Probability Distributions

Ms. Ashiyana M. Makandar
Assistant Professor
Department of Statistics
Vivekanand College, Kolhapur

What is distribution?

- Frequency distribution:
  - Binomial distribution
  - Poisson distribution
  - Normal Distribution

What is probability distribution?

- If rolling a dice, probability of getting every single output is 1/6
- P(1)=1/6
- P(2)=1/6
- P(3)=1/6
- P(4)=1/6
- P(5)=1/6
- P(6)=1/6

Flipping a coin

- P(H)=1/2
- P(T)=1/2
- Flipping two coins together/ flipping one coin two times
- In above case what is the probability of getting zero heads.

1st event	2nd event
H	H
H	T
T	H
T	T

Observation

- When there are both occurrences of tail that means there are not head
- From above observation table ¼ chance not to get head in both evidences
- P(H,H)=1/4 = 0.25
- Probability of getting one head is P(T,H) or P(H,T)
- P(T,H) or P(H,T)=2/4 = 0.50
- P(T,T)=1/4 = 0.25

Glimpse

- If the same flipping coins or rolling dice event occurs many times then calculating probability will become a difficult task.
- The same challenge can be handled using statistical concepts, software and coding.
- The distribution graph will help to see the probability distribution.

Glimpse

- Sum of the area under probability distribution is always 1.
- Total probability is always 1.
- Type of data can be discrete or continuous

Binomial distribution

Properties

- More and more experiments should be happened.
- Each trial can result in just two possible outcome. Like success or failure , sing in or log out, accepted or rejected etc…..
- Probability of success is denoted by p
- Every trial is independent. That means it is not dependant on previous outcome.

Binomial distribution

- Simply count number of event and Outcome of experiment is in two forms.
- Gives probability of success out of total number of experiments
- If P(w)=0.75 is probability for winning
  - P(l)=0.25 is probability for loosing
- Then out of 5 events what is the probability to win ?

Binomial

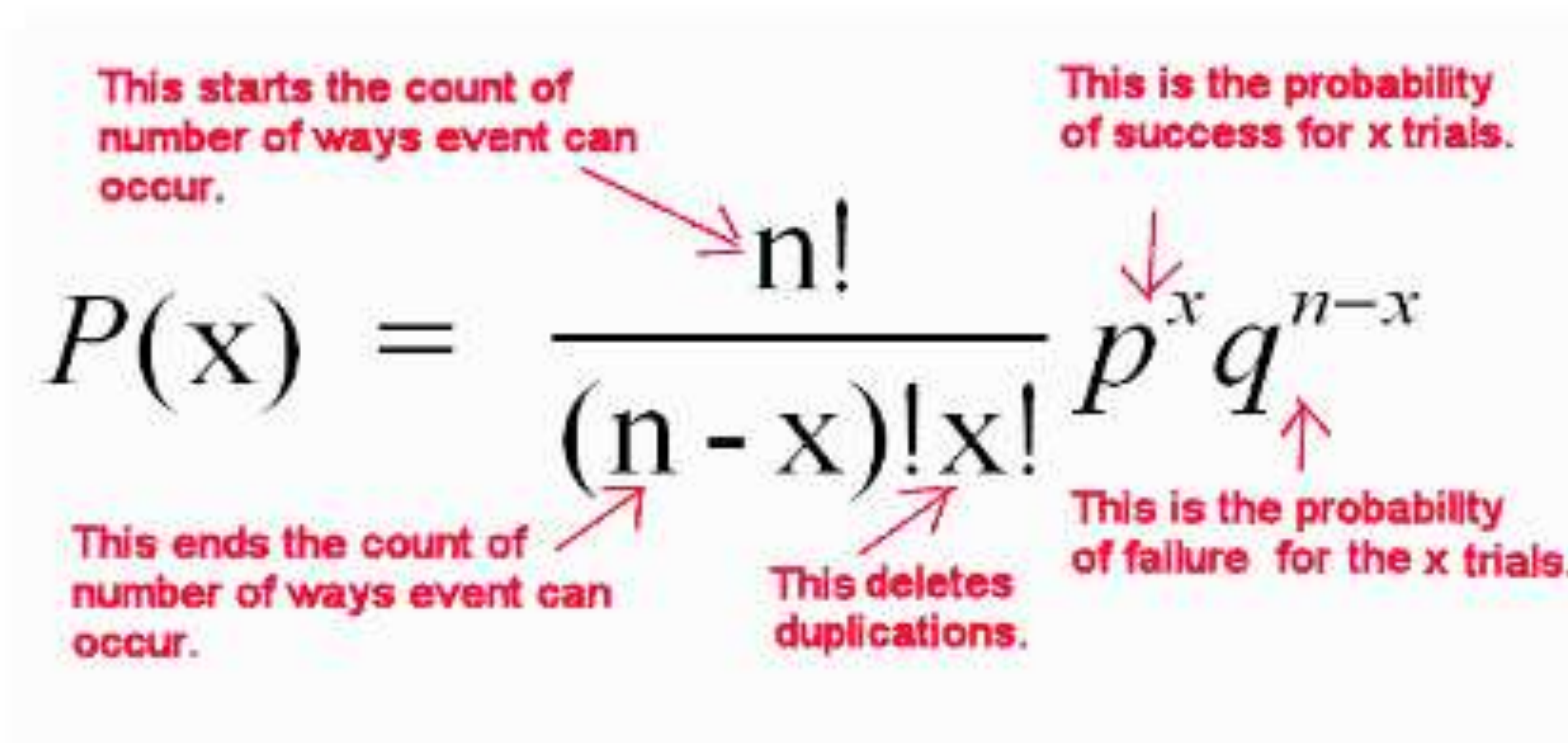
- K=number of successful event
- Probability of successes
- Out of 5 events what is the chance of success?

