

CHAPTER 4: Probabilistic Features of Certain Data Distribution (Probability Distributions)

August 2024

NOTE: This presentation is based on the presentation prepared thankfully by Professor Abdullah al-Shiha.

4.1 Introduction

4.2 Probability Distributions of Discrete Random Variables

4.3 Binomial Distribution

4.4 Poisson Distribution

4.5 Continuous Probability Distributions

4.6 Normal Distribution

4.7 Normal Distribution Application

4.8 t-Distribution

Introduction

The concept of random variables is very important in Statistics. Some events can be defined using random variables.

There are two types of random variables:

1. Discrete Random variables
2. Continuous Random Variables

Probability Distributions of Discrete Random Variables

Definition :

The probability distribution of a discrete random variable is a table, graph, formula or other device used to specify all possible values of the random variable along with their respective probabilities.

Examples of discrete r v.'s :

- The no. of patients visiting KKUH in a week.
- The no. of times a person had a cold in last year.

Example:

Consider the following discrete random variable :

X = The number of times a Saudi person had a cold in January 2010

Suppose we are able to count the no. of Saudis which $X = x$:

x (no. of colds a Saudi person had in January 2010)	Frequency of x (no. of Saudi people who had a cold x times in January 2010)
0	10,000,000
1	3,000,000
2	2,000,000
3	1,000,000
Total	$N = 16,000,000$

Note that the possible values of the random variable X are:

$$x = 0, 1, 2, 3$$

Experiment:

Selecting a person at random

Define the event:

$(X = 0)$ = The event that the selected person had no cold.

$(X = 1)$ = The event that the selected person had 1 cold.

$(X = 2)$ = The event that the selected person had 2 colds.

$(X = 3)$ = The event that the selected person had 3 colds.

In general

$(X = x)$ = The event that the selected person had x colds.

For this experiment, there are $n(\Omega) = 16,000,000$ equally likely outcomes .

The number of elements of the event $(X=x)$ is :

$n(X=x)$ = no. of Saudi people who had a cold x times in January 2010 .

= frequency of x .

The probability of event $(X=x)$ is :

$$P(X=x) = \frac{n(X=x)}{n(\Omega)} = \frac{n(X=x)}{16,000,000}, \quad \text{for } x=0, 1, 2, 3$$

x	freq. of x $n(X=x)$	$P(X=x) = \frac{n(X=x)}{16000}$ (Relative frequency)
0	10000000	0.6250
1	3000000	0.1875
2	2000000	0.1250
3	1000000	0.0625
Total	16000000	1.0000

Note :

$$P(X=x) = \frac{n(X=x)}{16,000,000} = \text{Relative frequency} = \frac{\text{frequency}}{16,000,000}$$

The probability distribution of the discrete random variable X is given by the following table:

x	$P(X = x) = f(x)$
0	0.6250
1	0.1874
2	0.1250
3	0.0625
Total	1.0000

Notes:

- The probability distribution of any discrete random variable X must satisfy the following two properties:
 - (1) $0 < P(X = x) < 1$
 - (2) $\sum_x P(X = x) = 1$
- Using the probability distribution of a discrete r.v. we can find the probability of r.v. X .

Example:

Consider the discrete r.v. X in the previous example.

$$(1) P(X \geq 2) = P(X = 2) + P(X = 3) = 0.1250 + 0.0625 = 0.1875$$

$$(2) P(X > 2) = P(X = 3) = 0.0625 \quad [\text{note: } P(X > 2) \neq P(X \geq 2)]$$

$$(3) P(1 \leq X < 3) = P(X = 1) + P(X = 2) = 0.1875 + 0.1250 = 0.3125$$

$$(4) P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \\ = 0.6250 + 0.1875 + 0.1250 = 0.9375$$

Another solution :

$$P(X \leq 2) = 1 - P(X > 2) \\ = 1 - P(X = 3) \\ = 1 - 0.0625 \\ = 0.9375$$

$$(5) P(-1 \leq X < 2) = P(X = 0) + P(X = 1) = 0.6250 + 0.1875 = 0.8125$$

x	$P(X = x)$
0	0.6250
1	0.1875
2	0.1250
3	0.0625
Total	1.0000

$$(6) P(-1.5 \leq X < 1.3) = P(X = 0) + P(X = 1) = 0.6250 + 0.1875 = 0.8125$$

$$(7) P(X = 3.5) = P(\phi) = 0$$

$$(8) P(X \leq 10) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = P(\Omega) = 1$$

(9) The probability that the selected person had at least 2 colds :

$$P(X \geq 2) = P(X = 2) + P(X = 3) = 0.1875$$

(10) The probability that the selected person had at most 2 colds :

$$P(X \leq 2) = 0.9375$$

(11) The probability that the selected person had more than 2 colds :

$$P(X > 2) = P(X = 3) = 0.0625$$

(12) The probability that the selected person had less than 2 colds :

$$P(X < 2) = P(X = 0) + P(X = 1) = 0.8125$$

x	$P(X = x)$
0	0.6250
1	0.1875
2	0.1250
3	0.0625
Total	1.0000

Graphical Presentation:

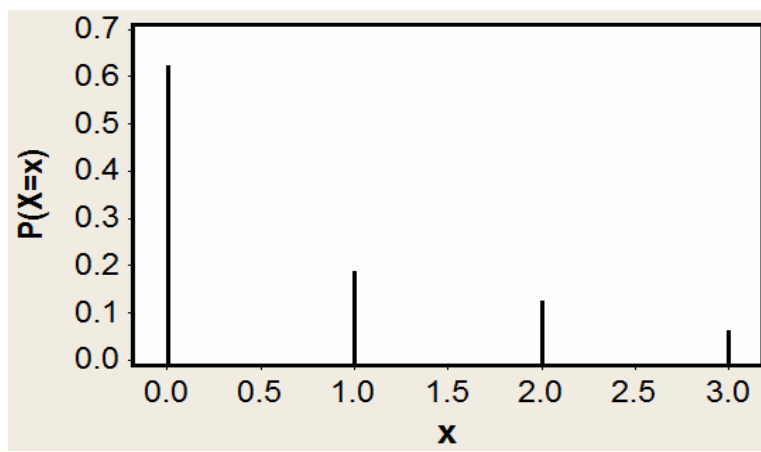
The probability distribution of a discrete r. v. X can be graphically represented .

Example:

The probability distribution of the random variable in the previous example is:

x	$P(X = x)$
0	0.6250
1	0.1875
2	0.1250
3	0.0625

The graphical presentation of this probability distribution is given by the following figure:



Find the Probability
that the selected
person had no cold
in January 2010 ?

Mean and Variance of a Discrete Random Variable :

Mean :

The mean (or expected value) of a discrete random variable X is denoted by μ_X or μ .

It is defined by:

$$\mu = \sum_x x P(X = x)$$

Variance:

The variance of a discrete random variable X is denoted by σ_X^2 or σ^2

It is defined by:

$$\sigma^2 = \sum_x (x - \mu)^2 P(X = x)$$

Example:

We wish to calculate the mean and the variance of the discrete r. v. X whose probability distribution is given by the following table:

x	$P(X = x)$
0	0.05
1	0.25
2	0.45
3	0.25

Solution:

x	$P(X = x)$	$x P(X = x)$
0	0.05	0
1	0.25	0.25
2	0.45	0.9
3	0.25	0.75
Total	1	$\sum x P(X = x)$ = 1.9

$$\mu = \sum_x x P(X = x)$$

$$= 0 \cdot 0.05 + 1 \cdot 0.25 + 2 \cdot 0.45 + 3 \cdot 0.25 = 1.9 = \text{Mean}$$

$$\sigma^2 = \sum_x (x - 1.9)^2 P(X = x)$$

$$= 0.69$$

Cumulative Distributions:

The cumulative distribution function of a discrete r. v. X is defined by :

$$P(X \leq x) = \sum_{a \leq x} P(X = a)$$

Example:

Calculate the cumulative distribution of the discrete r. v. X whose probability distribution is given by the following table:

x	$P(X = x)$
0	0.05
1	0.25
2	0.45
3	0.25

Use the cumulative distribution to find:

$$P(X \leq 2), P(X < 2), P(X \leq 1.5), P(X < 1.5), P(X > 1), P(X \geq 1) .$$

Solution :

The cumulative distribution of X is :

Using the cumulative distribution

$$P(X \leq 2) = 0.75$$

$$P(X < 2) = P(X \leq 1) = 0.30$$

$$P(X \leq 1.5) = P(X \leq 1) = 0.30$$

$$P(X < 1.5) = P(X \leq 1) = 0.30$$

$$P(X > 1) = 1 - P(X \leq 1)$$

$$= 1 - 0.3$$

$$= 0.7$$

$$P(X \geq 1) = 1 - P(X < 1)$$

$$= 1 - P(X \leq 0)$$

$$= 1 - 0.05$$

$$= 0.95$$

x	$P(X \leq x)$
0	0.05
1	0.30
2	0.75
3	1.0000

$$P(X \leq 0) = P(X = 0)$$

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

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

probability distribution

x	$P(X = x)$
0	0.05
1	0.25
2	0.45
3	0.25

cumulative distribution

x	$P(X \leq x)$
0	0.05
1	$0.05+0.25= 0.30$
2	$0.05+0.25+0.45= 0.75$
3	$0.05+0.25+0.45+0.25=1$

cumulative distribution

x	$P(X \leq x)$
0	0.05
1	0.30
2	0.75
3	1

probability distribution

x	$P(X = x)$
0	0.05
1	$0.30-0.05= 0.25$
2	$0.75 -0.30 = 0.45$
3	$1 - 0.75 =0.25$

Complement of probability:

- $P(X \leq a) = 1 - P(X > a)$
- $P(X < a) = 1 - P(X \geq a)$
- $P(X \geq a) = 1 - P(X < a)$
- $P(X > a) = 1 - P(X \leq a)$

Example: (Reading Assignment)

Given the following probability distribution of a discrete random variable X representing the number of defective teeth of the patient visiting a certain dental clinic :

A- Find the value of K .

