


INNOMATICS RESEARCH LABS

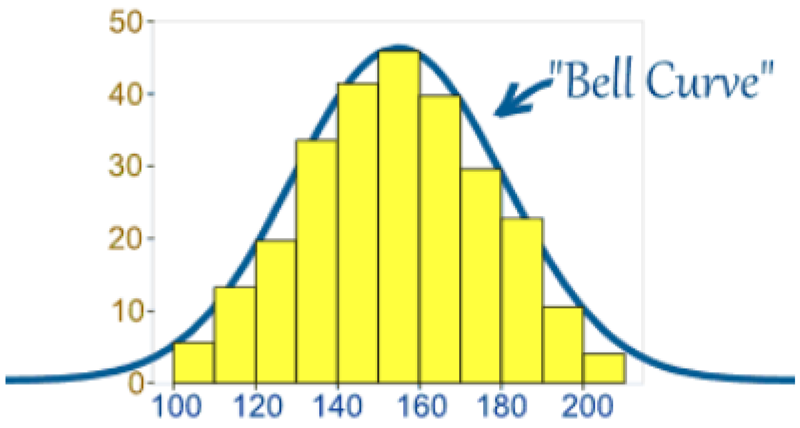


Normal Distribution

Normal Distribution

- One of the most important continuous distributions most of the real life phenomena can be modeled using a normal distribution
  - Height, weight, grades, salary ….

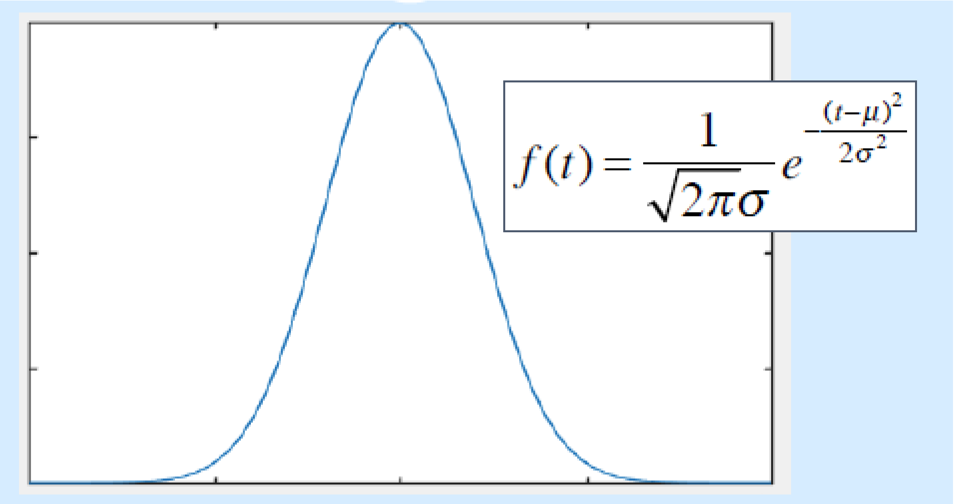
- A normal distribution can graphically be represented as a bell shaped curve



Normal Distribution Equation

- Just like we can specify a Binomial distribution by n and p, we can specify a normal distribution by $\boldsymbol{\mu}$ **and** $\boldsymbol{\sigma}$
- (μ is mean of normal random variable and σ is standard deviation)
- Formally, **X ~ N(μ ,σ2)**

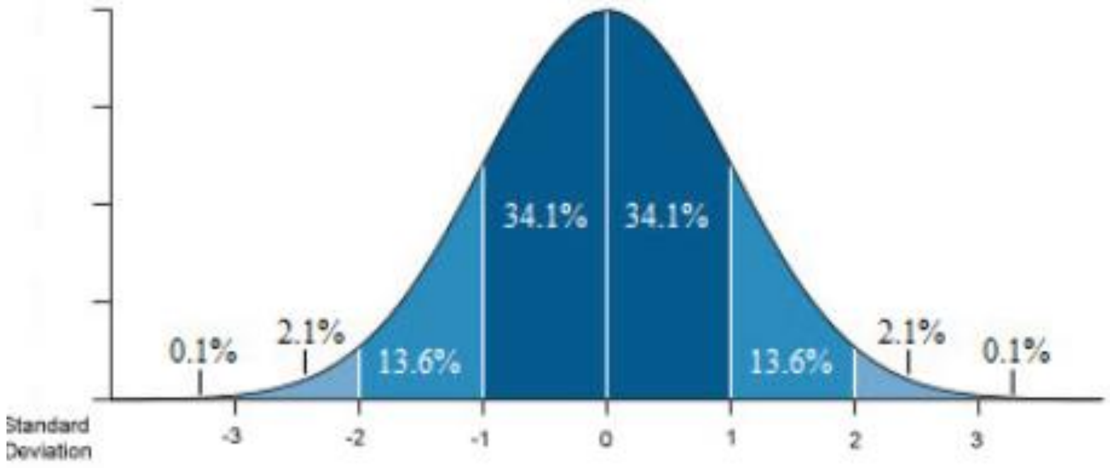
- The curve is **pdf** of a normal random variable

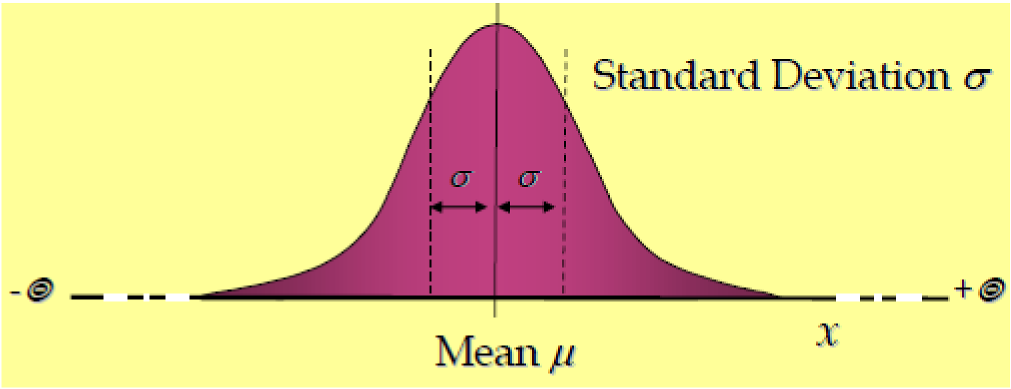


$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

Normal Distribution contd.

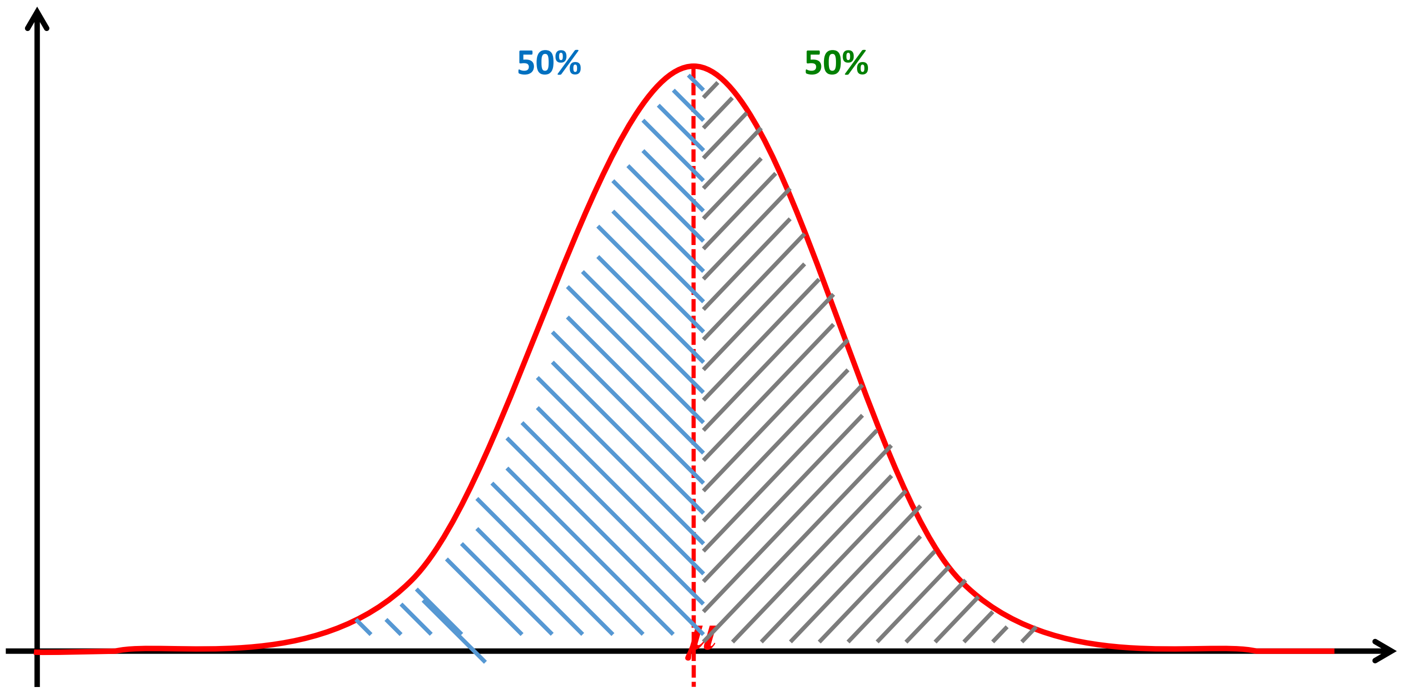
- **X ~ N(μ ,σ2)**
- A normal random variable can be completely specified by μ ,σ
  - Takes values from - ∞ to +∞
  - μ specifies location of centrality, and σ specifies width of curve
  - Symmetric around mean (mean=median=mode)
  - 68.2% area is within one s.d (σ) away from mean (μ) and 95% within 2 s.d away from mean