$$f(k;p) = p^k(1-p)^{1-k} \quad \text{for } k \in \{0,1\}$$

Formula for binomial probability

- P(x)=nc_x * p^x * (1-p)^(n-x)
- P(x)= (n!/(x!(n-x)!)) * P^x *(1-p) ^(n-x)
- X:number of success result from binomial experiment.
- N: the number of trials in the binomial experiment.
- p: The probability of success on an individual trial
- P(x) the probability that n trial binomial experiments results in exactly x successes, when the probability of success on an individual trial is p.

Formula


This starts the count of number of ways event can occur.
This is the probability of success for x trials.
$P(x) = \frac{n!}{(n-x)!x!} p^x q^{n-x}$
This ends the count of number of ways event can occur.
This deletes duplications.
This is the probability of failure for the x trials.

Example -1

- If we are flipping 4 coins then what is the probability of getting 1 head?
- P(x)= (n!/(x!(n-x)!)) * P^x *(1-p) ^(n-x)
- P(1)= (4!/ 1! * 3!) * 0.5*1 * (1- 0.5) ^(4-1)
- P(1) = 4 * 0.5 * 0.5^3
- P(1)=4* 0.5^4 = 0.25

Example-2

- If win is 0 means losing of 5 matches ie. [L,L,L,L,L]
- Probability is (0.25)**5
- If 1 win event then Probability = (0.75)*(0.25)**4
- Hence we can say
- P(x=3)+P(x=4)+P(x=5)
- The same can be done as
- 5C3*(0.75)**3*(0.25)**(5-3)
- Outcome is the probability to win the series.

Example-3

- Suppose a toy production company has 12% defect rate. Buyer decided to purchase after testing random 20 toys. Buyer will accept this if there are 2 or less defective toys. What is the probability of getting accepted?
- p=0.12 , n = 20, x = 0,1,2
- P(0) = 20!/(0!*(20-0)!) * 0.12^0 * (1-0.12)^(20-0)
- P(1) = 20!/(1!*(20-1)!) * 0.12^1 * (1-0.12)^(20-1)
- P(2) = 20!/(2!*(20-2)!) * 0.12^2 * (1-0.12)^(20-2)
- P(o)= P(0)+p(1)+p(2) =0.563132

outcome

- Once we are having binomial distribution, we can find out mean, variance and standard deviation of the distribution.
- Mean= $n*p$
- Variance = $n*p*(1-p)$
- Standard deviation = $sqrt(n*p*(1-p))$

Poisson distribution

Binomial vs. Poisson

Binomial	Poisson
Similarities:	
1) Used for discrete distribution	
2) Measures the number of success	
Differences	
Possibility of success is finite	Possibility of success are infinite
e.g. out of 20 defective toys 20 can be defective but not more than 20	e.g. How many people are purchasing movie tickets

Properties

- Experiment results in outcomes that can be classified as success or failure, true or false, yes or no etc…
- Average number of success in the specific region is known.
- Outcomes are random. Occurrence of one outcome does not influence the chance of another outcome of interest.
- The outcome of interest is rare relative to the possible outcomes
- E.g. Queue at the counter, road accident, etc

Poisson's distribution

- outcome is in count, time or range
- Poisson’s distribution is used when we want probability between time period.
- E.g. how many of students will be placed?
- How many fraud cases except to see?

