The Central Limit Theorem

MAT 152 – STATISTICAL METHODS I
LECTURE 1
INSTRUCTOR: DUSTIN ROTEN
FALL 2020

A Quick Review

DISCRETE DISTRIBUTIONS

Binomial Distribution

- What is the probability of x successes in n trials?

Geometric Distribution

- What is the probability of x trails before a success occurs?

Poisson Distribution

- Given a rate of events per interval, μ, what is the probability of x events?

CONTINUOUS DISTRIBUTIONS

Uniform Distribution

- Along an interval, values are all equally likely to occur.

Exponential Distribution

- The probability of observing certain waiting times between events.

Normal Distribution

- The probability of obtaining certain values that are centered about a mean.

The Central Limit Theorem

Suppose we wanted to find the average (mean) dollar amount that Americans spend each week at the grocery store. The population of the United States is over 300 million individuals; therefore, conducting a census is not possible. (The population in this study is the entire United States.)

Instead of trying to collect a value for everyone in the United States, what if the means of “bits and pieces” of the population were calculated instead?

What properties would these bits and pieces have?

The Central Limit Theorem

Population

Sample 3 (100 people)
$\bar{x}_3 = \$53.42$

Sample 6 (100 people)
$\bar{x}_6 = \$92.06$

Sample 9 (100 people)
$\bar{x}_9 = \$67.54$

Sample 2 (100 people)
$\bar{x}_2 = \$97.50$

Sample 5 (100 people)
$\bar{x}_5 = \$79.31$

Sample 8 (100 people)
$\bar{x}_8 = \$143.69$

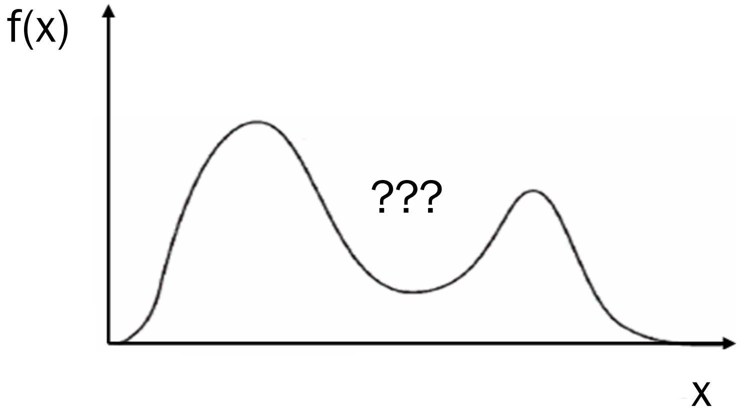
Sample 1 (100 people)
$\bar{x}_1 = \$103.75$

Sample 4 (100 people)
$\bar{x}_4 = \$133.85$

Sample 7 (100 people)
$\bar{x}_7 = \$121.37$

The Mean of the Means

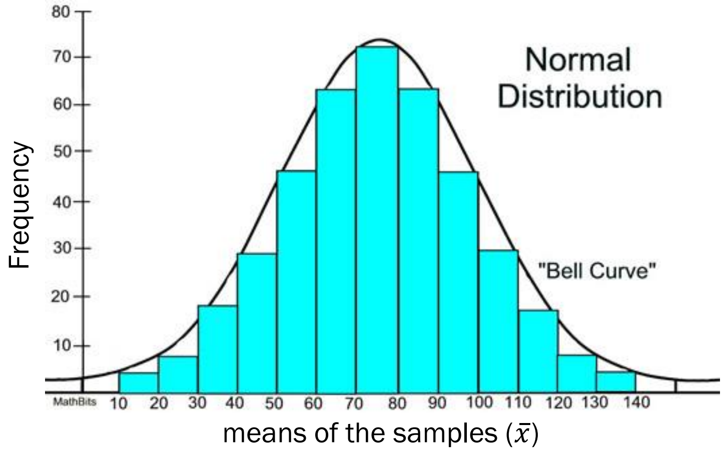
POPULATION (UNKNOWN)



