Fundamentals of Business Statistics

Probability Distribution

Dr. P.K.Viswanathan

What is Probability Distribution?

In precise terms, a **probability distribution** is a total listing of the various values the random variable can take along with the corresponding probability of each value. A real life example could be the pattern of distribution of the machine breakdowns in a manufacturing unit.

- The random variable in this example would be the various values the machine breakdowns could assume.

- The probability corresponding to each value of the breakdown is the relative frequency of occurrence of the breakdown.

- The probability distribution for this example is constructed by the actual breakdown pattern observed over a period of time. Statisticians use the term "observed distribution" of breakdowns.

Binomial Distribution

- The Binomial Distribution is a widely used probability distribution of a discrete random variable.

- It plays a major role in **quality control** and **quality assurance** function. Manufacturing units do use the binomial distribution for **defective** analysis.

- Reducing the number of defectives using the proportion defective control chart (p chart) is an accepted practice in manufacturing organizations.

- Binomial distribution is also being used in **service organizations** like banks, and insurance corporations to get an idea of the proportion customers who are satisfied with the service quality.

Conditions for Applying Binomial Distribution (Bernoulli Process)

- Trials are independent and random.
- There are fixed number of trials (n trials).
- There are only two outcomes of the trial designated as *success* or *failure*.
- The probability of success is uniform through out the n trials

Binomial Probability Function

The probability of getting x successes out of n trials is indeed the definition of a Binomial Distribution. The Binomial Probability Function is given by the following expression

$$P(x) = \binom{n}{x} \pi^{x} (1-\pi)^{n-x}$$

x can take values 0, 1, 2, ………, n

Where P(x) is the probability of getting x successes in n trials

$\binom{n}{x}$ is the number of ways in which x successes can take place out of n trials

$$= \frac{n!}{x!(n-x)!}$$

π is the probability of success, which is the same through out the n trials.

π is the parameter of the Binomial distribution

Example for Binomial Distribution

A bank issues credit cards to customers under the scheme of Master Card. Based on the past data, the bank has found out that 60% of all accounts pay on time following the bill. If a sample of 7 accounts is selected at random from the current database, construct the Binomial Probability Distribution of accounts paying on time.

Spreadsheet Showing the Solution

	A	B	C	D	
1		Example Problem-Master Card			
2					
3	x	P(x)	Cumulative		
4			Probability		
5	0	0.0016384	0.0016384		
6	1	0.0172032	0.0188416		
7	2	0.0774144	0.0962560		
8	3	0.1935360	0.2897920		
9	4	0.2903040	0.5800960		
10	5	0.2612736	0.8413696		
11	6	0.1306368	0.9720064		
12	7	0.0279936	1.0000000		

Mean and Standard Deviation of the Binomial Distribution

The mean μ of the Binomial Distribution is given by $\mu = E(x) = n\pi$

The Standard Deviation σ is given by

$$\sigma = \sqrt{n\pi(1-\pi)}$$

For the example problem in the previous two slides, Mean $= 7 \times 0.6 = 4.2$.

Standard Deviation $= \sqrt{4.2(1-0.60)} = 1.30$

Poisson Distribution

Poisson Distribution is another discrete distribution which also plays a major role in **quality control** in the context of reducing the number of defects per standard unit.

- Examples include number of defects per item, number of defects per transformer produced, number of defects per 100 m^2 of cloth, etc.

- Other real life examples would include 1) The number of cars arriving at a highway check post per hour; 2) The number of customers visiting a bank per hour during peak business period.

Poisson Process

- The probability of getting exactly one success in a continuous interval such as length, area, time and the like is constant.

- The probability of a success in any one interval is independent of the probability of success occurring in any other interval.

- The probability of getting more than one success in an interval is 0.

Poisson Probability Function

Poisson Distribution Formula

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

where

$P(x)$ = Probability of x successes given an idea of λ

λ = Average number of successes

e = 2.71828(based on natural logarithm)

x = successes per unit which can take values 0, 1, 2, 3,…………..∞

λ is the Parameter of the Poisson Distribution.