- Suppose there are 10 cars/ hrs in garage then what is the probability that there are 100 cars/day
- So here, we are considering hrs to calculate the probability for day.
- e: a constant equal to 2.71828
- μ: the number of success occurred in specific time interval or specific region
- X: is number of success
- P: P(x; μ): The Poisson probability that exactly x success occur in a poisson experiment when the mean number of success is μ.

$$\left\{ P = \frac{e^{-\mu} \times \mu^{x}}{x!} \right\}$$

Example 1

- On a booking counter on the average 3.6 people come every 10 minutes on weekends. What is the probability of getting 7 people in 10 minute?
- μ= 3.6
- X= 7
- P(X, μ)=(e^-3.6)* (3.6^7)/7!
- P(X, μ)=0.02732*7836.41/5040
- P(X, μ)=0.0424
- This means 4% chance of getting 7 people in 10 minute

Example-2

- Real estate company is selling 2 flats/ day what is the probability that company will sell 3 flats/days
- Here $\mu = 2$ and $x = 3$ is the goal
- After using the formula is gives probability as 0.18

Observation

- The mean of distribution is μ
- The variance is equal to μ

Normal distribution

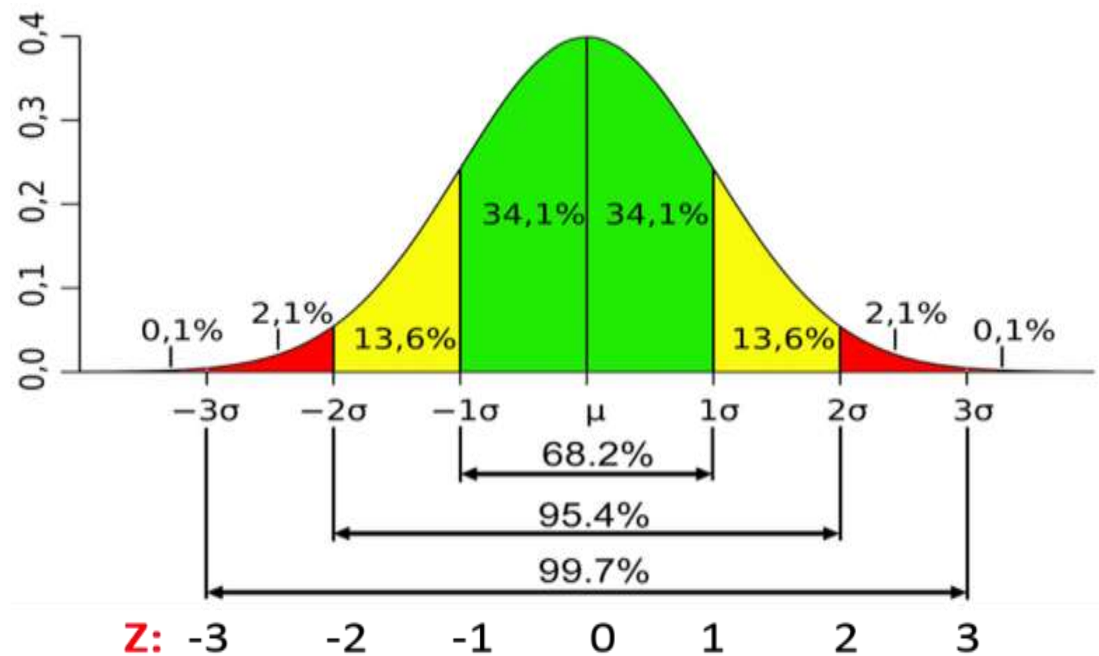
Properties

- Continuous data
- Normal distribution is symmetric
- Long tail, bell shaped
- Mean, median/mode are same
- Shape of the curve defined by
  - Mean
  - Standard deviation

Properties

- If we have skewness in data set tail of the data is not distributed evenly.
- The probability of any particular value is 0
- The probability that X is greater than or less than a value = area under the normal curve in that direction
- For normal distribution of the data
  - mean and variance could be any value
  - range is from -∞ to + ∞
  - Entire area = 1

- Normal distribution shows the bell shape curve



Normal Distribution

- Most occurring distribution
- Standard normal distribution is also known as *z-distribution*

- Equal amount of data on both sides then we call is as balanced data set
- ML algorithm is trained based on normally distributed algorithm
- Formula to find probability of single value is

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

Formula explained

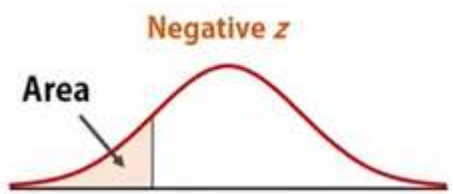
- μ: mean
- σ: standard deviation
- π: 3.14159
- e: approximately 2.71828
- For z-distribution $\mu = 0$ and $\sigma = 1$

