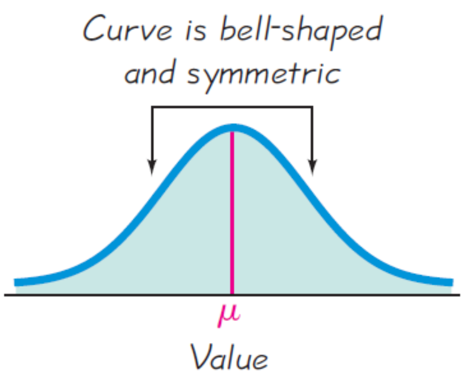
Chapter 6 - Normal/Continuous Probability Distributions:

If a continuous random variable has a distribution with a graph that is symmetric and bell-shaped and it can be described by the equation given as given below, we say that it has a normal distribution.



$$y = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$$

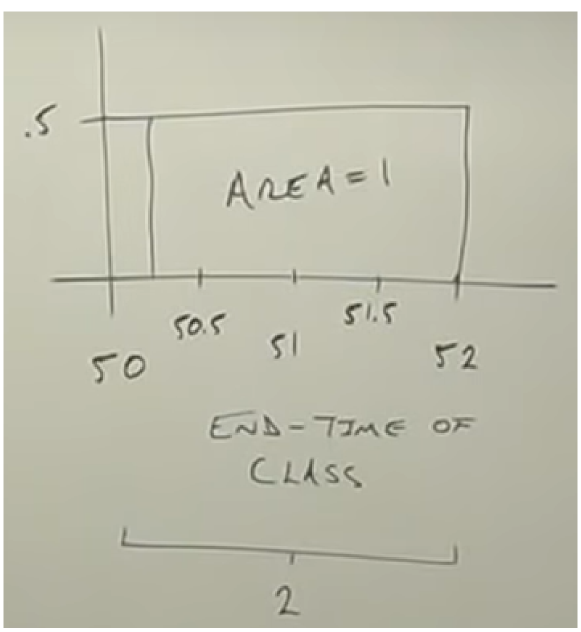
Uniform Distributions:

All the values have the same probability of occurrence.

The uniform distribution allows us to see two very important properties:

1. The area under the graph of a probability distribution is equal to 1.

2. There is a correspondence between area and probability (or relative frequency), so some probabilities can be found by identifying the corresponding areas.



The graph of a continuous probability distribution, such as in Figure 6-2, is called a density curve. A density curve must satisfy the following two requirements.

Requirements for a Density Curve:

1. The total area under the curve must equal 1.

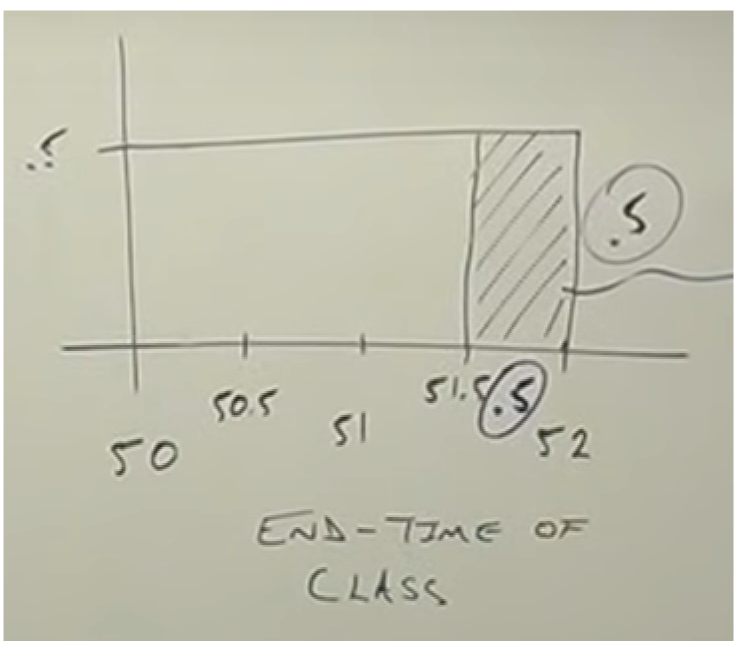
2. Every point on the curve must have a vertical height that is 0 or greater and less than 1. (That is, the curve cannot fall below the x-axis.)

By setting the height of the rectangle in the above graph to be 0.5, we force the enclosed area to be 2 * 0.5 = 1, as required. (In general, the area of the rectangle becomes 1 when we make its height equal to the value of 1/range) The requirement that the area must equal 1 makes solving probability problems simple, so the following statement is important:

" Because the total area under the density curve is equal to 1, there is a correspondence between area and probability ".