$\mu_X = ???$

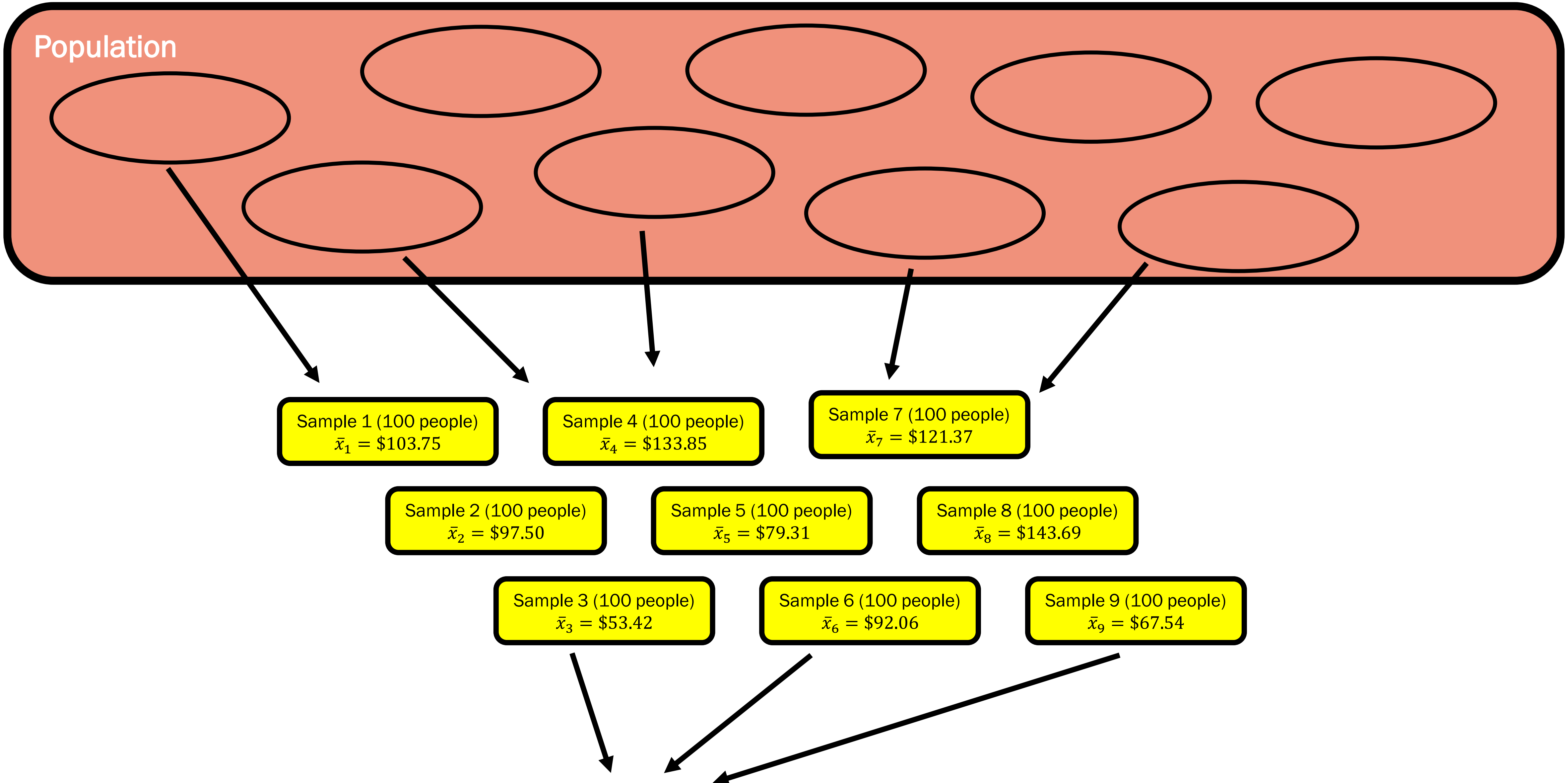
$\sigma_X = ???$

SAMPLES



$\mu_{\bar{X}} = \$75.00$

$\sigma_{\bar{X}} = \$25.00$

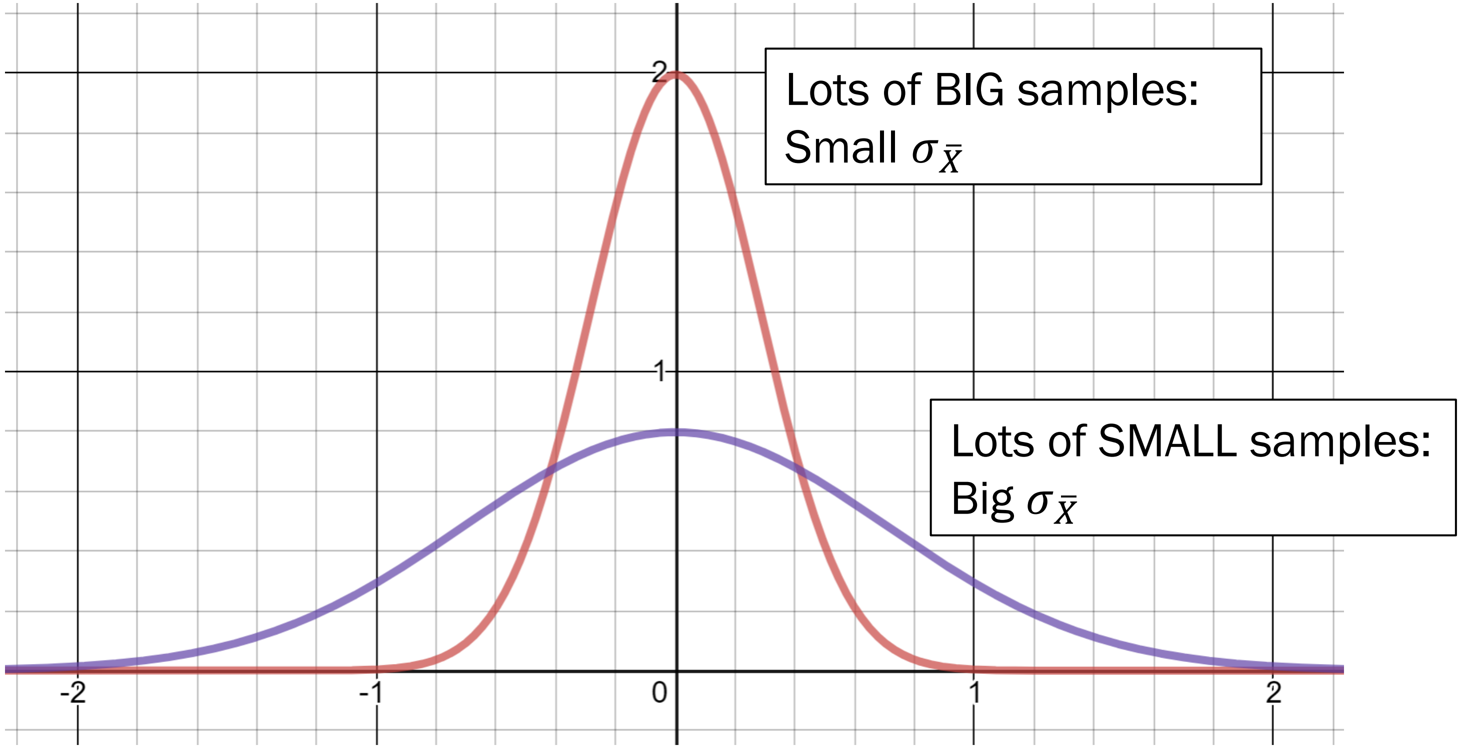


Mean of the means: $\frac{1}{9}\sum_{i=1}^{9} \bar{x}_i \approx \mu_X$

The mean of the sampled means will approximate the mean of the population

