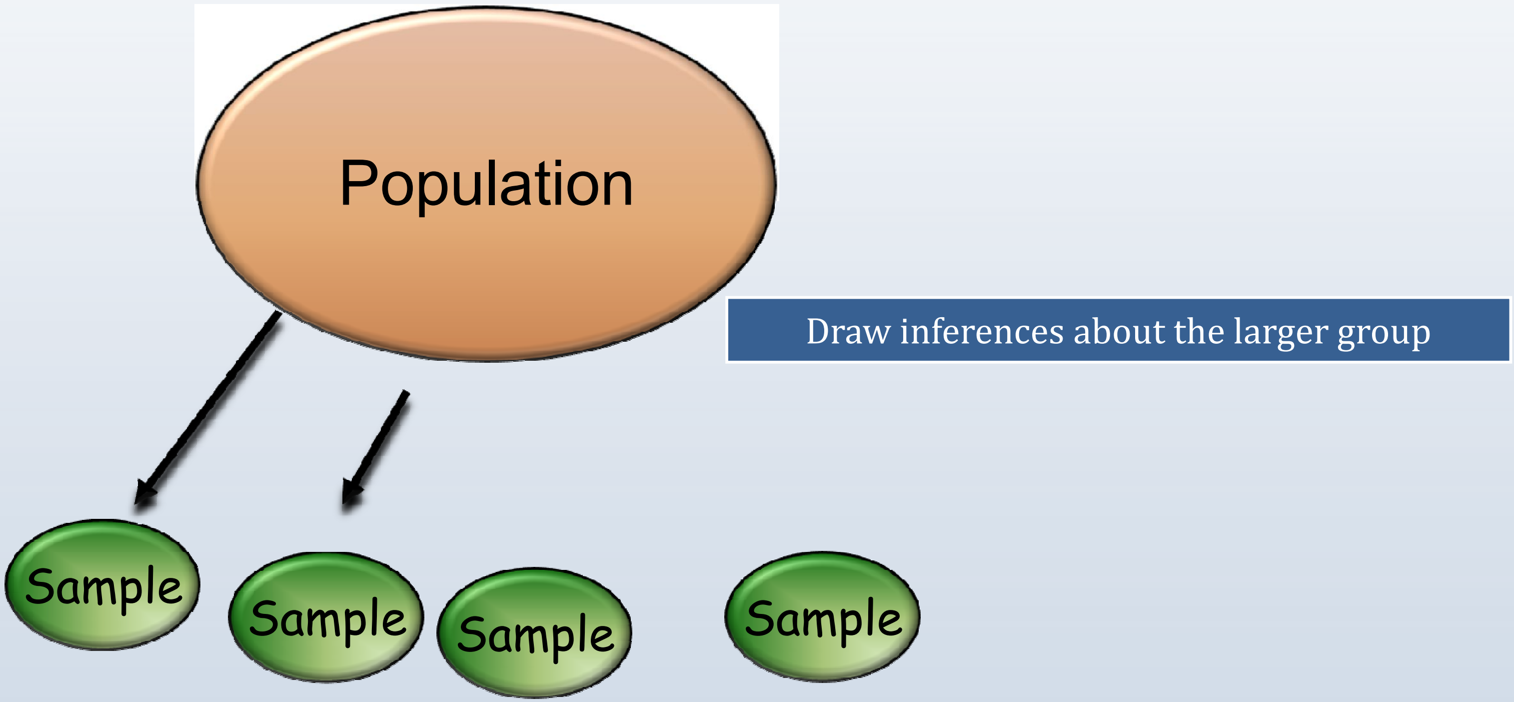
DS Part I I - Inferential Statistics

Sasken Training, Adyar , Chennai – 600 020

Agenda

- Introduction to Inferential Statistics
- Probability
- Binomial Distribution
- Negative Binomial Distribution
- Poison Distribution
- Geometric Distribution
- Hyper geometric Distribution
- Normal Distribution
- Sampling ,Sample Distributions and CLT

- Inferential statistics entail **drawing conclusion about a parameter** (e.g., industry a client works in) **from the sample** (e.g., of clients) to the **population** (all clients). However, the sample must be representative of the population for accuracy.



Descriptive Vs Inferential Statistics

Descriptive Statistics

Organize
Summarize
Simplify
Presentation of data

Inferential Statistics

Generalize from samples to Populations
Hypothesis Testing
Relationship among Variables

Describing data

Make predictions

values that describe the characteristics of population

values that infer results of a sample to the population from which the sample is drawn

Population : Complete set of individuals
Sample : Subset of a Population

Statistics come from Samples (sample statisitc is referred to as the point estimator of corresponding population parameter)
And **P**arameters come from Populations (a characteristic of a population)

	Sample	Population
Mean	X bar	mu
Standard Deviation	s	sigma

Parameter
Statistic

What kind of inferences can we make in statistics ?

1. We can estimate unknown population parameters based on properties of A sample
2. We can test hypothesis about a population based on sample parameters

Samples and Population

Sample Vs Population

Why do I need to separate these two ?

- It's cumbersome to gather population data as its very very large , making it difficult(or impossible) to collect
- Its easier to take a subset of the population, and then make inferences about the population

Sample Vs Population

We can draw as many samples as from a population, a good sample is that which is chosen

Without bias: (Not choosing only high income respondents)

Complete coverage : All segments in the population are correctly represented

Non response inclusive : if 20% of your population are defaulters, your sample ideally should also have 20% defaulters

Sampling method

Probability Sampling

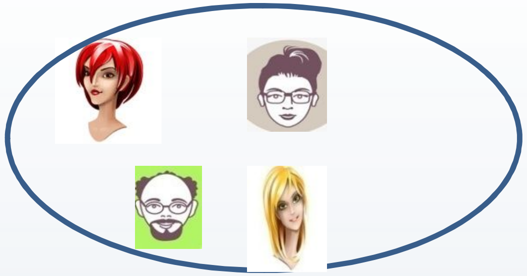
– Simple random Sampling
– Systematic
– Cluster
– Stratified

Non-Probability Sampling

- Convenience Sampling
- Quota sampling

Simple Random Sampling

Employees



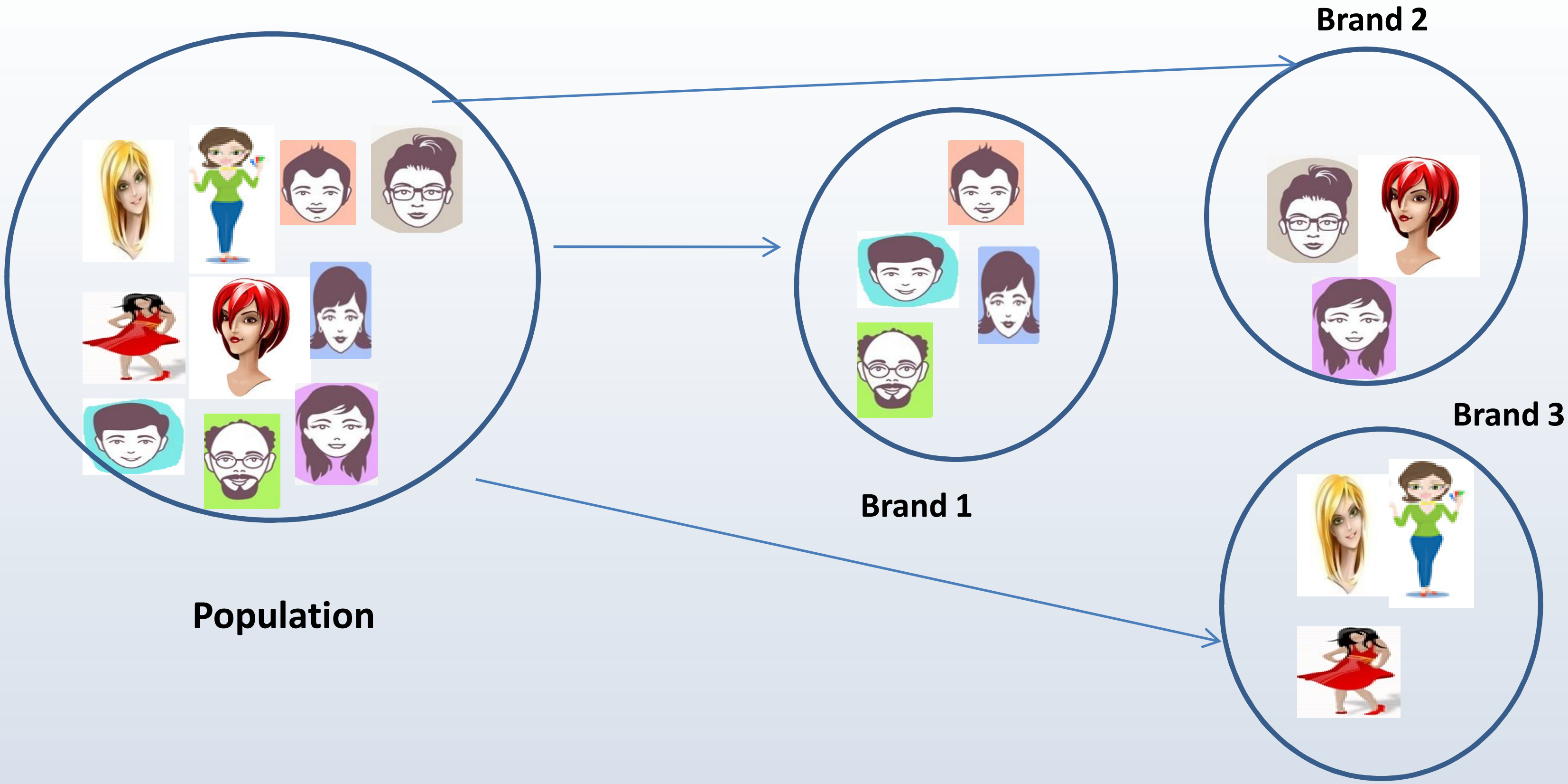
Sample of employees

Consists of n individuals from the population chosen in such a way that every individual has a equal chance of being selected