Eg:

Find the probability that class will end between 51.5 to 52 minutes



A = 0.5 * 0.5 = 0.25

P = 25 %   (Area and probability are same for continuous random variable)

Standard Normal Distribution:

The density curve of a uniform distribution is a horizontal line, so we can find the area of any rectangular region by applying this formula: Area = Height * Width. Because the density curve of normal distribution has a complicated bell shape as shown in the bell curve figure, it is more difficult to find areas. However, the basic principle is the same: There is a correspondence between area and probability. The figure below shows that for a standard normal distribution, the area under the density curve is equal to 1.

The standard normal distribution, also called the z-distribution, is a special normal distribution where the mean is 0 and the standard deviation is 1. Any normal distribution can be standardised by converting its values into z-scores. Z-scores tell you how many standard deviations from the mean each value lies.

The standard normal distribution is a normal probability distribution with μ = 0 and σ = 1. The total area under its density curve is equal to 1.

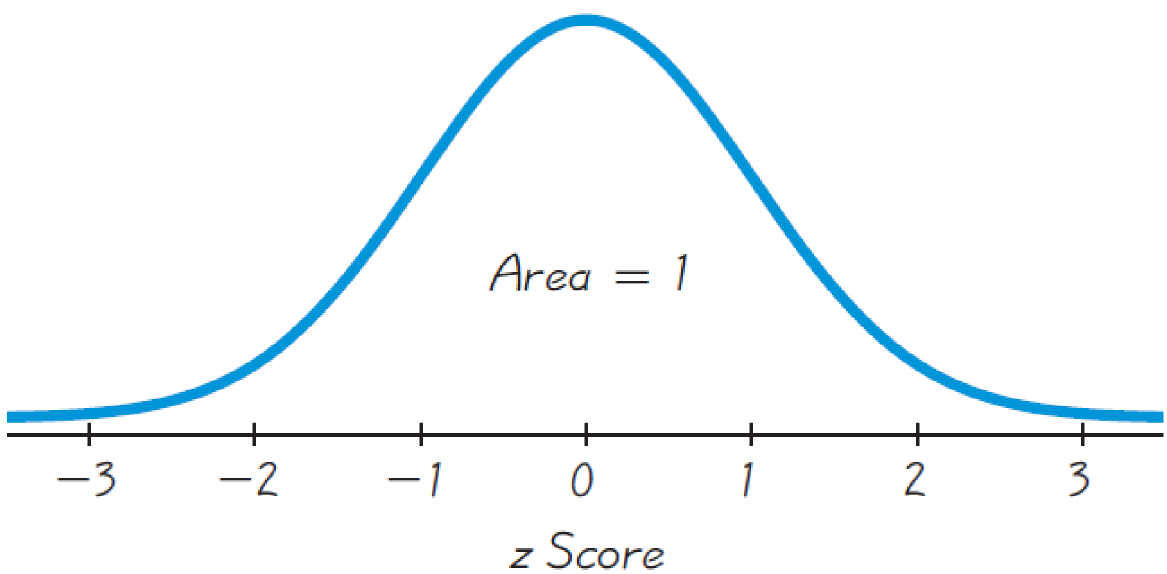


Figure 6-4 Standard Normal Distribution: Bell-Shaped Curve with $\mu = 0$ and $\sigma = 1$

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

x = value (continuous random variable).
z = standard normal distribution.

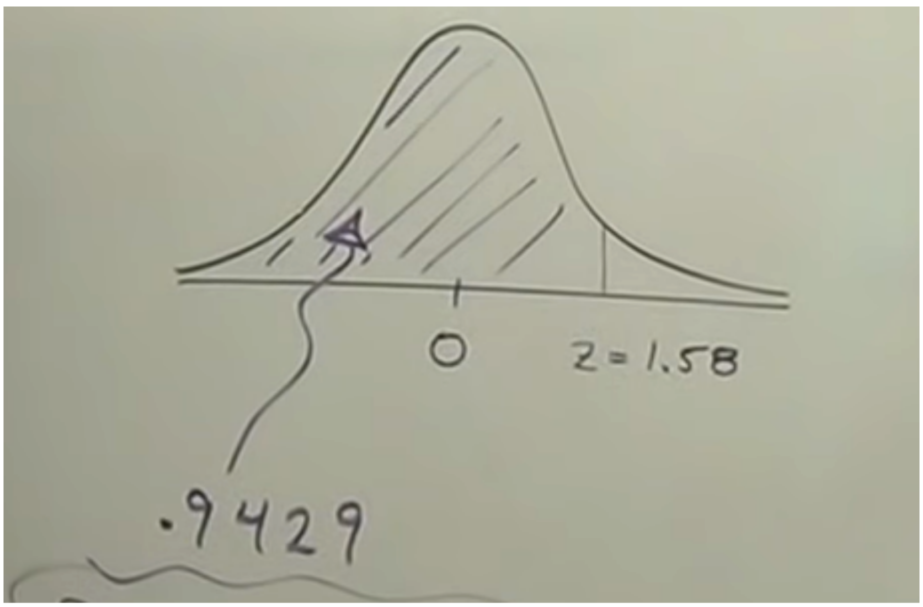
Eg:

Finding Areas from Z scores

1. The company tested the thermometers and found that mean of the thermometers was 0, and their SD was 1. Readings were normally distributed. Find the probability that a thermometer will have a reading of fewer than 1.58 degrees.