What About the Unknown Standard Deviation?

Now that the mean of a population can be found with the Central Limit Theorem, what about the standard deviation of the population?

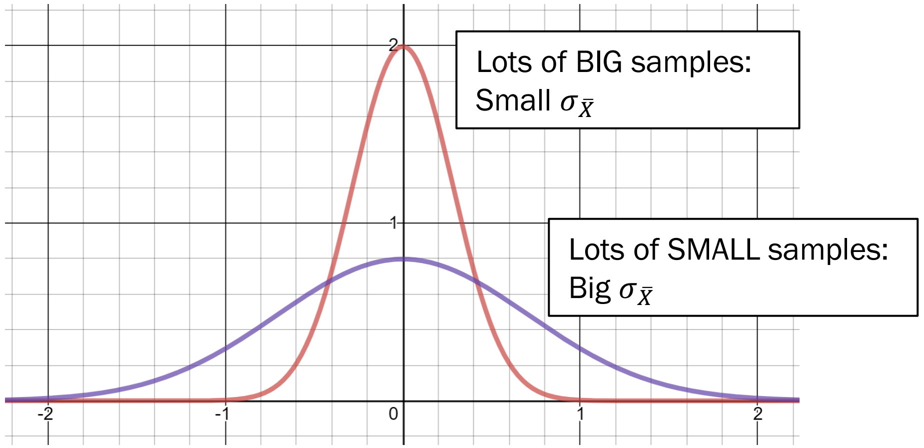


The Central Limit Theorem

$\bar{X} \sim N(\mu_{\bar{X}}, \sigma_{\bar{X}}) \approx N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right)$

