8.1 Distribution of the Sample Mean

Statistics such as $\bar{x}$ are random variables since their value varies from sample to sample. As such, they have probability distributions associated with them. In this chapter we focus on the shape, center and spread of statistics such as $\bar{x}$.

The **sampling distribution** of a statistic is a probability distribution for all possible values of the statistic computed from a sample of size *n*.

The **sampling distribution of the sample mean** $\bar{x}$ is the probability distribution of all possible values of the random variable $\bar{x}$ computed from a sample of size *n* from a population with mean μ and standard deviation σ.

Steps for Illustrating Sampling Distributions

Step 1: Obtain a simple random sample of size *n*.
Step 2: Compute the sample mean.
Step 3: Assuming we are sampling from a finite population, repeat Steps 1 and 2 until all simple random samples of size *n* have been obtained.

Distribution of the sample mean: normal population

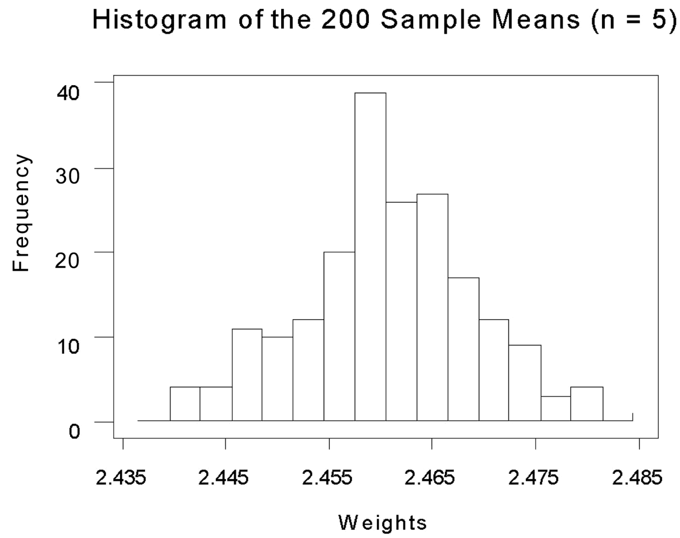
Example: *Sampling Distribution of the Sample Mean-Normal Population*

The weights of pennies minted after 1982 are approximately normally distributed with mean 2.46 grams and standard deviation 0.02 grams.
Approximate the sampling distribution of the sample mean by obtaining 200 simple random samples of size $n = 5$ from this population.

The data here represent the sample means for the 200 simple random samples of size $n = 5$.
For example, the first sample of $n = 5$ had the following data:
2.493 2.466 2.473 2.492 2.471
Note: = 2.479 for this sample
Sample Means for Samples of Size *n* = 5

2.479	2.463	2.459	2.455	2.468	2.467	2.464	2.480	2.467	2.468
2.465	2.463	2.463	2.459	2.465	2.461	2.459	2.467	2.475	2.455
2.466	2.475	2.465	2.470	2.449	2.474	2.458	2.471	2.461	2.463
2.451	2.446	2.470	2.465	2.461	2.460	2.460	2.454	2.466	2.466
2.458	2.454	2.467	2.471	2.453	2.468	2.467	2.476	2.456	2.473
2.446	2.459	2.460	2.472	2.467	2.471	2.466	2.459	2.454	2.459
2.471	2.468	2.460	2.461	2.468	2.463	2.467	2.446	2.469	2.458
2.472	2.457	2.444	2.480	2.443	2.459	2.452	2.458	2.460	2.460
2.466	2.465	2.442	2.461	2.466	2.446	2.454	2.466	2.463	2.455
2.459	2.451	2.451	2.458	2.461	2.459	2.457	2.464	2.459	2.459
2.445	2.459	2.474	2.458	2.447	2.451	2.478	2.465	2.470	2.438
2.469	2.456	2.445	2.450	2.473	2.449	2.450	2.463	2.447	2.453
2.457	2.457	2.446	2.447	2.456	2.465	2.453	2.458	2.455	2.460
2.460	2.460	2.465	2.466	2.475	2.460	2.450	2.472	2.455	2.462
2.454	2.467	2.463	2.474	2.460	2.465	2.465	2.460	2.466	2.483
2.456	2.453	2.453	2.446	2.465	2.461	2.466	2.463	2.457	2.442
2.457	2.459	2.461	2.442	2.471	2.465	2.459	2.448	2.466	2.460
2.479	2.461	2.462	2.459	2.453	2.473	2.463	2.461	2.442	2.478
2.446	2.467	2.455	2.449	2.462	2.461	2.464	2.457	2.460	2.469
2.461	2.462	2.460	2.456	2.458	2.471	2.459	2.455	2.465	2.456