B- Find the following probabilities:

1- $P(X < 3)$

2- $P(X \leq 3)$

3- $P(X < 6)$

4- $P(X < 1)$

5- $P(X=3.5)$

C- Find the probability that the patient has at least 4 defective teeth.

D- Find the probability that the patient has at most 2 defective teeth.

E- Find the expected number of defective teeth (mean of X).

F- Find the variance of X .

x	$P(X = x)$
1	0.25
2	0.35
3	0.20
4	0.15
5	K

Solution:

$$A- \sum_x P(X = x) = 1$$

$$0.25+0.35+ 0.20+0.15+k=1 \implies 0.95+k=1 \implies k=0.05$$

The probability distribution of X is:

x	P(X = x)
1	0.25
2	0.35
3	0.20
4	0.15
5	0.05
Total	1.00

B- Finding the probabilities:

$$1- P(X < 3) = P(X=1)+P(X=2) = 0.25+0.35 = 0.60$$

$$2- P(X \leq 3) = P(X=1)+P(X=2)+P(X=3) = 0.25+0.35+0.20 = 0.8$$

$$3- P(X < 6) = P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5) = 1$$

$$4- P(X < 1) = 0$$

$$5- P(X=3.5) = 0$$

C- The probability that the patient has at least 4 defective teeth :

$$P(X \geq 4) = P(X=4) + P(X=5) = 0.15 + 0.05 = 0.20$$

D- The probability that the patient has at most 2 defective teeth :

$$P(X \leq 2) = P(X=1) + P(X=2) = 0.25 + 0.35 = 0.6$$

E- The expected number of defective teeth (mean of X) :

x	P(X = x)
1	0.25
2	0.35
3	0.20
4	0.15
5	K=0.05

X	P(X = x)	x P(X = x)
1	0.25	0.25
2	0.35	0.70
3	0.20	0.60
4	0.15	0.60
5	0.05	0.25
Total	$\sum P(X = x) = 1$	$\mu = \sum x P(X = x) = 2.4$

$$\mu = \sum x P(X = x) = (1)(0.25) + (2)(0.35) + (3)(0.2) + (4)(0.15) + (5)(0.05) = 2.4$$

F- The variance of X:

x	$P(X = x)$	$(x - \mu)$	$(x - \mu)^2$	$(x - \mu)^2 P(X = x)$
1	0.25	-1.4	1.96	0.49
2	0.35	-0.4	0.16	0.056
3	0.20	0.6	0.36	0.072
4	0.15	1.6	2.56	0.384
5	0.05	2.6	6.76	0.338
Total	$\mu = 2.4$			$\sigma^2 = \sum (x - \mu)^2 P(X = x)$ $= 1.34$

The variance is :

$$\sigma^2 = \sum (x - \mu)^2 P(X = x) = 1.34$$

Combinations:

Notation (n!) :

n! is read (n factorial) .

It defined by :

$$n! = n(n-1)(n-2) \dots (2)(1) \quad \text{for } n \geq 1$$

Examples :

$$5! = (5)(4)(3)(2)(1) = 120$$

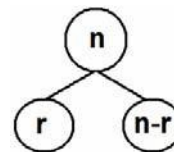
$$0! = 1$$

Combinations:

The number of different ways for selecting r objects from n distinct objects is denoted by C_r^n or $\binom{n}{r}$ and is given by :

$$C_r^n = \frac{n!}{r!(n-r)!}$$

for $n=0,1,2,\dots,n$



Notes:

1. C_r^n is read as “ n “ choose “ r ”.
2. $C_n^n = 1$, $C_0^n = 1$
3. $C_r^n = C_{n-r}^n$ (for example : $C_3^{10} = C_7^{10}$)
4. C_r^n = number of unordered subsets of a set of (n) objects such that each subset contains (r) objects.

Example: For $n = 4$ and $r = 2$:

$$C_2^4 = \frac{4!}{2!(4-2)!} = \frac{4!}{2! \times 2!} = \frac{4 \times 3 \times 2 \times 1}{(2 \times 1)(2 \times 1)} = 6$$

$C_2^4 = 6$ = The number of different ways for selecting 2 objects from 4 distinct objects.

Example: Suppose that we have the set {a, b, c, d} of ($n=4$) objects.

We wish to choose a subset of two objects. The possible subsets of this set with 2 elements in each subset are:

{a , b}, {a , c}, {a , d}, {b , d}, {b , c}, {c , d}

The number of these subset is : $C_2^4 = 6$

4.3 Binomial Distribution:

- Bernoulli Trial:

Is an experiment with only two possible outcomes: S = success and F = failure (Boy or girl, Saudi or non-Saudi, sick or well, dead or alive).

- Binomial distribution is a discrete distribution.
- Binomial distribution is used to model an experiment for which:
 1. The experiment has a sequence of n Bernoulli trials.
 2. The probability of success is $P(S) = p$, and the probability of failure is $P(F) = 1 - p = q$.
 3. The probability of success $P(S) = p$ is constant for each trial.
 4. The trials are independent; that is the outcome of one trial has no effect on the outcome of any other trial.

In this type of experiment, we are interested in the discrete r. v. representing the number of successes in the n trials.

X = The number of successes in the n trials

The possible values of X (number of success in n trials) are:

$$x = 0, 1, 2, \dots, n$$

The r.v. X has a binomial distribution with parameters n and p , and we write:

$$X \sim \text{Binomial}(n, p)$$

The probability distribution of X is given by:

$$P(X = x) = \begin{cases} C_x^n p^x q^{n-x} & ; \quad \text{for } x=0,1,2,\dots,n \\ 0 & ; \quad \text{otherwise} \end{cases}$$

Where : $C_x^n = \frac{n!}{x!(n-x)!}$

We can write the probability distribution of X as a table as follows :

x	P(X=x)
0	$C_0^n p^0 q^{n-0} = q^n$
1	$C_1^n p^1 q^{n-1}$
2	$C_2^n p^2 q^{n-2}$
⋮	⋮
n-1	$C_{n-1}^n p^{n-1} q^1$
n	$C_n^n p^n q^0 = p^n$
Total	1

Result : (Mean and Variance for Binomial distribution)

If $X \sim \text{Binomial}(n, p)$, then

The mean: $\mu = np$ (expected value)

The variance: $\sigma^2 = npq$

Example:

Suppose that the probability that a Saudi man has high blood pressure is 0.15. Suppose that we randomly select a sample of 6 Saudi men.

1. Find the probability distribution of the random variable (X) representing the number of men with high blood pressure in the sample.
2. Find the expected number of men with high blood pressure in the sample (mean of X).
3. Find the variance X.
4. What is the probability that there will be exactly 2 men with high blood pressure?
5. What is the probability that there will be at most 2 men with high blood pressure?
6. What is the probability that there will be at least 4 men with high blood pressure?

Solution:

We are interested in the following random variable:

X = The number of men with high blood pressure in the sample of 6 men.

Notes:

- Bernoulli trial: diagnosing whether a man has a high bloodpressure or not. There are two outcomes for each trial:

S = Success: The man has high blood pressure

F = failure: The man does not have high blood pressure.

- Number of trials = 6 (we need to check 6 men)

- Probability of success: $P(S) = p = 0.15$

- Probability of failure: $P(F) = q = 1 - p = 0.85$

- Number of trials: $n = 6$

- The trials are independent because of the fact that the result of each man does not affect the result of any other mansince the selection was made at random.

