Section 4.8 The Central Limit Theorem.

Every time we've calculated a probability, we needed to know the distribution of the population so we had a probability structure in our hands.

Random Variable?

Discrete?

Continuous?

The amount of toothpaste in a tube has a mean of 181 g and a standard deviation of 2 g.

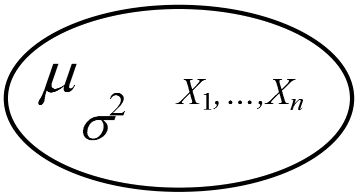
If one tube is selected at random, find the probability it contains less than 180 grams.

Most of the time we take a sample and look at the sample mean.

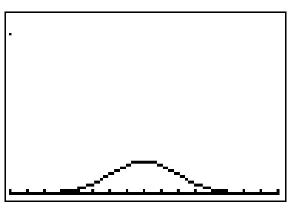
The sample mean is a random variable too...

We need to know which probability distribution to use so we can find probabilities involving the sample mean.

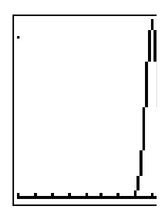
We talked about this already... $\overline{x}$ **is a random variable**

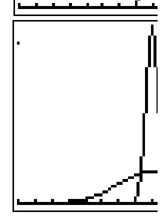


The amount of toothpaste in a tube has a mean of 181 g and a standard deviation of 2 g.
We assume volume is normally distributed.



If 30 tubes are selected at random, find the probability the mean weight is less than 180 grams.





So...if the population is normally distributed then the sample mean is a random variable that is also normally distributed.

BUT...

What if the population is not normal? (Oh no!)

What if we have no clue what the population distribution is?

(Insert dramatic music now.)

The Central Limit Theorem

Let $X_1, \ldots, X_n$ be a *large enough* random sample from a population with mean μ and variance σ^2.

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

No need to know the population distribution!

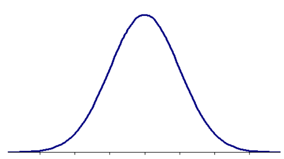
A manufacturer claims the lifetime of a component has a mean of 2 years and a standard deviation of 2 years.

Find the probability a component lasts at most 1.9 years.

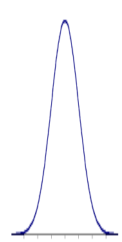
In a sample of 35 components, find the probability the sample mean is less than 1.9 years.

If you took a sample of 35 and found the sample mean lifetime to be 1.9 years, then would you conclude the manufacturer's claim seems reasonable?

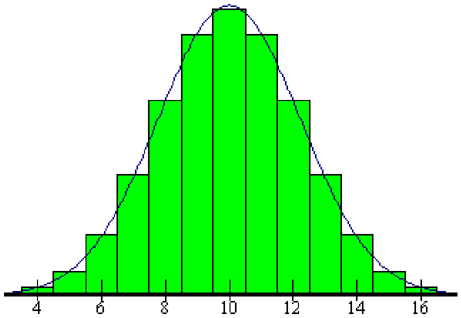
Human body temperatures are normally distributed with a mean of 98.2°F and a standard deviation of 0.7°F.



How large a sample should be taken so that only 5% of sample means will be less than 98 °F ?



Binomial Probabilities can be approximated using the Normal Distribution



As long as np and npq are both greater than 10 we get a good approximation when using the normal to approximate the binomial.

However...I do not want you to use the normal approximation. Just calculate binomial approximations directly. The normal approximation to the binomial is an outdated statistical practice.

Why is the Central Limit Theorem so important?

We need to describe properties of populations...called parameters.

The sample mean is one measure we often use...it is a statistic.

It would be pretty useless if sample means varied wildly.

http://www.intuitor.com/statistics/CentralLim.html

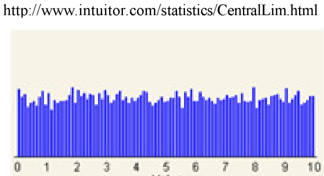


Fig. 1A) Histogram of Population - Uniform Distribution: all values in the population are randomly determined but all equally likely. This approximates a uniform distribution. Data points in population = 16,000; mean = 4.996 std dev 2.882

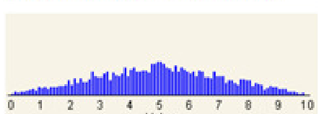


Fig. 1B) Sampling Distribution (from a uniform population) **n = 2:** number of samples = 4000; mean = 5.01; std dev 2.048; std error = 2.037

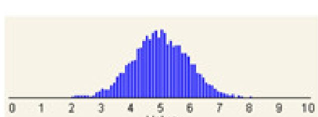


Fig. 1C) Sampling Distribution (from a uniform population) **n = 10:** number of samples = 2010; mean = 5.015; std dev 0.906; std error = 0.911