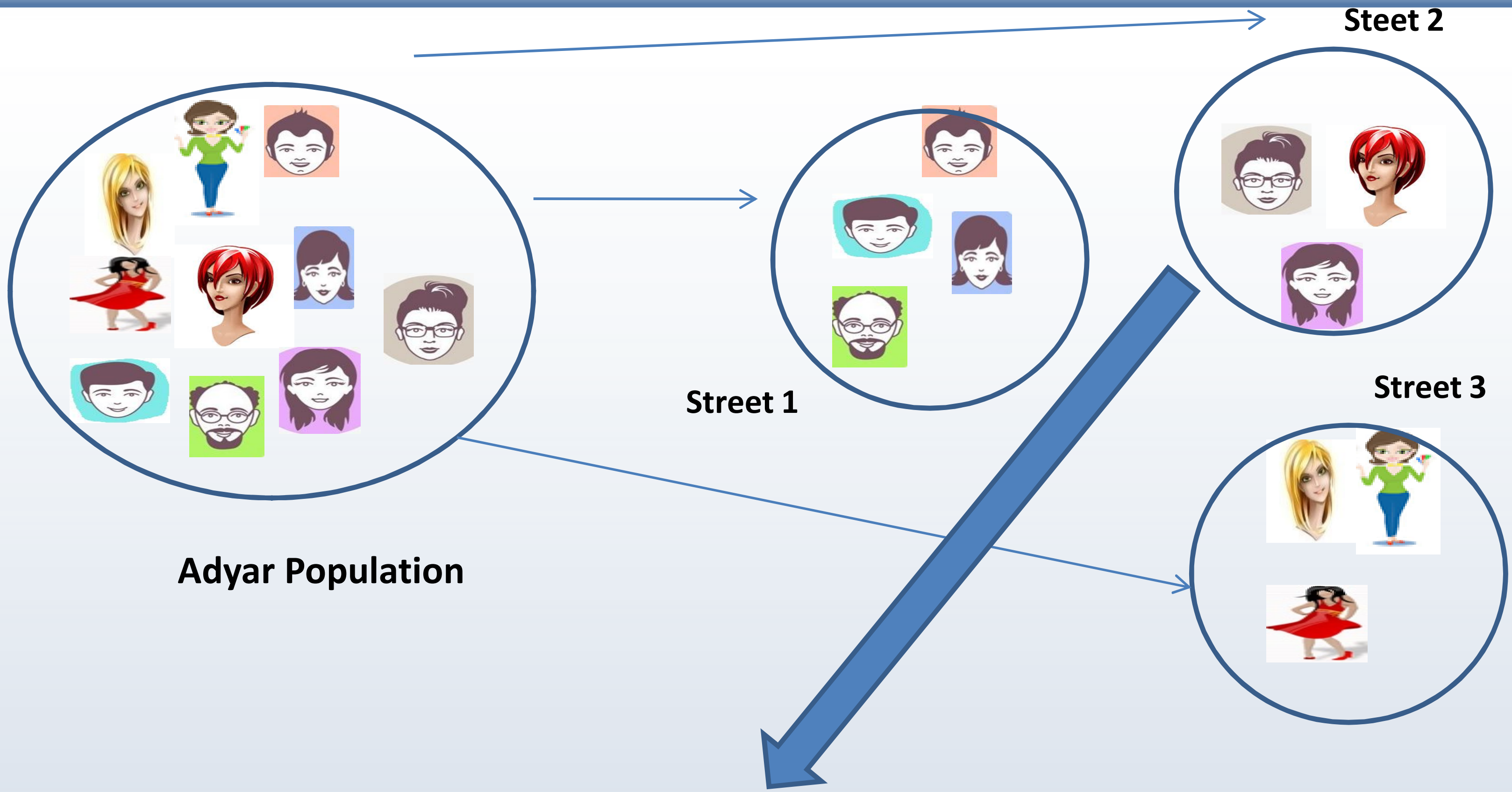
<u>Advantages</u>
Unbiased
Easy

<u>Disadvantages</u>
Large variance
May not be representative
Must have sampling frame (list of population)

Stratified Sampling



Cluster Sampling



Choose one street randomly and run the survey with all individuals

A cluster sample is obtained by selecting all the individuals , within a randomly collected group

Bootstrap

A method of Resampling :

Creating many samples from a single sample. Estimates the population distribution by using the information based on a number of re samples from the simple Estimation of the standard error of your statistic of interest.

The bootstrap is a general method for doing statistical analysis without making strong parametric assumptions

Quiz

- Our data come from _______, but we really care most about ______.

  theories; mathematical models

  samples; populations

  populations; samples

  subjective methods; objective methods

Quiz

- A random sample

  is more likely to be representative of the population than any other kind of sample.

  is always representative of the population.

  allows you to directly calculate the parameters of the population.

  all of the above are true.

  all of the above are false.

Quiz

- Which of the following is (are) true? Using a random sample

  is to accept some uncertainty about the conclusions.

  enables you to calculate statistics.

  biases your results.

Quiz

- A random sample is one

  that is haphazard.

  that is unplanned.

  in which every sample of a particular size has an equal probability of being selected.

  that ensures that there will be no uncertainty in the conclusions.

Quiz

- A biased sample is one that

  is too small.

  will always lead to a wrong conclusion.

  will likely have certain groups from the population over-represented or under-represented due only to chance factors.

  is always a good and useful sample.

Sampling in R

- To generate random samples of a given population,
  - Sample(1:10)
- To specify the number of items returned
  - Sample(1:10, size =5)

Quiz

- How can we compare a sample in a population to other samples in that population ?
  - By finding the mean of this sample
  - By finding the means of other samples in this population
  - By comparing the mean of this sample to the others

Sampling Distribution

Central Limit Theorem

What is Sampling Distribution ?

Suppose that we draw all possible samples of size *n* from a given population and then we compute a statistic (e.g., a mean, proportion, standard deviation) for each sample. The probability distribution of this statistic is called a **sampling distribution**. And the standard deviation of this statistic is called the **standard error**.

Central Limit Theorem

Sampling distribution of the sample mean will be approximately normal

$$\text{As } n \text{ increases, } \bar{x} \rightarrow \mu$$

Confidence Interval

A confidence interval is always centered around the mean of your sample. In order to construct this, add a margin of error. It can be found by multiplying the standard error of the mean by the z score of the percent confidence interval

$$\text{confidence interval} = \bar{x} \pm \text{margin of error}$$

$$\text{margin of error} = Z \times \frac{\sigma}{\sqrt{n}}$$

Note

Common choices for C.I are 90%, 95% , and 99% . The interval constructed with 99% confidence will have higher chance of containing the true mean than an interval constructed with 95% confidence

C.I - Exercise

- Suppose the average height of a sample of 100 women is 5’5’’, in other words, X bar = 5’5’’. Within what range of heights can we expect the population mean to be, with 95% confidence? Assume a standard deviation for the population of 1.5’’

Our sample mean is 5’5”

The standard deviation of the population is 1’5“

Our sample size is 100.

We are asked for 95% confidence, so we want to use $z = \pm 1.96$.

$$5.5 \pm 1.96 * 1.5/ \sqrt{100} = 5.5 \pm .294 = (5.21, 5.79)$$

We can conclude, with 95% confidence, that the true average height for women is between 5’2" and 5’8".

Probability Distribution

- What is inferential statistics ?
- Random variables and probability
- Joint and conditional probability

Probability distributions

- Discrete Distribution
- Continuous Distribution
- Binomial Distribution
- Negative Binomial Distribution
- Poisson Distribution
- Normal Distribution
- Standard Normal Distribution

Random Variables and Probability

Random Variables and Probability

- We introduce the concept of **random variables and how they apply to statistical theory**

  A **random variable is one that takes a numerical value whose outcome is determined by an experiment.**
  - Number of products a customer orders
  - Time spent by a visitor on your website
  - Purchase by a visitor via your website

Types of Variables

Continuous Data	Discrete Data
Data that can be expressed in either fractions or whole numbers	Data that can take a limited number of values
Often obtained by use of a physical measurement	Discrete process outputs typically describe an event
Examples: – Temperature of the room – Exchange rate of a currency – Lateness of a delivery – Height of a person – Elapsed time to complete transaction – average length of phone calls	Examples: – Number of bugs/ defects in the code – Proportion of late applications, incorrect invoices – CMMI levels – ACSAT ratings (1,2,3,4,5)

Probability

What is Probability ?

It is a quantitative measure of probability , in other words it's a measure of uncertainty

Probability = Chance = Likelihood

$$P(E) = \frac{\text{\# of favorable outcomes}}{\text{Total \# of possible outcomes}}$$

Rule 1 : For any event A, the probability P(A) satisfies

$0 <= P(A) <= 1$

Rule 2: The sum of the probabilities of all basic outcomes in the sample space must equal 1

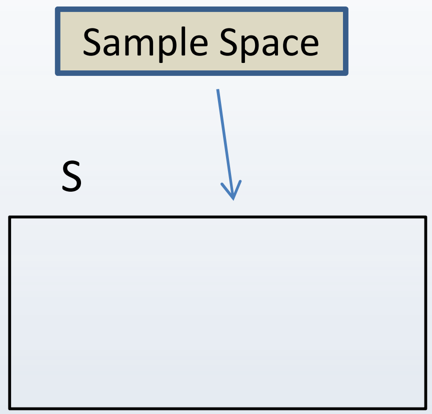
$P(S)=P(e_1)+P(e_2)+P(e_3)+....+P(e_n)=1$

Components of Probability

- Experiment - is a process that leads to one of several possible outcomes
  - ✓ Produces outcomes
  - ✓ Another example,
    - ✓ tossing a coin
    - ✓ Rolling a die
    - ✓ Tossing a coin 2 times
- Event
  - ✓ Its an outcome of an experiment (for ex, head or tail)
  - ✓ For any event A, the probability of A is between 0 and 1,

    $0 <= P(A) <= 1$

    Experiment : Toss 2 coins , note down the faces
    Event -> Getting 1 Head & 1 Tail (outcomes in event)



- Sample Space
  - ✓ The probability of sample space is **1**
  - ✓ A listing of all possible elementary events for an experiment

    S = { Head, Tail } -- Sample space for flipping a coin once

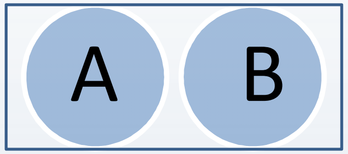
    S = { (H,H),(H,T),(T,H),(T,T) } - Sample space for flipping a coin twice

Probability – Rule

Addition Rule 1: (Independent case/disjoint/mutually exclusive events)

When two events, A and B, are mutually exclusive, the probability that A or B will occur is the sum of the probability of each event.

$$P(A \text{ or } B) = P(A) + P(B) \quad \text{(or)} \quad P(A \cup B) = P(A) + P(B)$$



If events A and B are mutually exclusive , then $p(A \cap B) = 0$

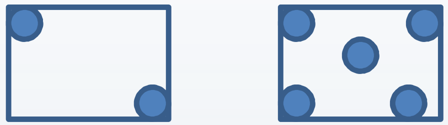
Multiplication Rule 1: (Independent case/disjoint events)

$$P(A \text{ and } B) = P(A) * P(B)$$

$P(A \cap B)$ **=** $P(A) * P(B)$, ,if A & B are independent

Experiment 1: A single 6-sided die is rolled. What is the probability of rolling a 2 or a 5?

- Probabilities:
- $P(2) = \frac{1}{6}$
- $P(2 \text{ or } 5) = P(2) + P(5)$



$$= \frac{1}{6} + \frac{1}{6}$$

$$= \frac{2}{6}$$

$$= \frac{1}{3}$$

A glass jar contains 1 red, 3 green, 2 blue, and 4 yellow marbles. If a single marble is chosen at random from the jar, what is the probability that it is yellow or green?

- P(yellow) = $\frac{4}{10}$
- P(green) = $\frac{3}{10}$
- P(yellow or green) = P(yellow) + P(green)

$$= \frac{4}{10} + \frac{3}{10}$$

$$= \frac{7}{10}$$

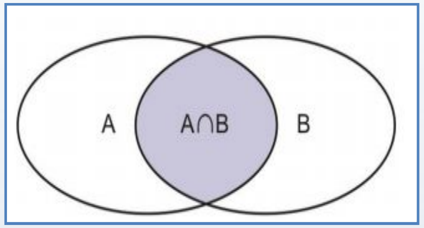
Addition Rule 2: (Dependent case/non-disjoint events)

- When two events, A and B, are non-mutually exclusive, there is some overlap between these events. The probability that A or B will occur is the sum of the probability of each event, minus the probability of the overlap

P(A or B) = P(A) + P(B) – P(A and B)

Or

$P(A \cup B) = P(A) + P(B) - P(A \cap B))$



To calculate the probability of the union of two events A and B, we add their marginal probabilities and subtract their joint probability from this sum. We must subtract the joint probability of A and B from the sum of their marginal **probabilities to avoid double counting** because of common outcomes in A and B.

Multiplication Rule 2 – (Dependent case)

P(A and B) = P(A) * P(B|A)

$P(A \cap B) = P(A) * P(B|A)$

Assignment

On a festival eve, the probability of a person meeting with a car accident is 0.07. The probability of driving while drunk is 0.42 and probability of a person having a car accident while drunk is 0.25. What is the probability of a person driving while drunk or having a car accident?

P(drunk or accident) = P(drunk) + P(accident) - P(drunk and accident)

= 0.42 + 0.07 - 0.25

= 0.24

Assignment

What is the probability of rolling a “5” and then a “3” with a normal 6 sided die ?

Ans : P(3 and 5) = ?

What is the probability of drawing a “king” and then drawing a “queen” from a deck of cards, without putting the king back ?

Ans : P(king) =

p(Queen) =

P(King and Queen) =

Without replacement

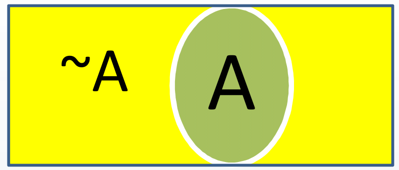
A box contains 4 apples and 3 oranges. Two fruits are drawn. What is the probability that both fruits are apples if the fruits are drawn ?