1. **Convert the value into a Z score.**
2. **Draw the picture (bell curve).**
3. **Find Area from table**

$$z = \frac{x - \mu}{\sigma} = \frac{1.58 - 0}{1} = 1.58$$



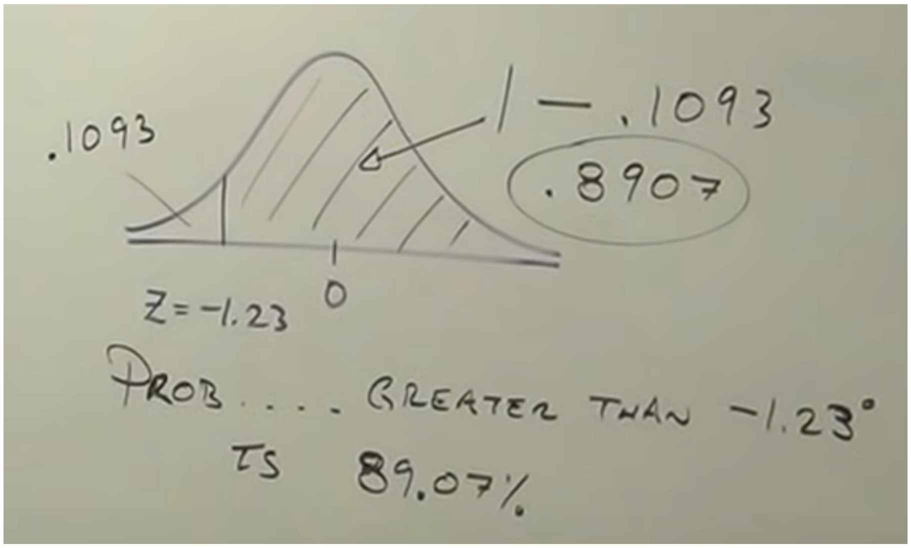
A = 94.29 % = 0.9429 (from SND table)

Normal CDF - Cumulative distribution function.

The probability of selecting a thermometer with a reading of fewer than 1.58 degrees is 0.94

2. The company tested the thermometers and found that mean of the thermometers was 0, and their SD was 1. Readings were normally distributed. Find the probability that a thermometer will have a reading of more than -1.23 degrees.

$$z = \frac{x - \mu}{\sigma} = \frac{-1.23 - 0}{1} = -1.23$$



A = .10935 = 10.93 %

A = 1 - A = 1 - 0.1093 = 89.07

The probability of selecting a thermometer with a reading of more than -1.23 degrees is 89.07 %

3. The company tested the thermometers and found that mean of the thermometers was 0, and their SD was 1. Readings were normally distributed. Find the probability that a thermometer will have a reading between -2 degrees and 1.50 degrees.

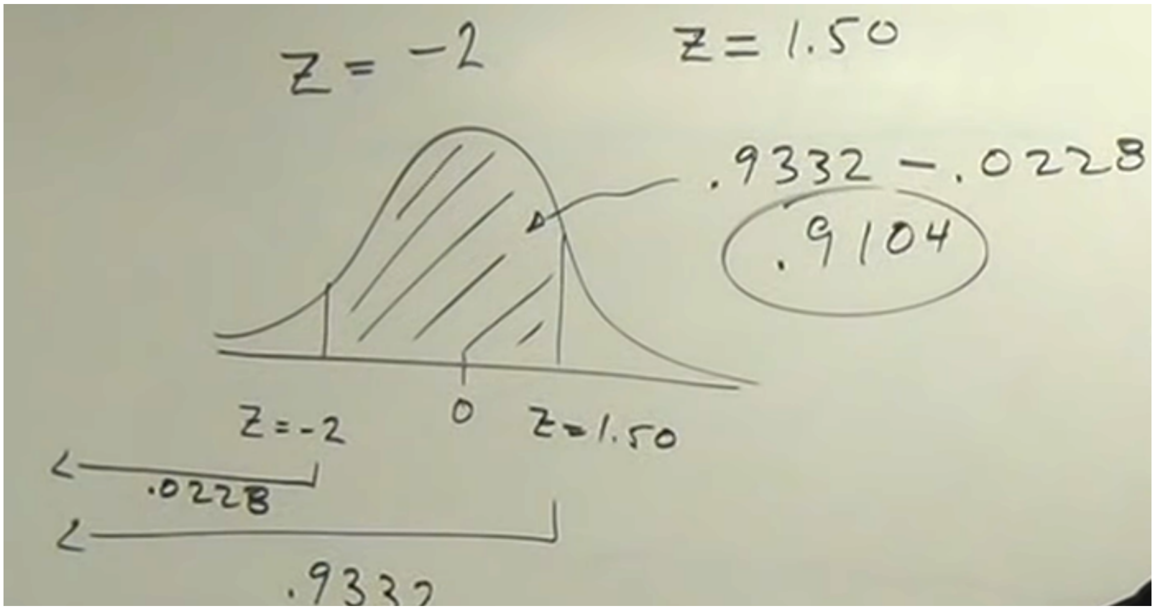
$$z = \frac{x - \mu}{\sigma} = \frac{-2 - 0}{1} = -2$$

