Data Analytics

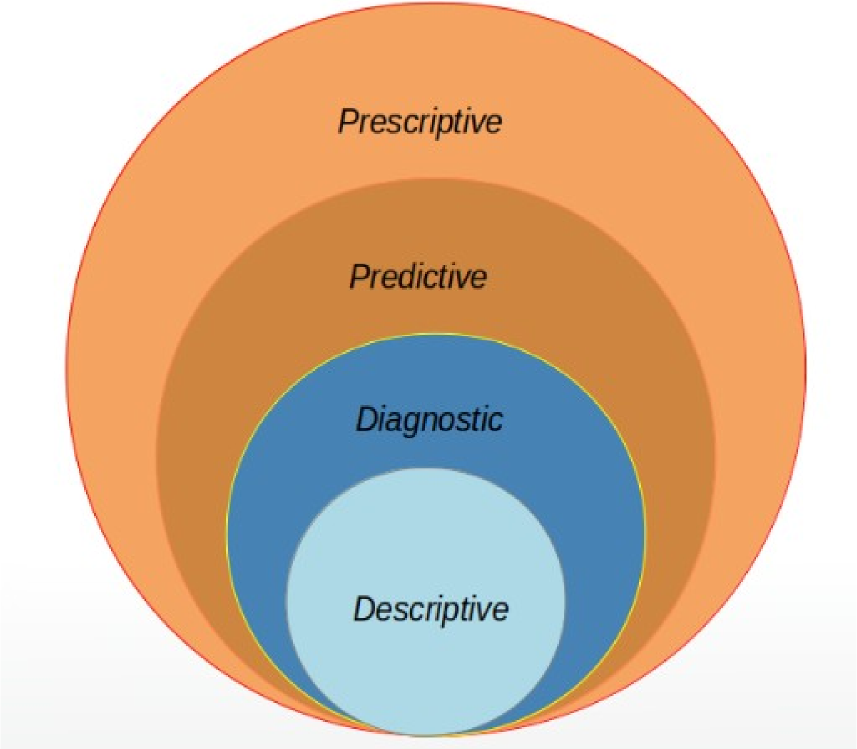
Why Data Analytics

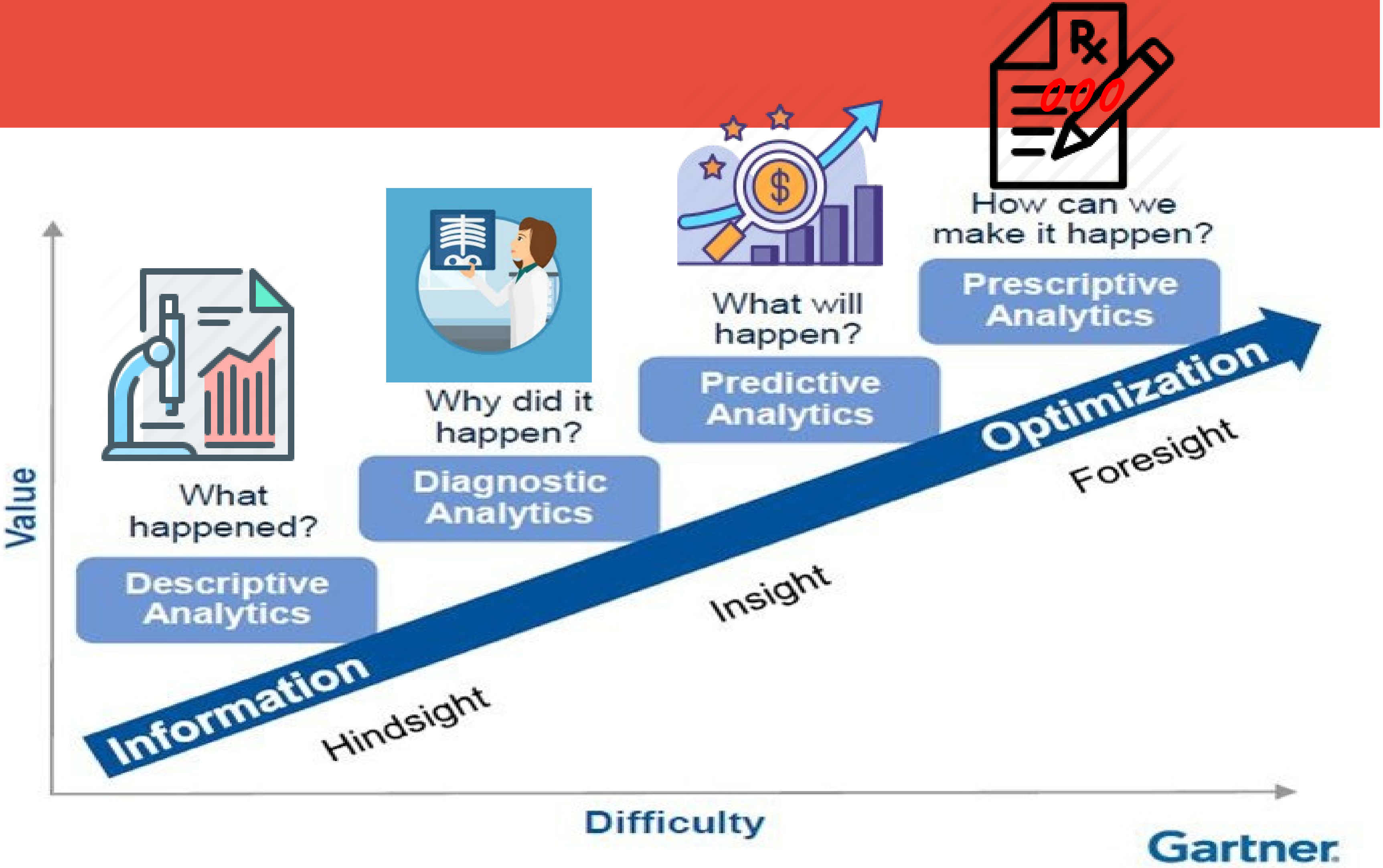
- it helps businesses to optimize their performances
- reduce costs
- To make better business decisions
- Analyze customer trends and satisfaction
- better products and services.

Data Analytics

Qualitative and quantitative techniques and processes used to business analysis to get meaningful conclusions, enhance productivity and profit gain.

There are four types of Data Analytics




000
How can we make it happen?
Prescriptive Analytics
What will happen?
Predictive Analytics
Why did it happen?
Diagnostic Analytics
What happened?
Descriptive Analytics
Value
Information
Optimization
Hindsight
Insight
Foresight
Difficulty
Gartner

Types of Data Analytics

Descriptive analytic:

tell you what happened in the past.

helps a business understand how it is performing

This can be in the form of data visualizations like graphs, charts, reports.

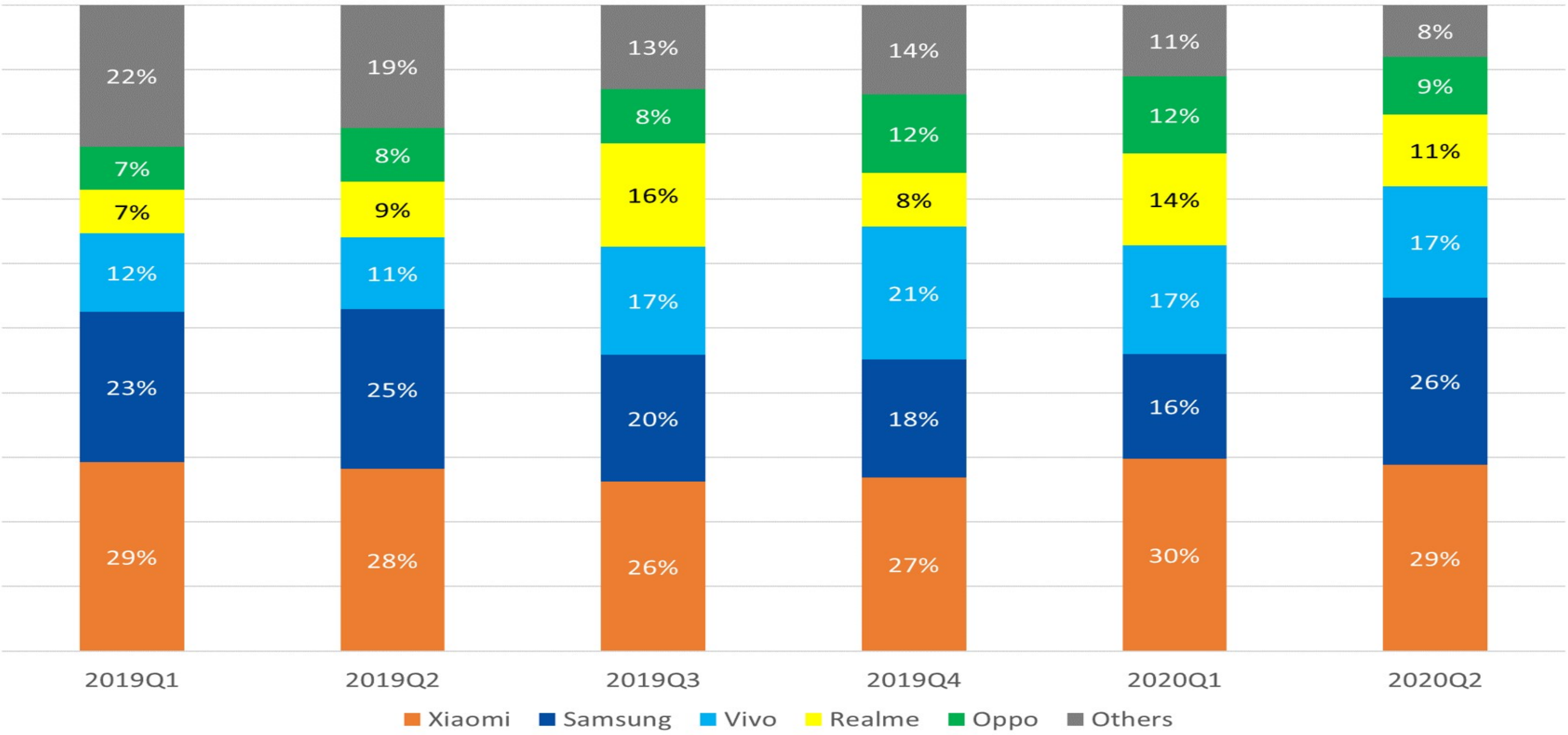
Diagnostic analytics

provides deeper analysis to answer the question: Why did this happen?

diagnostic analytics would explore the data and make correlations.

It mostly uses probabilities, likelihoods, and the distribution of outcomes for the analysis. Hypothesis testing is the major technique used for this analytics

India Smartphone Market Share (2019 Q1- 2020 Q2)



Types of Data Analytics

Predictive analytics

Takes historical data and feeds it into a machine learning model.

The model is then applied to current data to predict what will happen next.

This can use machine learning algorithms like random forests, SVM, etc. and statistics for learning and testing the data.

Prescriptive analytics

Takes predictive data to the next level. Now that you have an idea of what will likely happen in the future, what should you do? It suggests various courses of action and outlines what the potential implications would be for each.

Natural Language Processing, Sentiment Analysis, Deep learning, Stochastic modelling are some of the techniques used for prescriptive analysis.

