JAMES COOK UNIVERSITY AUSTRALIA

Basic Statistics

Sample vs. Population Distributions

JAMES COOK UNIVERSITY AUSTRALIA

Learning, Teaching and Student Engagement

Revision

Remember:

- Sample vs Population
- Variance
- Standard Deviation
- Standard Normal Distribution

Image accessed: http://intouchacquisitions.co.uk/in-touch-acquisitions-review-the-importance-of-business-statistics/

Sample vs Population

Remember:

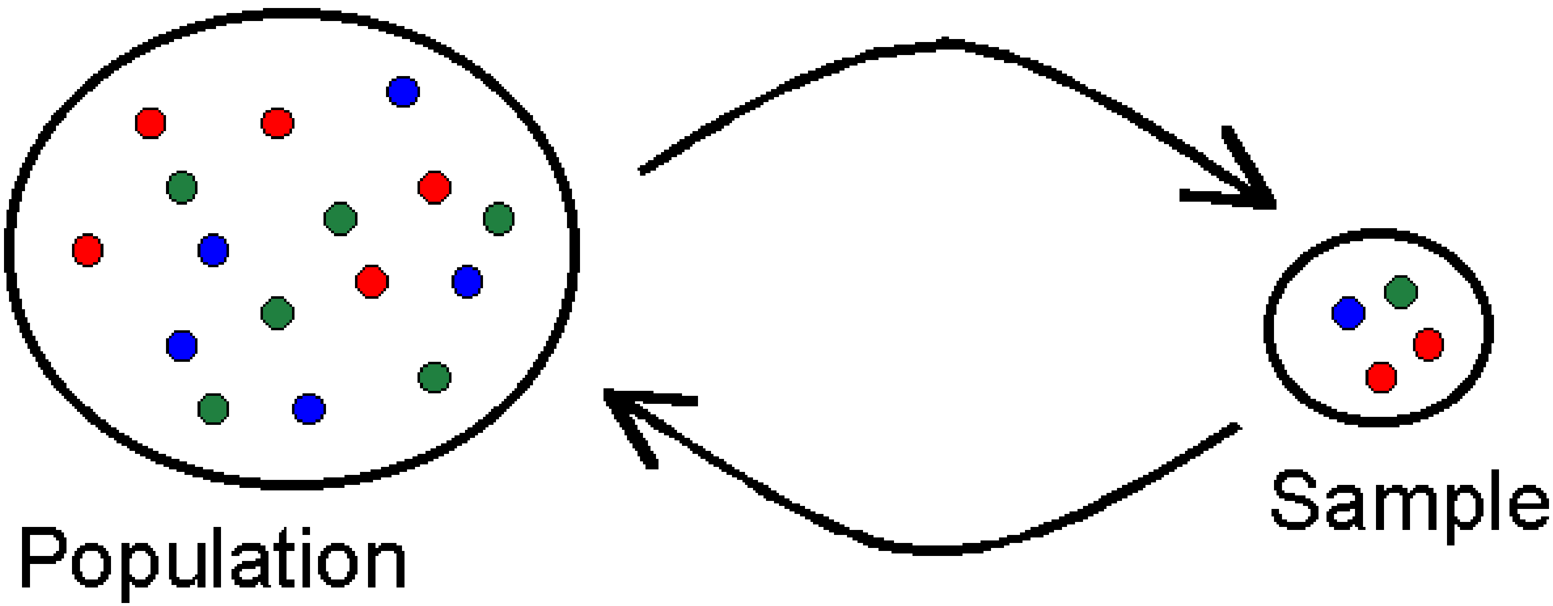


Image accessed: https://scholar.vt.edu/access/content/group/43c8db00-e78f-4dcd-826c-ac236fb59e24/STAT5605/normal01.htm

Population Variance

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

Where:

σ^2 = variance of the population (pronounced sigma squared)

x_i = the measurement of each data unit in the population

μ = the population mean

n = the size of the population

Sample Variance

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

Where:

s^2 = the variance of the sample

x_i = the measurement of each data unit in the sample

$\bar{x}$ = the sample mean

n = the size of the sample (the number of data values)

Standard Deviation - Population

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

Where:

σ = the standard deviation of the population

x_i = the measurement of each data unit in the population

μ = the population mean

n = the size of the population

Standard Deviation - Sample

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Where:

s = the standard deviation of the sample

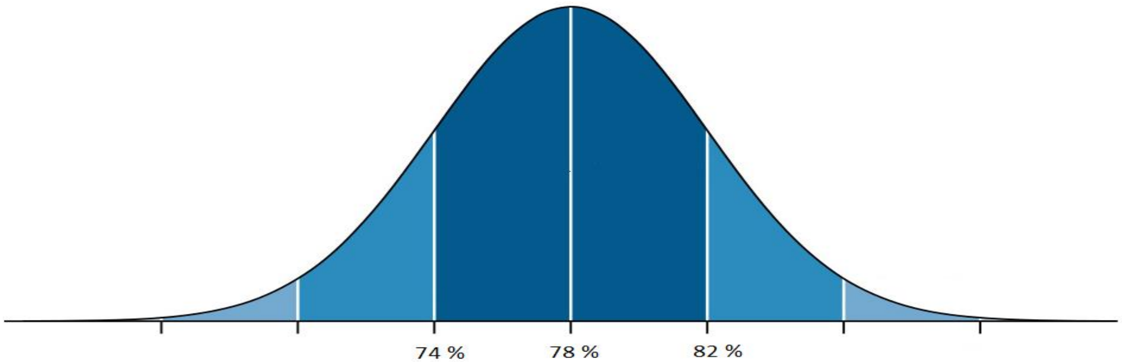
x_i = the measurement of each data unit in the sample

$\bar{x}$ = the sample mean

n = the size of the sample (the number of data values)