$$z = \frac{x - \mu}{\sigma} = \frac{1.50 - 0}{1} = 1.50$$

z = -2 z = 1.50

A for z = -2 = 0.02275

A for z = 1.50 = 0.9331



A = 0.9331 - 0.02275 = 0.90535 = 90.53%

Finding Z scores from Areas. (Work backwards by using an Area to Find the distance from the mean (Z score))

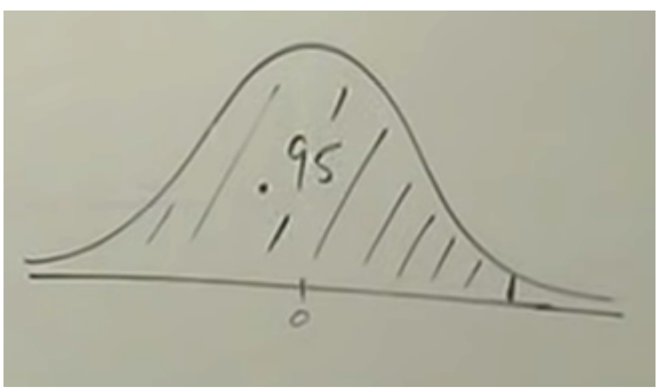
Steps:

1. Draw the Picture (Area of a bell curve)

2. Use table in reverse

https://www.math.arizona.edu/~rsims/ma464/standardnormaltable.pdf

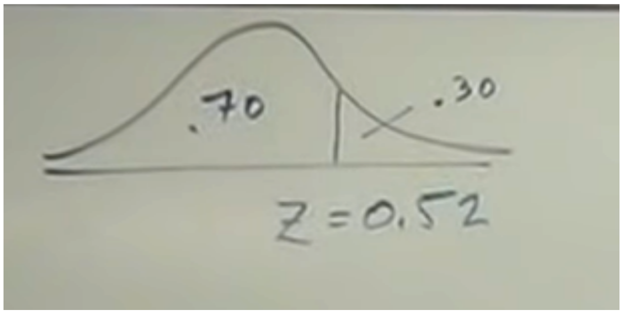
1. For thermometer: Find the Z score that represents the bottom 95%.



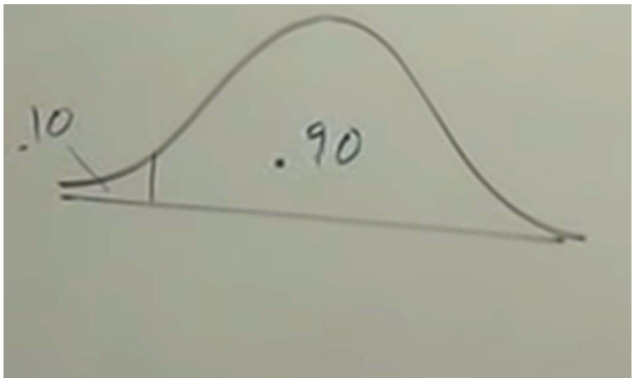
From table - 1.645

So, 95% of the thermometers will have a reading of fewer than 1.645 degrees.

