STA 1013 : Statistics through Examples

Lecture 17: Central Limit Theorem

Hwiyoung Lee

October 11, 2019

Department of Statistics, Florida State University

Motivational Example : Rolling one dice

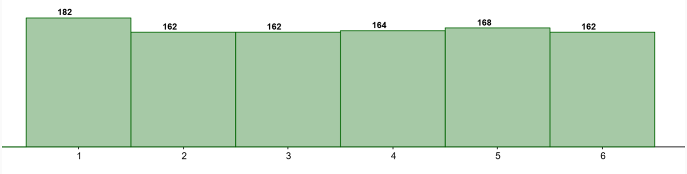
- Suppose we roll one dice 1,000 times and record the outcome of each roll
- The probabilitiy table for dice rolling

Outcome	1	2	3	4	5	6
Prob	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- https://www.geogebra.org/m/UsoH4eNl

Motivational Example : Rolling one dice

- A histogram of outcomes



- All six outcomes have roughly the same relative frequency
- Because the die is equally likely to land in each of the six possible ways
- Is the histogram a normal distribution ? **No !!**
- The histogram follows a (nearly) **uniform distribution**

Rolling two dice

- Now suppose we roll **two dice** 1,000 times and record the mean of the two numbers that appear on each roll

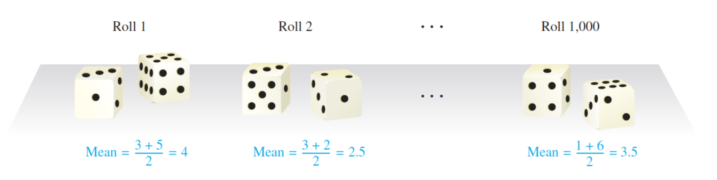


Figure 1: This diagram represents the idea of rolling two dice 1,000 times and recording the mean on each roll. The mean of the values on two dice is their sum divided by 2.

- For example,
  - if the two dice come up 3 and 5, then the mean is 4
  - if the two dice come up 1 and 2, then the mean is 1.5
- This mean can range from ______ to ______

Rolling two dice

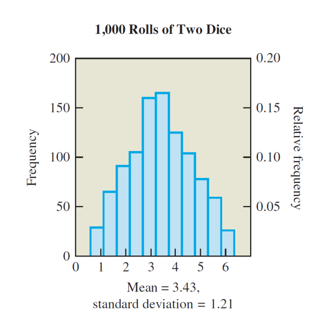
- High and low values occur less frequently because they can occur in fewer ways.
    - a roll can have a mean of 1 : if (,)
    - a roll can have a mean of 6 : if (,)
- Central values (ex : 3, 3.5, 4) are common because they can occur in several ways.
    - 3.5 can occur if the two dice land as :
      (,), (,), (,),
      (,), (,), (,).

Rolling two dice

- Possible outcomes and probabilities when we roll **two dice**

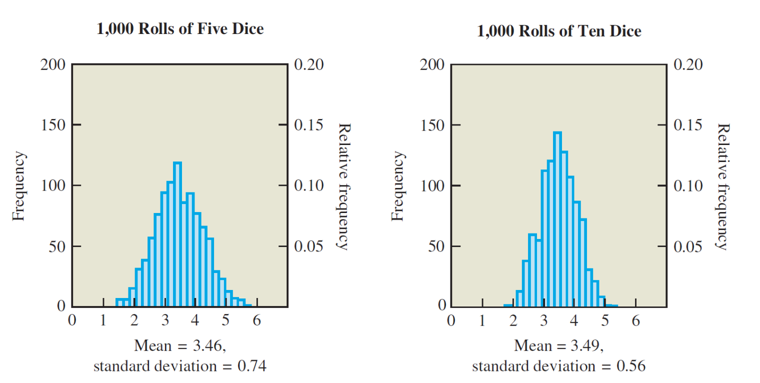
Mean							Prob
1	(1,1)						1/36
1.5	(1,2)	(2,1)					2/36
2	(1,3)	(2,2)	(3,1)				3/36
2.5	(1,4)	(2,3)	(3,2)	(4,1)			4/36
3	(1,5)	(2,4)	(3,3)	(4,2)	(5,1)		5/36
3.5	(1,6)	(2,5)	(3,4)	(4,3)	(5,2)	(6,1)	6/36
4	(2,6)	(3,5)	(4,4)	(5,3)	(6,2)		5/36
4.5	(3,6)	(4,5)	(5,4)	(6,3)			4/36
5	(4,6)	(5,5)	(6,4)				3/36
5.5	(5,6)	(6,5)					2/36
6	(6,6)						1/36

A result of rolling two dice 1,000 time



- Is the shape of distribution **still Uniform ?** **No !!**
- Looks similar to a normal distribution
- **Central tendancy** : The most common values in this distribution are the central values 3.0, 3.5, and 4.0.