Data Types

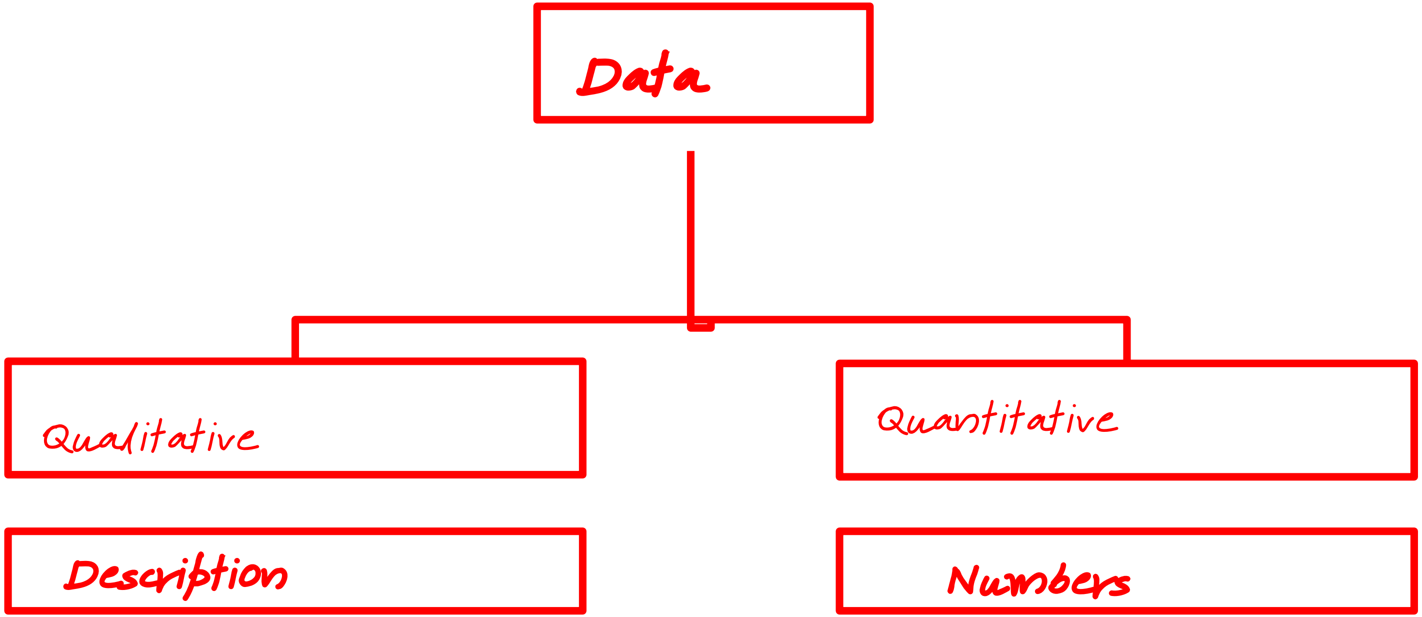
Structured Data

IBM
DB2
Microsoft SQL Server
teradata.
MySQL
X

Unstructured Data

f
in
g+
You Tube

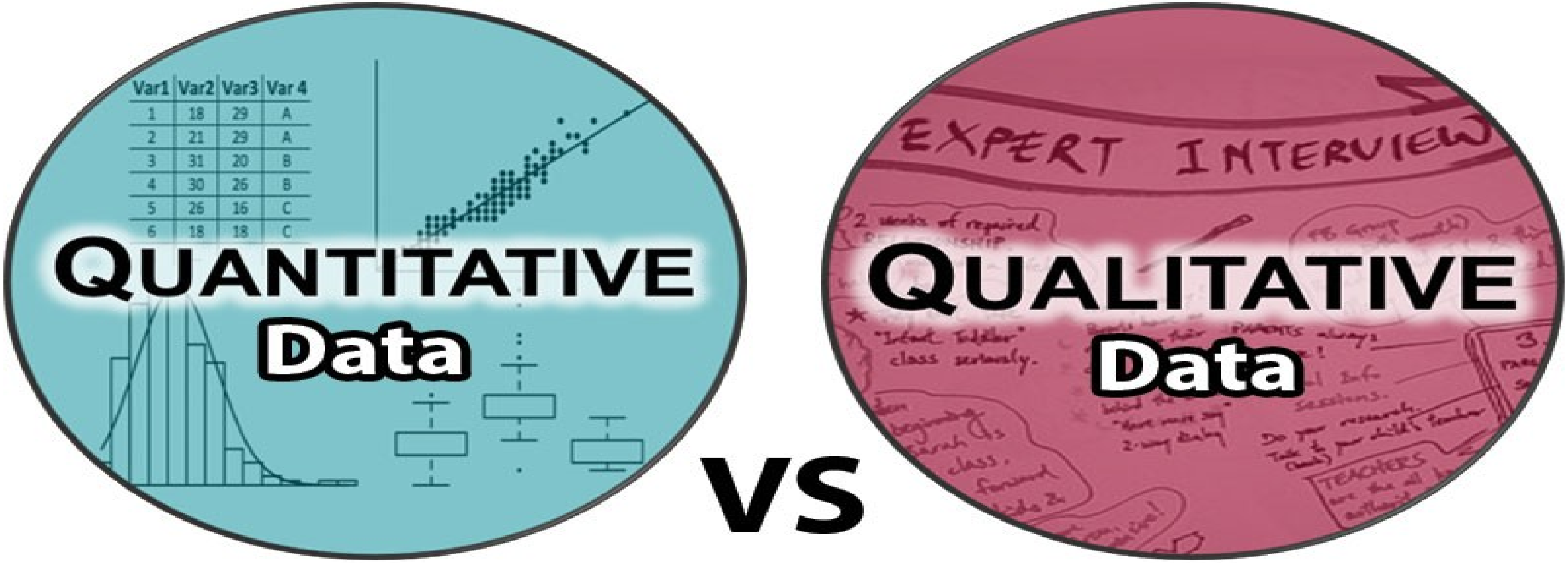
Quantitative vs Qualitative



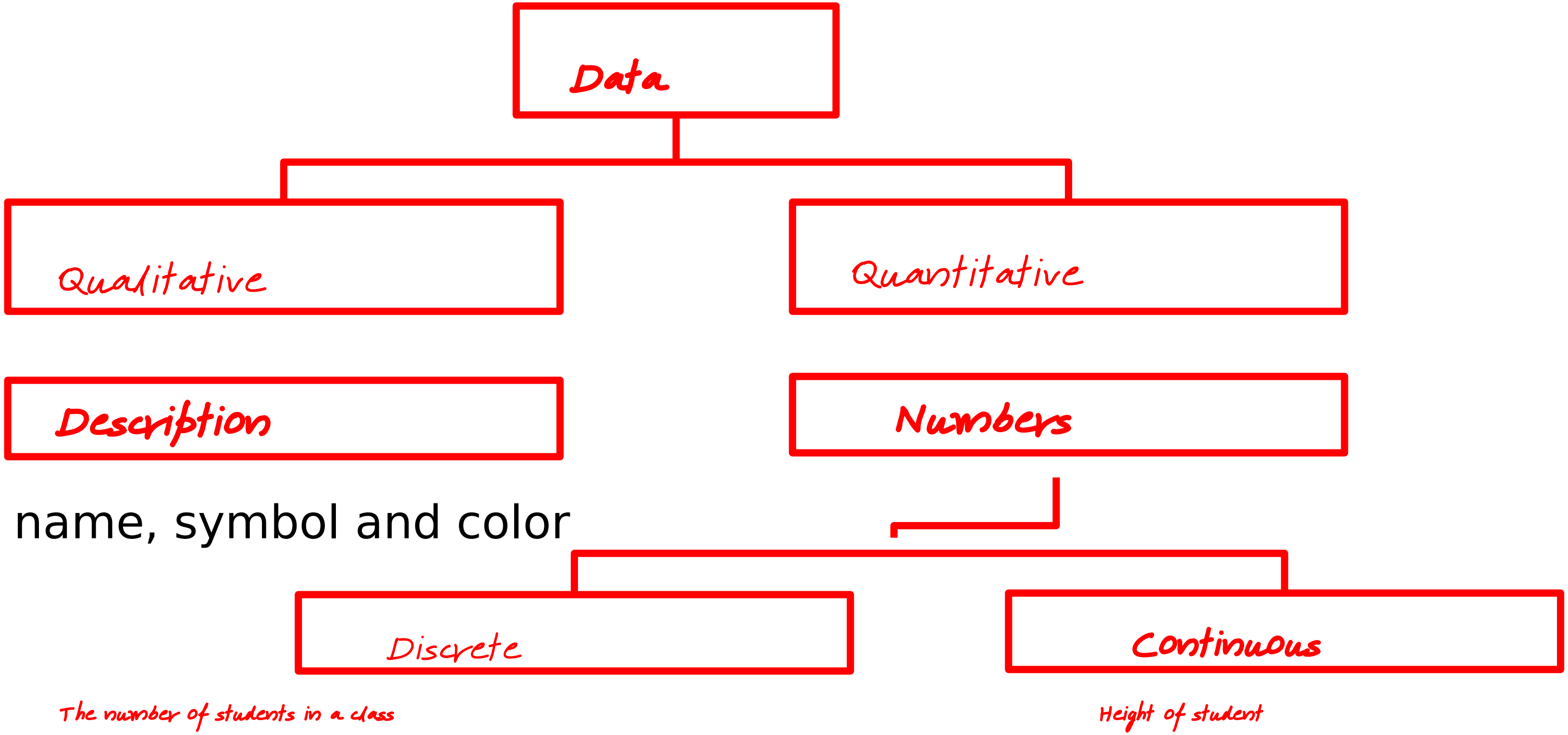
name, symbol and color

Height, weight and
temparature

Quantitative vs Qualitative



Quantitative vs Qualitative



Data

Qualitative

Quantitative

Description

Numbers

name, symbol and color

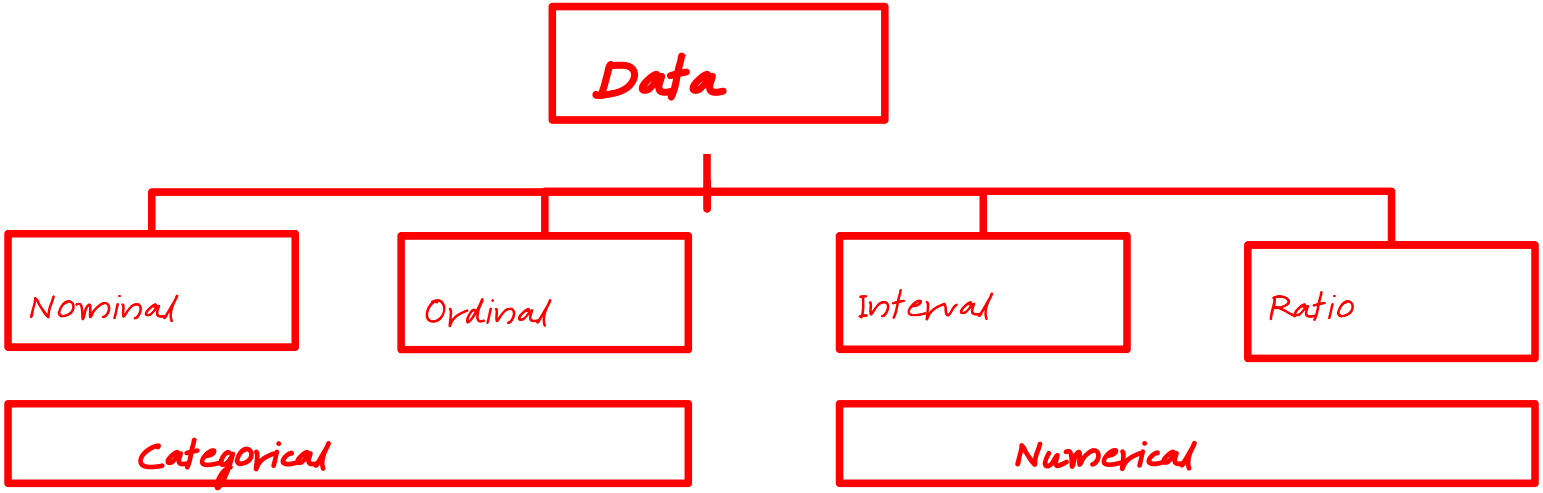
Discrete

Continuous

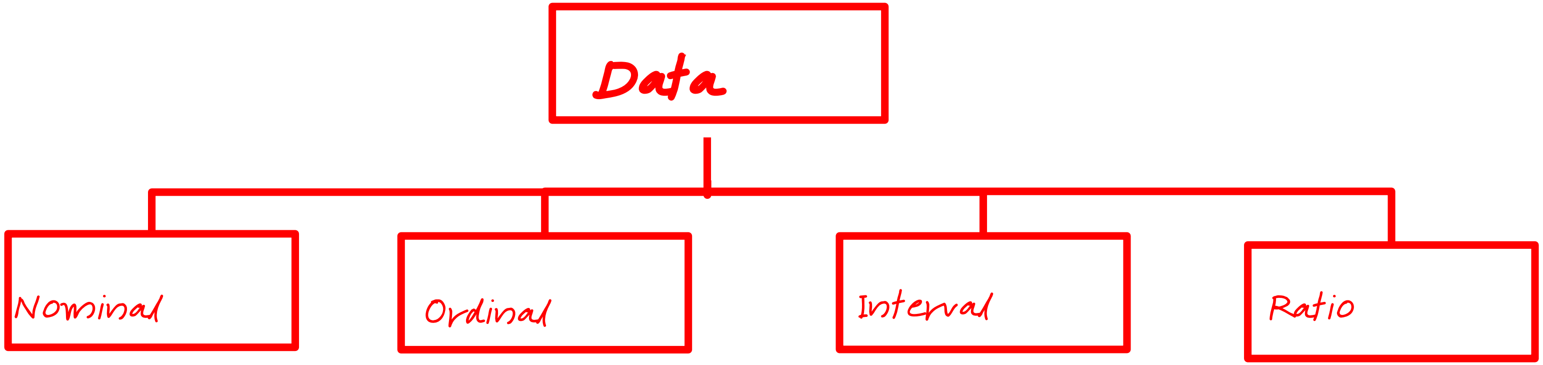
The number of students in a class

Height of student

Measurement Scales



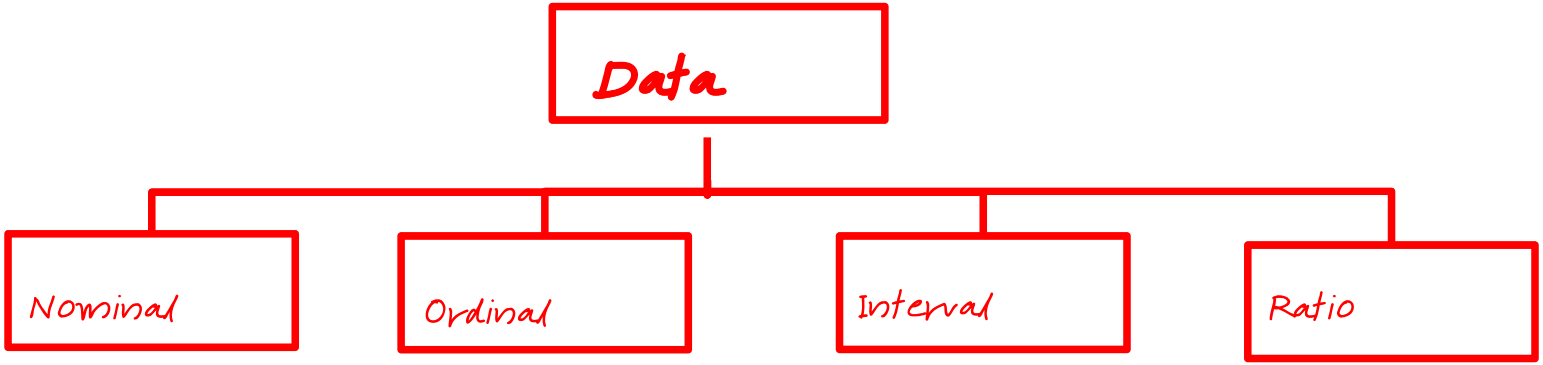
Measurement Scales



Color: Yellow, Green and Blue

Marital Status: single, married

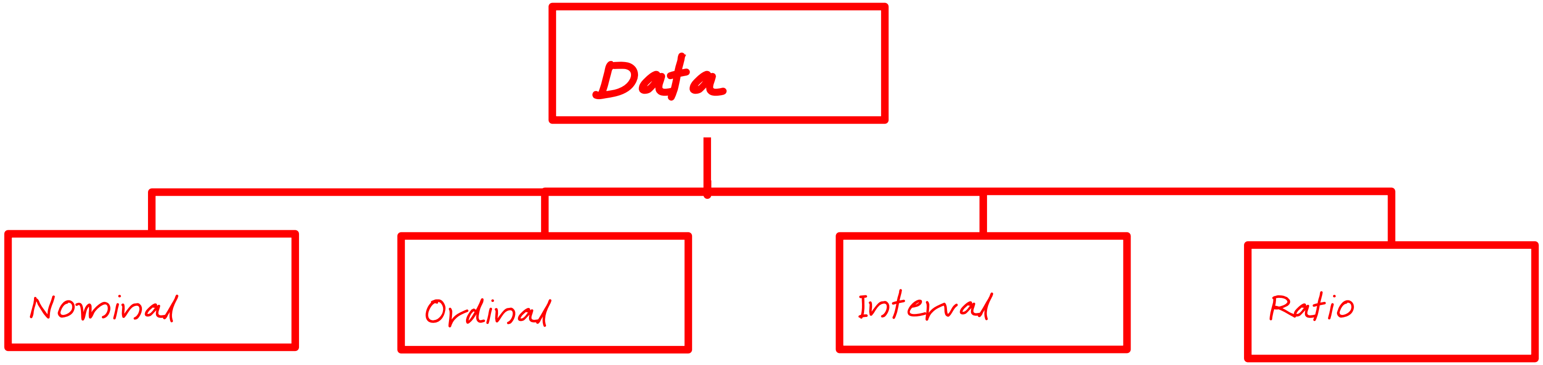
Measurement Scales



Small,Medium and Large

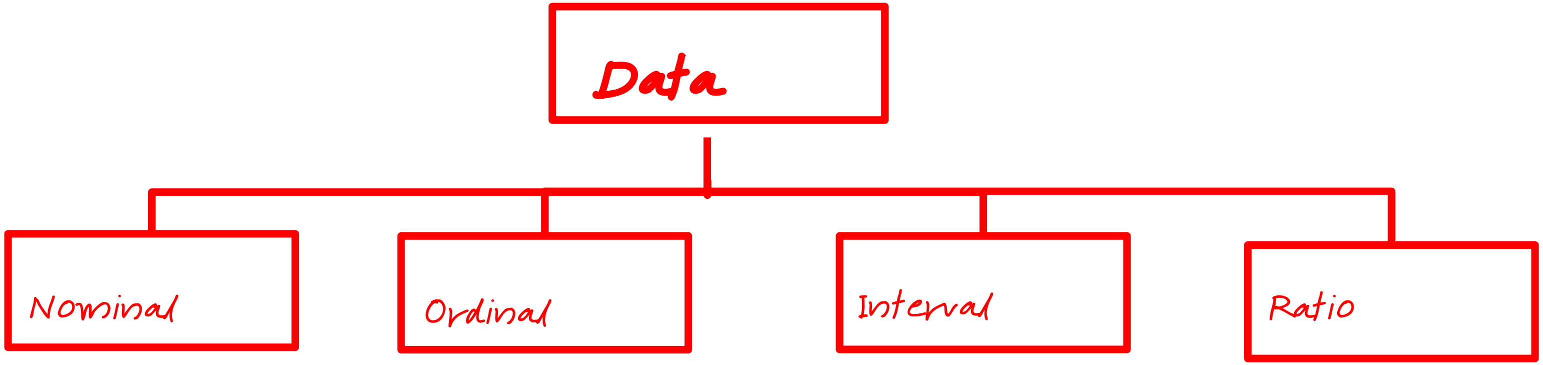
Rating on 5 point scale

Measurement Scales



Date and temparature

Measurement Scales

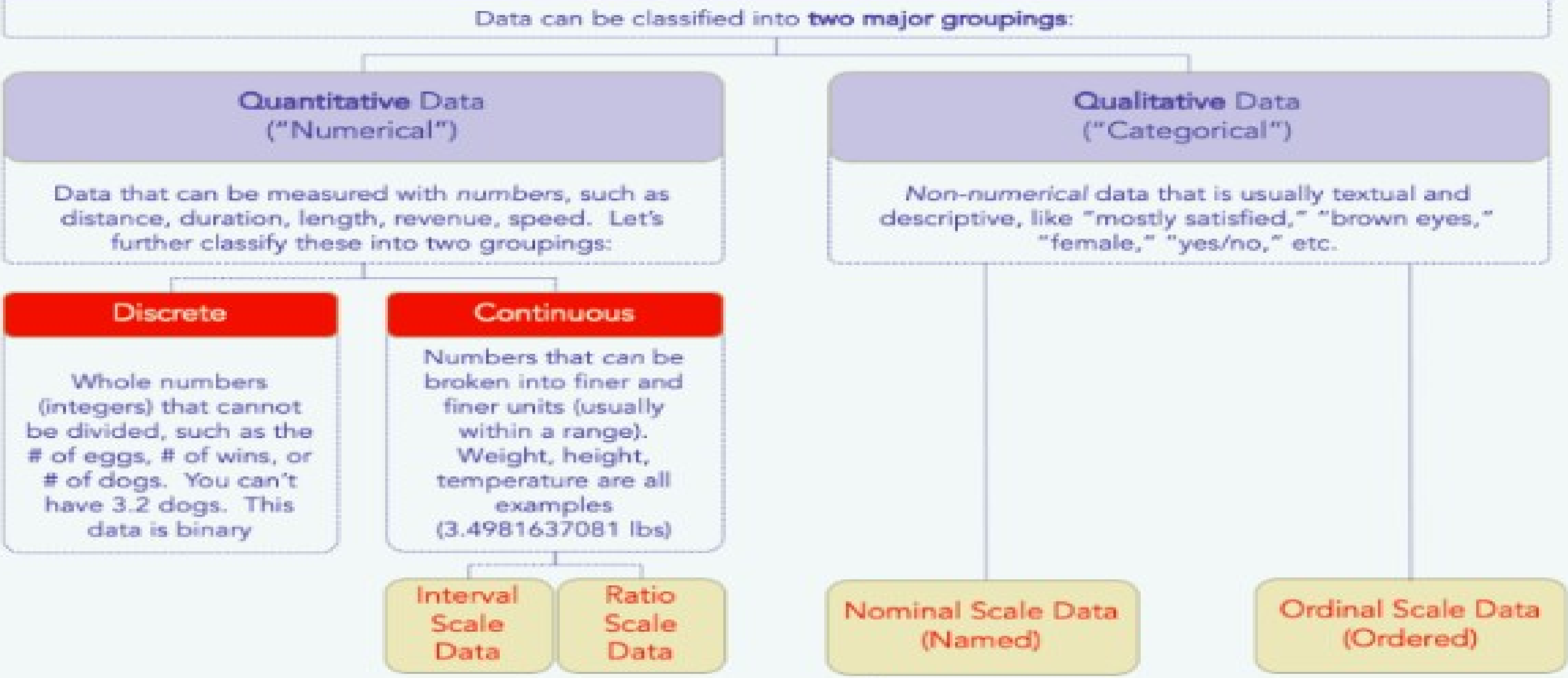


Height, weight and Volume

Measurement Scales

	Nominal	Ordinal	Interval	Ratio
Ordered	No	Yes	Yes	Yes
Difference	No	No	Yes	Yes
Absolute Zero	No	No	No	Yes

Types of Data in Statistics



Data can be classified into **two major groupings**:

Quantitative Data ("Numerical")

Data that can be measured with *numbers*, such as distance, duration, length, revenue, speed. Let's further classify these into two groupings:

Discrete

Whole numbers (integers) that cannot be divided, such as the # of eggs, # of wins, or # of dogs. You can't have 3.2 dogs. This data is binary

Continuous

Numbers that can be broken into finer and finer units (usually within a range). Weight, height, temperature are all examples (3.4981637081 lbs)

Interval Scale Data

Ratio Scale Data

Qualitative Data ("Categorical")

Non-numerical data that is usually textual and descriptive, like "mostly satisfied," "brown eyes," "female," "yes/no," etc.

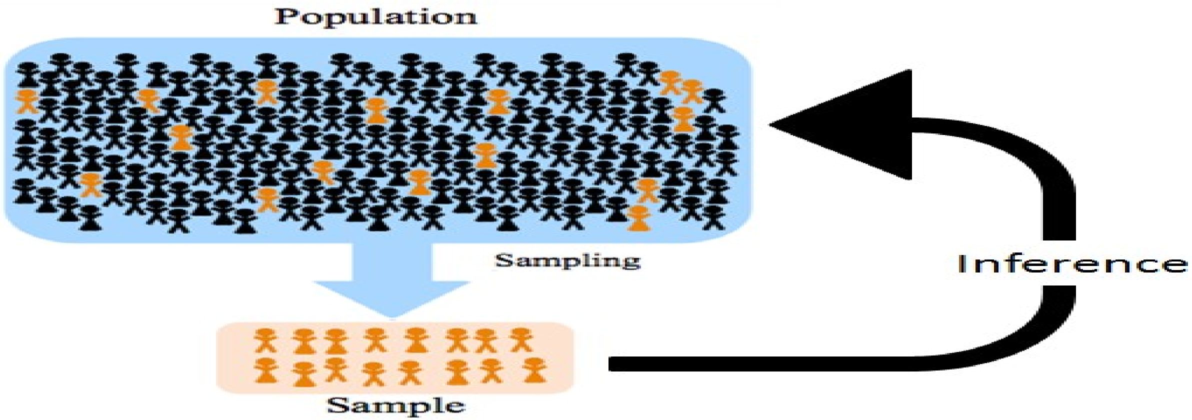
Nominal Scale Data (Named)

Ordinal Scale Data (Ordered)

Population and Sample

Population: A population is the entire group that you want to draw conclusions about.

Sample: A sample is the specific group with in the population that you will collect data from



Descriptive Analytics

Descriptive Analytics reprasent complex data in user friendly manner

Descriptive Analytics are important to extract important information

Descriptive Analytics helps to identify what had happend in the past and to understand business well

Descriptive Analytics

Descriptive Analysis can be perfomed by using following techiques

1. Data cleaning
2. Summarization of data by using Descriptive Statistics
3. Data Visualization

Descriptive Statisticss

To summarize data we will use descriptive statistics

Distribution

Central tendency

Dispersion

Distribution

The distribution is a summary of the frequency of individual values or ranges of values for a variable.

For instance, The frequency of each response to a survey question is depicted

Rank	Degree of agreement	Frequency
1	Strongly agree	20
2	Agree somewhat	30
3	Not sure	20
4	Disagree somewhat	15
5	Strongly disagree	15

Central Tendency

The central tendency of a distribution is an estimate of the “center” of a distribution of values.

There are three major types of estimates of central tendency:

- Mean
- Median
- Mode

Central Tendency

Mean: Arithematical average of the data

It is probably the most commonly used method of describing central tendency.

Assume that the data has n observations in sample

Let X_i be the value of the i^{th} observation

Mean $= \overline{x} = x_1+x_2+.....+x_n/ n =$ $$\sum_{i=1}^{n} Xi/n$$

Here $\overline{x}$ sample mean

Note: Summation of deviation of observations from mean is Zero

$$\sum_{i=1}^{n} (Xi - \overline{X}) = 0$$

Central Tendency

Mean

If the entire population is avilable

Then the average value is called population Mean(μ)

Let X_i be the value of the i^{th} observation

$$\text{Mean} = \mu = X_1 + X_2 + \ldots + X_N / N \quad = \quad \frac{\sum_{i=1}^{N} Xi}{N}$$

The Mean is a phenomenon called “Wisdom of crowd”

collective wisdom of of people is better than any individual person’s knowledge

Central Tendency

Mean

If the data is captured in frequencies then average could be

$$\overline{X} = \frac{\sum_{i=1}^{n} fi\,xi}{fi}$$

Mean is affected significantly by presence of outliers

Median is good when data is continuous numeric and distribution of data is symetric

Central Tendency

Median

The Median is the score found at the exact middle of the set of values.

One way to compute the median is to list all scores in numerical order, and then locate the score in the center of the sample.

If n is odd

$$Median = \left\{ \frac{(n+1)}{2} \right\}$$

If n is even

$$Median = \left\{ \frac{\left(\frac{n}{2} + \frac{(n+2)}{2} \right)}{2} \right\}$$

Central Tendency

Median

It is stable than Mean

Adding new observation may not change the median significantly

When observations are skewed median is very good measure of central tendency

If data is ordinal then and distribution od data is skewed median is good

Drawback is it not caliculated using entire data like in case of mean

Central Tendency

Mode

Mode is the most frequently occurring value in the set of scores.

To determine the mode, you might order the scores and then count each one. The most frequently occurring value is the mode.

For example 10,20,30,20,30,20

20 is appeared 3 times so 20 is the mode

Mode is the only measure of central tendency which is valid for qualitative(nominal) data

Central Tendency

Percentile, Decile and Quartile

These are used to identify the position of the observationin the dataset

Percentile denoted as P_x is the value of data at which x percentage of data lie below that value

To find P_x we have to arrange the data in the increasing order and the value of P_x is the position

Position corresponding to P_x = $\frac{x(n+1)}{100}$

Here n is number of observations

Central Tendency

Percentile, Decile and Quartile

Decile special value of percentile that devides the data into 10 equal parts

First decile contains first 10% of data $P_{10} = \frac{10(n+1)}{100}$

second decile contains first 20% of data $P_{20} = \frac{20(n+1)}{100}$

Central Tendency

Percentile, Decile and Quartile

Quartile devides the data into 4 equal parts

First quartile(Q_1) contains first 25% of data

second quartile(Q_2) contains first 50% of data and it is also median

Third quartile(Q_3) contains first 75% of data

$$Q_1 = P_{25} = \frac{25(n+1)}{100}$$

$$Q_2 = P_{50} = \frac{50(n+1)}{100}$$

$$Q_3 = P_{75} = \frac{75(n+1)}{100}$$

Measuring spread

Dispersion

Dispersion refers to the spread of the values around the central tendency.

Most common metrics to measure variation in the data are:

1. Range
2. Inter-Quartile Distance(IQR)
3. Standard Deviation and Variance

Dispersion or Measure of Variation

Range

Range is difference between maximum and minimum values of the data

It capture the data spread

Eg: 10,50,89,20,90

Range = 90-10 = 80

Dispersion or Measure of Variation

Inter Quartile Range(IQR)

IQR is measure of distance between Quartile1(Q_1) and Quartile3(Q_3)

It capture the data spread

Eg: The temperatures for a city over a week are 35,23,28,25,28,32,33

To caliculate Quartiles short the data

23,25,28,28,32,33,35

Dispersion or Measure of Variation

Inter Quartile Range(IQR)

To caliculate Quartiles short the data

23,25,28,28,32,33,35

Position to $Q_1 = P_{25} = \frac{25(n+1)}{100} = \frac{25(7+1)}{100} = \frac{8}{4} = 2$

Position to $Q_3 = P_{75} = \frac{75(n+1)}{100} = \frac{75(7+1)}{100} = \frac{3(8)}{4} = 6$

Value at 2nd position(Q_1) = 25

Value at 4th position(Q_3) = 33

IQR = Q_3- Q_1 = 33-25 = 8

Dispersion or Measure of Variation

Inter Quartile Range(IQR)

If data is numerical and skewed then IQR is the good measure of variation

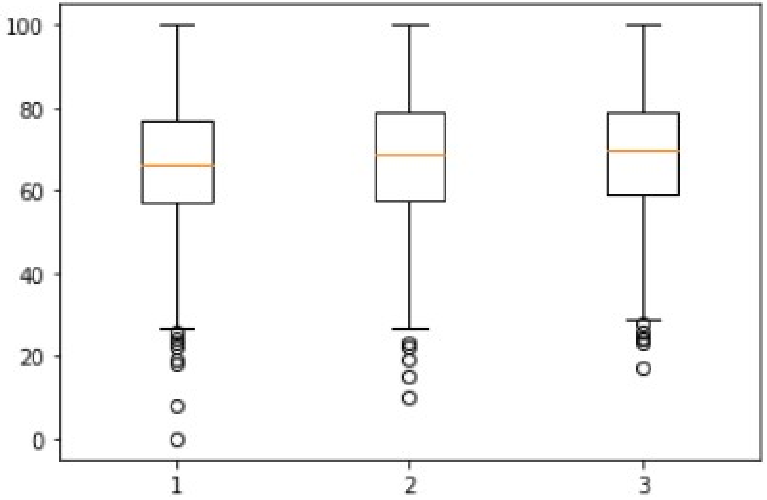
we can use IQR to identify values that are much farther from the center than usual.

IQR tell us how far a typical value could be from the average, so anything much more than the typical distance can be identified

Dispersion or Measure of Variation

Lower bound for Olutliers = $Q_1 - 1.5(IQR)$

Upper bound for Olutliers = $Q_3 + 1.5(IQR)$



Dispersion or Measure of Variation

Variance

Measure of variability in the data from the mean value

Variance of population $$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

Variance of sample $$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}$$

Dispersion or Measure of Variation

Standard Deviation

Population Standard Deviation $$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}$$

Sample Standard Deviation $$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}}$$

If data is numeric and have the symetric distribution then Standard deviation or variance is the good measure of variation

Visualization

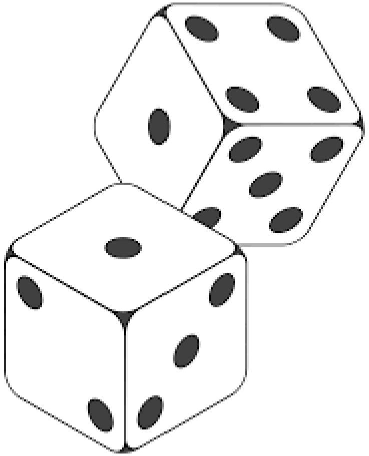
Data visualization is the graphical representation of information and data by using charts, graphs, and maps

Data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

Random Experiment

It is an experiment whose outcome can’t be predicted with certainity before the experiment is run

Examples: Tossing coin, rolling dice



Random Variable

A function which maps real number for each sample point in the sample space S in random experiment

Example: Tossing 3 coins

S = {HHH, HHT,HTH,HTT,THH,THT,TTH,TTT}

how many heads?

X = The number of Tails is the Random Variable

X = f(#T) = {0,1,1,1,2,2,2,3}

Random Variable Types

1. Discreate Random Variable

Takes only finite no of values in finite observation interval

Example: rolling dice

2. Continuous Random Variable

Takes only infinite no of values in finite observation interval

Example: time taken to fill water tank

Probability distribution

Mathematical function that gives the probabilities of each possible value that a random variable can take is called its probability distribution.

Example: if random variable X is used to denote the outcome of a coin toss

then the probability distribution of X

would take the value 0.5 for X = heads,

and 0.5 for X = tails (assuming that the coin is fair)

Probability mass function

Probability mass function is the probability distribution of a discrete random variable, and provides the possible values and their associated probabilities.

The probabilities associated with each possible values must be positive and sum up to 1.

