Continuous Distributions

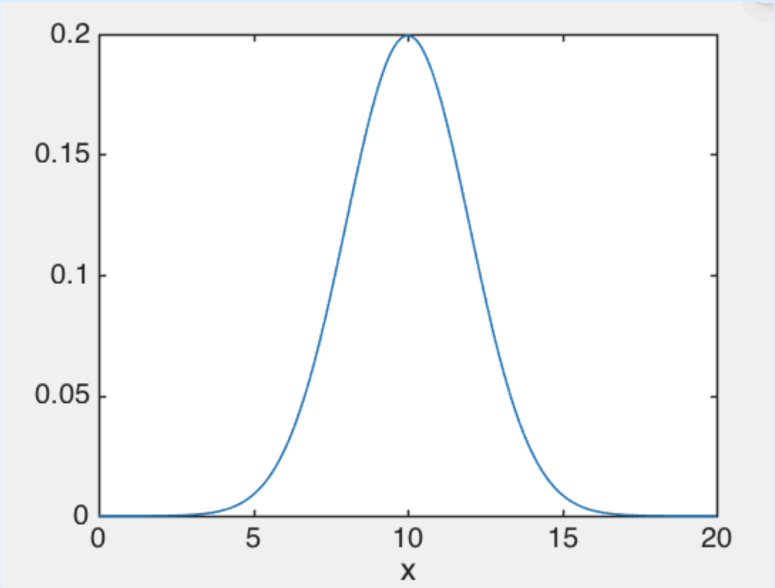
Agenda

- Continuous Distributions
- Probability Density Function
- Normal Distribution

Continuous Distributions

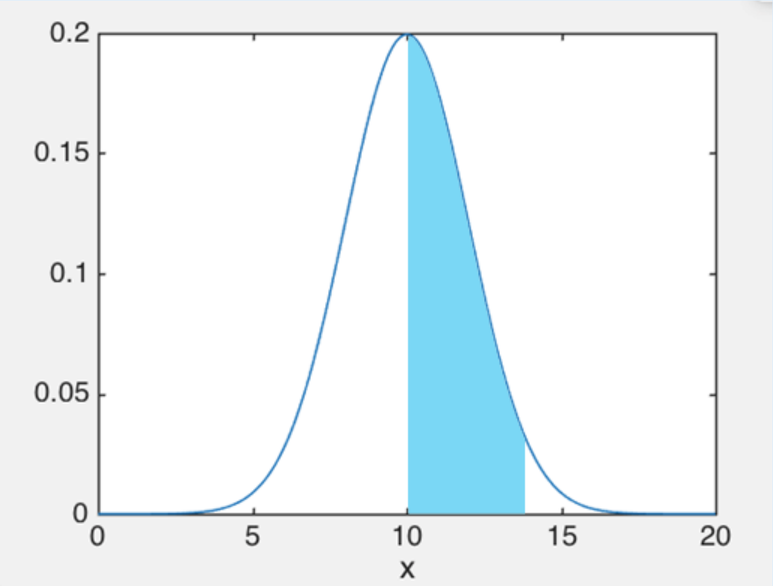
- Amount of rainfall in Mumbai in any month
  - Continuous Random Variable
- We have a range but it can take any values in the range (say 0-20 cm)
- Prob that rainfall in a month would be 10 cm?
  - Zero.
  - Infinite possibilities and prob of any single event is zero.
- PMF is no longer applicable. Replaced by Probability Density Function (PDF)
  - Prob of rainfall between 9.9 – 10.1 cm?
  - We can find prob in a particular area

Continuous Probability Distributions



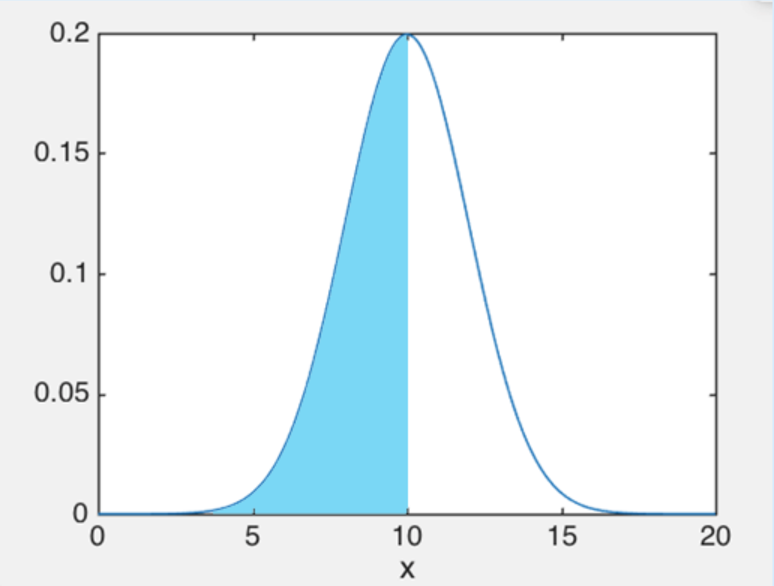
- Amount of rainfall (in cm) to fall in Mumbai during month of July.