$\bar{X}$ is the sampling distribution of the mean.

$\frac{\sigma_X}{\sqrt{n}}$ is the standard error of the mean.



If many samples are generated from a population, the means of the values in the sample will reflect the mean of the values in the population.

The standard deviation of the sample means is reflective (but not equal to) of the population's standard deviation.

Z-Scores for the CLT

Recall that a z-score is calculated as follows:

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} = \frac{x - \mu}{\sigma}$$

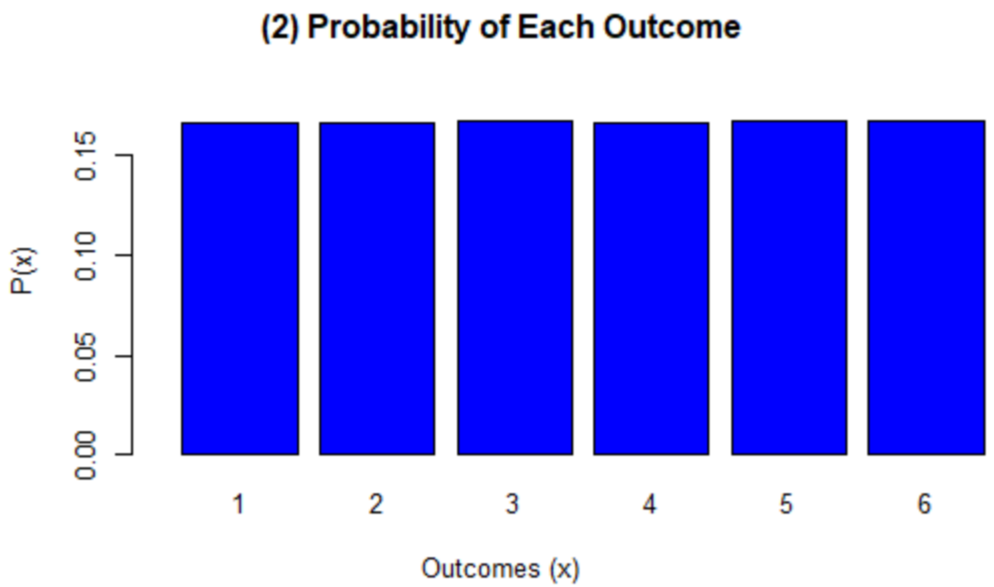
When using the central limit theorem, A z-score may be calculated for a sample mean. This works a bit differently. The standard error must be used!

$$z = \frac{\text{value} - \text{mean}}{\text{standard } error} = \frac{x - \mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)}$$

Example

Answer the following questions by considering a fair six-sided dice.