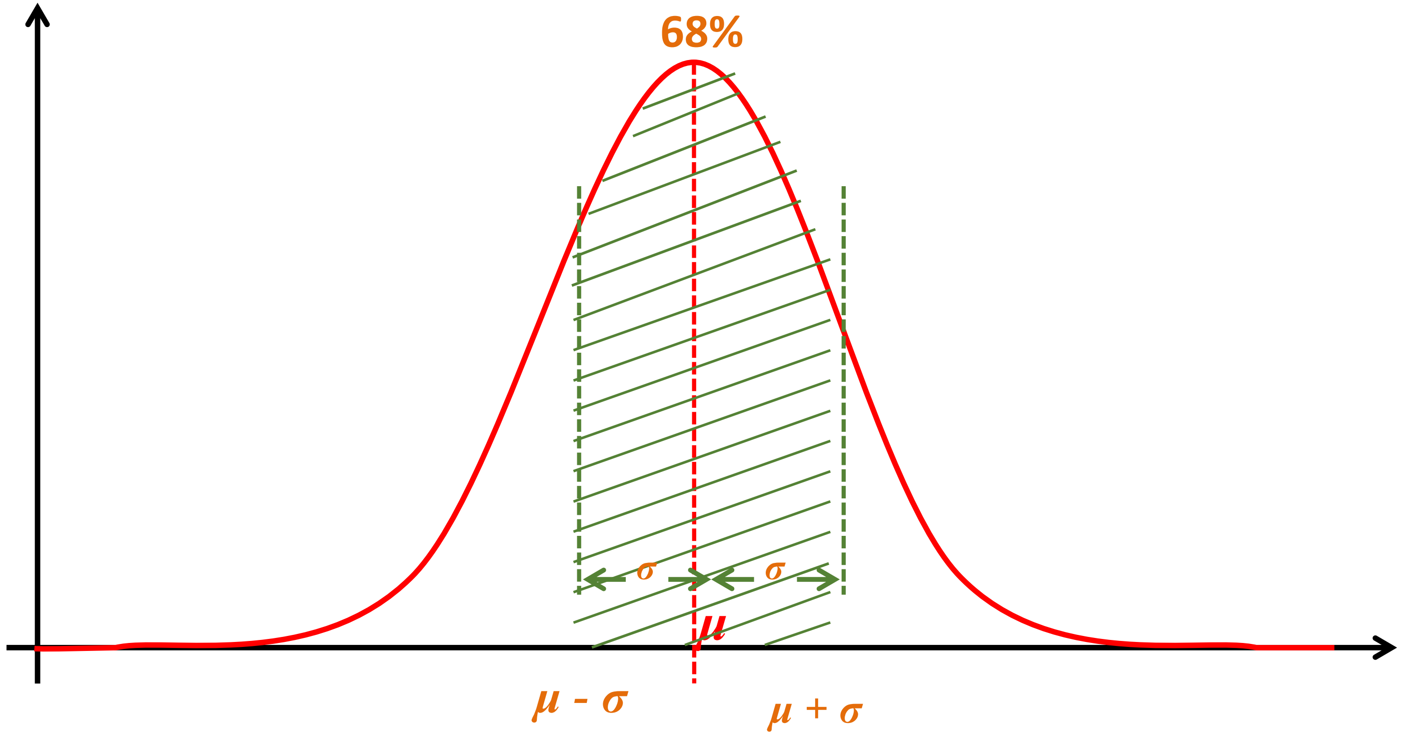
The properties of a normal distribution:

- It is a bell-shaped curve.
- It is symmetrical about the mean, μ. (The mean, the mode and the median all have the same value).
- The total area under the curve is 1 (or 100%).
- 50% of the area is to the left of the mean, and 50% to the right.



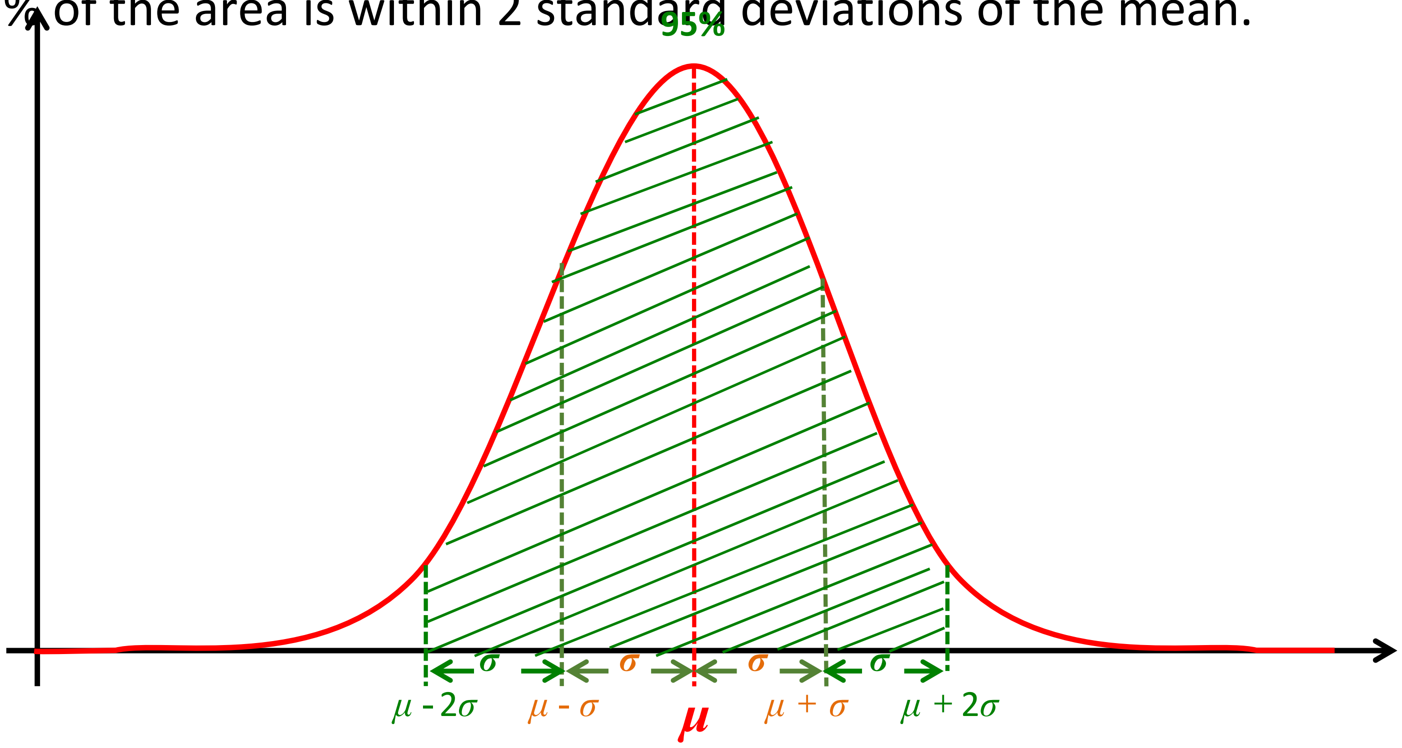
The properties of a normal distribution:

- It is a bell-shaped curve.
- It is symmetrical about the mean, μ. (The mean, the mode and the median all have the same value).
- The total area under the curve is 1 (or 100%).
- 50% of the area is to the left of the mean, and 50% to the right.
- Approximately 68% of the area is within 1 standard deviation, σ, of the mean.



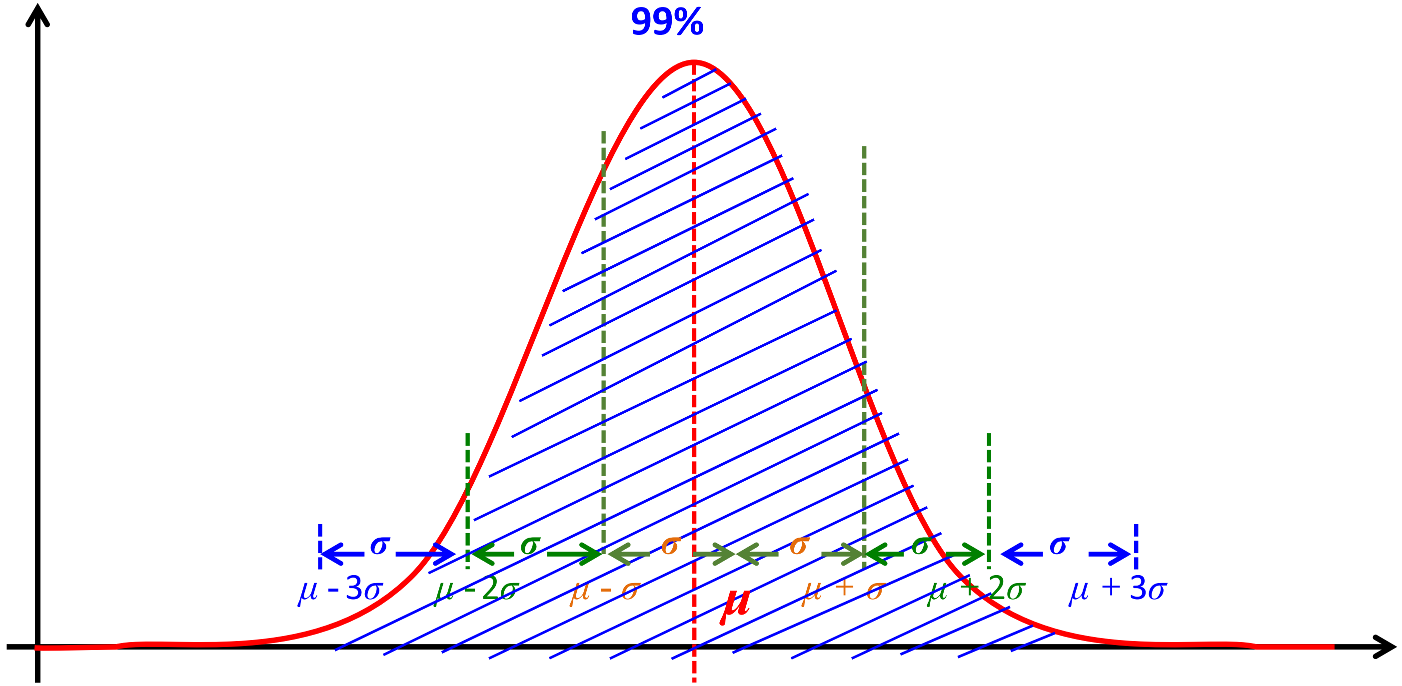
The properties of a normal distribution:

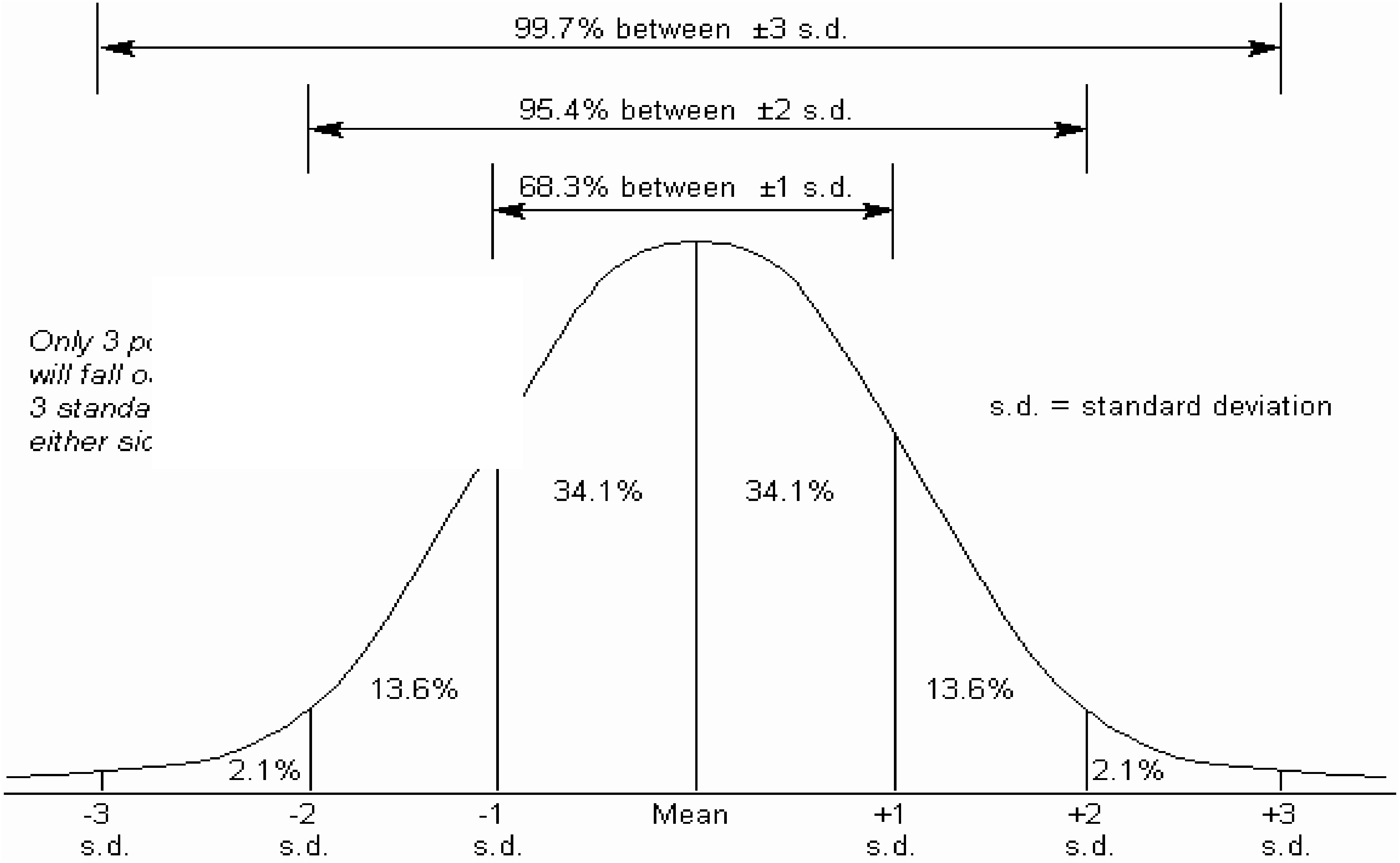
- It is a bell-shaped curve.
- It is symmetrical about the mean, μ. (The mean, the mode and the median all have the same value).
- The total area under the curve is 1 (or 100%).
- 50% of the area is to the left of the mean, and 50% to the right.
- Approximately 68% of the area is within 1 standard deviation, σ, of the mean.
- Approximately 95% of the area is within 2 standard deviations of the mean.



The properties of a normal distribution:

- It is a bell-shaped curve.
- It is symmetrical about the mean, μ. (The mean, the mode and the median all have the same value).
- The total area under the curve is 1 (or 100%).
- 50% of the area is to the left of the mean, and 50% to the right.
- Approximately 68% of the area is within 1 standard deviation, σ, of the mean.
- Approximately 95% of the area is within 2 standard deviations of the mean.
- Approximately 99% of the area is within 3 standard deviations of the mean.



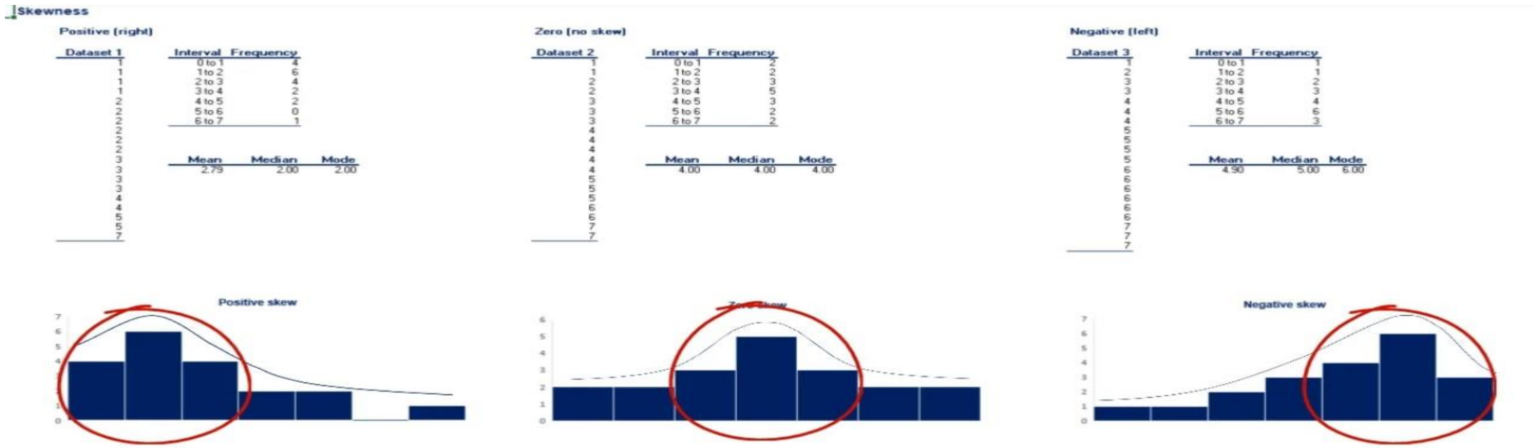

99.7% between ±3 s.d.
95.4% between ±2 s.d.
68.3% between ±1 s.d.
Only 3 pc
will fall o.
3 standa
either sic
s.d. = standard deviation
34.1%
34.1%
13.6%
13.6%
2.1%
2.1%
-3 s.d.
-2 s.d.
-1 s.d.
Mean
+1 s.d.
+2 s.d.
+3 s.d.

Skewness:

- Skewness basically gives the shape of normal distribution of values.
- Skewness is asymmetry in a statistical distribution, in which the curve appears distorted or skewed either to the left or to the right.
- Skewness can be quantified to define the extent to which a distribution differs from a normal distribution.

Skewness:

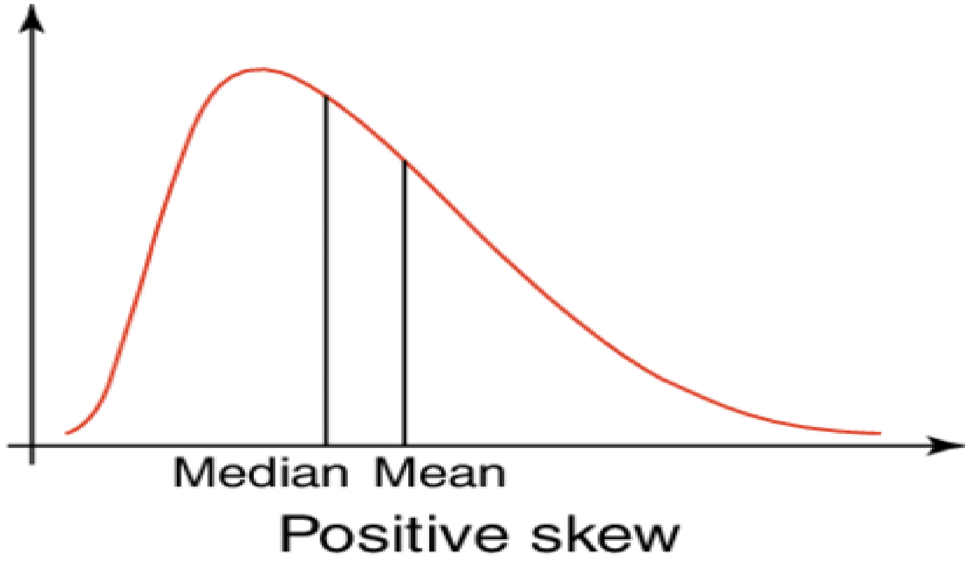
- Skewness tells us a lot about where the data is situated.



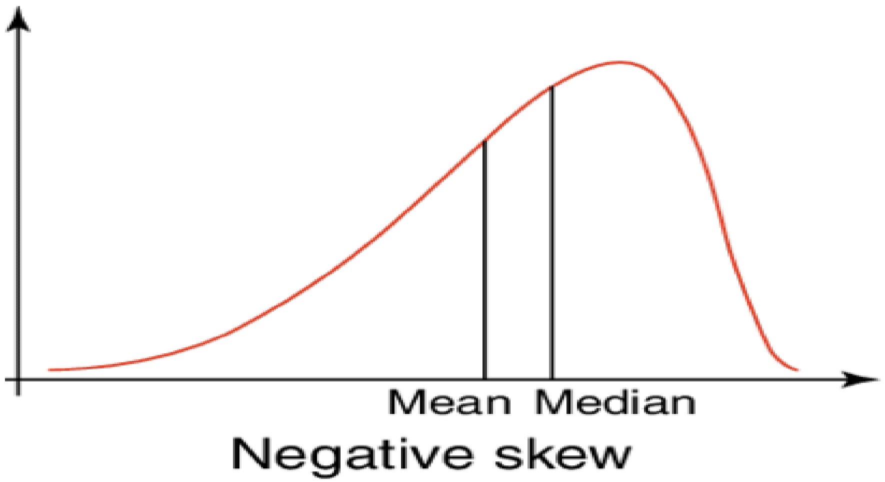
- In fact, the mean, median and mode should be used together to get a good understanding of the dataset.
- Measures of asymmetry like skewness are the link between central tendency measures and probability theory.
- This ultimately allows us to get a more complete understanding of the data we are working with.