(a) With replacement

(b) Without replacement

Ans:

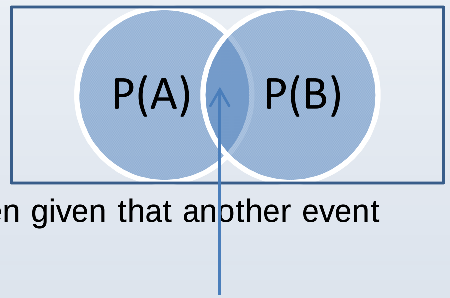
- **Rule of subtraction/Complementation Law**
  - P(A') = 1 - P(A)
  - **The yellow colored one is Complement of A**



The probability that Mr. Lal will graduate from college is 0.75. What is the probability that Lal will not graduate from college?

Based on the rule of subtraction, the probability that Lal will not graduate is 1.00 - 0.75 or 0.25.

- **Conditional Probability**



  - A conditional probability is the likelihood that an event will happen given that another event has already happened

P(A and B)

$$\mathbf{P(A|B) = \frac{P(A\ AND\ B)}{P(B)} = \frac{P(A \cap B)}{P(B)}}$$

Conditional Probability – Assignment

At ST Johns School, the probability that a student takes Technology and Statistics is 0.087. The probability that a student takes Technology is 0.68. What is the probability that a student takes Statistics given that the student is taking Technology?

$$P(\text{Statistics} \mid \text{Technology}) = \frac{P(\text{Technology and Statisitcs})}{P(\text{Technology})} = \frac{0.087}{0.68} = 0.13 = 13\%$$

Contingency Table

- Provides insight into the relationship between two categorical variables
- Contingency can represent more than 2 values
- Find the expected frequencies using probability

	Male	Female	
Project manager	8	3	11
Project Leader	31	13	44
Senior developers	52	17	69
Administrators	9	22	31
	100	55	155

Joint and Marginal Probability

- **Joint Probability (when 2 events both occur)** ∩ = intersection
  - Is the likelihood that two or more events will happen **at the same time**
  - Its like a contingency table but here we don't assign frequencies instead probabilities will be assigned

Contingency table

	Male	Female	Total
Promoted	288 (.24)	36 36/1200 = .03	324 .24+.03 = .27
Not Promoted	672 672/1200 = .56	204 204/1200 = .17	876 .56+.17 = .73
	960 (.24+.56) .80	240 (.03+.17) = .20	1200 1

Divide all areas by 1,200 to find the probability associated with each area.

	Male	Female	Total
Promoted	.24	.03	.27
Not Promoted	.56	.17	.73
	.80	.20	1

Marginal Probabilities appears in the margin of the table

Joint and Marginal Probability

Lets say

M = Being a male } F = Being a female

P = Being Promoted P' = Not being promoted

	Male	Female	Total
Promoted	.24	.03	.27
Not Promoted	.56	.17	.73
	.80	.20	1

- What is the probability of being male and being promoted?
- What is the probability of being female and is promoted ?
- What is the probability of being male and not promoted ?
  What is the probability being a female and is not promoted ?
  What is the probability that employees being promoted ?

- Pr(M n P) = .24
- Pr(F n P) = .03
- Pr(M n P') = .56
- Pr(F n P') = . 17
- Pr(E n P) = .27

The comparison we want to make is

Pr(P|M) vs. Pr(P|F)

Pr(P|M) = Pr(P and M) / Pr(M)

= 0.24 / (0.56 + 0.24) = 0.3

Pr(P|F) = Pr(P and F) / Pr(F)

= 0.03 / (0.03 + 0.17) = 0.15

What do we conclude

→ Males are promoted at <u>2 times</u> the frequency of females.

Conditional Probability - Assignment

- In Europe, 78% of all households have a television. 61% of all households have a television and a Blue ray. What is the probability that a household has a Blue ray given that it has a television?

- At a middle school, 18% of all students play football and basketball and 32% of all students play football. What is the probability that a student plays basketball given that the student plays football?

- A probability that a data engineer will remain with a company is .7. The probability that an employee earns more than Rs.30,000 per month is .6. The probability that an employee who is a data engg remained with the company or who earns more than Rs.30,000 per month is 0.7. What is the probability that an employee earns more than Rs.30,000 per month given that he is a data engineer who stayed with the company ?

Joint Probability – Employee Turnover Case Study

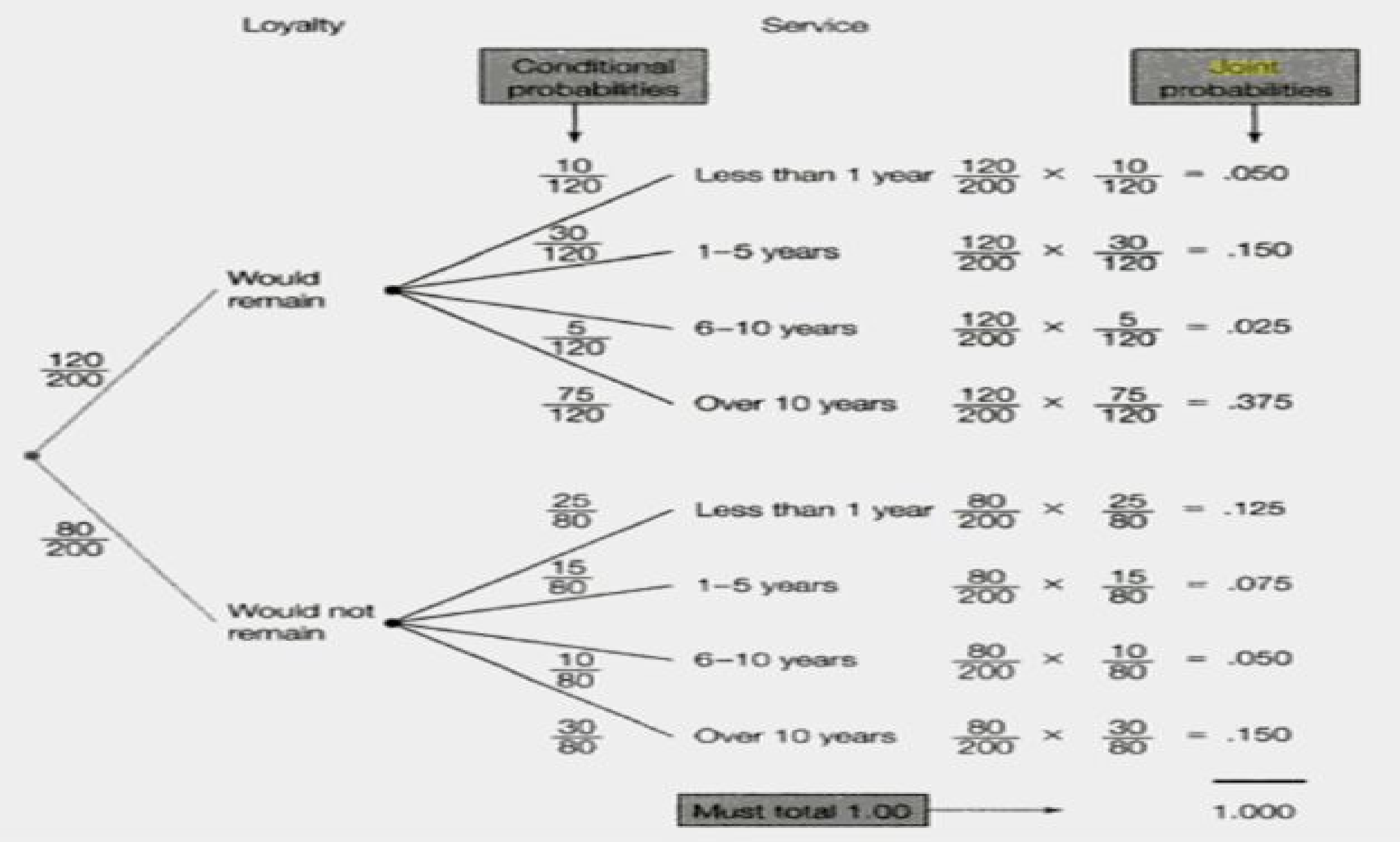
Problem Scenario
When a MNC company ran survey with their employees to find out whether or not they would remain with the company if they received a better offer from other employer

A CONTINGENCY TABLE is a table used to classify sample observations according to two or more identifiable characteristic

Loyalty	Length of Service				
	Less than 1 year	1-5 years	6-10 years	More than 10 years	Total
Would remain	10	30	5	75	120
Would not remain	25	15	10	30	80
Total	35	45	15	105	200

1. What is the probability of employees randomly selected who would remain **and** who has more than 10 years experience ? Ans: .375
2. What is the probability of employees who would not remain **and** who has more than 10 years experience ?
3. What is the probability of randomly selected employees with 1-5 years ?

Conditional Probabilities with Decision Tree



Marginal	Union	Joint	Conditional
P(A)	P(A U B)	P(A ∩ B)	P(A | B)
The probability of A occurring	The probability of A or B occurring	The probability of A and B occurring	The probability of A occurring given that B has occurred

What is a probability distribution ?

A probability distribution is a table or an equation that links each possible value that a random variable can assume with its probability of occurrence

Steps for building a probability distribution

a. Define random variable
b. Build frequency distribution
c. Calculate relative frequency
d. Check requirements : P(x) >= 0 and P(x) = 1
e. Create a chart to visualize the distribution
f. Make predictions

Supposing a coin is tossed for two times, what is the probability of

(i) **obtaining 2 heads**

(ii) **obtaining 1 , or 0 heads**

(iii) **2 or 1 heads**

(iv) **More than 1 head**

Probability Distribution (probability mass function)

- The probability of any specific outcome for a discrete random variable must be between 0 and 1
- The sum of the probabilities over all possible values of a discrete random variable must equal 1.

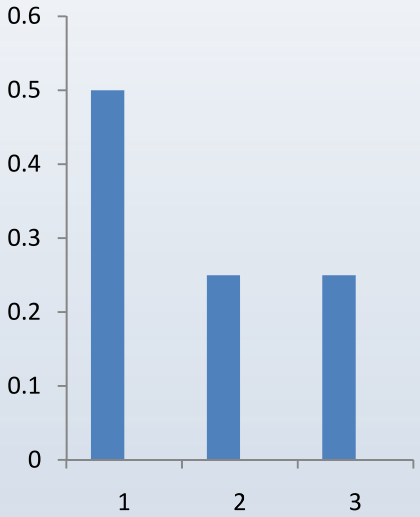
Probability Distribution - Exercise

Random Experiment : tossing a coin 2 times

Event : an event is a subset of sample space E = {H,T}

random discrete variable : # of Heads **Sample space:** { (H,H),(H,T),(T,H),(T,T)}

Outcomes : 2 =➔ 2 * 2



1st toss	2nd toss	# Heads	Probability
H	H	2	2/4 .5
H	T	1	.1/4 25
T	H	1	¼ .25
T	T	0	0
			1

- Hilton hotel has 4 ball rooms. Based on the week end historical room usage data
  - On two days 0 rooms were used,
  - On 21 days 1 room was used,
  - 42 days 2 rooms were used
  - 27 days 3 rooms were used
  - On 4 days 8 rooms were used

  Let x = # of ball rooms used during day

# of rooms used	# of days Frequency	Relative frequency P(x)
0	2	
1	21	
2		
3		
4		

Obtain the probability of

(i) $p(x <= 1)$

(ii) $p(x = 2 \text{ or } x = 3)$

(iii) $p(x > 0)$

Expected Value

- Expected value is an anticipated value for a given risk or investment or given event.
- The expected value is calculated by multiplying each of the possible outcomes by the likelihood each outcomes will occur, and summing all of those values

Expected value = ∑ Probability * value

Example

It costs $3 to play a game of dice. You can expect to win $7 if you roll a 5 and $20 if you roll a 6. What is the expected value of this game?

