Sampling Distributions

MATH 2300

Chapter 7

Sample Means

- Suppose you sample from a population 10 times.
- You record the following sample means:

10.1	9.5	9.6	10.2	9.5
9.2	10.4	9.3	8.5	11.0

- Why aren't the values all the same?

The Population

- The entire group of individuals about which we want to get information.
- Often, it is too time-consuming or expensive to obtain data for the entire population.
- How do we get around these constraints?

The Sample

- Part of the population from which we actually collect information
- We use this information to draw conclusions about the population.
  - We call this process **inference**.

How do we describe samples?

- First, we collect raw data.
  - The actual measurements or observations.
- We consolidate or summarize the raw data to describe the various characteristics of the sample.
  - These numerical summaries are called **statistics**.

How do we describe populations?

- We use our statistics to make *estimates* about the characteristics of the population.
- The characteristics of a population are called **parameters**.
- The value that a parameter takes is almost always *unknown*.

Summary

Parameters	Statistics
Describe Populations	Describe Samples
Fixed Values for a Given Population	Changes from Sample to Sample
Value Unknown in Practice	Value is Calculated for a Given Sample

A Population Distribution

- For a given variable, this is the distribution of values the variable can take among all the **individuals** in the ***population***.
- IMPORTANT:
  - Describes the **individuals** in the population.

A Sampling Distribution

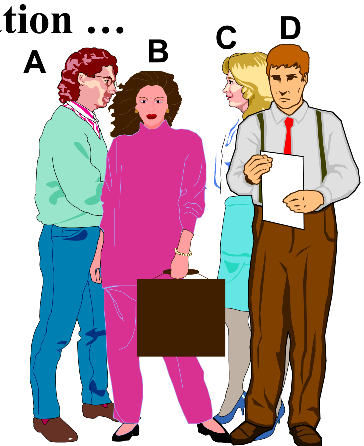
- The distribution of values taken by a ***statistic*** in all possible samples of the **same size** from the **same population**.
- IMPORTANT:
  - Describes a statistic calculated from **samples** from a given population.

Random Sampling Error

- The deviation between the statistic and the parameter.
- Caused by chance in selecting a random sample.
- Margin of error includes only random sampling error.
  - NOT errors associated with choosing bad samples.

Developing a Sampling Distribution

- **Assume there is a population …**
- Population size n = 4
- Measurement of interest is age of individuals
- Values: 18, 20, 22, 24 (years)



Consider All Possible Samples of Size n = 2

1st Obs	2nd Observation 18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 possible samples (sampling with replacement)

16 Sample Means

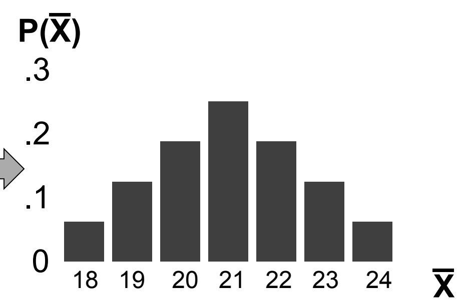
1st Obs	2nd Observation 18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Displaying the Sampling Distribution

16 Sample Means

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

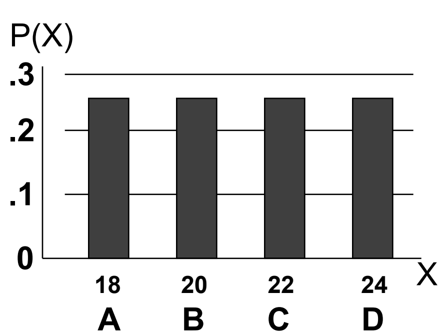
Sample Means Distribution

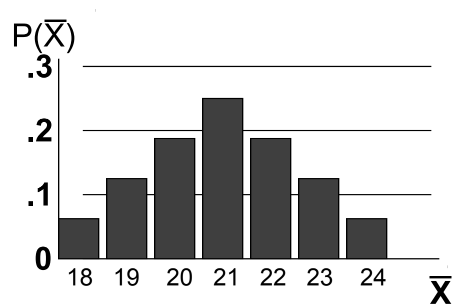


Population Distribution vs. Sampling Distribution

Population

Sample Means Distribution
n = 2





The Law of Large Numbers

- A sample is drawn at random from any population with mean μ.
- As the number of observations goes up, the sample mean $\bar{x}$ tends to get closer to the population mean μ.