Positive and Negative Skewness:

A positively skewed distribution means that the extreme data results are larger. This skews the data in that it brings the mean (average) up. The mean will be larger than the median in a Positively skewed distribution.

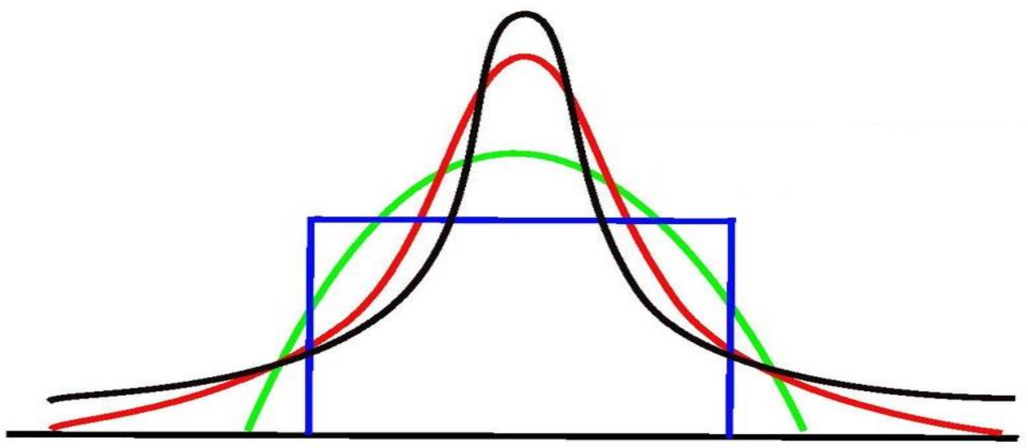


A negatively skewed distribution means the opposite: that the extreme data results are smaller. This means that the mean is brought down, and the median is larger than the mean in a negatively skewed distribution.



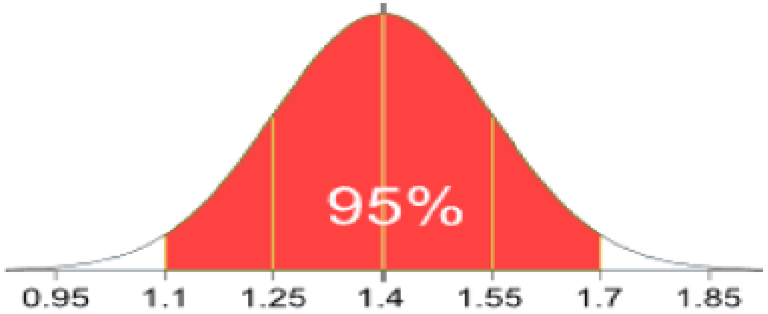
Kurtosis:

The exact interpretation of the measure of Kurtosis used to be disputed, but is now settled. It's about existence of outliers. Kurtosis is a measure of whether the data are heavy-tailed (profusion of outliers) or light-tailed (lack of outliers) relative to a normal distribution.



Normal Distribution: Example

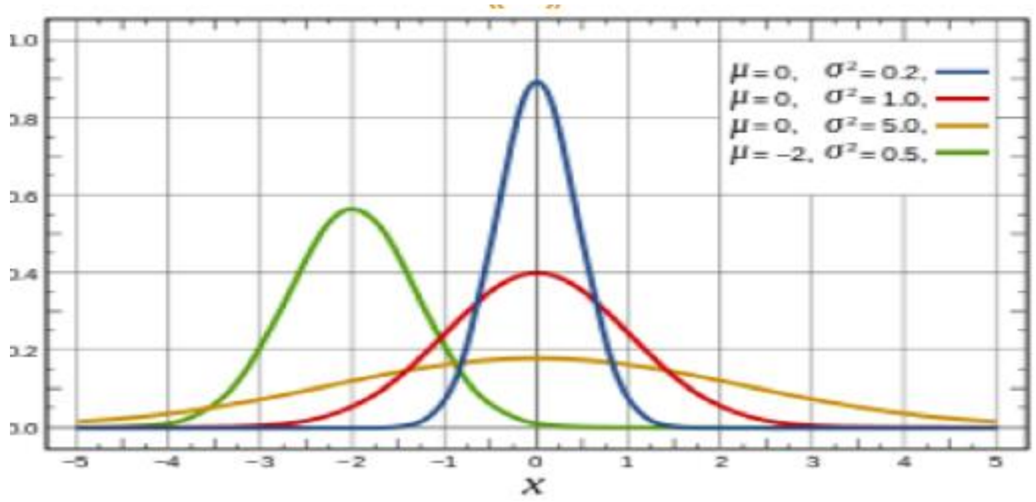
Height of 95% of Innomatics students varies between 1 m and 1.6 m. Assume that height is normally distributed.



- **What is the mean?**
  - Mean = (1+1.6)/2 = 1.3 m
- **What is standard deviation?**
  - We know 95% represents 2 sd away from mean (on both sides)
    - (1.6-1)/4 = 0.15 m is standard deviation
- **What does it mean?**
  - It says that out of every 1000 values 680 should be within 1 sd (likely)
  - For every 1000, 950 should be within two sd (very likely)
  - For ever 1000, 997 within 3 sd (almost certainly)

Normal Distribution contd.

- There can be many different shapes (either because of change in mean or standard deviation).

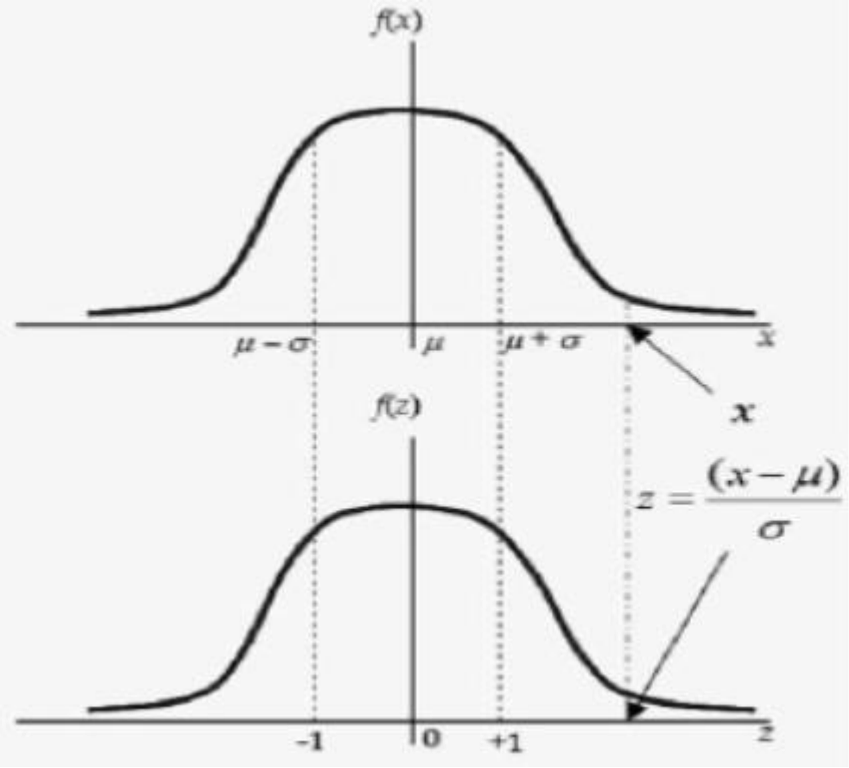


Standardizing the Normal Distribution

- Assume X1 ~ N(10 ,106), and X ~ N(7 ,56)
- If you have to compare these two, how would you do that?
  - **Standardizing** comes to rescue

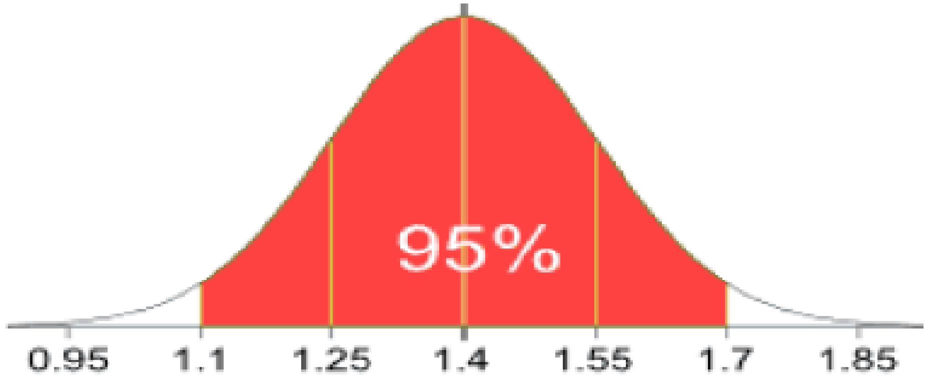
We define standard normal variable (Z) with mean 0 and sd 1

- $Z=(X-\mu)/\sigma$
- Z also has a normal distribution
- Z score tells position of a point relative to other points in distribution
- Alternatively, it tells how many s.d. away from mean a point is