If X is random variable the function $f(x) = P(X = x)$

$$1.\, f(x) \geq 0 \; for\, all \; x \in X$$

$$2.\, \sum f(x) = 1$$

Probability mass function

If we flip coin two times

X = no of heads

X ={ (t,t),(h,t),(t,h),(h,h)} = {0,1,1,2}

x	0	1	2
f(x)	0.25	0.5	0.25

$$f(1)=P(X=1)=P(t,h)+P(h,t)=\frac{1}{4}+\frac{1}{4}=0.5$$

$$f(2)=P(X=2)=P(h,h)=\frac{1}{4}=0.25$$

$$1.\ \frac{1}{4},\frac{2}{4},\frac{1}{4}\geq 0$$

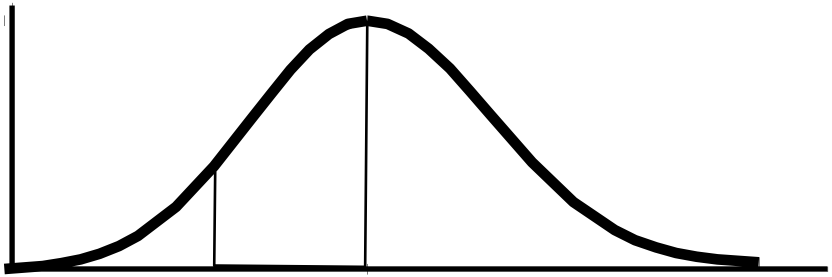
$$2.\ \frac{1}{4}+\frac{2}{4}+\frac{1}{4}=1$$

Probability Density function

Probability density function is the probability distribution of a continuous random variable

If X is continuous random variable the function f(x) is called Probability density function that defined as the probability that X lies between a and b is the area under the curve between any real constants a and b such that a<= b

$$f(x) = P(a \le X \le b) = \int_a^b f(x)\,dx$$



Such that

$$1.\ f(x) \ge 0$$

$$2.\ \int_{-\infty}^{\infty} f(x)\,dx = 1$$

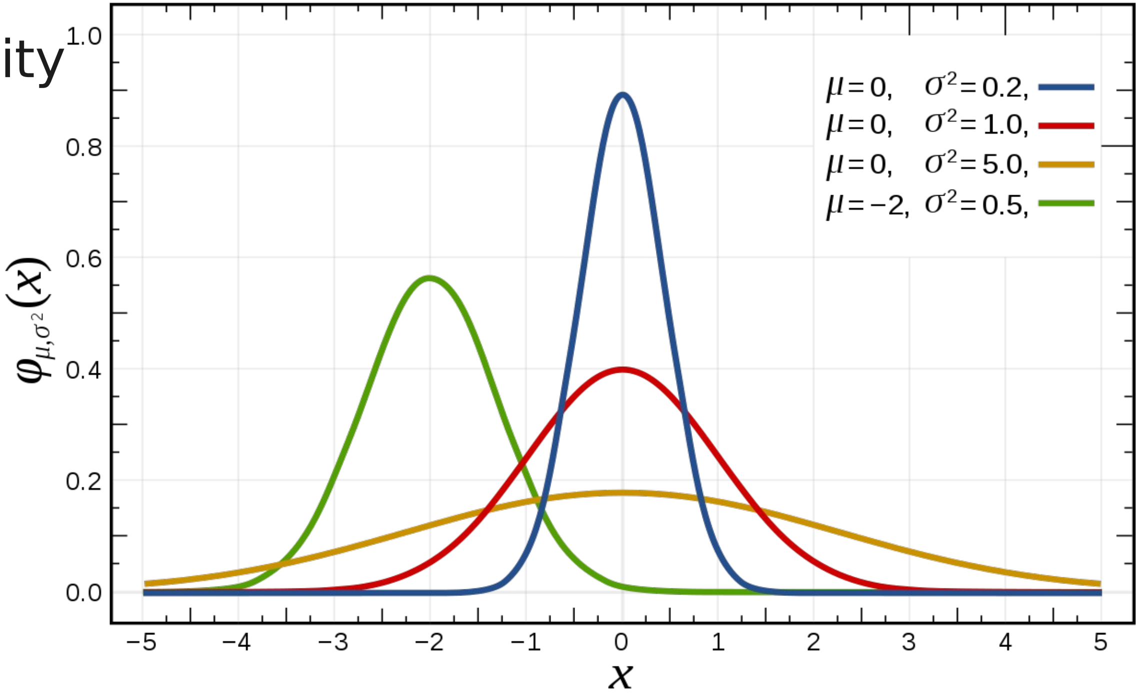
total area under the curve is equal to 1

Probability Density function

It is common for probability density functions (and probability mass functions) to be parametrized

For example, the normal distribution is parametrized in terms of the mean and the variance, denoted by μ and σ^2 respectively

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



Cumilative distribution function

PMF cant be defined for continuous Random variable

PDF cant be defined for discreate Random variable

CDF of a random variable is another method to describe the distribution of a random variable Can be defined for

- Discreate
- Continuous
- mixed

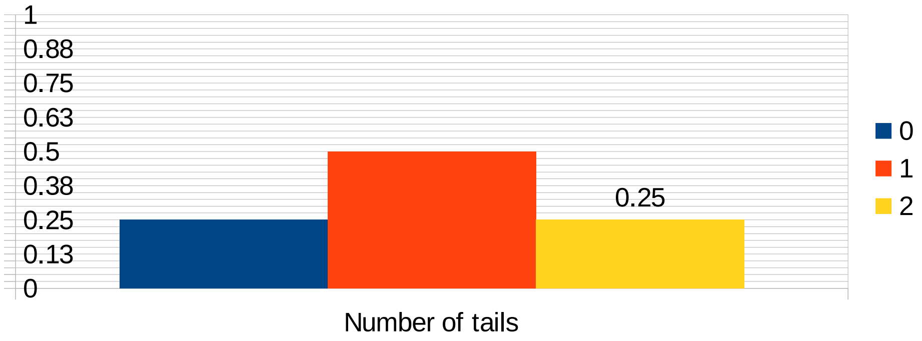
Cumilative distribution function

The probability that the random variable X takes a value “less than or equal to x”

$$CDF = F_X(x) = P(X \le x)$$

$$f(a) = P(X \le a) = \sum_{X \le a} f(X)$$

$$f(1) = P(X \le 1) = \sum_{X \le 1} f(X)$$



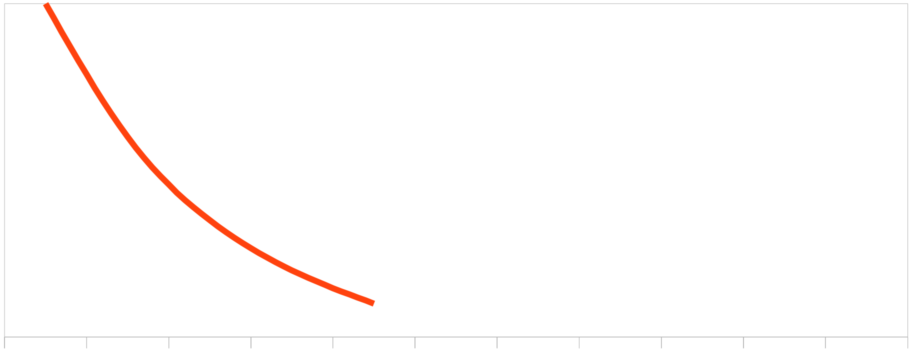
Cumilative distribution function

For continuous

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

$$f(3) = P(X \leq 3) = \int_{-\infty}^{3} 4e^{-4x}\,dx$$

$$\int_{0}^{3} 4e^{-4x}\,dx$$



Popular Discrete Probability Distributions

1. Binomial
2. Poisson

Binomial Probability Distribution

The binomial is a type of distribution that has two possible outcomes

For example, a coin toss has only two possible outcomes: heads or tails and taking a test could have two possible outcomes: pass or fail.

Binomial distributions must also meet the following three criteria:

1. The number of observations or trials is fixed.
2. Each observation or trial is independent.
3. The probability of success is exactly the same from one trial to another.

Binomial Probability Distribution Function

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

- P(X) is probability of X sucesses given n and p
- X is total numbre of successes.
- p probability of a success on an individual trial.
- n is number of trails.

Binomial Distribution Example

Example: number of heads in two tosses of coin

$P(0)= 2!/(2-0)!0! \; . \; (0.5)^{0}.(0.5)^{2}$

$P(1)= 2!/(2-1)!1! \; . \; (0.5)^{1}.(0.5)^{1}$

$P(2)= 2!/(2-2)!2! \; . \; (0.5)^{2}.(0.5)^{0}$

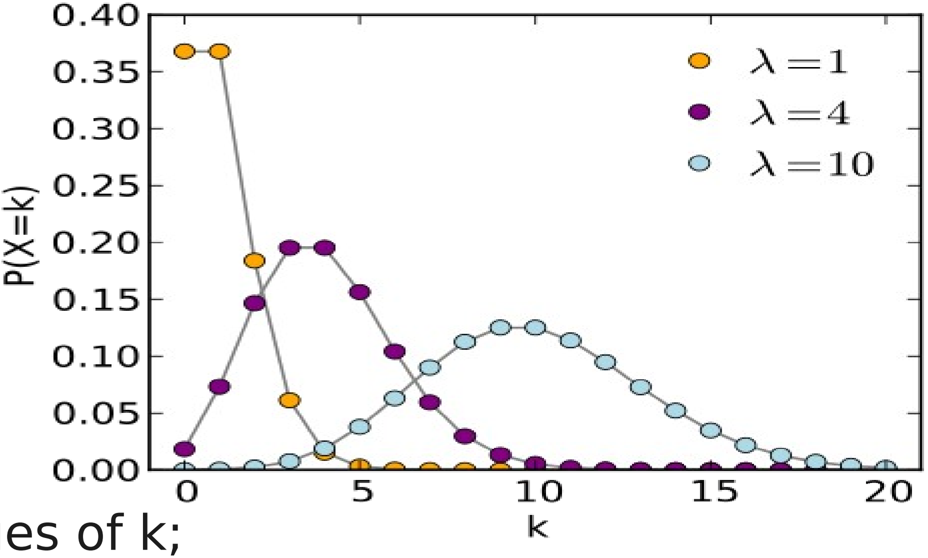
x	0	1	2
P(X)	0.25	0.5	0.25

Poisson Probability Distribution

Predicts the probability of a given number of events occurring in a fixed interval of time.

if these events occur with a known constant mean rate and independently of the time since the last event.

- The horizontal axis is the index k, the number of occurrences.
- λ is the expected rate of occurrences.
- The vertical axis is the probability of k occurrences given λ.
- The function is valid only for integer values of k;



Poisson Probability Distribution Function

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

where

- e is Euler's number (e = 2.71828...)
- k is the number of occurrences
- The positive real number λ is equal to the expected value of X

Poisson Distribution Assumptions

k is the number of times an event occurs in an interval and k can take values 0, 1, 2, ….

The occurrence of one event does not affect the probability that a second event will occur. That is, events occur independently.

The average rate at which events occur is independent of any occurrences.

Two events cannot occur at exactly the same instant; instead, at each very small sub-interval exactly one event either occurs or does not occur.

If these conditions are true, then k is a Poisson random variable, and the distribution of k is a Poisson distribution.

Poisson Distribution Example

Example: On a particular river, overflow floods occur once every 100 years on average.

Calculate the probability of k = 0, 1, 2 overflow floods in a 100-year interval, assuming the Poisson model is appropriate.

the average event rate is one overflow flood per 100 years, so λ = 1

$$P(k \text{ overflow floods in 100 years}) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{1^k e^{-1}}{k!}$$

$$P(k = 0 \text{ overflow floods in 100 years}) = \frac{1^0 e^{-1}}{0!} = \frac{e^{-1}}{1} \approx 0.368$$

$$P(k = 1 \text{ overflow flood in 100 years}) = \frac{1^1 e^{-1}}{1!} = \frac{e^{-1}}{1} \approx 0.368$$

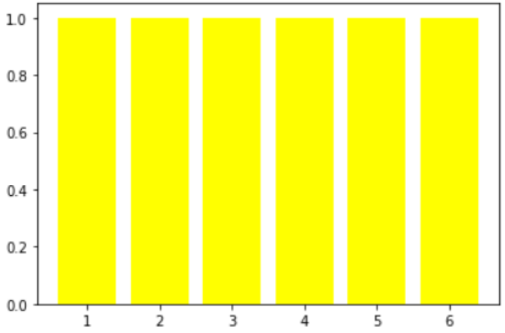
$$P(k = 2 \text{ overflow floods in 100 years}) = \frac{1^2 e^{-1}}{2!} = \frac{e^{-1}}{2} \approx 0.184$$

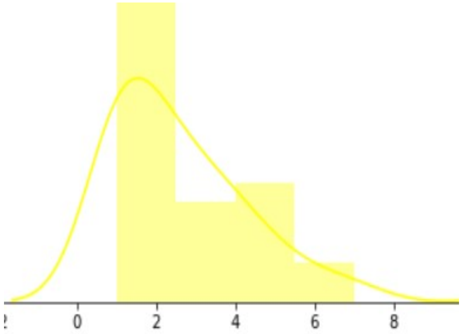
Popular continuous Probability Distributions

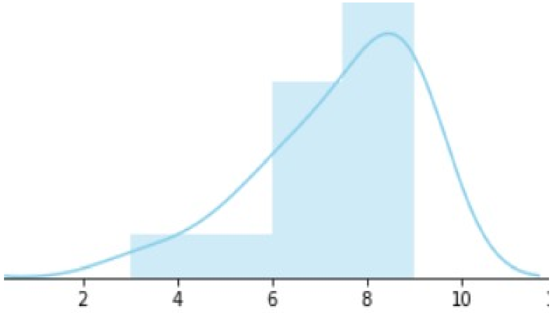
1. Normal Disribution
2. Students t Distribution
3. Chi square Distribution
4. F Distribution

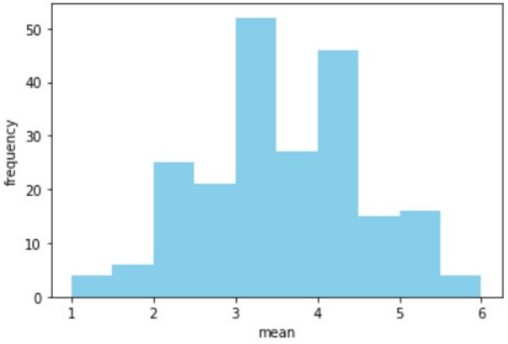
Normal Distribution

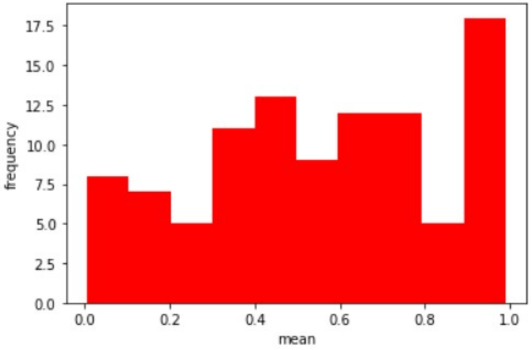
Data can be distributed in different ways











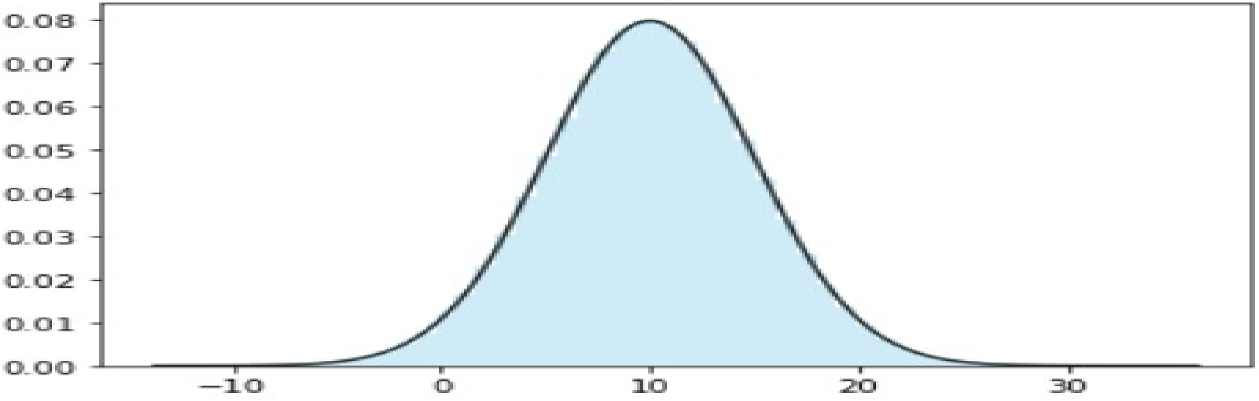
Normal Distribution

Data distribution centered around mean and forms bell shaped curve it is called normal distribution

Most of the real world data follows normal distribution

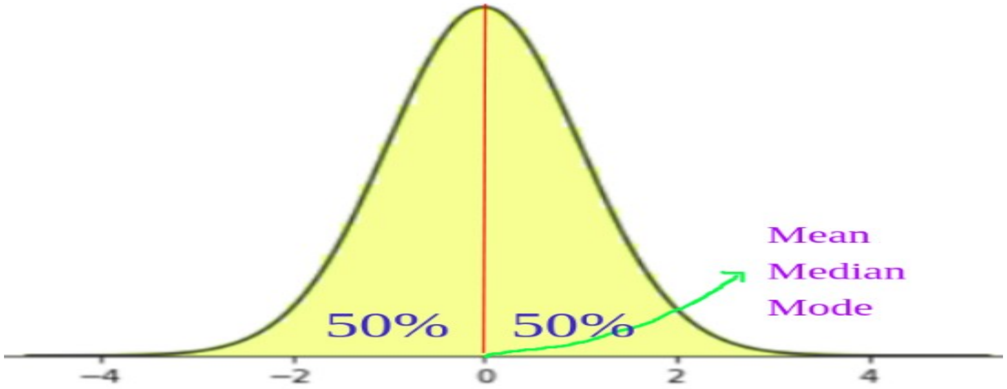
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Students Marks in test
- Heights of people



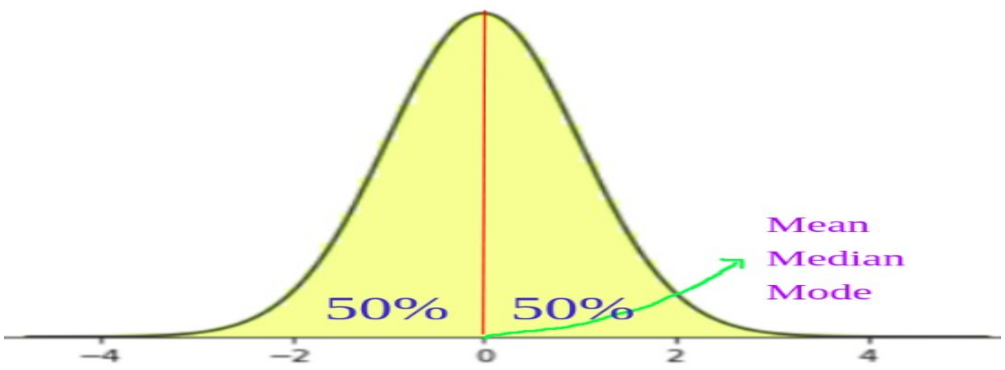
Why Normal Distribution

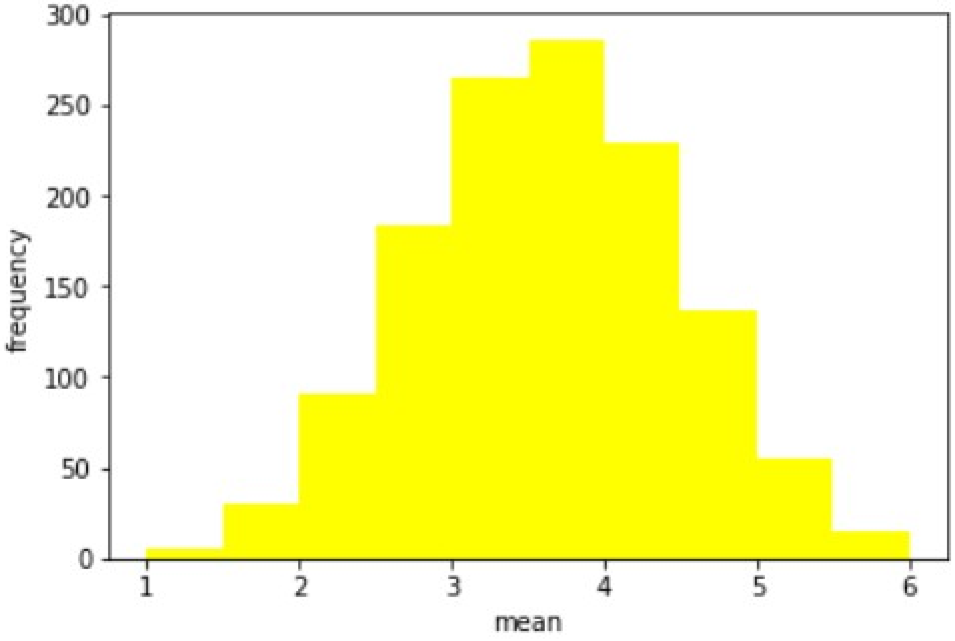
- It is symetric around center
- Mean = Median = Mode
- Allows Parametric statistics to analyze data
- Parametric statistics have lot of tools and it is well developed system



Practical Insight

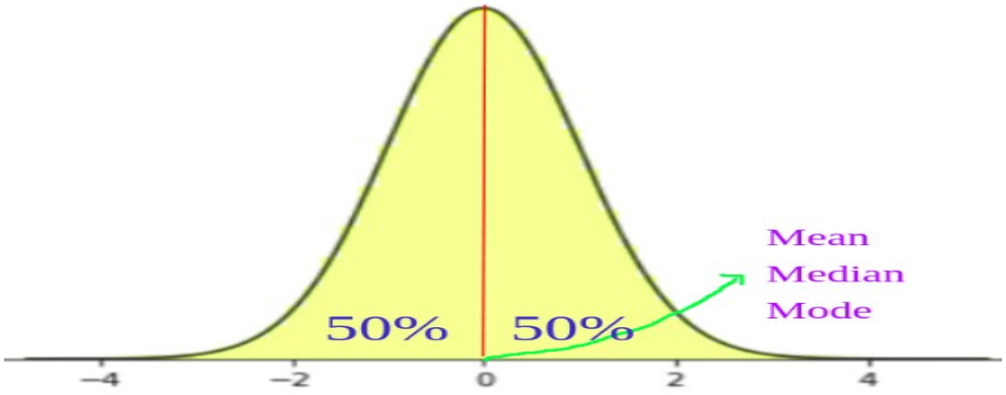
- Inferential Statistics allow some fluctuation in the data

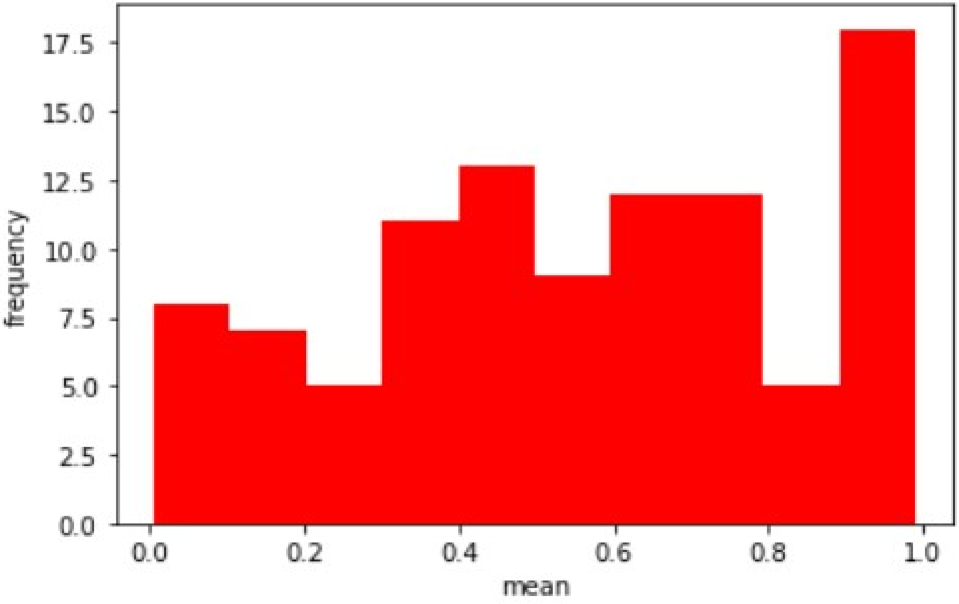




Assessing Normality Voilation

- If the voilation from normal distribution is sufficiently large
- Statistical tests can’t be applied on that data





