NORMAL DISTRIBUTIONS

a symmetrical, bell-shaped histogram with a number of significant statistical properties

Marlon's class has been challenged to guess their teacher's height in centimetres. Listed below are the estimates, submitted anonymously. Calculate the mean and standard deviation, and then create a histogram.

~~183~~ ~~183~~ ~~174~~ ~~212~~ ~~178~~ ~~207~~ ~~186~~ ~~178~~ ~~204~~ ~~172~~ ~~189~~ ~~183~~
~~184~~ ~~190~~ ~~184~~ ~~168~~ ~~190~~ ~~180~~ ~~183~~ ~~190~~ ~~185~~ ~~162~~ ~~200~~
~~206~~ ~~196~~ ~~187~~ ~~204~~ ~~185~~ ~~206~~ ~~175~~

Class Interval	Frequency
[160 – 168)	1
[168 – 176)	4
[176 – 184)	7
[184 – 192)	10
[192 – 200)	1
[200 – 208)	6
[208 – 216]	1

[inclusive

) exclusive

$\frac{\text{Range}}{\#\text{ of intervals}} = \frac{212-162}{7}$

$\div 7.14$

(at least 7.14)

~ 8

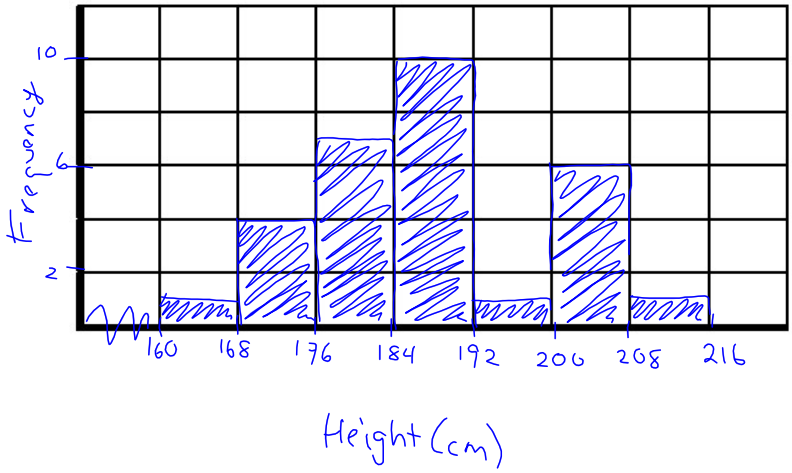
Excel

$\bar{x}$ =AVERAGE(...data...)

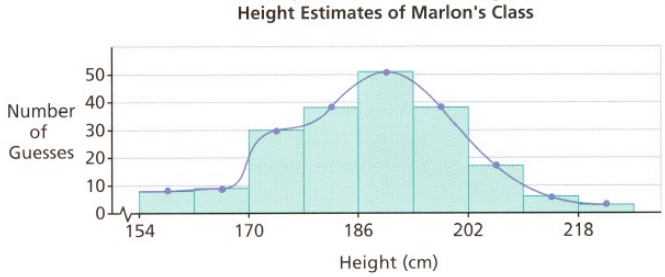
$\div 187.5$

σ =STDEV.P(...data...)

$\div 12.15$



Marlon decides to collect 120 more samples and the histogram becomes the following:



If you draw a smooth curve close to or through the tops of the rectangles in the histogram, you get a normal curve, like the one shown below.

Given enough data and small enough intervals, Marlon would eventually get a perfectly symmetrical bell-shaped curve. A distribution with a histogram that follows a normal curve is called a normal distribution.

CHARACTERISTICS OF NORMAL DISTRIBUTION

A normal distribution has the following properties:
• It is symmetrical; the mean, median, and mode are equal and fall at the line of symmetry for the curve.
• It is shaped like a bell, peaking in the middle and sloping down toward the sides.
• Approximately 68% of the data is within one standard deviation of the mean.
• Approximately 95% of the data is within two standard deviations of the mean.
• Approximately 99.7% of the data is within three standard deviations of the mean.