Recalling the example

- Assume following distribution for height of students in Innomatics class

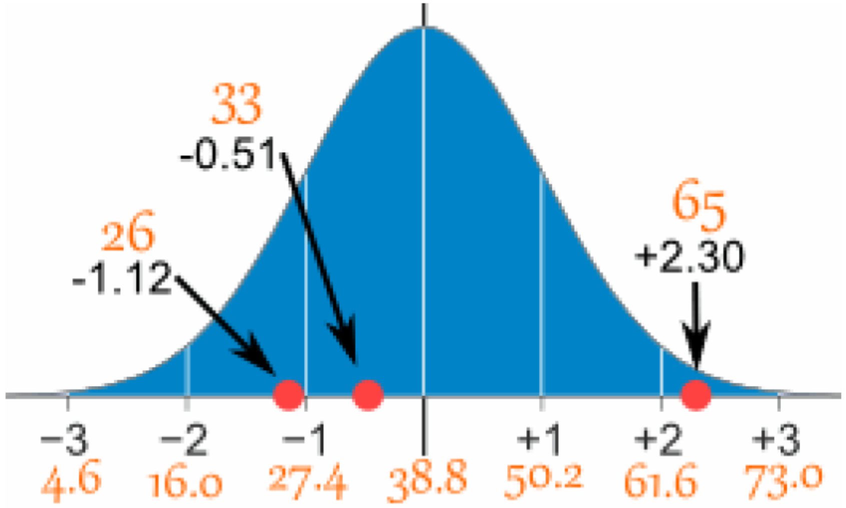


- Say, height of a student is 1.85m.
  - Student is 3 s.d. away from mean
  - Height of student has **z-score of 3**

Another Example

Assume following daily temperatures in **Hawaii**

- 26, 33, 65, 28, 34, 55, 25, 44, 50, 36, 26, 37, 43, 62, 35, 38, 45, 32, 28, 34
- Mean?
  - **38.8**
- S.D?
  - **11.4**
- Convert them to z-scores Z=(X-μ)/σ
  - 26 -> (26-38.8)/11.4 = -1.12
  - 33 -> (33-38.8)/11.4 = -0.51
  - 65 -> (65-38.8)/11.4 = 2.3 and so on



INNOMATICS RESEARCH LABS

Z Score & Probability

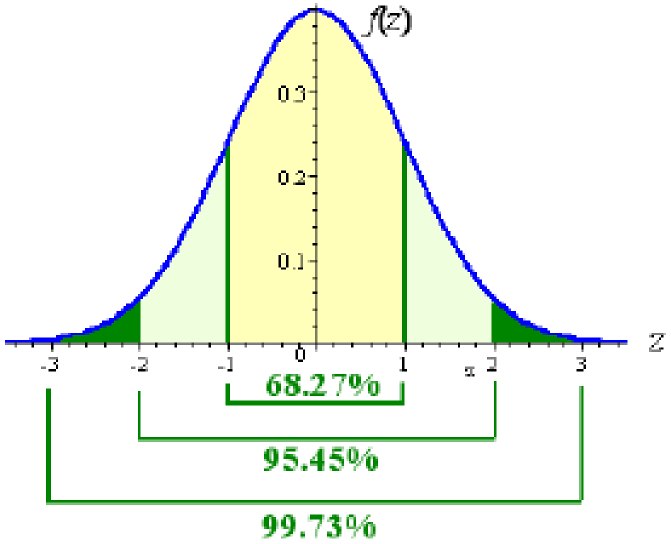
- $X \sim N(\mu, \sigma 6) \Rightarrow Z = (X - \mu)/\sigma$

  - for x1 and x2 belonging to X, we have corresponding z1 and z2 (belonging to Z)

  - Area under curve between X=x1 and X=x2 is same as that under Z for Z=z1 and Z=z2

  - $P(x1 <= X <= x2) = P(z1 <= Z <= z2)$

  - $-1 <= Z <= 1 \Rightarrow 68\%$
  - $-2 <= Z <= 2 \Rightarrow 95\%$
  - $-3 <= Z <= 3 \Rightarrow 99.7\%$



PMF (Probability Mass Function) vs PDF (Probability Density Function)

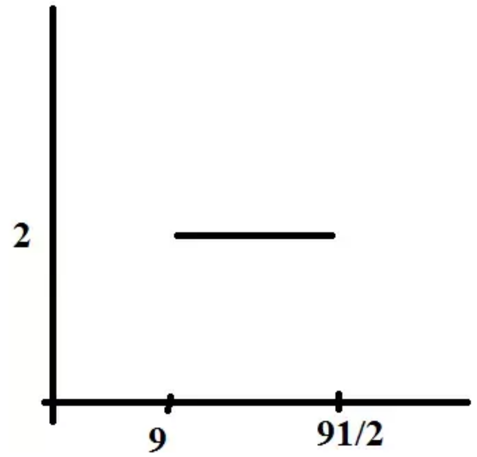
Consider **rolling a dice.** First let's define the Sample space: Sample space is the set of all possible outcomes. So in the case of rolling a dice once, the sample space is {1,2,3,4,5,6}.

Now what is the **probability** of getting 1? well, 1/6 (assuming unbiased dice).

List all the outcomes/events and their corresponding probabilities. That's **PMF** (Probability Mass Function). 1-1/6, 2-1/6, 3-1/6, 4-1/6, 5-1/6, 6-1/6. This is a function (mapping) from the events to the probability values.

PDF is analogous to **PMF**, but it's defined for continuous random variables. For example, if I say that I will come between 9pm and 9.30pm today, but expect me at any time in between with equal probability, then the pdf corresponding to that random variable is:

Here the range of pdf of any value between 9 and 91/2 is 2. But we are not interested in that value, rather we want to answer questions like what is the probability that I will come between 9 and 9.15. Then you can find the area under the curve between 9 and 91/4, which is 1/2.



Gaussian For Blind date

Julie is a student, and her best friend keeps trying to get her fixed up on blind dates in the hope that she'll find that special someone.

- She want to date only tall guys

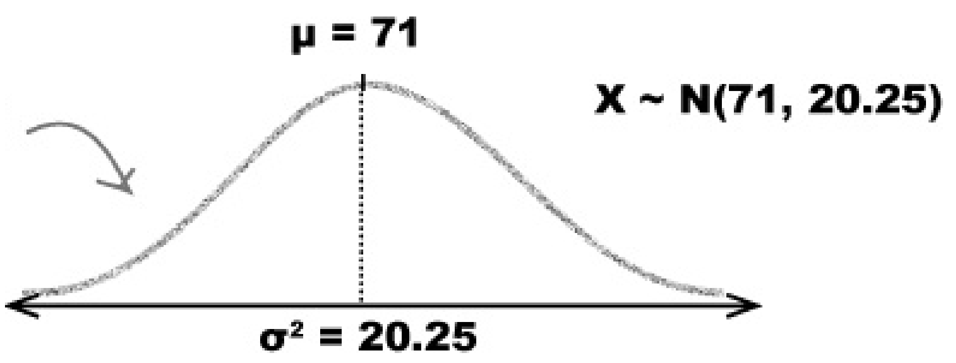
Oh! By the way Julie is 64" tall.

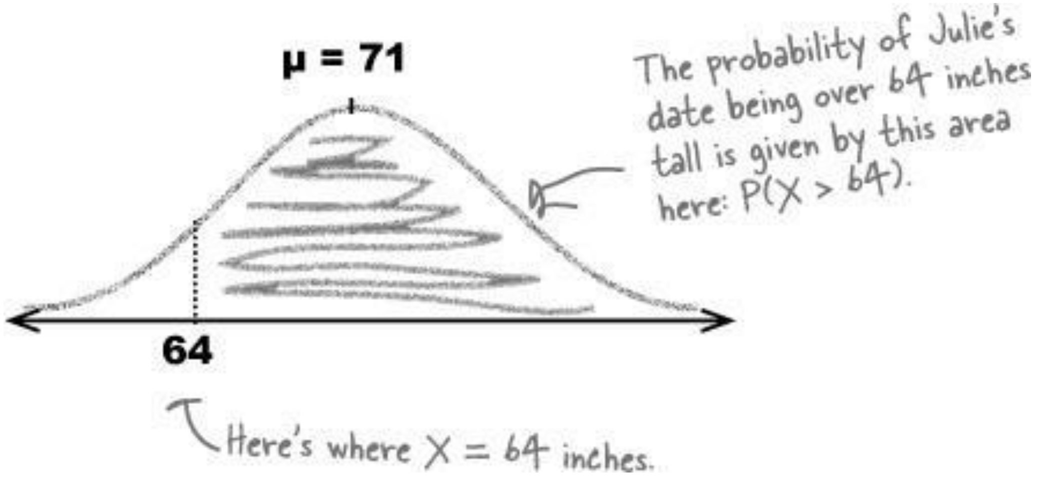
Probability Distribution

- Step 1: Determine the distribution

Julie wants to marry a person taller than her and is going on blind dates. The mean height of the 'available' guys is 71" and the variance is 20.25 $inch^2$.

Julie is 64" tall.

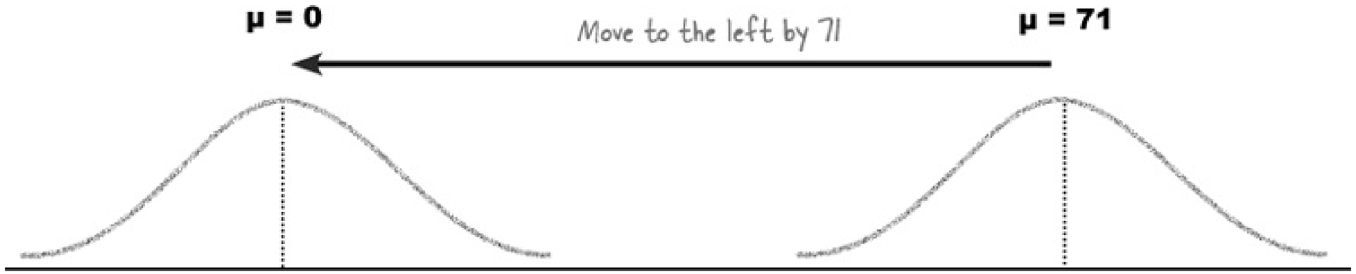




Calculating Normal Probabilities

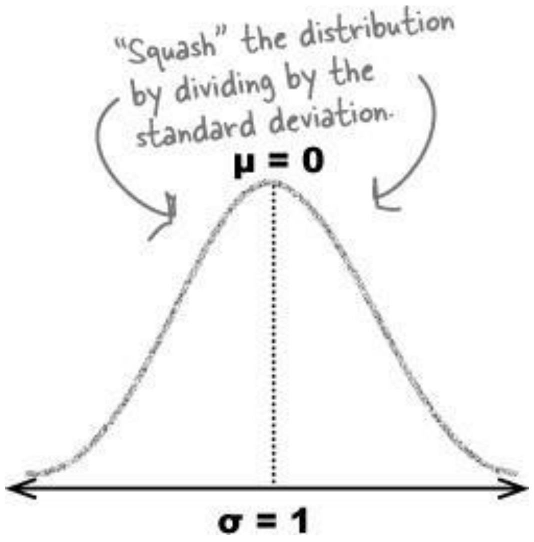
- Step 2: Standardize to Z ~ N(0,1)

