Math 243 **Sections 7.1 and 7.3** **Sampling Distribution of a Mean**

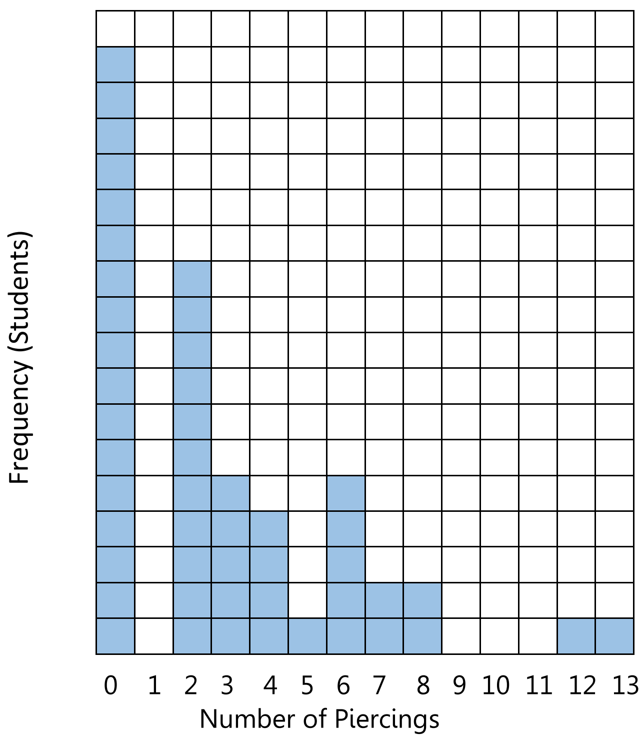
Piercings data, both classes

Sampling Models - Random samples have their own distribution models and we need to understand what they look like before we can make inferences.

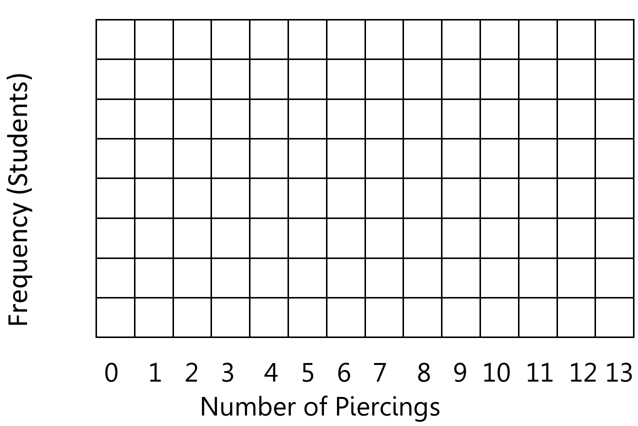
Remember the number of piercings data from the first day of class? Let's take a random sample of 2 students from our class and take the average. If we take many samples of 2 students, what will the histogram look like?

Population mean, μ=

Original Class Histogram **Random Sample** **Sample Mean, $\bar{x}$**

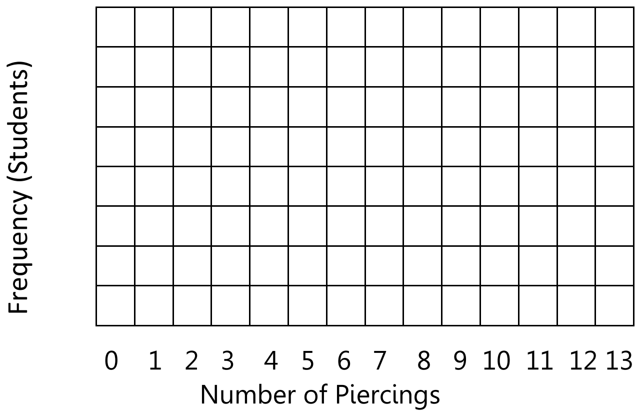


Histogram of Sample Means with n=2



A sample size of 2 is pretty small. Let's take a bigger sample of 5 students. What do you think this histogram look like?