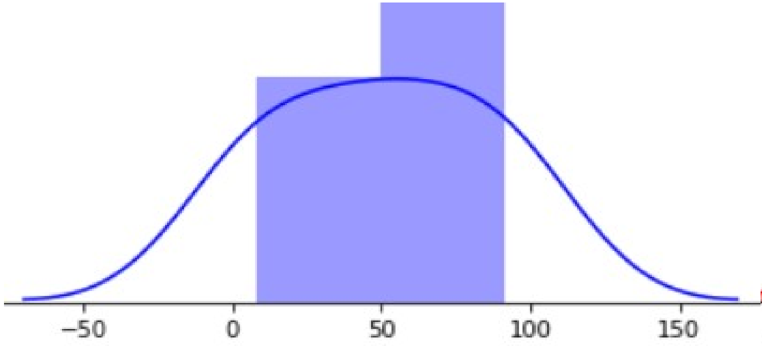
Assessing Normality Voilation

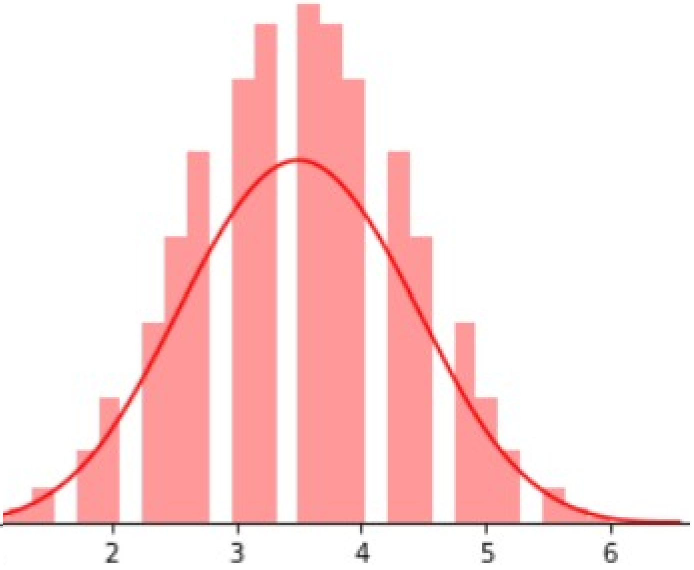
1. Shape of distribution
2. Sample size

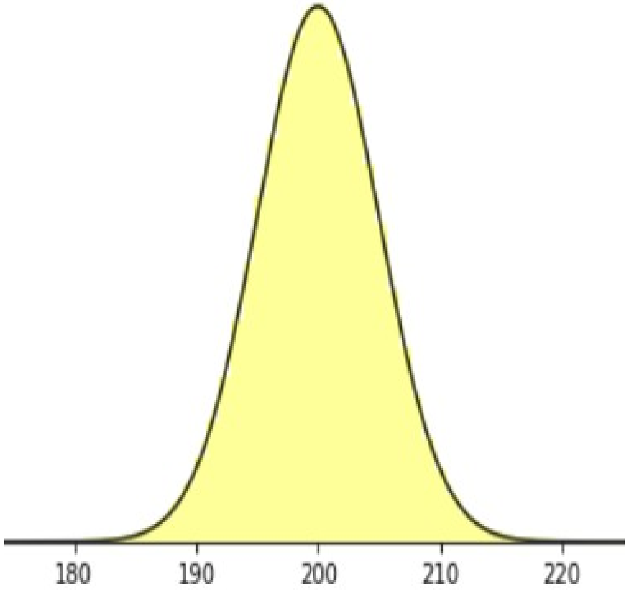
Impact due to Shape of Distribution

1. **Kurtosis:** Kurtosis measures the peakedness(flatness) of the distribution compared   with normal distribution

2. Skewness







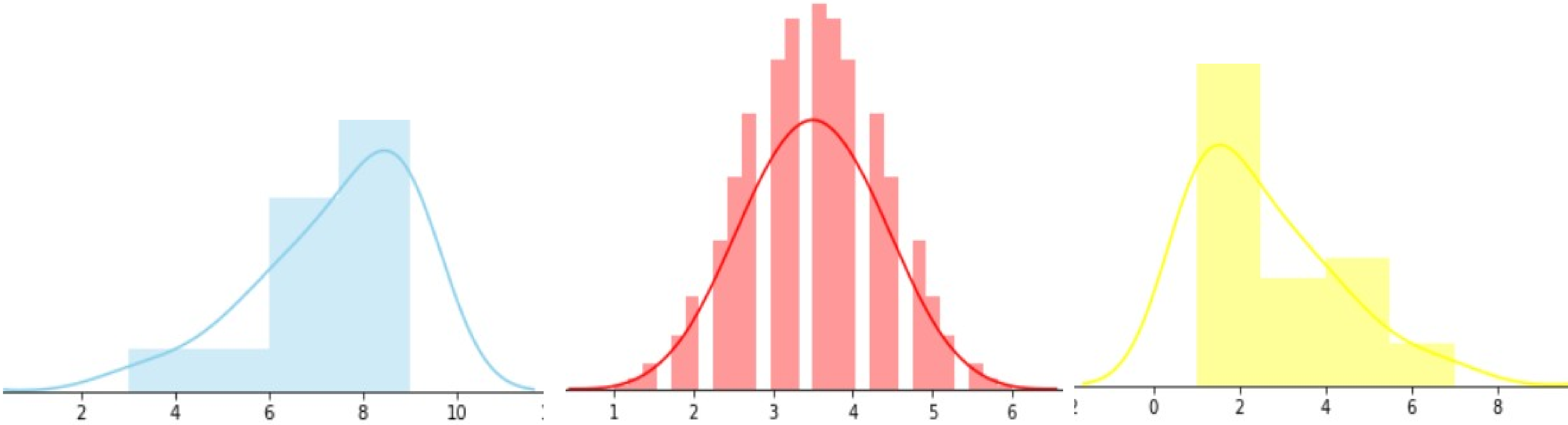
Evaluating Kurtosis

$$Kurtosis \leq 3\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$Kurtosis < 3\left(\frac{\sigma}{\sqrt{n}}\right)$$

Skewness

Measure of symetry(balence of the distribution)



Evaluating Skewness

$$Skewness \leq 3\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$Skewness < 3\left(\frac{\sigma}{\sqrt{n}}\right)$$

Impact due to Sample Size

Sample size <30

Sample size > 30

$$SampleSize \propto StatisticalPower \propto \frac{1}{Error}$$

Sampling

Essence of Sampling:

Subset of population to make inference about population parameters such as mean, praportion and standard deviation etc..

Sampling is nesessary when it is difficult or expensive to collect data on the entire population

Incorect sample may lead to wrong inference about population

Benifits of Sampling

1. Reduced Cost
2. Speed

Example

In 1973 USA presidential elections the Litaray Digest conducted Opinion polls

	Frankline Roosevelt	Alfred Landon	Others
Literary Digest	41%	55%	4%
Actual Votes Polled	61%	37%	2%

Sample Size was 2.4 million

Reson for huge error is sampling method

Example

Two major issues

1. Voter names taken from telephone directory

2. They conducted for 10 million voters only 2.4 million responded

Another servey reveled that 67% roosevelt supporters claimed they did not get call by the Literary Digest.

George Gallup predicted correct result with just 50000 samples

Note: larger sample will not improve prediction if sampling process is incorrect

Steps in sampling process

1. Identification of target population
2. Decide tha sampling frame
3. Determine sample size
4. Sampling Method
    a. Probablistic Sampling
    b. Non Probalistic Sampling

Probablistic sampling

The individual observations in sample are selected based on probability distribution

1. Random Sampling
2. Stratified Sampling
3. Clustered Sampling
4. Bootstrap Sampling

Non Probablistic sampling

The individual observations in sample does not follow probability distribution

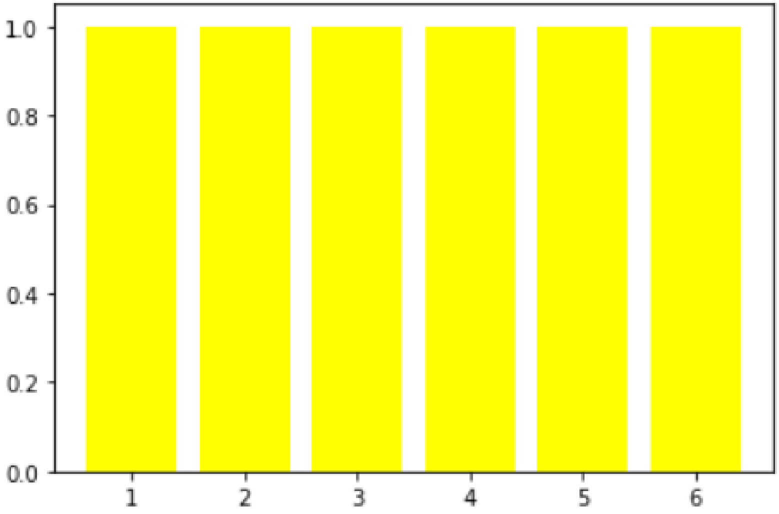
1. Convenience Sampling

2. Voluntary Sampling

Sampling distribution

The probabilty distribution of a statistic such as sample mean, standard deviation computed from several random samples

A fair die is rolled



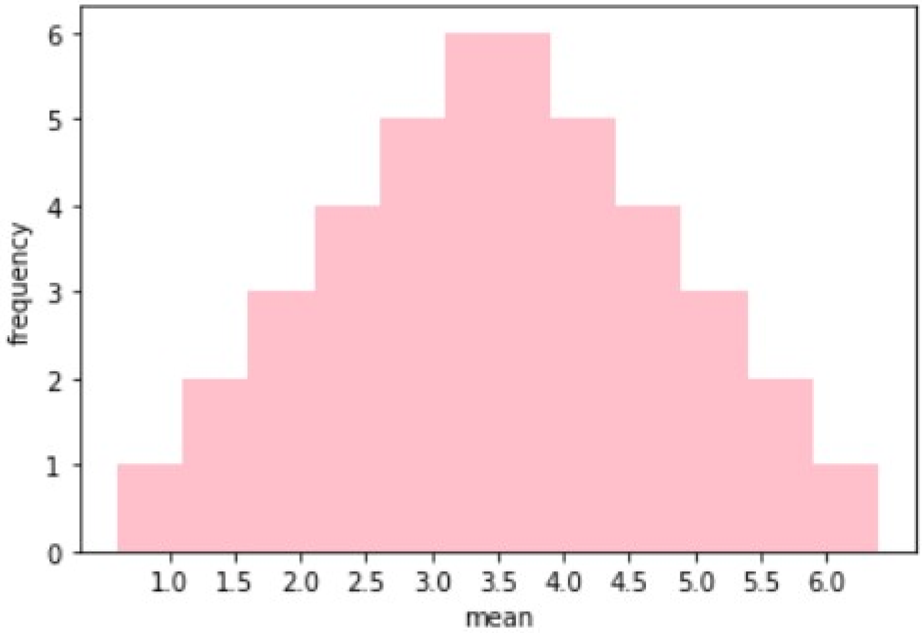
It follows uniform distribution

Mean μ = 3.5

Standard Deviation σ = 1.7

Sampling distribution

if you roll two dice

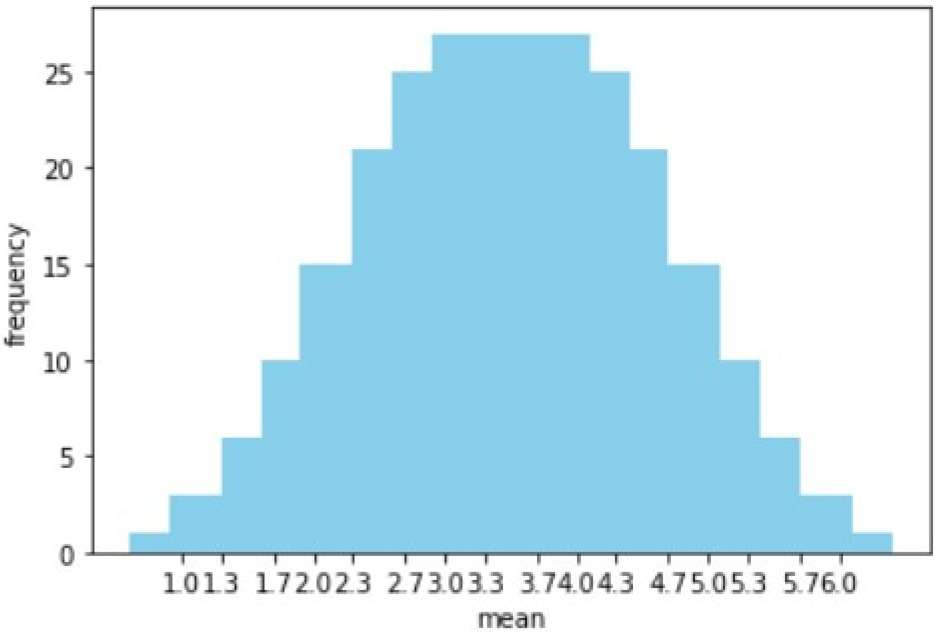


Mean $\overline{X}$ = 3.5

Standard Deviation   = 1.20

Sampling distribution

if you roll three dice

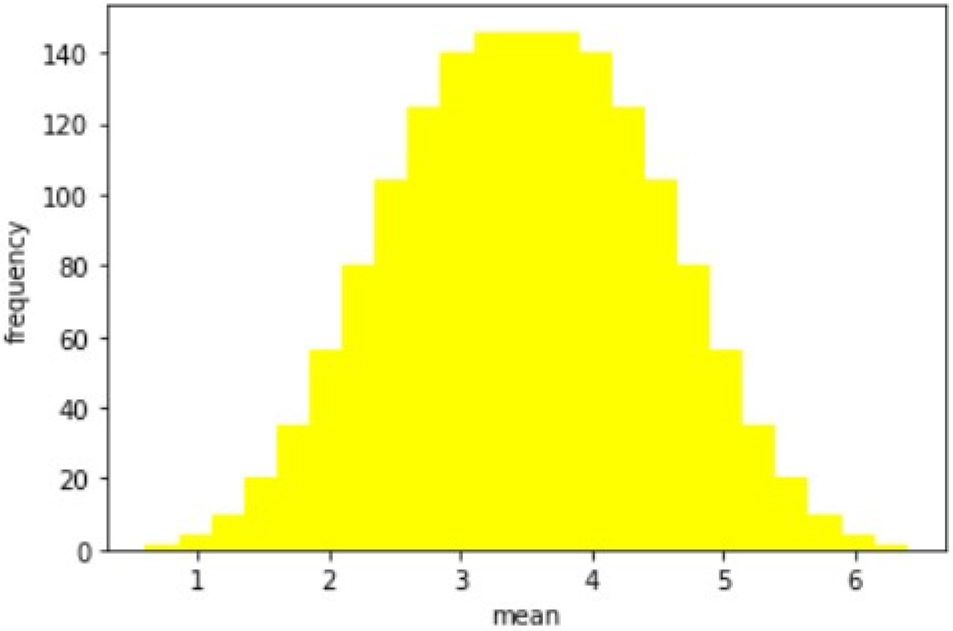


Mean $\overline{X} = 3.5$

Standard Deviation = 0.99

Sampling distribution

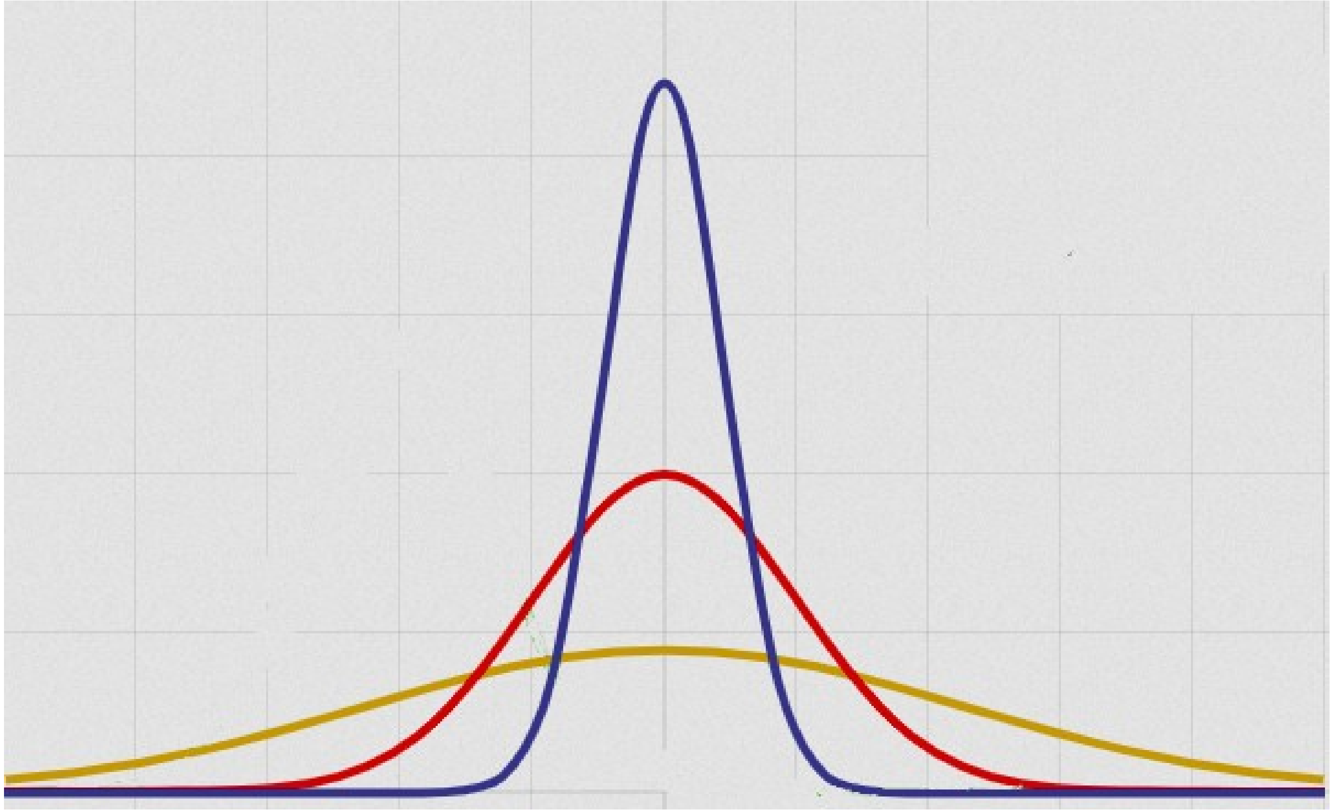
if you roll four dice



Mean $\overline{X}$ = 3.5

Standard Deviation = 0.85

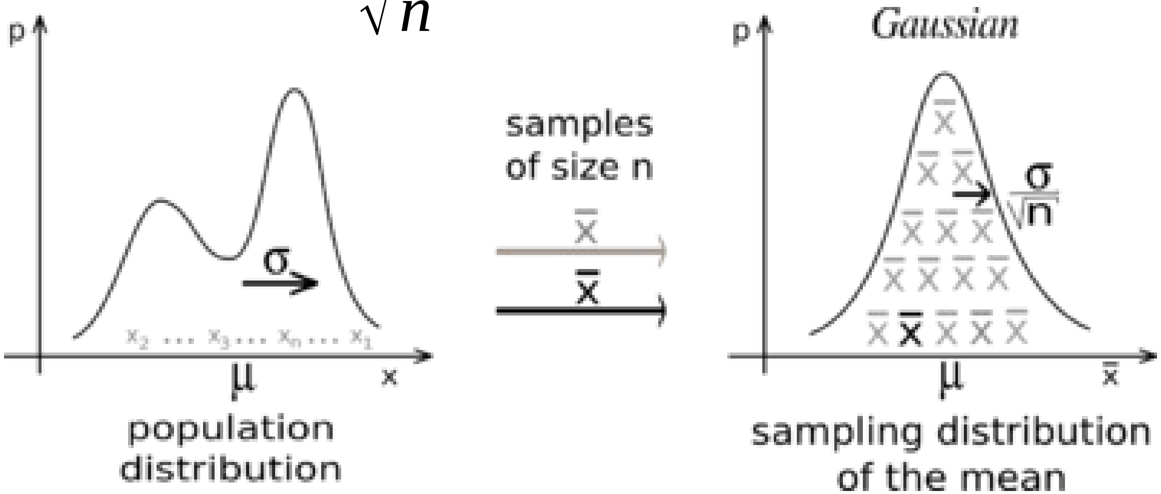
Sampling distribution



Central Limit theorem

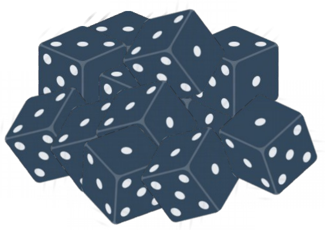
For large sample drawn from population with mean μ and standard deviation σ

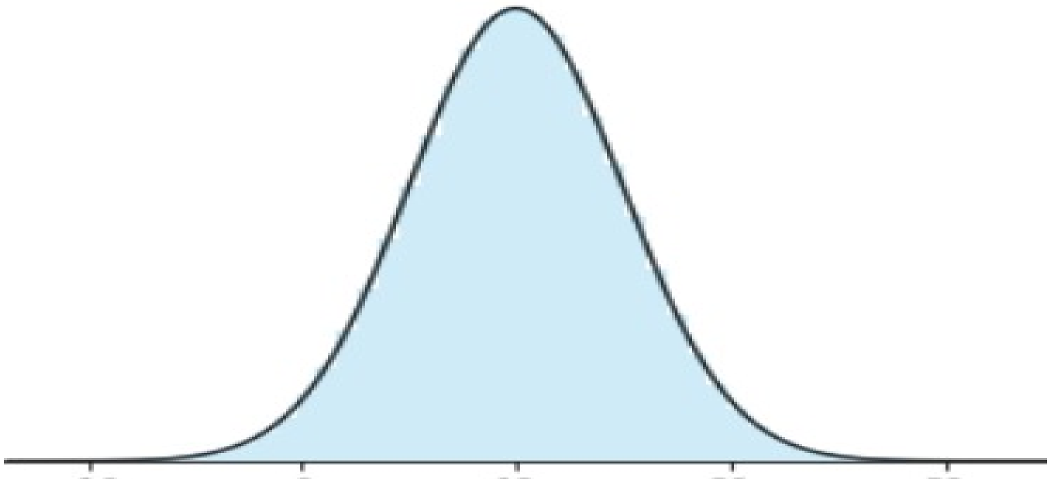
The sampling distribution of mean follows an approximate normal distribution with mean $\overline{X}$ is close to the normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ irrespective of distribution of the population


p
σ
x_2 ... x_3 ... x_n ... x_1
μ
x
population distribution
samples of size n
$\bar{x}$
$\bar{x}$
p
Gaussian
$\frac{\sigma}{\sqrt{n}}$
$\bar{x}$
μ
$\bar{x}$
sampling distribution of the mean

Sampling distribution

If you increase sample size(#dice), the distribution of possible averages (the sampling distribution) looks like a bell curve (a normal distribution)