Dice roll	Earnings	Probability	Expected value
1			
2			
3			
4			
5	$7	1/6	.33 * $7 = 1.19
6	$20	1/6	.33 * $20 = 3.4

Expected value – Portfolio Analysis

For the given below Portfolio, compute the expected value

Prob of State of economy	TCS Stock Return	Infosys Stock Return	CTS Stock Return
0.15	0.17	0.32	0.1
0.30	-0.03	-0.12	0.04
0.55	-0.02	0.43	0.11

Types of Probability Distributions

Types of Distributions

Discrete Distributions	**Continuous Distributions**
Binomial Distribution	Uniform Distribution
Negative Binomial Distribution	Normal Distribution
Poisson Distribution	Exponential Distribution
Hyper geometric Distribution	

Discrete Probability Distributions

Discrete Distribution - Binomial

Binomial Probability Distribution

- It's a discrete probability distribution
- A single trial of a binomial experiment contains only 2 possible outcomes

$P(x) = n\,C_x\,P^x\,q^{n-k}$

n = number of trials, k = number of successes desired,
p = probability of getting success in one trial.
q=1-p = probability of getting failure in one trial

$$nC_x = \frac{n\,!}{k!\,(n-k)\,!}$$

$$P(x) = \frac{n!}{x!(n-x)!}\,p^x(1-p)^{n-x}$$

for *x = 0, 1, 2, …… n.*

BINOMIAL DISTRIBUTION

Binomial Distribution is an example of a probability distribution of a discrete random variable

The coin toss example is most widely used, but there are many other examples:

- Gender of babies delivered in a hospital
- Whether a car is Defective
- Whether a Customer is default

The common feature across all these examples are:

1. There are only two possible outcomes: Win or Lose, 1 or 0, Male or Female
2. There are no external factors influencing the probability of each outcome over time
3. The chances of each outcome are independent of previous results

A salesperson has appointments with 4 prospective clients Tomorrow. From the past he knows that the probability of making a sale on any one appointment is 1 in 5 (1/5=.2). What is the likelihood that she will sell 3 policies in 4 tries ?

what is the probability of making exactly 3 sales in 4 attempts , p = .2 ?

$$f(3) = P(3) = \frac{4!}{3!\,(4-3)!} * .2^{3} * .8^{(4-3)} = 4 * .008 * .8 = .0256$$

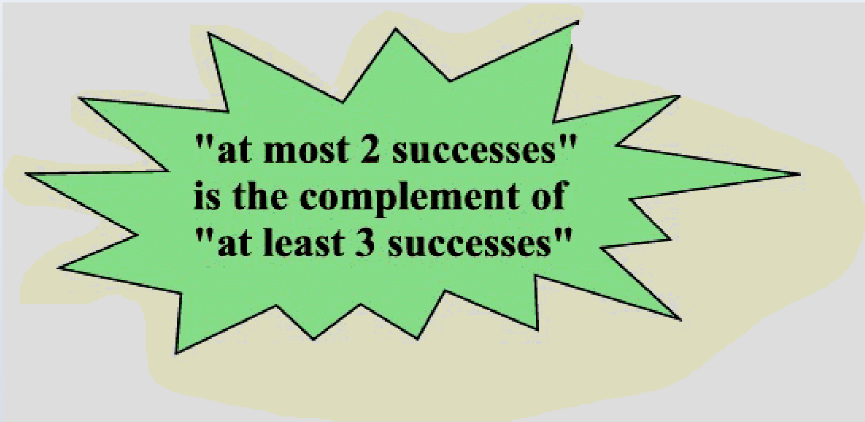
= BINOMDIST(3,4,.02,0)

= BINOMDIST(x=3,size=4,prob=.2)

dbinom(x=3,size=4,p=.2)

When computing **"at least"** and **"at most"** probabilities, it is necessary To consider, in addition to the given probability,

- All probabilities larger than the given probability ("at least")
- All probabilities smaller than the given probability ("at most")
- At least 2 is also 2 or more , the minimum no. of events are 2 but not less than 2
- At most 2 is also two or less



Discrete Distribution Binomial – Case Study

You are the Finance Manager for Company X, and you are looking at your AR balances. Based on past data, you know that on average 40% of your customers are more than 60 days late with payments

Let's say you have a total of 5 customers, and you want to create a contingency for situations where > 50% of your customers are late

Probability of 0 customer being late ?

Overall probability of being late (p): 0.4

Probability of not being late (1-p) : (1-0.4) 0.6

Total number of trials (n) : 5

Total successes (x) : **0**

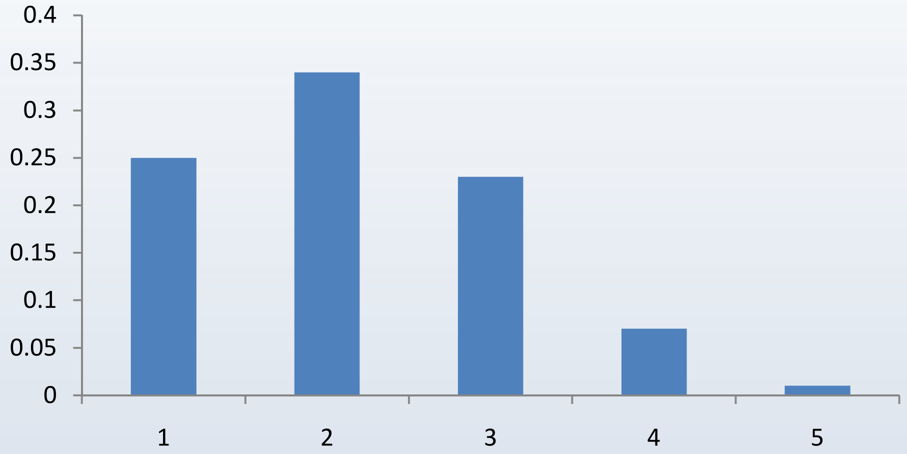
```
dbinom(x=0,size=5,p=.4)
```

If you calculate this you will get : **0.0776**

Binomial Distribution

- If you recalculate for 1 late customer out of 5:
  p = 0.2592
- 2 Late customers:
  P = 0.3456
- 3 late customers:
  P = 0.2304
- 4 late customers:
  P = 0.0768
- 5 late customers:
  P = 0.0120

P of > 50% customers being late = 0.2304+0.0768+0.0120 = 0.3192



Binomial Distribution

1. Irrespective of the actual experiment, as long as number of trials are 5 and Probability of success is 0.4 , p=0 is always 0.0776

2. If probability of success is 0.4, it makes intuitive sense that probability of success = 0 is a low number relative to probability of success, as is probability of success =5

3. We can calculate the probabilities for a family of binomials, where n=5 but p changes from 0.1 to 0.9

4. We can also calculate the probabilities for any n and success combination , and these will be fixed values

Binomial Distribution – Assignment

A quality control staff is in charge of testing whether or not 90% of the DVD players produced by his company conform to specifications. To do this, the engineer randomly selects a batch of 12 DVD players from each day's production. The day's production is acceptable provided no more than 1 DVD player fails to meet specifications. Otherwise, the entire day's production has to be tested.

(i) What is the probability that the engineer incorrectly passes a day's production as acceptable if only 80% of the day's DVD players actually conform to specification?

(ii) What is the probability that the engineer unnecessarily requires the entire day's production to be tested if in fact 90% of the DVD players conform to specifications?

Answer

Let X denote the number of DVD players in the sample that fail to meet specifications. In part (i) we want $P(X \le 1)$ with binomial parameters
$n = 12$, $p = 0.2$.

$$P(X \le 1) = P(X = 0) + P(X = 1)$$

$$= \binom{12}{0} (0.2)^0 (0.8)^{12} + \binom{12}{1} (.2)^1 (0.8)^{11} = 0.069 + 0.206 = 0.275$$

n = number of trials, k = number of successes desired,

p = probability of getting success in one trial.

q=1-p = probability of getting failure in one trial

Answer 2

(ii) We now want $P(X > 1)$ with parameters $n = 12$, $p = 0.1$.

$$P(X \le 1) = P(X = 0) + P(X = 1)$$

$$= \binom{12}{0} (0.1)^0 (0.9)^{12} + \binom{12}{1} (0.1)^1 (0.9)^{11} = 0.659$$

So $P(X > 1) = 0.341$.

n = number of trials, k = number of successes desired,
p = probability of getting success in one trial.
q=1-p = probability of getting failure in one trial

A manufacturer produces PenDrives in batches of 10,000. On average, 2% of the PenDrives are defective.

A retailer purchases PenDrives in batches of 1,000. The retailer will return any shipment if 3 or more PenDrives are found to be defective out of 30 batches. For each batch received, the retailer inspects thirty PenDrives. What is the probability that the retailer will return the batch?

N = 30 trials

k = 3 successes

p = 0.02

$$\Pr(\text{3 successes out of 30 trials}) = \binom{30}{3} 0.02^3 (1-0.02)^{30-3} = 0.019 = 1.9\%$$

<u>Error</u>

The formula gives us the probability of <u>exactly</u> 3 successes out of 30 trials. But, the retailer will return the shipment if it finds <u>at least</u> 3 defective PenDrives. What we want is

Pr(3 out of 30) + Pr(4 out of 30) + … + Pr(30 out of 30)

Binomial Distribution

N = 30 trials
x = 3 successes
p = 0.02

$$\Pr(3 \text{ successes out of 30 trials}) = \binom{30}{3} 0.02^3 (1-0.02)^{30-3} = 0.019 = 1.9\%$$

N = 30 trials
x = 4 successes
p = 0.02

$$\Pr(4 \text{ successes out of 30 trials}) = \binom{30}{4} 0.02^4 (1-0.02)^{30-4} = 0.003 = 0.3\%$$

N = 30 trials
x = 5 successes
p = 0.02

$$\Pr(5 \text{ successes out of 30 trials}) = \binom{30}{5} 0.02^5 (1-0.02)^{30-5} = 0.0003 = 0.03\%$$

Etc. out to x = 30 successes.

Alternatively

Because Pr(0 or more successes) = 1, we have an easier path to the answer:

Pr(3 or more successes) = 1 – Pr(2 or fewer successes)

Binomial Distribution

N = 30 trials
x = 0 successes
p = 0.02

N = 30 trials
x = 1 successes
p = 0.02

$$\Pr(1 \text{ successes out of 30 trials}) = \binom{30}{1} 0.02^{1} (1-0.02)^{30-1} = 0.334$$

N = 30 trials
x = 2 successes
p = 0.02

$$\Pr(2 \text{ successes out of 30 trials}) = \binom{30}{2} 0.02^{2} (1-0.02)^{30-2} = 0.099$$

→ Pr(2 or fewer successes) = 0.545 + 0.334 + 0.099 = 0.978

→ Pr(3 or more successes) = 1 – 0.978 = 0.022 = 2.2%

Negative Binomial (also known as Pascal) Distribution

- A **negative binomial random variable** is the number *X* of repeated trials to produce *r* successes in a negative binomial experiment. The probability distribution of a negative binomial random variable is called a **negative binomial distribution**.
  - Tossing a coin till you get to 3 heads
  - Number of customers that visit a website till they order 5 items

$$P(X=r) = {n-1}\,C_{r-1}\,P^{r}\,(1-p)^{n-r}$$

n = number of events , r = number of successful events ,

p = probability of success on a single trial. 1-p = probability of failure

$$nCr = \frac{n!}{(n-r)!\,/\,r!}$$

times on heads.
  - Each trial can result in just two possible outcomes - heads or tails.
  - The probability of success is constant - 0.5 on every trial.
  - The trials are independent; that is, getting heads on one trial does not affect whether we get heads on other trials.
  - The experiment continues until a fixed number of successes have occurred; in this case, 5 heads.

Negative binomial distribution - Example

- A company runs a survey over phone has found based on the past data that there is a 0.40 probability that a call made between 2:30 pm and 5.30 pm will be answered.