Histogram of Sample Means with n=5



Random Sample	Sample Mean, $\bar{x}$

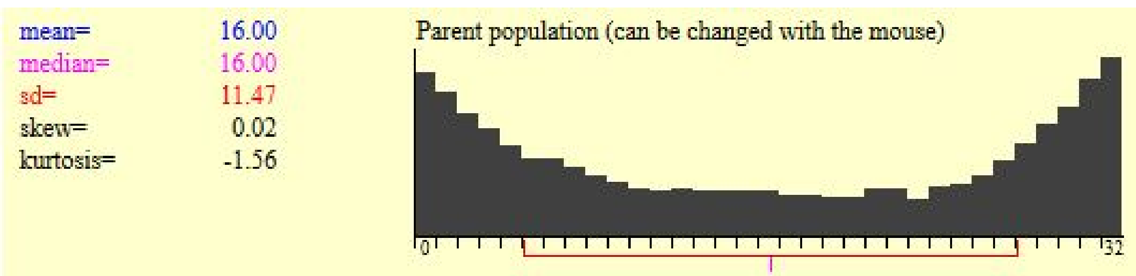
Math 243 Lecture Notes

Sections 7.1 and 7.3 Sampling Distribution of the MEAN

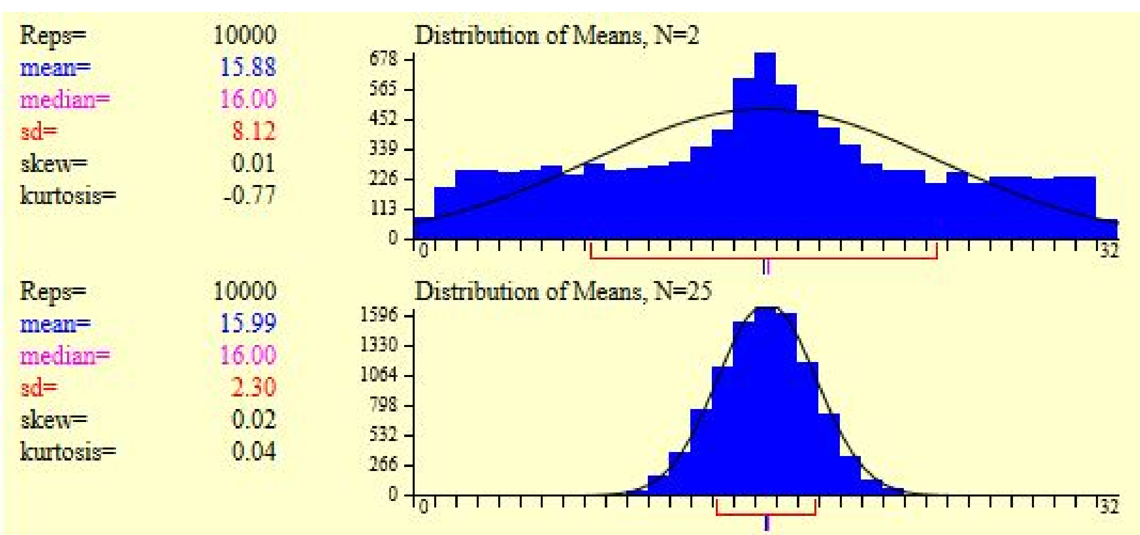
In this chapter we'll tie together what we know about distributions and sampling methods. Our primary focus will be on the distributions of a sample of means and the distribution of a sample of proportions. The surprising and underlying property we'll discover is that the sampling distributions for both means and proportions are approximately Normal (given certain conditions are met) *regardless of the shape of the original distribution.*

In exploring this, we'll use an interactive applet[1]. There's also a Khan Academy video that demonstrates the use of this applet that you may find helpful to reference later[2].

Example 1. Let's start by looking at a distribution that is definitely <u>not</u> Normal.