1. Move the mean This gives a new distribution X-71 ~ N(0,20.25)



2. Rescale the width by dividing by the standard deviation

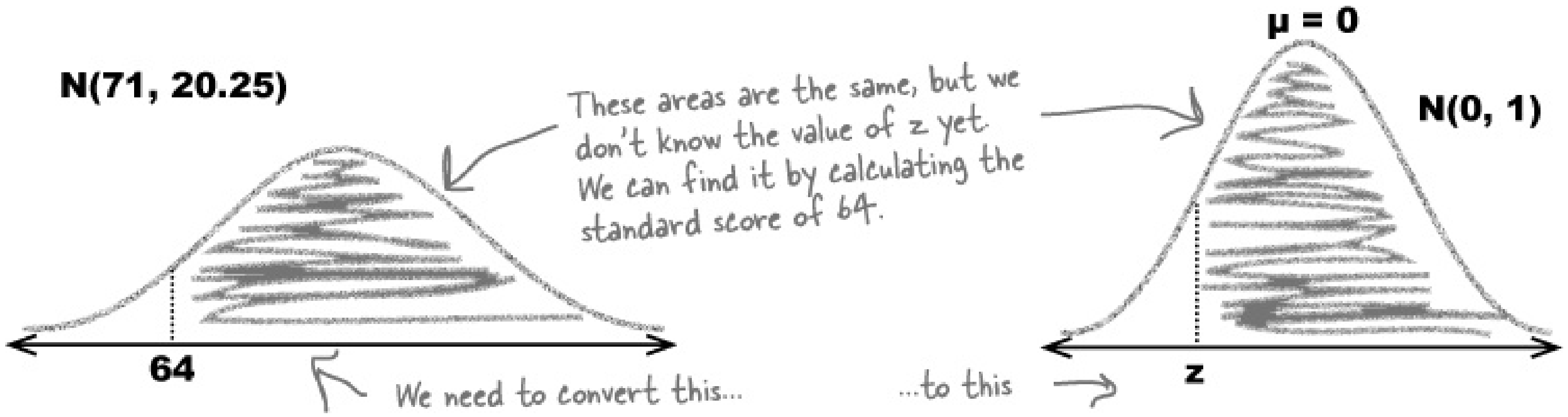
This gives us $\frac{X-71}{4.5}$



Z = $\frac{X-\mu}{\sigma}$ is called standard score or the Z-score

Calculating Normal Probabilities

- Step 2: Standardize to Z ~ N(0,1)



$$Z = \frac{X - 71}{4.5} = -1.56 \text{ for height}$$

64”

Calculating Normal Probabilities

- Step 3: Look up the
- probability in the tables

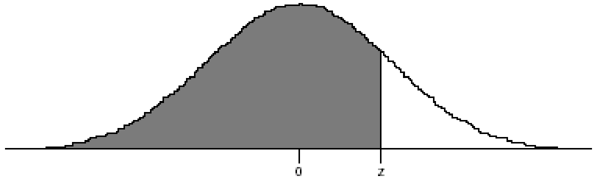
Note the tables give P(Z<z)

$$Z = \frac{X - 71}{4.5} = -1.56$$

for height 64”

$$P(Z > -1.56) = 1 - P(Z < -1.56)$$
$$= 1 - 0.0594$$
$$= 0.9406$$

Z-table



Normal Deviate z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-4.0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.7	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379

Attention Check

Find the probability **Julie** finds a man in the between 66” to 76”.

Z-Score @ 66” = (66 – 71)/4.5 = - 1.11

P(Z = -1.11) = 0.1131

Z-Score @ 76” = (76 - 71)/4.5 = 1.11

P(Z = 1.11) = 0.8665

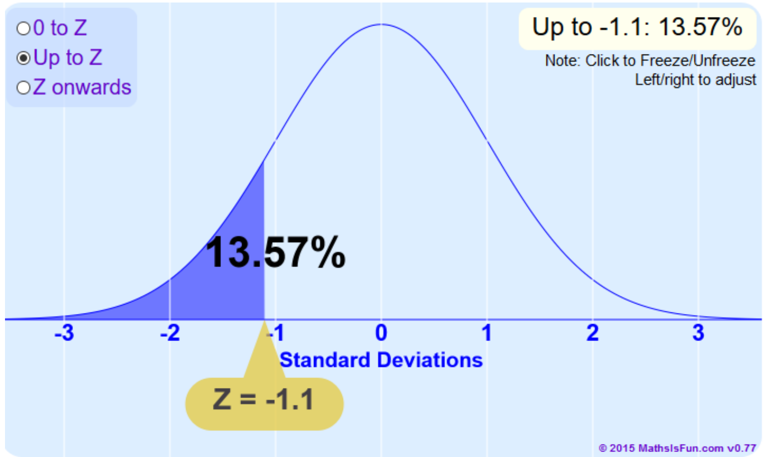
P(66” < X < 76”) = P(X = 76”) – P(X = 66”)

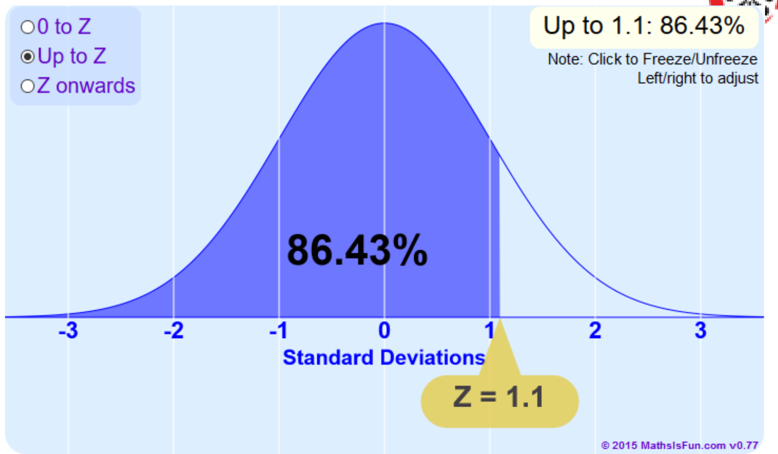
= 0.8665 – 0.1131

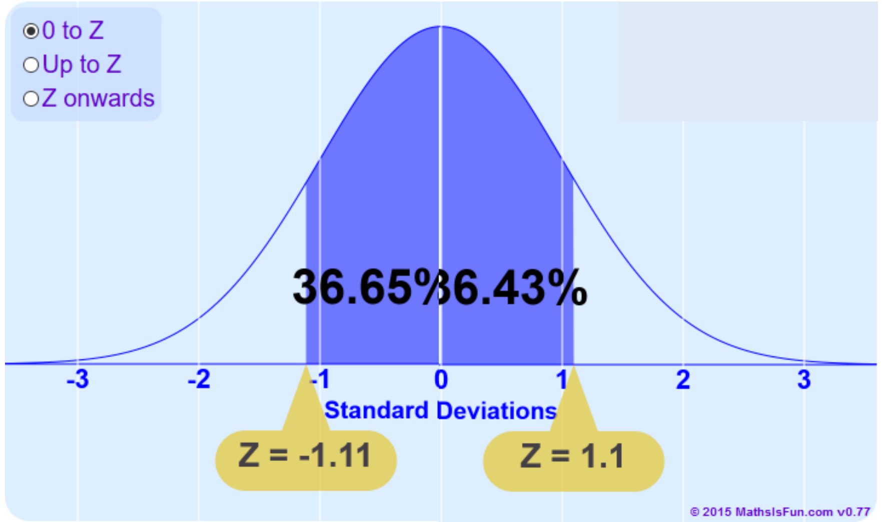
= 0.7534

Z-table

https://www.mathsisfun.com/data/standard-normal-distribution-table.html







P(-1.11 < Z < 1.11)

= P(Z = 1.11) – P(Z = -1.11)
= 0.8665 – 0.1131
= 0.7534

Attention Check

Q. What is the standard score for N(20,9), value 6?

A. $z = \frac{6 - 20}{3} = -4.66$

Q. The standard score of value 40 is 1. If the variance is 25, what is the mean?

A.

$$1 = \frac{40 - \mu}{5}$$

$\therefore \mu = 40 - 5 = 35$

Attention Check

Q. Julie just realized that she wants her date to be taller when she is wearing her heels, which are 5" high. Find the new probability that her date will be taller.

A. Julie Height = 64"

$$Z = \frac{(64 + 5) - 74}{4.5} = -0.44$$

$P(Z < -0.44) = ?$

Z-table

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
–3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
–3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
–3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
–3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
–3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
–2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
–2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
–2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
–2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
–2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
–2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
–2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
–2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
–2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
–2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
–1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
–1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
–1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
–1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
–1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
–1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
–1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
–1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
–1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
–1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
–0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
–0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
–0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
–0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
–0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
–0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
–0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
–0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
–0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247

Attention Check

Q. Julie just realized that she wants her date to be taller when she is wearing her heels, which are 5” high. Find the new probability that her date will be taller.