Continuous Probability Distributions



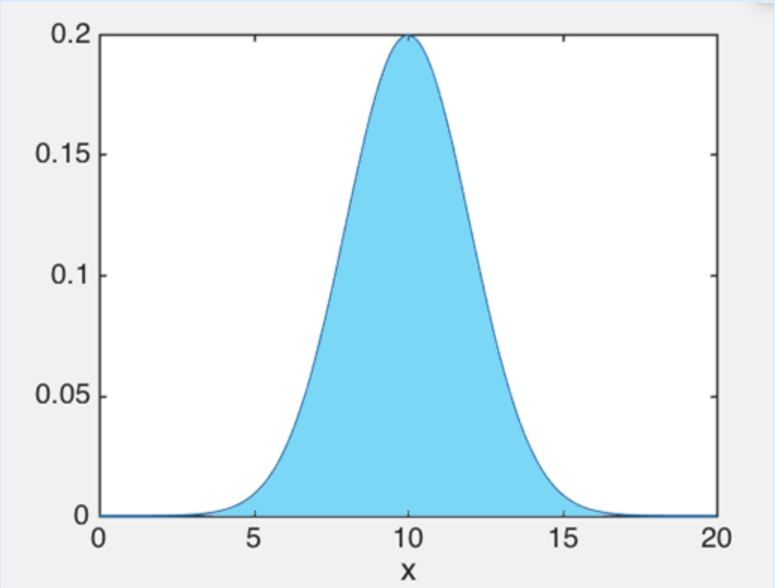
$$P(10 \leq X \leq 14)$$

Continuous Probability Distributions



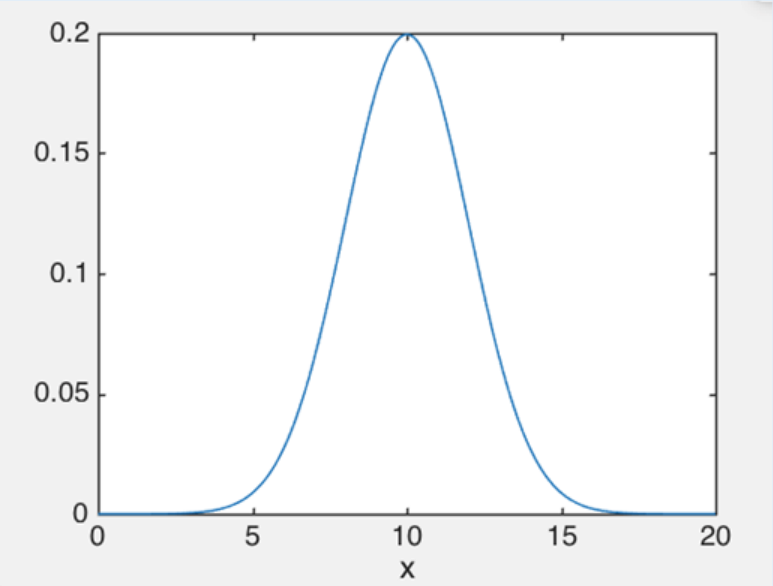
$$P(X \leq 10)$$

Continuous Probability Distributions



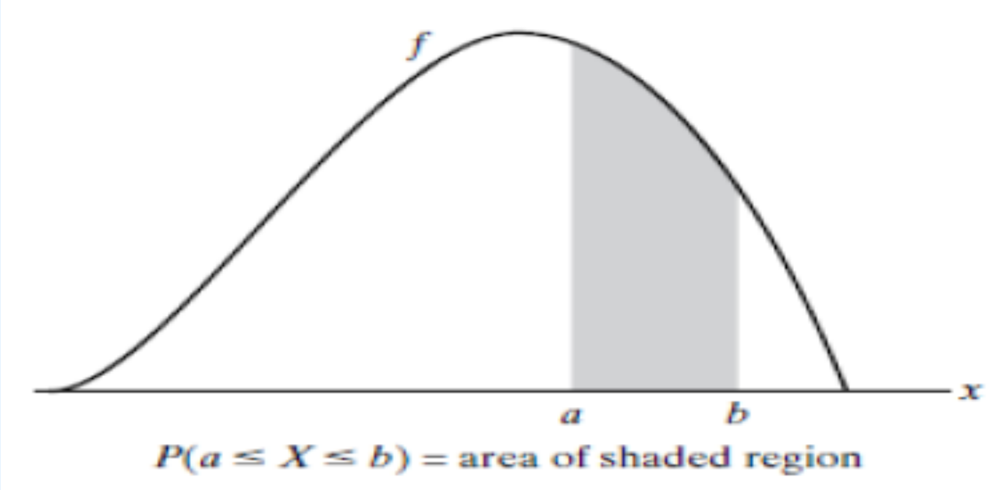
$$P(X \leq 20)$$

Continuous Probability Distributions



- This is a probability density function (pdf).

Technically Speaking…

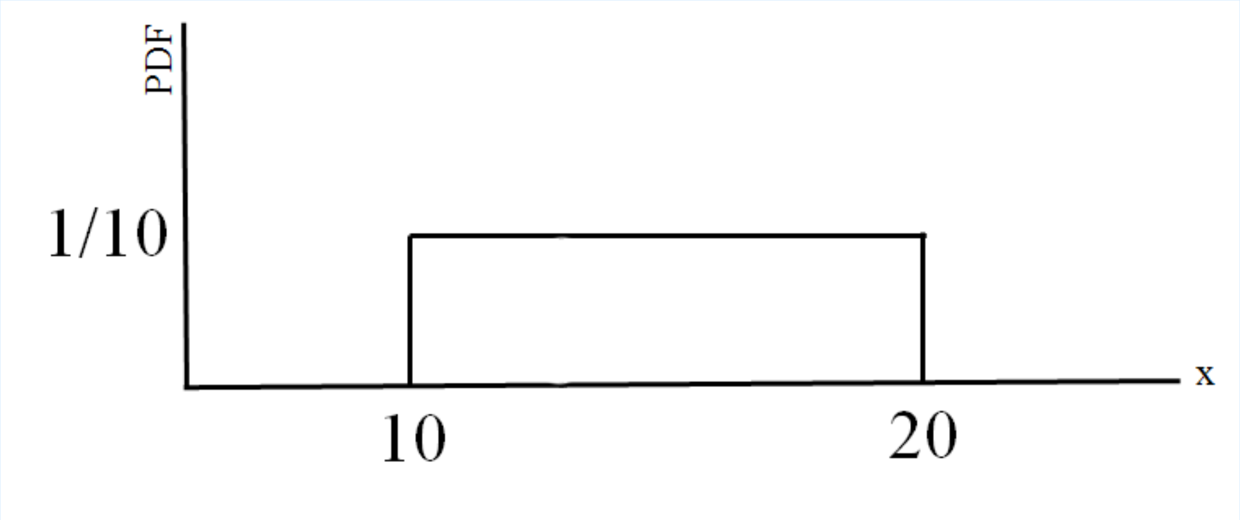


- For a function f(x), P(a<=x<=b) = $\int_a^b f(x)dx$
- Like PMF, PDF should also satisfy two conditions
  - f(x) >= 0 for every x
  - $\int_{-\infty}^{+\infty} f(x)dx = 1$

Discrete and Continuous

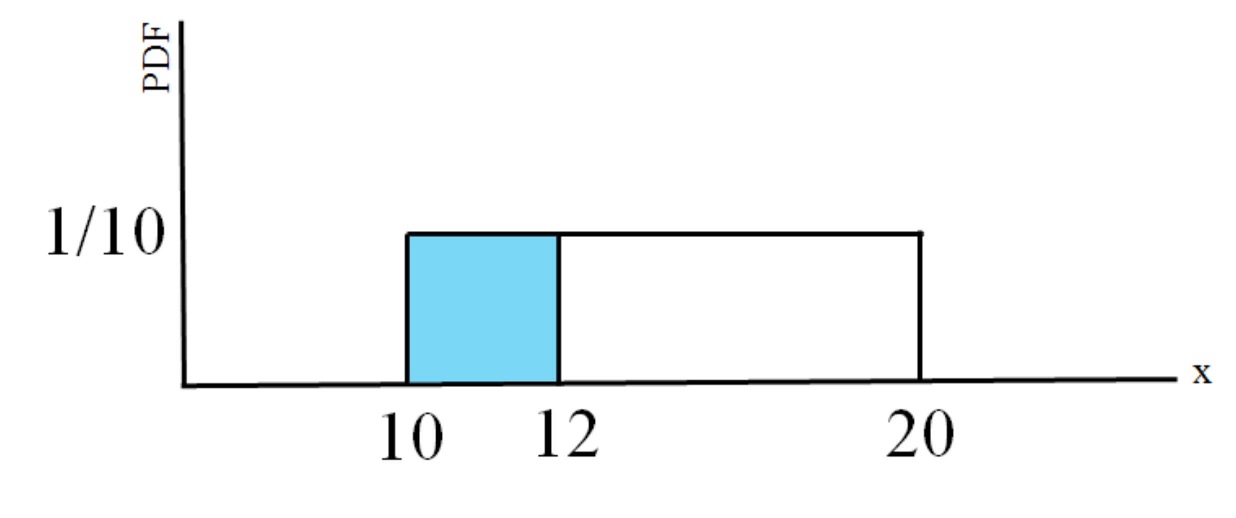
- Both are close cousins, concepts of mean and variance remain the same
- Σ is replaced by $\int$
- $\mu = E(X) = \int_{-\infty}^{\infty} x * f(x)dx$
- $V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx$

Continuous Probability Distributions



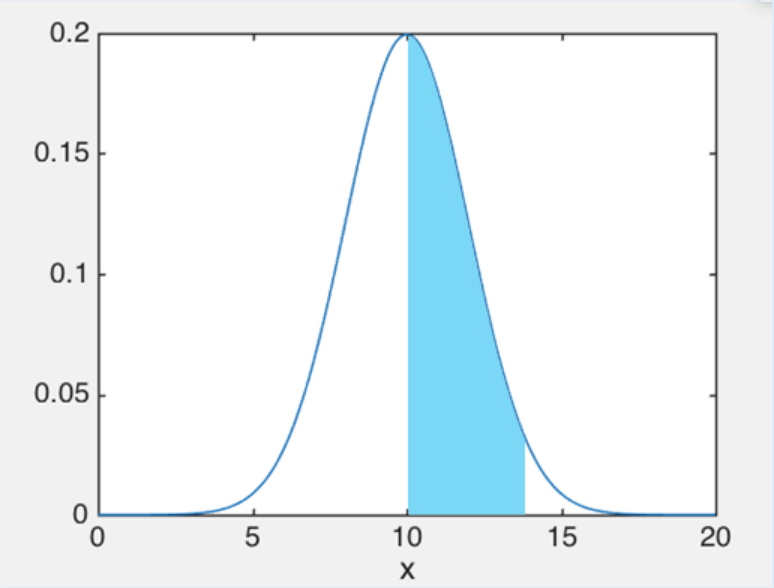
- Pdf for the uniform distribution.