The mean of the 200 sample means is 2.46, the same as the mean of the population. The standard deviation of the sample means is 0.0086, which is smaller than the standard deviation of the population.

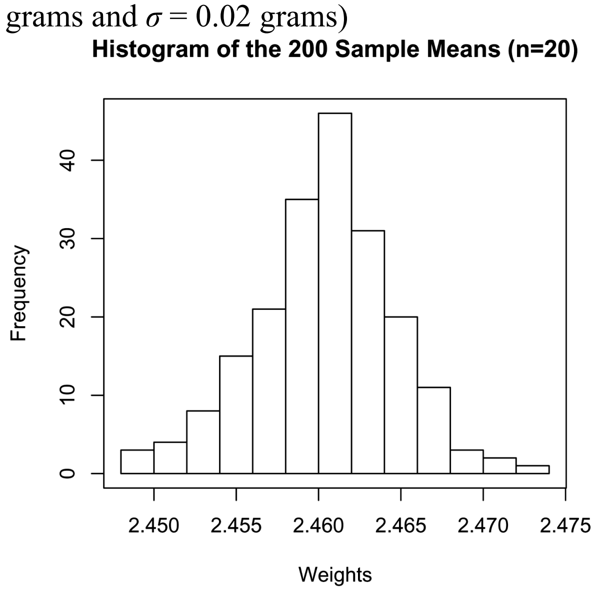


What role does n, the sample size, play in the standard deviation of the distribution of the sample mean?

As the size of the sample increases, the standard deviation of the distribution of the sample mean decreases.

<u>Example 2:</u> The Impact of Sample Size on Sampling Variability

Approximate the sampling distribution of the sample mean by obtaining 200 simple random samples of size $n = 20$ from the population of weights of pennies minted after 1982 ($\mu = 2.46$ grams and $\sigma = 0.02$ grams)



The mean of the 200 sample means for n = 20 is still 2.46, but the standard deviation is now 0.0045 (0.0086 for n = 5). As expected, there is less variability in the distribution of the sample mean with n =20 than with n = 5.

The Mean and Standard Deviation of the Sampling Distribution of $\bar{x}$

Suppose that a simple random sample of size n is drawn from a large population with mean μ and standard deviation σ. The sampling distribution of $\bar{x}$ will have

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

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

The standard deviation of the sampling distribution of $\bar{x}$ is called the **standard error of the mean** and $\sigma_{\bar{x}}$

The Shape of the Sampling Distribution of $\overline{x}$ If X is Normal

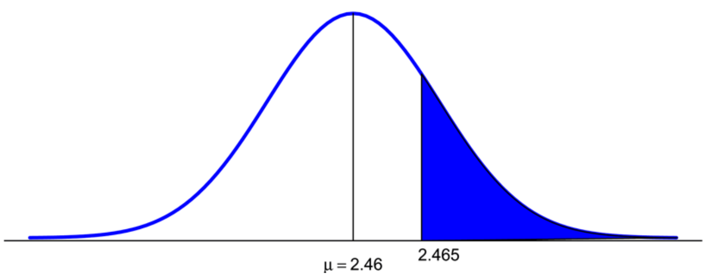
If a random variable X is normally distributed, the distribution of the sample mean $\overline{x}$ is normally distributed.

Example
The weights of pennies minted after 1982 are approximately normally distributed with mean 2.46 grams and standard deviation 0.02 grams.
What is the probability that in a simple random sample of 10 pennies minted after 1982, we obtain a sample mean of at least 2.465 grams?

