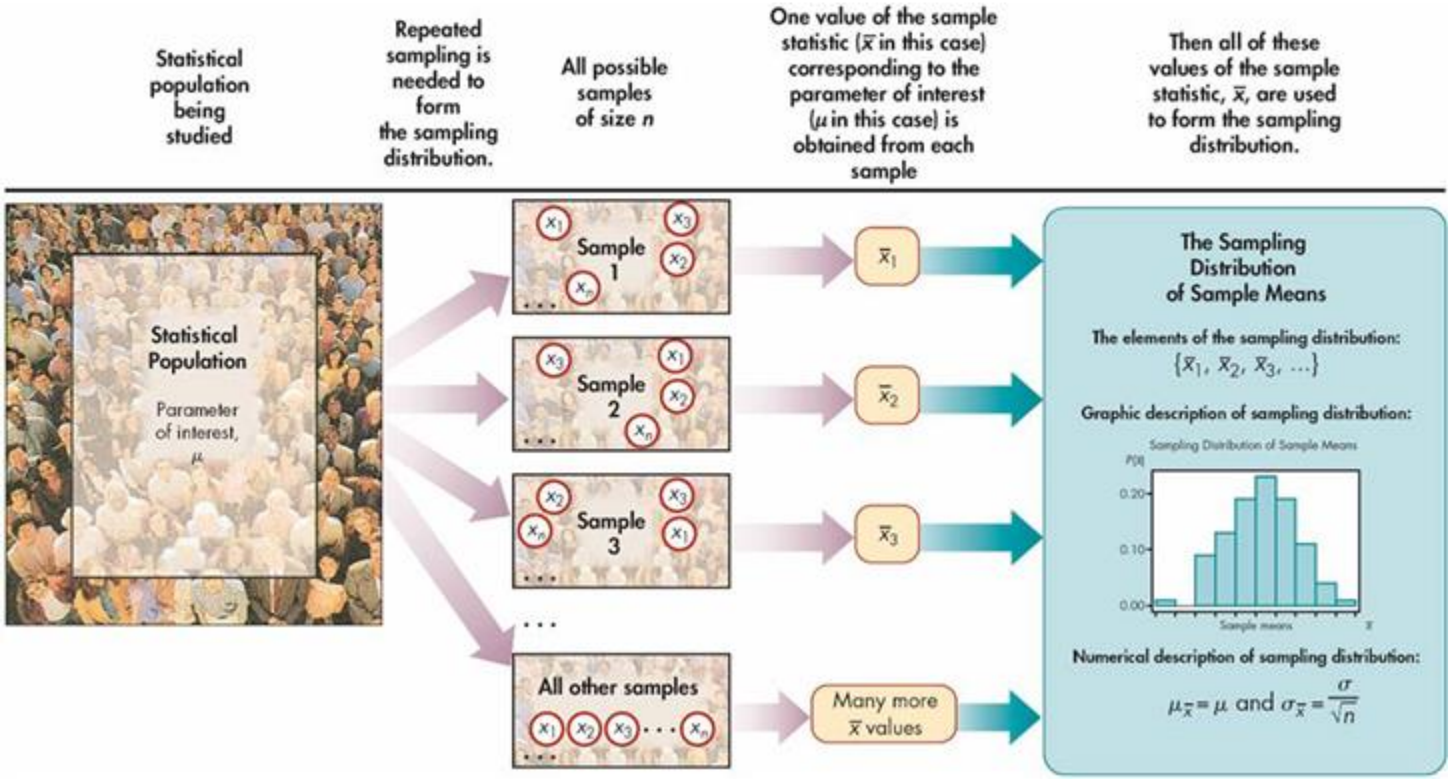
EXAMPLE 2 – CREATING A SAMPLING DISTRIBUTION OF SAMPLE MEANS



- **Looks like, under certain assumptions, the Sampling distribution of the sample mean approaches the normal distribution. (Sampling distribution applet)**