Obtain the probability that an interviewers 10th answer comes on his 20th call

n = number of events , r = number of successful events ,
p = probability of success on a single trial. 1-p = probability of failure

$P(X=r) = {}^{n-1}C_{r-1}\, P^{r}\, (1-p)^{n-r}$

$= {}^{19}C_9\, (.4)^{10}\, (.6)^{10}$ **$n = 20, r=10$**

$= .058$

Binomial Table

$$B(x;n,p) = \sum_{0 \le y \le x} b(y;n,p)$$

The values of $B(x;n,p)$ for $0.5 < p < 1.0$ are obtained by using the formula

$$B(x;n,1-p) = 1 - B(n-1-x;n,p)$$

		p									
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
5	0	0.774	0.590	0.444	0.328	0.237	0.168	0.116	0.078	0.050	0.031
	1	0.977	0.919	0.835	0.737	0.633	0.528	0.428	0.337	0.256	0.188
	2	0.999	0.991	0.973	0.942	0.896	0.837	0.765	0.683	0.593	0.500
	3	1.000	1.000	0.998	0.993	0.984	0.969	0.946	0.913	0.869	0.813
	4	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.990	0.982	0.969
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0	0.599	0.349	0.197	0.107	0.056	0.028	0.013	0.006	0.003	0.001
	1	0.914	0.736	0.544	0.376	0.244	0.149	0.086	0.046	0.023	0.011
	2	0.988	0.930	0.820	0.678	0.526	0.383	0.262	0.167	0.100	0.055
	3	0.999	0.987	0.950	0.879	0.776	0.650	0.514	0.382	0.266	0.172
	4	1.000	0.998	0.990	0.967	0.922	0.850	0.751	0.633	0.504	0.377
	5	1.000	1.000	0.999	0.994	0.980	0.953	0.905	0.834	0.738	0.623
	6	1.000	1.000	1.000	0.999	0.996	0.989	0.974	0.945	0.898	0.828
	7	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.988	0.973	0.945
	8	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.989
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
15	0	0.463	0.206	0.087	0.035	0.013	0.005	0.002	0.000	0.000	0.000
	1	0.829	0.549	0.319	0.167	0.080	0.035	0.014	0.005	0.002	0.000
	2	0.964	0.816	0.604	0.398	0.236	0.127	0.062	0.027	0.011	0.004
	3	0.995	0.944	0.823	0.648	0.461	0.297	0.173	0.091	0.042	0.018
	4	0.999	0.987	0.938	0.836	0.686	0.515	0.352	0.217	0.120	0.059
	5	1.000	0.998	0.983	0.939	0.852	0.722	0.564	0.403	0.261	0.151
	6	1.000	1.000	0.996	0.982	0.943	0.869	0.755	0.610	0.452	0.304
	7	1.000	1.000	0.999	0.996	0.983	0.950	0.887	0.787	0.654	0.500
	8	1.000	1.000	1.000	0.999	0.996	0.985	0.958	0.905	0.818	0.696
	9	1.000	1.000	1.000	1.000	0.999	0.996	0.988	0.966	0.923	0.849
	10	1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.991	0.975	0.941
	11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.994	0.982
	12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996
	13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Discrete Distribution – Poisson

Poisson Distribution

- Another discrete probability distribution, that is used to model number of events occurring in a fixed time frame

- Examples include
  - Number of warranty claims in a month
  - Swine flu spread in a day
  - Number of call in an hour made by customer
  - Number of patients needing emergency services in a day

- The following conditions apply to appropriately use Poisson Distribution
  - Events have to be counted as whole numbers
  - Events are independent : so if one event occurs, it does not impact the chances of the second event occurring
  - Average frequency of occurrence for the given time period is known
  - Number of events that have already occurred can be counted

Poisson Distribution

Poisson probabilities are calculated as

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

Where, lambda is the mean number of occurrences in a given interval of time.

Notice that there is no n (sample size) impact.

We can calculate the Poisson probabilities

e is a constant value

e = 2.718281828.

The average number of homes sold by a realtor is 2 homes per day. What is the probability that exactly 3 homes will be sold tomorrow?

λ = 2; since 2 homes are sold per day, on average.

x = 3; since we want to find the likelihood that 3 homes will be sold tomorrow.

We plug these values into the Poisson formula as follows:

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

$P(3; 2) = (2.71828^{-2}) (2^{3})!$

$P(3; 2) = (0.13534) (8) / 6$

$P(3; 2) = 0.180$

Thus, the probability of selling 3 homes tomorrow is 0.180 .

Assignment 2

You work as a manager in a call center. You have a staff of 55 people, who on average handle 330 calls in an hour. A major holiday is coming up and 5 resources want leave. You estimate that the 50 remaining resources can manage 20% greater calls, but want to plan for the chances if greater than 20% increased call volume.

What are the chances that the number of calls on that day will go up by more than 20% ?

r = (330)/55 = 6 calls an hour
20% greater calls with 5 less resources = (330*1.2)/50 = 8 calls an hour

So we need probability of seeing 8 or more calls an hour when average is 6.
You can either calculate it manually or look it up in the Poisson Distribution

```
1- sum(dpois(0:8,6))
```

Suppose a life insurance company insures the lives of 5000 persons aged 52. As per the studies show the probability that any 52 year old person will die in a year to be 0.001, find the probability that the company will have to pay **at least** two claims during a given year.

Solution

n = 5000, p=0.001 and lambda = np = n * p

P(x >= 2) = 1 – P(x < 2) = 1 – [P(x=0) + P(x=1)]
= 1 – [e to the power lambda + lambda e to the power minus lamba]
= 1- [e to the power -5 + 5 e to the power -5]
= 1- 6 e to the power -5
= 1 – 6 * .0067

Poisson Distribution – Assignment 4

A website can expect 2 customers every 3 minutes, on average. What is the prob that 4 or fewer customers will enter the website in a 9 minute period.

- Tip : lambda = 6

Poisson Distribution – Assignment 5

- Suppose a life insurance company insures the lives of 5000 persons aged 52. If studies show the probability that any 52 year old person will die in give year to be 0.001, find the probability that the company will have to pay **at least** two claims during a given year.

Poisson distribution – Assignment 6

- A leading reality builder Olympia opaline sells homes on an average of 2 per day. What is the probability that exactly 3 homes will be sold tomorrow?

Poisson Distribution – Assignment 7

- A company produces Blue ray player that 0.1 percent of the blue rays are defective. The Blue rays are packed in boxes containing 500 blue rays. A stockiest buys 100 boxes from the producer of blue rays. Using Pois distribution, find how many boxes will contain (a) no defectives (b) atleast 2 defectives

Geometric Distribution

It's a special case of binomial distribution, and computes the probability of seeing the first success in r trials

Example

1. Toss a coin till it lands on heads
2. Customers walk-in till first purchase

$$g(x ;P) = P * Q^{x-1} \text{, where } Q = (1-P)$$

- **Characteristics of a Geometric Setting:**
  - Each observation (or trial) has two categories: success or failure.
  - The observations are all independent.
  - The probability of success (p) is the same for each trial.
  - We wish to find the number of trials needed to obtain the <u>first</u> success.

- **What are the Differences between the Geometric and the Binomial Distributions?**
  - The most obvious difference is that the Geometric Distribution <u>does</u> <u>not</u> have a set number of observations, n.
  - The second most obvious difference is the question being asked:
    - ➢ Binomial: Asks for the probability of a certain number of successes.
    - ➢ Geometric: Asks for the probability of the <u>first</u> success.

Geometric Distribution

- It's a distribution of number of trials needed to get the **first** success in repeated independent trials

For example, if a coin is repeatedly tossed what is the probability the first time head appears occurs on the 8^{th} toss

Geometric Distribution – Example

- There are 20 red marbles, 10 blue marbles, and 5 white marbles in a jar. Select a marble without looking, note the color, and then replace the marble in the jar. We're interested in the number of marbles you would have to draw in order to be sure you have a red marble.

  - Two Categories?
    - Yes, in this case type red marble = success and any other color = failure.
  - Independent Observations?
    - Yes, we are drawing blindly and replacing the marbles – so results are independent.
  - p is the same?
    - Yes, for each trial p = 20/35.
  - Are we looking for the first success?
    - Yes, we are looking for the number of trials before we obtain the first red marble.

Since all 4 conditions are met, this distribution is GEOMETRIC

Geometric distribution – Assignment 1

Among employed women, 25% have never been married. Suppose we randomly Sample women in a particular business office.

What is the probability that the first woman who says she has never been married is the fourth woman in a sample?

dgeom(1,.25)
Ans: .1875

Hyper geometric Distribution

It describes the number of successes in a sequence of n draws from a finite population **without replacement**

$$P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$$

Hypergeometric in R

Usage

dhyper(x, m, n, k, log = FALSE) phyper(q, m, n, k, lower.tail = TRUE, log.p = FALSE) qhyper(p, m, n, k, lower.tail = TRUE, log.p = FALSE) rhyper(nn, m, n, k)

Arguments

x, qvector of quantiles representing the number of white balls drawn without replacement from an urn which contains both black and white balls.

m the number of white balls in the urn.

n the number of black balls in the urn.

k the number of balls drawn from the urn.

p probability, it must be between 0 and 1.

n number of observations. If length(nn) > 1, the length is taken to be the number required.

log, log.plogical; if TRUE, probabilities p are given as log(p).

lower.taillogical; if TRUE (default), probabilities are *P[X ≤ x]*, otherwise, *P[X > x]*

Hypergeometric Distribution – Example

A Jar contains black and white balls. The variable *N* describes the number of **all balls in the basket** and ***m*** describes the number of **white marbles**, then ***N – m*** corresponds to the number of **black balls**.

Now, assume that there are 5 white and 45 black balls in the basket. Close your eyes and draw **10 balls** without replacement.
What is the probability that exactly 4 of the 10 are white?

Solution :

N be the size of the population ,

m be the total number of elements having a certain characteristic (called success) and the remaining N – m do not have it

N= ? m=? m=? k=?

N=50, m=5, n=N-m , k=10

dhyper(x=4,m=5,n=45,k=10)

Cipla is a pharma company it produces neurozan, FDA arrives and wants to conduct a test on a week of production (20 batches) by taking 5 batches. Your internal testing has indicated that 17 of the 20 batches are good .

What is the probability that at least 4 of the 5 batches tested as good ?

Solution

N=20, m = K=17 no. of good (success), n = N-m = 3 r= 4 , k = 5

Linkedin resource manager randomly selects 4 individuals from a group of 10 employees for a special assignment; Assuming that 4 of the employees were assigned to a similar assignment previously, determine the prob. that exactly 2 of the 3 employees have had previous experience

Answer

N = 10 , m = 4, r = 2, N-m = 6

```
dhyper(x=2,m=4,n=6,k=3) # correct
```

Hypergeometric Distribution – Assignment 3

Suppose a particular industrial product is shipped in lots of 20. to determine whether an item is defective a sample of 5 items from each lot is drawn. A lot is rejected if more than one defective item is observed.

If the lot is rejected, each item in the lot is then tested).
If a lot contains 4 defectives, what is the probability that it will be accepted ?