Continuous Probability Distributions



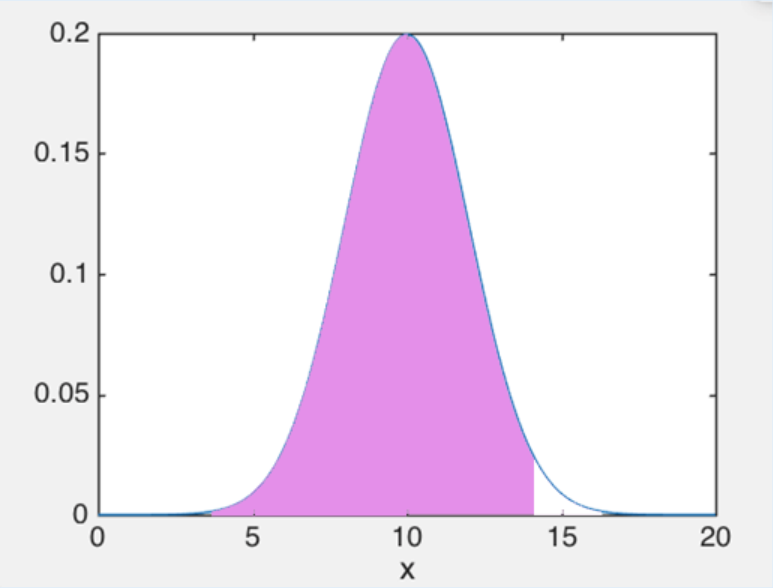
$$P(10 \leq X \leq 12)$$

Continuous Probability Distributions



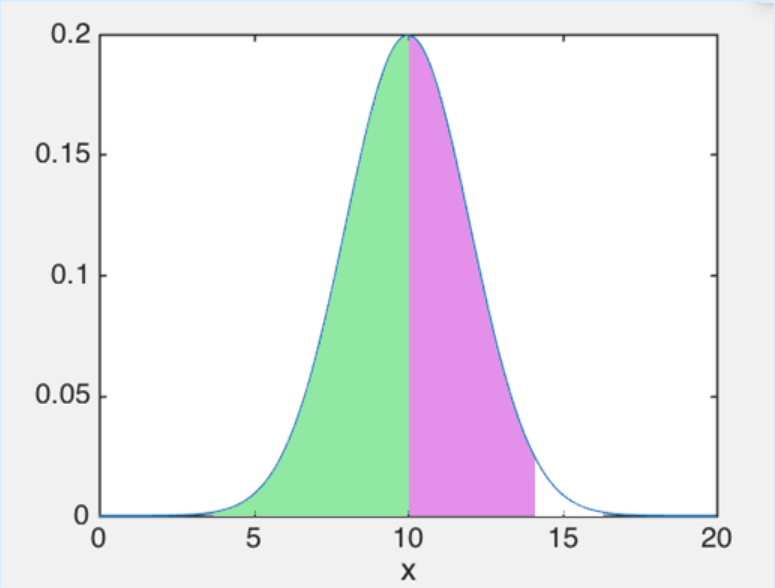
$$P(10 \leq X \leq 14)$$

Continuous Probability Distribution



- $P(10 \leq X \leq 14) = P(X \leq 14) - ???$

Continuous Probability Distributions



$$P(10 \le X \le 14) = P(X \le 14) - P(X \le 10)$$

Example

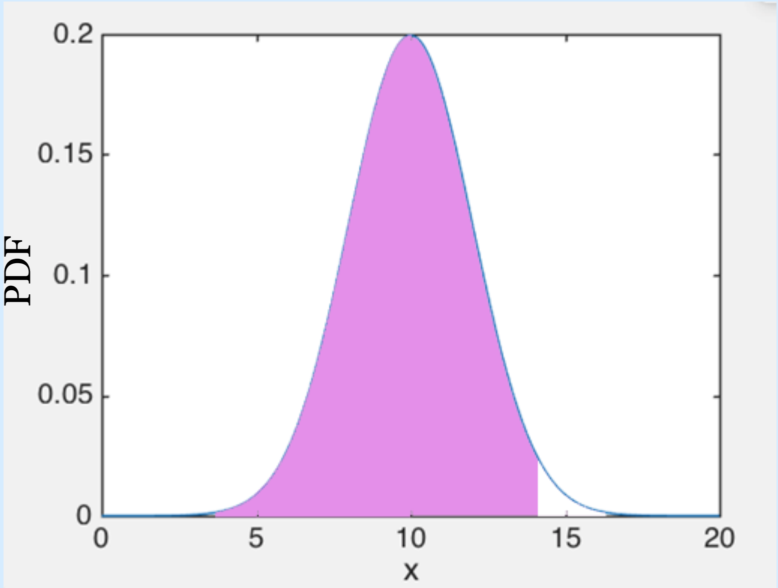
- Income (in lacs) is given by a continuous distribution with density f(x) = $2x^{-2}$ if x ≥ 2 and 0 if x < 2
  - What is probability that a person has salary between 3 Lacs and 5 Lacs.
    - P(3<=X<=5) = $\int_3^5 f(x)dx = \int_3^5 2x^{-2}dx = \frac{-2}{5}$ - $(\frac{-2}{3}) = \frac{4}{15}$
  - What is probability that a person has salary of minimum 8 Lacs.
    - P(X>=8) = $\int_8^\infty f(x)dx = \int_8^\infty 2x^{-2}dx = \frac{-2}{\infty}$ - $(\frac{-2}{8}) = \frac{1}{4}$
  - What is probability that a person has salary of maximum 5Lacs
    - P(X<=5) = $\int_{-\infty}^5 f(x)dx = \int_{-\infty}^2 0\ dx + \int_2^5 2x^{-2}dx = \frac{-2}{5}$ - $(\frac{-2}{2}) = \frac{3}{5}$

Example

- Suppose your office starts at 9 AM but you are generally late and I can model late arrival (in number of minutes) by a random variable with following pdf:
  - $f(x) = \begin{cases} 2(1-x)\ if 0 \leq x \leq 1 \\ 0\ otherwise \end{cases}$
  - What is the mean value and standard deviation of your late arrival
    - $\mu = E(X) = \int_{-\infty}^{\infty} x * f(x) dx = \int_0^1 x * 2(1-x)\ dx = \frac{1}{3}$
    - Variance is 1/18. (Verify this)

