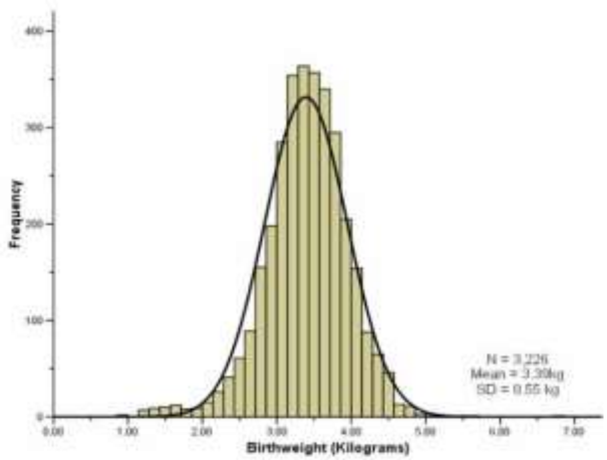
8.2
The Normal distribution curve

For the most part, the data is distributed symmetrically (and uni-modular) about the mean (like peoples height or weight or IQ)

This is called the Bell Curve and it looks like this.
It is also known as Normal distribution



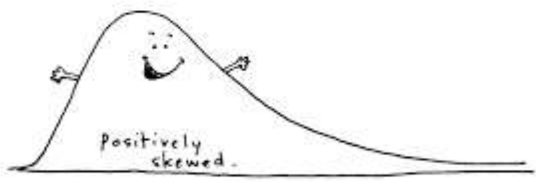
www.mathsisfun.com/data/quincunx.html

Basic Properties of the Standard Normal Curve:

Property 1: The total area under the standard normal curve is 1.
Property 2: The standard normal curve extends indefinitely in both directions, approaching, but never touching, the horizontal axis as it does so.
Property 3: The standard normal curve is symmetric about 0.
Property 4: Almost all the area under the standard normal curve lies between $z=-3$ and 3.

