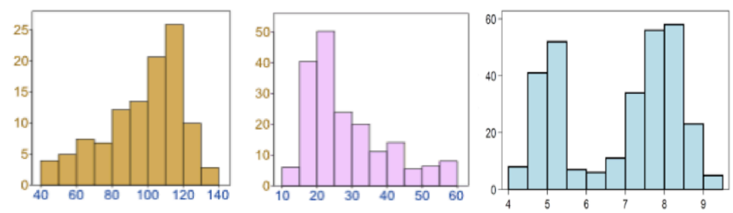
THE ERINDALE ACADEMY
1576 Dundas Street West, Mississauga ON L5C 1E5
www.teacademy.ca Tel: 905-232-1576
Email: info@teacademy.ca

Teacher: Ella

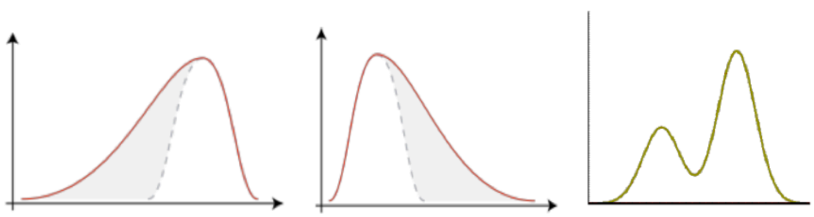
Unit 7: Normal distribution

Lesson 7.1: Normal distribution and application

Data can be “distributed” (spread out) in different ways. It can be spread out more on the left (Graph A), or more on the right (Graph B), or it can be all jumbled up (Graph C).



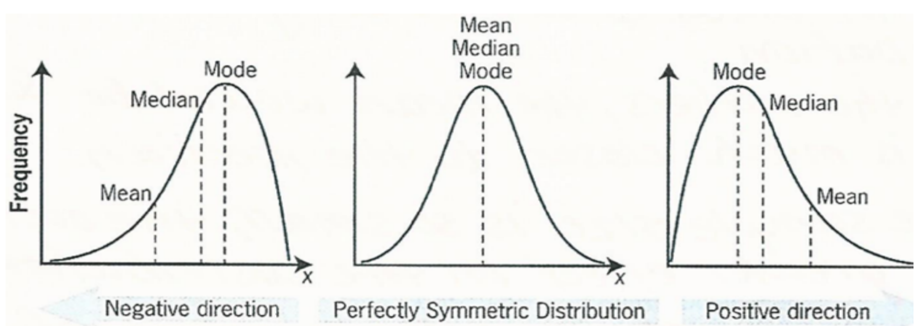
A **B** **C**

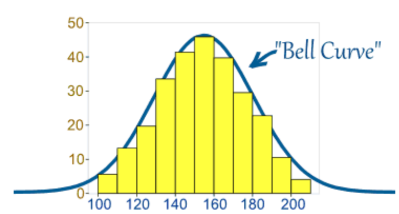


A: It has a “tail” going to the left with one mode on the right (most of the data falls to the right), called negative skew or left skew distribution.

B: it has a “tail” going to the right with one mode on the left (most of the data falls to the left), called positive skew or right skew distribution.

C: It has more than one mode, would not be not normal distribution





We say the data is “normally distributed” when: 1) Mean = median = mode;

2) Symmetrical about the center;

3) 50% of values less than the mean and 50% greater.