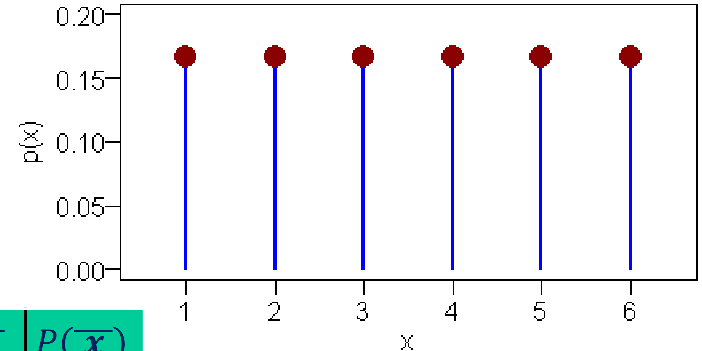
SAMPLING DISTRIBUTIONS

- **Notes**
  - The variable for the sampling distribution is $\bar{x}$; therefore, the mean of the $\bar{x}$'s is $\bar{\bar{x}}$ and the standard deviation of $\bar{x}$ is $s_{\bar{x}}$.
  - The theory involved with sampling distributions that will be described in the remainder of this chapter requires *random sampling*.
  - **Random sample:** A sample obtained in such a way that each possible sample of fixed size n has an equal probability of being selected.

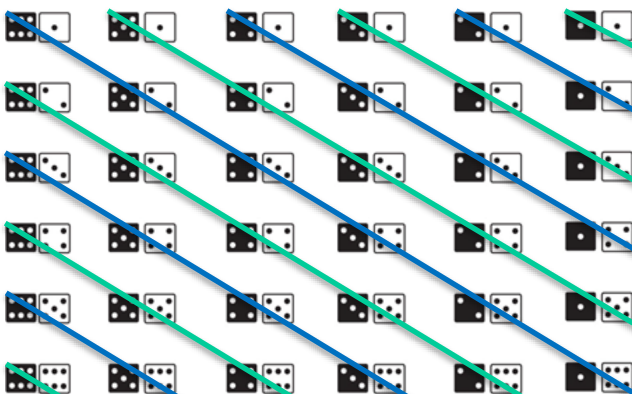


EXAMPLE – ROLLING A PAIR OF DICE

- **Roll a die. Let** $x =$ the number we see. **The distribution of** x**:**

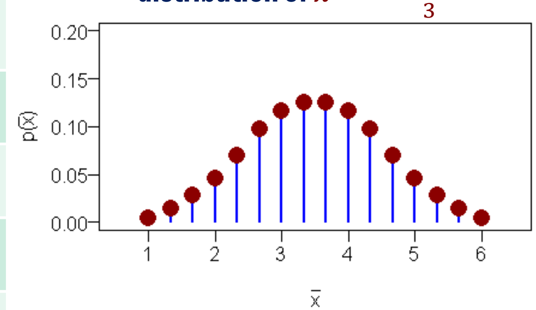


- **Roll two dice, say** x_1, x_2
  - **What is the distribution of** $\bar{x} = \frac{x_1+x_2}{2}$ **?**

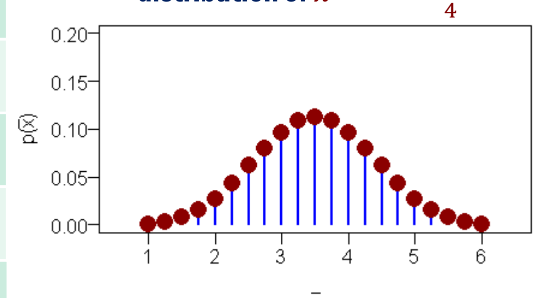


$\bar{x}$	$P(\bar{x})$
1	$^1/_{36}$
1.5	$^2/_{36}$
2	$^3/_{36}$
2.5	$^4/_{36}$
3	$^5/_{36}$
3.5	$^6/_{36}$
4	$^5/_{36}$
4.5	$^4/_{36}$
5	$^3/_{36}$
5.5	$^2/_{36}$
6	$^1/_{36}$