This data in not normal

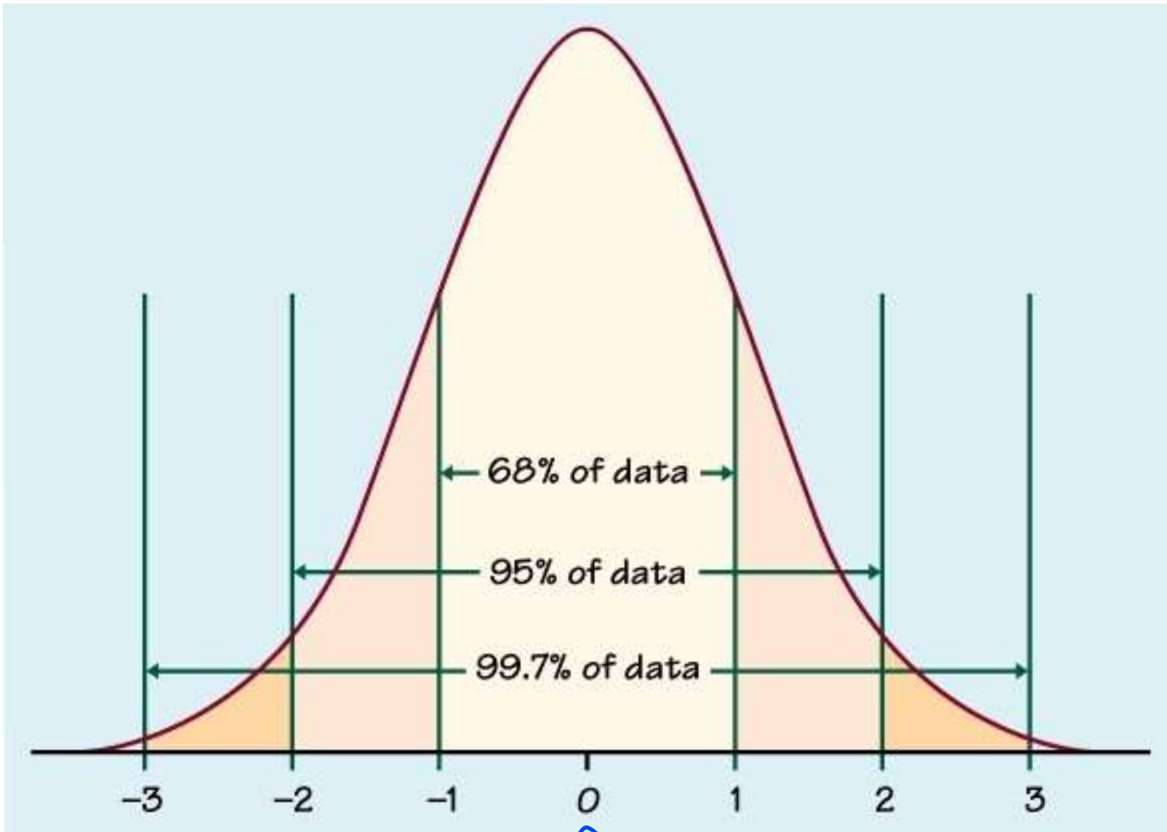
mean left of median



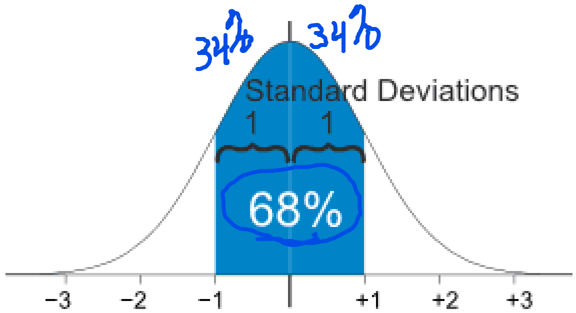
mean right of median

All normal curves can be used to calculate the probabilities for a range of outcomes
A population that has normal distribution can be totally described by knowing its MEAN (μ) and its standard deviation (σ)

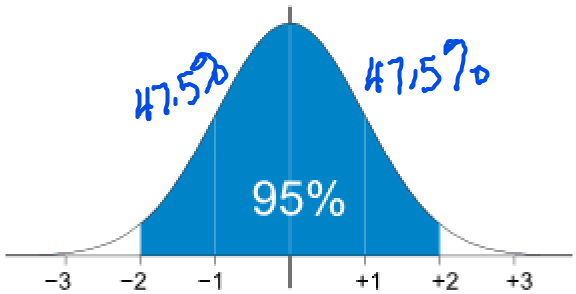
For symmetrical and uni-modular distribution the mode=mean, also the larger σ the more spread out the bell shape.



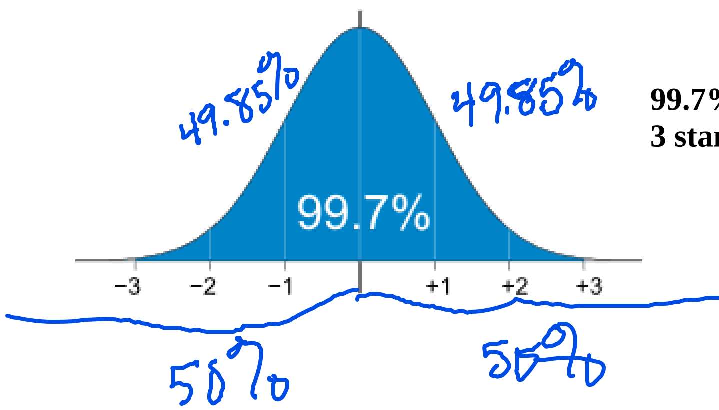
mean



68% of values are within
1 standard deviation of the mean



95% of values are within
2 standard deviations of the mean



99.7% of values are within
3 standard deviations of the mean

In general:

Probability of a range of x that is within 1 standard deviation from the mean:

$P(\mu - \sigma < x < \mu + \sigma)$ = about 68%

Probability of a range of x that is within 2 standard deviation from the mean:

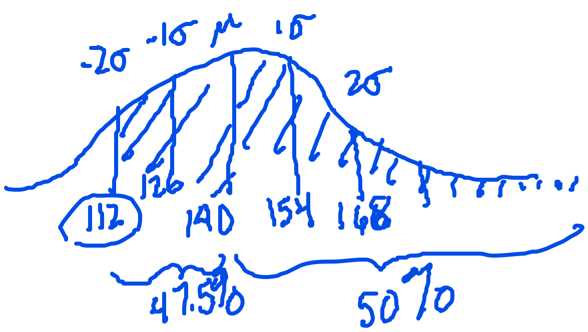
$P(\mu - 2\sigma < x < \mu + 2\sigma)$ = about 95%

Probability of a range of x that is within 3 standard deviation from the mean:

$P(\mu - 3\sigma < x < \mu + 3\sigma)$ = about 99.7%

Also

1. Rugby player IQ players has normal distribution. Josh has an IQ of 112 but the normal distribution curve has an IQ of 140 and a standard deviation of 14 points. What is the probability that Josh meets a Rugby player that is smarter than he is?