Z-value

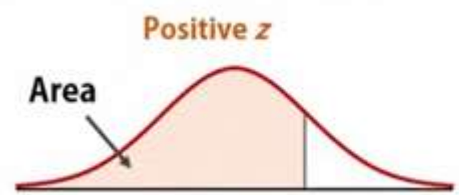
- Z-value is also known as standard score
- $Z = (x-\mu)/\sigma$
- A bottle is having average volume as 150cc and the standard deviation is 2 cc. What percentage of bottles will have volume more than 153 cc
- μ =150cc
- σ = 2cc
- Z=(153-150)/2 = 1.5
- Hence z-score is 1.5
- Probability of area selection is 0.9332
- More than 153 bottles are 1 – 0.9332 = 0.06680 = 6.6%

Z score

- Normal distribution calculating probability
- Area under normal curve is 1



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294



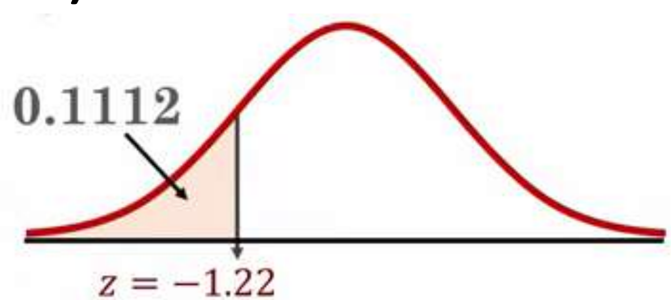
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	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	0.7852
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

Example-1

- Scores in an exam are normally distributed with mean 65 and standard deviation of 9. find the percent of the score.
  – Less than 54
  – At least 80
  – Between 70 and 86

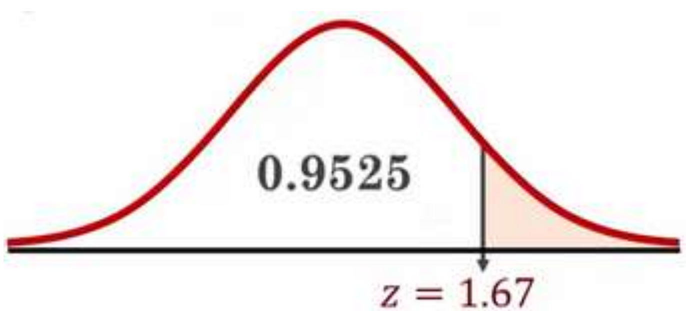
Less than 54

- This means $P(x<54)$
- $Z= (x-\mu)/\sigma = (54-65)/ 9 = -1.2222$
- Check in the negative z-score for row with -1.2 and column 0.02
- Corresponding area is 0.1112 i.e. Area to the left of z=-1.222 is 0.1112
- $P(x<54)= P(z< -1.222) = 0.1112 =11.12\%$



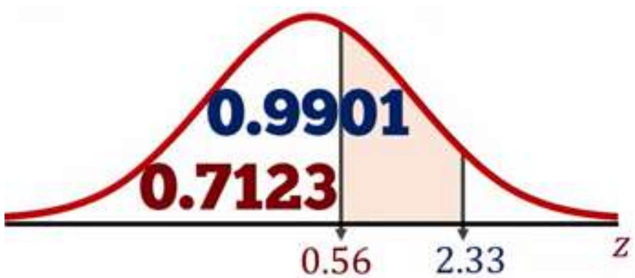
At least 80

- P(x>=80) = P(x>80)
- Z= (x-μ)/σ = (80-65)/9 = 1.67
- In z table, check 1.6 against 0.07 equal to 0.9525
- P(x>80)> P(z>1.67) = 1- 0.9525 = 0.0475 = 4.75%