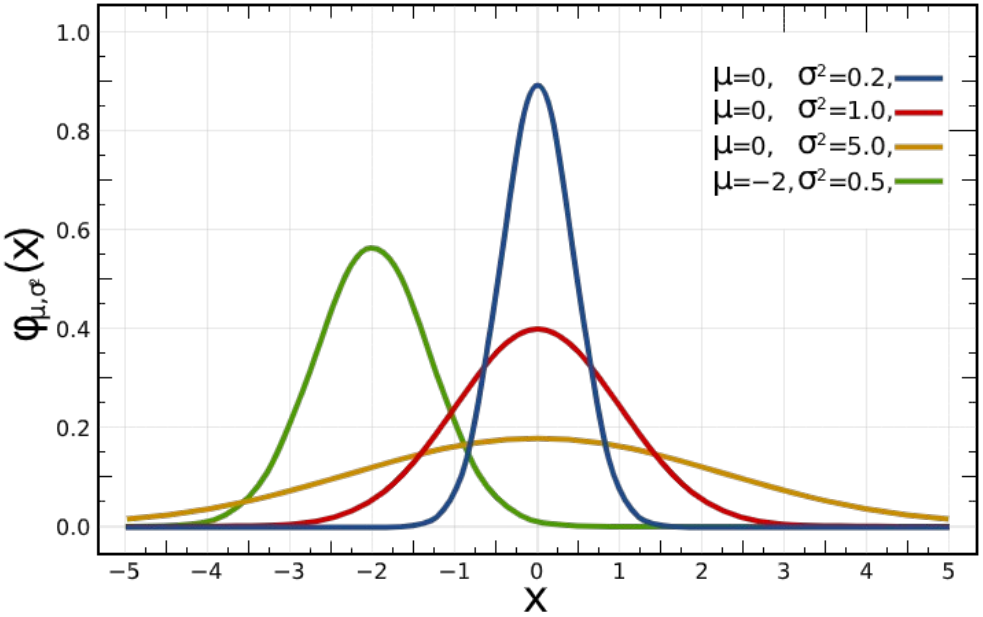




Standard Normal Distribution N(0,1)

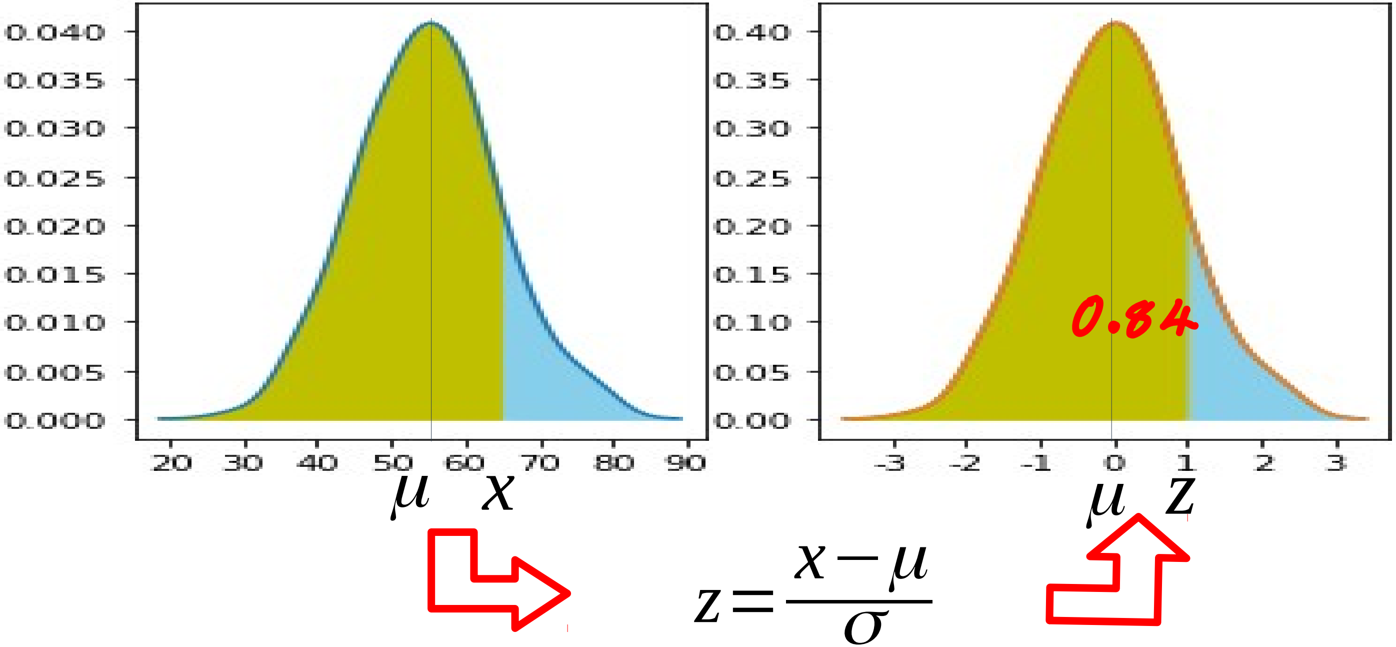
The simplest case of a normal distribution is known as the standard normal distribution. This is a special case when μ = 0 and σ = 1, and it is described by this probability density function

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



Standard Normal Distribution N(0,1)

Normally distributed test scores with μ = 55 and σ = 10, Caliculate the probability of randomly picked test score below 65?

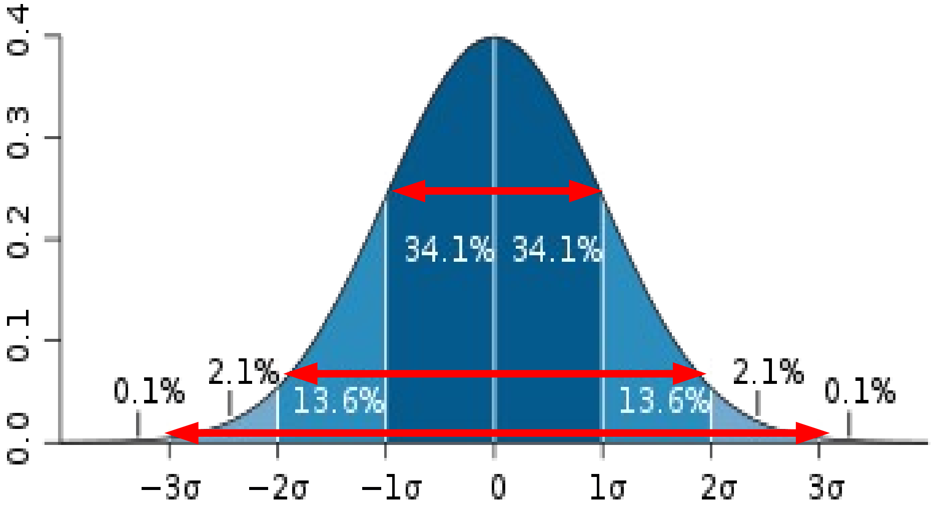


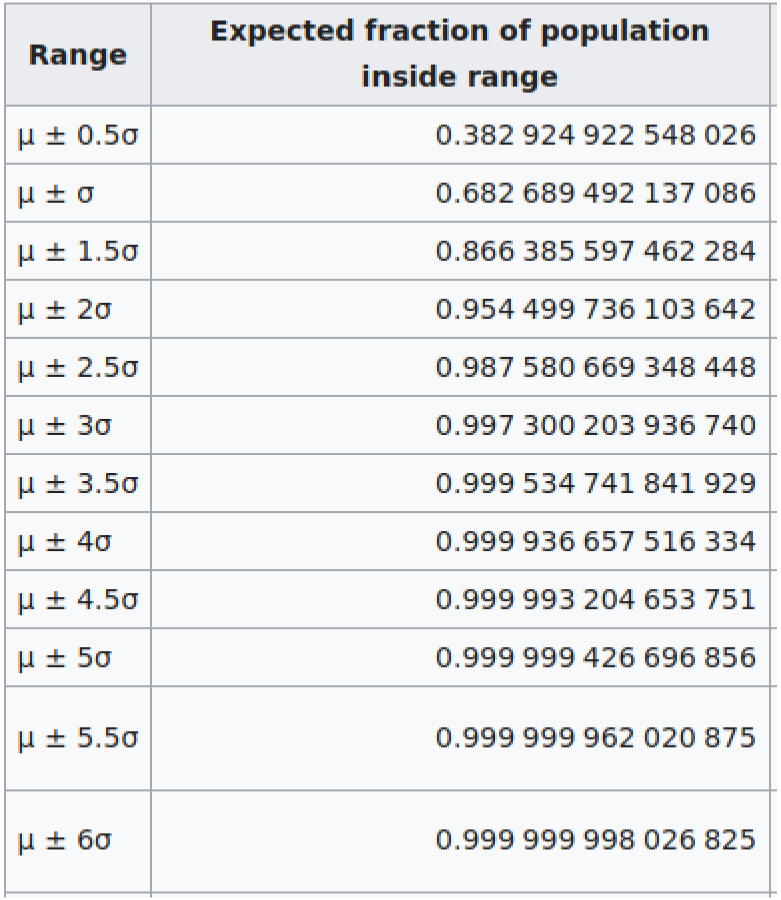
$$z = \frac{x - \mu}{\sigma}$$

z	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41308	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520

Standard deviation and coverage

3 sigma or 68–95–99.7 rule





Range	Expected fraction of population inside range
μ ± 0.5σ	0.382 924 922 548 026
μ ± σ	0.682 689 492 137 086
μ ± 1.5σ	0.866 385 597 462 284
μ ± 2σ	0.954 499 736 103 642
μ ± 2.5σ	0.987 580 669 348 448
μ ± 3σ	0.997 300 203 936 740
μ ± 3.5σ	0.999 534 741 841 929
μ ± 4σ	0.999 936 657 516 334
μ ± 4.5σ	0.999 993 204 653 751
μ ± 5σ	0.999 999 426 696 856
μ ± 5.5σ	0.999 999 962 020 875
μ ± 6σ	0.999 999 998 026 825

Confidence Interval

If you were asked to find the average maths score of inter students in Telangana?

Take a sample and find the mean ($\bar{x}$) this can be used to estimate the population parameter

The sample mean ($\bar{x}$) is a **point estimate** of the population mean (μ)

If sample is large $\bar{x}$ will be good estimate for μ(Central Limit Theorem)

It will be very difficult to estimate exact value and we don’t know how good it is

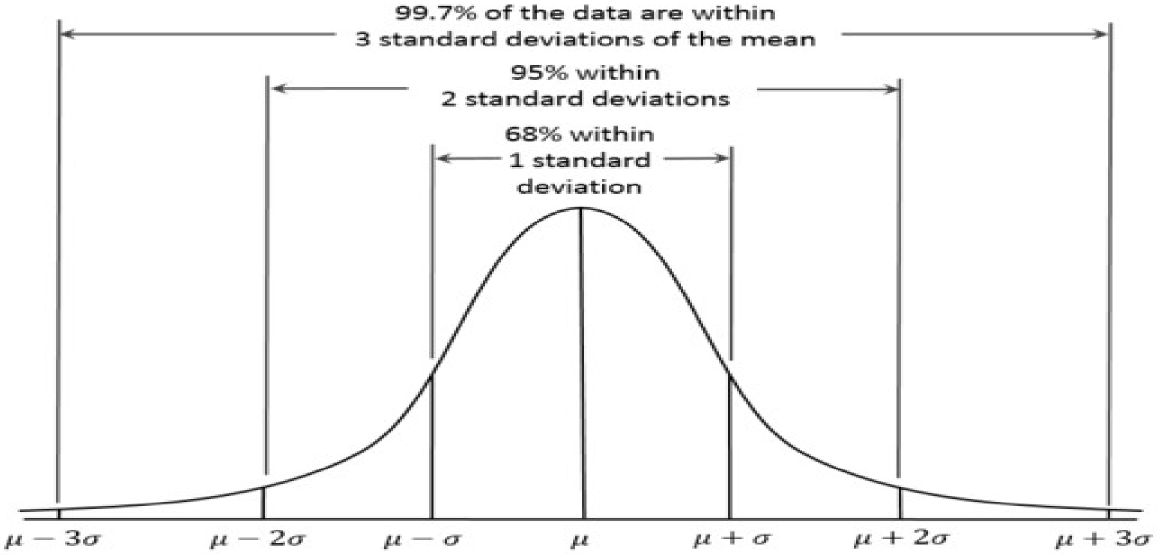
If our estimation is in terms of range of values there will be more chance of success in our estimate.

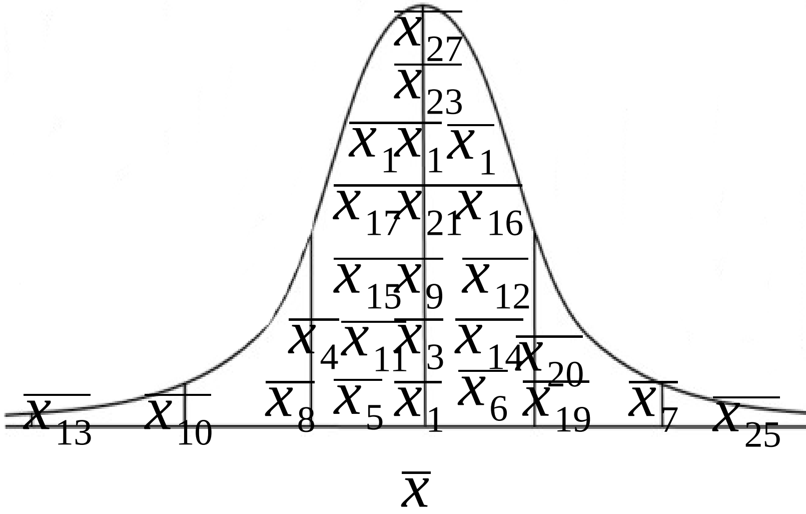
This type of estimation is called **interval estimate**

Margin of Error

The margin of error is a statistic expressing the amount of random sampling error in the results of a study

If we repeatedly take the samples of size n and caliculated those sample means $\overline{x_1}, \overline{x_2}, \ldots\ldots.$ these sample means will be normally distributed with mean $\overline{x}$

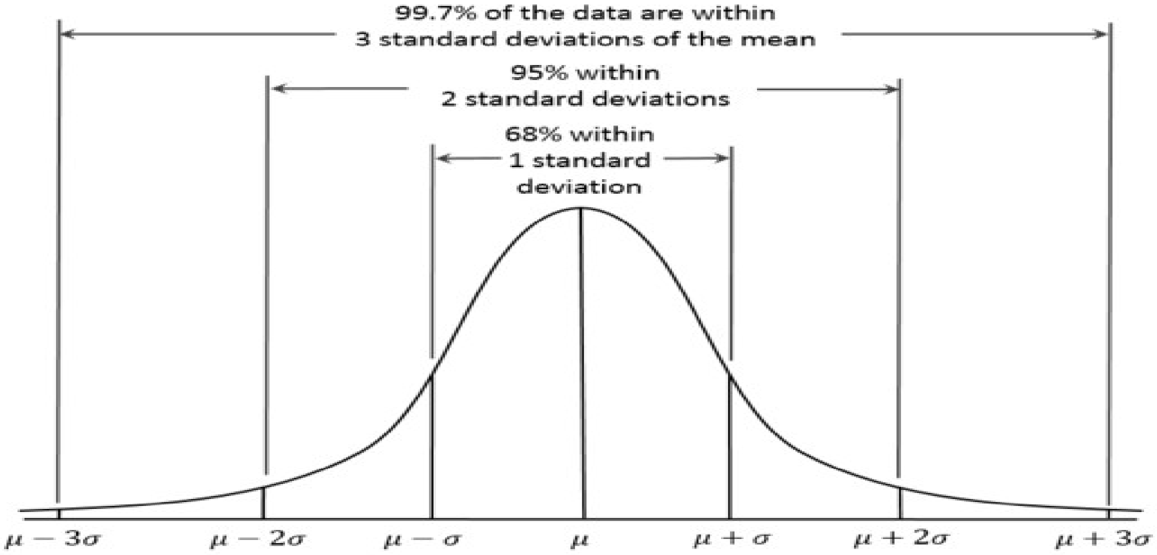


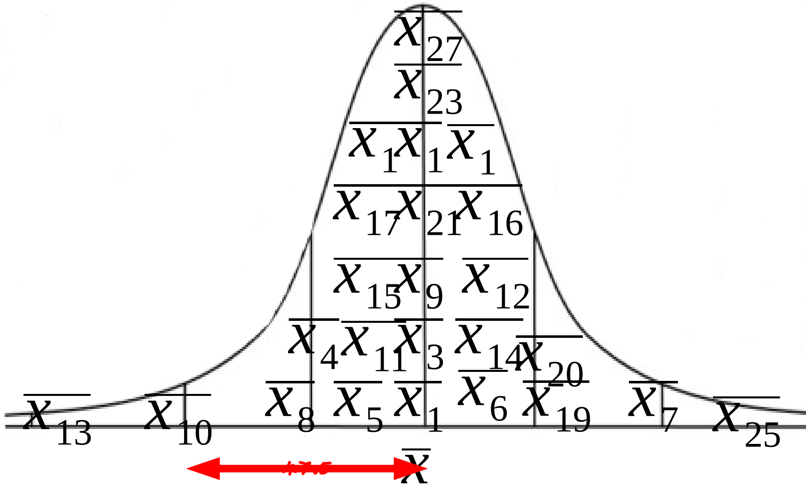

$\overline{x}_{27}$
$\overline{x}_{23}$
$\overline{x}_{1}$ $\overline{x}_{1}$ $\overline{x}_{1}$
$\overline{x}_{17}$ $\overline{x}_{21}$ $\overline{x}_{16}$
$\overline{x}_{15}$ $\overline{x}_{9}$ $\overline{x}_{12}$
$\overline{x}_{4}$ $\overline{x}_{11}$ $\overline{x}_{3}$ $\overline{x}_{14}$ $\overline{x}_{20}$
$\overline{x}_{8}$ $\overline{x}_{5}$ $\overline{x}_{1}$ $\overline{x}_{6}$ $\overline{x}_{19}$ $\overline{x}_{7}$
$\overline{x}_{13}$ $\overline{x}_{10}$ $\overline{x}_{25}$
$\overline{x}$

Margin of Error

The margin of error is a statistic expressing the amount of random sampling error in the results of a study

If we repeatedly take the samples of size n and caliculated those sample means $\overline{x}_1, \overline{x}_2, ...$ these sample means will be normally distributed with mean $\overline{x}$




$\overline{x}_{27}$
$\overline{x}_{23}$
$\overline{x}_1$ $\overline{x}_1$ $\overline{x}_1$
$\overline{x}_{17}$ $\overline{x}_{21}$ $\overline{x}_{16}$
$\overline{x}_{15}$ $\overline{x}_9$ $\overline{x}_{12}$
$\overline{x}_4$ $\overline{x}_{11}$ $\overline{x}_3$ $\overline{x}_{14}$ $\overline{x}_{20}$
$\overline{x}_8$ $\overline{x}_5$ $\overline{x}_1$ $\overline{x}_6$ $\overline{x}_{19}$ $\overline{x}_7$
$\overline{x}_{13}$ $\overline{x}_{10}$ $\overline{x}_{25}$
$\overline{x}$

Margin of Error

Generally, at a confidence level γ , a sample sized n of a population having expected standard deviation σ

$$MOE_{\gamma} = Z_{\gamma} * \left(\frac{\sigma}{\sqrt{n}}\right)$$

Here Z_{γ} is z score

$\frac{\sigma}{\sqrt{n}}$ is standard error

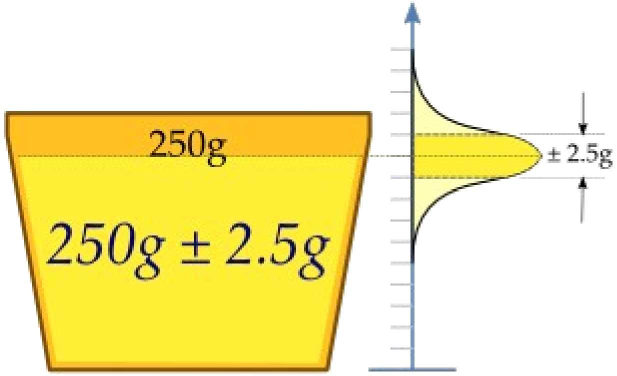
The value used to caliculate the upper limit and lower limit of the sample statistic before caliculating Confidence Interval(CI) choose 90% ,95% or 99% confidence level.

Confidence Interval

A machine fills cups with a milk, and is supposed to be adjusted so that the content of the cups is 250 g of liquid.

As the machine cannot fill every cup with exactly 250.0 g, the content added to individual cups shows some variation, and is considered a random variable X.

This variation is assumed to be normally distributed around the desired average of 250 g, with a standard deviation, σ, of 2.5 g.



Confidence Interval

To determine if the machine is adequately filled

a sample of n = 25 cups of milk is chosen at random and the cups are weighed.

The resulting measured masses of milk are x_1, ..., x_{25}

To get guess the population mean μ, it is sufficient to give an estimate. The appropriate estimator is the sample mean

The sample shows actual weights x_1, ..., x_{25}, with mean:

$$\bar{x} = \frac{1}{25} \sum_{i=1}^{25} x_i = 250.2\, grams$$

If we take another sample of 25 cups, we could easily expect to find mean values like 250.4 or 251.1 grams.

Confidence Interval

A sample mean value of 280 grams however would be extremely rare if the mean content of the cups is in fact close to 250 grams.

There is a whole interval around the observed value 250.2 grams of the sample mean within which, if the whole population mean actually takes a value in this range, the observed data would not be considered particularly unusual.

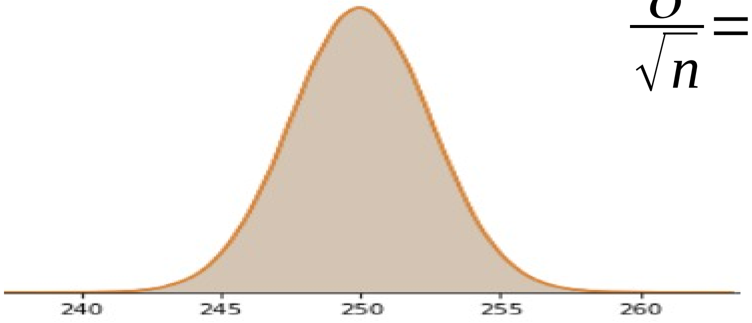
interval is called a confidence interval for the parameter μ.

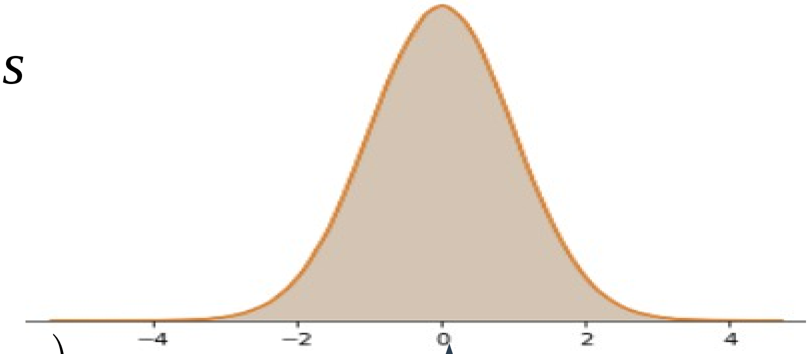
How do we calculate such an interval?

We can determine the endpoints by using the sample mean $\bar{x}$ which is normally distributed, with the same expectation μ, but with a standard error of: $\frac{\sigma}{\sqrt{n}}$

Confidence Interval

$$\frac{\sigma}{\sqrt{n}}=\frac{2.5\,g}{\sqrt{25}}=0.5\,grams$$





$$Z=\frac{(\bar{x}-\mu)}{\sigma/\sqrt{n}}=\frac{(\bar{x}-\mu)}{0.5}$$

it is possible to find numbers $-Z_{\frac{\alpha}{2}}$ and $Z_{\frac{\alpha}{2}}$, independent of μ, between which Z lies with probability 1 − α, a measure of how confident we want to be.