What happens if we increase the number of dice we roll?



- Suppose we roll five, and ten dice 1,000 times.
- If we further increase the number of dice on each of 1,000 rolls, we find the histogram would be even narrower.

Results of dice rolling

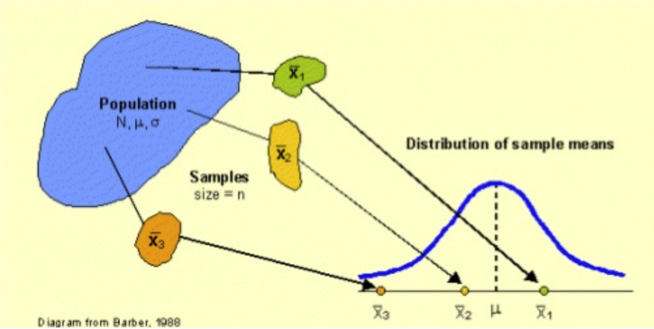
TABLE 2 Summary of Dice Rolling Experiments

Number of dice rolled each time	Mean of the distribution of means	Standard deviation of the distribution of means
1	3.41	1.73
2	3.43	1.21
5	3.46	0.74
10	3.49	0.56

- Mean converges to 3.5
- Standard deviation goes smaller

Central Limit Theorem

Sampling distribution



A sampling distribution is a probability distribution of a statistic obtained through a large number of samples drawn from a specific population.

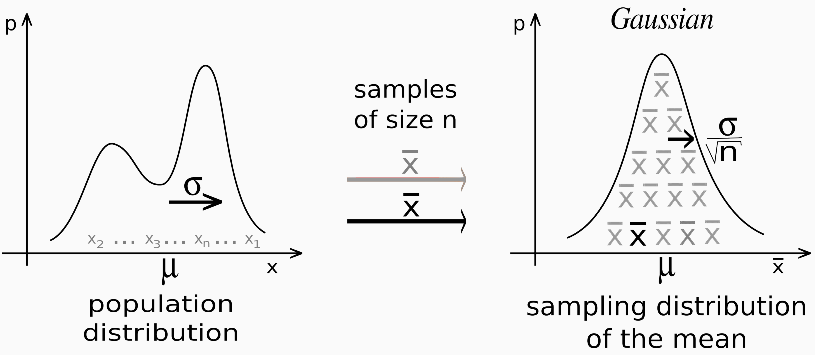
Central Limit Theorem

Central Limit Theorem (CLT)

Suppose the population distribution (**Not necessarily a normal distribution**) has population mean μ, and population standard deviation σ , then the distribution of sample means (Sampling distribution of $\bar{X}$) converges to $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

- The distribution of sample means will be approximately a normal distribution for large sample sizes ($n \geqslant 30$).