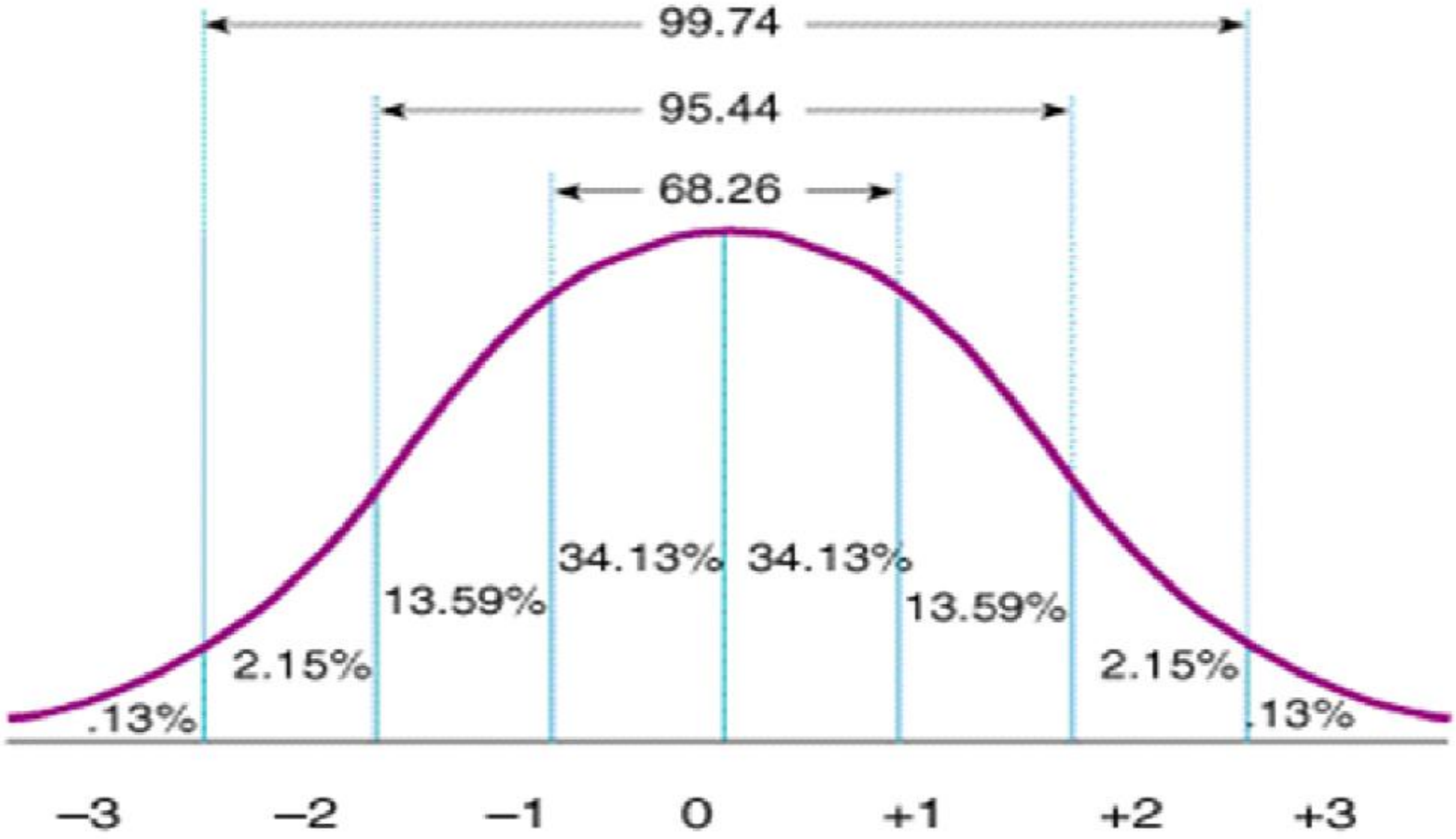
Normal Probability Distribution

- Because the area under the curve = 1 and the curve is symmetrical, we can say the probability of getting more than 78 % is 0.5, as is the probability of getting less than 78 %



- To define other probabilities (ie. The probability of getting 81 % or less) we need to define the **standard normal distribution**

Standard Normal Distribution



- Normal distribution with μ = 0 and SD = 1

Normal Probability Distribution

- We do this using the following formula

$$z = \frac{x - \mu}{\sigma}$$

x = the normally distributed random variable of interest

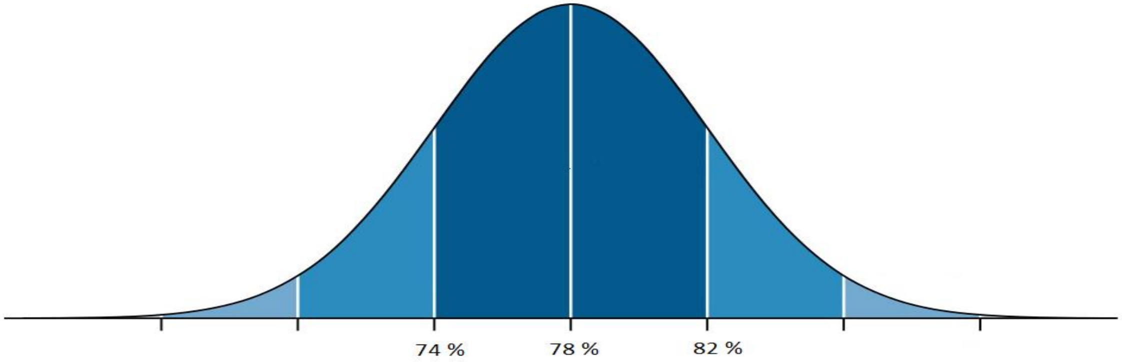
μ = the mean for the normal distribution

σ = the standard deviation of the normal distribution

z = the z-score (the number of standard deviations between x and μ)

Normal Probability Distribution

- To determine the probability of getting 81 % or less



$$z = \frac{x - \mu}{\sigma} = \frac{81 - 78}{4} = 0.75$$

Normal Probability Distribution

- Now that you have the standard z-score (0.75), use a z-score table to determine the probability