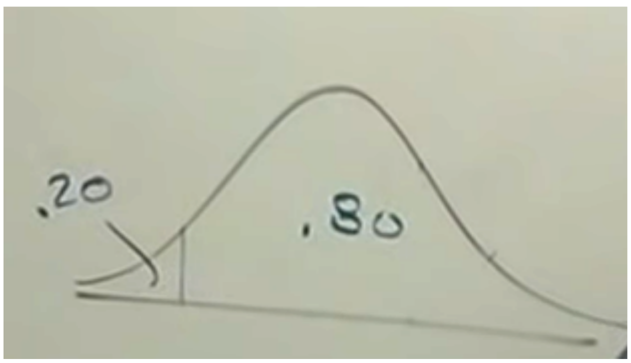
2. Draw the diagrams for the top 30%



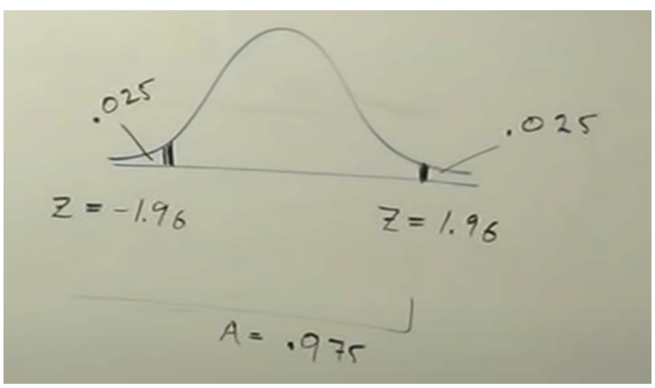
3. Draw the diagram for the bottom 10%



4. Draw the diagram for the top 80 %



5. Find the z scores that give the area between the top 2.5 % and the bottom 2.5 %.



Applications of Normal Distributions:

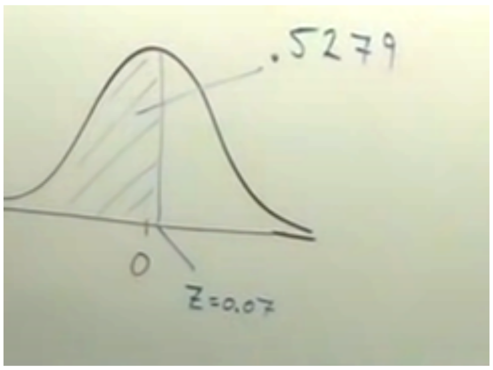
$$z = \frac{x - \mu}{\sigma}$$

Eg:

1. A population of men has a mean weight of 172 lbs and a standard deviation of 29 lbs, find the probability that a randomly selected man will have a weight of less than 174 lbs.

$$z = \frac{x - \mu}{\sigma} = \frac{174 - 172}{29} = 0.068 = 0.07$$

Check table for 0.07

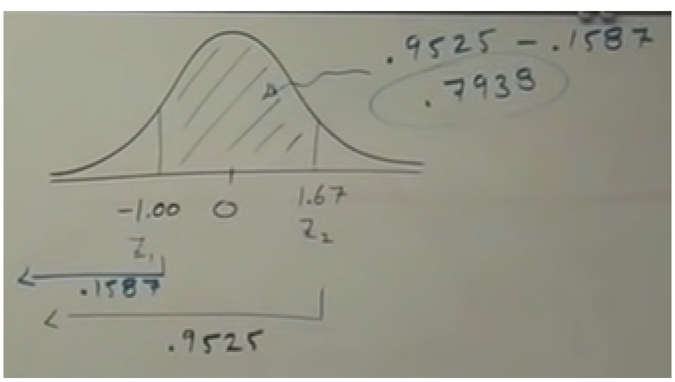


2. IQ is normally distributed with a mean of 100 and a standard deviation of 15. What percentage of people have an IQ between 85 & 125?

μ **= 100 ,** $\sigma = 15$ **, x1 = 85 , x2 = 125**

$$z1 = \frac{x - \mu}{\sigma} = \frac{85 - 100}{15} = -1.00$$

$$z2 = \frac{x - \mu}{\sigma} = \frac{125 - 100}{15} = 1.67$$



79.38% of people have an UQ between 85 and 125.

Key Points:

1. **Z score is not a distance.**
2. **Area we are finding is a probability (Z score can be negative but probability cannot).**
3. **Area cannot be negative.**

We can find data values from z score:

$z = \frac{x - \mu}{\sigma}$ **x is the data value**

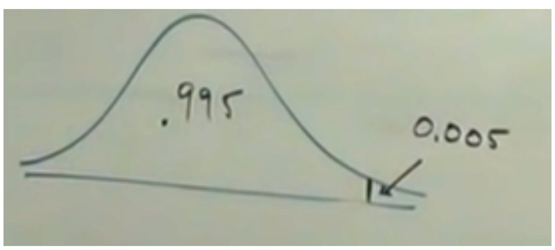
$z * \sigma = x - \mu$

$\mu + \sigma * z = x$

$x = \mu + \sigma * z$

Eg:

1. A population of men has a mean weight of 172 lbs and standard deviation of 29 lbs, find the weight that separates the lightest 99.5% from the heaviest 0.5%?



z = 2.575

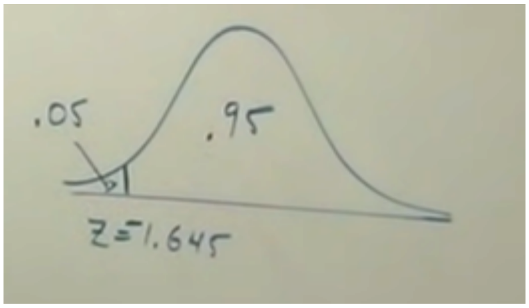
$x = \mu + \sigma * z$

x = 172 + 29 * 2.575

x = 246.68 lbs

246.68 lbs separates the lightest 99.5% from the heaviest 0.5%

2. Grip reach for women is normally distributed and the mean grip reach is 27.0 inches with a standard deviation of 1.3 inches, Find the grip reach that represent the longest 95% of women.

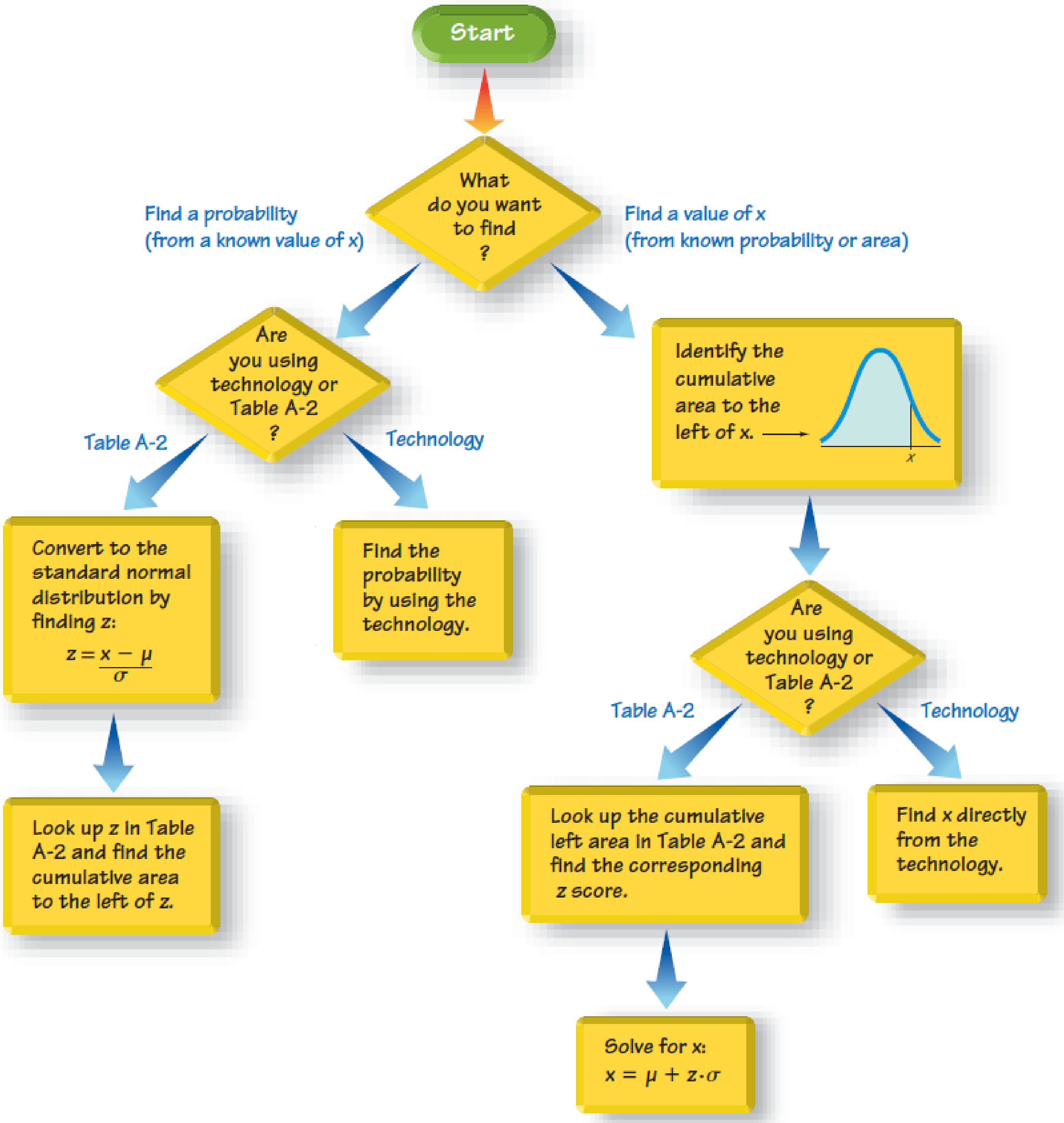


$x = \mu + \sigma * z$

x = 27 + 1.3 * -1.645

x = 24.86 inches

Applications with Normal Distributions



(**Prepared by** - Pradeepchandra Reddy S C - LinkedIn - https://www.linkedin.com/in/pradeepchandra-reddy-s-c/)