Central Limit Theorem

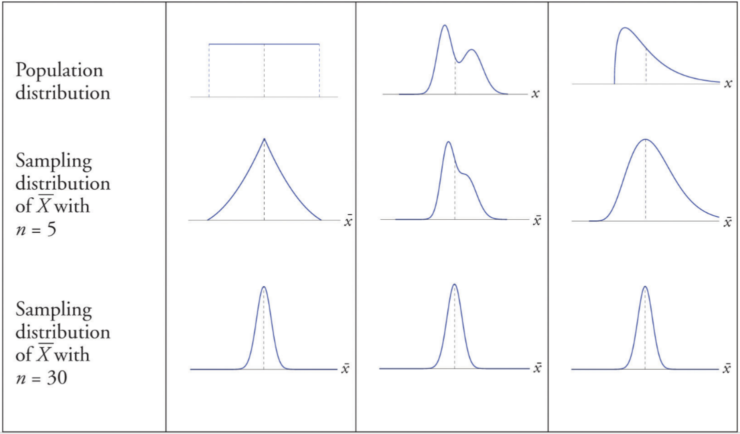


Central Limit Theorem

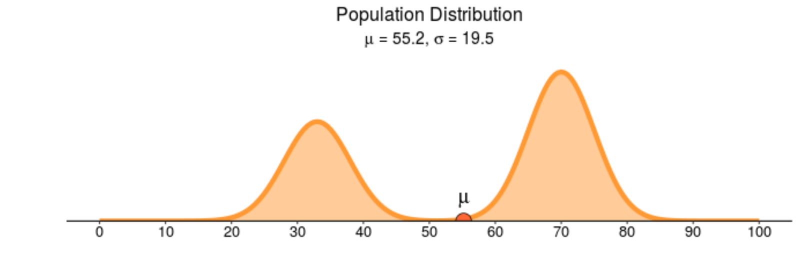
- The central limit theorem states that if you have a population with mean μ and standard deviation σ and take **sufficiently large random samples** $(n \geqslant 30)$ from the population, then the distribution of the sample means will be **approximately normally distributed**.

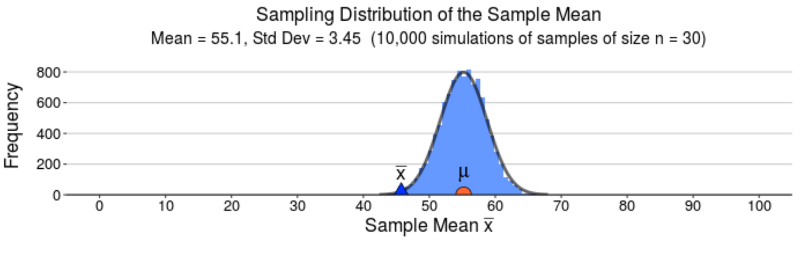
- This will hold true **regardless of** the shape of population distributions (Bell, skew, bimodalm uniform)

Central Limit Theorem

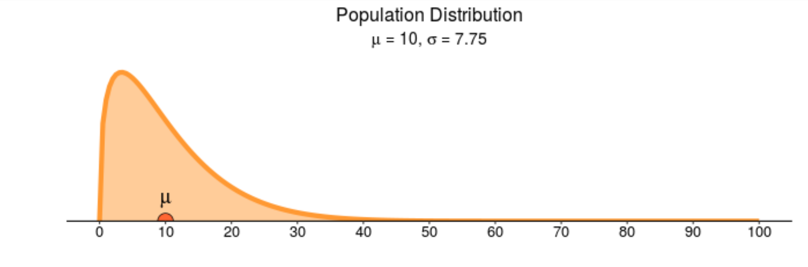


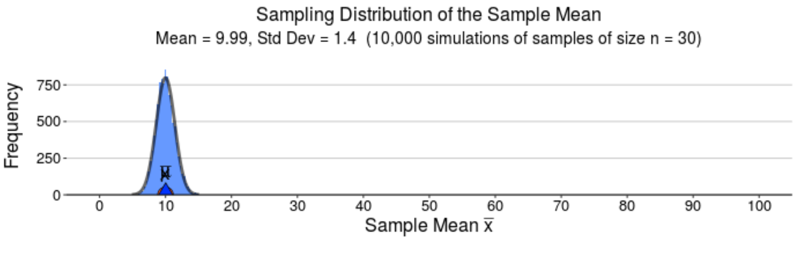
CLT Simulation (Bimodal)