The two graphics below show what happens if you sample 10,000 times. The first takes only 2 values from the original sample and finds their mean. The second takes 25 values from the original sample and finds their mean. Each of these means are computed 10,000 times and result in the distributions below. The Normal curves have also been included in each.

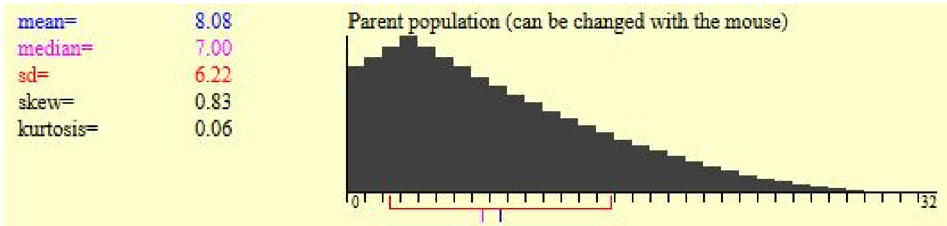


---

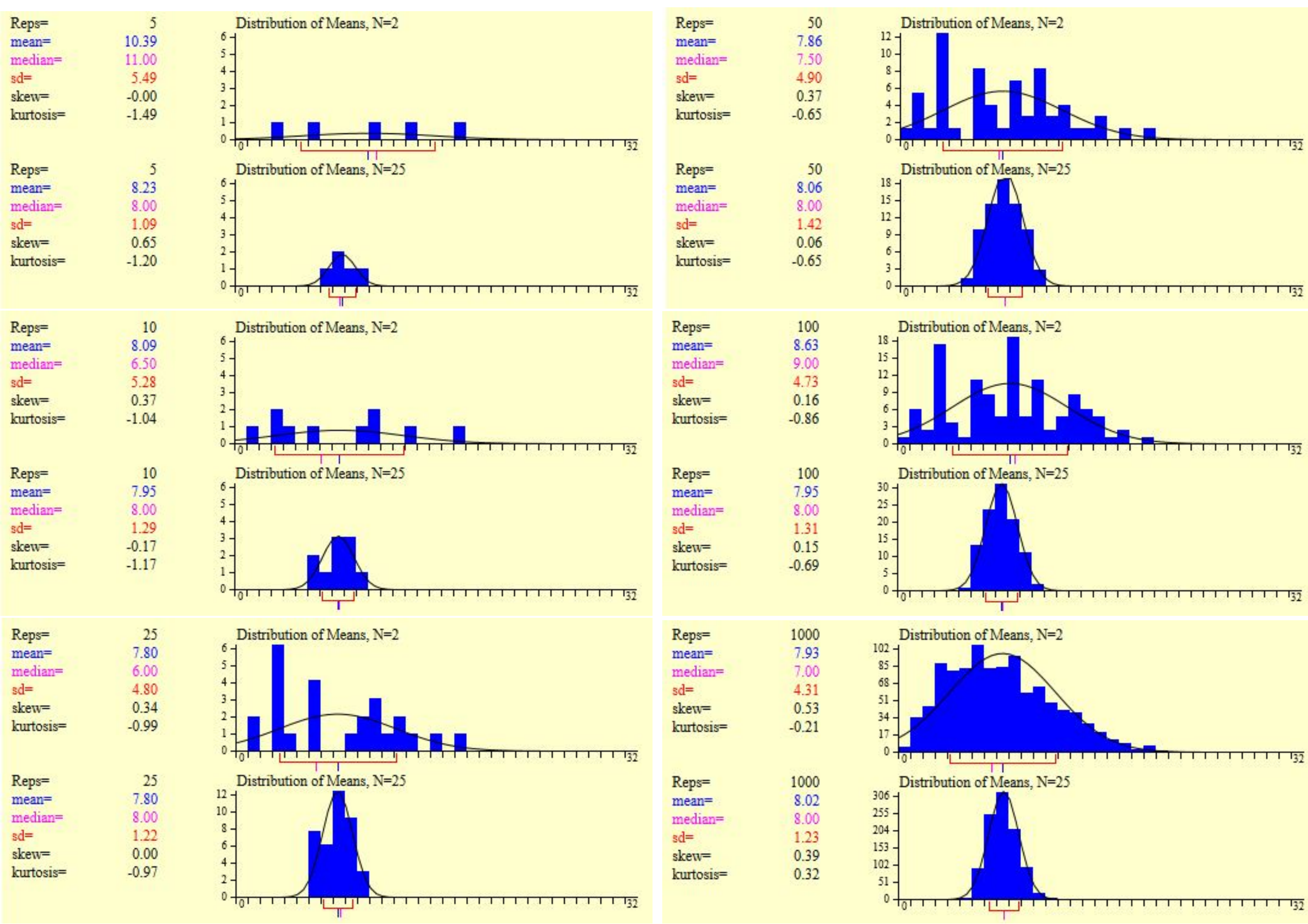
[1] http://onlinestatbook.com/stat_sim/sampling_dist/index.html