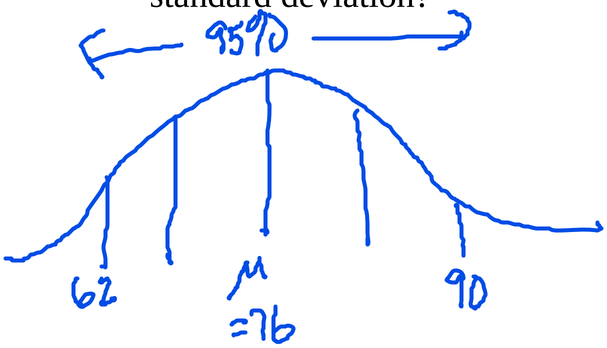


∴ 97.5% are above Josh

∴ smarter than Josh.

2. 95% of students at school weigh between 62 kg and 90 kg. Assuming this data is normally distributed, what are the mean and standard deviation?

from 2σ below to 2σ above the mean



mean is halfway

$$\therefore \frac{62+90}{2} = 76$$

$$76 - 2\sigma = 62$$
$$76 - 62 = 2\sigma$$
$$14 = 2\sigma$$
$$7 = \sigma$$

3. A machine produces electrical components. 99.7% of the components have lengths between 1.176 cm and 1.224 cm. Assuming this data is normally distributed, what are the mean and standard deviation?

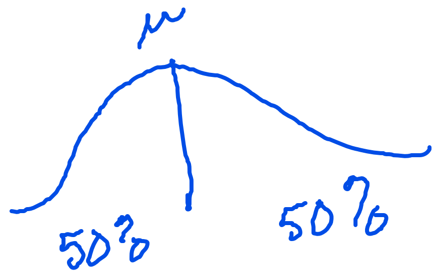
3σ above $\therefore 3\sigma$ below mean

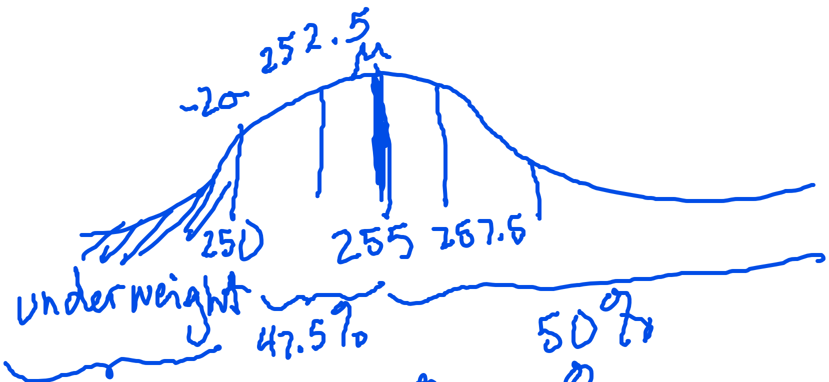
$$\mu = \frac{1.176 + 1.224}{2}$$
$$= \frac{2.4}{2}$$
$$= 1.2 \text{ cm}$$

$$1.2 + 3\sigma = 1.224$$
$$3\sigma = 1.224 - 1.2$$
$$3\sigma = 0.024$$
$$\sigma = .008$$

4. The Fresha Tea Company pack tea in bags marked as 250 g
A large number of packs of tea were weighed and the mean and standard deviation were calculated as 255 g and 2.5 g respectively.
Assuming this data is normally distributed, what percentage of packs are underweight?