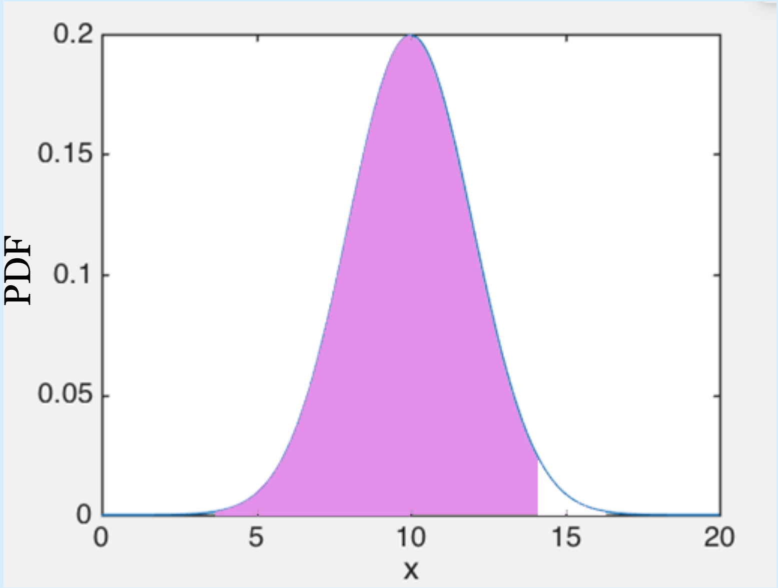
Cumulative Distribution Function

Cumulative Distribution Function



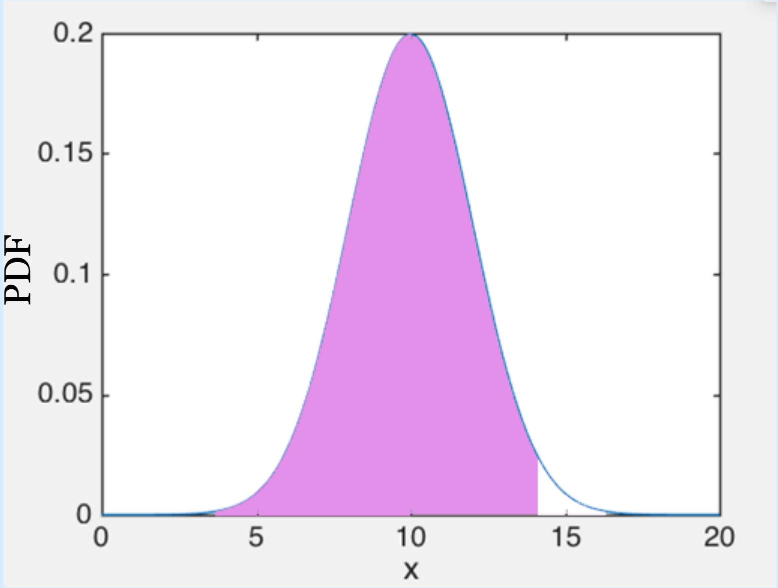
$$F(14) = P(X \leq 14)$$

Cumulative Distribution Function



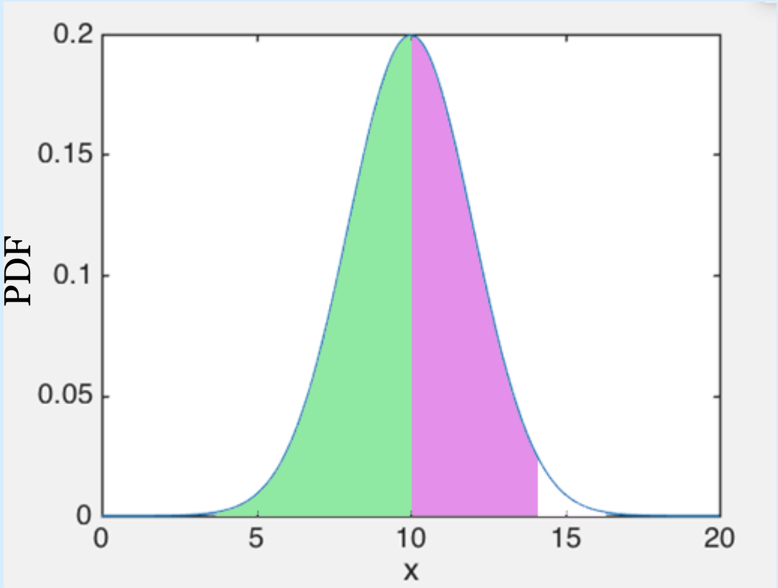
$$F(x) = P(X \leq x)$$

Cumulative Distribution Function



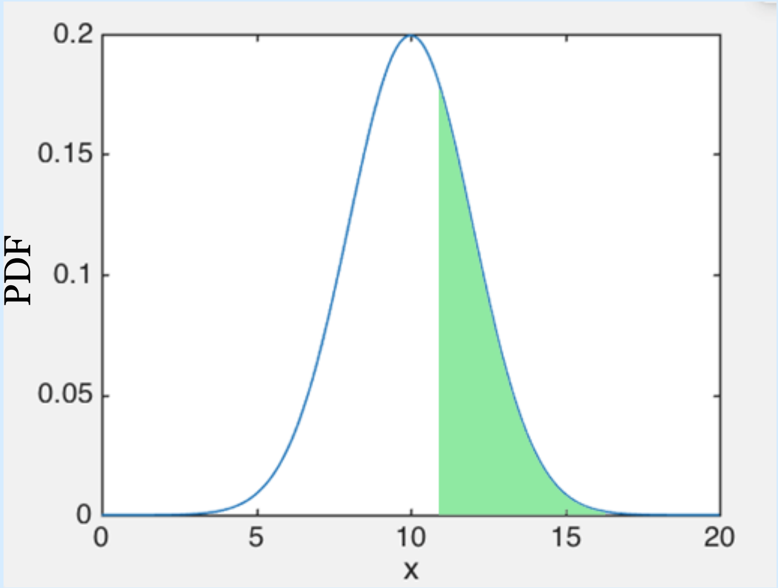
$$F(x) = P(X < x)$$

Cumulative Distribution Function



$$P(10 \le X \le 14) = P(X \le 14) - P(X \le 10)$$
$$= F(14) - F(10)$$

Cumulative Distribution Function



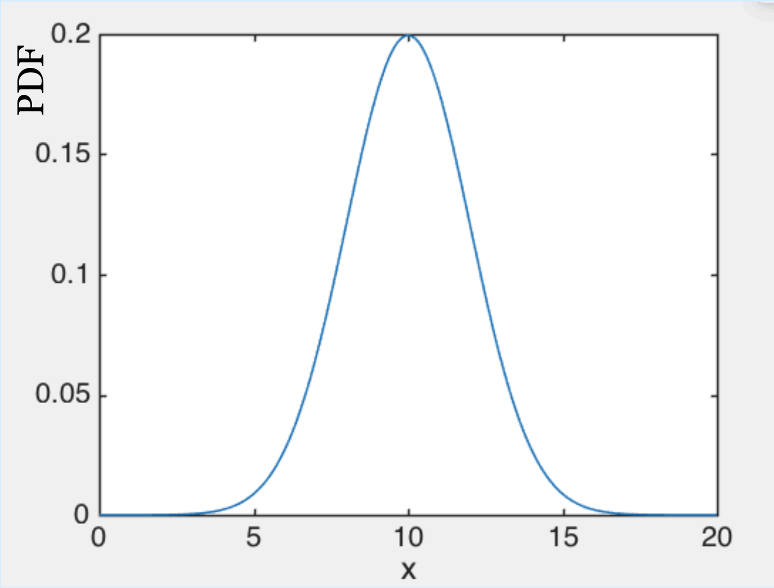
$$P(X \geq 11) = 1 - F(11)$$

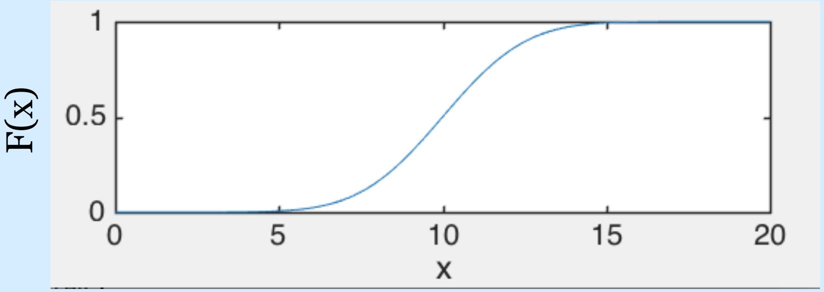