Find $\mu_{\bar{x}}$:

Find $\sigma_{\bar{x}}$:

Use normalcdf to find the probability.

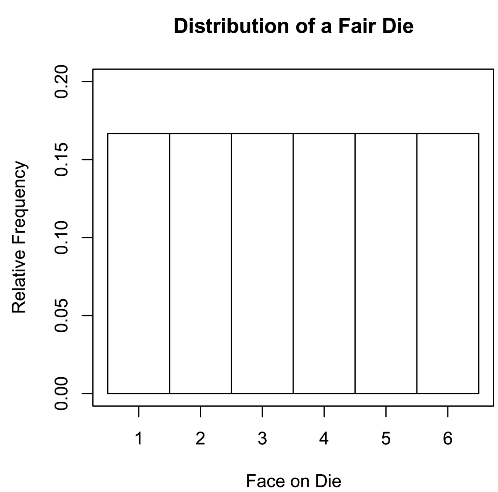


Describe the Distribution of the Sample Mean: Nonnormal Population

Example: *Sampling from a Population that is Not Normal*
The following table and histogram give the probability distribution for rolling a fair die:

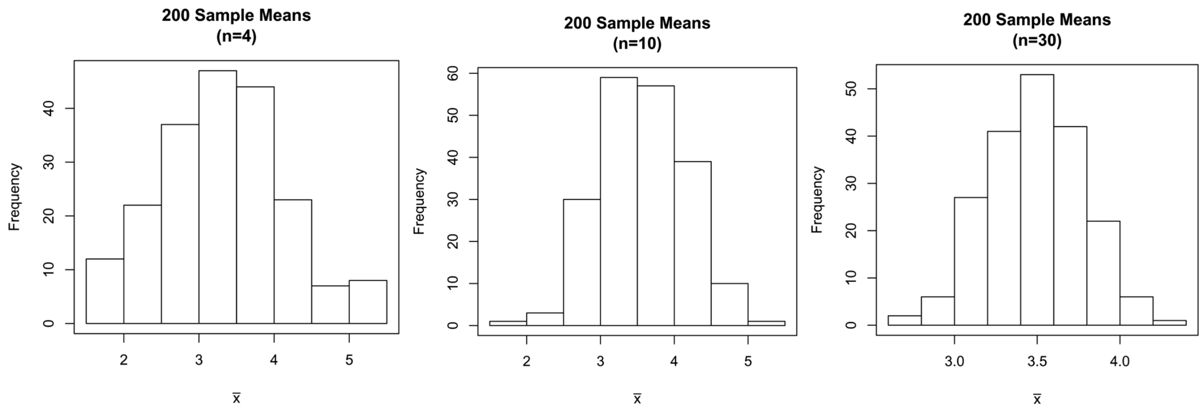
Face on Die	Relative Frequency
1	0.1667
2	0.1667
3	0.1667
4	0.1667
5	0.1667
6	0.1667

μ = 3.5, σ = 1.708
Note that the population distribution is NOT normal



Estimate the sampling distribution of $\bar{x}$ by obtaining 200 simple random samples of size *n* = 4 and calculating the sample mean for each of the 200 samples. Repeat for *n* = 10 and 30.

Histograms of the sampling distribution of the sample mean for each sample size are given on the next slide.



Key Points:

- The mean of the sampling distribution is equal to the mean of the parent population and the standard deviation of the sampling distribution of the sample mean is $\frac{\sigma}{\sqrt{n}}$ regardless of the sample size.
- **The Central Limit Theorem:** the shape of the distribution of the sample mean becomes approximately normal as the sample size *n* increases, regardless of the shape of the population.

Example: Using the Central Limit Theorem

Suppose that the mean time for an oil change at a "10-minute oil change joint" is 11.4 minutes with a standard deviation of 3.2 minutes.

(a) If a random sample of *n* = 35 oil changes is selected, describe the sampling distribution of the sample mean.

(b) If a random sample of *n* = 35 oil changes is selected, what is the probability the mean oil change time is less than 11 minutes?