$\mu = 255g \quad \sigma \; 2.5g$





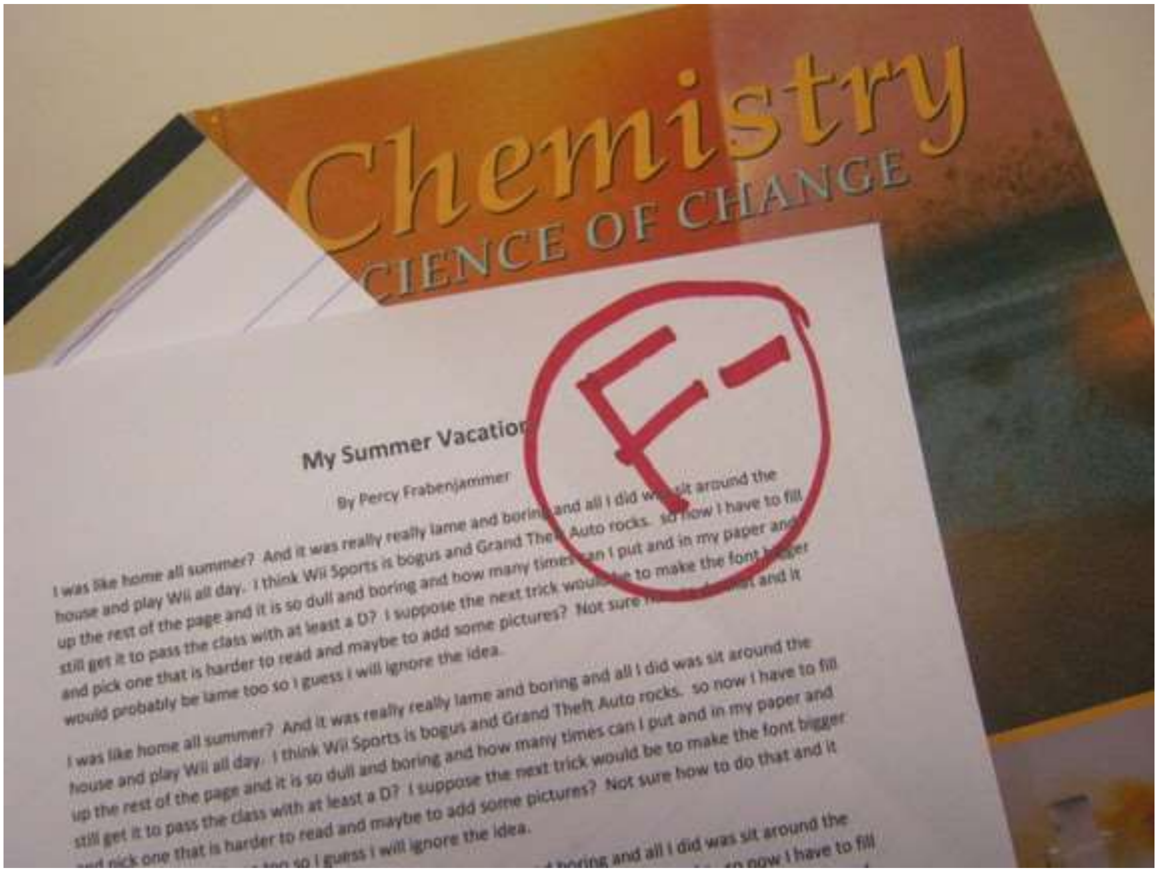
$100\% - 47.5\% - 50\%$

$= 2.5\%$ is probably underweight.

OR top half ⇒ 50%
bottom 2σ ⇒ 47.5%
97.5% is OK

∴ underweight = 100% − 97.5%
= 2.5%

WHAT??? How did that happen???



So how DID that happen?

You get an F- on a paper and the teacher stands up in class and announces, "I grade on the curve - deal with it."

What does it mean to "grade on the curve"?

Do you remember standard deviation? the empirical rule? the standard normal distribution?

The standard deviation is a measure of how closely grouped or how widely spaced a set of data appears. The empirical rule says in a standard normal distribution, 68% of the data points will fall within ± one standard deviation from the mean and 95% will fall within ± two standard deviations.

A standard normal distribution has a mean of 0 and a standard deviation of 1. Got it?

The normal distribution is important because lots of variables studied in education and psychology are normally distributed. Interesting stuff like reading ability, job satisfaction and memory, to name a few. Knowing data is normally distributed means all sorts of nifty statistical tests can be invented. Even better, the tests work pretty well even if the distributions are only **approximately** normally distributed (meaning sort of close to normal but maybe off a little).

"60% of children learn to read by age 7" might be a result of a statistical test...