Mean of the Poisson Distribution is = λ

Standard Deviation of the Poisson Distribution is = $\sqrt{\lambda}$

Example

If on an average, 6 customers arrive every two minutes at a bank during the busy hours of working, a) what is the probability that exactly four customers arrive in a given minute? b) What is the probability that more than three customers will arrive in a given minute?

6 customers arrive every two minutes. Therefore , 3 customers arrive every minute. That implies my lambda=3

$P(X=4)=?$

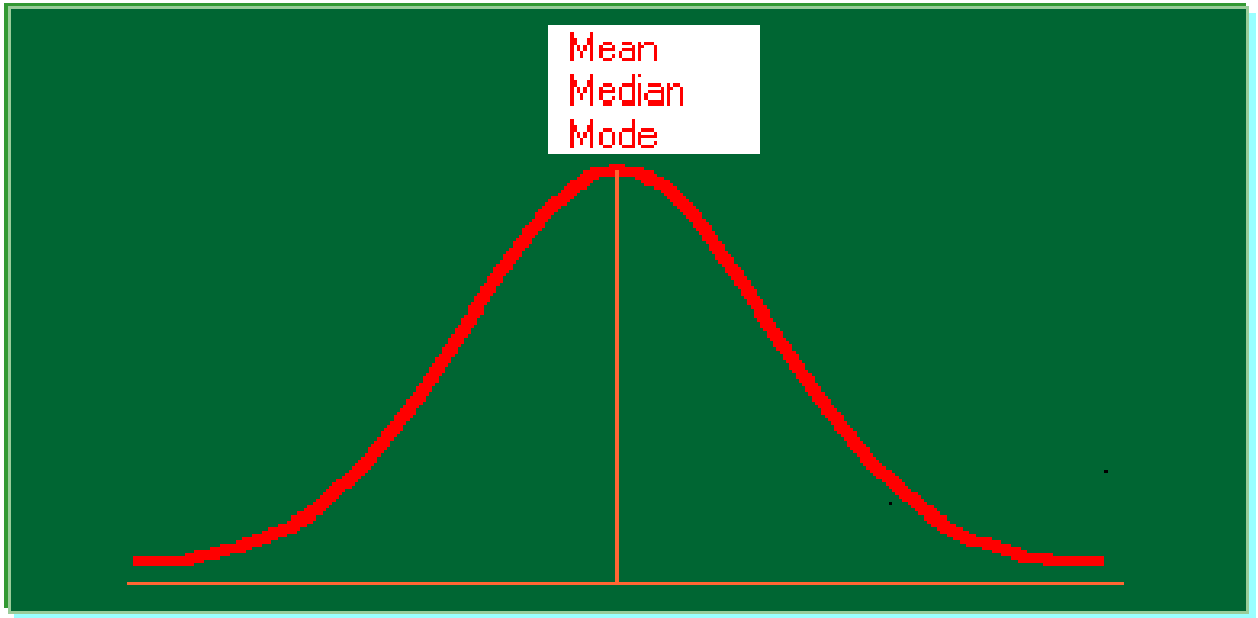
$P(X>3)=?$ Implies $1-P(X<=3)$

Spreadsheet showing Solution

	A	B	C	D	E	F	G	H
1			Poisson Distribution for the Example(x rages from 0 to 10)					
2						$\lambda = 3$		
3				Cumulative				
4		x	P(x)	Probability				
5		0	0.049787	0.049787				
6		1	0.149361	0.199148				
7		2	0.224042	0.423190				
8		3	0.224042	0.647232				
9		4	0.168031	0.815263				
10		5	0.100819	0.916082				
11		6	0.050409	0.966491				
12		7	0.021604	0.988095				
13		8	0.008102	0.996197				
14		9	0.002701	0.998898				
15		10	0.000810	0.999708				