N= 20 , m = 4, n= N-m=16

P(accept the lot) = p(x <=1) = p(0) + P(1)

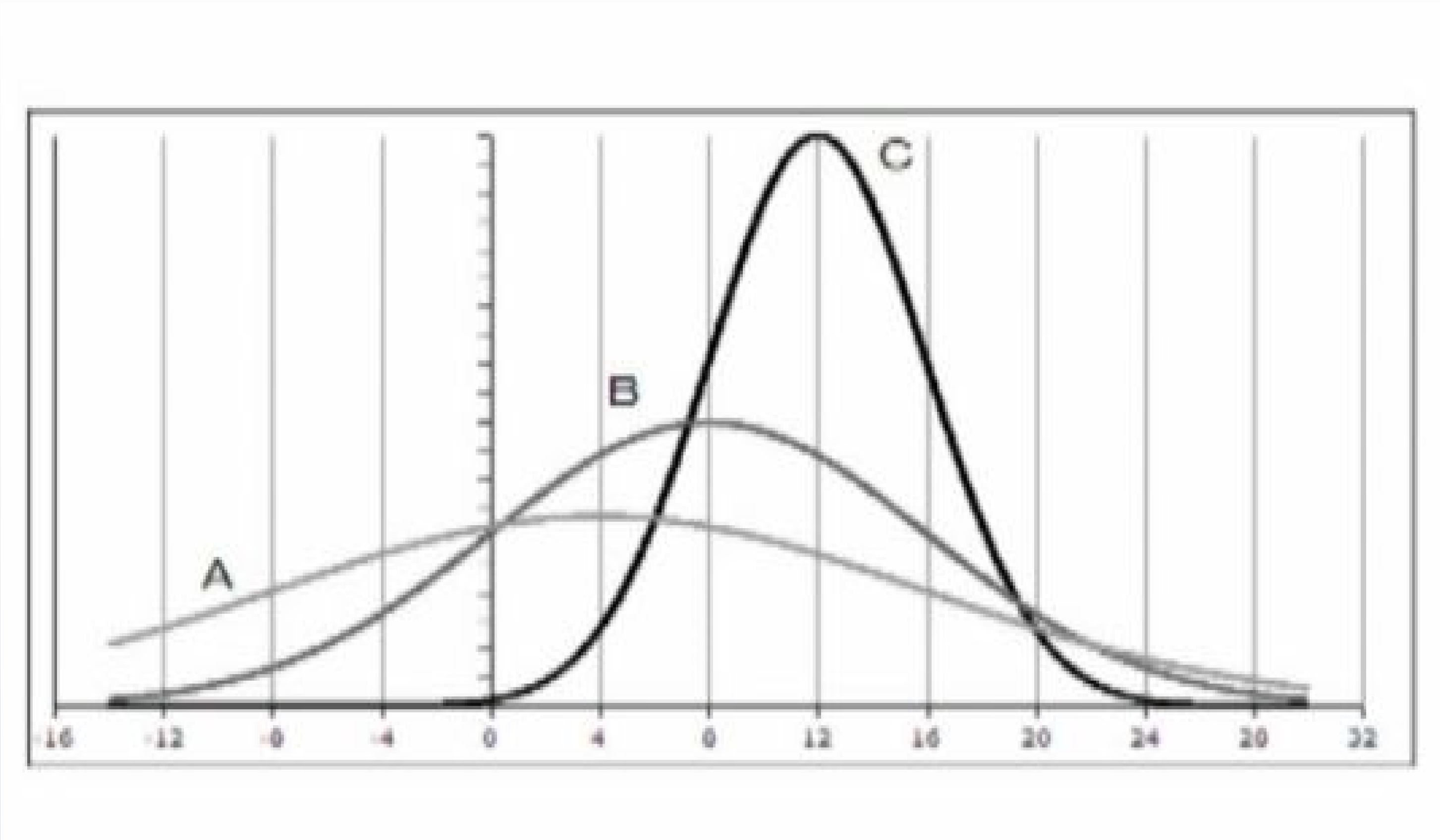
```
sum(dhyper(x=0:1,m=4,n=16,k=5)) # correct
```

Discrete Distributions - Summary

Distribution	Application	Examples
Binomial	X successes in n trials	5 faulty parts in 100, p(faulty part) = 0.2 15 customers purchase product out of 100 that visit the online
Negative Binomial	Xth success in Nth trial	Second faulty part will be in the 5th box
Geometric	1st success in Nth trial	First purchase will not happen till 50th customer
Poisson	X counts in a fixed time period when average for time period is N	50 patients in an hour when average is 35
Hypergeometric	X successes in n sample out of a population of N	If 5 defects in a sample of 1000, should shipment of 20,000 be rejected ?

Continuous Distribution – Normal

Normal Distribution



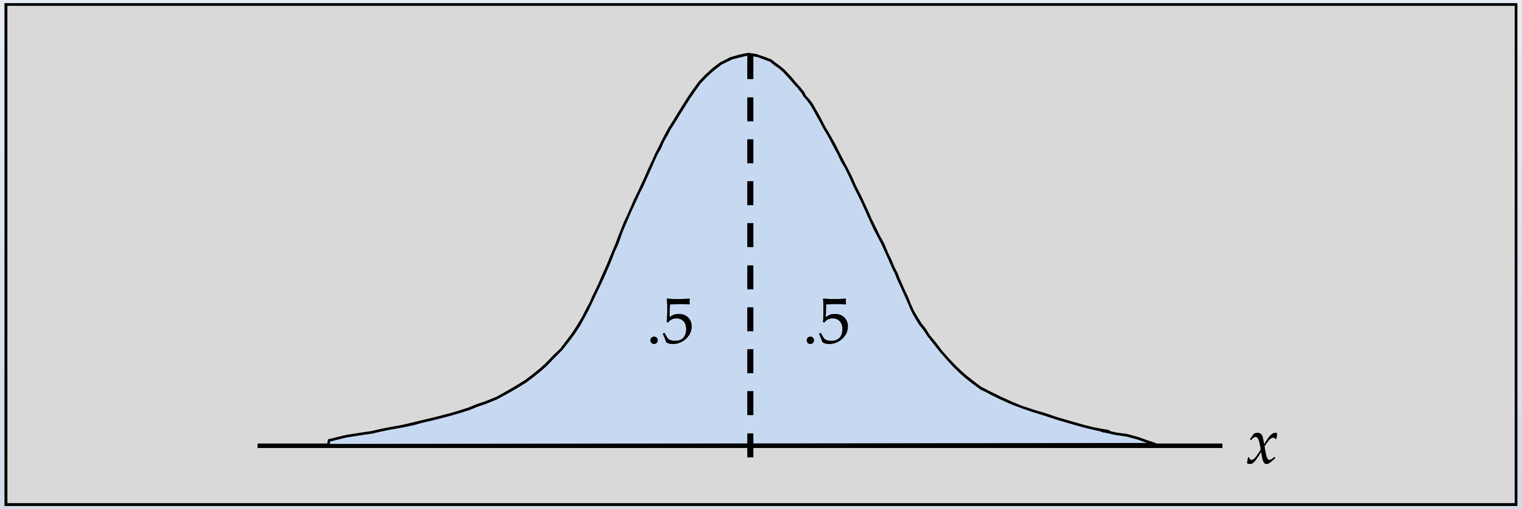
Normal Distribution

- Bell-shaped and also known as the Gaussian curve
- Symmetric around mean
- Continuous
- Never touches x-axis
- The mean, median and mode of a normal distribution are equal
- Total area under the curve is equal to 1
- Normal distributions are denser in the center and less dense in the tails
- Normal distributions are defined by two parameters, the mean, and the standard deviation

$$p(x, \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Normal Distribution

- Characteristics
  - Probabilities for the normal random variable are given by areas under the curve. The total area under the curve is 1 (.5 to the left of the mean and .5 to the right).

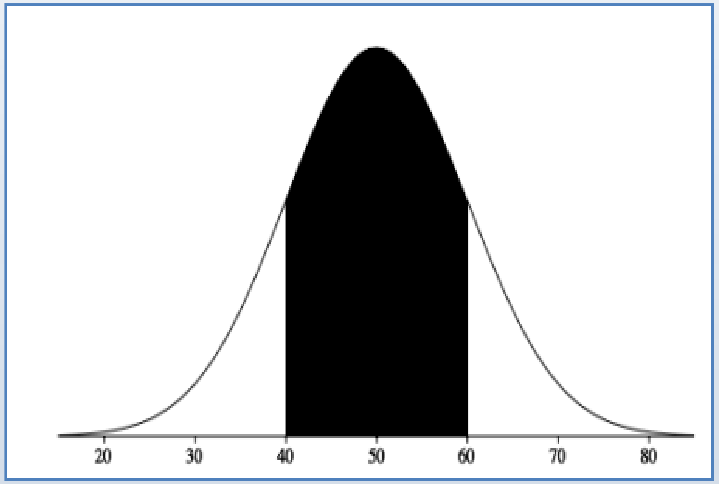


Normal Distribution

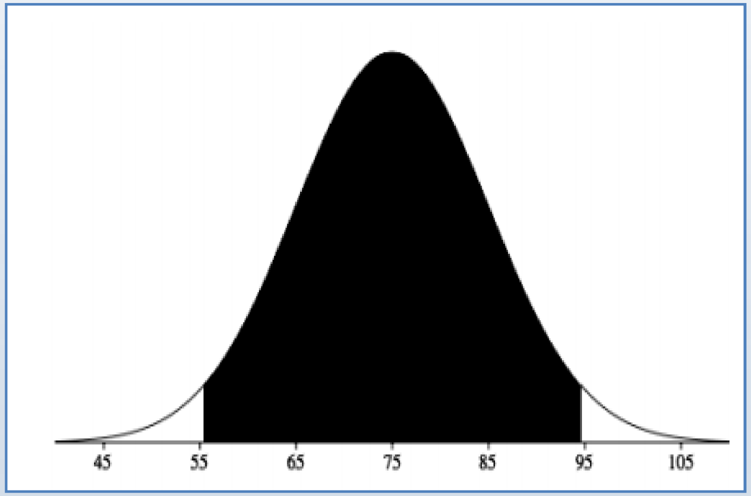
- Characteristics
  - 68.26% of values of a normal random variable are within +/- 1 standard deviation of its mean.
  - 95.44% of values of a normal random variable are within +/- 2 standard deviations of its mean.
  - 99.72% of values of a normal random variable are within +/- 3 standard deviations of its mean.

Normal Distribution

Normal distribution with a mean of 50 and standard deviation of 10. The shaded Area between 40 and 60 contains 68% of the area is within one standard deviation (10) of the mean (50)



Normal distribution with a mean of 75 and standard deviation of 10. The shaded Area between 55 and 95 contains 96% of the area is within two standard deviation of the mean

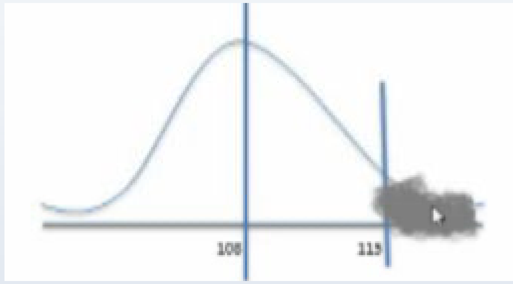


Normal Distribution

Let’s say we are interested in IQ of mightyvel students. We test 100 students and find that IQ is normally distributed with an average of 108, with a standard deviation of 7.

Supposing you pick a random person from those 100 students. What are the chances that student has an IQ > 115 ?

What we know:
IQ Score: **Random variable**
Distribution: **Continuous normal**
Mean = **108**
Std Dev: **7**



What we need : P(Score > 115)

Which is nothing but 1 – P(Score <= 115)

We can of course rely on excel
NORMDIST(Outcome, Mean,Std Dev, Cuml)

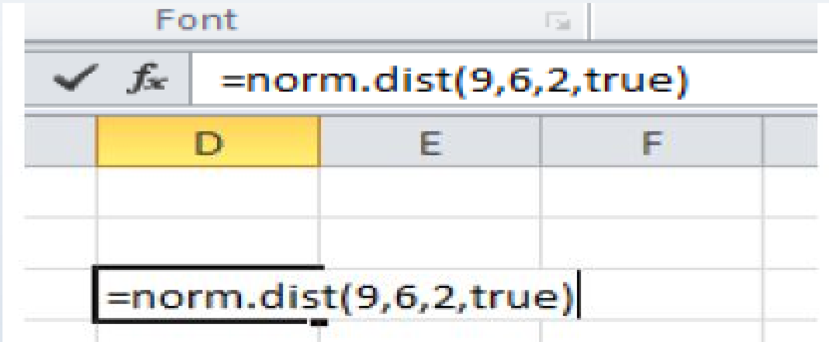
Normal Distribution

We can calculate probabilities of any X given mean and std deviation If total delivery time is normally distributed with a mean of 6 Days and a stddeviation of two days, what is the probability that a random delivery takes less than 9days?

Use Excel:

Formula: **NORM.DIST(Outcome x, Mean, StdDev, cumulative)= NORM.DIST(9,6,2,?)**

Is this the right answer?