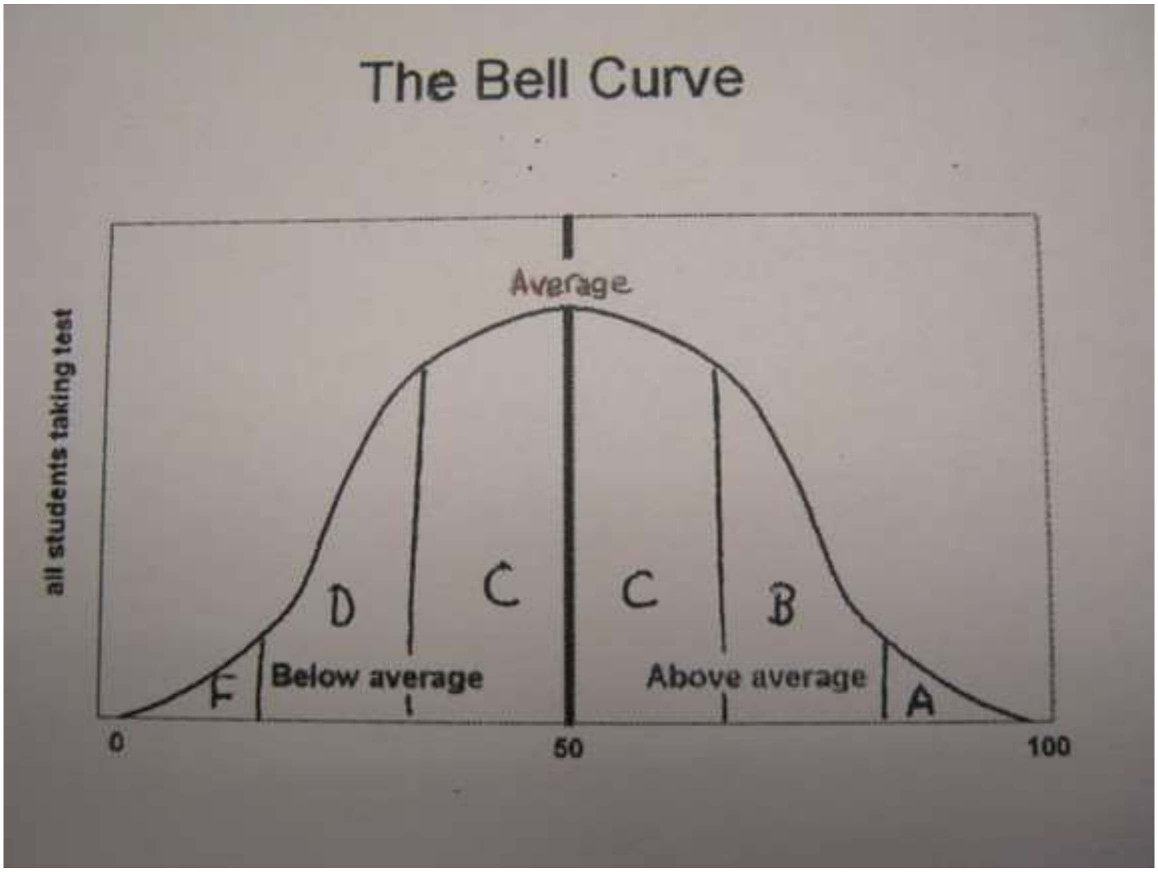
Back to grading on the curve

Below is an example of a bell shaped curve.

Let's say a teacher gives a test to a class of students and the mean (or average) score is **80** and the standard deviation is **5**.

According to the empirical rule, **68%** of the students should fall within ± one standard deviation of the mean. If you look at the curve, those students will get **C's** (C is meant to show "average performance".

Move out to two standard deviations away from the mean, and you have the **B's** to the right and the **D's** to the left. Move out to three standard deviations from the mean and you have the **A's** to the right and the **F's** to the left.



You can set it up as a grading scale:

100 - 91 = A

90 - 85 = B

84 - 76 = C

75 - 70 = D

69 - 00 = F

The key to grading on the curve means you will always have at least one **A** and one **F** and the majority of a class will be **C's**.

And the issue of someone "blowing the curve"? It means a student gets such a high score that the rest of the curve will be skewed to the right, meaning the range for **C's** and **D's** will be much higher than the usual 60 - 70%.

Are you getting irritated?

Say you are in a class of 17 students and the mean and standard deviation follow the curve above.

17 x 68% = 11.56 or 12 students will get a C and B's, D's, A's and F's will be spread among 5 students. Depending on the instructor's preferences, usually it works out to one F, one A, two B's and one D. **Not much chance of improving your gradepoint average!**