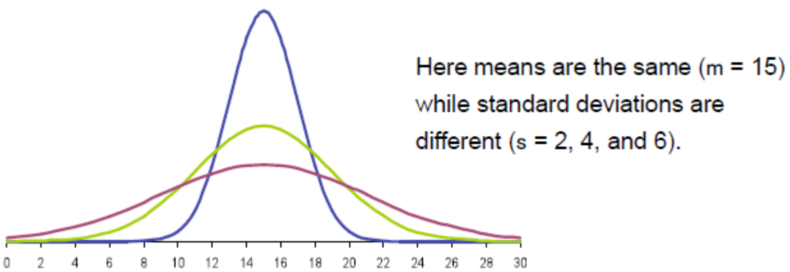
Many things closely follow a **normal distribution** (as long as the sample size is big enough):

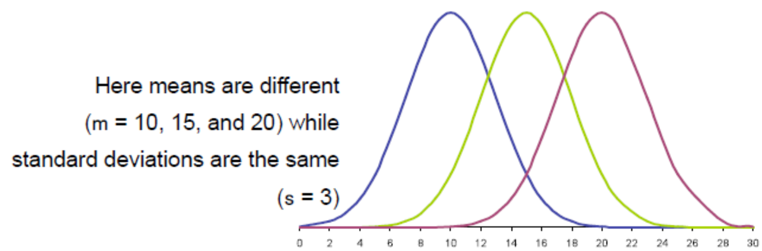
Heights of people; Blood pressure; Marks on a test.

Teacher: Ella

Thus, **normal distributions** (or Gaussian) distributions are a family of symmetrical, bell shaped density curves defined by: a mean (μ) and a standard deviation (σ): **N (μ, σ)**

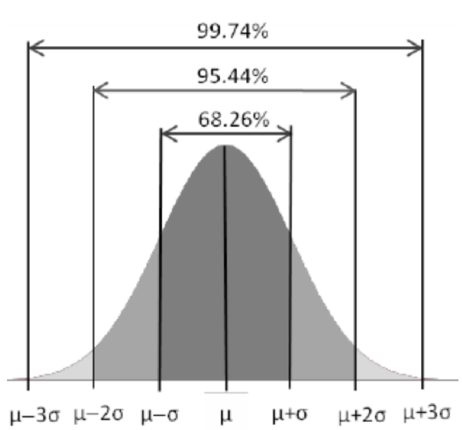
The total area under the curve, by definition, is equal to 1, or 100%. The area under the curve for a range of values is the proportion of all observations for that range.

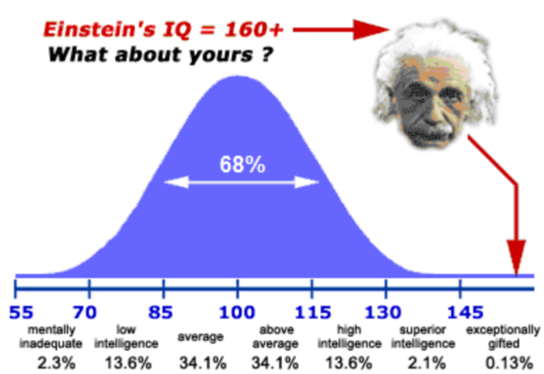




- Although every normal distribution has its own mean and standard deviation, for any normal distribution:
  i) Approx. 68% of the data will lie within 1 standard deviation of the mean ($\mu \pm \sigma$)
  ii) Approx. 95% of the data will lie within 2 standard deviation of the mean ($\mu \pm 2\sigma$)
  iii) Approx. 99.7% of the data will lie within 3 standard deviation of the mean ($\mu \pm 3\sigma$)

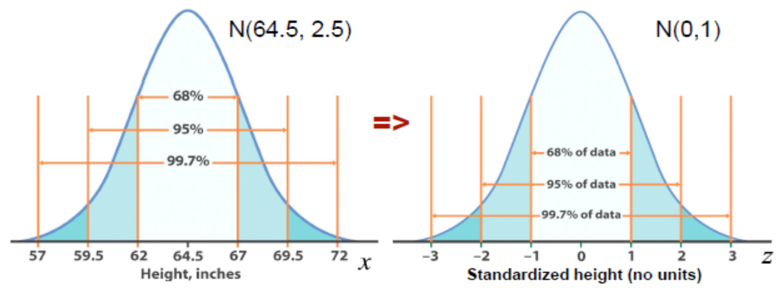
- So once we know the mean and standard deviation of a normal distribution, we can say that any value is:
  i) **Likely** to be within 1 standard deviation (68 out of 100 should be)
  ii) **Very likely** to be within 2 standard deviations (95 out of 100 should be)
  iii) **Certainly** within 3 standard deviation (997 out of 1000 should be)





Teacher: Ella

Standardize normal distribution: Because all normal distribution shares the same properties, we can standardize our data to transform any Normal Curve N (μ, σ) into the standard normal curve N (0, 1).