[2] https://www.khanacademy.org/math/probability/statistics-inferential/sampling_distribution/v/sampling-distribution-of-the-sample-mean

Example 2. Let's look at another distribution that is also definitely not Normal; in this case, it's skewed right.



The graphics below show the distributions if you sample 5 times, 10 times, 25 times, 100 times, and 1000 times. For each, the mean is first computed from two values and in the second the mean is computed from 25 values.



What (in general) happens to the sample mean as the number of samples gets larger?

What (in general) happens to the sample standard deviation as the number of samples gets larger?

Central Limit Theorem
The mean of a random sample is a random variable whose sampling distribution can be approximated by a Normal model. The larger the sample, the better the approximation will be.

The Central Limit Theorem requires the following conditions hold:

- Independence Assumption: The sampled values must be independent of each other.
- Randomization Condition: The samples need to be randomly chosen, or it's not safe to assume independence.
- Sample Size Condition: A Normal model is appropriate if a sample is **large enough**. This is very vague and is determined on a case-by-case basis. In general, the more skewed a distribution is, the larger the sample needs to be.

The Sampling Distribution Model for a Mean
When a random sample is drawn from any population with mean μ and standard deviation σ, we have the following mean and standard deviation for the sampling distribution:

$$\mu_{\bar{y}} = \mu \qquad\qquad \sigma_{\bar{y}} = SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$$

Regardless of the shape of the population distribution, the shape of the sampling distribution will be approximately Normal as long as the sample size is large enough.

Example 3. Service call lengths are measured with a population mean of 173.95 seconds and a population standard deviation of 184.81 seconds.

(a) If we choose a SRS of 20 calls, what's the mean of their mean length?

(b) If we choose a SRS of 20 calls, what's the standard deviation of their mean length?

(c) If we choose a SRS of 80 calls, what's the standard deviation of their mean length?

Example 4. A person's measured glucose level one hour after ingesting a sugary drink varies according to the Normal distribution with $\mu = 125$ mg/dl and $\sigma = 10$mg/dl.

(a) If a single glucose measurement is made, what's the probability that it's greater than 140mg/dl?

(b) If measurements are made on three different occasions and the mean result is computed, what's the probability that the mean is greater than or equal to 140mg/dl?

(c) What is the 95th percentile for the mean blood glucose level measurements?

Example 5. Restaurant bills at a given restaurant have an assumed population mean of \$32.40 and a population standard deviation of \$8.16. This data is heavily skewed to the left.

(a) Explain why you cannot determine that a given bill will be at least \$35.

(b) Can you estimate the probability that the next 5 bills will average at least \$35?

(c) How likely is it that the next 50 bills have an average of at least \$35?

Practice. A survey of college students found that their total sleep each night was approximately Normal with a mean of 6.78 hours and standard deviation of 1.24 hours.

(a) What's the probability that a single randomly chosen student gets 6.9 hours of sleep or fewer?

(b) For a SRS of 150 students, what is the probability that the average is below 6.9 hours?

(c) What is the mean amount of sleep that the top 5% of this sample of 150 students get?