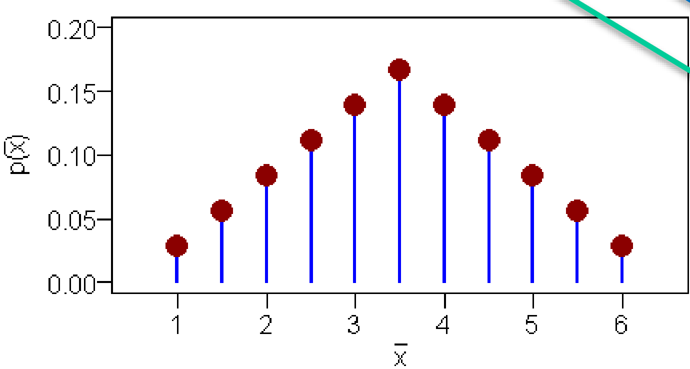
- **How about three dice?**
- **distribution of** $\bar{x} = \frac{x_1+x_2+x_3}{3}$



- **How about four dice?**
- **distribution of** $\bar{x} = \frac{x_1+x_2+x_3+x_4}{4}$



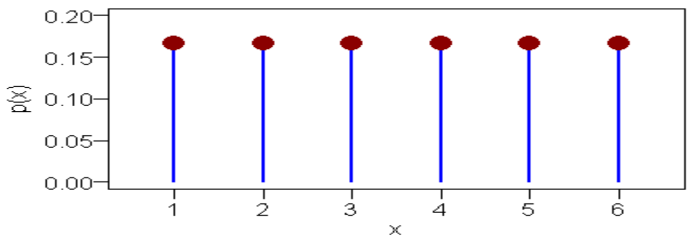
- **The distribution of** $\bar{x}$ **is:**



EXAMPLE – ROLLING A PAIR OF DICE
LET'S COMPARE THE DIST. OF X AND $\bar{X}$

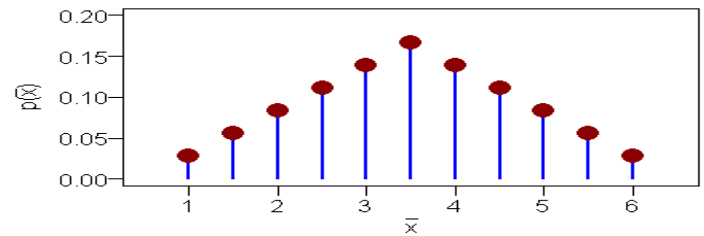
MARQUETTE UNIVERSITY
Be The Difference.

- **dist. of** x = the # we see.



- $\mu_x = \sum_{i=1}^{6} x_i P(x_i)$
- $= \frac{1}{6}(1+2+3+4+5+6)$
- $= 3.5$

- **distribution of** $\bar{x} = \dfrac{x_1+x_2}{2}$



- $\mu_{\bar{x}} = \sum_{i=1}^{11} \bar{x}_i P(\bar{x}_i)$
- $= 1\times\frac{1}{36} + 1.5\times\frac{2}{36} + \cdots + 6\times\frac{1}{36}$
- $= 3.5$

$\bar{x}$	$P(\bar{x})$
1	$1/36$
1.5	$2/36$
2	$3/36$
2.5	$4/36$
3	$5/36$
3.5	$6/36$
4	$5/36$
4.5	$4/36$
5	$3/36$
5.5	$2/36$
6	$1/36$