x	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621

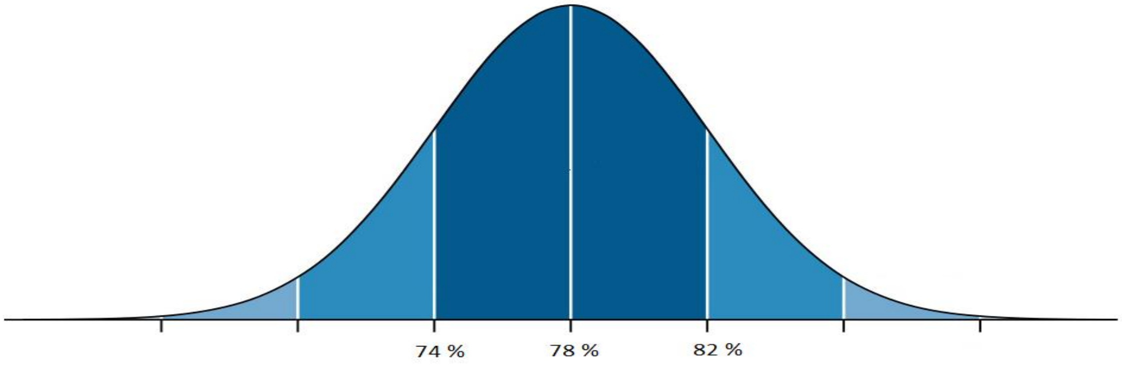
Normal Probability Distribution

- Z = 0.75, in this example, so we go to the 0.7 row and the 0.05 column

x	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621

Normal Probability Distribution

- The probability that the z-score will be equal to or less than 0.75 is 0.7734
- Therefore, the probability that the score will be equal to or less than 81 % is 0.7734



- There is a 77.34 % chance I will get 81 % or less on my test

Sample vs Population Distributions

Learning Intentions

Today we will understand:

- Using sampling distributions of the mean and proportion
- Working with the central limit theorem
- Using standard error of the mean

What is a Sampling Distribution?

- The **sampling distribution of the mean** refers to the pattern of sample means that will occur as samples are drawn from the population at large

Example

I want to perform a study to determine the number of kilometres the average person in Australia drives a car in one day. It is not possible to measure the number of kilometres driven by every person in the population, so I randomly choose a sample of 10 people and record how far they have driven.

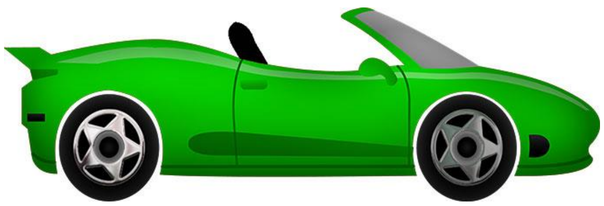


Image accessed: http://www.fg-a.com/autos.htm

What is a Sampling Distribution?

I then randomly sample another 10 drivers in Australia and record the same information. I do this a total of 5 times. The results are displayed in the table below.

Sample Number	Average number of Kilometres
1	25.6
2	50.2
3	15.1
4	43.9
5	36.8

What is a Sampling Distribution?

- Each sample has its own mean value, and each value is different
- We can continue this experiment by selecting and measuring more samples and observe the pattern of sample means
- This pattern of sample means represents the **sampling distribution** for the number of kilometres a person the average person in Australia drives

Image accessed: http://www.carclipart.com/free_car_clipart/cute_little_green_cartoon_car_0071-1006-2115-2057.html

Sampling Distribution of the Mean

- The distribution from this example represents the **sampling distribution of the mean** because the mean of each sample was the measurement of interest

- **What happens to the sampling distribution if we increase the sample size?**

Don't confuse sample size (n) and the number of samples. In the previous example, the sample size equals 10 and the number of samples was 5.

The Central Limit Theorem

- **What happens to the sampling distribution if we increase the sample size?**

- As the sample size (n) gets larger, the sample means tend to follow a normal probability distribution

- As the sample size (n) gets larger, the sample means tend to cluster around the true population mean

- Holds true, regardless of the distribution of the population from which the sample was drawn

Image accessed: http://resultsci.com/risking-business-success-coin-toss/

Standard Error of the Mean

- As the sample size increases, the distribution of sample means tends to converge closer together – to cluster around the true population mean
- Therefore, as the sample size increase, the standard deviation of the sample means decreases
- The standard error of the mean is the standard deviation of the sample means

Image accessed: http://www.heightdb.com/blog/best-time-of-day-to-measure-your-height

Standard Error of the Mean

- Standard error can be calculated as follows:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Where:

$\sigma_{\bar{X}}$ = the standard deviation of the sample means (standard error)

σ = the standard deviation of the population

$\sqrt{n}$ = the sample size

Standard Error of the Mean

- In many applications, the true value of σ (the SD of the population) is unknown

- SE can be estimated using the sample SD

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Where:

$SE_{\bar{x}}$ = the standard deviation of the sample means (standard error)

s = the sample standard deviation (the sample based estimate of the SD of the population