1) Complete the table



Value x	Probability of Value $P(x)$	$x \cdot P(x)$	$(x-\mu)^2 \cdot P(x)$
1	$P(1) = \frac{\{1 \text{ outcome}\}}{\{6 \text{ possibilities}\}} = \frac{1}{6} = 0.1\overline{6}$	$1 \cdot \left(\frac{1}{6}\right) = \frac{1}{6} = 0.1\overline{6}$	$(1-3.5)^2 \cdot \left(\frac{1}{6}\right) \approx 1.042$
2	$P(2) = \frac{\{1 \text{ outcome}\}}{\{6 \text{ possibilities}\}} = \frac{1}{6} = 0.1\overline{6}$	$2 \cdot \left(\frac{1}{6}\right) = \frac{2}{6} = 0.3\overline{3}$	$(2-3.5)^2 \cdot \left(\frac{1}{6}\right) = 0.375$
3	$P(3) = \frac{\{1 \text{ outcome}\}}{\{6 \text{ possibilities}\}} = \frac{1}{6} = 0.1\overline{6}$	$3 \cdot \left(\frac{1}{6}\right) = \frac{3}{6} = 0.50$	$(3-3.5)^2 \cdot \left(\frac{1}{6}\right) \approx 0.042$
4	$P(4) = \frac{\{1 \text{ outcome}\}}{\{6 \text{ possibilities}\}} = \frac{1}{6} = 0.1\overline{6}$	$4 \cdot \left(\frac{1}{6}\right) = \frac{4}{6} = 0.6\overline{6}$	$(4-3.5)^2 \cdot \left(\frac{1}{6}\right) \approx 0.042$
5	$P(5) = \frac{\{1 \text{ outcome}\}}{\{6 \text{ possibilities}\}} = \frac{1}{6} = 0.1\overline{6}$	$5 \cdot \left(\frac{1}{6}\right) = \frac{5}{6} = 0.8\overline{3}$	$(5-3.5)^2 \cdot \left(\frac{1}{6}\right) \approx 0.375$
6	$P(6) = \frac{\{1 \text{ outcome}\}}{\{6 \text{ possibilities}\}} = \frac{1}{6} = 0.1\overline{6}$	$6 \cdot \left(\frac{1}{6}\right) = \frac{6}{6} = 1$	$(6-3.5)^2 \cdot \left(\frac{1}{6}\right) \approx 1.042$
	$\sum P(x) = 1$	Expected value $\mu = \sum xP(x) = 3.5$	Variance $\sigma^2 = \sum(x-\mu)^2 P(x) = 2.918$

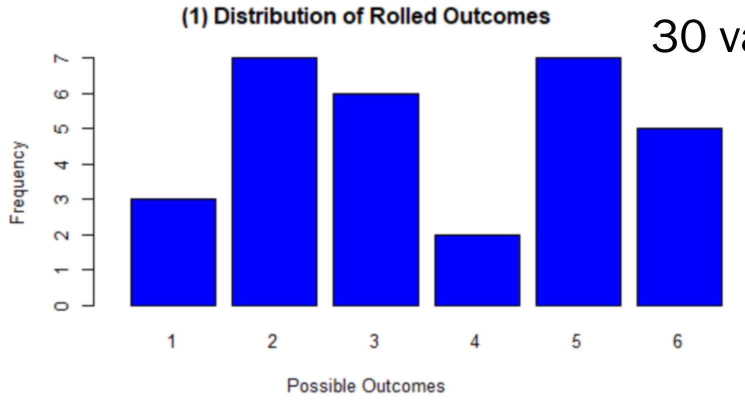
Standard Deviation

$\sigma = \sqrt{\sigma^2} = \sqrt{2.918} \approx 1.708$

Example (Cont.)

What if we collect 1 sample of 30 rolls? (n=30)

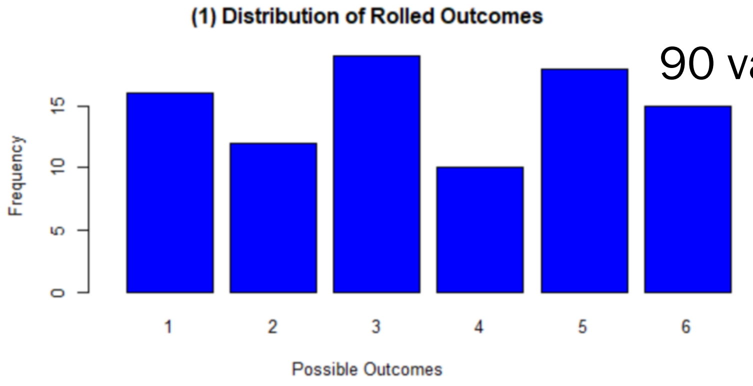
Sample	Values	Sample Mean
S1	3 4 3 6 2 3 3 1 6 2 2 1 5 5 6 2 5 1 6 4 2 2 3 5 2 3 5 5 5 6	3.6
	(30 total values)	



30 values

What if we collect 3 samples of 30 rolls? (n=30)