Confidence Level usually written as (1-α) 100%

α is probability of not observing true population mean in interval

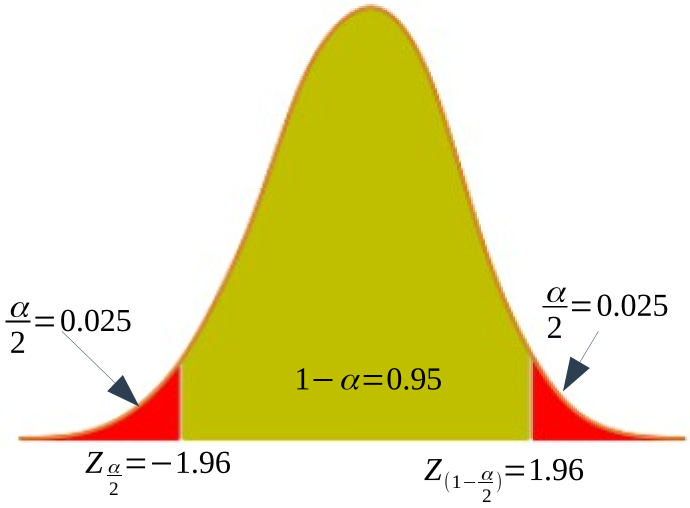
Caliculating Confidence Interval

- $P\left(Z \leq z_{\frac{\alpha}{2}}\right) \quad = \quad \frac{\alpha}{2} \quad = \quad 0.025$

$$P\left(Z \leq z_{\left(1-\frac{\alpha}{2}\right)}\right) \quad = \quad 1-\frac{\alpha}{2} \quad = \quad 0.975$$

$$z_{\frac{\alpha}{2}} \quad = z_{0.025} \quad = \quad -1.96$$

$$z_{\left(1-\frac{\alpha}{2}\right)} \quad = z_{0.975} \quad = \quad 1.96$$



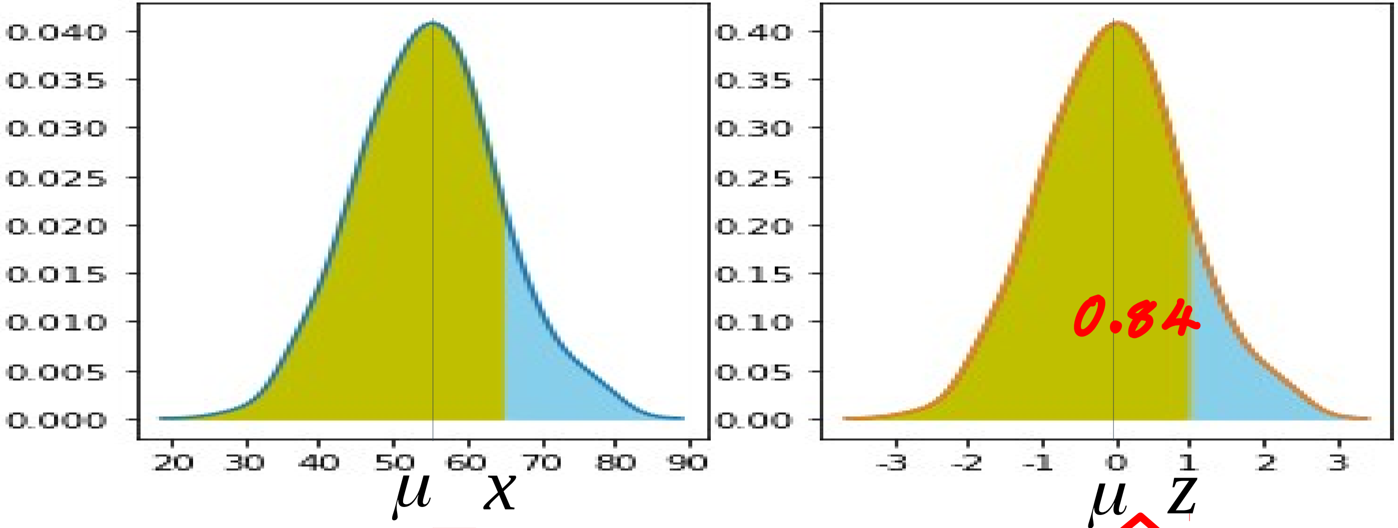
$\frac{\alpha}{2}=0.025$
$\frac{\alpha}{2}=0.025$
$1-\alpha=0.95$
$Z_{\frac{\alpha}{2}}=-1.96$
$Z_{\left(1-\frac{\alpha}{2}\right)}=1.96$

$$P\left(z_{\frac{\alpha}{2}} \leq Z \leq z_{\left(1-\frac{\alpha}{2}\right)}\right)=1-\alpha=0.95$$

$$P\left(z_{\frac{\alpha}{2}} \leq Z \leq z_{\left(1-\frac{\alpha}{2}\right)}\right)=P\left(-1.96 \ \leq \frac{(\bar{x}-\mu)}{\sigma/\sqrt{n}} \leq \ 1.96\right)=P\left(\bar{x}-1.96\,\frac{\sigma}{\sqrt{n}} \ \leq \mu \leq \bar{x}+1.96\,\frac{\sigma}{\sqrt{n}}\right)$$

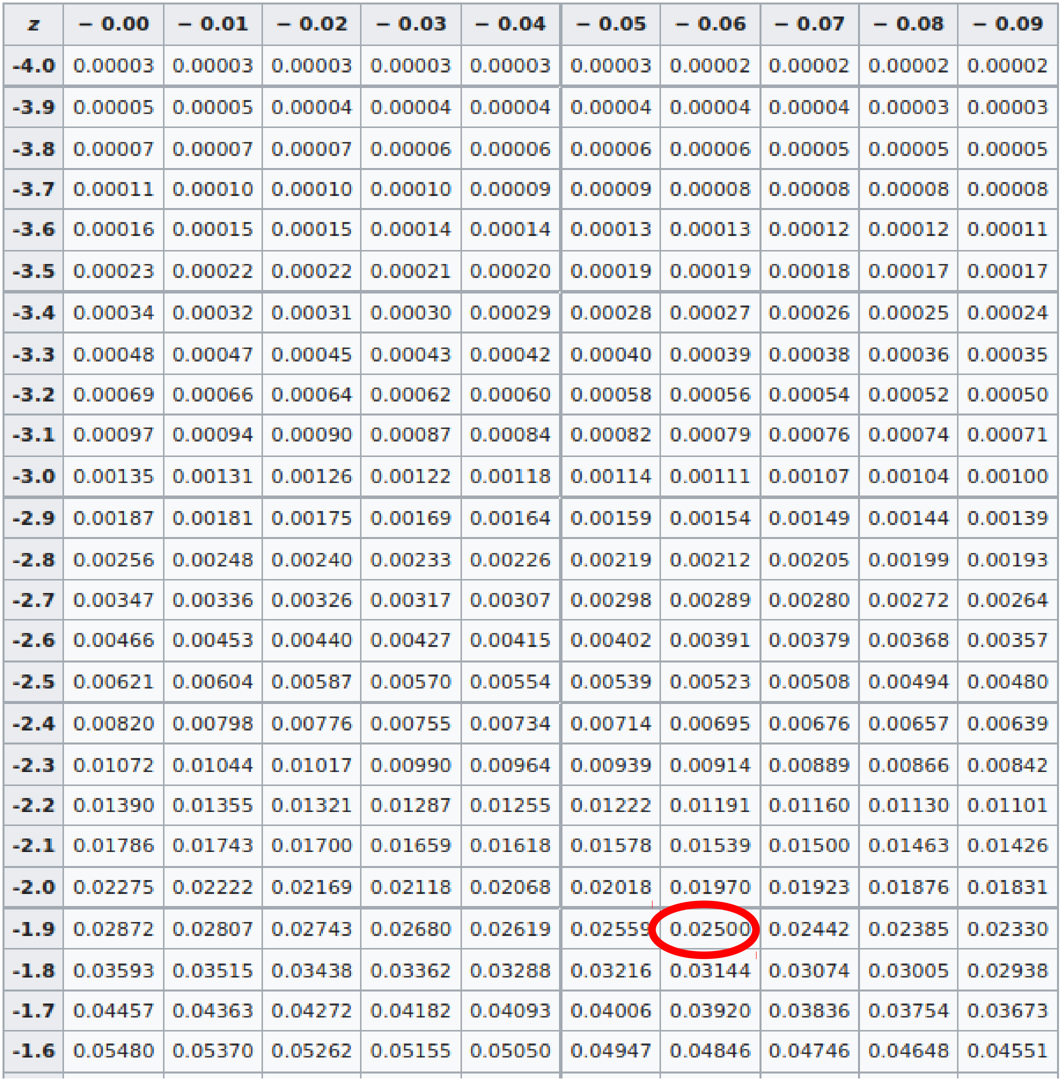
Standard Normal Distribution N(0,1)

Normally distributed test scores with μ = 55 and σ = 10, Caliculate the probability of randomly picked test score below 65?

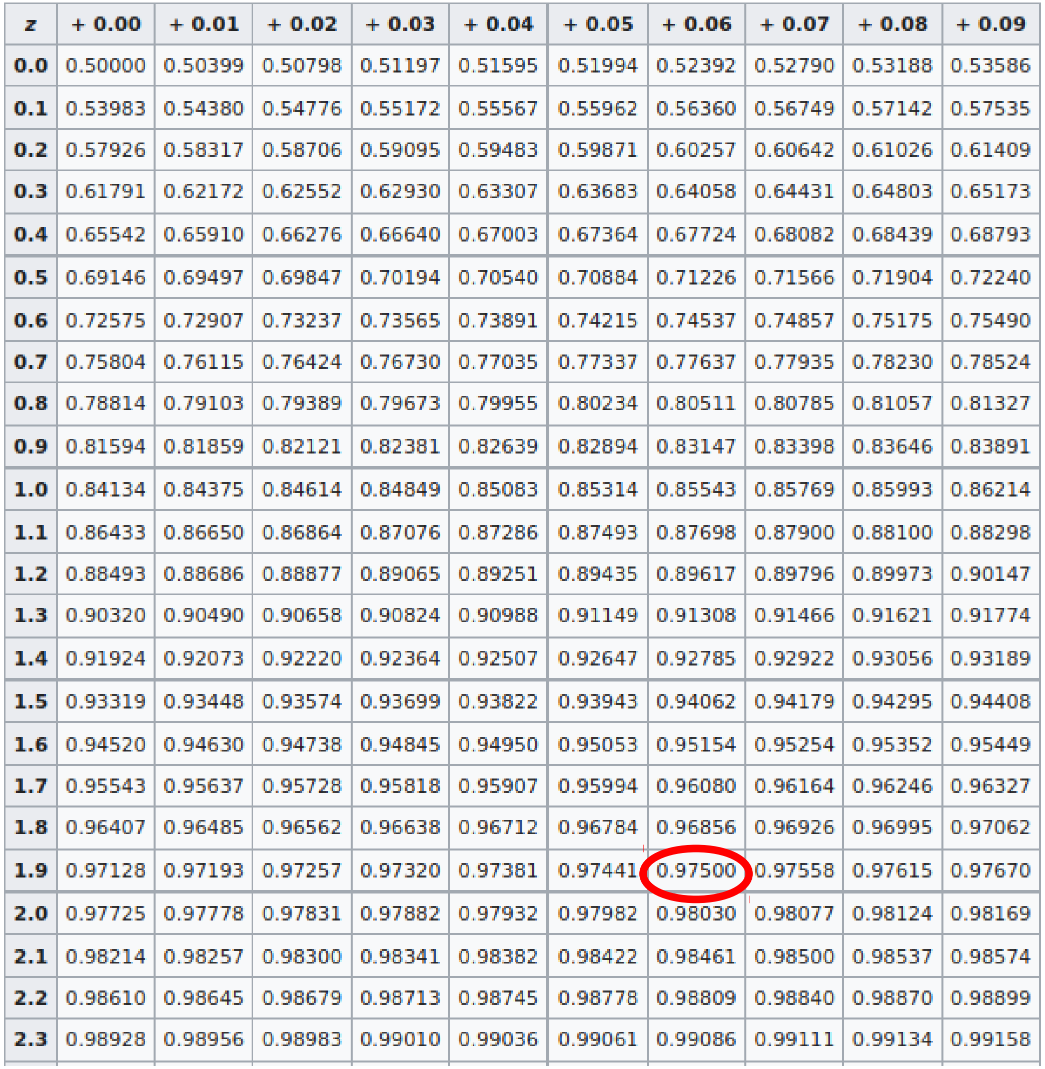


$$z = \frac{x - \mu}{\sigma}$$

Cumilative probabilities - statistic is less than Z (i.e. between negative infinity and Z)



z	− 0.00	− 0.01	− 0.02	− 0.03	− 0.04	− 0.05	− 0.06	− 0.07	− 0.08	− 0.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551



z	+ 0.00	+ 0.01	+ 0.02	+ 0.03	+ 0.04	+ 0.05	+ 0.06	+ 0.07	+ 0.08	+ 0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158

Caliculating Confidence Interval

In other words, the lower endpoint of the 95% confidence interval is

$$Lower\ Limit\ \ =\ \ \bar{x} - 1.96\frac{\sigma}{\sqrt{n}}$$

and the upper endpoint of the 95% confidence interval is:

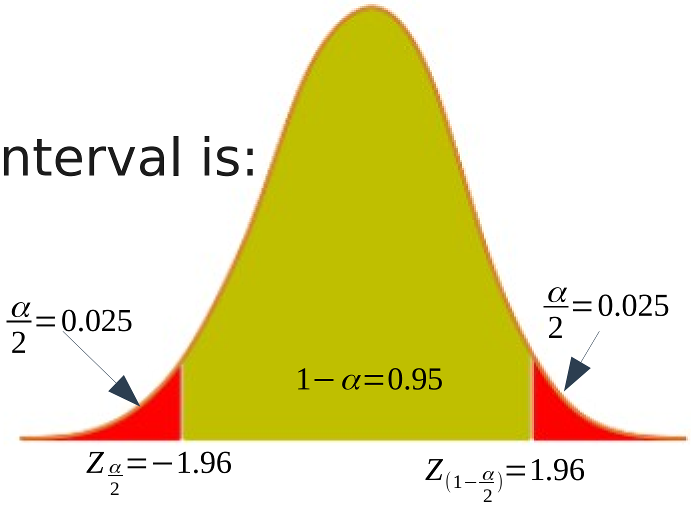
$$Upper\ Limit\ \ =\ \ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

$$0.95\ \ =\ \ P\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}\ \ \le \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

$$0.95 = \ \ P\left(\bar{x} - 1.96 * 0.5\ \ \le \mu \le \bar{x} + 1.96 * 0.5\right)$$

$$0.95 = \ \ P\left(250.2 - 0.98\ \ \le \mu \le 250.2 + 0.98\right)$$

$$0.95 = \ \ P\left(249.22\ \ \le \mu \le 251.18\right)$$



so 95% confidence interval is: (249.22, 251.18)

Confidence Interval

Confidence Interval captures a “true” (yet unknown) measure of a population using sample data.

A confidence interval is an interval of values instead of a single point estimate.

For repeated measurements with the same sample sizes, taken from the same population, X% of times the CI obtained will contain the true parameter value.

Confidence Interval interpretation

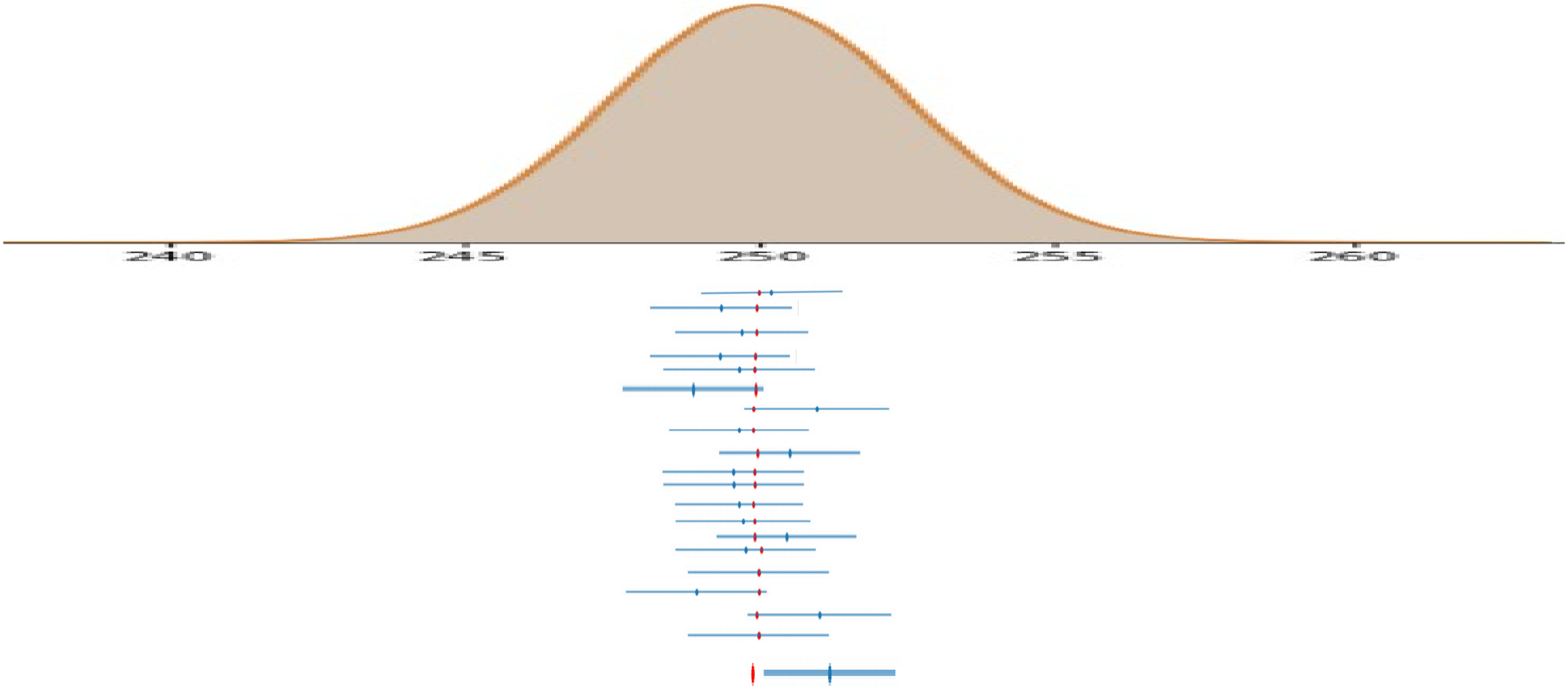
Confidance Interval of 95%

does not mean there is 0.95 probability that the value of parameter μ is in the interval obtained by using the currently computed value of the sample mean,

$$\left(\bar{x}-0.98\,,\bar{x}+0.98\right)$$

Instead, every time the measurements are repeated, there will be another value for the mean X of the sample. In 95% of the cases μ will be between the endpoints calculated from this mean, but in 5% of the cases it will not be.

Confidence Interval



Confidence Interval

Width of interval is decided by margin of error

Wider interval captures population mean accurately but less precise

$$Margin\, of\, Error = Z_{\gamma} * \left(\frac{\sigma}{\sqrt{n}}\right)$$

Insted of increasing confidence level increase sample size(n) it will reduce width of margin and will give you more precision

Caliculating Confidence Interval

Example: Suppose we measure the 35 students height the mean height is 165cm and standard deviation is 20cm. caliculate 95% confidence interval for population mean

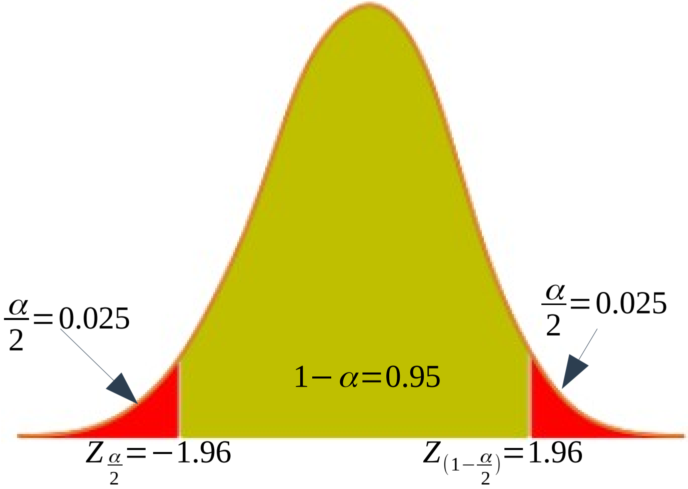
$$CI = \overline{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$CI = \overline{X} \pm 1.96 \frac{s}{\sqrt{n}}$$

$$CI = 165 \pm 1.96 \frac{20}{\sqrt{40}}$$

$$CI = 165 \pm 6.20$$

$$CI = [158.8, 171.2]$$



$\frac{\alpha}{2} = 0.025$
$\frac{\alpha}{2} = 0.025$
$1 - \alpha = 0.95$
$Z_{\frac{\alpha}{2}} = -1.96$
$Z_{(1-\frac{\alpha}{2})} = 1.96$

we can be 95% confident that the population mean (μ) falls between 158.8 and 171.2.

Caliculating Confidence Interval

Example: A sample of 49 students choosen to estimate average test scores if sample mean is 65 and standard deviation of population is 10 caliculate 95% confidence interval for population mean

$$n = 40, \overline{X} = 65, \sigma = 10, \alpha = 0.5$$

$$CI = \overline{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$CI = 65 \pm 1.96 \frac{10}{\sqrt{49}}$$

$$CI = 65 \pm 1.96 * 1.43$$

$$CI = 65 \pm 2.8$$

$$CI = [62.2, 67.8]$$

Caliculating Confidence Interval

Example: A sample of 49 students choosen to estimate average test scores if sample mean is 65 and standard deviation of population is 10 find the probability for population mean is greater than 68

$$n=40, \overline{X}=65, \sigma=10, \alpha=0.5$$

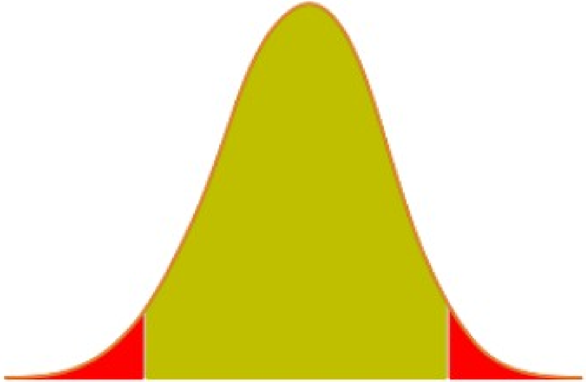
$$CI=\overline{X}\pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$CI=65\pm1.96\frac{10}{\sqrt{49}}$$

$$CI=65\pm1.96*1.43$$

$$CI=65\pm2.8$$

$$CI=[62.2, 67.8]$$



Caliculating Confidence Interval

Example: A sample of 49 students choosen to estimate average test scores if sample mean is 65 and standard deviation of population is 10 find the probability for population mean is greater than 68

$$n=40, \overline{X}=65, \sigma=10, \alpha=0.5$$

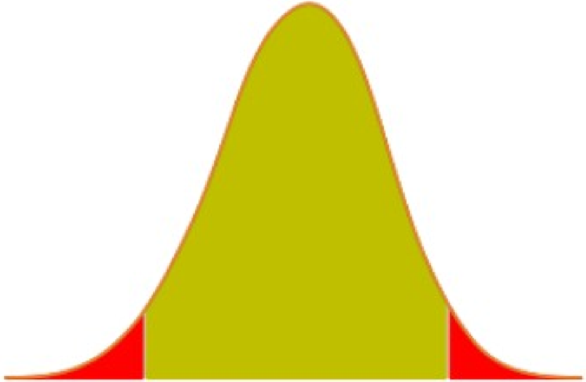
$$CI=\overline{X}\pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$CI=65\pm 1.96\frac{10}{\sqrt{49}}$$

$$CI=65\pm 1.96*1.43$$

$$CI=65\pm 2.8$$

$$CI=[62.2, 67.8]$$



67.8 is the upper limit of 95% confidence interval so aproximate probability is 0.025

Hypothesis

Hypothesis is basically an idea that must be put to the test

An assumption about certain characteristic of population

Hypothesis Testing

The statistical method of statistical inference.