$$\boxed{\mu_x = \mu_{\bar{x}}}$$

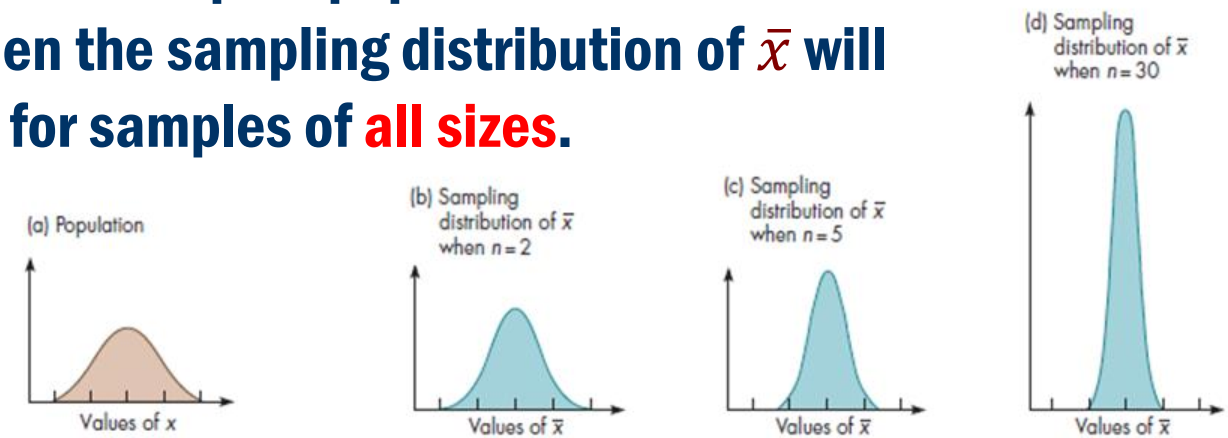
- $\sigma_x^2 = \sum_{i=1}^{6} x_i^2 P(x_i) - \mu_x^2$
- $= \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) - 3.5^2$
- $= \frac{35}{12} = 2.92$

- $\sigma_{\bar{x}}^2 = \sum_{i=1}^{11} \bar{x}_i^2 P(\bar{x}_i) - \mu_{\bar{x}}^2$
- $= 1^2\times\frac{1}{36} + 1.5^2\times\frac{2}{36} + \cdots + 6^2\times\frac{1}{36} - 3.5^2$
- $= \frac{35}{24} = 1.46$

$$\boxed{\sigma_{\bar{x}}^2 = \sigma_x^2/2}$$

THE SAMPLING DISTRIBUTION OF SAMPLE MEANS

- **Sampling distribution of sample means ($SDSM$) If all possible random samples, each of size n, are taken from any population with mean μ and standard deviation σ, then the sampling distribution of sample means will have the following:**
  1. **A mean $\mu_{\bar{x}}$ equal to μ**
  2. **A variance $\sigma_{\bar{x}}^2$ equal to $\frac{\sigma^2}{n}$**
  3. **A standard deviation $\sigma_{\bar{x}}$ equal to $\frac{\sigma}{\sqrt{n}}$**
- **Furthermore, if the sampled population has a normal distribution, then the sampling distribution of $\bar{x}$ will also be normal for samples of all sizes.**
- **[Sampling distribution applet]**



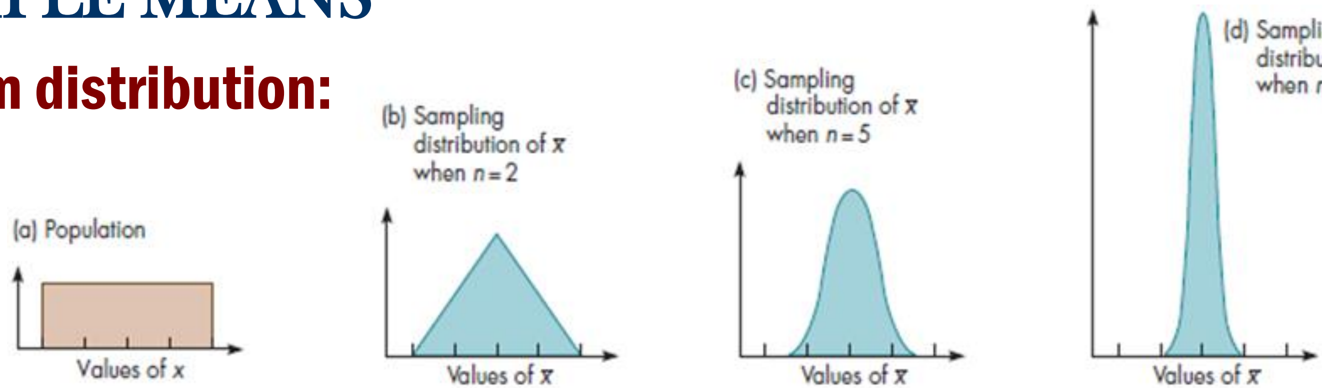
THE SAMPLING DISTRIBUTION OF SAMPLE MEANS