We use z-scores to standardize normal curve.

$$\mathbf{z} = \frac{(\boldsymbol{x} - \boldsymbol{\mu})}{\boldsymbol{\sigma}}$$

- when x is 1 standard deviation larger than the mean, then z = 1
- when x is 2 standard deviations less than the mean, then z = - 2

Using Normal distribution curve table:

Since we know the area under the curve is 1, we can calculate the probability that a given data value falls within a certain range of values.

- Step 1: calculate the z-score for a certain datum, keep 2 decimal places.
- Step 2: find the Z-score on a table. This table will give you the probability of finding a **value less than (to the left)** of your data.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616

<u>Example 1:</u> Women heights follow the N (64.5'', 2.5'') distribution. What percent of women are shorter than 67 inches tall?

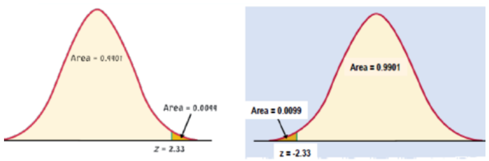
Conclusion: 84.13% of women are shorter than 67 inches. By subtraction, 15.87% of women are taller than 67''.

THE ERINDALE ACADEMY
1576 Dundas Street West, Mississauga ON L5C 1E5
www.teacademy.ca Tel: 905-232-1576
Email: info@teacademy.ca

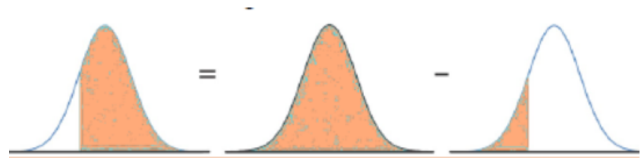
Teacher: Ella

Tips on using the table: because the normal distribution is symmetrical, there are 2 ways that you can calculate the area under the standard Normal curve to the right of a z-value.

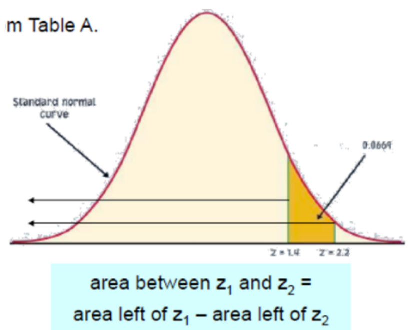
Method 1: **area right of z score = area left of – z**



Method 2: **area right of z score = 1 – area left of z**



To calculate the area between 2 z-scores, first get the area under N (0, 1) to the left for each z-score from Table. Then subtract the smaller area from the larger area.



Example 2: The national collegiate athletic association (NCAA) requires Division I athletes to score at least 820 on the combined math and verbal SAT exam to compete in their first college year. The SAT scores of 2003 were approximately normal with mean 1026 and standard deviation 209.

a) What proportion of all students would be NCAA qualifiers (SAT ≥ 820)?

b) The NCAA defined a "partial qualifier" eligible to practice and receive an athletic scholarship, but not to compete. As a combined SAT score is at least 720. What proportion of all students who take SAT would be partial qualifiers? That is, what proportion have scores between 720 and 820?

Teacher: Ella

Example 3: Gestation time in malnourished mothers

What is the effects of better maternal care on gestation time? The goal is to obtain pregnancies 240 days (8 months) or longer. Under each treatment, what percent of mothers failed to carry their babies at least 240 days?