The random variable X has a binomial distribution with parameters: $n=6$ and $p=0.15$, that is:

$$X \sim \text{Binomial}(n, p)$$

$$X \sim \text{Binomial}(6, 0.15)$$

The possible values of X are: $x = 0, 1, 2, 3, 4, 5, 6$

1. The probability distribution of X is:

$$P(X=x)=\begin{cases} C_x^6 (0.15)^x (0.85)^{6-x} & ; \quad \text{for } x=0,1,2,3,4,5,6 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

The probabilities of all values of X are:

$$P(X=0)= C_0^6 (0.15)^0 (0.85)^{6-0} = 0.37715$$

$$P(X=1)= C_1^6 (0.15)^1 (0.85)^{6-1} = 0.39933$$

$$P(X=2)= C_2^6 (0.15)^2 (0.85)^{6-2} = 0.17618$$

$$P(X=3)= C_3^6 (0.15)^3 (0.85)^{6-3} = 0.04145$$

$$P(X=4)= C_4^6 (0.15)^4 (0.85)^{6-4} = 0.00549$$

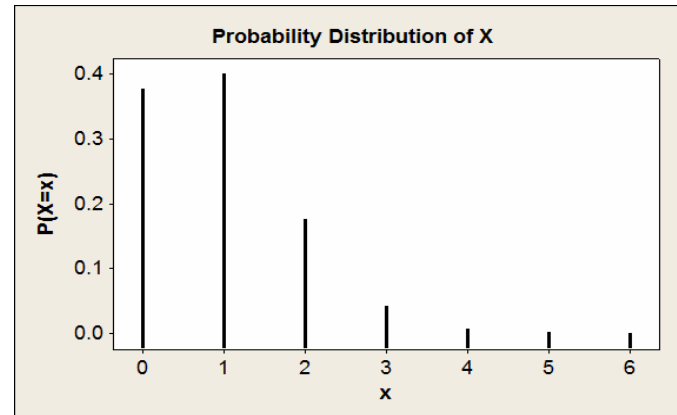
$$P(X=5)= C_5^6 (0.15)^5 (0.85)^{6-5} = 0.00039$$

$$P(X=6)= C_6^6 (0.15)^6 (0.85)^{6-6} = 0.00001$$

The probability distribution of X can be presented by the following table:

x	$P(X = x)$
0	0.37715
1	0.39933
2	0.17618
3	0.04145
4	0.00549
5	0.00039
6	0.00001
Total	1

The probability distribution of X can be presented by the following graph:



2. The mean of the distribution (the expected number of men out of 6 with high blood pressure) is:

$$\mu = np = (6)(0.15) = 0.9$$

3. The variance is:

$$\sigma^2 = npq = (6)(0.15)(0.85) = 0.765$$

4. The probability that there will be exactly 2 men with high blood pressure is:

$$P(X = 2) = 0.17618$$

5. The probability that there will be at most 2 men with high blood pressure is:

$$\begin{aligned} P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= 0.37715 + 0.39933 + 0.17618 \\ &= 0.95266 \end{aligned}$$

6. The probability that there will be at least 4 men with high blood pressure is:

$$\begin{aligned} P(X \geq 4) &= P(X=4) + P(X=5) + P(X=6) \\ &= 0.00549 + 0.00039 + 0.00001 \\ &= 0.00589 \end{aligned}$$

Example: (Reading Assignment)

Suppose that 25% of the people in a certain population have low hemoglobin levels. The experiment is to choose 5 people at random from this population. Let the discrete random variable X be the number of people out of 5 with low hemoglobin levels .

- Find the probability distribution of X .
- Find the probability that at least 2 people have low hemoglobin levels.
- Find the probability that at most 3 people have low hemoglobin levels.
- Find the expected number of people with low hemoglobin levels out of the 5 people.
- Find the variance of the number of people with low hemoglobin levels out of the 5 people.

Solution:

X = the number of people out of 5 with low hemoglobin levels
The Bernoulli trial is the process of diagnosing the person

Success = the person has low hemoglobin

Failure = the person does not have low hemoglobin

$$\begin{array}{ll} n = 5 & \text{(no. of trials)} \\ p = 0.25 & \text{(probability of success)} \\ q = 1 - p = 0.75 & \text{(probability of failure)} \end{array}$$

a) X has a binomial distribution with parameter $n = 5$ and $p = 0.25$
 $X \sim \text{Binomial}(5, 0.25)$

The possible values of X are : $x=0, 1, 2, 3, 4, 5$

The probability distribution is:

$$P(X=x)=\begin{cases} C_x^5 (0.25)^x (0.75)^{5-x} & ; \quad \text{for } x=0,1,2,3,4,5 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

x	P(X = x)
0	$C_0^5 (0.25)^0 (0.75)^{5-0}=0.23730$
1	$C_1^5 (0.25)^1 (0.75)^{5-1}=0.39551$
2	$C_2^5 (0.25)^2 (0.75)^{5-2}=0.26367$
3	$C_3^5 (0.25)^3 (0.75)^{5-3}=0.08789$
4	$C_4^5 (0.25)^4 (0.75)^{5-4}=0.01465$
5	$C_5^5 (0.25)^5 (0.75)^{5-5}=0.00098$
Total	1

b) The probability that at least 2 people have low hemoglobin levels:

$$\begin{aligned} P(X \geq 2) &= P(X=2)+P(X=3)+P(X=4)+P(X=5) \\ &= 0.26367+ 0.08789+ 0.01465+ 0.00098 \\ &= 0.36719 \end{aligned}$$

c) The probability that at most 3 people have low hemoglobin levels:

$$\begin{aligned} P(X \leq 3) &= P(X=0)+P(X=1)+P(X=2)+P(X=3) \\ &= 0.23730+ 0.39551+ 0.26367+ 0.08789 \\ &= 0.98437 \end{aligned}$$

d) The expected number of people with low hemoglobin levels out of the 5 people (the mean of X):

$$\mu = np = 5(0.25) = 1.25$$

e) The variance of the number of people with low hemoglobin levels out of the 5 people (the variance of X) is:

$$\sigma^2 = npq = 5(0.25)(0.75) = 0.9375$$

4.4 The Poisson Distribution:

- It is a discrete distribution.
- The Poisson distribution is used to model a discrete r.v representing the number of occurrences of some random event in an interval of time or space (or some volume of matter).
- The possible values of X are: $x = 0, 1, 2, 3, \dots$
- The discrete r. v. X is said to have a Poisson distribution with parameter (average or mean) λ if the probability distribution of X is given by :

$$P(X = x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & ; \quad \text{for } x = 0, 1, 2, \dots \\ 0 & ; \quad \text{otherwise} \end{cases}$$

Where $e = 2.71828$

We write : $X \sim \text{Poisson}(\lambda)$

Result: (Mean and Variance of Poisson distribution)

If $X \sim \text{Poisson}(\lambda)$, then:

The mean (average) of X is : $\mu = \lambda$ (Expected value)

The variance of X is: $\sigma^2 = \lambda$

The standard deviation is : $\sigma = \sqrt{\lambda}$

Example :

Some random quantities that can be modeled by Poisson distribution:

- No. of patients in a waiting room in an hours.
- No. of surgeries performed in a month.
- No. of rats in each house in a particular city.

Note:

- λ is the average (mean) of the distribution.
- If X = The number of patients seen in the emergency unit **in a day** , and if $X \sim \text{Poisson } (\lambda)$ then :
 1. The average (mean) of patients seen every day in the emergency unit = λ .
 2. The average (mean) of patients seen every month in the emergency unit = 30λ .
 3. The average (mean) of patients seen every year in the emergency unit = 365λ .
 4. The average (mean) of patients seen every hour in the emergency unit = $\lambda/24$.

Also, notice that:

- If Y = The number of patients seen every month, then:

$Y \sim \text{Poisson } (\lambda^*)$, where $\lambda^* = 30\lambda$

- W = The number of patients seen every year, then:

$W \sim \text{Poisson } (\lambda^*)$, where $\lambda^* = 365\lambda$

- V = The number of patients seen every hour, then:

$V \sim \text{Poisson } (\lambda^*)$, where $\lambda^* = \frac{\lambda}{24}$

Example:

Suppose that the number of snake bites cases seen at KKHU in year has a Poisson distribution with average 6 bite cases.

- (1) What is the probability that in a year:
 - (i) The no. of snake bite cases will be 7 ?
 - (ii) The no. of snake bite cases will be less than 2 ?
- (2) What is the probability that there will be 10 snake bite cases in 2 years?
- (3) What is the probability that there will be no snake bite cases in a month?

Solution:

- (1) X = no. of snake bite cases in a year.