Sampling Distributions:

The sampling distribution of a statistic (such as a sample mean or sample proportion) is the distribution of all values of the statistic when all possible samples of the same size n are taken from the same population. (The sampling distribution of a statistic is typically represented as a probability distribution in the format of a table, probability histogram, or formula.)

Eg:

1. How many samples of size "n" are possible out of a population of size "N".

$N\ C\ n$

If there were 27 people in a class room we have to select in a 5 group then it will be 27 C 5

27 C 5 = 80730.

Sampling distribution of the mean:

The sampling distribution of the mean is the distribution of sample means, with all samples having the same sample size n taken from the same population. (The sampling distribution of the mean is typically represented as a probability distribution in the format of a table, probability histogram, or formula.)

Sampling distribution of the variance:

The sampling distribution of the variance is the distribution of sample variances, with all samples having the same sample size n taken from the same population. (The sampling distribution of the variance is typically represented as a probability distribution in the format of a table, probability histogram, or formula.)

Sampling distribution of the proportion:

The sampling distribution of the proportion is the distribution of sample proportions, with all samples having the same sample size n taken from the same population.

If you organise all of those statistics of each different sample into a table, this is a sampling distribution.

Eg:

1. Assume there are only numbers 1, 2, 5. Find the sample distribution of the proportion.

Proportion of ODD numbers (Population Proportion (p)): p = ⅔

Take all the possible samples of size 2:

N = 3
n = 2

3C2 = 3 (3 unique combinations)

Sample Proportion of ODD numbers (Sample Proportion ($\widehat{p}$)): $\widehat{p}$ =

Sample , n = 2	$\widehat{p}$	**Probability**	$\widehat{p} * p(\widehat{p})$
1,1	**1.0**	**1/9**	**1/9**
1,2	**0.5**	**1/9**	**1/18**
1,5	**1.0**	**1/9**	**1/9**
2,1	**0.5**	**1/9**	**1/18**
2,2	**0**	**1/9**	**0**
2,5	**0.5**	**1/9**	**1/18**
5,1	**1.0**	**1/9**	**1/9**
5,2	**0.5**	**1/9**	**1/18**
5,5	**1.0**	**1/9**	**1/9**

$\mu = \Sigma [\widehat{p} * p(\widehat{p})]$

μ = **6/9 = 2/3**

Sample proportion targets (estimates) population proportion.

2. Assume there are only numbers 1, 2, 5. Find the sample distribution of the mean.

Mean of the population:

ω = 8/3

Sample , n = 2	$\overline{x}$ **(Sample)**	**Probabilit y**	$\overline{x} * p(\overline{x})$
1,1	**1.0**	**1/9**	**1/9 = 2/18**
1,2	**1.5**	**1/9**	**3/18**
1,5	**3.0**	**1/9**	**3/9 = 6/18**
2,1	**1.5**	**1/9**	**3/18**
2,2	**2**	**1/9**	**2/9 = 4/18**
2,5	**3.5**	**1/9**	**7/18**
5,1	**3**	**1/9**	**1/9 = 2/18**
5,2	**3.5**	**1/9**	**7/18**
5,5	**5**	**1/9**	**5/9 = 10/18**

$\mu = \Sigma[\overline{x} * p(\overline{x})]$

μ = 48/18 = 8/3

Sample mean targets (estimates) population mean.
SD is not a great estimate for Population SD

The Central Limit Theorem:

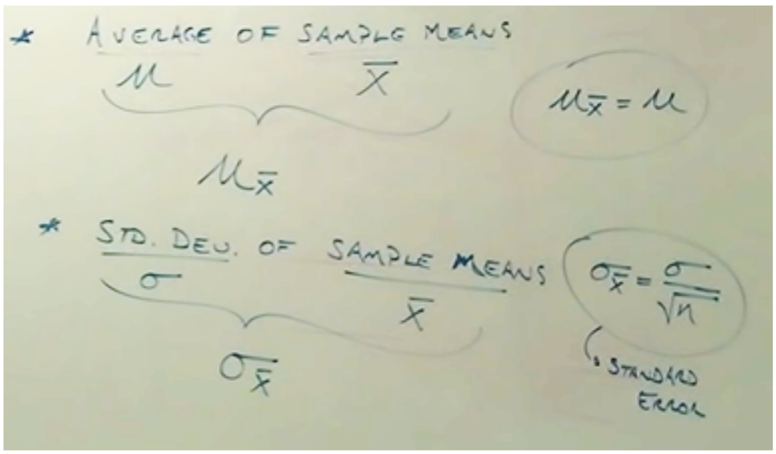
The central limit theorem states that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement , then the distribution of the sample means will be approximately normally distributed. When selecting a simple random sample of n subjects from a population with mean μ and standard deviation σ it is essential to know these principles:

n = size
N = population

1. For a population with any distribution, if n > 30 , then the sample means have a distribution that can be approximated by a normal distribution with mean μ **and standard deviation** $\sigma/\sqrt{n}$.

2. If $n \leq 30$ **and the original population has a normal distribution, then the sample means have a normal distribution with mean** μ **and standard deviation** $\sigma/\sqrt{n}$.

3. If and the original population does not have a normal distribution, then the methods of this section do not apply.



The Central Limit Theorem and the Sampling Distribution of $\bar{x}$

Given

1. The random variable x has a distribution (which may or may not be normal) with mean μ and standard deviation σ.
2. Simple random samples all of the same size n are selected from the population. (The samples are selected so that all possible samples of size n have the same chance of being selected.)

Conclusions

1. The distribution of sample means $\bar{x}$ will, as the sample size increases, approach a *normal* distribution.
2. The mean of all sample means is the population mean μ.
3. The standard deviation of all sample means is $\sigma/\sqrt{n}$.

Practical Rules Commonly Used

1. If the original population is *not normally distributed,* here is a common guideline: For $n > 30$, the distribution of the sample means can be approximated reasonably well by a normal distribution. (There are exceptions, such as populations with very nonnormal distributions requiring sample sizes larger than 30, but such exceptions are relatively rare.) The distribution of sample means gets closer to a normal distribution as the sample size n becomes larger.
2. If the original population is *normally distributed,* then for *any* sample size n, the sample means will be normally distributed.

Before : To find Z for an Individual Value

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

Now: To find Z for an average of Group of Values a sample.

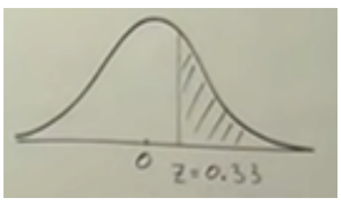
$$z = \frac{\bar{x} - \mu\bar{x}}{\sigma\bar{x}}$$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Eg:

1. A population of men have a mean of 172 lbs and standard deviation of 10 lbs. Find probability of randomly selected Man will weigh more than 175 lbs.

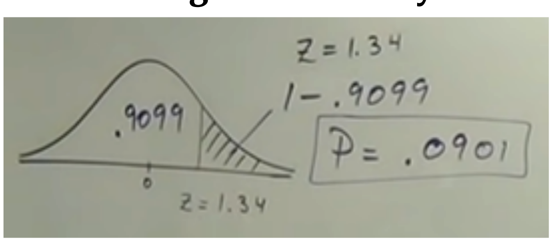
$$z = \frac{x - \mu}{\sigma} = \frac{175 - 172}{10} = 0.3$$



= 1 - 0.61791 = 0.38209

The probability of randomly selecting a man will weigh more than 175 lbs is 38.20%.

2. A population of men have a mean of 172 lbs and standard deviation of 10 lbs. Find probability that a group of 20 men will have an average weight of more than 175 lbs. Assume weight is normally distributed.



$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{175 - 172}{10/\sqrt{20}} = \frac{3}{10/4.472} = \frac{3}{2.2364} = 1.34$$

The probability of randomly selecting a group of men whose average weight is more than 175 pounds is approximately 9.01%

Introduction to Hypothesis Testing:

Rare Event Rule for Inferential Statistics:

If, under a given assumption, the probability of a particular observed event is exceptionally small (such as less than 0.05), we conclude that the assumption is probably not correct.

Eg:

1. Body temp has an average of 98.6 degrees and a standard deviation of 0.62 degrees. A sample of 106 people were selected. What's the probability that the average temp for the sample will be 98.2 degrees or lower

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{98.2 - 98.6}{0.62/\sqrt{106}} = \frac{-0.4}{0.62/10.2956} = \frac{-0.4}{0.06022} =- 6.64$$

From table = 0.000000000016

Ch. 6: Normal Distribution

$$z = \frac{x - \bar{x}}{s} \text{ or } \frac{x - \mu}{\sigma} \quad \text{Standard score}$$

$$\mu_{\bar{x}} = \mu \quad \text{Central limit theorem}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \text{Central limit theorem (Standard error)}$$