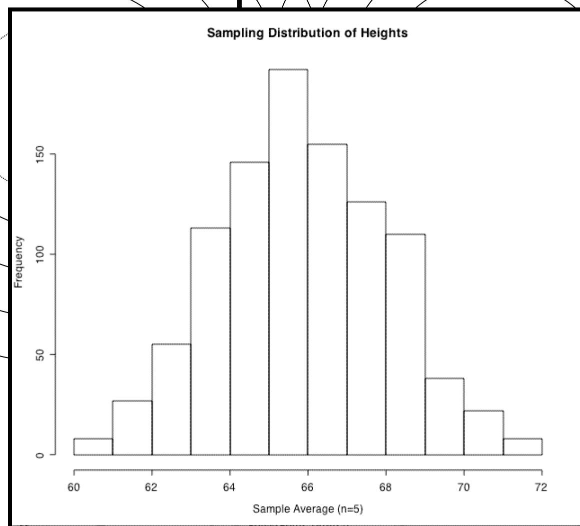
Example: Class of 20 Students

- Suppose there are 20 people in a class and you are interested in the average height of the class.
- The heights (in inches):

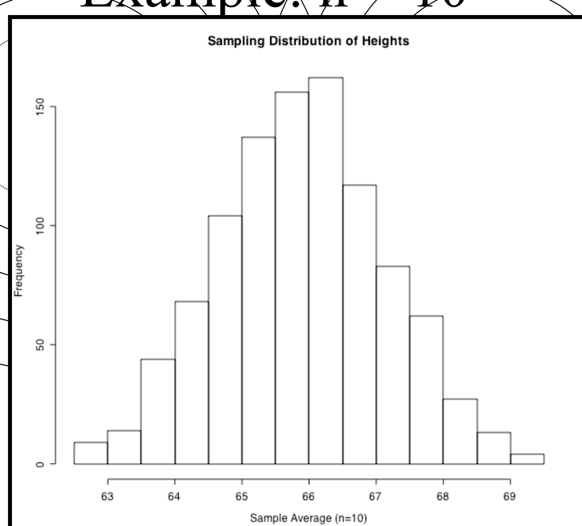
  72 64 75 63 62 61 68 76 59 73

  67 66 65 64 60 65 70 56 71 62
- The average height is 65.95 in.

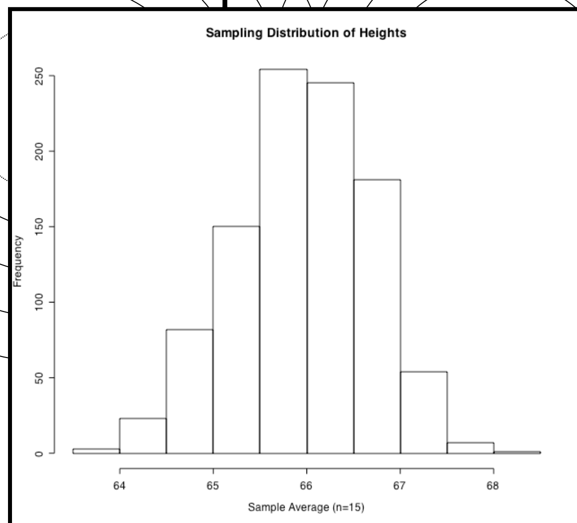
Example: n = 5



Example: n = 10



Example: n = 15



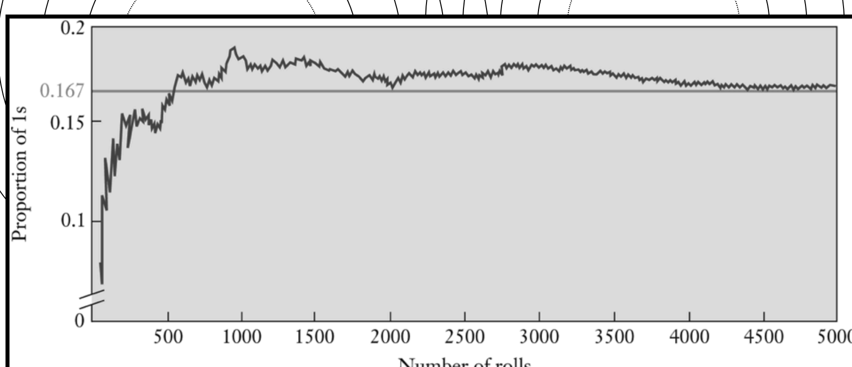
What the Law of Large Numbers Tells Us

- It tells us that our estimate of the population mean will get better and better as we take bigger and bigger samples.
- This means the variability (or spread) of the sample mean decreases as *n* increases.
- However, there are some cautions regarding misusing it.