$X \sim \text{Poisson}(6)$

$$P(X = x) = \frac{e^{-6} 6^x}{x!} \quad x = 0, 1, 2, \dots$$

i. $P(X=7) = \frac{e^{-6} 6^7}{7!} = 0.13678$

ii. $P(X < 2) = P(X=0) + P(X=1) = \frac{e^{-6} 6^0}{0!} + \frac{e^{-6} 6^1}{1!} = 0.00248 + 0.01487 = 0.01735$

(2) Y= no of snake bite cases in 2 years

$$X \sim \text{Poisson (12)} \quad (\lambda^* = 2\lambda = 6(2) = 12)$$

$$P(Y = y) = \frac{e^{-12} 12^y}{y!} \quad y=0,1,2,\dots$$

$$P(Y = 10) = \frac{e^{-12} 12^{10}}{10!} = 0.1048$$

(3) W= no of snake bite cases in a month

$$W \sim \text{Poisson (0.5)} \quad (\lambda^* = \frac{\lambda}{12} = \frac{6}{12} = 0.5)$$

$$P(W = w) = \frac{e^{-0.5} 0.5^w}{w!} \quad w=0,1,2,\dots$$

$$P(W = 0) = \frac{e^{-0.5} 0.5^0}{0!} = 0.6065$$

Extra questions :

(4) Find the probability that there will be more than or equal one snake bite cases in a month

$$\lambda^* = \frac{\lambda}{12} = \frac{6}{12} = 0.5$$

$$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + \dots \quad (\text{Since } x = 0, 1, 2, 3, 4, 5, \dots)$$

$$= 1 - P(x < 1)$$

$$= 1 - P(X = 0)$$

$$= 1 - \frac{e^{-0.5} 0.5^0}{0!}$$

$$= 1 - 0.6065 = 0.3935$$

(5) The mean of snake bite cases in a year

$$\mu = \lambda = 6$$

(6) The variance of snake bite cases in a month

$$\sigma^2 = \lambda^* = \frac{\lambda}{12} = \frac{6}{12} = 0.5$$

(7) The standard deviation of snake bite cases in 2 years :

$$\sigma = \sqrt{\lambda^*} = \sqrt{2\lambda} = \sqrt{2(6)} = \sqrt{12} = 3.4641$$

(8) Find the probability that there will be more than 3 snake bite cases in 2 years :

$$\lambda^* = 2\lambda = 2(6) = 12$$

$$P(X > 3) = 1 - P(X \leq 3)$$

$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$$

$$= 1 - 0.0023$$

$$= 0.9977$$

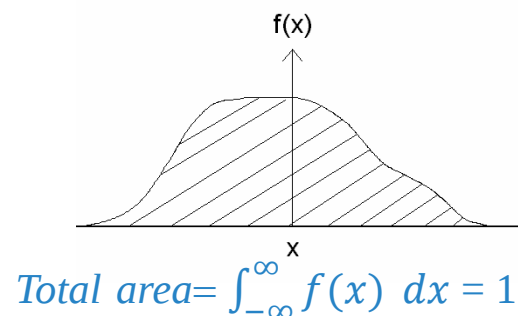
By calculator :

$$\sum_{x=0}^3 \left(\frac{e^{-12} \times 12^x}{x!} \right) = 2.29 \times 10^{-3} = 0.0023$$

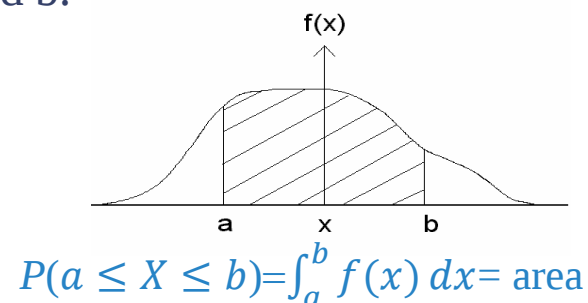
4.5 Continuous Probability Distributions:

For any continuous r.v. X , there exists a function $f(x)$, called the probability density function (pdf) of X , for which:

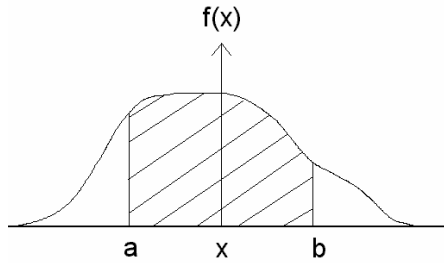
1. The total area under the curve of $f(x)$ equals to 1.



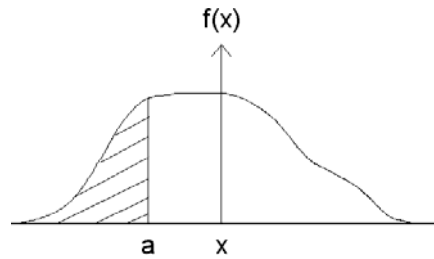
2. The probability that X is between the points (a) and (b) equals to the area under the curve of $f(x)$ which is bounded by the point a and b .



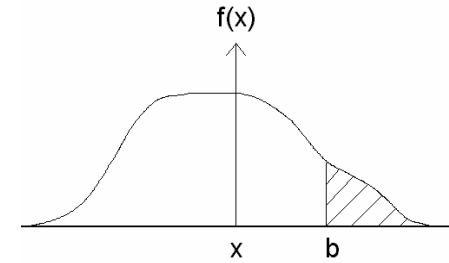
3. In general, the probability of an interval event is given by the area under the curve of $f(x)$ and above that interval.



$$P(a \leq X \leq b) = \int_a^b f(x) dx = \text{area}$$



$$P(X \leq a) = \int_{-\infty}^a f(x) dx = \text{area}$$



$$P(X \geq b) = \int_b^{\infty} f(x) dx = \text{area}$$

Note:

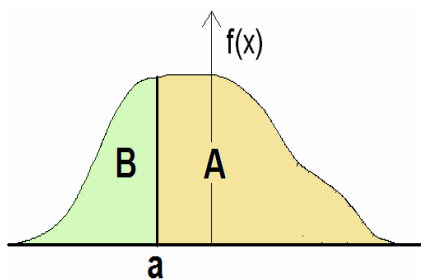
If X is continuous r.v. then:

1. $P(X = a) = 0$ for any a
2. $P(X \leq a) = P(X < a)$
3. $P(X \geq b) = P(X > b)$
4. $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$

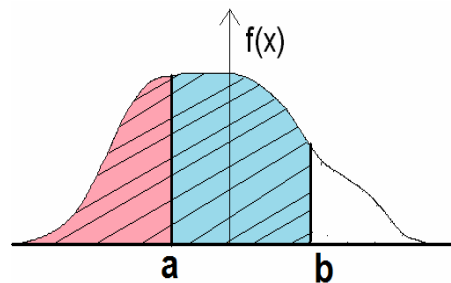
5. $P(X \leq x)$ = cumulative probability

6. $P(X \geq a) = 1 - P(X < a) = 1 - P(X \leq a)$

7. $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$



$$P(X \geq a) = 1 - P(X < a)$$



$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$$

$$\int_a^b f(x) dx = \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx$$

4.6 The Normal Distribution:

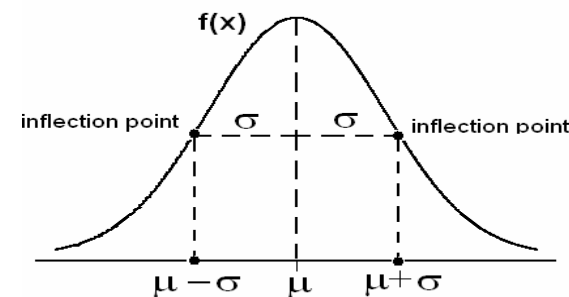
- One of the most important continuous distributions.
- Many measurable characteristics are normally or approximately normally distributed. (Examples: height, weight, ...)
- The probability density function of the normal distribution is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; -\infty < x < \infty$$

where (e=2.71828) and ($\pi=3.14159$)

- The parameters of the distribution are : the mean (μ) and the standard deviation (σ).
- The continuous r.v. X which has a normal distribution has several important characteristics:

1. $-\infty < X < \infty$
2. The density function of X , $f(x)$, has a bell-Shaped curve:



Mean = μ
Standard deviation = σ
Variance = σ^2

3. The highest point of the curve of $f(x)$ at the mean.

4. The curve of $f(x)$ is symmetric about the mean μ .

$$\text{Mean} = \text{median} = \text{mode} = \mu$$

5. The normal distribution depends on two parameters :

Mean = μ (determines the location)

Standard deviation = σ (variance σ^2) (determines the shape)

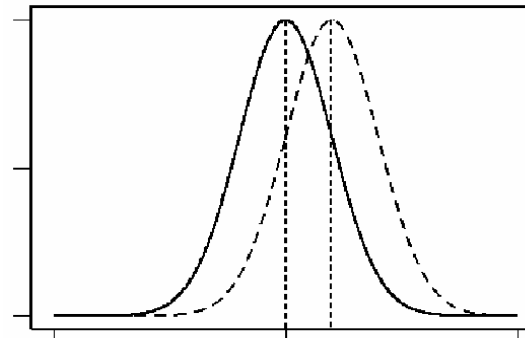
6. If the r.v. X is normally distributed with mean μ and standard deviation σ (variance σ^2), we write:

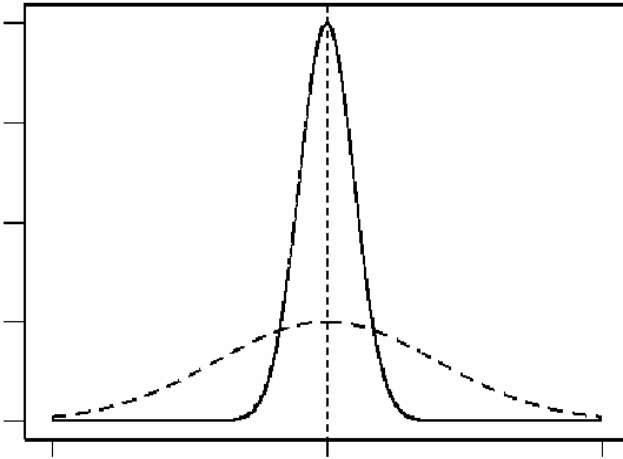
$$X \sim \text{Normal}(\mu, \sigma^2) \quad \text{or} \quad X \sim N(\mu, \sigma^2)$$