A. Julie Height = 64”

$$Z = \frac{(64+5) - 74}{4.5} = -0.44$$

$P(Z < -0.44) = 0.33$

$\therefore P(Z > -0.44) = 0.67$ or 67%

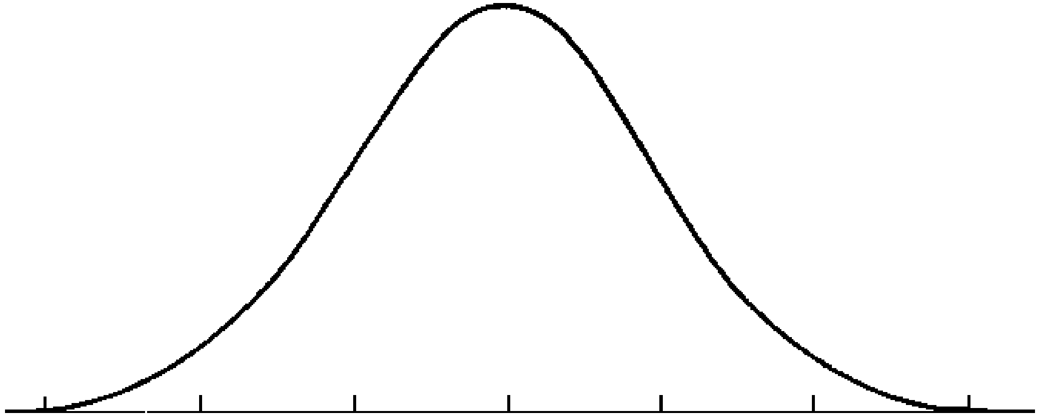
Z-table

LET’S PRACTICE!

You can use a white board or a scrap piece of paper!

WE DO: 95 % of students at school are between 1.1 m and 1.7 m tall.

- Assuming this data is normally distributed can you calculate the mean and standard deviation?

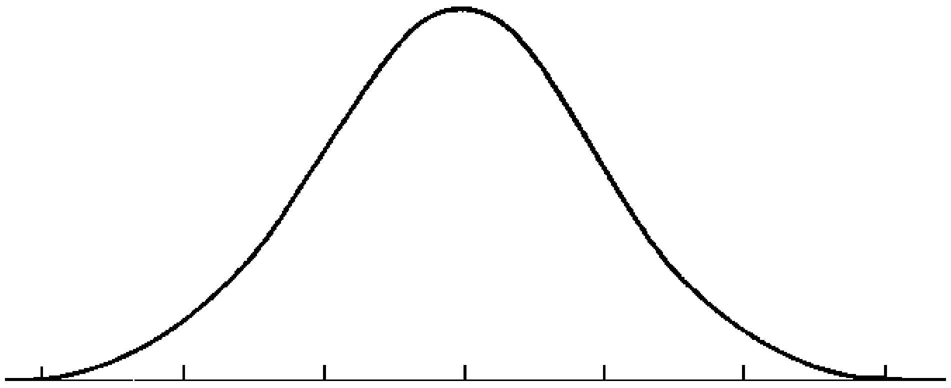


YOU DO: 68% of American's own a dog between 4 and 6 years old.

Assuming this data is normally distributed can you calculate the mean?

WE DO: The reaction times for a hand-eye coordination test administered to 1800 teenagers are normally distributed with a mean of .35 seconds and a standard deviation of .05 seconds.

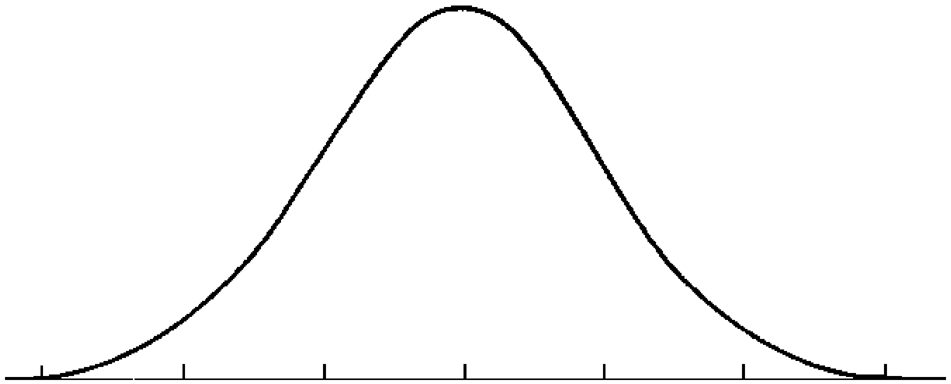
- Represent this information on a bell curve:



- About how many teens had reaction times between .25 and .45 seconds?
- What is the probability that a teenager selected at random had a reaction greater than .4 seconds?

YOU DO: The waiting times for an elevator are normally distributed with a mean of 1.5 minutes and a standard deviation of 20 seconds.

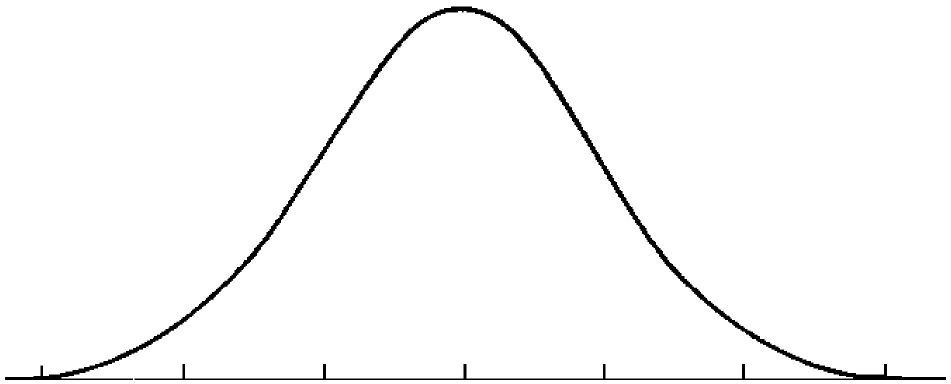
- Represent this information on a bell curve:



- Find the probability that a person waits longer than 2 minutes 10 seconds for the elevator.

YOU DO: Mrs. Smith gave a test in her Algebra 2 class. The scores were normally distributed with a mean of 85 and a standard deviation of 3.

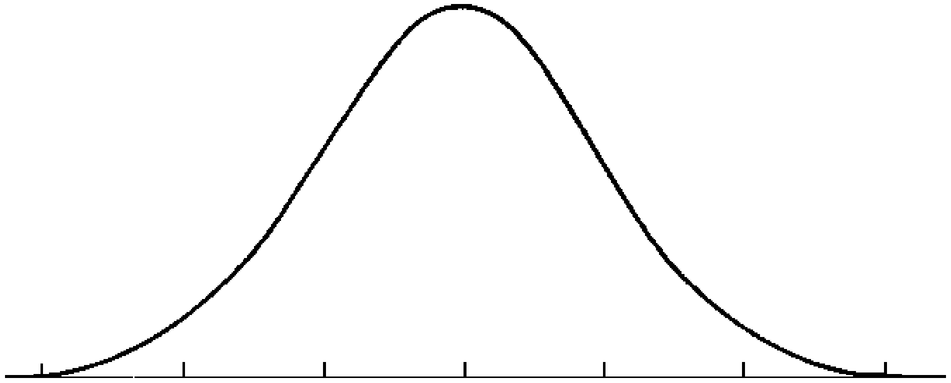
- Represent this information on a bell curve:



- What percent would you expect to score between 82 and 88?

YOU DO: The heights of 250 twenty-year-old women are normally distributed with a mean of 1.68 m and standard deviation of 0.06 m.

- Represent this information on a bell curve.



- Find the probability that a woman has a height between 1.56 m and 1.74 m

Normal Approximation of Binomial distribution

Using the normal distribution to approximate the binomial distribution

If n, p and q are such that:

np and nq

are both greater than 5.

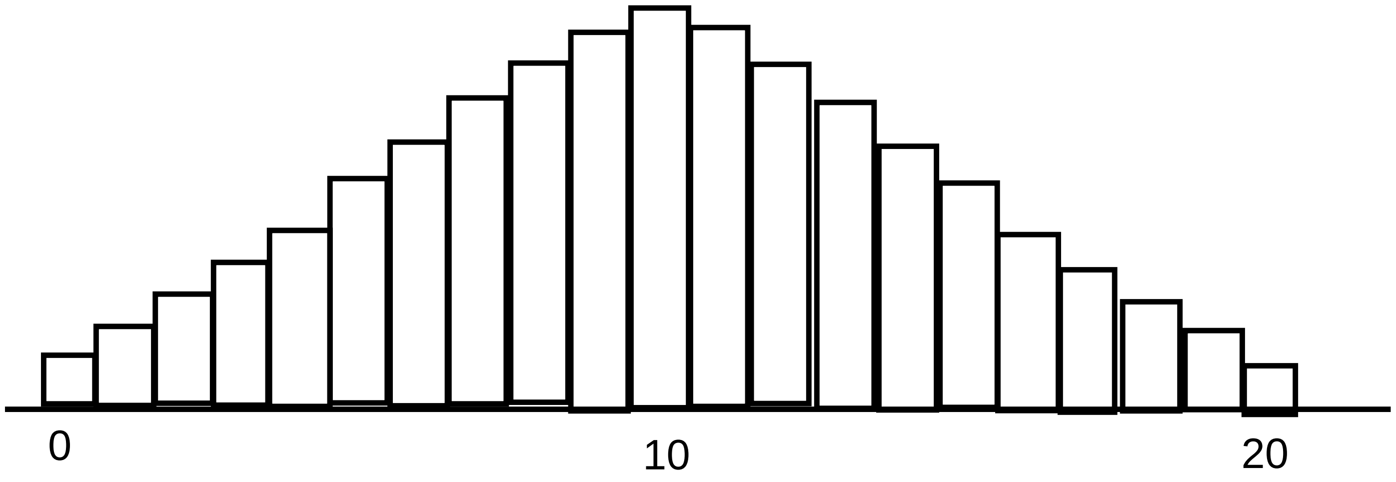
Mean and Standard Deviation: Binomial Distribution

$$\mu = np \text{ and } \sigma = \sqrt{npq}$$

Experiment: tossing a coin 20 times

Problem: Find the probability of getting exactly 10 heads.