Cumulative Distribution Function

$$F(-10) \approx 0$$

$$F(10) \approx .5$$

$$F(20) \approx 1$$

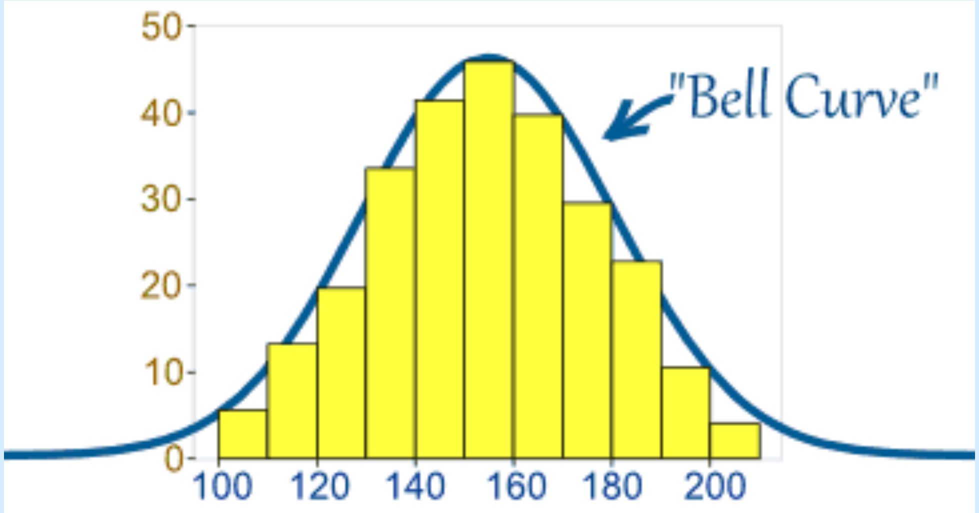




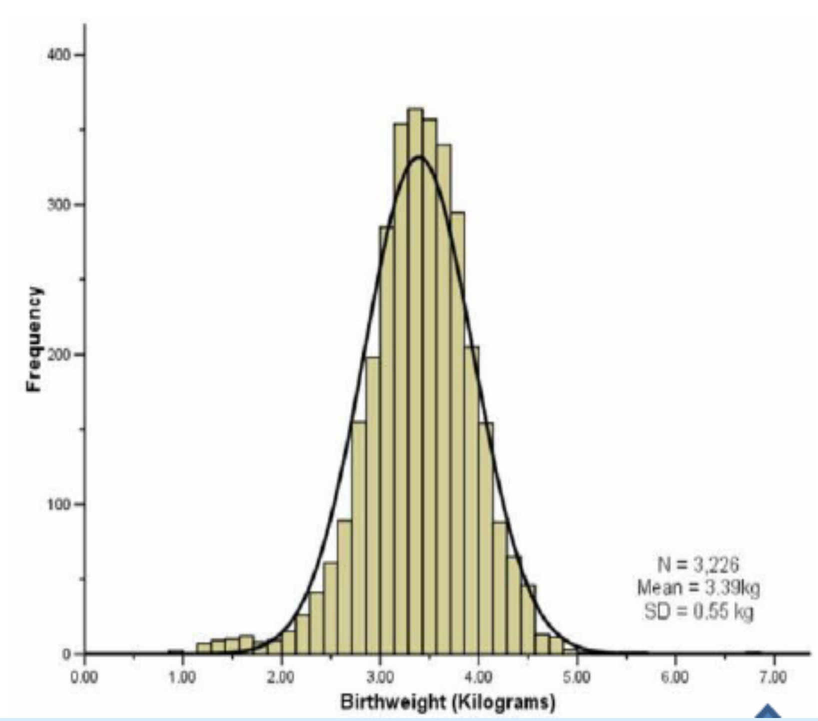
Normal Distribution

- 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

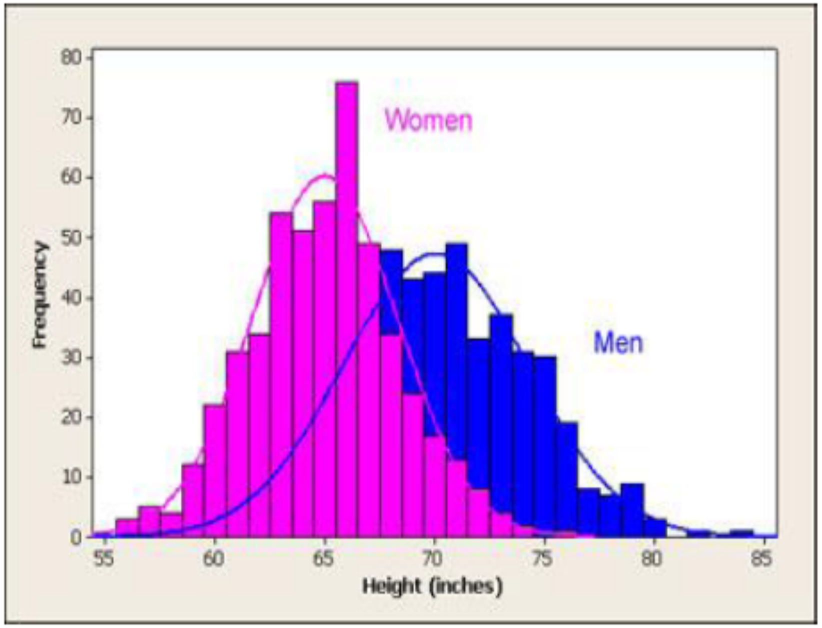
Say Hi to the Bell Curve



Examples

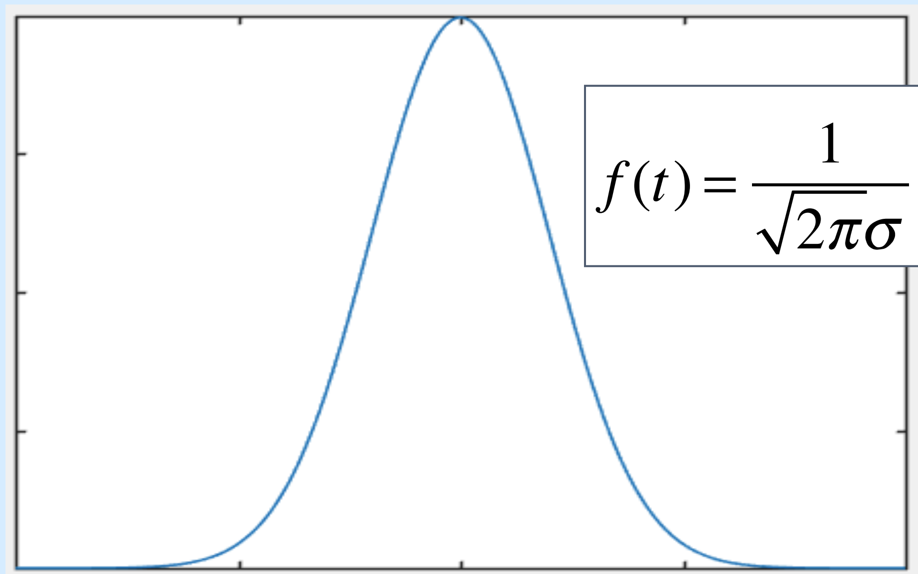


Frequency
400
300
200
100
0
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
Birthweight (Kilograms)
N = 3,226
Mean = 3.39kg
SD = 0.55 kg



Frequency
80
70
60
50
40
30
20
10
0
Women
Men
55
60
65
70
75
80
85
Height (inches)