7. The location of the normal distribution depends on μ and The shape of the normal distribution depends on σ .

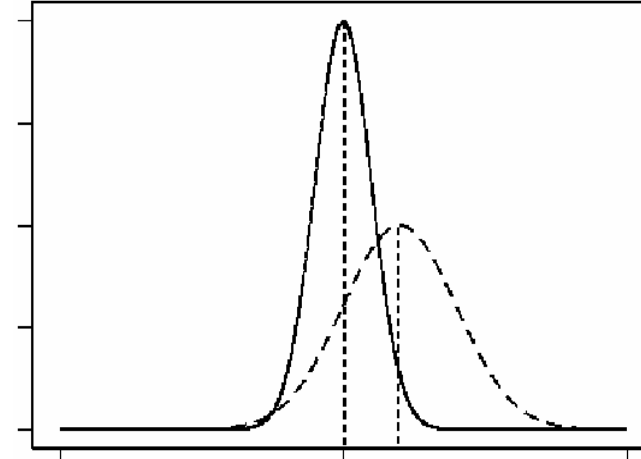
Suppose we have two normal distributions:

$$\begin{array}{l} \text{—————} N(\mu_1, \sigma_1) \\ \text{-----} N(\mu_2, \sigma_2) \end{array}$$





$$\mu_1 = \mu_2, \sigma_1 < \sigma_2$$



$$\mu_1 < \mu_2, \sigma_1 < \sigma_2$$

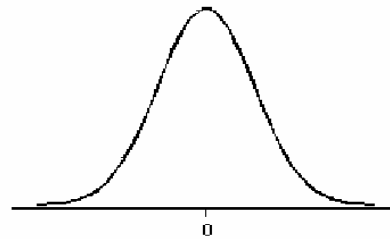
The Standard Normal Distribution:

The normal distribution with mean $\mu=0$ and variance $\sigma^2=1$ is called standard normal distribution and is denoted by Normal(0,1) or N(0,1) . The standard normal random variable is denoted by (Z) , and we write :

$$Z \sim N(0,1)$$

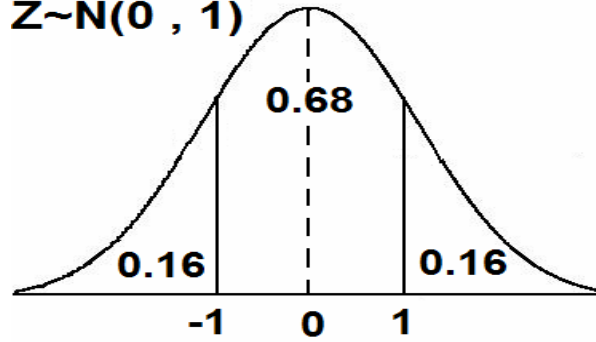
The probability density function (pdf) of $Z \sim N(0,1)$, is given by :

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

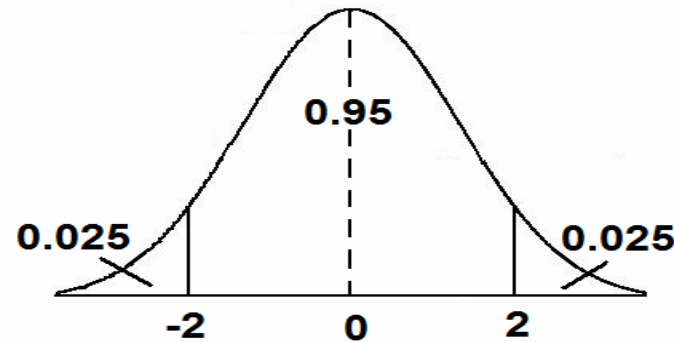


The standard normal distribution, Normal (0,1), is very important because probabilities of any normal distribution can be calculated from the probabilities of the standard normal distribution.

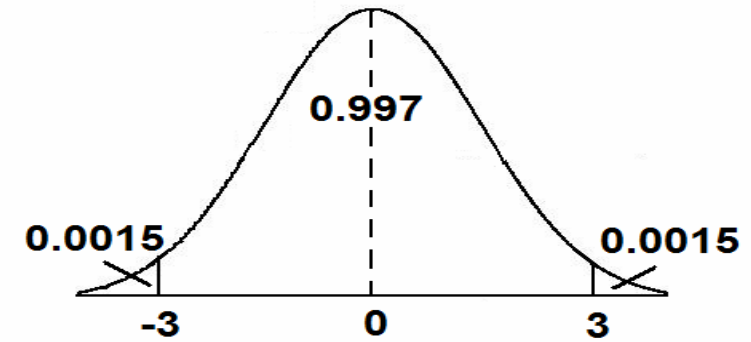
$Z \sim N(0, 1)$



68% of the area is between
-1 and 1
(approximately)



95% of the area is between
-2 and 2
(approximately)



99.7% of the area is between
-3 and 3
(approximately)

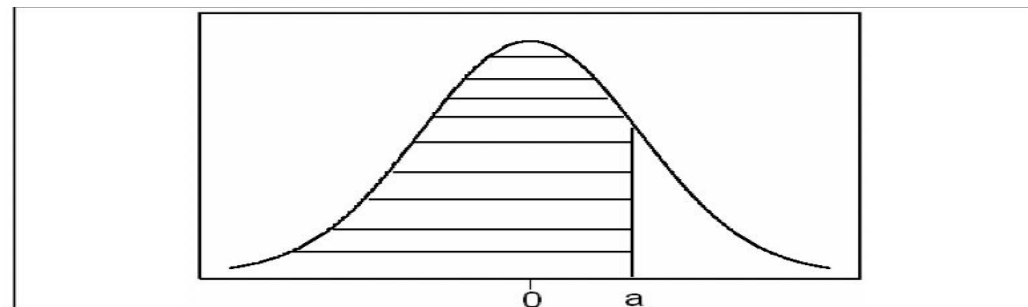
Result :

If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{x - \mu}{\sigma} \sim N(0, 1)$

Calculating Probabilities of Normal (0,1): (Table Z)

Suppose $Z \sim \text{Normal}(0,1)$. For the standard normal distribution $Z \sim N(0,1)$, there is a special table used to calculate probabilities of the form: $P(Z \leq a)$

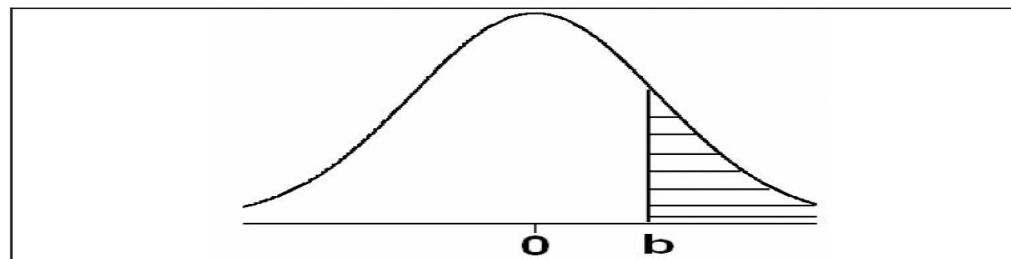
(i) $P(Z \leq a) = \text{From the table}$



(ii) $P(Z \geq b) = 1 - P(Z \leq b)$

Where:

$P(Z \leq b) = \text{From the table}$

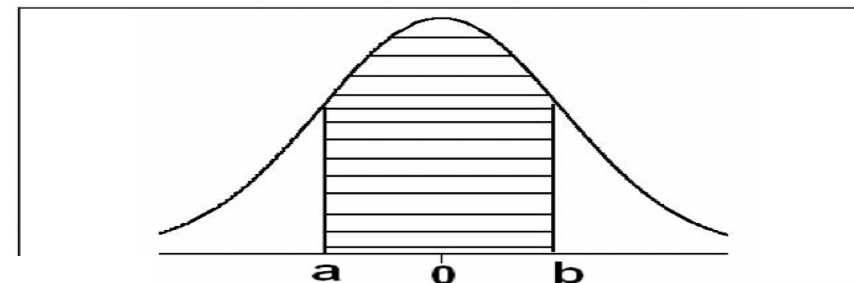


(iii) $P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$

Where:

$P(Z \leq b) = \text{from the table}$

$P(Z \leq a) = \text{from the table}$

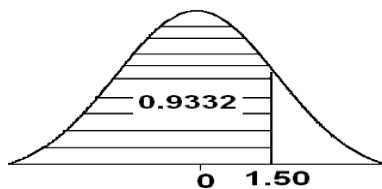


(iv) $P(Z = a) = 0$ for every a .