Between 70 and 86

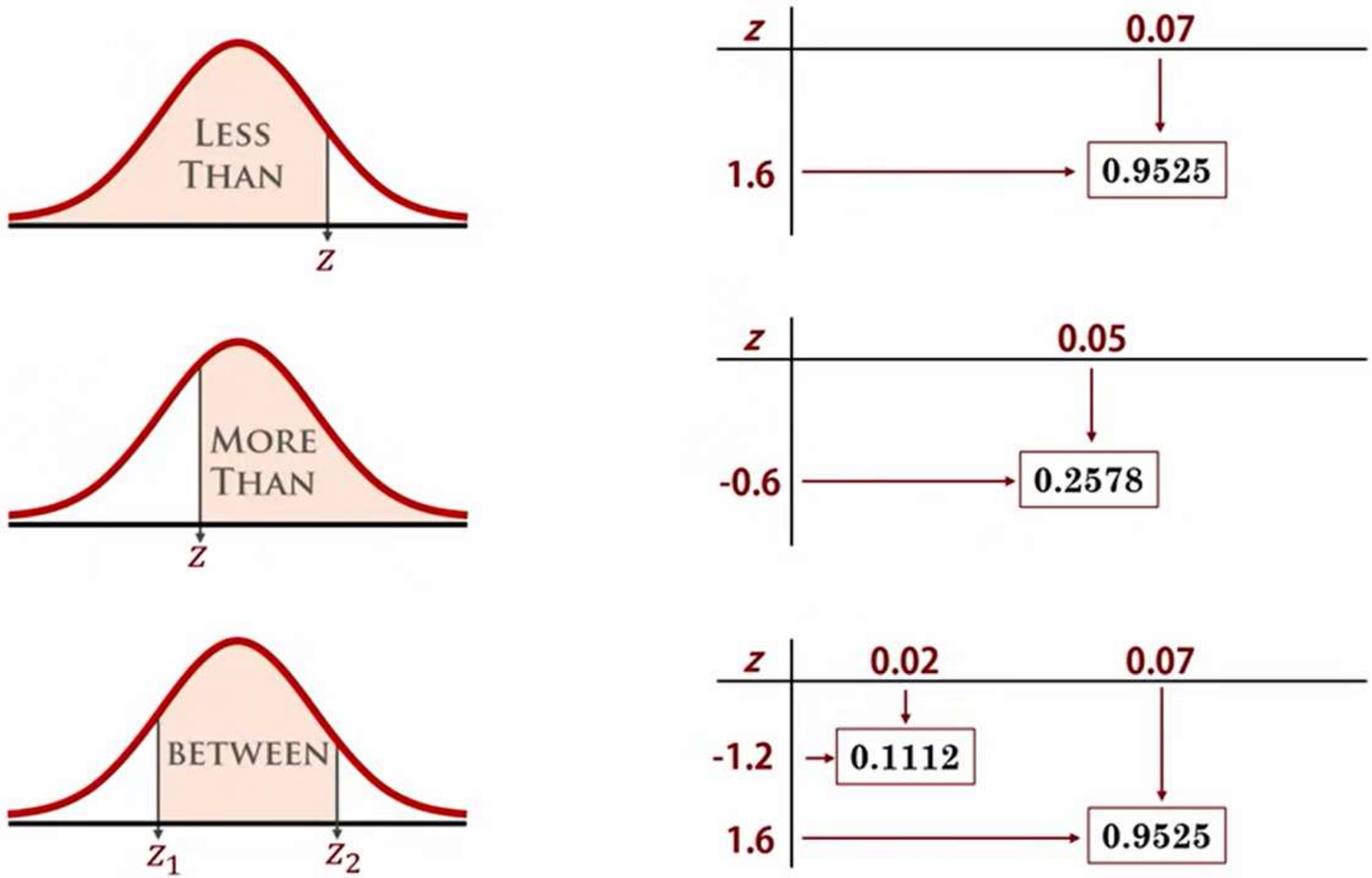
- Z= (x-μ)/σ
- For x=70
  - Z=(70 – 65)/9 = 0.56
  - In z table, check 0.5 against 0.06 equal to 0.7123
- For x = 86
  - Z= (86-65)/9 = 2.33
  - In z table, check 2.3 against 0.03 equal to 0.991



Between 70 and 86

- Finding the are between two z values from cumulative table is subtracting from large amount
- $P(70 < x < 85) = P(0.56 < z < 2.33)$
- $= P(z < 2.33) - P(z < 0.56)$
- $= 0.9901 - 0.7123$
- $= 0.2778 = 27.78\%$

Summary


LESS THAN
z
MORE THAN
z
BETWEEN
z1
z2
z
0.07
1.6
0.9525
z
0.05
-0.6
0.2578
z
0.02
0.07
-1.2
0.1112
1.6
0.9525

Example 2

- Perfume bottle filled with the average volume of 150 cc and the standard deviation of 2 cc.
- What percentage of bottles will have the volume between 148 and 152 cc?
- $Z= (x-\mu)/\sigma$
- Here $\mu = 150$ cc and $\sigma= 2$cc
- $Z1=(148-150)/2 = -1$
- $Z2 =(152 – 150)/2 = 1$
- $P(x)=1 – 0.1587 – 0.1587 = 0.6826$

Thank you