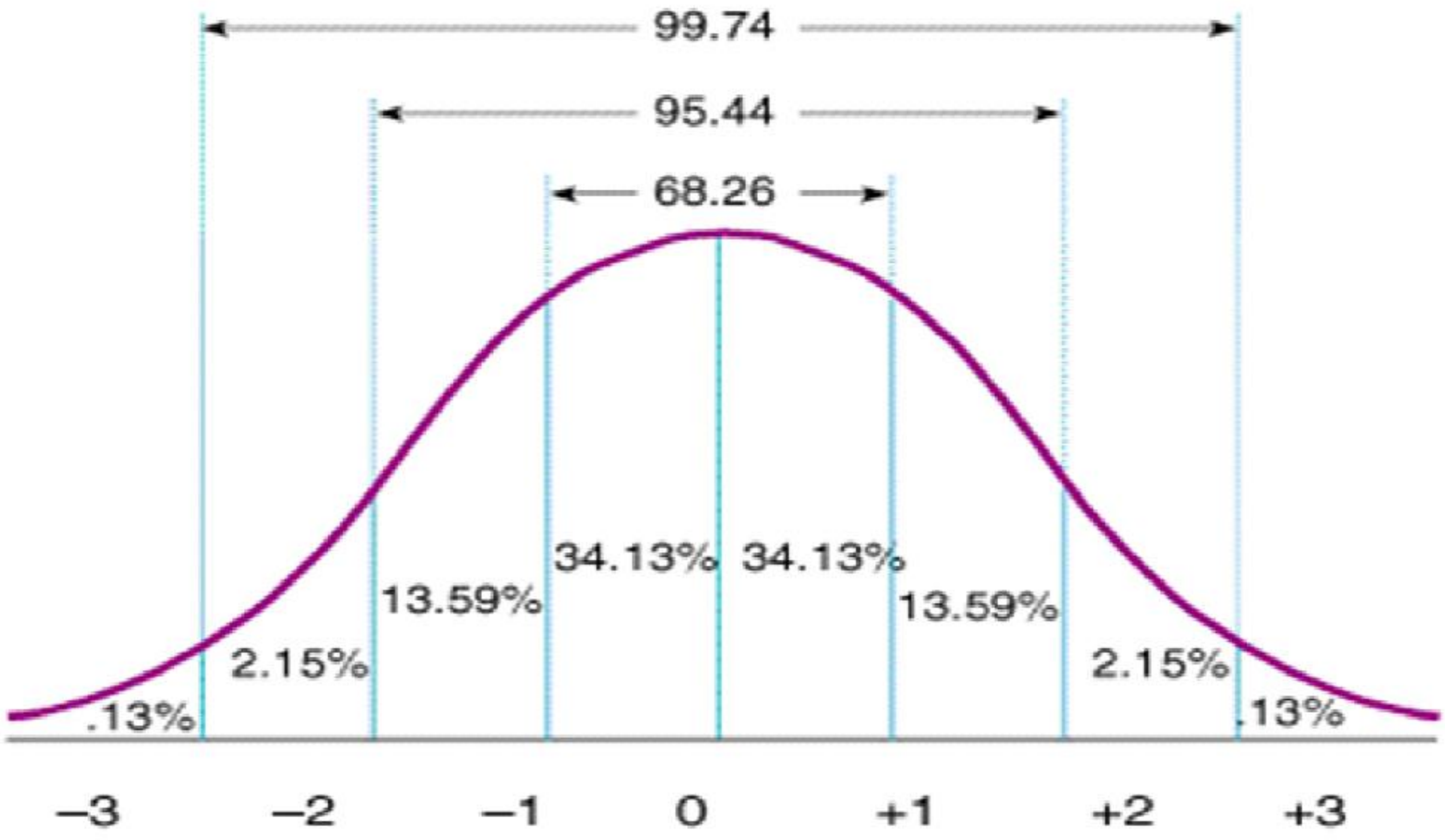
$\sqrt{n}$ = the sample size

Using the Central Limit Theorem

- If we know that the sample means follow the normal probability distribution
- And we can calculate the mean and standard deviation of that distribution
- We can predict the likelihood that the sample means will be more or less than certain values

Image accessed: http://www.free-training-tutorial.com/math-games/probability-fair.html?1&

Remember! - Standard Normal Distribution



- Normal distribution with μ = 0 and SD = 1

Using the Central Limit Theorem

- As we did last week, we can calculate the z-score
- Determine probability using a standard z table

x	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621