The Normal Distribution

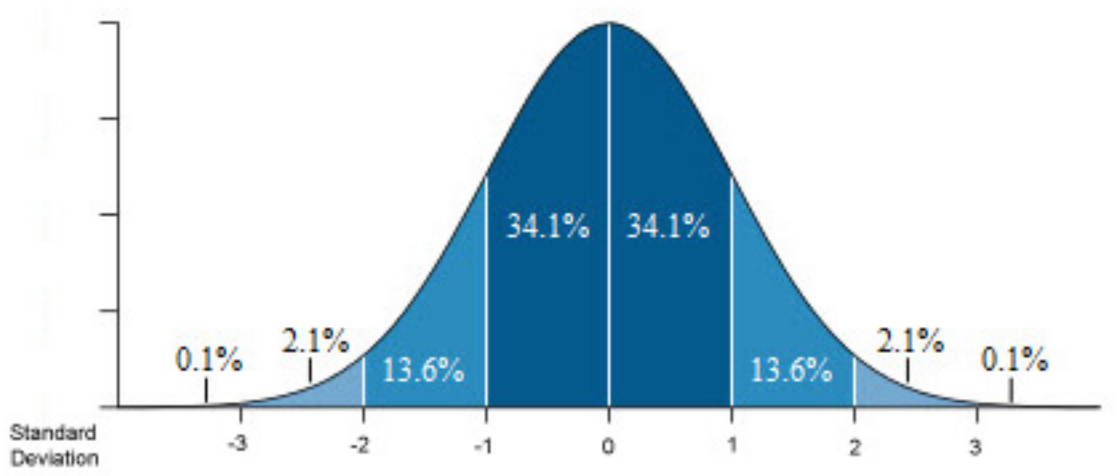


$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

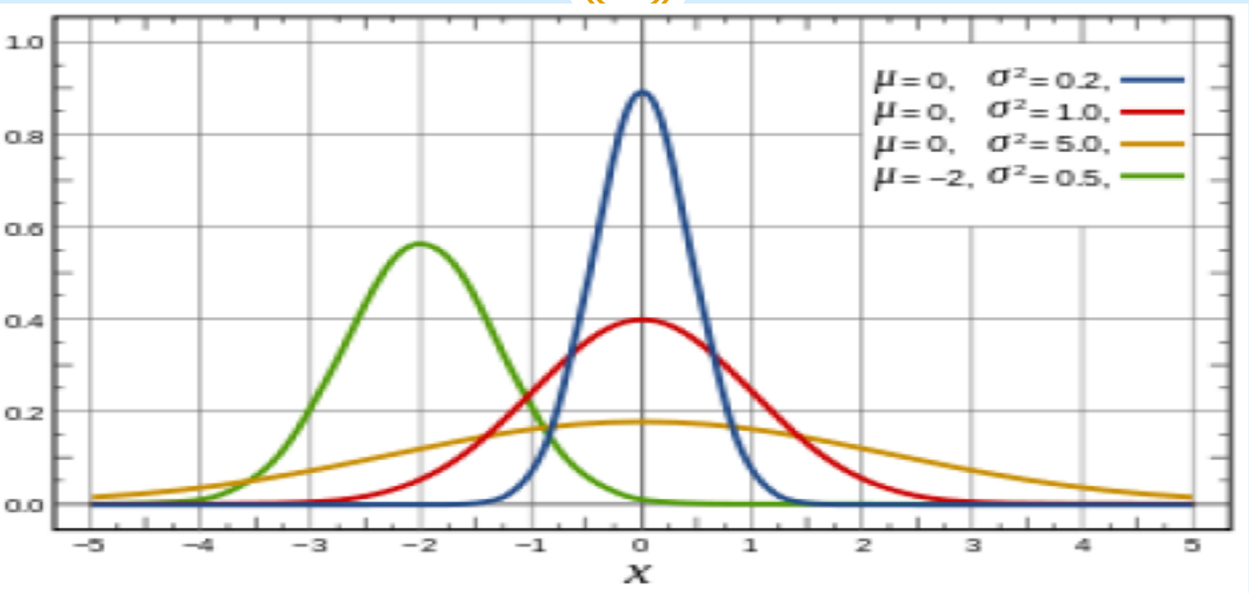
- Just like we can specify a Binomial distribution by n and p, we can specify a normal distribution by μ and σ (μ is mean of normal random variable and σ is standard deviation)
- Formally, X ~ N(μ ,σ^2)
- The curve is pdf of a normal random variable

Let us examine closely

- $X \sim N(\mu, \sigma^2)$
  - A normal random variable can be completely specified by μ ,σ
    - Takes values from - ∞ to +∞
  - Has pdf as specified
  - 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



Many shapes of Normal Distribution



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

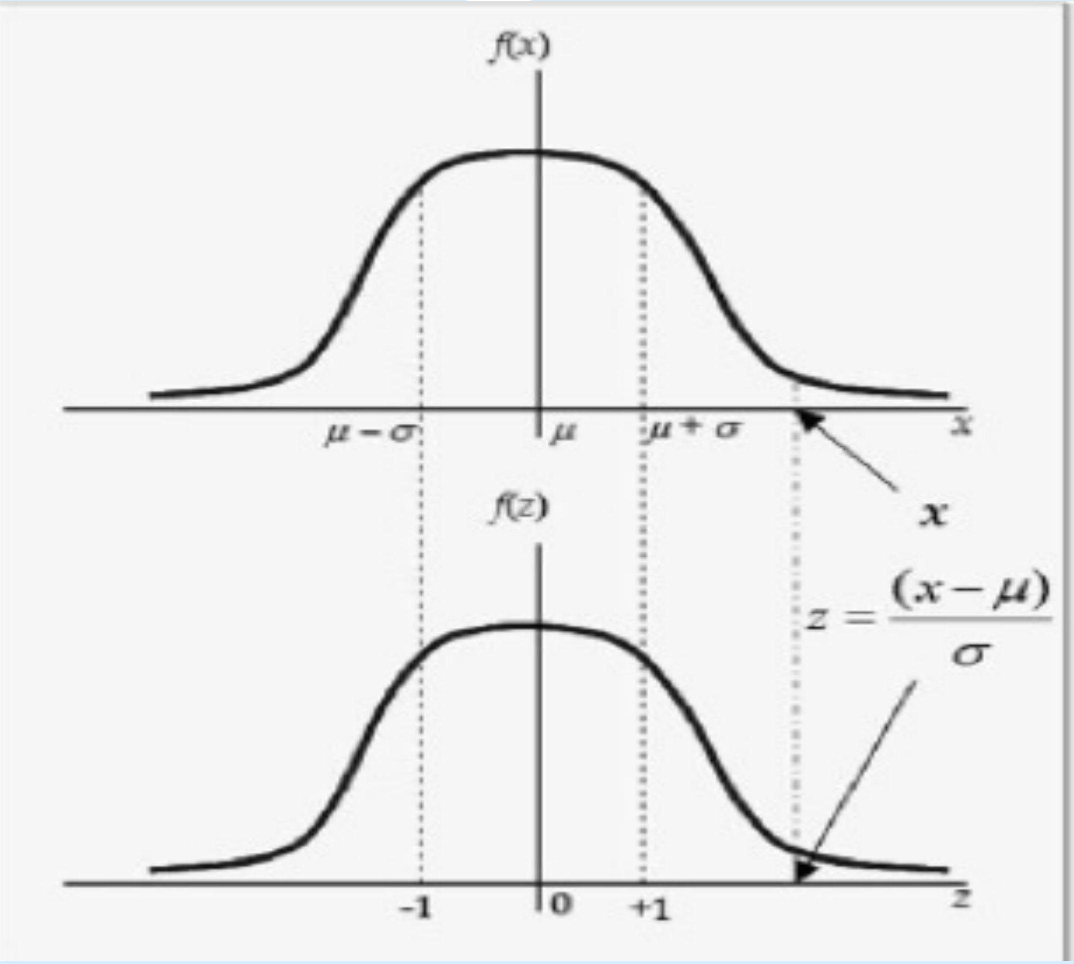
Example

- Height of 95% of AMPBA 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 one sd (likely)
  - For every 1000, 950 should be within two sd (very likely)
  - For ever 1000, 997 within 3 sd (almost certainly)