Hypothesis testing

1. Significance testing
2. Confidence Intervals

Significance-based hypothesis testing is the most common framework for statistical hypothesis testing

Hypothesis Testing

Suppose that in a particular geographic region in a reading test the mean and standard deviation of scores are 100 points, and 12 points, respectively.

In a particular school The mean score of 55 students is 96.

We can ask whether this mean score is significantly lower than the regional mean?

that is, are the students in this school comparable to the region students, or are their scores surprisingly low?

Hypothesis Testing

Error might be because of sampling bias

Hypothesis testing tests an assumption regarding a population parameter.

The **null hypothesis** is usually a hypothesis of equality between population parameters

e.g., a null hypothesis may state that the difference between population means is equal to zero

H_0 : μ >= 100 Null Hypothesis(No difference)

The scool students are comparable to the population of test-takers

Hypothesis Testing

The alternative hypothesis is effectively the opposite of a null hypothesis;

e.g., Alternative hypothesis may state that the difference between population means is not equal to zero.

Thus, they are mutually exclusive, and only one can be true. However, one of the two hypotheses will always be true

H_1 : $\mu < 100$ Alternate Hypothesis(there is difference)

The school students have an unusually low mean test score from the population of test-takers.

Statistical Significance

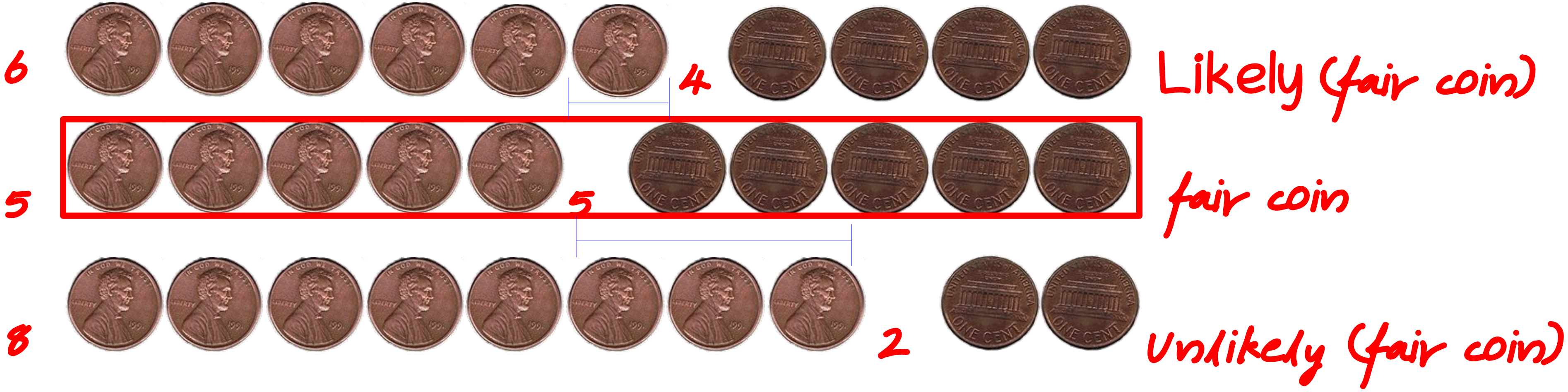
Suppose you are trying to decide weather a coin is fair or biased

H_0: coin is fair

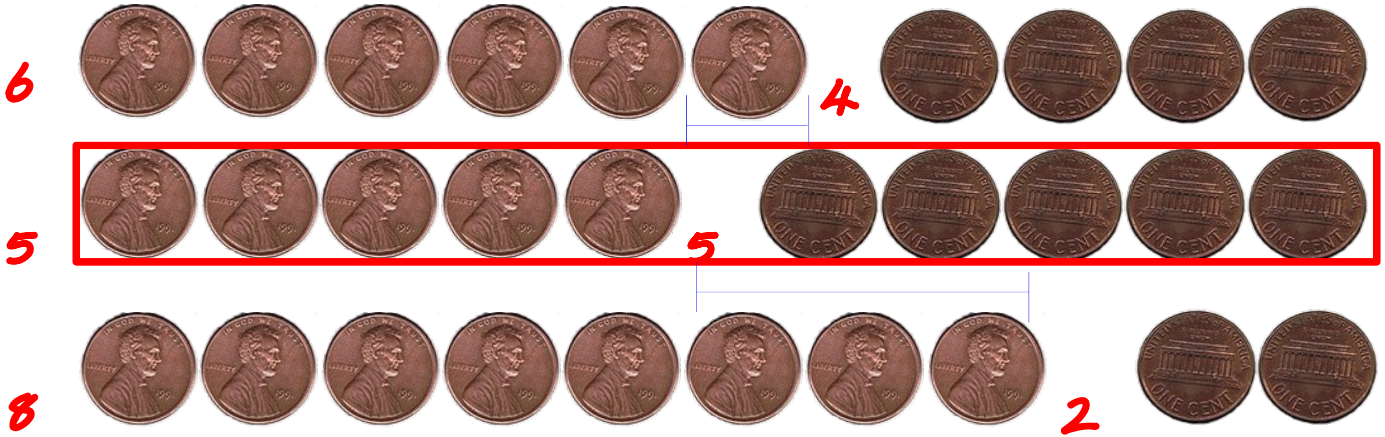
H_1: coin is biased

$$H_0: \; p_H = 0.5$$
$$H_1: \; p_H \neq 0.5$$

Statistical Significance



Statistical Significance



Likely (fair coin)

fair coin

Unlikely (fair coin)

Small Variation from expected value is considered as random chance

Huge Variation from expected value is not considered as random chance

Statistical Significance

H_0: coin is fair

H_1: coin is biased

$$H_0: \ p_H = 0.5$$

$$H_1: \ p_H \neq 0.5$$

you would reject the null hypothesis in favor of the alternative hypothesis if the observed number of heads difference is huge from 5.

How to decide defference is huge or not?

We need some cut of

$$no\ of\ heads \leq 2$$

$$no\ of\ heads \geq 8$$

Statistical Significance

Observations in statistical study is said to be **statistically significant** if it is unlikely to have occured by **random chance**?

We need to detrermine how unlikely an outcome was to occur by random chance, and then deciding if that probability is unlikely enough that we can conclude something other than random chance caused that outcome.

Statistical Significance

You are trying to determine if the health drink makes the children gain height.

To test that a study was conducted, 100 children were selected for study.

You need some standard to compare this study results.

Suppose average height of 5 year children is 100 cm with standard deviation is 2cm(in normal conditions)

Health drink was given daily for them(assume except this health drink other conditions are same) and you measured their average height when their age is 5 years.

you found that the children who took the health drink gained more height than 100 cm.

Statistical Significance

Does that mean the health drink caused the increased height gain?

For instance, may be the health drink is the factor for change in the results.

Or may be that change in results is just random chance.

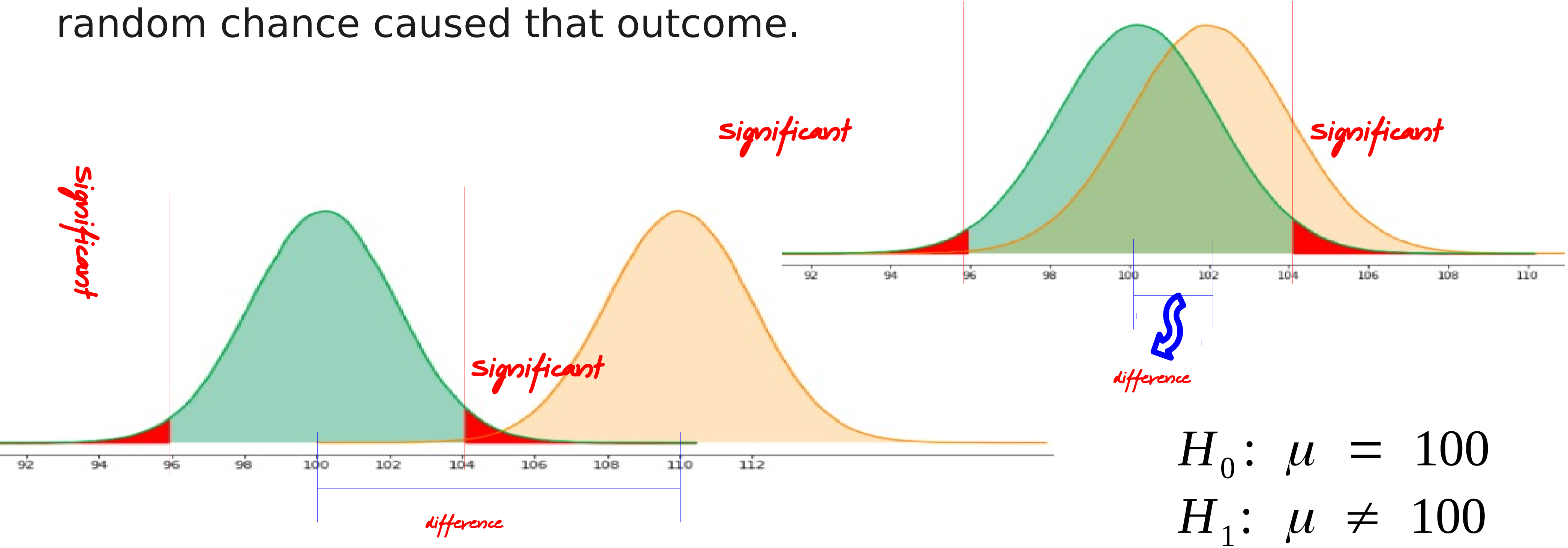
Statistical Significance is a way of determining which of those possibilities is true.

1. Height is not dependent on health drink(change in results is just random variation)

2. Height is dependent on health drink(change in results is not random variation really because of health drink)

Statistical Significance

We need to detrermine how unlikely an outcome was to occur by random chance in null hypothesis, and then deciding if that probability is **unlikely enough** in null hypothesis that we can conclude something other than random chance caused that outcome.



$$H_0: \ \mu \ = \ 100$$
$$H_1: \ \mu \ \neq \ 100$$

Hypothesis Testing

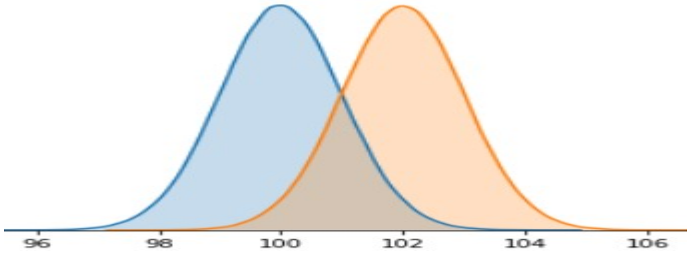
A statistical test can have one of two outcomes

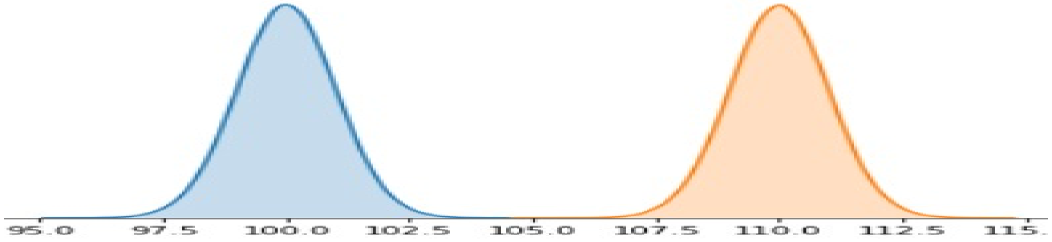
1. Null hypothesis is rejected and alternate Hypothesis accepted
2. Null hypothesis is not rejected based on the evidence

Statistical Significance is used to determine whether the null hypothesis should be rejected or not rejected based on the evidence

Statistical Significance

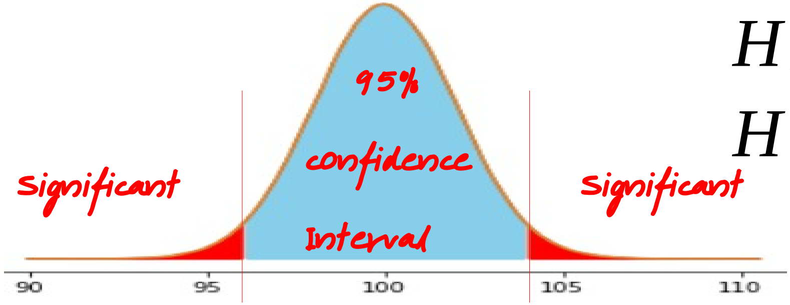
Statistical significance helps us to identify change in results due to random chance or due to some factor(like health drink).

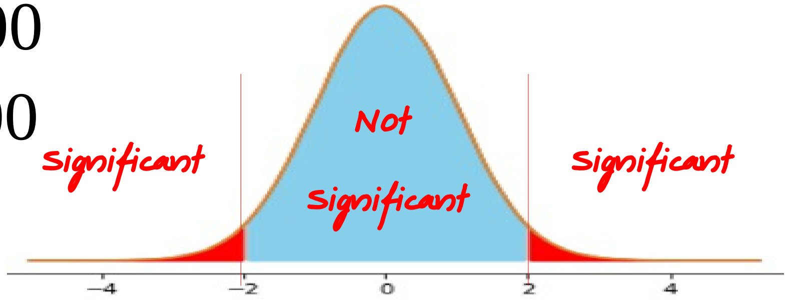




$$H_0: \ \mu \ = \ 100$$

$$H_1: \ \mu \ \neq \ 100$$





Hypothesis Testing

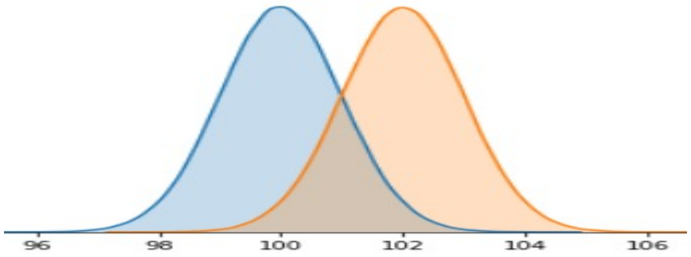
1. Null hypothesis is rejected and alternate Hypothesis accepted

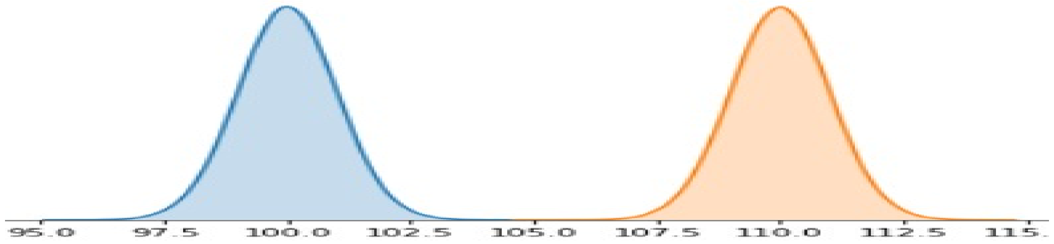
For the null hypothesis to be rejected, an observed result has to be statistically significant.

a. If observed statistic value does not falls in the pre-specified confidance interval region, we can say that the null hypothesis is very unlikely(statistically significant)

b. If observed p-value is less than the pre-specified significance level(α) (statistically significant)

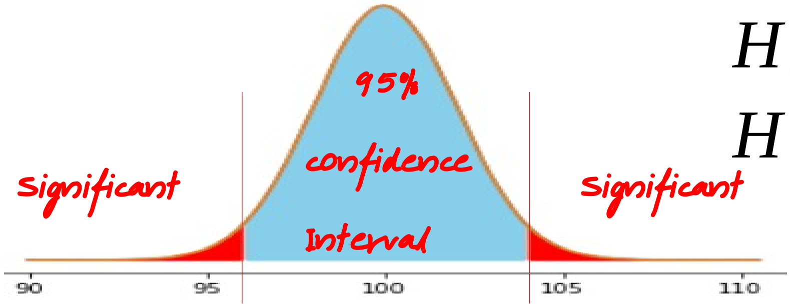
Statistical Significance

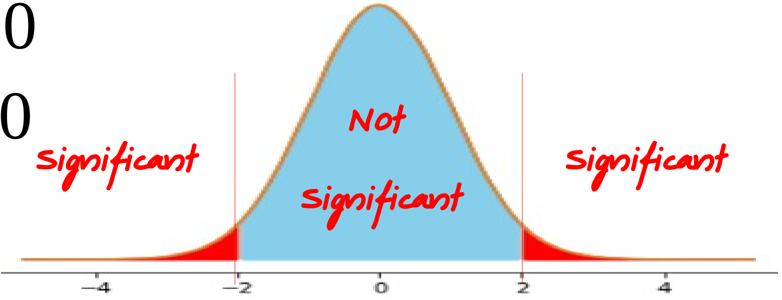




$$H_0:\ \mu\ =\ 100$$

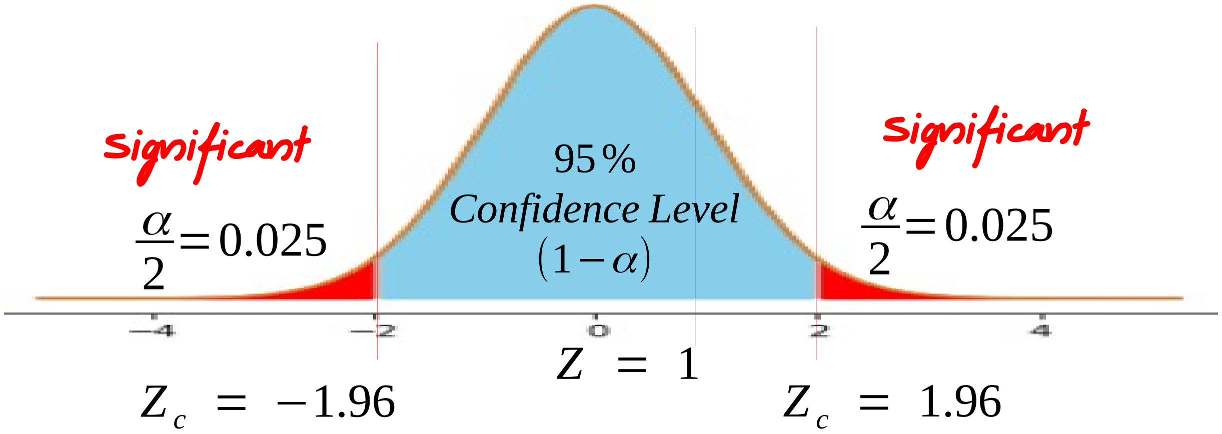
$$H_1:\ \mu\ \neq\ 100$$





Hypothesis Testing

1. Null hypothesis is rejected and alternate Hypothesis accepted



$\frac{\alpha}{2}=0.025$ 95% Confidence Level $(1-\alpha)$ $\frac{\alpha}{2}=0.025$

$Z_c = -1.96$ $Z = 1$ $Z_c = 1.96$

$$H_0: \mu = 100$$
$$H_1: \mu \neq 100$$

Observed statistic value falls in the pre-specified confidance interval region, we can say that the null hypothesis is very likely

Result is not statistically significant

Can’t Reject null hypothesis

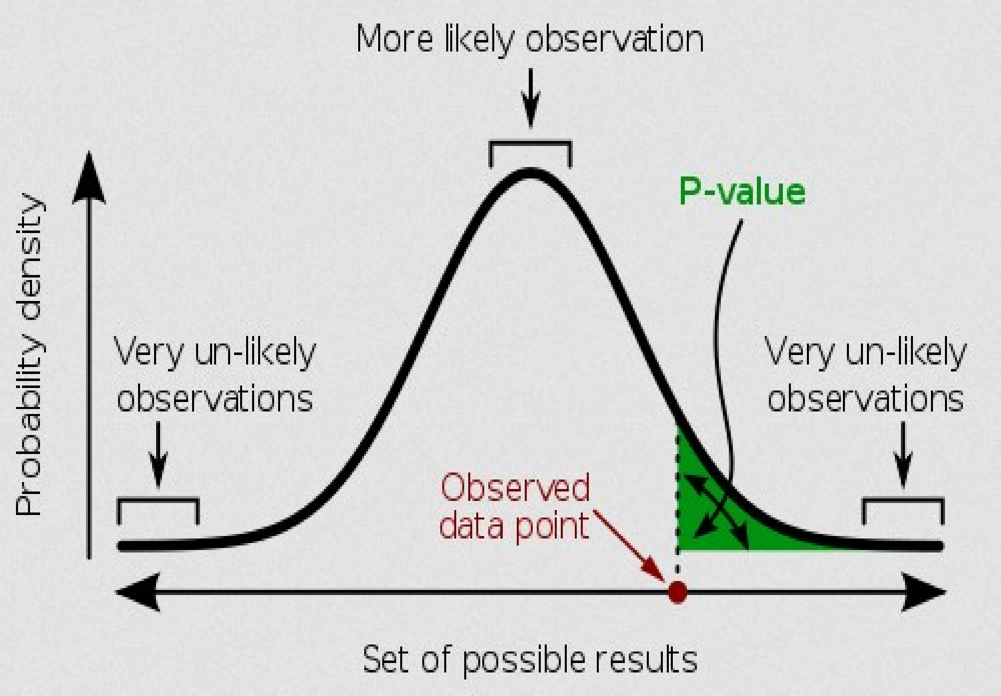
p value

p-value is the probability of obtaining test results at least as extreme as the results actually observed, when the null hypothesis is correct

The P-valueis frequently called the "tail probability."

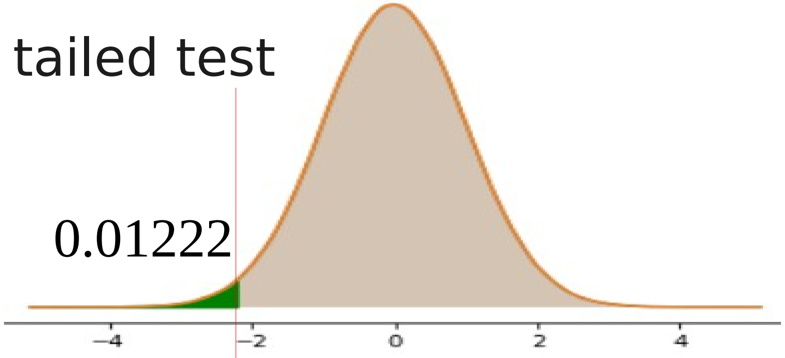
A p-value is a measure of the probability that an observed difference could have occurred just by random chance.