Sample	Values	Sample Mean
S1	1 2 2 6 6 6 1 6 5 4 3 5 1 3 2 5 1 5 3 1 2 6 3 3 5 4 5 3 4 2	3.6
S2	1 2 1 1 4 3 4 6 1 6 5 3 6 3 5 5 5 5 3 3 5 1 4 6 3 1 2 1 5 2	3.4
S3	1 6 2 2 2 3 3 3 1 2 4 1 6 3 6 4 5 6 6 5 3 4 5 5 5 4 6 3 1 3	3.67
	(90 total values)	Mean of the means = 3.56 (close to 3.5)

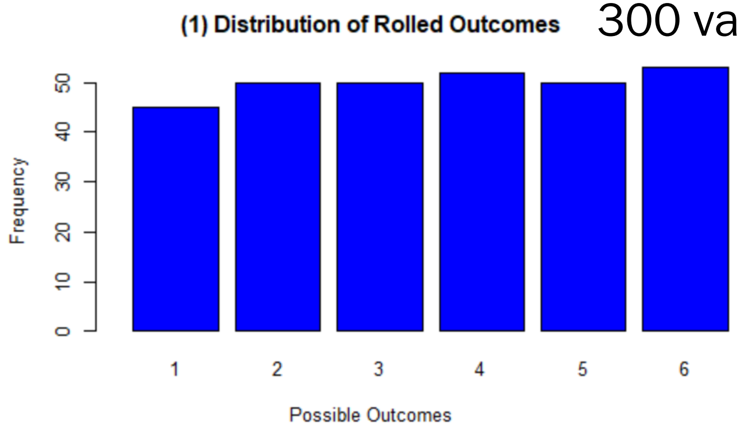


90 values

Example (Cont.)

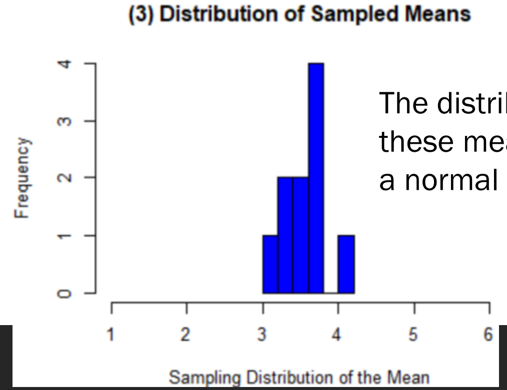
What if we collect 10 samples of 30 rolls? (n=30)

Sample	Values	Sample Mean ($\bar{X}$)
S1	2 4 1 5 4 3 5 5 1 2 5 1 2 2 2 5 5 4 5 2 3 1 3 4 2 6 2 6 2 6	3.33
S2	4 4 4 2 1 3 6 1 2 6 3 3 2 5 6 6 6 6 3 6 4 3 6 6 6 6 1 4 3 5	4.1
S3	5 2 4 1 1 3 6 3 4 4 3 3 6 4 1 3 2 1 3 6 6 6 5 5 5 2 3 4 6 3	3.67
S4	4 1 3 2 2 4 6 2 3 4 1 1 2 3 4 4 3 6 5 3 4 3 6 3 6 2 3 5 2 6	3.43
S5	5 2 4 4 4 2 5 2 4 5 1 5 2 2 4 3 6 3 5 6 4 3 3 3 4 5 3 4 6 5	3.8
S6	1 6 2 6 5 1 3 4 6 5 6 5 5 1 3 4 4 6 1 1 6 1 4 3 6 2 6 3 2 2	3.67
S7	5 1 5 4 1 3 5 2 4 3 1 1 6 1 4 3 4 6 1 5 1 6 1 2 6 3 2 6 5 4	3.37
S8	6 3 5 5 1 5 4 5 3 1 3 4 2 3 1 6 6 1 4 6 2 5 3 6 4 4 3 3 2 4	3.67
S9	1 2 1 2 2 4 5 5 3 4 2 6 5 4 5 2 1 3 2 2 5 5 1 6 1 5 2 5 2 1	3.13
S10	6 4 1 6 1 5 1 5 4 5 2 6 2 4 5 6 5 1 1 4 2 6 5 1 4 3 3 2 4 2	3.53



300 values

Mean of the means = 3.57 (close to 3.5)



The distribution of these means looks like a normal distribution.

Example (Cont.)

