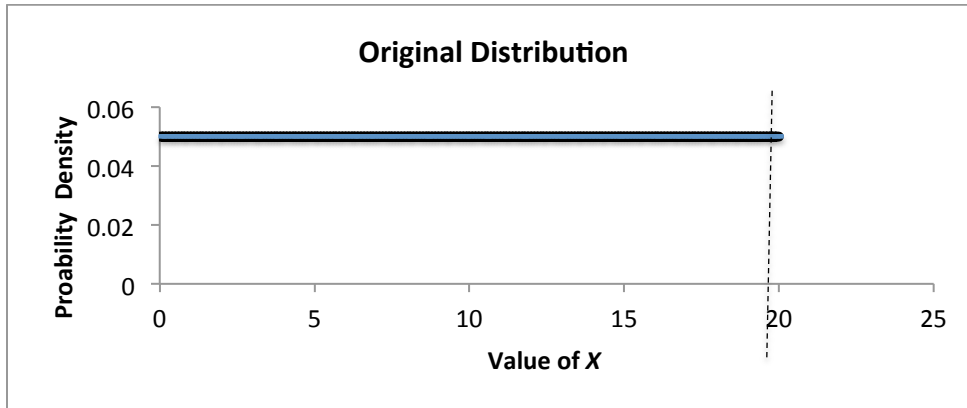


Class 11: Sampling Distributions and the Central Limit Theorem (5.1)

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.

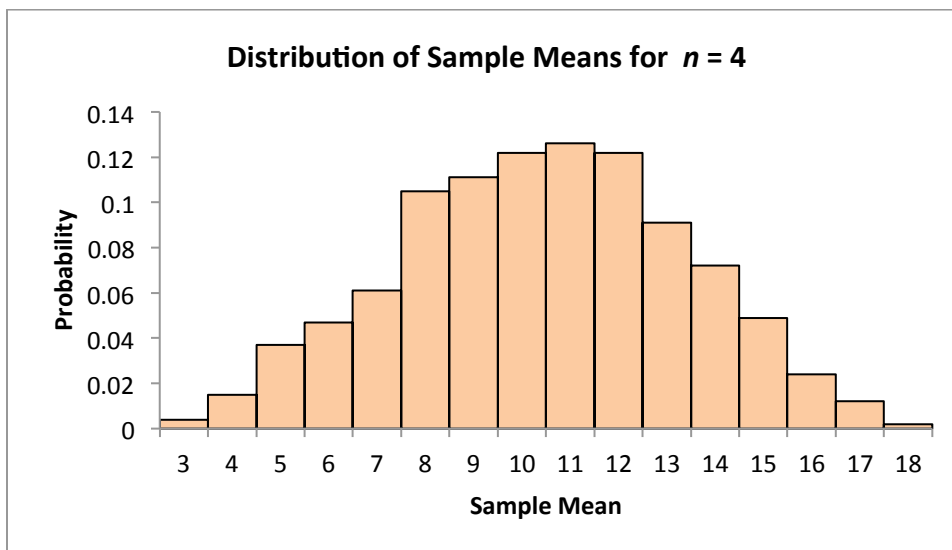
Distribution of means of 1000 samples of size 4 :

Mean about 10

Standard deviation a bit less than 3

One case: Mean = 10.11

One case: Standard deviation = 2.90



Shape of this distribution is more scrunched, approximately normal
 Mean of this distribution (the mean of all the means) is close to 10, the original mean
 Standard Deviation of this distribution is smaller than the original SD

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.

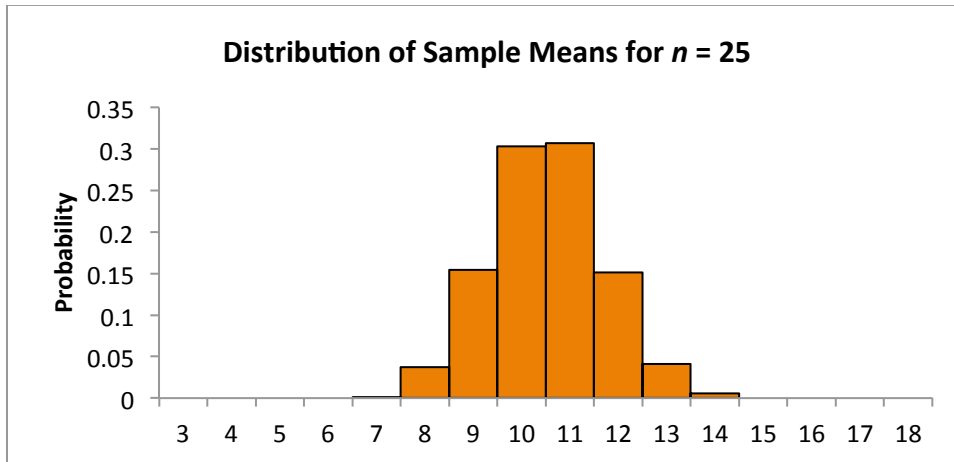
Distribution of means of 1000 samples of size 25:

Mean about 10

Standard deviation a bit more than 1

One case: Mean = 10.05

One case: Standard deviation = 1.14



Shape of this distribution is more scrunched than the previous one

Mean of this distribution (the mean of all the means) is close to 10, the original mean

Standard Deviation of this distribution is smaller than the previous one

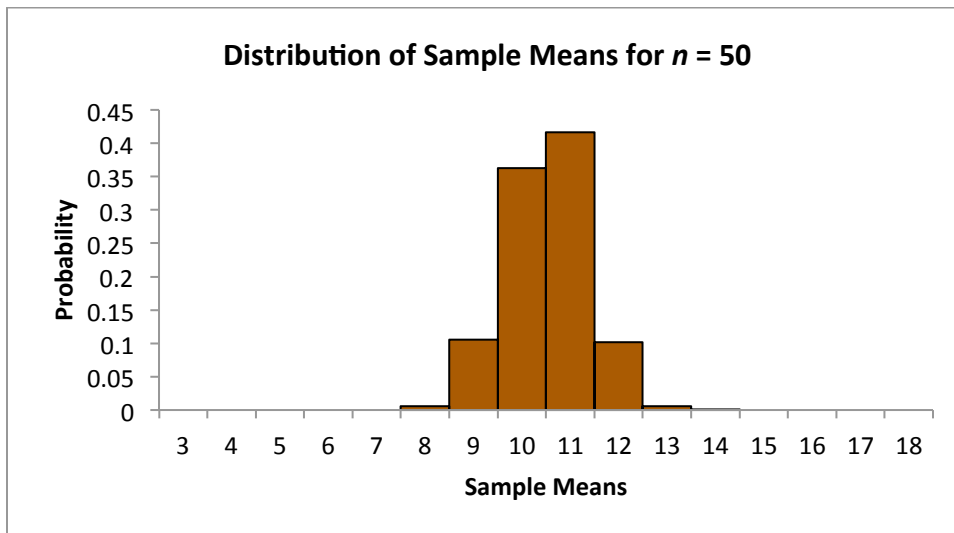
Distribution of means of 1000 samples of size 50:

Mean about 10

Standard deviation a bit less than 1

One case: Mean = 9.97

One case: Standard deviation = 0.83



Shape of this distribution is more scrunched still than the previous one, approximately normal

Mean of this distribution (the mean of all the means) is close to 10, the original mean

Standard Deviation of this distribution is even smaller than the previous one

Overall

Shape of this distribution is approximately normal

Mean of this distribution (the mean of all the means) is close to the original mean

Standard Deviation of this distribution is smaller than the original SD, decreasing with sample size

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 10
- Standard deviation of \bar{x} is $\frac{5.77}{\sqrt{n}}$ so
 - For $n = 4$, $SE = \frac{5.77}{\sqrt{4}} = 2.89$,
 - for $n = 25$, $SE = \frac{5.77}{\sqrt{25}} = 1.15$, and
 - for $n = 50$, $SE = \frac{5.77}{\sqrt{50}} = 0.82$.

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:

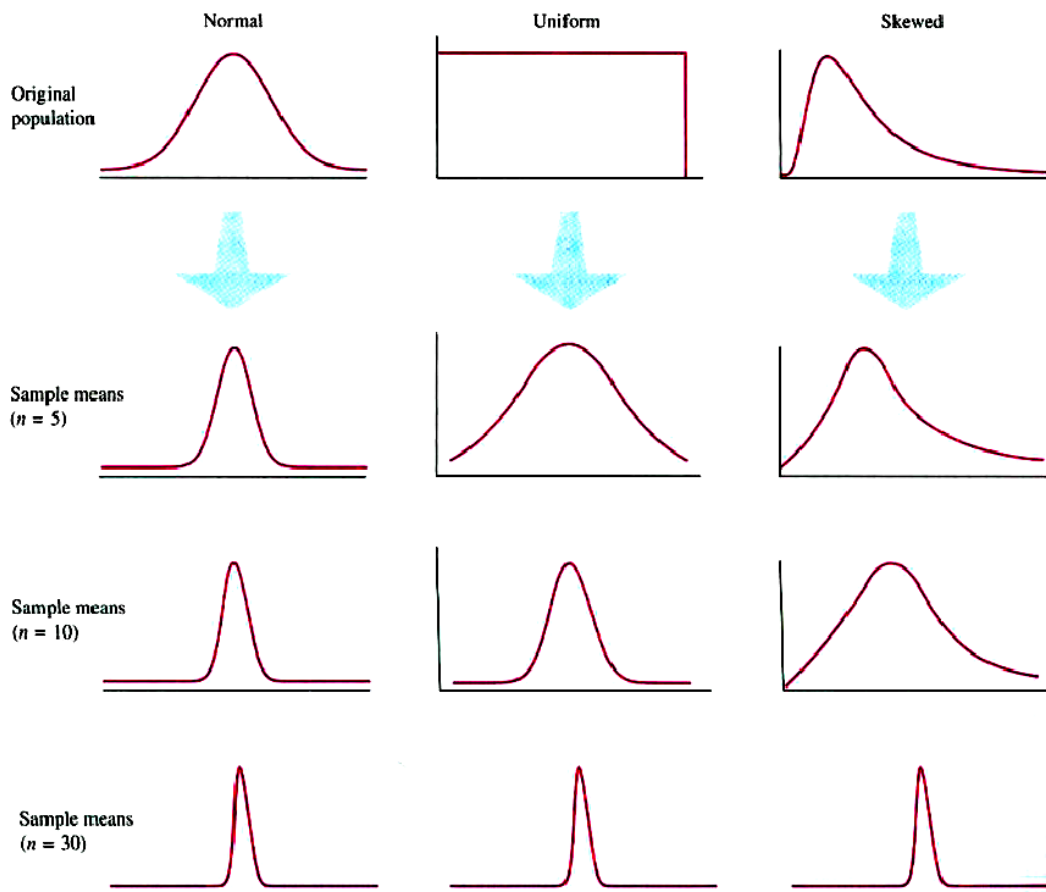


Figure 5.27 As the sample size increases ($n = 5, 10, 30$), the distribution of sample means approaches a normal distribution, regardless of the shape of the original distribution. The larger the sample size, the smaller is the standard deviation of the distribution of sample means.

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?

$$Z = \frac{8.7 - 3.43}{0.48} = 11$$

Thus Muhammad's weight was 11 standard deviations above the mean—a staggeringly unlikely event.³

Now we look at some more usual birth weights.

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

$$Z = \frac{4 - 3.43}{0.48} = 1.18$$

From the table, $Z = 1.2$ corresponds to a probability of 0.8810. So the percentage above $Z = 1.2$ is $1 - 0.8810 = 0.1190 = 11.9\%$. Thus about 12 % of newborns weigh 4 kg or more—not uncommon.

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?

The weights of individual newborns are more variable than the average weight of samples, because extreme values in the sample (very high or very low) tend to average out. Thus the probability of an individual newborn weighing more than 4 kg is *larger* than the probability of a sample of three newborns having 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?

By the Central Limit Theorem, the mean weights, \bar{x} , of samples of three newborns are distributed normally with mean 3.43 kg and standard deviation $0.48 / \sqrt{3} = 0.277$ kg.

The Z-value for a random sample of three newborns with mean weight of 4 kg is

$$Z = \frac{4 - 3.43}{0.277} = 2.06$$

From the table, $Z = 2.06$ corresponds to a probability of 0.9803, so the likelihood of an observation above $Z = 2.06$ is $1 - 0.9803 = 0.0197 = 1.97\%$.

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?

The percentage would decrease still further if the sample size were increased.

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

From the table, the 75th percentile corresponds to a value of $z = 0.67$. Thus

$$0.67 = \frac{x - 3.43}{0.48}$$

$$x = 3.43 + 0.67(0.48) = 3.75 \text{ kg.}$$

Thus 75% of babies weigh less than 3.75 kg.

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?

The answer to this problem will be smaller, as the means of samples are less variable than the individual weights. Thus 75% of the means of samples will be closer to 3.43.

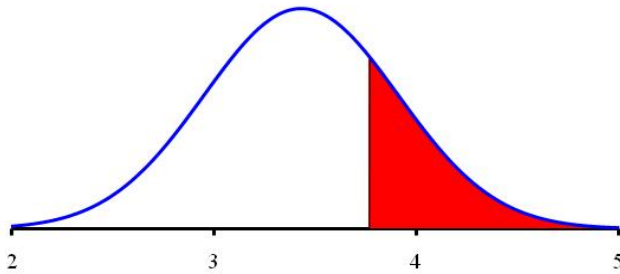
The standard deviation is now 0.277 kg, so

$$0.67 = \frac{x - 3.43}{0.277}$$

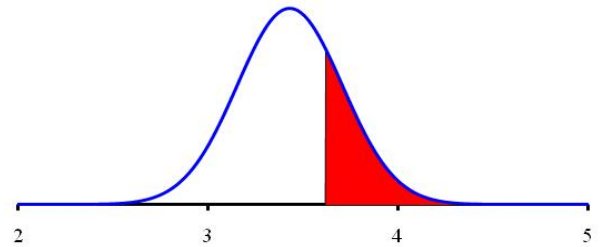
$$x = 3.43 + 0.67(0.277) = 3.62 \text{ kg.}$$

Thus 75% of the samples of three babies have mean weight less than 3.62 kg.

Show your answers on diagrams of the two distributions:



Weights of individual babies



Mean weights of samples of three babies