Standardized Normal Distribution

Normal Distribution Vs Standard Normal distribution

Standard Normal	Non Standard Normal
Mean = 0	mean is not 0
Variance = 1	variance is not 1

Normal Distribution

- To make life easier, all normal distributions can be converted to a standard normal distribution. A standard normal distribution has a mean of 0 and a standard deviation of 1
- The area under the normal curve between any two points represents the probability

- Mean is zero
- Variance is 1
- Standard deviation is 1
- Data values represented by z
- Probability function given by

$$p(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2}$$

Suppose that X is normally distribtued with mean mu and variance sigma^2 , then the probability density function of the normal RV is given by

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

- If mean is zero and variance is one, the normal RV is referred as the standard normal RV, and the standard is to denote it by Z

- **Uses of Normal Distribution**
  - It helps managers/management make decisions
  - The major use of normal distribution is the role of it plays in statistical inference.

Standard Normal distribution

The letter Z is used to designate the standard normal random variable
z standardized normal table ,

Converting to the standard normal distribution requires the use of this formula

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

z indicates how many standard deviations away from the mean the point x lies

Example Problem

Test scores of a special examination administered to all potential employees of a firm are normally distributed with a mean of 500 points and a standard deviation of 100 points. What is the probability that a score selected at random will be higher than 700?

$P(x > 700) = ?$

If we convert this normal variable, x, to a standard normal variable, z,

$z = (x - \quad) / \sigma = (700 - 500) / 100 = 2$

--------------500----------700   x-scale

$P(x > 700) = P(z > 2)$ ----------------0-----------2   z-scale

Applications of Standard Normal Distribution

- If the behavior of a continuous random variable X is described by the distribution N((μ , σ^2) then the behavior of the random variable z = (X-μ)/ σ is described by the standard normal distribution
- We call the Z standardized normal table

- If the random variable X is described by the distribution N(45,0.000625) then

  what is the transformation to obtain the standardized normal variable?

Assignments

- Assuming the normal heart rate (H.R) in normal healthy individuals is normally distributed with Mean = 70 and Standard Deviation =10 beats/min

  1) What area under the curve is above 80 beats/min?
  2) What area of the curve is above 90 beats/min ?
  3) What area of the curve is between 50-90 beats/min ?
  4) What area of the curve is above 100 beats/min ?
  5) What area of the curve is below 40 beats/min or above 100 beats/min ?

<u>In R</u>

- 1 - pnorm(80,70,10)
- 1 - pnorm(90,70,10)

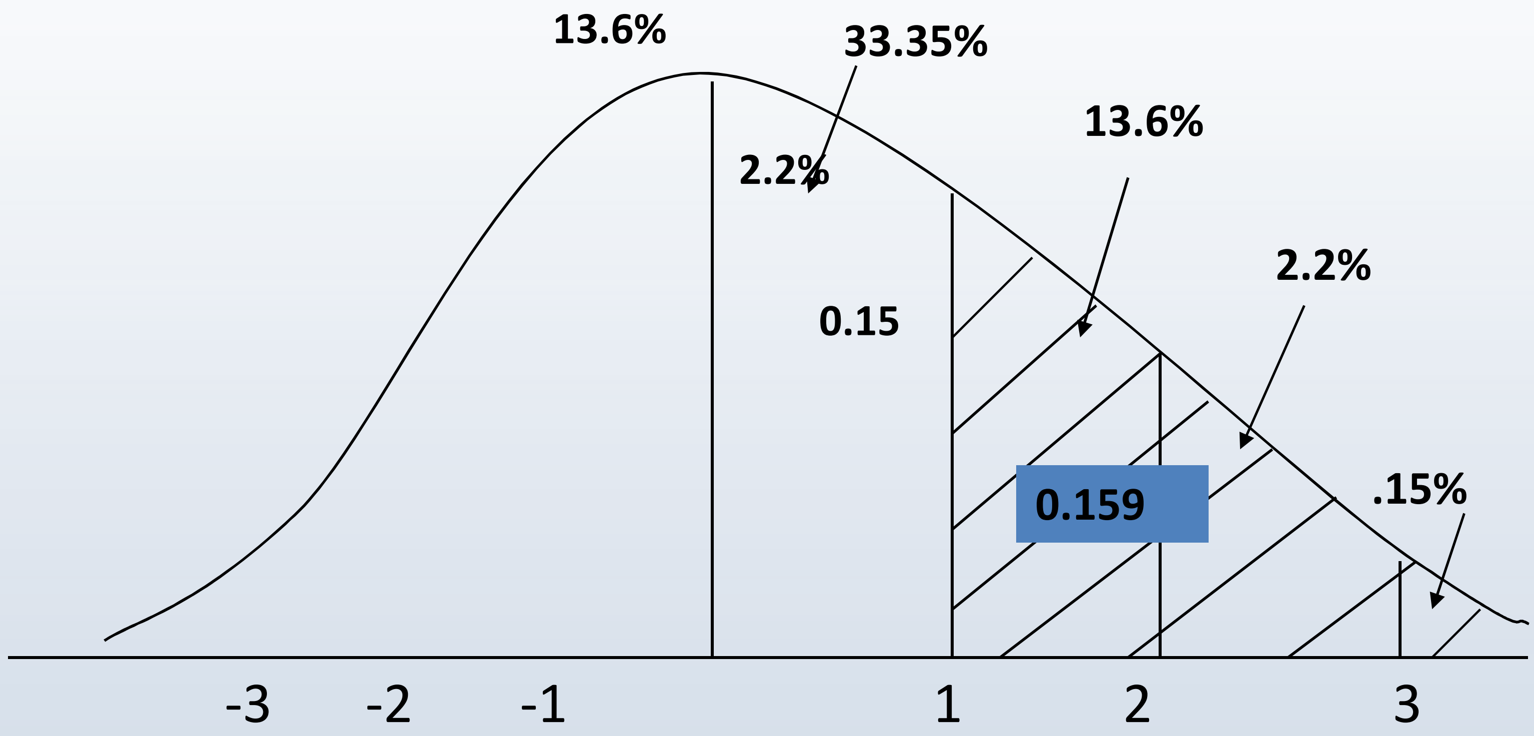
z score probability table

Table A-1 The Standard Normal Distribution

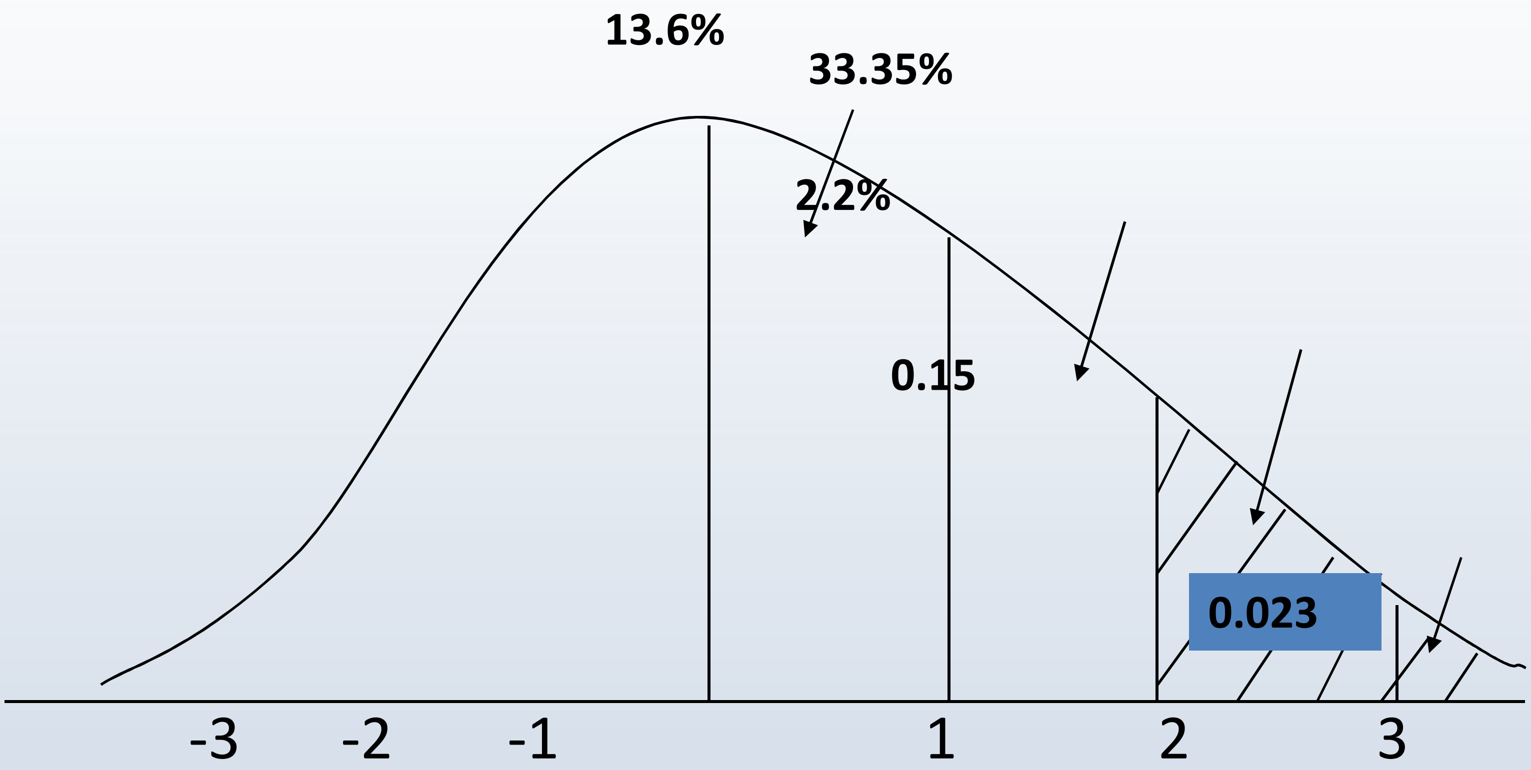
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233

1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

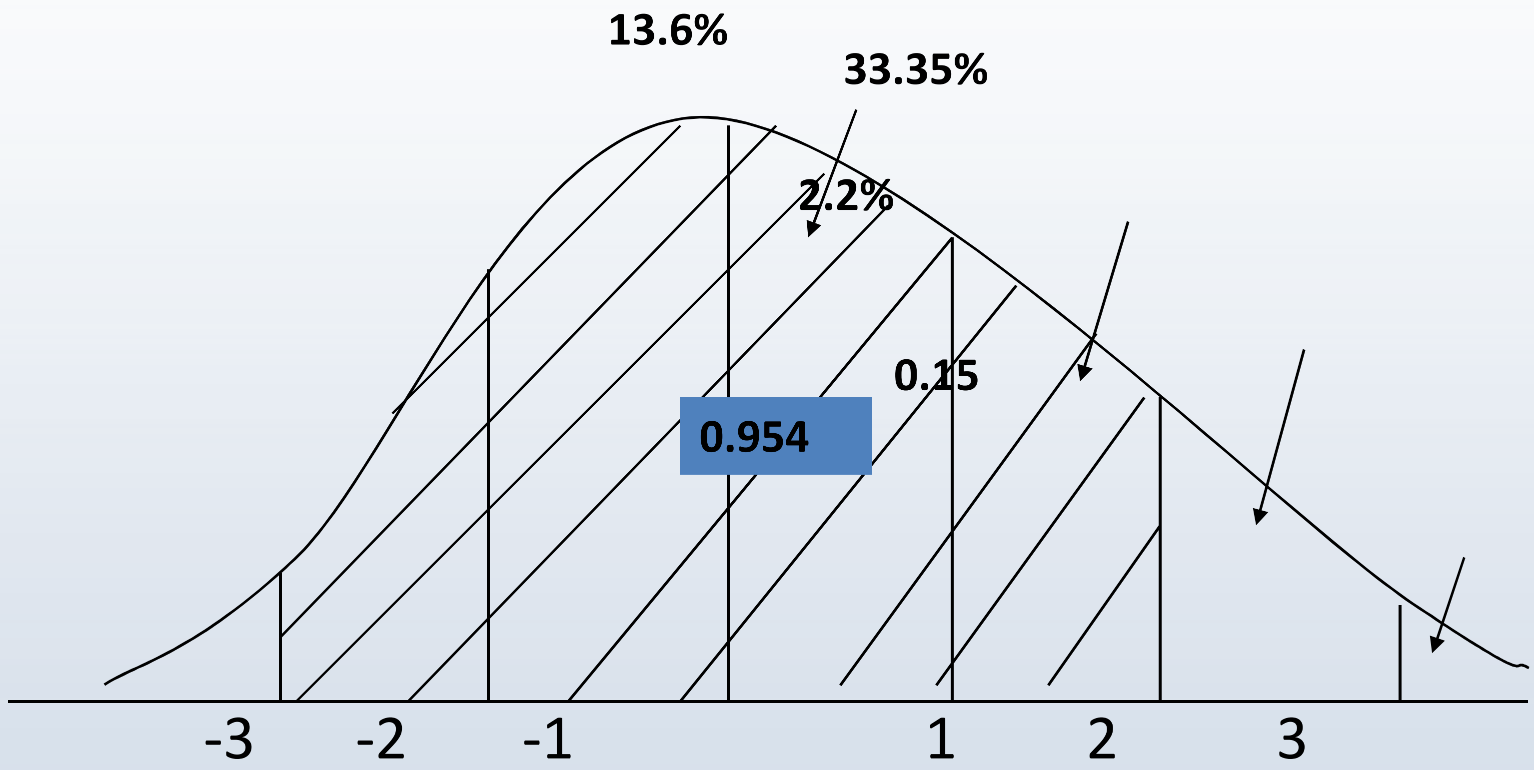
Applications of Standard Normal Distribution