- **Central limit theorem (CLT): The sampling distribution of sample means will more closely resemble the normal distribution as the sample size increases.**
- **The CLT is extremely important (the most) in Statistics.**

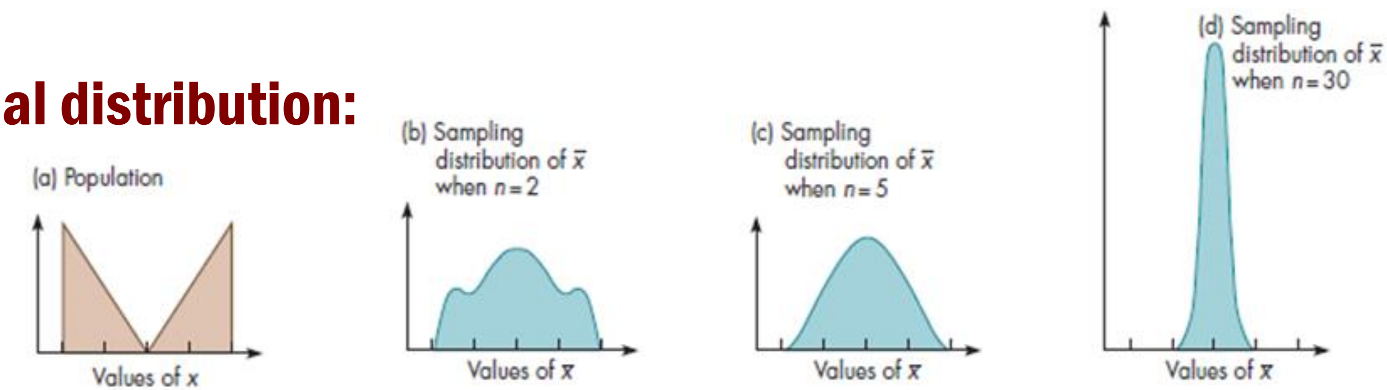
- **Notes:**
- **If the sampled distribution is normal, then the sampling distribution of sample means ($SDSM$) is normal, as stated previously, and the central limit theorem (CLT) is not needed.**
- **But, if the sampled population is not normal, the CLT tells us that the sampling distribution will still be approximately normally distributed under the right conditions.**
- **Let's consider the following examples in the next slide.**
- **Sampling distribution applet**

THE SAMPLING DISTRIBUTION OF SAMPLE MEANS

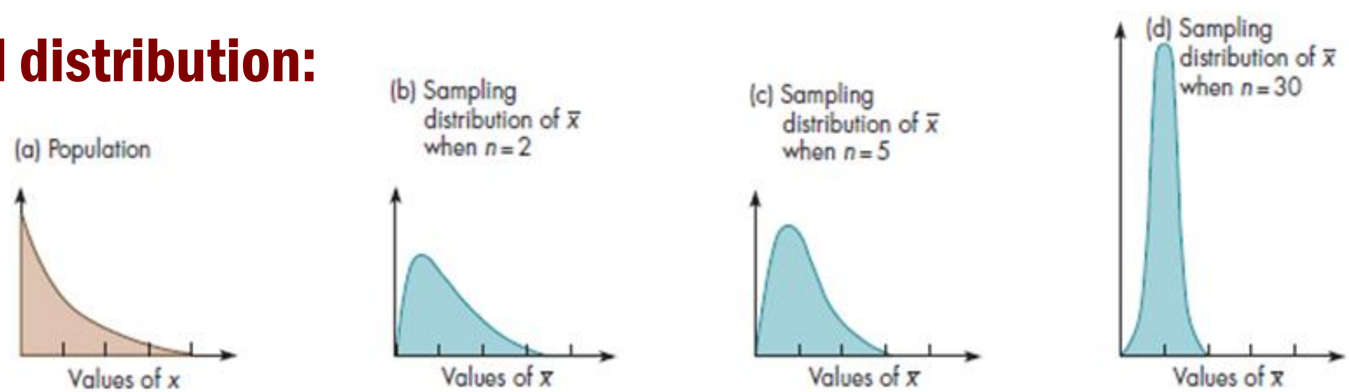
- **Uniform distribution:**



- **Bimodal distribution:**



- **Skewed distribution:**



THE SAMPLING DISTRIBUTION OF SAMPLE MEANS

- **Notes:**
  - If the sampled population distribution is nearly normal, the distribution is approximately normal for fairly small n (possibly as small as 15).
  - When the sampled population distribution lacks symmetry, n may have to be quite large (maybe 50 or more) before the normal distribution provides a satisfactory approximation.
  - All three nonnormal population distributions seem to verify the $\boldsymbol{CLT}$; the sampling distributions of sample means appear to be approximately normal for all three when samples of size 30 are used.