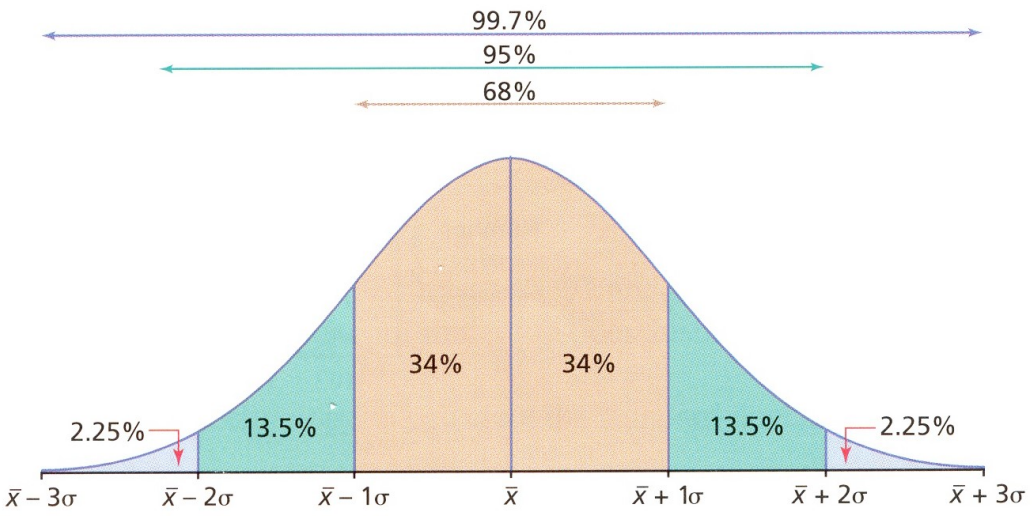
The notation used to describe a normal distribution is $X \sim N(x, \sigma^2)$, where x is the mean and σ^2 is the variance (the square of the standard deviation).

Example

Suppose a population's heights are normally distributed with a mean height of **170 cm** and a standard deviation of **5 cm**. The notation for this distribution would be:

$$X \sim N(170, 25)$$

- X is the random variable representing height.
- The mean $x = 170$ cm.
- The variance $s^2 = 5^2 = 25$ cm² (since the standard deviation is 5 cm).



The graph of the normal distribution $X \sim N(\bar{x}, \sigma^2)$

Examples

Julie is an engineer who designs roller coasters. She wants to develop a ride that 95% of the population can ride. The average adult in North America has a mass of 71.8 kg, with a standard deviation of 13.6 kg.

(a) What range of masses should she be prepared to anticipate?

In a normal distribution, 95% of the data is within 2σ of $\bar{x}$

$\bar{x} = 71.8$

$\sigma = 13.6$

$\bar{x} + 2\sigma = 71.8 + 2(13.6)$
$= 99$

$\bar{x} - 2\sigma = 71.8 - 2(13.6)$
$= 44.6$

∴She should prepare for 44.6kg - 99 kg

(b) If she wanted to provide for 99.7% of the general population, what range of masses should she be prepared to anticipate?

Within 3σ

$\bar{x} + 3\sigma = 71.8 + 3(13.6)$
$= 112.6$

$\bar{x} - 3\sigma = 71.8 - 3(13.6)$
$= 31$

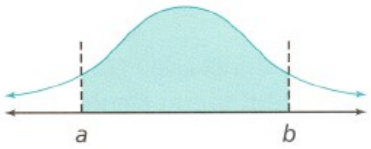
∴She should prepare for 31 kg - 112.6 kg

(c) What assumption is Julie making in this example that could cause problems?

She is assuming this is a normal distribution!
(likely bi-modal, due to two sexes)

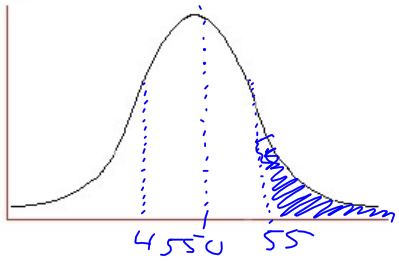
She should calculate the two data sets separately.

The area under every normal curve equals 1. As a result, in any normal distribution, the percent of the data that lies between two specific values, a and b, is the area under the normal curve between endpoints a and b.

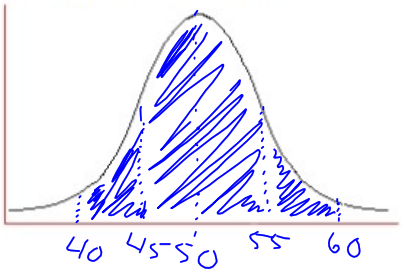


If $X \sim (50, 5^2)$, draw a diagram that represents the percent of data that have these values for X:

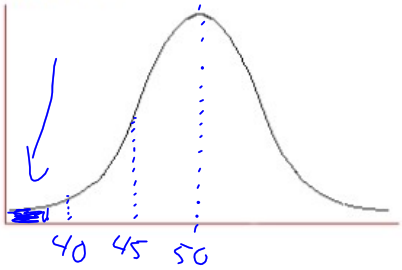
(a) $x > 55$



(b) $40 < x < 60$



(c) $x < 38$



Practice: #1, 3b, 6, 8, 9, 14