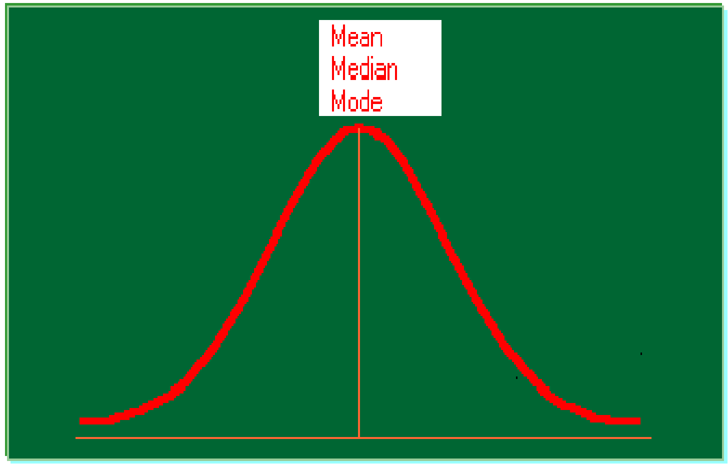
Look at the table and you will find that corresponding to x =4 in column B, the probability P(x=4) is 0.168031 found in column C. You want P(x > 3) = 1- P(x < = 3) = 1-0.647232 = 0.352768 (Please note that P(x< = 3) is found in Column D corresponding to the value of x =3 in column B. You could see 0.647232 in row 8 and column D.

Normal Distribution



Properties of Normal Distribution

greatlearning



- The normal distr … ooking like a bell. Statisticians use the expression '

- It is a beautiful distribution in which the mean, the median, and the mode are all equal to one another.

- It is symmetrical about its mean.

- If the tails of the normal distribution are extended, they will run parallel to the horizontal axis without actually touching it. (asymptotic to the x-axis)

- The normal distribution has two parameters namely the mean μ and the standard deviation σ

Normal Probability Density Function

In the usual notation, the probability density function of the normal distribution is given below:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]}$$

x is a continuous normal random variable wit[illegible] the property $-\infty < x < \infty$ meaning x can take all real numbers in the interval $-\infty < x < \infty$.

Standard Normal Distribution

The Standard Normal Variable is defined as follows:

$$Z = \frac{X - \mu}{\sigma}$$

Please note that Z is a pure number independent of the unit of measurement. The random variable Z follows a normal distribution with mean=0 and standard deviation =1.

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\left[\frac{Z^2}{2}\right]}$$

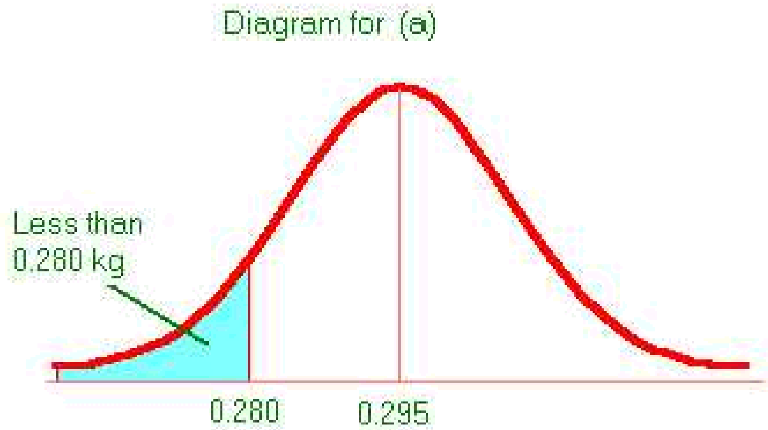
Example Problem

The mean weight of a morning breakfast cereal pack is 0.295 kg with a standard deviation of 0.025 kg. The random variable weight of the pack follows a normal distribution.

a) What is the probability that the pack weighs less than 0.280 kg?
b) What is the probability that the pack weighs more than 0.350 kg?
c) What is the probability that the pack weighs between 0.260 kg to 0.340 kg?