  - Standard error of the mean ($\sigma_{\bar{x}} = \sigma/\sqrt{n}$): is used to denote to the standard deviation of the sampling distribution of sample means.
  - The n referred to is the size of each sample in the sampling distribution. (The number of repeated samples used in an empirical situation has no effect on the standard error.)

THE SAMPLING DISTRIBUTION OF SAMPLE MEANS

- **Final Notes:**

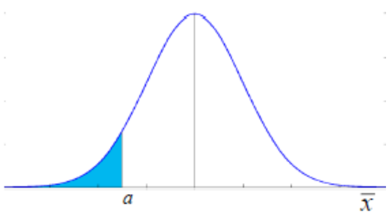
- **With the normal population, the sampling distributions of the sample means for all sample sizes appear to be normal.**

- **Also, you have seen an amazing phenomenon: No matter what the shape of a population, the sampling distribution of sample means either is normal or becomes approximately normal when n becomes sufficiently large.**

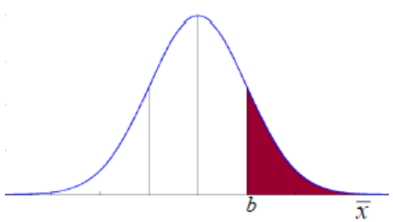
- **You should notice one other point: The sample mean becomes less variable as the sample size increases. Notice that as n increases from 2 to 30, all the distributions become narrower and taller.**

APPLICATION OF THE SAMPLING DISTRIBUTION OF SAMPLE MEANS

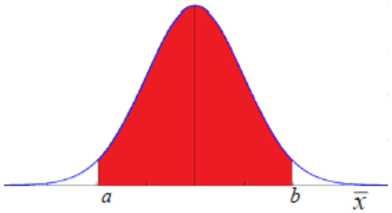
- **When the sampling distribution of sample means is normally distributed, or approximately normally distributed, we will be able to answer probability questions with the aid of the standard normal distribution.**
- **Given $\bar{x}$ is approximately normal with**
  - **mean:** $\mu_{\bar{x}} = \mu$, **and**
  - **standard deviation:** $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- $P(\bar{x} < a) = P(z < z_a)$, **where** $z_a = \frac{a - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$



- $P(\bar{x} > b) = 1 - P(z < z_b)$, **where** $z_b = \frac{b - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$



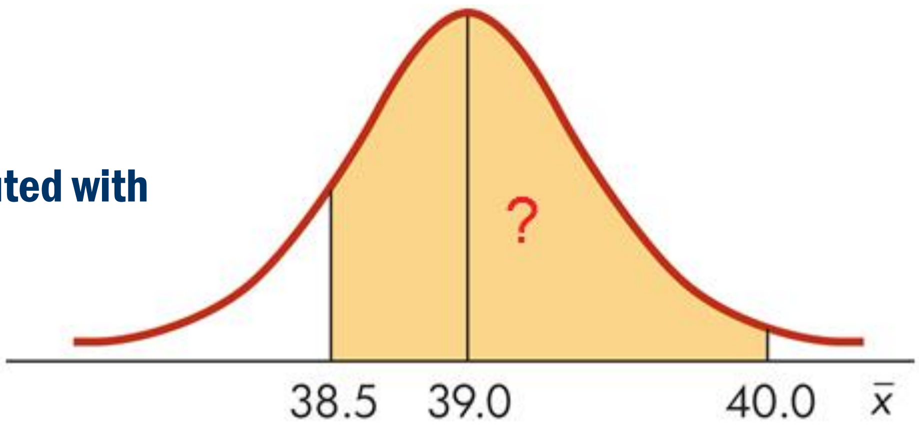
- $P(a < \bar{x} < b) = P(z < z_b) - P(z < z_a)$



