

Packet 8: Sampling Distribution of the Sample Mean

Textbook pages: 24 – 30; 701 – 704

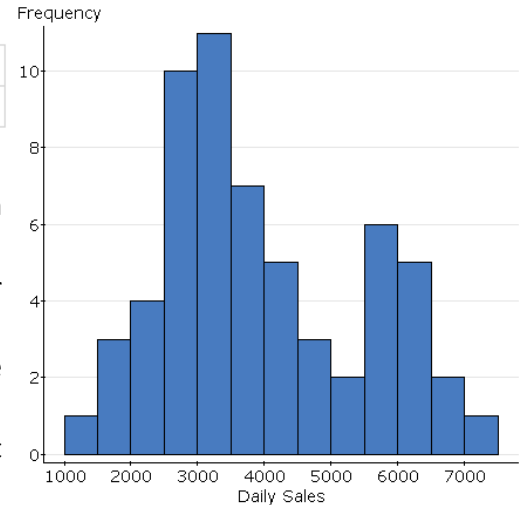
After completing this material, you should be able to:

- explain what the symbols $\mu_{\bar{y}}$ and $\sigma_{\bar{y}}$ represent.
- describe the sampling distribution of the sample mean by discussing its shape, mean, and standard deviation.
- find probabilities associated with various sample means based on the sampling distribution.
- make inferences from the probability and explain the reasoning.

A local pub is interested in learning more about their daily sales. In order to do this, daily sales (measured in dollars) from the past two months are collected. A histogram and summary statistics are given below. Use these to fill in the description of the distribution below – include units with the values where appropriate.

Summary statistics:

Column	n	Mean	Std. dev.	Median	Min	Max	Q1	Q3
Daily Sales	60	3917.4433	1512.3071	3611.45	1210.5	7068.2	2853.25	5318.05



The distribution of _____ has a shape that is _____. The average sales for these _____ days is _____, and 50% of the time daily sales are _____ or less. The smallest sales amount was _____, while the largest was _____.

The middle 50% of daily sales fell between _____ and _____. According to _____ Rule, we expect daily sales to fall between _____ and _____ (three standard deviations from the mean) _____% of the time.

Let's assume this is our population of sales values – what happens if we start to take samples of size 14 from this population of values?

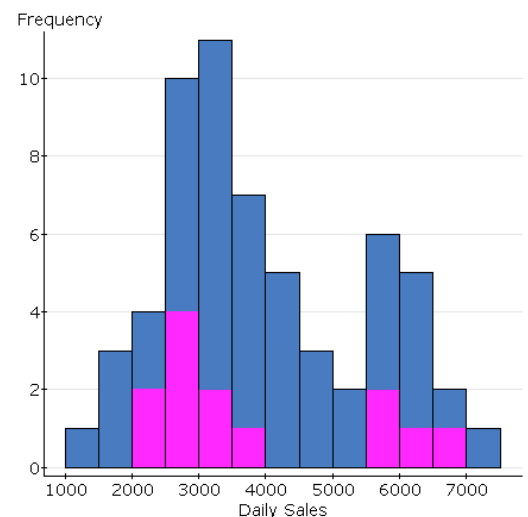
Below is the sample which was taken – the values selected in the sample have been highlighted in pink on the original histogram.

Sample Sales:

2062	2142	2503	2566	2814	2827	3189
3446	3748	4908	5788	5816	6105	6799

What is the mean for this sample of 14 days?

If we took another sample of 14 days would we expect the same mean?

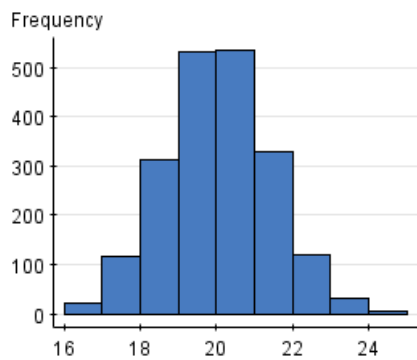


Recall: What is a *sampling distribution* and why is it important?

