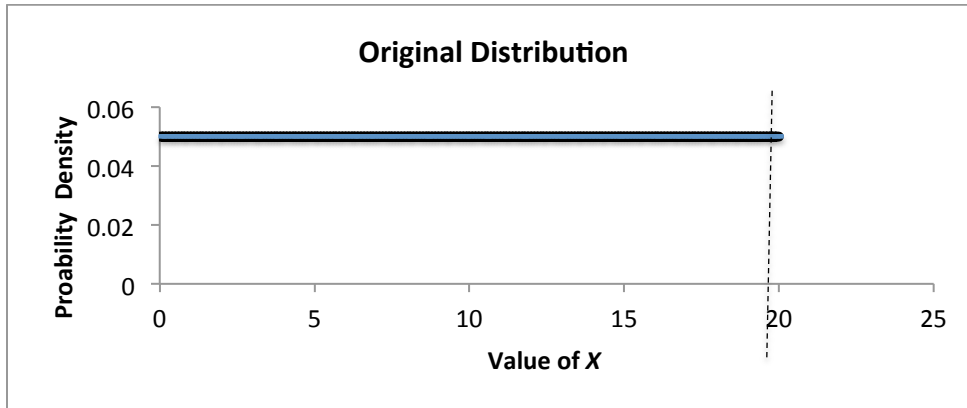


Class 11: Sampling Distributions and the Central Limit Theorem (5.1)

SAMPLING DISTRIBUTIONS for MEANS

In Excel Assignment #4, you took samples from the random variable X , with probability distribution:



This distribution is uniform, with mean 10 and standard deviation 5.77.

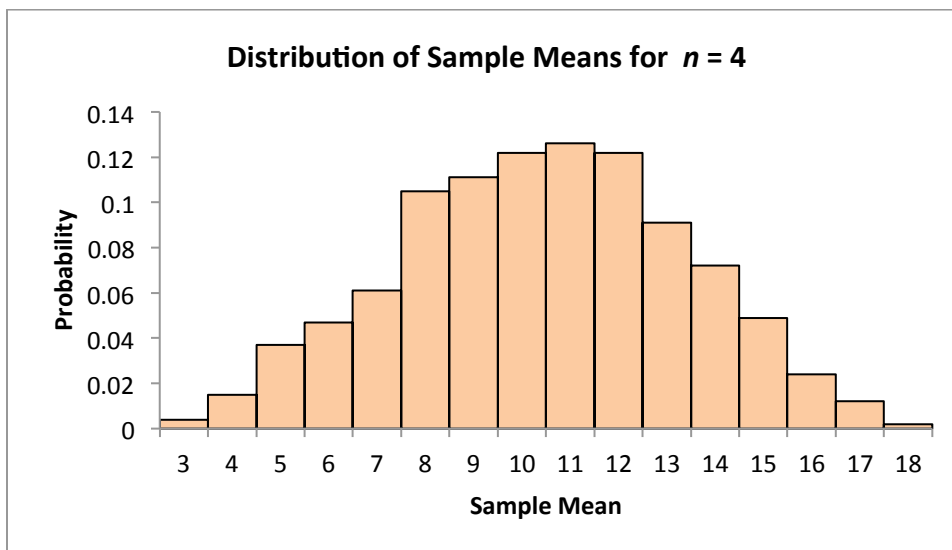
Distribution of means of 1000 samples of size 4 :

Mean about 10

One case: Mean

Standard deviation a bit less than 3

One case: Standard deviation



Shape of this distribution is

Mean of this distribution (the mean of all the means) is

Standard Deviation of this distribution is

Why is the standard deviation smaller? Taking the mean of a sample averages out outliers, so the means vary less than the original data. Thus the SD of the means is lower and the distribution more scrunched.

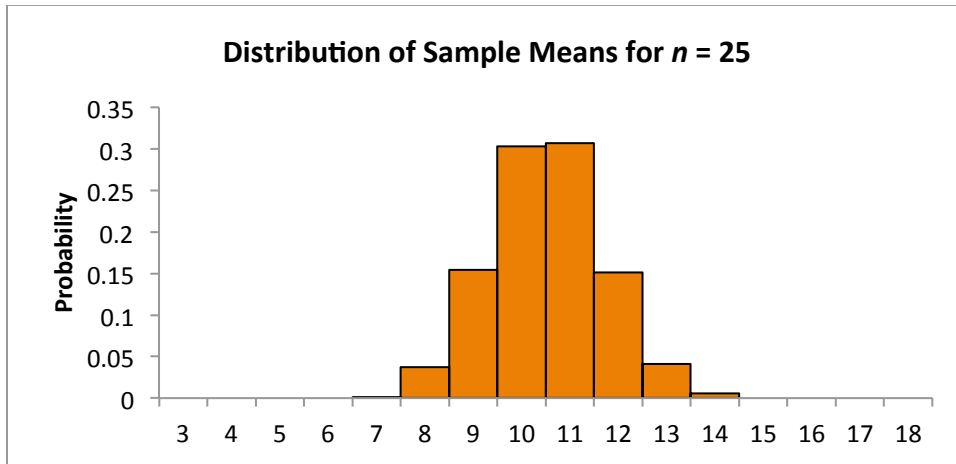
Distribution of means of 1000 samples of size 25:

Mean about 10

Standard deviation a bit more than 1

One case: Mean

One case: Standard deviation



Shape of this distribution is

Mean of this distribution (the mean of all the means) is

Standard Deviation of this distribution is

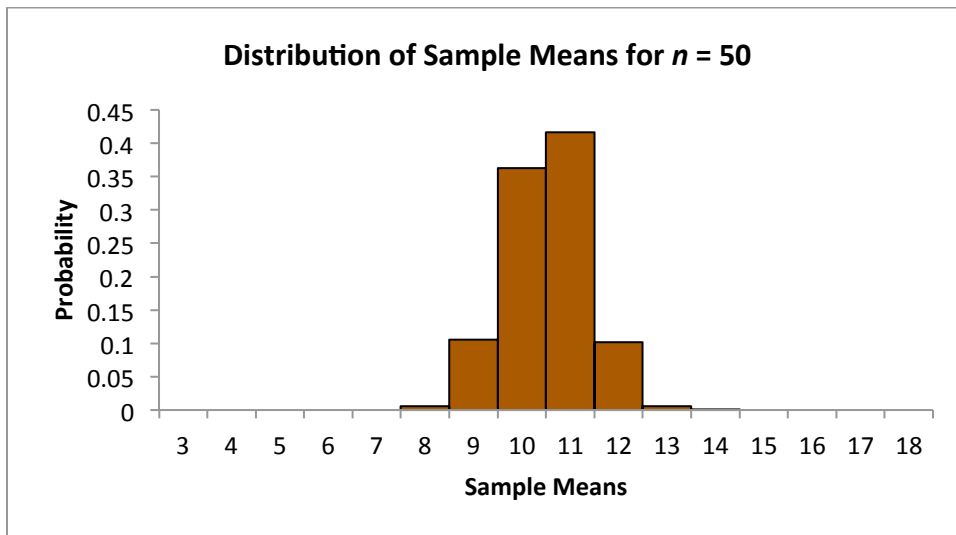
Distribution of means of 1000 samples of size 50:

Mean about 10

Standard deviation a bit less than 1

One case: Mean

One case: Standard deviation



Shape of this distribution is

Mean of this distribution (the mean of all the means) is

Standard Deviation of this distribution is

Overall

Shape of this distribution is

Mean of this distribution (the mean of all the means) is

Standard Deviation of this distribution is

The **Central Limit Theorem for Means** tells us that when we take random samples of a fixed size n from a population with mean μ and standard deviation σ , and calculate \bar{x} , the sample mean, then

- Distribution of \bar{x} is approximately normal
- Mean of \bar{x} is the population mean μ
- Standard deviation of \bar{x} is $\frac{\sigma}{\sqrt{n}}$

Ex: For Excel #4, we have

- Distribution of \bar{x} is approximately normal
- Mean of \bar{x} is the population mean
- Standard deviation of \bar{x} is

Note : The CLT for means applies for samples of size greater than 30, even if the original population distribution is not normal. If the population distribution is normal, then the sampling distribution is normal for any n . See the figures¹:

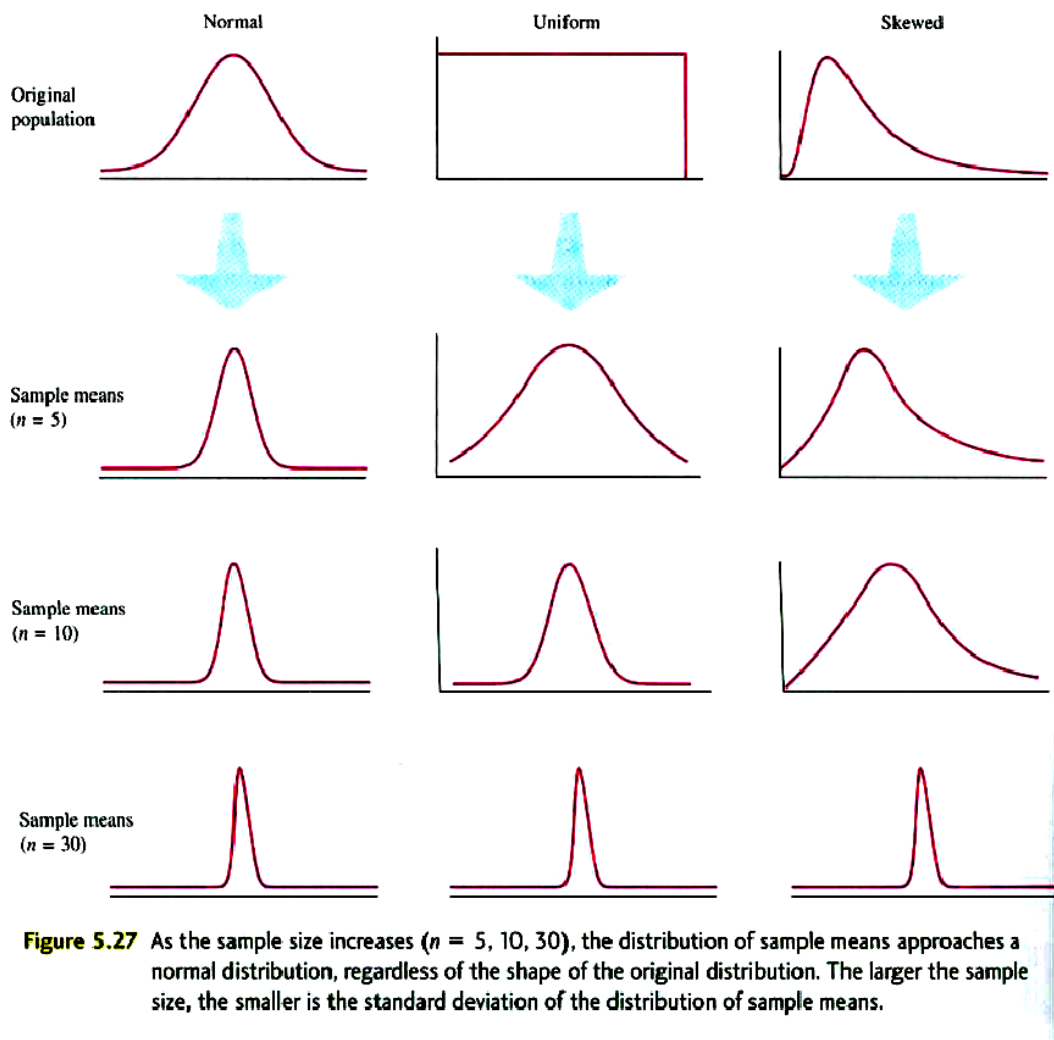


Figure 5.27 As the sample size increases ($n = 5, 10, 30$), the distribution of sample means approaches a normal distribution, regardless of the shape of the original distribution. The larger the sample size, the smaller is the standard deviation of the distribution of sample means.

¹ From *Statistical Reasoning for Everyday Life*, by Bennett, Briggs, and Triola (Addison Wesley 2nd edn)

What the Central Limit Theorem Tells Us Likely and Unlikely Population Values

On September 20, 2009, an Indonesian woman gave birth to a baby, Muhammad Akbar Risuddin, weighing 8.7 kg (19.2 pounds) and thought to be the heaviest baby ever born in Indonesia—more than double the weight of a normal newborn.² The weights of healthy newborns are normally distributed with mean 3.43 kg and standard deviation 0.48 kg.

Ex. What is Z-Score of Muhammad's weight? What does this number tell us?

Now we look at some more usual birth weights.

Ex. What is the likelihood of a newborn weighing 4 kg or more?

Populations Versus Samples

Ex: Which is more likely: An individual newborn weighing more than 4 kg, or a random sample of three newborns with mean weight more than 4 kg?

To make this argument more precise, we use the Central Limit Theorem for means.

Ex: What is the likelihood that the mean weight of three randomly selected newborns is 4 kg or above?

Thus, 11.5 % of individual newborns weigh 4 kg or more, but only 1.97% of samples of three newborns have a mean weight of 4 kg or more.

² <http://edition.cnn.com/2009/WORLD/asiapcf/09/25/indonesia.baby/index.html?iref=mpstoryview>

“Are Babies Normal” T.Demons, M.Pagano, *The American Statistician*

³ In 2005, a 7.73 kg (17 lb) baby boy was born in Brazil. The heaviest babies recorded were a 10.2 kg (22.5 lb) boy in Italy in 1955, and 10.8 Kg (23.8 lb) boy in the US in 1879 who died soon after birth.

Ex. How would the probability (now 1.97%) change if the sample size (now 3) were increased?

Ex: What is the weight of a baby at the 75th percentile of all healthy newborn babies?

Ex: What is the mean weight of a random sample of three healthy newborn babies at the 75th percentile of all such samples? Will the answer be large or smaller than the answer to the previous problem? Why?