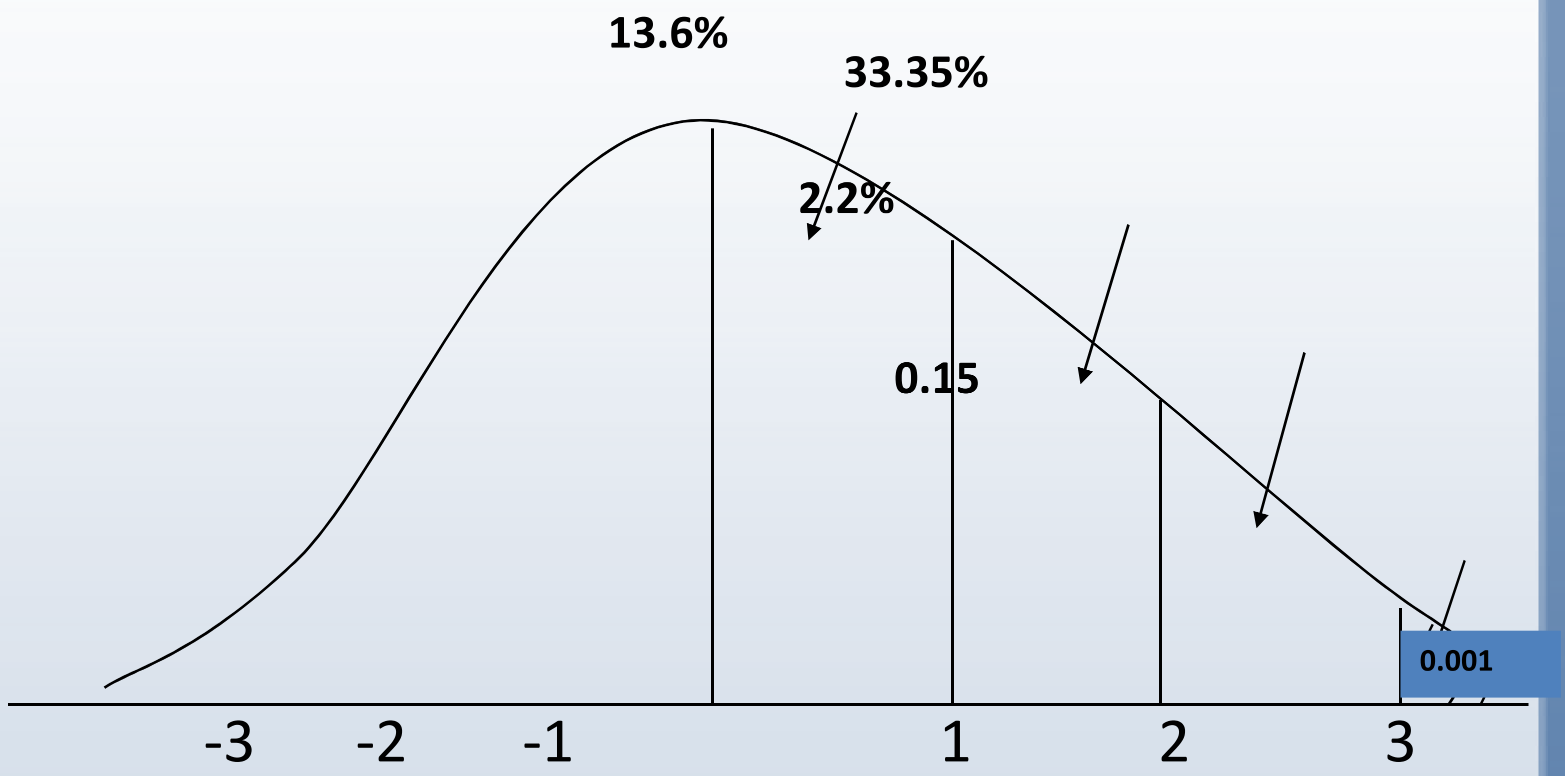
Exercise # 2 solution



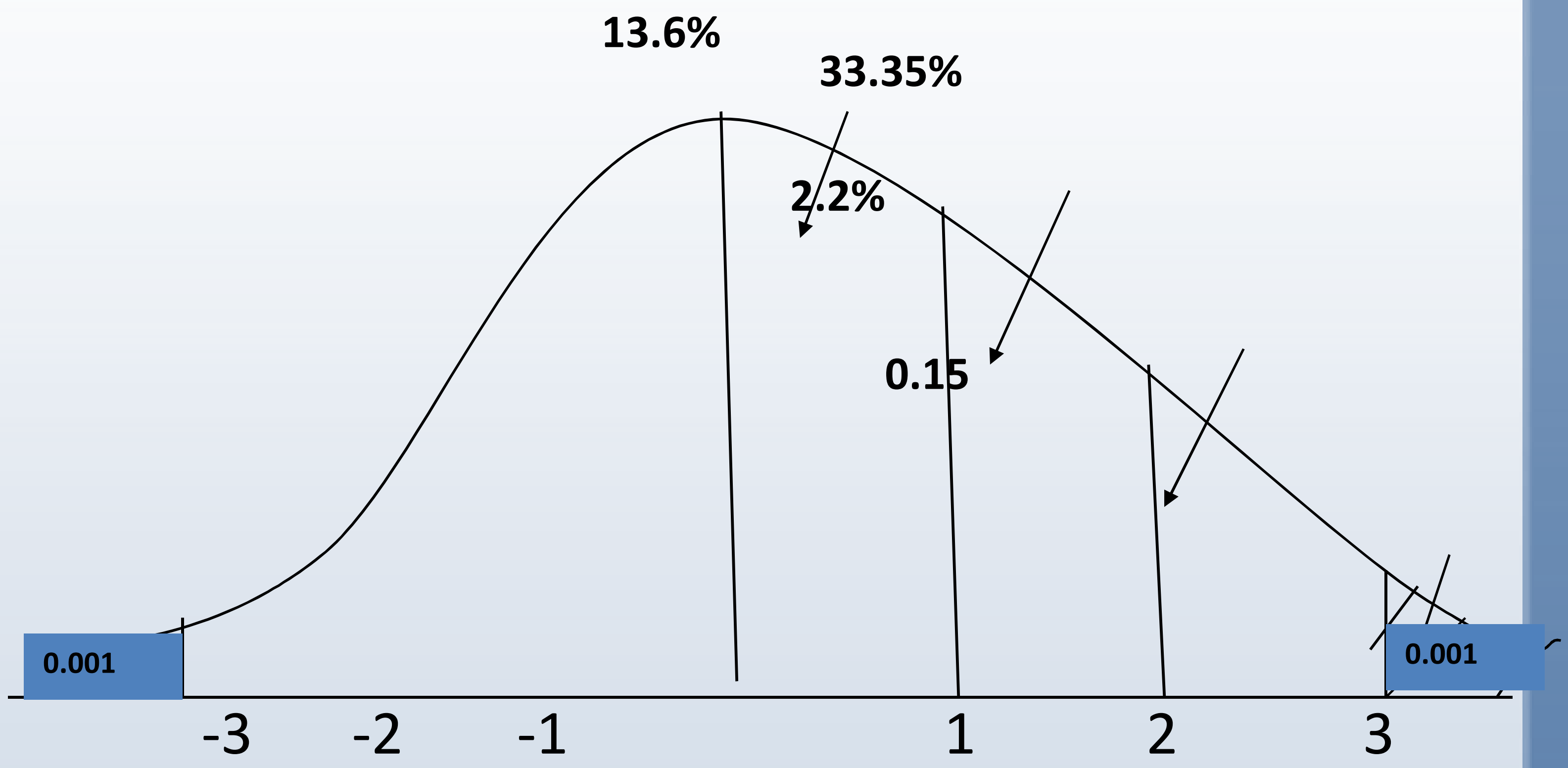
Exercise # 3



Exercise # 4 - Solution



Exercise # 5 - Solution



Solutions

Calculation of the problems and Interpretation of results:

1) For calculation of exercise # 1 see earlier slide. The result of exercise # 1 is 15.9%. This means that 15.9% of normal healthy individuals have a heart rate above one standard deviation (greater than 80 beats per minute).

2) **Calculation for exercise #2**

$Z = \frac{X-}{\sigma}$ $z = \frac{90-70}{10} = 20/10 = 2.00.$ If we look at the normal distribution tables, then the z value of 2.00 corresponds to 0.023 or 2.3%

This means that 2.3% of normal healthy individuals have a heart rate above two standard deviation (greater than 90 beats per minute).

3) **Calculation for exercise # 3**

$Z = \frac{X-}{\sigma}$ $Z1 = \frac{50-70}{10} = -20/10 = -2.00$ and $Z2 = 90-70/10 = 2.00.$

The area between -2 standard deviations and +2 standard deviations from The z tables is 0.954 or 95.4%.

This means that 95.4% have a heart rate between -2 and +2 standard deviations (between 50 -90 beats per minute).

4) Calculation for exercise #4

- Again, $z = \frac{X-}{\sigma} = \frac{100-70}{10} = 30/10 = 3.00$ From the z tables the value of 3.00 corresponds to 0.015 or 0.15%
- This means that only 0.15% have a heart rate above 3 standard deviations (greater than 100 beats per minute).

5) For calculations for this question, please see the earlier slide on this problem and diagram. The answer is 0.3%. This means that only 0.3% have a heart rate either below or above 3 standard deviations (less than 40 or greater than 100 beats per minute).

Normal Distribution – using R

- To find the probability density at a given value, use the dnorm function:

  pnorm(x, mean = 0, sd = 1)

  pnorm(x, mean = 0, sd = 1, lower = FALSE) <← Complement of the above

Assignments

Assignment 1

In a call centre 180 calls are received at random over a period of 12 hours. The chances of receiving a call is 1/3. Then the probability to receive calls between 50 to 70 is

(a) 0.06

(b) 0.89

(c) 0.95

(d) 0

Assignment 2

- The chances of Mike winning a game are 55%. What is the probability of winning the game in 3rd trial ?

Assignment 3

The average weight of the rice packets of a company is 1000 gms with a standard deviation of 20gms. Then assuming the data is normally distributed

Which one is correct from the below

(a) About 50% of the packets are less than 1000 gms
(b) About 50% of the packets are less than 990 gms
(c) All of these
(d) About 85% of the packets are less than 1020 grams