Solution-a)

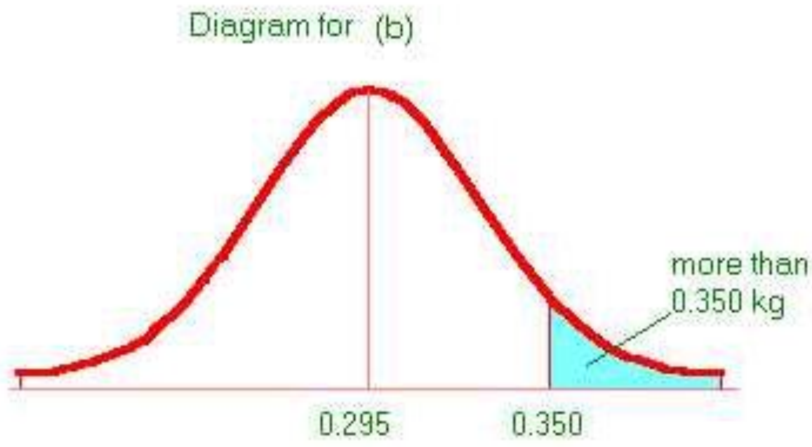
greatlearning



$z = \frac{x - \mu}{\sigma}$ = (0.280-0.295)/0.025 = -0.6. Click "Paste Function" of Microsoft Excel, then click the "statistical" option, then click the standard normal distribution option and enter the z value. You get the answer. Excel accepts directly both the negative and positive values of z. Excel always returns the cumulative probability value. When z is negative, the answer is direct. When z is positive, the answer is =1- the probability value returned by Excel. The answer for part a) of the question = 0.2743(Direct from Excel since z is negative). This says that 27.43 % of the packs weigh less than 0.280 kg.

Solution-b)

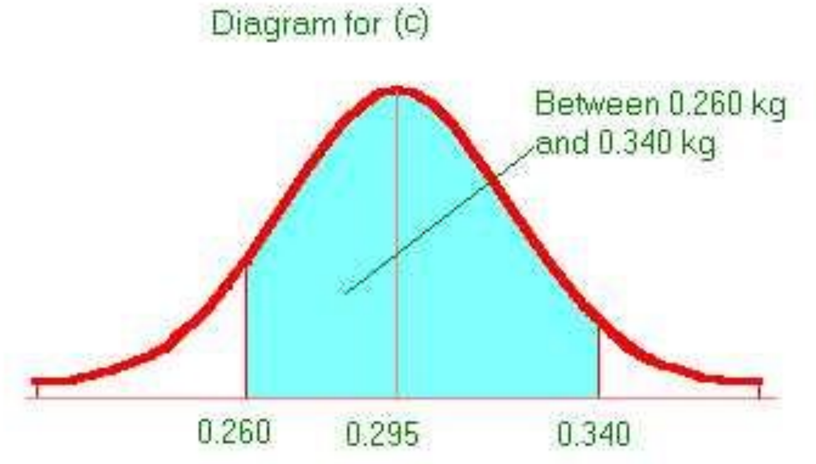
greatlearning



$z = \frac{x - \mu}{\sigma}$ = (0.350-0.295)/0.025 =2.2. Excel returns a value of 0.9861. Since z is positive, the required probability is = 1-0.9861 =0.0139. This means that 1.39% of the packs weigh more than 0.350 kg.

Solution-c)

greatlearning



For this part, you have to first get the cumulative probability up to 0.340 kg and then subtract the cumulative probability up to 0.260. $z = \frac{x - \mu}{\sigma}$ = (0.340-0.295)/0.025 =1.8(up to 0.340). $z = \frac{x - \mu}{\sigma}$ = (0.260-0.295)/0.025 = -1.4(up to 0.260). These two probabilities from Excel are 0.9641 and 0.0808 respectively. The answer is = 0.9641-0.0808 = 0.8833. This means that 88.33% of the packs weigh between 0.260 kg and 0.340 kg.

Thinking Problem

- What weight of the pack produced would have been exceeded by 90% of the packs produced?