Standardizing a normal distribution

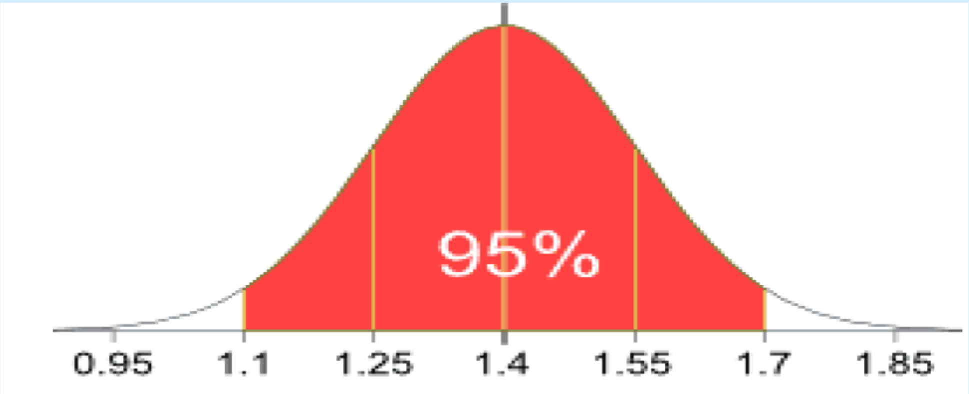
- Assume X1 ~ $N(5, 3^2)$, and X ~ $N(6, 4^2)$
- If you have to compare these two, how would you do that?
- Standardizing comes to rescue
- We define standard normal variable (Z) with mean 0 and sd 1
  - $Z=(X-\mu)/\sigma$
  - Z also has a normal distribution
  - Z score tells position of a point relative to other points in distribution
  - Alternatively, it tells how many s.d. away from mean a point is

Standardizing a Normal Distribution



Example

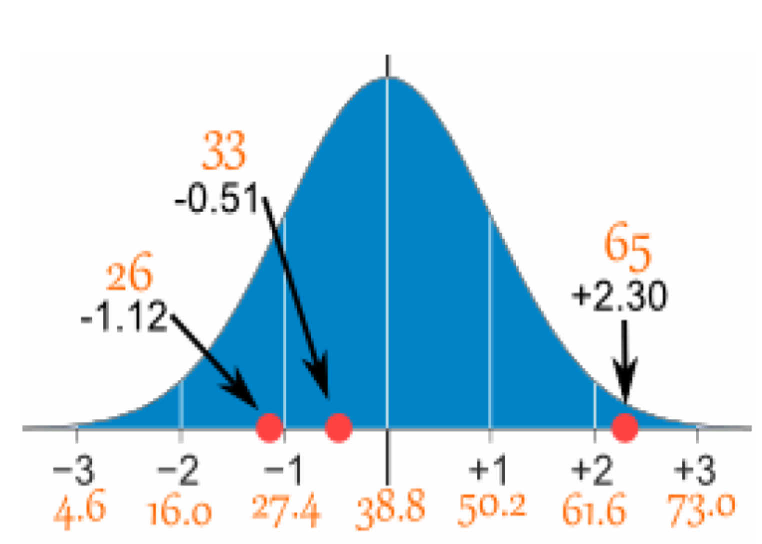
- Assume following distribution for height of students in CBA class.



- 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 $Z=(X-\mu)/\sigma$
  - 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

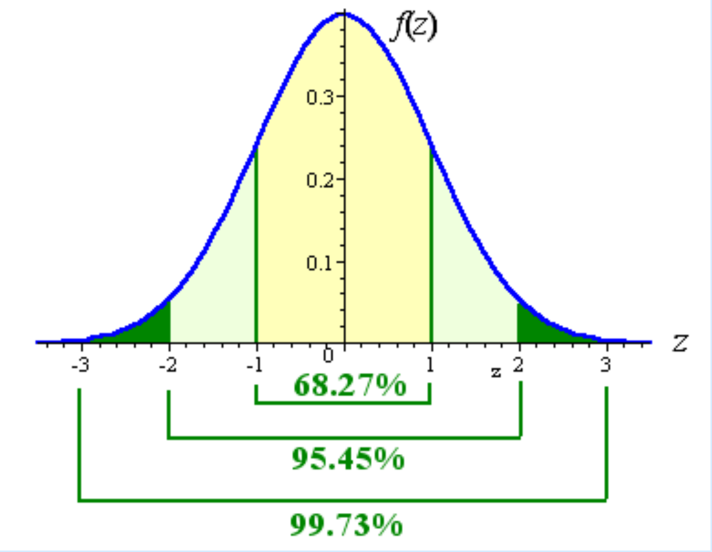


Why do we standardize?

- Helps us compare the data
  - Find area under curve using a table (normal table)
- Imagine X is conducting a test on Prob&Stats for AMPBA students. Maximum marks is out of 60. Consider the sample:
  - 20, 15, 26, 32, 18, 28, 35, 14, 26, 22, 17
  - From the scores, it looks like most of them are going to fail
  - X is kind hearted, and decides test is really hard. So decides to standardize it and only fail if student scores below 2 s.d.
  - Mean: 23. SD: 6.6. Conversion to z-score gives:
  - -0.45, -1.21, 0.45, 1.36, -0.76, 0.76, 1.82, -1.36, 0.45, -0.15, -0.91
  - Thus, no one fails. Z-score saves the day. What if criteria was 0.5 SD? What about 1 SD?

Z-score and probability

- $X \sim N(\mu, \sigma^2) \Rightarrow Z = (X - \mu)/\sigma$
  - => 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)
  - Eg. -1 <= Z <= 1 => 68%
  - -2 <= Z <= 2 => 95%
  - -3 <= Z <= 3 => 99.7%



Standard Normal Table

- Find P(z<=0.56)?
- We refer to standard normal table
- Gives cumulative prob upto z score

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389

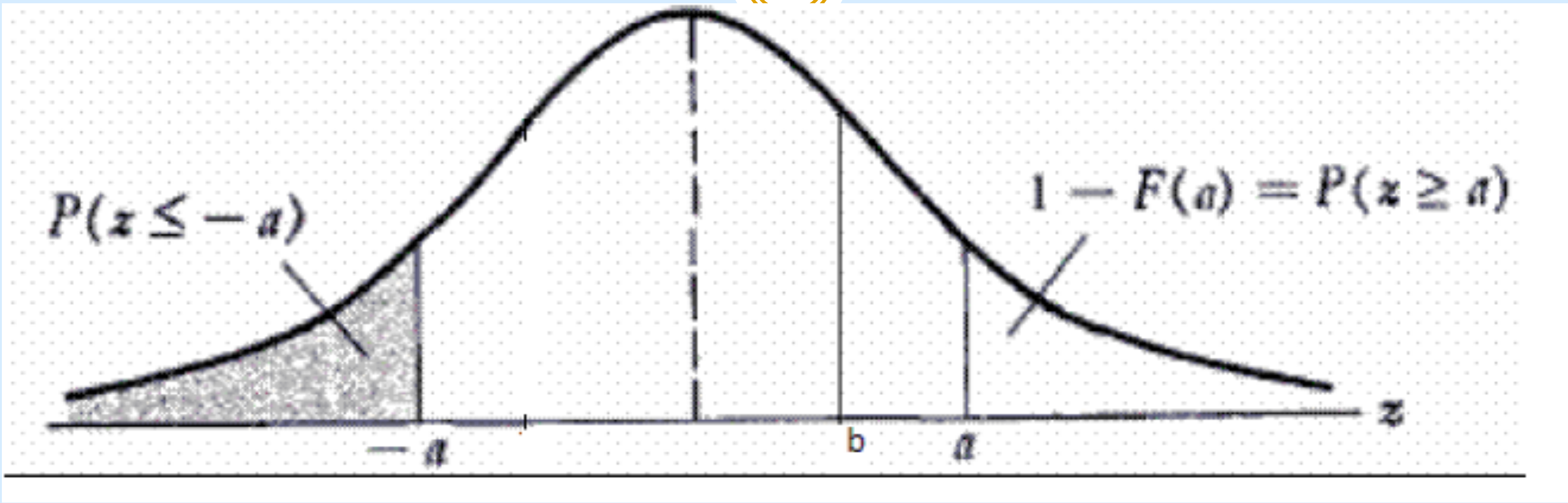
More Complete Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Digging in deeper