EXAMPLE 6 – CALCULATING MEAN HEIGHT LIMITS FOR THE MIDDLE 90%

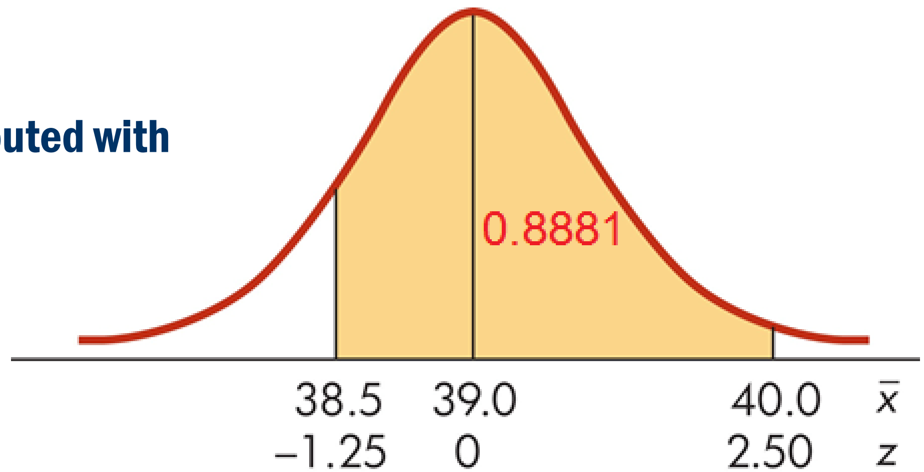
- Kindergarten children have heights that are approximately normally distributed about a mean of 39 inches and a standard deviation of 2 inches.
- A random sample of size 25 is taken, and the mean is calculated. What is the probability that this mean value will be between 38.5 and 40.0 inches?

- **Solution:**
- Given information
  - x is approximately normally distributed with
  - mean: $\mu = 39$, and
  - standard deviation: $\sigma = 2$
  - sample size: $n = 25$
  - What is $P(38.5 < \bar{x} < 40)$?



EXAMPLE 6 – SOLUTION

- **Given information**
  - x **is approximately normally distributed with**
  - **mean:** $\mu = 39$**, and**
  - **standard deviation:** $\sigma = 2$
  - **sample size:** $n = 25$
  - **What is** $P(38.5 < \bar{x} < 40)$**?**
  - $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$
  - $\bar{x} = 38.5\text{:}\;\; z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{38.5-39}{2/\sqrt{25}} = \frac{-0.5}{0.4} = -1.25$
  - $\bar{x} = 40\text{:}\;\; z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{40-39}{2/\sqrt{25}} = \frac{1}{0.4} = 2.5$
  - $P(38.5 < \bar{x} < 40) = P(-1.25 < z < 2.5)$
  - $= P(z < 2.5) - P(z < -1.25)$
  - $= 0.9938 - 0.1057$
  - $= \mathbf{0.8881}$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1094	0.1075	0.1057	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1563	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952

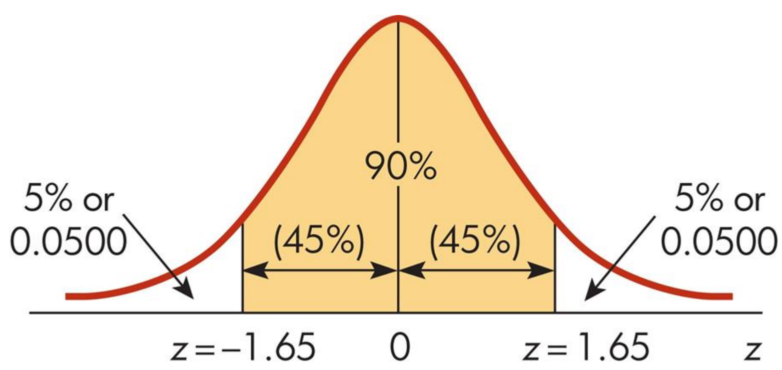
EXAMPLE 7 – CALCULATING MEAN HEIGHT LIMITS FOR THE MIDDLE 90%

MARQUETTE UNIVERSITY
Be The Difference.