Applying the LLN to Probabilities

- A process is repeated through many trials.
- Proportion of times event *A* occurs will be close to the probability P(*A*).
- The larger the number of trials the closer the proportion should be to P(*A*).

Example: Rolling a Die



Caution: Football Example

- A quarterback has a career pass completion ratio of 0.60.
- In today's game, he has so far thrown 10 passes and only completed 2 of them.
- The commentator says, "He's bound to complete the next few by the law of averages!"
  - Faulty logic!

Sampling Distribution of a Mean

- A SRS of *n* observations is taken from a large population with a mean μ and a standard deviation σ.
- The sampling distribution of the sample mean has the same mean: $(\mu_{\bar{x}} = \mu)$
- The standard deviation is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Example: Sodium Measurements

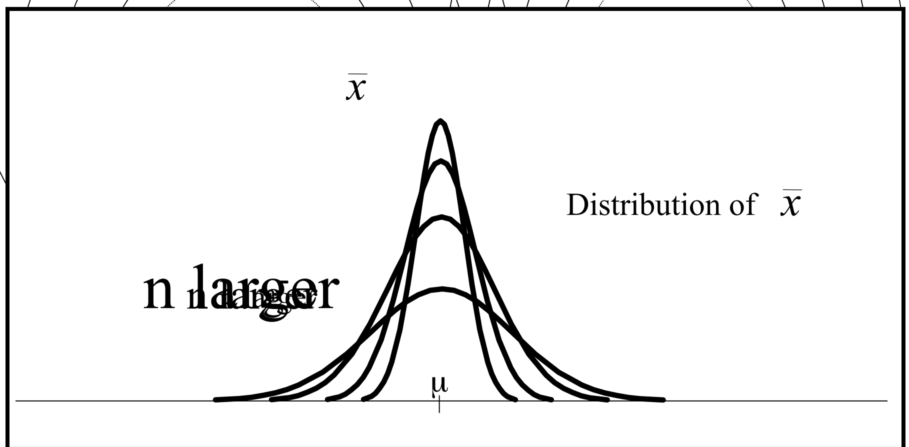
- Standard deviation of sodium content 10 mg.
- Measure 3 times and the mean of these 3 measurements is recorded.
- What is the standard deviation of the mean result?
- How many measurements are needed to get a standard deviation of the mean equal to 5?

Sampling Distribution of a Mean

- If the distribution of the population is N(μ, σ)
- Then the sample mean of n independent observations has the distribution:

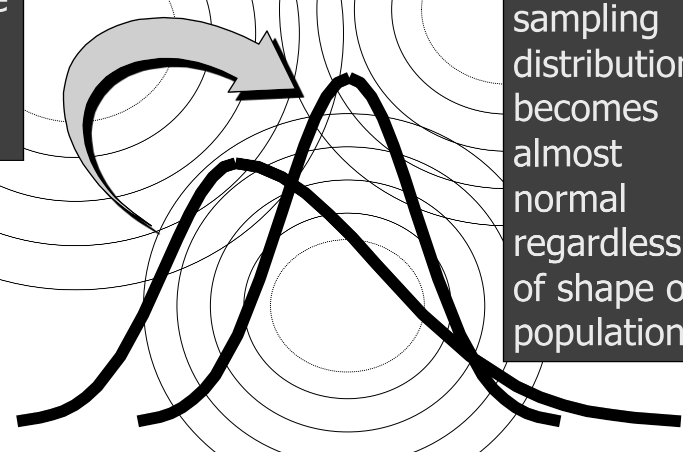
$$N(\mu, \frac{\sigma}{\sqrt{n}})$$

Graphical Depiction



The Central Limit Theorem

As sample size gets large enough…

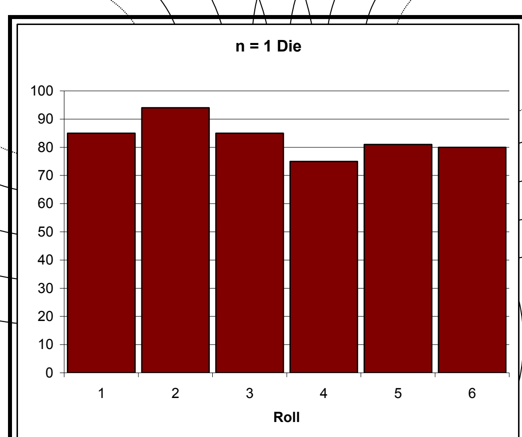


the sampling distribution becomes almost normal regardless of shape of population

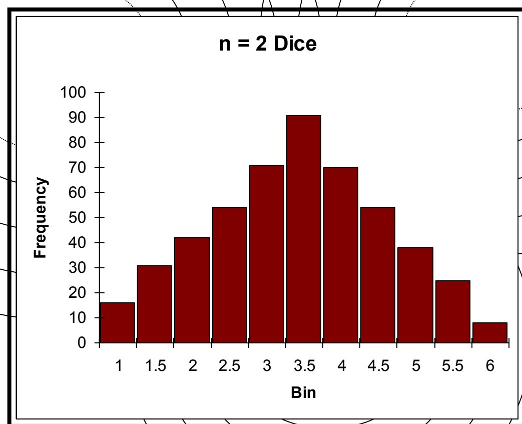
The Central Limit Theorem

- A SRS of n observations taken from a population with mean μ and standard deviation σ
- Regardless of the population's distribution the distribution of the sample mean will be approximately normal (provided the sample size is sufficiently large).
- The larger the sample size, the better the approximation to the normal distribution.

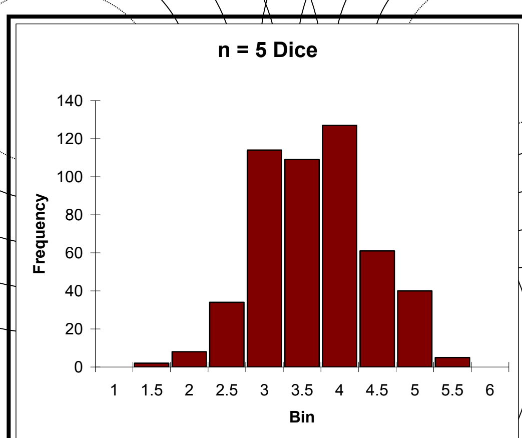
Simulating 500 Rolls of n Dice



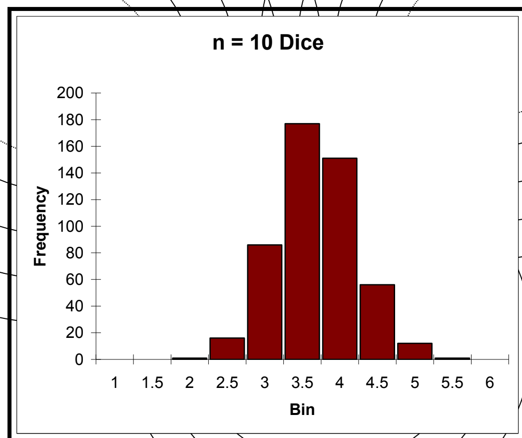
Simulating 500 Rolls of n Dice



Simulating 500 Rolls of n Dice



Simulating 500 Rolls of n Dice



Z-Score for Distribution of the Mean

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

where:

$\bar{x}$ = Sample mean
μ = Population mean
σ = Population standard deviation
n = Sample size

Example Calculation

What is the probability that a sample of 100 automobile insurance claim files will yield an average claim of $4,527.77 or less if the average claim for the population is $4,560 with standard deviation of $600?

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}} = \frac{(4{,}527.77 - 4{,}560)}{\frac{600}{\sqrt{100}}} = \frac{-32.23}{60} = -0.537$$

P(Z < -0.54) = 0.2946

Example: ACT Exam

- Scores on the ACT exam are distributed N(18.6, 5.9)
- What is the probability that a single student scores 21 or higher?
- What is the probability that the mean score of 50 students is 21 or higher?

Summary

- Means of random samples are **less variable** than individual observations.
- Means of random samples are **more Normal** than individual observations.