Example:

Suppose that $Z \sim N(0,1)$

(1) $P(Z \leq 1.50) = 0.9332$

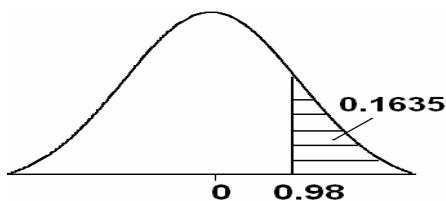


Z	0.00	0.01	...
:	↓		
1.50 ⇒	0.9332		
:			

(2) $P(Z \geq 0.98) = 1 - P(Z \leq 0.98)$

$= 1 - 0.8365$

$= 0.1635$



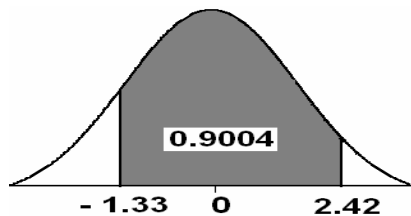
Z	0.00	...	0.08
:	:	:	↓
:	...	...	↓
0.90 ⇒	⇒	⇒	0.8365

$$(3) P(-1.33 \leq Z \leq 2.42)$$

$$= P(Z \leq 2.42) - P(Z \leq -1.33)$$

$$= 0.9922 - 0.0918$$

$$= 0.9004$$



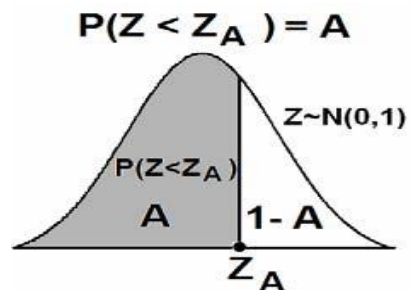
Z	...		-0.03
:	:		↓
-1.30	⇒		0.0918
:			

$$(4) P(Z \leq 0) = P(Z \geq 0) = 0.5$$

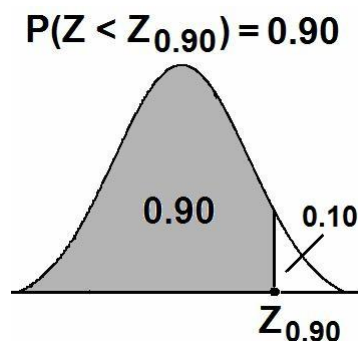
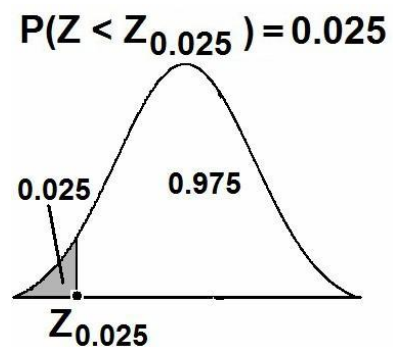
$$(5) P(Z=1.25) = 0$$

Notation:

$$P(Z \leq Z_A) = A$$



For example :



Result:

Since the pdf of $Z \sim N(0,1)$ is symmetric about 0, we have:

$$Z_A = -Z_{1-A}$$

For example:

$$Z_{0.35} = -Z_{1-0.35} = -Z_{0.65}$$

$$Z_{0.86} = -Z_{1-0.86} = -Z_{0.14}$$

$$Z_{0.975} = 1.96$$

$$Z_{0.35} = -Z_{0.975} = -1.96$$

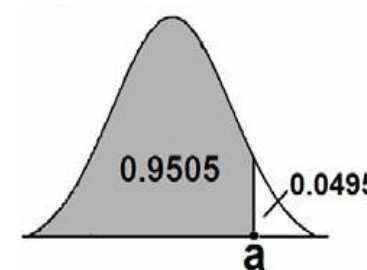
Example:

Suppose that $Z \sim N(0,1)$.

If $P(Z \leq a) = 0.95053$

Then $a = 1.65$

Z	...	0.05	...
:		↑↑	
1.60	⇐	0.95053	
:			



$$\begin{aligned} P(Z < a) &= 0.9505 \\ P(Z < Z_{0.9505}) &= 0.9505 \\ a &= Z_{0.9505} \\ a &= Z_{0.9505} = 1.65 \end{aligned}$$

Example:

Suppose that $Z \sim N(0,1)$.

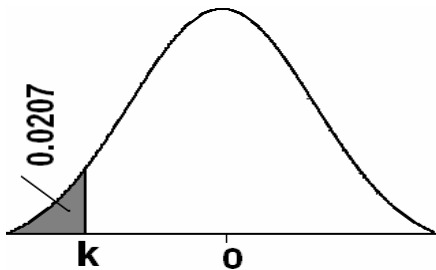
Find the value of k such that

$$P(Z \leq k) = 0.0207.$$

Solution:

$$k = -2.04$$

Notice that $k = Z_{0.0207} = -2.04$



Z	...	-0.04	
:	:	↑↑	
-2.0	⇐⇐	0.0207	
:			

Example:

If $Z \sim N(0,1)$, then:

$$Z_{0.90} = 1.285$$

$$Z_{0.95} = 1.645$$

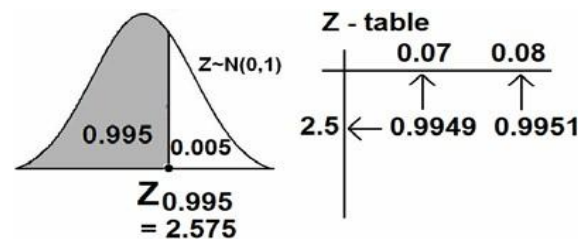
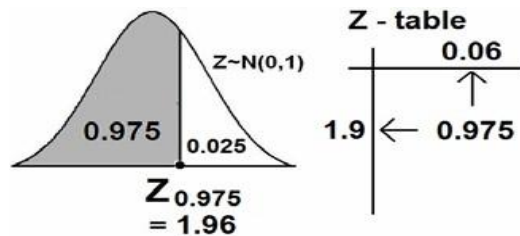
$$Z_{0.975} = 1.96$$

$$Z_{0.99} = 2.325$$

$$Z_{0.90} = (Z_{0.89973} + Z_{0.90147})/2 = (1.28 + 1.29)/2 = 1.285$$

$$Z_{0.95} = (Z_{0.94950} + Z_{0.95053})/2 = (1.64 + 1.65)/2 = 1.645$$

$$Z_{0.99} = (Z_{0.98983} + Z_{0.99010})/2 = (2.32 + 2.33)/2 = 2.325$$



Using the result:

$$Z_A = - Z_{1-A}$$

$$Z_{0.10} = - Z_{0.90} = - 1.285$$

$$Z_{0.05} = - Z_{0.95} = - 1.645$$

$$Z_{0.025} = - Z_{0.975} = - 1.96$$

$$Z_{0.01} = - Z_{0.99} = - 2.325$$

Calculating Probabilities of Normal (μ, σ^2) :

Recall the result : If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{x-\mu}{\sigma} \sim N(0,1)$

$$X \leq a \iff \frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma} \iff Z \leq \frac{a-\mu}{\sigma}$$

$$1. P(X \leq a) = P(Z \leq \frac{a-\mu}{\sigma})$$

$$2. P(X \geq a) = P(Z \geq \frac{a-\mu}{\sigma})$$

$$3. P(a \leq X \leq b) = P(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma})$$

$$= P(Z \leq \frac{b-\mu}{\sigma}) - P(Z \leq \frac{a-\mu}{\sigma})$$

$$4. P(X = a) = 0, \text{ for every } a.$$

4.7 Normal Distribution Application:

Example :

Suppose that the hemoglobin levels of healthy adult males are approximately normally distributed with a mean of 16 and a variance of 0.81.

A. Find that probability that a randomly chosen healthy adult male has a hemoglobin level less than 14.

B. What is the percentage of healthy adult males who have hemoglobin level less than 14 ?

C. In a population of 10,000 healthy adult males, how many would you expect to have hemoglobin level less than 14 ?

Solution :

X = hemoglobin level for healthy adults males

Mean: $\mu = 16$

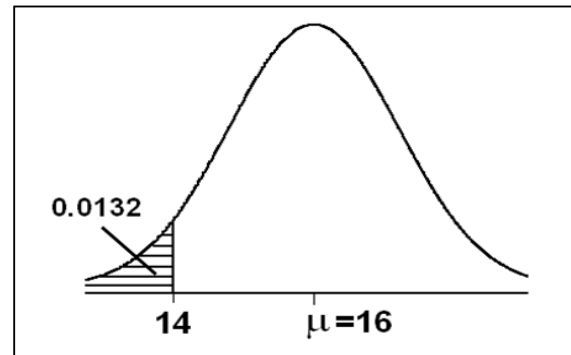
Variance: $\sigma^2 = 0.81$

Standard deviation : $\sigma = 0.9$

$X \sim \text{Normal}(16, 0.81)$

A. The probability that a randomly chosen healthy adult male has hemoglobin level less than 14 is :

$$\begin{aligned} P(X \leq 14) &= P\left(Z \leq \frac{14 - \mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{14 - 16}{0.9}\right) \\ &= P(Z \leq -2.22) \\ &= 0.0132 \end{aligned}$$



B. The percentage of healthy adult males who have hemoglobin level less than 14 is :

$$\text{Percentage} = \text{Probability} \times 100\%$$

$$P(X \leq 14) \times 100 \% = 0.0132 \times 100 \% = 1.32 \%$$

C. In a population of 10000 healthy adult males, we would expect that the number of males with hemoglobin level less than 14 to be: (N =population size =100000)

$$\text{Expect number} = \text{Probability} \times N$$

$$P(X \leq 14) \times 10000 = 0.0132 \times 10000 = 132 \text{ males.}$$

Example:

Suppose that the birth weight of Saudi babies has a normal distribution with mean $\mu=3.4$ and standard deviation $\sigma =0.35$.