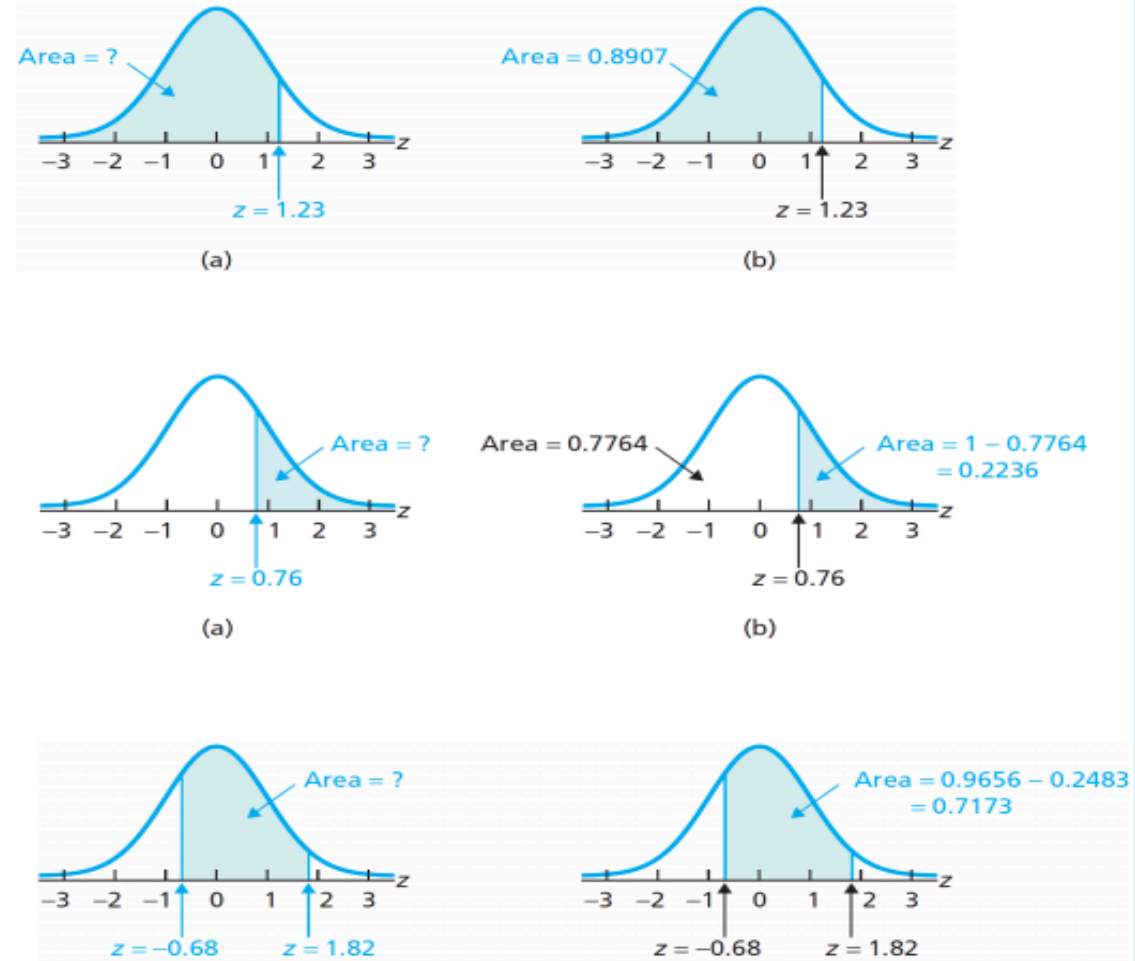
- The table is good for left of standard normal variable
- What about finding right areas?
- We make use of some properties

Properties



- P(z<=a) = F(a)
- P(z>=a) = 1-P(z<=a) = 1-F(a) = F(-a)
- P(z>=-a) = 1-F(-a) = F(a) = P(z<=a)
- P(b<=Z<=a) = P(Z<=a) – P(Z<=b) = F(a) – F(a)

Example



Area = ?
z = 1.23
(a)
Area = 0.8907
z = 1.23
(b)
Area = ?
z = 0.76
(a)
Area = 0.7764
Area = 1 – 0.7764
= 0.2236
z = 0.76
(b)
Area = ?
z = –0.68
z = 1.82
Area = 0.9656 – 0.2483
= 0.7173
z = –0.68
z = 1.82

Example

- A survey indicates that for each trip to the supermarket, a shopper spends an average $\mu=45$ minutes with a standard deviation of $\sigma=12$ minutes. The length of time spent in the store is normally distributed and is represented by the variable x. A shopper enters the store.
  - Find the probability that the shopper will be in the store for each interval of time listed below.
  - If 200 shoppers enter the store, how many shoppers would you expect to be in the store for each interval of time listed below?
  - 1) Between 24 and 54 minutes 2) More than 39 minutes

- Source: http://esminfo.prenhall.com/samplechps/larson/pdfs/ch05.pdf

Example

- The z-scores corresponding to x=24 and x=54 are: Z1= (24-45)/12= -1.75, Z2 = (54-45)/12= .75
- Thus, probability that a shopper will be in the store between 24 and 54 minutes is P(-1.75<=Z<=.75)= F(.75)-F(-1.75)= F(.75)- [1-F(1.75)] = F(.75)+F(1.75)-1 =.7333 (from the standard normal table)
- Another way of interpreting this probability is to say that 73.33% of shoppers will be in the store between 24 and 54 minutes after entering. So if 200 shoppers enter the stop, we expect (200*.7333)=146.66 or 147 shoppers to stay between 24 and 54 minutes.

- The z-score corresponding to 39 mins is Z= (39-45)/12= -.5
- P(Z> (-.5))= 1- P(Z<=.5)= 1-.3085= .6915
- If 200 shoppers enter the store, you would expect 200*(.6915)= 138.3 shoppers to stay in the store for more than 39 minutes

Example

- The amount of fuel consumed by the engines of a jetliner on a flight between two cities is a normally distributed random variable *X with mean of 5.7 tons and standard* deviation of 0.5. Carrying too much fuel is inefficient as it slows the plane. If, however, too little fuel is loaded on the plane, an emergency landing may be necessary. The airline would like to determine the amount of fuel to load so that there will be a 0.99 probability that the plane will arrive at its destination.

- We first find the value of Z such that $P(Z<= z)= .99$. From the standard normal table we see that the value of z corresponding to .99 is 2.33
- Transforming the z value to an x value, we get
- $x = \mu+\sigma z = 5.7 + (0.5) * (2.33) = 6.865$. Thus, the plane should be loaded with 6.865 tons of fuel to give a 0.99 probability that the fuel will last throughout the flight.
- Source: Complete Business Statistics by Aczel-Sounderpandian