Distribution of the number of heads appearing should look like:



Using the Binomial Probability Formula

$n = 20$

$x = 10$

$p = 0.5$

$q = 1 - p = 0.5$

$P(10) = \underline{0.176197052}$

Normal Approximation of the Binomial Distribution

First calculate the mean and standard deviation:

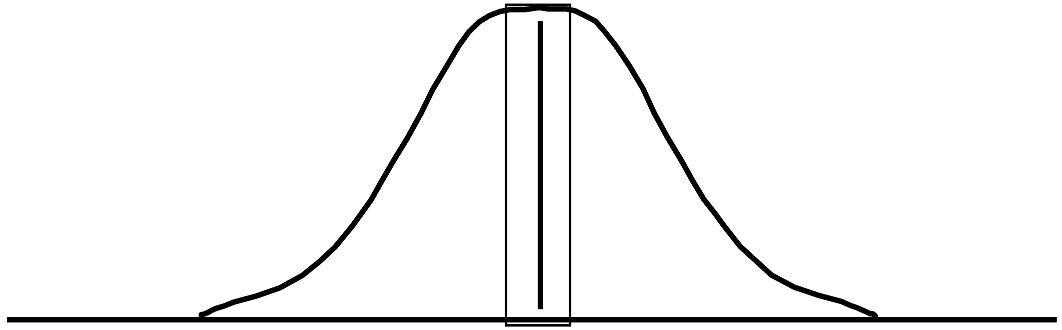
Mean = np = 20 (.5) = 10

$$\sigma = \sqrt{np(1-p)} = \sqrt{20(.5).5} = 2.24$$

The Continuity Correction

- Continuity Correction is needed because we are approximating a discrete probability distribution with a continuous distribution.

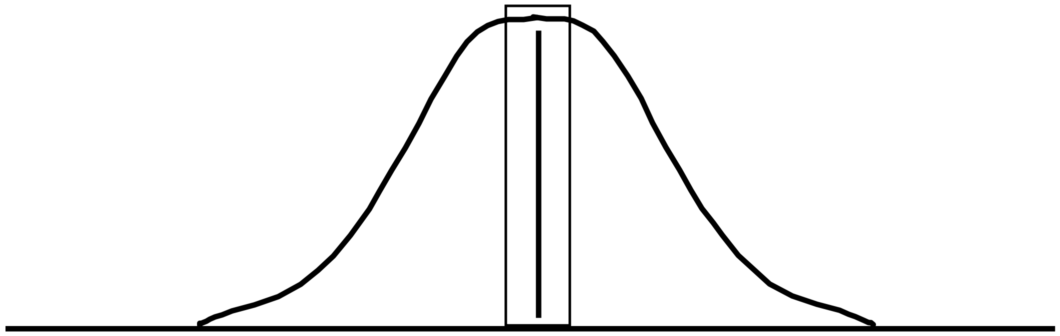
- We are using the area under the curve to approximate the area of the rectangle.



The Continuity Correction

Continuity Correction: To compute the probability of getting exactly 10 heads, find

The probability of getting between 9.5 and 10.5 heads.



Using the Normal Distribution

$P(9.5 \leq x \leq 10.5) = ?$

for $x = 9.5$: $z = -0.22$

$P(z < -0.22) = .4129$

for $x = 10.5$: $z = = 0.22$

$P(z < .22) = .5871$

$P(9.5 \leq x \leq 10.5) = .5871 - .4129 = .1742$

Application of Normal Distribution

If 22% of all patients with high blood pressure have side effects from a certain medication, and 100 patients are treated, find the probability that at least 30 of them will have side effects.

Using the Binomial Probability Formula we would need to compute:

$$P(30) + P(31) + \ldots + P(100) \text{ or } 1 - P(x \leq 29)$$

Using the Normal Approximation to the Binomial Distribution

Is it appropriate to use the normal distribution?

Check: **$n\ p =$**

$n\ q =$

$n\ p = 22$

$n\ q = 78$

Both are greater than five.

Find the mean and standard deviation

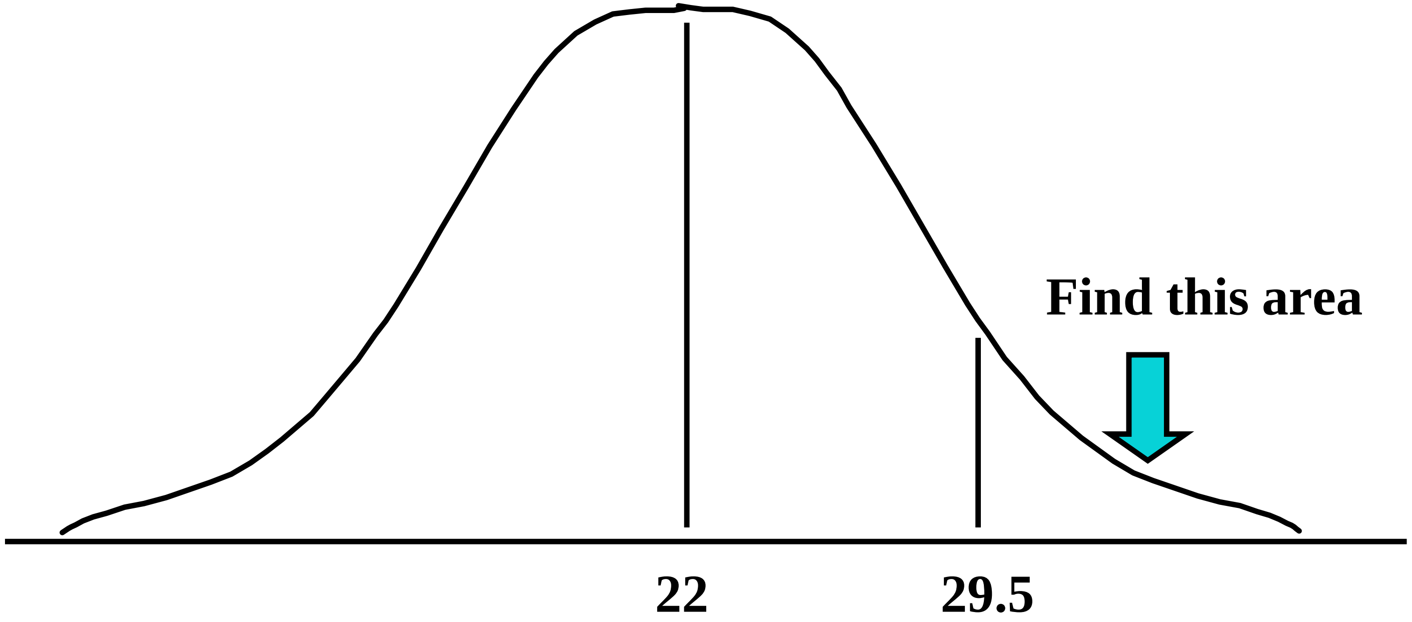
Mean = 100(.22) = 22

and σ = $\sqrt{\mathbf{100(.22)(.78)}}$

$= \sqrt{17.16} = 4.14$

Applying the Normal Distribution

To find the probability that at least 30 of them will have side effects, find P(x = 29.5)

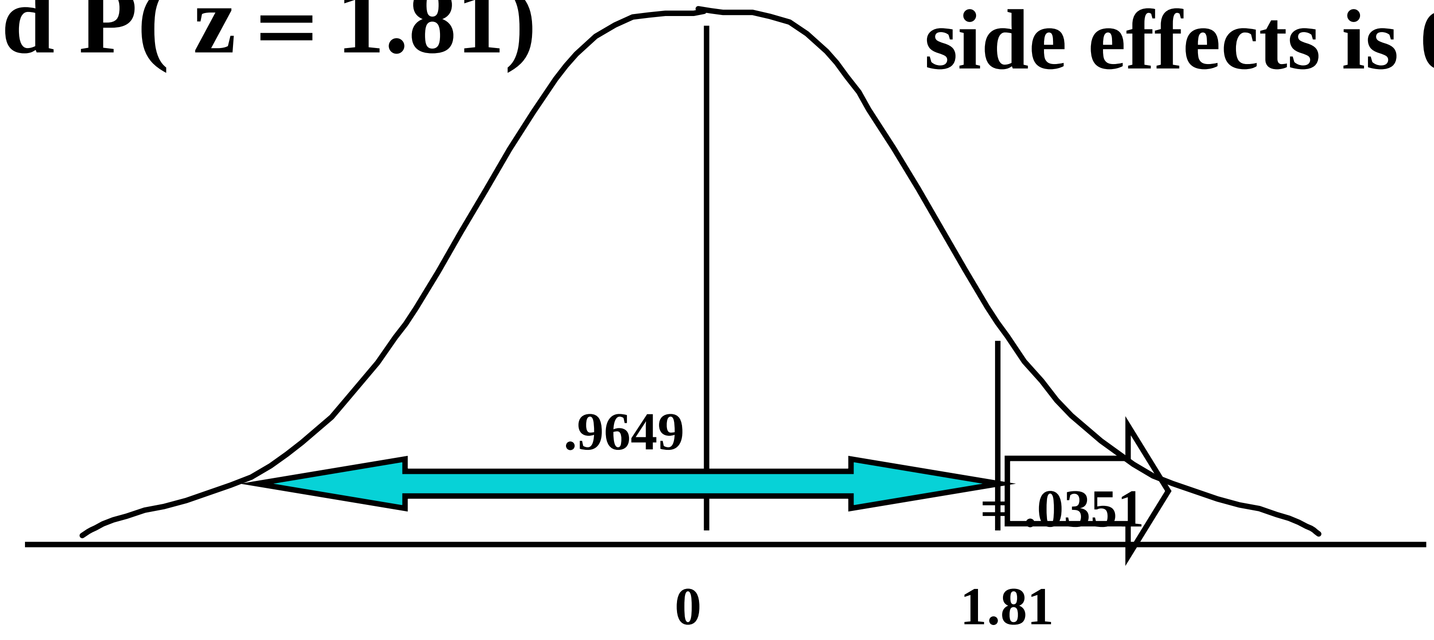


Applying the Normal Distribution

$$z = \frac{29.5 - 22}{4.14} = 1.81$$

Find P(z = 1.81)

The probability that at least 30 of the patients will have side effects is 0.0351.



Reminders:

- Use the normal distribution to approximate the binomial only if both np and nq are greater than 5.

- Always use the continuity correction when approximating the binomial distribution.

References:

- **Binomial Distribution:**
- https://www.khanacademy.org/video/binomial-distribution
- **Poisson Distribution:** https://www.khanacademy.org/video/poisson- process-1
- **Normal Distribution:** https://www.youtube.com/watch?v=RKdB1d5- OE0
- **Central Limit Theorem:** https://www.khanacademy.org/video/central- limit-theorem

THANK YOU