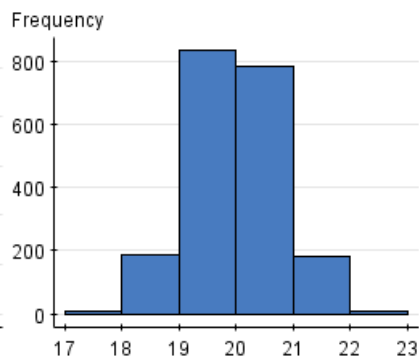
We need to understand how the samples vary. To do this, we need to **describe** the sampling distribution of the sample mean. To do this, the following three characteristics must be addressed:

-
-
-

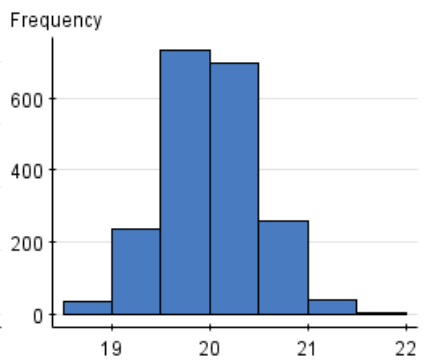
Example: Each of the following histograms represents a sampling distribution of the sample mean for various sample sizes. Samples of size 5, 15, and 40 were taken from some population. Match each histogram to the appropriate sample size.



sample size: _____



sample size: _____



sample size: _____

Based on the sampling distributions shown, what is the (approximate) population mean? **Explain.**

The population from which these samples were taken must have had what shape? **Explain.**

Example: It was reported that the average age for a Major League Baseball player is 27.2 years. With the end of the 2016 season, a sports enthusiast wants to know if the average age for players on the Cincinnati Reds is lower. To test this claim, a sample of 16 players on the Reds' roster will be taken. Assume the ages of all MLB players are normally distributed with a standard deviation of 3.12 years.

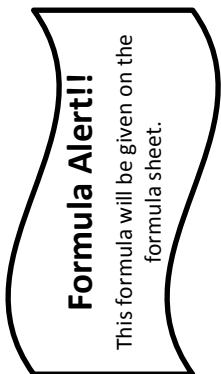
What variable will be recorded in the sample selected? Is this quantitative or categorical?

Assign the appropriate notation to the values reported in the example.

What conjecture has been made by the sports enthusiast?

Describe the sampling distribution of the sample mean age which will be obtained from a sample of 16 players.

The fact that the sampling distribution of the sample mean is *normally distributed* is important – we know how to find probabilities from the normal distribution. To do this, though, we will need to modify our formula for the z-score to reflect that we are now dealing with sample means:



Let's go back to the example and see how this formula is used ...

Back to the example: The sample of 16 players was taken and the average age was found to be 26.28 years. What is the probability of observing this sample mean or some smaller value?

Based on the probability found above, what conclusion can be drawn?

Example: According to Nielsen, the 2010 NCAA Men's Basketball Tournament (aka March Madness) averaged 10.19 million viewers (including online viewers). The NCAA has expanded its online coverage in recent years, and it is thought the mean number of viewers will have increased. Assume the population is right-skewed with a standard deviation of 6.15 million viewers.

What conjecture has been made by the NCAA?

A sample of 40 tournament games is taken (from 2011 – 2014), completely describe the sampling distribution of the sample mean number of viewers?

The sample of 40 games has a mean of 11.76 million viewers. What is the probability of observing an average of 11.76 or larger?

Based on the probability computed above, what can be concluded about the mean number of viewers for March Madness?

If the sample size were increased to 75 games, between what two values would we expect **68%** of the sample means to fall?

Example: An ambulance service reports that the mean time required for ambulances to reach their destinations was 10 minutes with a standard deviation of 4.72 minutes. Given traffic congestion it has been hypothesized that the mean time has increased.

If a sample of 35 ambulance runs is selected, how would the sampling distribution of the sample mean time to destination be described?

Suppose that a sample of 35 ambulance runs is taken, and they report a sample mean time to the destinations of 10.92 minutes. What is the probability of observing a sample mean time of 10.92 minutes or more?

Based on the probability above, what (if anything) can be inferred about the mean time required for ambulances to reach their destination?