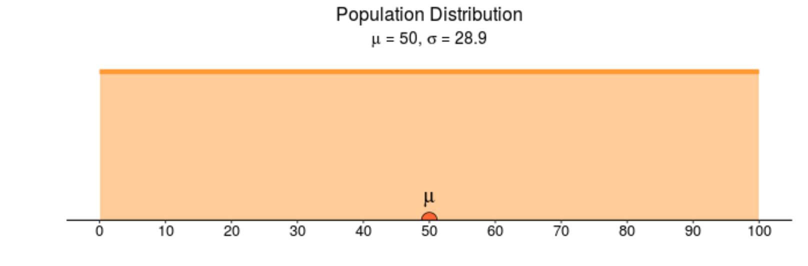


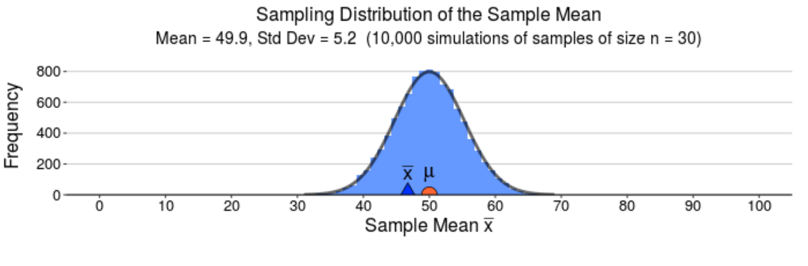
CLT Simulation (Skewed)



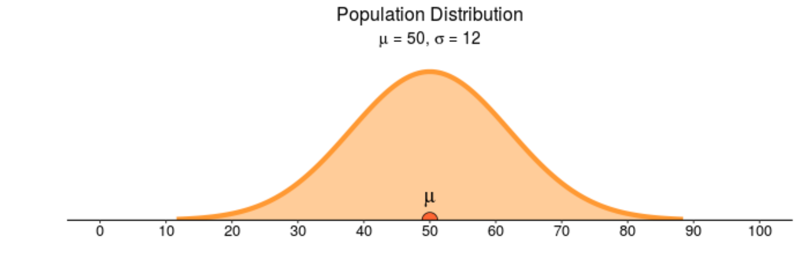


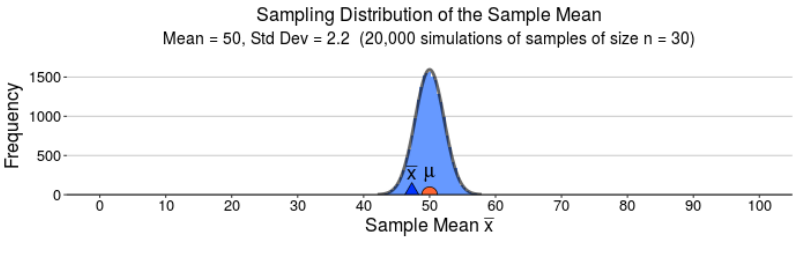
CLT Simulation (Uniform)





CLT Simulation (Bell-shape)





CLT : Simulation

https://istats.shinyapps.io/sampdist_cont/

Theoretically, the discrete uniform distribution Unif(a, b) has

- $\mu = \frac{a+b}{2}$
- $\sigma = \sqrt{\frac{(b-a+1)^2-1}{12}}$

Since in our case, $a = 1, b = 6$

- $\mu = 3.5$
- $\sigma = 1.71$

CLT : Dice rolling revisit

TABLE 2 Summary of Dice Rolling Experiments

Number of dice rolled each time	Mean of the distribution of means	Standard deviation of the distribution of means
1	3.41	1.73
2	3.43	1.21
5	3.46	0.74
10	3.49	0.56

From the CLT

- The mean of the sampling dist converges to 3.5
- The standard devistion of the sampling dist converges to $\frac{1.71}{\sqrt{n}}$

Examples

Example-SAT scores

Based on data from the College Board, assume that SAT scores are normally distributed with a mean of 1518 and a standard deviation of 325. Assume that many samples of size n are taken from a large population of students and the mean SAT score is computed for each sample.

a. If the sample size is $n = 100$, find the mean and standard deviation of the distribution of sample means.

b. If the sample size is $n = 2{,}500$, find the mean and standard deviation of the distribution of sample means.

Example-IQ scores

IQ scores are normally distributed with a mean of 100 and a standard deviation of 16. Assume that many samples of size n are taken from a large population of people and the mean IQ score is computed for each sample.

a. If the sample size is $n = 64$, find the mean and standard deviation of the distribution of sample means.

b. If the sample size is $n = 100$, find the mean and standard deviation of the distribution of sample means.

LSQA

Section 15 : MWF (10:10 $\sim$ 11:00)

- Test Date : Oct 21 (Mon) HCB 217

Section 05 : MWF (12:20 $\sim$ 1:10)

- Test Date : Oct 18 (Fri) OSB 108