- **Kindergarten children have heights that are approximately normally distributed about a mean of 39 inches and a standard deviation of 2 inches.**

- **Within what limits does the middle 90% of the sampling distribution of sample means for samples of size 100 fall?**

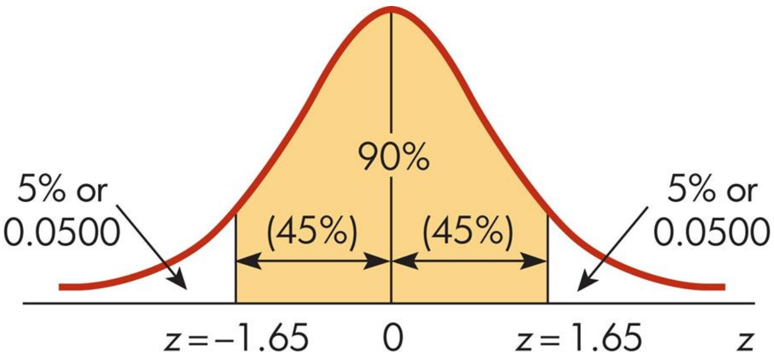
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7258	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7996	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936



EXAMPLE 7 – SOLUTION

MARQUETTE UNIVERSITY
Be The Difference.

- **First, using Table 3, we find that the middle 0.9000 is bounded by $z = \pm 1.65$.**
  - $P\ (-1.65 < z < 1.65) = 0.90$



- **Second, we use formula (7.2),**

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}:$$

$$z = -1.65: \quad -1.65 = \frac{\bar{x} - 39.0}{2/\sqrt{100}}$$
$$\bar{x} - 39 = (-1.65)(0.2)$$
$$\bar{x} = 39 - 0.33$$
$$= \mathbf{38.67}$$

$$z = 1.65: \quad 1.65 = \frac{\bar{x} - 39.0}{2/\sqrt{100}}$$
$$\bar{x} - 39 = (1.65)(0.2)$$
$$\bar{x} = 39 + 0.33$$
$$= \mathbf{39.33}$$

- **Thus,**

$$\boxed{P(38.67 < \bar{x} < 39.33) = 0.90}$$

DISCUSSION ON COURSE

- **Part I: Descriptive Statistics**
- Chapter 1: Statistics
  - Background material.
  - Definitions.
- Chapter 2: Descriptive Analysis and Presentation of single variable data
  - Graphs,
  - Central Tendency,
  - Dispersion,
  - Position
- Chapter 3: Descriptive Analysis and Presentation of bivariate data
  - Scatter plot,
  - Correlation,
  - Regression

DISCUSSION ON COURSE

- **Part II: Probability**
- Chapter 4: Probability
  - Conditional,
  - Rules,
  - Mutually Exclusive,
  - Independent
- Chapter 5: Probability Distributions (Discrete)
  - Random variables,
  - Probability Distributions,
  - Mean & Variance,
  - Binomial Distribution with Mean & Variance
- Chapter 6: Probability Distributions (Continuous)
  - Normal Distribution,
  - Standard Normal,
  - Applications,
  - Notation
- Chapter 7: Sample Variability
  - Sampling Distributions,
  - SDSM,
  - CLT

DISCUSSION ON COURSE

- **Part III: Inferential Statistics**
- **Chapter 8: Introduction to Statistical Inferences**
  - Confidence Intervals and
  - Hypothesis testing,
  - for Population mean μ (when σ is known),
- **Chapter 9: Inferences Involving One Population**
  - Mean μ (when σ is unknown),
  - Proportion p,
  - Variance σ^2
- **Chapter 10: Inferences Involving Two Populations**
  - Difference in means $\mu_1 - \mu_2$,
  - Proportions $p_1 - p_2$,
  - Variances σ_1^2/σ_2^2
- **Chapter 11: Applications of Chi-Square**
  - Chi-square statistics. .... We will discuss later.
- **Chapter 12: Analysis of Variance (ANOVA)**

QUESTIONS?

- ANY QUESTION?