A. Find the probability that a randomly chosen Saudi baby has a birth weight between 3.0 and 4.0 kg.

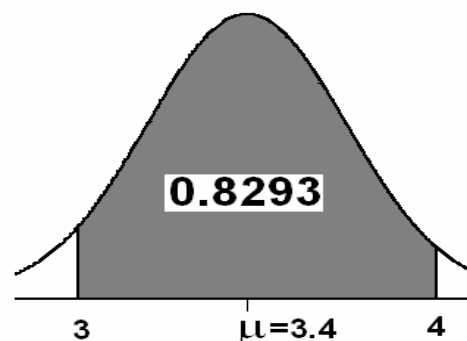
B. What is the percentage of Saudi babies who have a birthweight between 3.0 and 4.0 kg?

C. In a population of 100000 Saudi babies, how many would you expect to have birth weight between 3.0 and 4.0 kg?

Solution:

A. The probability that a randomly chosen Saudi baby has a birth weight between 3.0 and 4.0 kg is :

$$\begin{aligned}
 P(3.0 < X < 4.0) &= P\left(\frac{3-\mu}{\sigma} < Z < \frac{4-\mu}{\sigma}\right) \\
 &= P\left(\frac{3-3.4}{0.35} < Z < \frac{4-3.4}{0.35}\right) \\
 &= P(-1.14 < Z < 1.71) \\
 &= P(Z < 1.71) - P(Z < -1.14) \\
 &= 0.9564 - 0.1271 \\
 &= 0.8293
 \end{aligned}$$



B. The percentage of Saudi babies who have a birth weight between 3.0 and 4.0 kg is

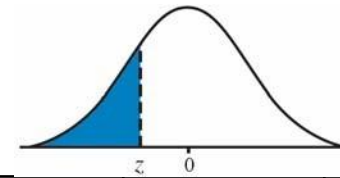
$$P(3.0 < X < 4.0) \times 100\% = 0.8293 \times 100\% = 82.93\%$$

C. In a population of 100,000 Saudi babies, we would expect that the number of babies with birth weight between 3.0 and 4.0 kg to be:

$$P(3.0 < X < 4.0) \times 100000 = 0.8293 \times 100000 = 82930 \text{ babies}$$

Standard Normal Table

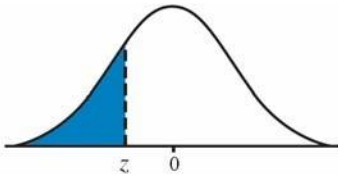
Areas Under the Standard Normal Curve



z	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	-0.00	z
-3.50	0.00017	0.00017	0.00018	0.00019	0.00019	0.00020	0.00021	0.00022	0.00022	0.00023	-3.50
-3.40	0.00024	0.00025	0.00026	0.00027	0.00028	0.00029	0.00030	0.00031	0.00032	0.00034	-3.40
-3.30	0.00035	0.00036	0.00038	0.00039	0.00040	0.00042	0.00043	0.00045	0.00047	0.00048	-3.30
-3.20	0.00050	0.00052	0.00054	0.00056	0.00058	0.00060	0.00062	0.00064	0.00066	0.00069	-3.20
-3.10	0.00071	0.00074	0.00076	0.00079	0.00082	0.00084	0.00087	0.00090	0.00094	0.00097	-3.10
-3.00	0.00100	0.00104	0.00107	0.00111	0.00114	0.00118	0.00122	0.00126	0.00131	0.00135	-3.00
-2.90	0.00139	0.00144	0.00149	0.00154	0.00159	0.00164	0.00169	0.00175	0.00181	0.00187	-2.90
-2.80	0.00193	0.00199	0.00205	0.00212	0.00219	0.00226	0.00233	0.00240	0.00248	0.00256	-2.80
-2.70	0.00264	0.00272	0.00280	0.00289	0.00298	0.00307	0.00317	0.00326	0.00336	0.00347	-2.70
-2.60	0.00357	0.00368	0.00379	0.00391	0.00402	0.00415	0.00427	0.00440	0.00453	0.00466	-2.60
-2.50	0.00480	0.00494	0.00508	0.00523	0.00539	0.00554	0.00570	0.00587	0.00604	0.00621	-2.50
-2.40	0.00639	0.00657	0.00676	0.00695	0.00714	0.00734	0.00755	0.00776	0.00798	0.00820	-2.40
-2.30	0.00842	0.00866	0.00889	0.00914	0.00939	0.00964	0.00990	0.01017	0.01044	0.01072	-2.30
-2.20	0.01101	0.01130	0.01160	0.01191	0.01222	0.01255	0.01287	0.01321	0.01355	0.01390	-2.20
-2.10	0.01426	0.01463	0.01500	0.01539	0.01578	0.01618	0.01659	0.01700	0.01743	0.01786	-2.10
-2.00	0.01831	0.01876	0.01923	0.01970	0.02018	0.02068	0.02118	0.02169	0.02222	0.02275	-2.00
-1.90	0.02330	0.02385	0.02442	0.02500	0.02559	0.02619	0.02680	0.02743	0.02807	0.02872	-1.90
-1.80	0.02938	0.03005	0.03074	0.03144	0.03216	0.03288	0.03362	0.03438	0.03515	0.03593	-1.80
-1.70	0.03673	0.03754	0.03836	0.03920	0.04006	0.04093	0.04182	0.04272	0.04363	0.04457	-1.70
-1.60	0.04551	0.04648	0.04746	0.04846	0.04947	0.05050	0.05155	0.05262	0.05370	0.05480	-1.60
-1.50	0.05592	0.05705	0.05821	0.05938	0.06057	0.06178	0.06301	0.06426	0.06552	0.06681	-1.50
-1.40	0.06811	0.06944	0.07078	0.07215	0.07353	0.07493	0.07636	0.07780	0.07927	0.08076	-1.40
-1.30	0.08226	0.08379	0.08534	0.08691	0.08851	0.09012	0.09176	0.09342	0.09510	0.09680	-1.30
-1.20	0.09853	0.10027	0.10204	0.10383	0.10565	0.10749	0.10935	0.11123	0.11314	0.11507	-1.20
-1.10	0.11702	0.11900	0.12100	0.12302	0.12507	0.12714	0.12924	0.13136	0.13350	0.13567	-1.10
-1.00	0.13786	0.14007	0.14231	0.14457	0.14686	0.14917	0.15151	0.15386	0.15625	0.15866	-1.00
-0.90	0.16109	0.16354	0.16602	0.16853	0.17106	0.17361	0.17619	0.17879	0.18141	0.18406	-0.90
-0.80	0.18673	0.18943	0.19215	0.19489	0.19766	0.20045	0.20327	0.20611	0.20897	0.21186	-0.80
-0.70	0.21476	0.21770	0.22065	0.22363	0.22663	0.22965	0.23270	0.23576	0.23885	0.24196	-0.70
-0.60	0.24510	0.24825	0.25143	0.25463	0.25785	0.26109	0.26435	0.26763	0.27093	0.27425	-0.60
-0.50	0.27760	0.28096	0.28434	0.28774	0.29116	0.29460	0.29806	0.30153	0.30503	0.30854	-0.50
-0.40	0.31207	0.31561	0.31918	0.32276	0.32636	0.32997	0.33360	0.33724	0.3409	0.34458	-0.40
-0.30	0.34827	0.35197	0.35569	0.35942	0.36317	0.36693	0.37070	0.37448	0.37828	0.38209	-0.30
-0.20	0.38591	0.38974	0.39358	0.39743	0.40129	0.40517	0.40905	0.41294	0.41683	0.42074	-0.20
-0.10	0.42465	0.42858	0.43251	0.43644	0.44038	0.44433	0.44828	0.45224	0.45620	0.46017	-0.10
-0.00	0.46414	0.46812	0.47210	0.47608	0.48006	0.48405	0.48803	0.49202	0.49601	0.50000	-0.00

Standard Normal Table(continued)

Areas Under the Standard Normal Curve



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
0.00	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586	0.00
0.10	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535	0.10
0.20	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409	0.20
0.30	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173	0.30
0.40	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793	0.40
0.50	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240	0.50
0.60	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490	0.60
0.70	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524	0.70
0.80	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327	0.80
0.90	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891	0.90
1.00	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214	1.00
1.10	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298	1.10
1.20	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147	1.20
1.30	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774	1.30
1.40	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189	1.40
1.50	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408	1.50
1.60	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449	1.60
1.70	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327	1.70
1.80	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062	1.80
1.90	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670	1.90
2.00	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169	2.00
2.10	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574	2.10
2.20	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899	2.20
2.30	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158	2.30
2.40	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361	2.40
2.50	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520	2.50
2.60	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643	2.60
2.70	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736	2.70
2.80	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807	2.80
2.90	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861	2.90
3.00	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900	3.00
3.10	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929	3.10
3.20	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950	3.20
3.30	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965	3.30
3.40	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976	3.40
3.50	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983	3.50

4.8 The t-Distribution:

- Student's t distribution.
- t-distribution is a distribution of a continuous random variable.
- Recall that, if $X_1, X_2, \dots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2 ,

i.e. $N(\mu, \sigma^2)$, then :

$$Z = \frac{x - \mu}{\sigma} \sim N(0, 1)$$

We can apply this result only when σ^2 is known!