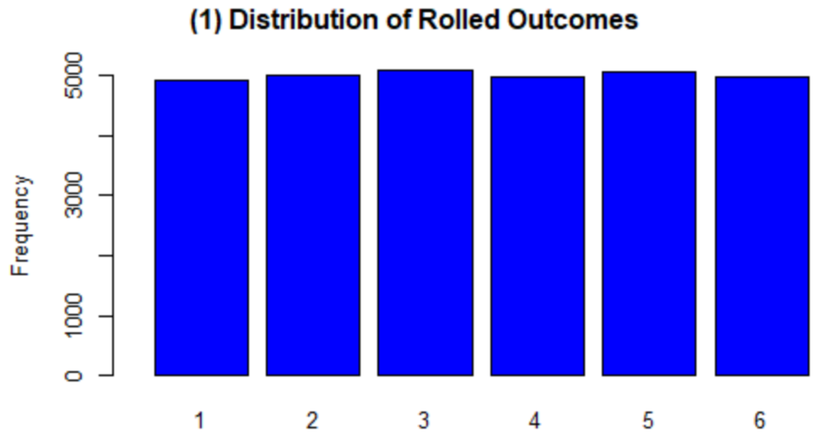
What if we collect 1000 samples of 30 rolls? (n=30)

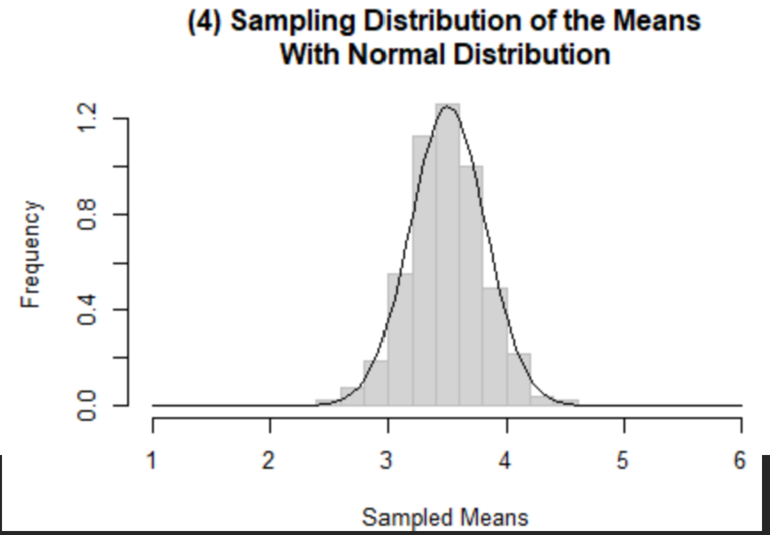
In total how many rolls were there?
$1000 \cdot 30 = 30000$

What is the mean of the means?
$\mu_{\bar{X}} = 3.5$ (Same as the expected value! $\mu_{\bar{X}} = \mu_X$)

What is the standard error of the mean?
$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{30}} = 0.32$ (represents spread of the distribution)

Can we approximate the standard deviation of the population?
Use $0.32 = \frac{\sigma_X}{\sqrt{30}}$ and solve for σ_X. Here, $\sqrt{30} \cdot 0.32 \approx 1.753$
(The ACTUAL standard deviation is 1.708)





Example (Cont.)

Given the previous scenario, what is the z-score of $\bar{x} = 3.13$?
(This is one mean value in our entire collection of 1000 means)

We have two ways to calculate our standard error:

1) From the distribution of $\bar{X}$. ($\frac{\sigma_X}{\sqrt{n}} = 0.32$)

2) Since we know σ_X for the population, we can calculate it directly ($\frac{\sigma_X}{\sqrt{n}} = \frac{1.708}{\sqrt{30}} \approx 0.312$)

$$z = \frac{\bar{x} - \mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)} = \frac{3.13 - 3.5}{0.32} = -1.15625$$

$$z = \frac{\bar{x} - \mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)} = \frac{3.13 - 3.5}{0.312} \approx -1.18652$$

Review

Central Limit Theorem: The collection of sample means from ANY distribution will form their own normal distribution. $\bar{X}$ is the sampling distribution of the mean.

$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right)$.

The for large sample sizes, the “mean of the means” is the same as the population mean $(\mu_X = \mu_{\bar{X}})$

$\frac{\sigma_X}{\sqrt{n}}$ is the standard error. It states how far away, on average, the sample mean will be from the population mean.

$z = \frac{\bar{x} - \mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)}$ (z-score for a mean)