If σ^2 is unknown, we replace the population variance σ^2 with the sample variance $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$

to have the following statistic

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Recall :

if $X_1, X_2, \dots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2 is unknown i.e. $N(\mu, \sigma^2)$, then the statistic :

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

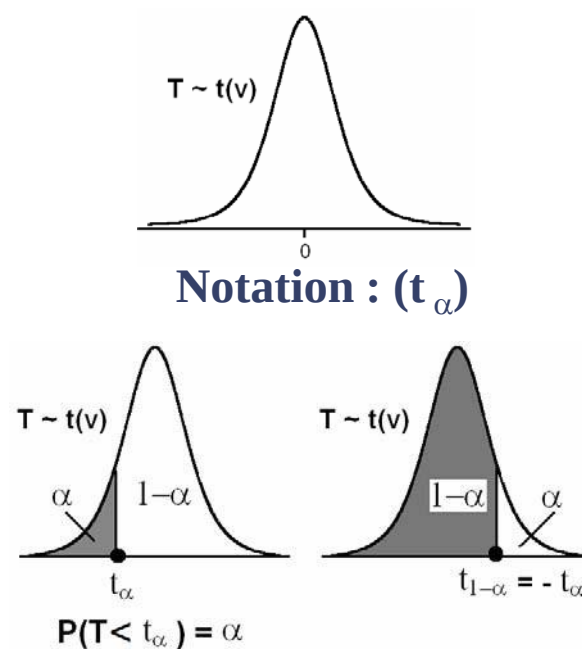
has a t-distribution with $(n-1)$ degrees of freedom ($df=n-1$) , and we write

$$T \sim t(v) \text{ or } T \sim t(n-1)$$

Note :

- t-distribution is a continuous distribution.
- The value of t random variable range from $-\infty$ to $+\infty$ (that is, $-\infty < t < \infty$).
- The mean of t distribution is 0.
- It is symmetric about the mean 0.

- The shape of t-distribution is similar to the shape of the standard normal distribution.
- t-distribution $\rightarrow$ Standard normal distribution as $n \rightarrow \infty$.



- t_α = The t-value under which we find an area equal to α = The t-value that leaves an area of α to the left.
- The value t_α satisfies: $P(T < t_\alpha) = \alpha$.

- Since the curve of the pdf of $T \sim t(v)$ is symmetric about 0, we have

$$t_{1-\alpha} = -t_{\alpha}$$

$$t_{0.1} = -t_{1-0.1} = -t_{0.9}$$

$$t_{0.975} = -t_{1-0.975} = -t_{0.025}$$

- Values of t_{α} are tabulated in a special table for several values of α and several values of degrees of freedom.

Example:

Find the t-value with $v=14$ (df) that leaves an area of:

- 0.95 to the left.
- 0.95 to the right.

Solution:

$v = 14$. (df); $T \sim t(14)$

- The t-value that leaves an area of 0.95 to the left is: $t_{0.95} = 1.761$.

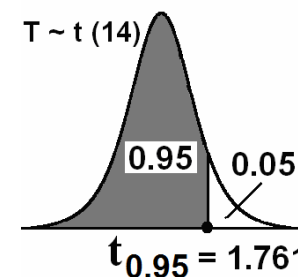
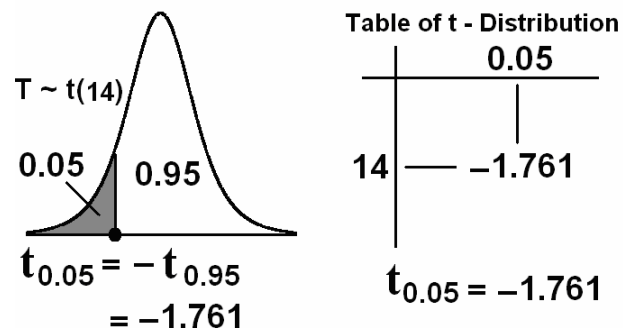


Table of t - Distribution

	0.95
14	1.761

$t_{0.95} = 1.761$

(b) The t-value that leaves an area of 0.95 to the right is : $t_{0.05} = -t_{1-0.05} = -t_{0.95} = -1.761$



Note:

Some t-tables contain values of α that are greater than or equal to 0.90. When we search for small values of α in these tables, we may use the fact that:

$$t_{1-\alpha} = -t_{\alpha}$$

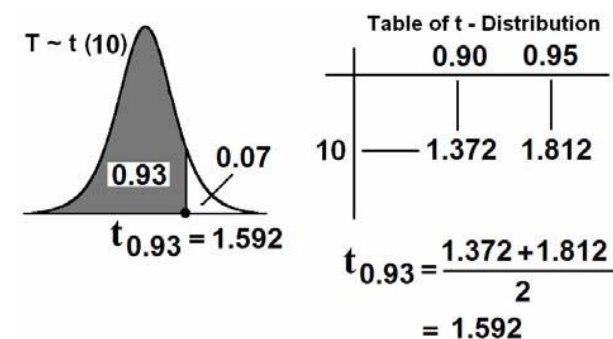
Example:

For $v = 10$ degrees of freedom (df), find $t_{0.93}$ and $t_{0.07}$

Solution

$$t_{0.93} = (1.372 + 1.812) / 2 = 1.592 \quad (\text{from the table})$$

$$t_{0.07} = -t_{1-0.07} = -t_{0.93} = -1.592 \quad (\text{using the rule: } t_{1-\alpha} = -t_{\alpha})$$



Examples

The t-value that leaves an area of 0.975 to the left (use $\nu = 12$) is : $t_{0.975} = 2.179$

- The t-value that leaves an area of 0.90 to the right (use $\nu = 16$) is : $t_{0.10} = - t_{1-0.10} = - t_{0.90} = -1.337$
- The t-value that leaves an area of 0.025 to the right (use $\nu = 8$) is : $t_{0.975} = 2.306$
- The t-value that leaves an area of 0.025 to the left (use $\nu = 8$) is : $t_{0.025} = - t_{1-0.025} = - t_{0.975} = -2.306$
- The t-value that leaves an area of 0.93 to the left (use $\nu = 10$) is : $t_{0.93} = \frac{t_{0.90} + t_{0.95}}{2} = \frac{1.372 + 1.812}{2} = 1.592$
- The t-value that leaves an area of 0.07 to the left (use $\nu = 10$) is : $t_{0.70} = - t_{0.93} = - \left(\frac{t_{0.90} + t_{0.95}}{2} \right) = -1.592$

- $P(T < K) = 0.90$, $df = 10$

$$K = 1.372$$

- $P(T \geq K) = 0.95$, $df = 15$

$$K = -1.753$$

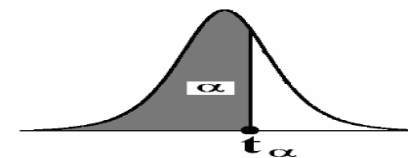
- $P(T < 2.110) = ?$, $df = 17$

$$P(T < 2.110) = 0.975$$

- $P(T \leq 2.718) = ?$, $df = 11$

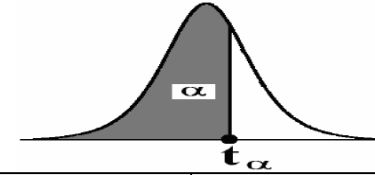
$$P(T \leq 2.718) = 0.99$$

Critical Values of the t -distribution (t_α)



$v=df$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787

Critical Values of the t -distribution (t_α)



$\nu=df$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
35	1.3062	1.6896	2.0301	2.4377	2.7238
40	1.3030	1.6840	2.0210	2.4230	2.7040
45	1.3006	1.6794	2.0141	2.4121	2.6896
50	1.2987	1.6759	2.0086	2.4033	2.6778
60	1.2958	1.6706	2.0003	2.3901	2.6603
70	1.2938	1.6669	1.9944	2.3808	2.6479
80	1.2922	1.6641	1.9901	2.3739	2.6387
90	1.2910	1.6620	1.9867	2.3685	2.6316
100	1.2901	1.6602	1.9840	2.3642	2.6259
120	1.2886	1.6577	1.9799	2.3578	2.6174
140	1.2876	1.6558	1.9771	2.3533	2.6114
160	1.2869	1.6544	1.9749	2.3499	2.6069
180	1.2863	1.6534	1.9732	2.3472	2.6034
200	1.2858	1.6525	1.9719	2.3451	2.6006
∞	1.282	1.645	1.960	2.326	2.576