The lower the p-value, the greater the statistical significance of the observed difference.



p value

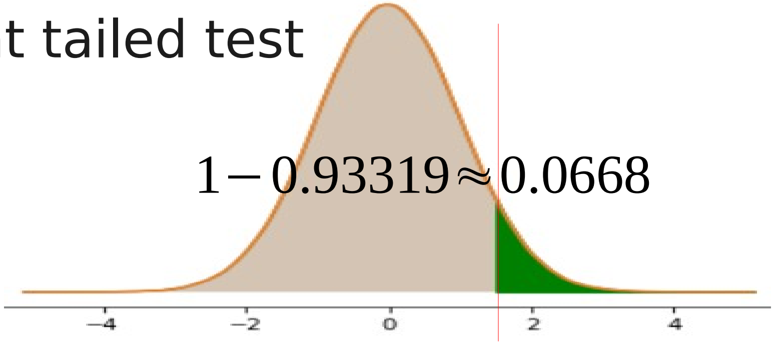
P value Left tailed test



0.01222

$Z = -2.25$

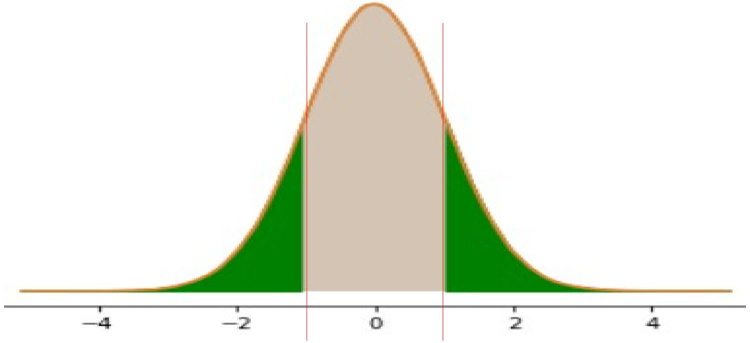
P value Right tailed test



$1 - 0.93319 \approx 0.0668$

$Z = 1.5$

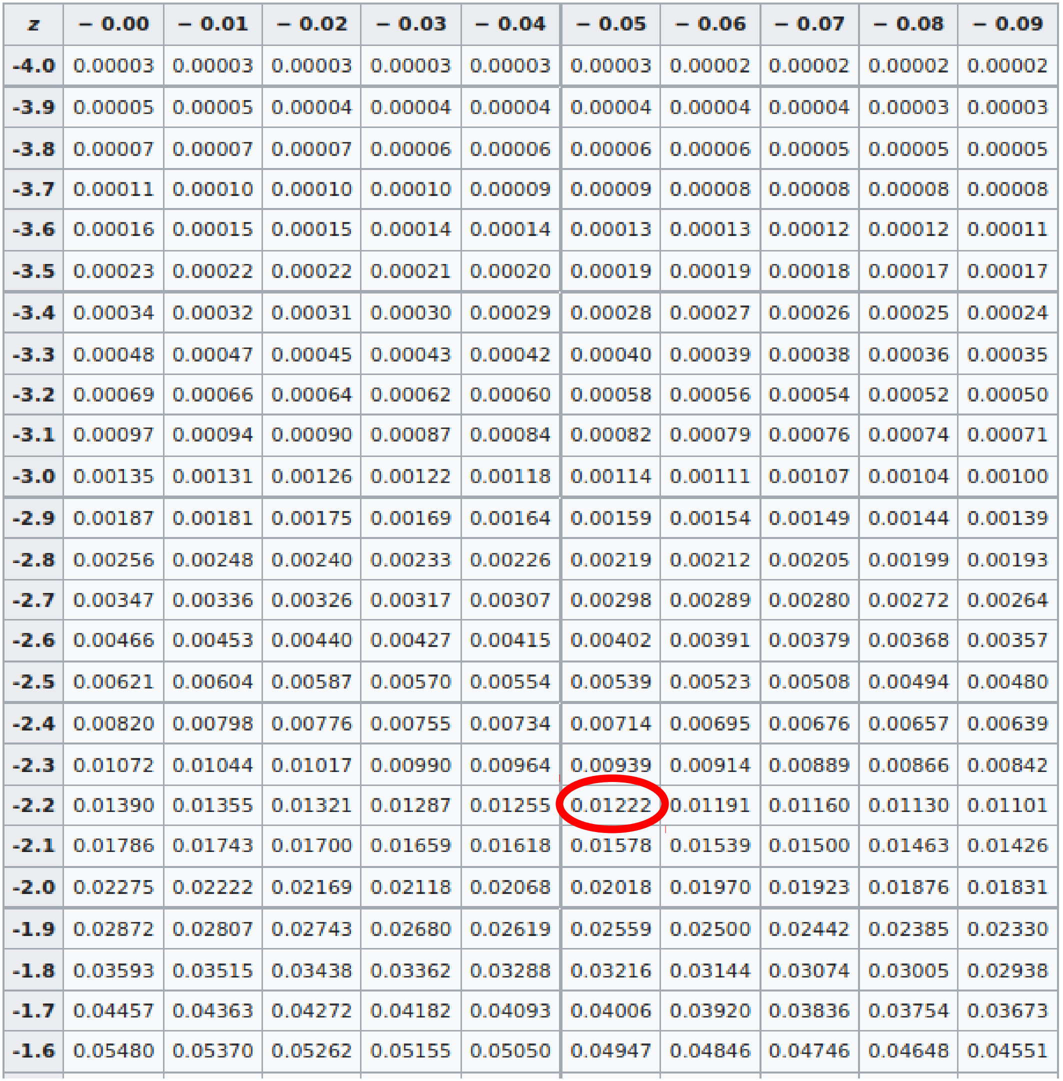
P value Two tailed test



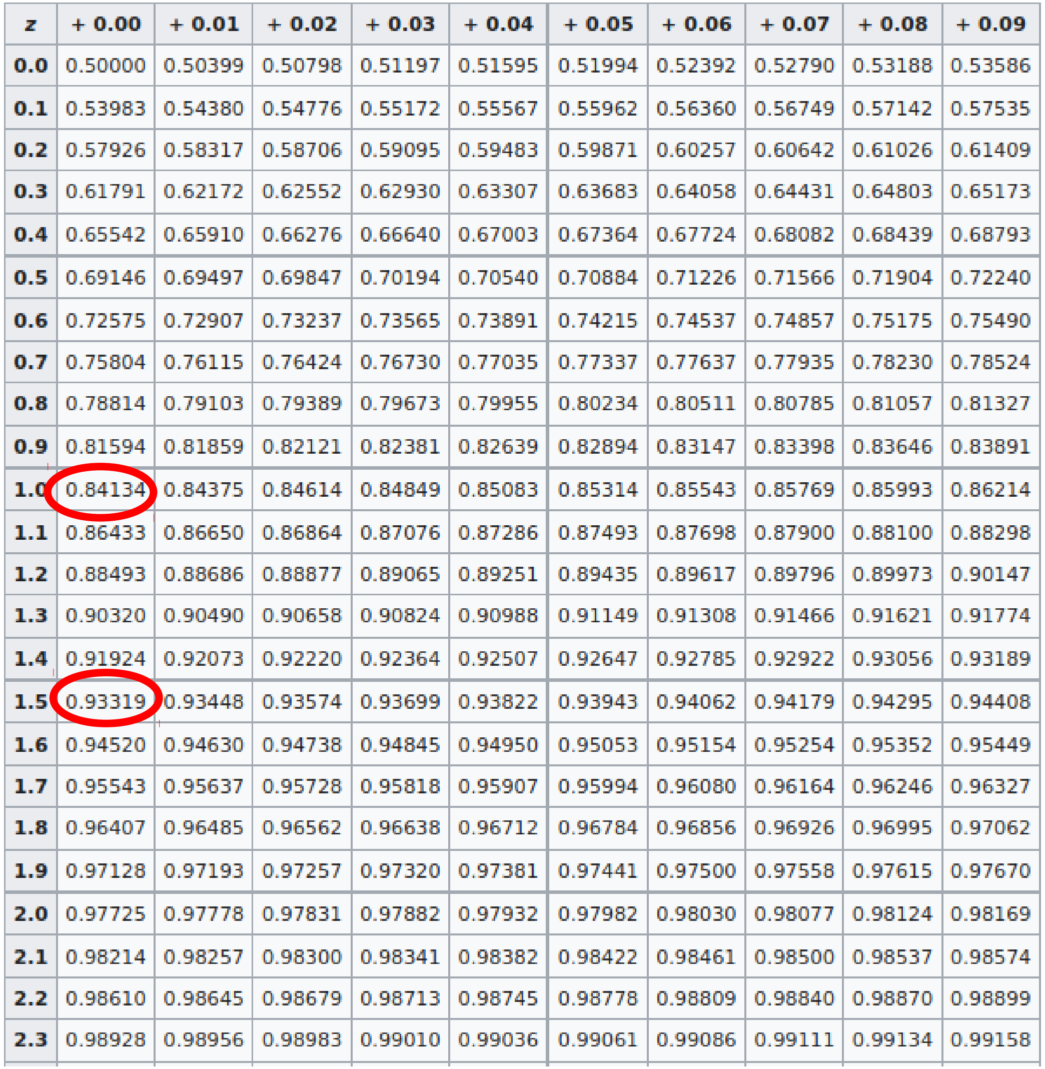
$Z = 1$

$2 * 0.158 \approx 0.3173$

Cumilative probabilities - statistic is less than Z (i.e. between negative infinity and Z)



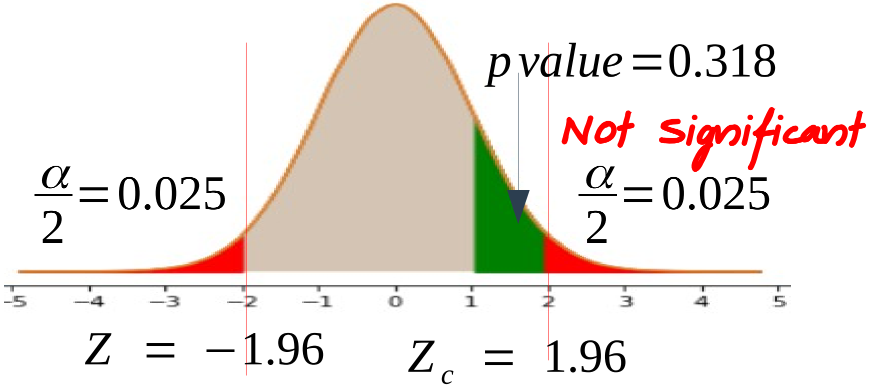
z	− 0.00	− 0.01	− 0.02	− 0.03	− 0.04	− 0.05	− 0.06	− 0.07	− 0.08	− 0.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551

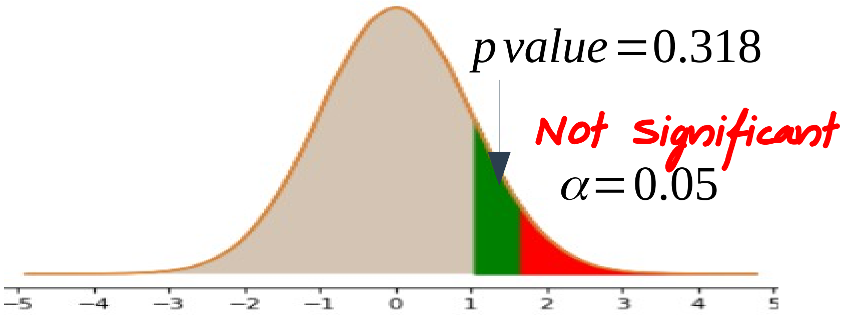


z	+ 0.00	+ 0.01	+ 0.02	+ 0.03	+ 0.04	+ 0.05	+ 0.06	+ 0.07	+ 0.08	+ 0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158

Hypothesis Testing

1. Null hypothesis is rejected and alternate Hypothesis accepted





$$p(0.318) \nless 0.05(\alpha)$$

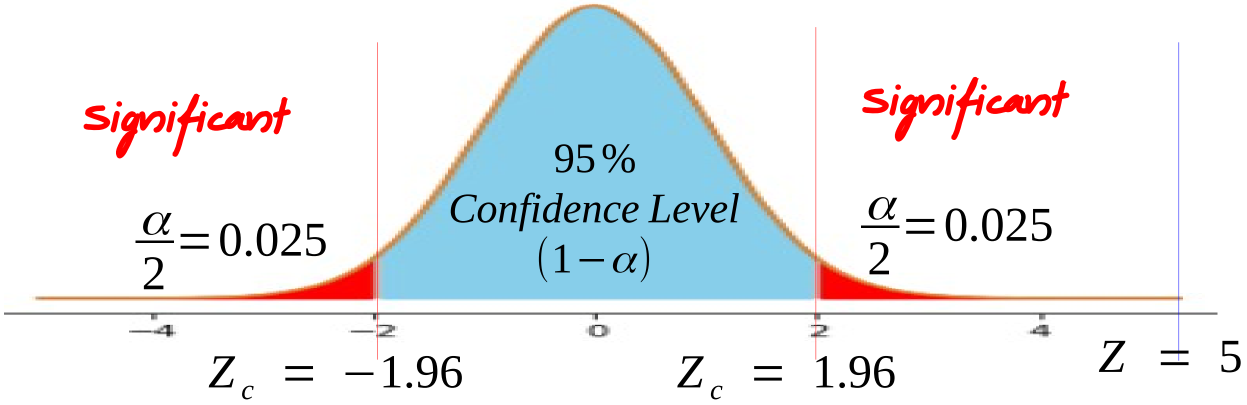
Observed p-value is not less than the pre-specified significance level(α)

Result is not statistically significant

Can’t Reject null hypothesis

Hypothesis Testing

2. Null hypothesis is not rejected based on the evidence



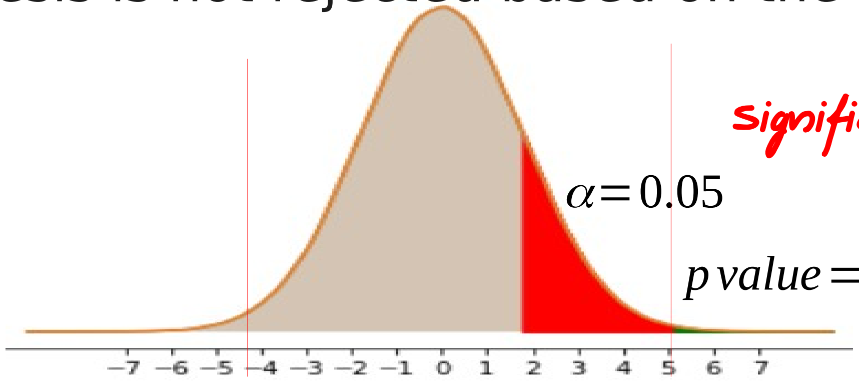
Observed statistic value do not falls in the pre-specified confidance interval region, we can say that the null hypothesis is very unlikely

Result is statistically significant

Can Reject null hypothesis

Hypothesis Testing

2. Null hypothesis is not rejected based on the evidence



significant

$\alpha = 0.05$

$p\,value = 0.0000006$

$Z = 5$

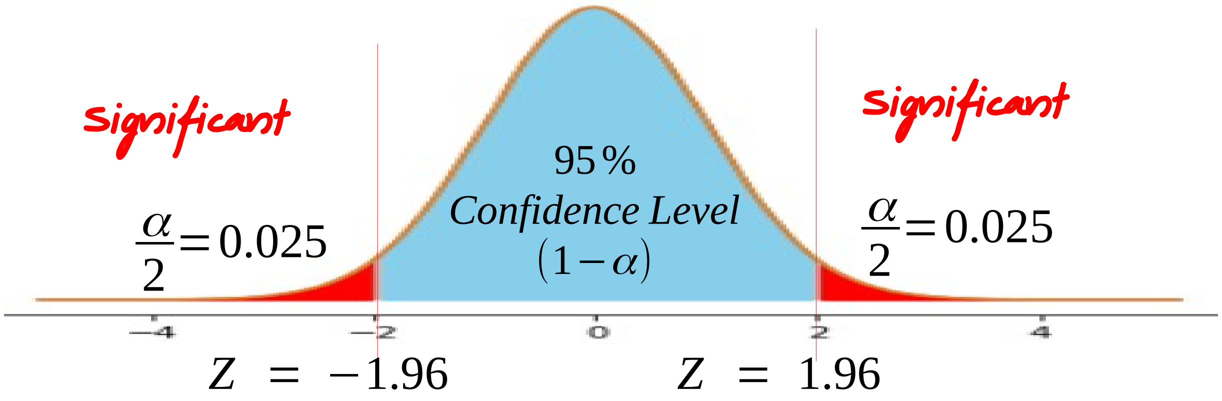
$p(0.0000006) < 0.05(\alpha)$

Observed p-value is less than the pre-specified significance level(α)

Result is statistically significant

Can’t Reject null hypothesis

Hypothesis Testing



It would be incorrect to conclude that the null hypothesis is not rejected it can be accepted as valid

Errors in Hypothesis Testing

A statistically significant result cannot prove that a research hypothesis is correct (as this implies 100% certainty).

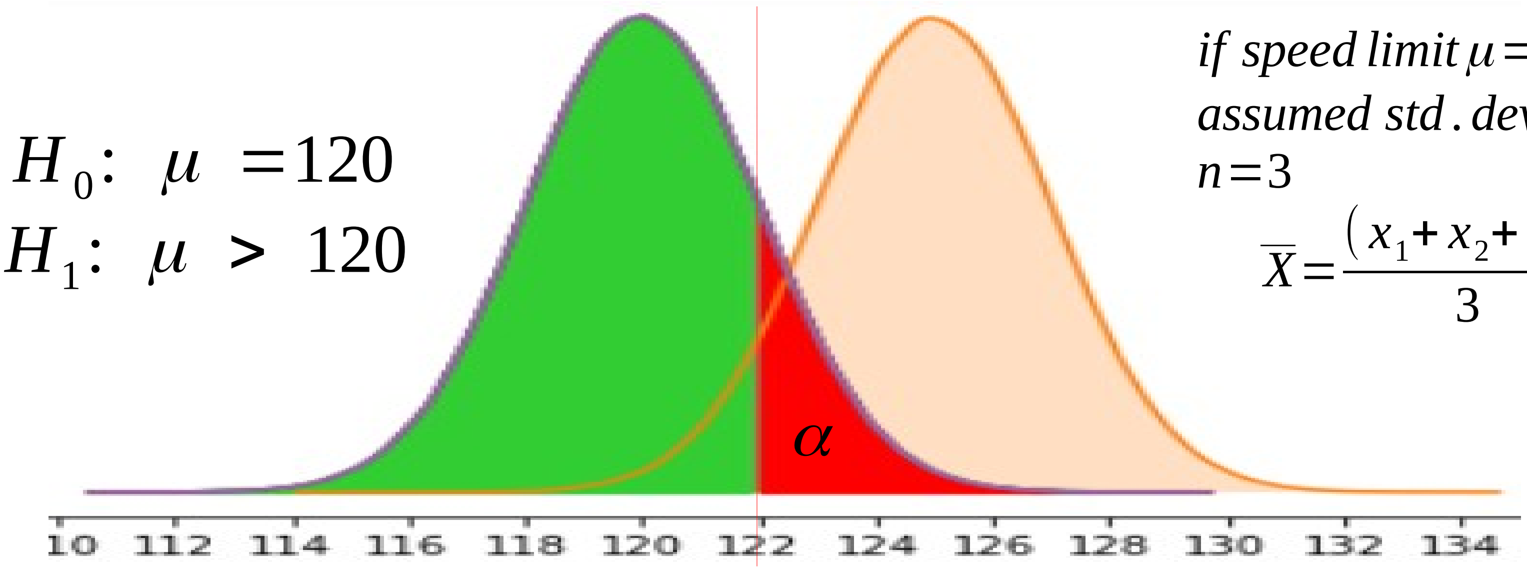
We do not have complete information about the population,and hence we work with a sample, which brings in an element of probability

Because a decision is based on probabilities, there is always a chance of making an incorrect conclusion regarding rejecting or not rejecting the null hypothesis (H0).

Errors in Hypothesis Testing

Type I Error(α): Conditional probability of false rejection of Null Hypotesis when it is true

$$Type\ I\ Error\ =\ \alpha\ =\ P\left(Rejecting\ Null\ Hypothesis\ \middle|\ H_0\ \ is\ \ true\right)$$



$$H_0:\ \ \mu\ =120$$
$$H_1:\ \ \mu\ >\ 120$$

$if\ speed\ limit\ \mu=120$

$assumed\ std.dev\ \sigma=2$

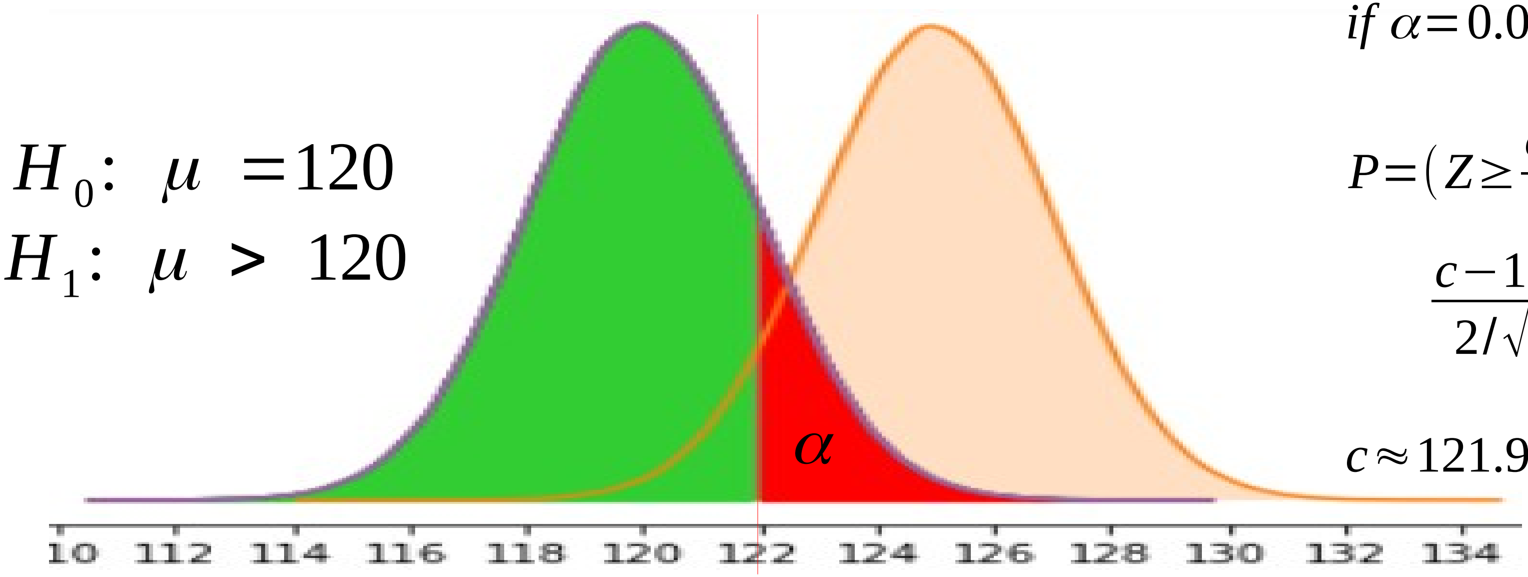
$n=3$

$$\overline{X}=\frac{\left(x_1+x_2+x_3\right)}{3}$$

Errors in Hypothesis Testing

Type I Error(α): Conditional probability of false rejection of Null Hypotesis when it is true

$$Type\ I\ Error\ =\ \alpha\ =\ P\left(Rejecting\ Null\ Hypothesis\ \middle|\ H_0\ \ is\ \ true\right)$$



$$H_0:\ \mu\ =120$$
$$H_1:\ \mu\ >\ 120$$

$$if\ \alpha=0.05$$

$$P=\left(Z\geq\frac{c-120}{2/\sqrt{3}}\right)=0.05$$

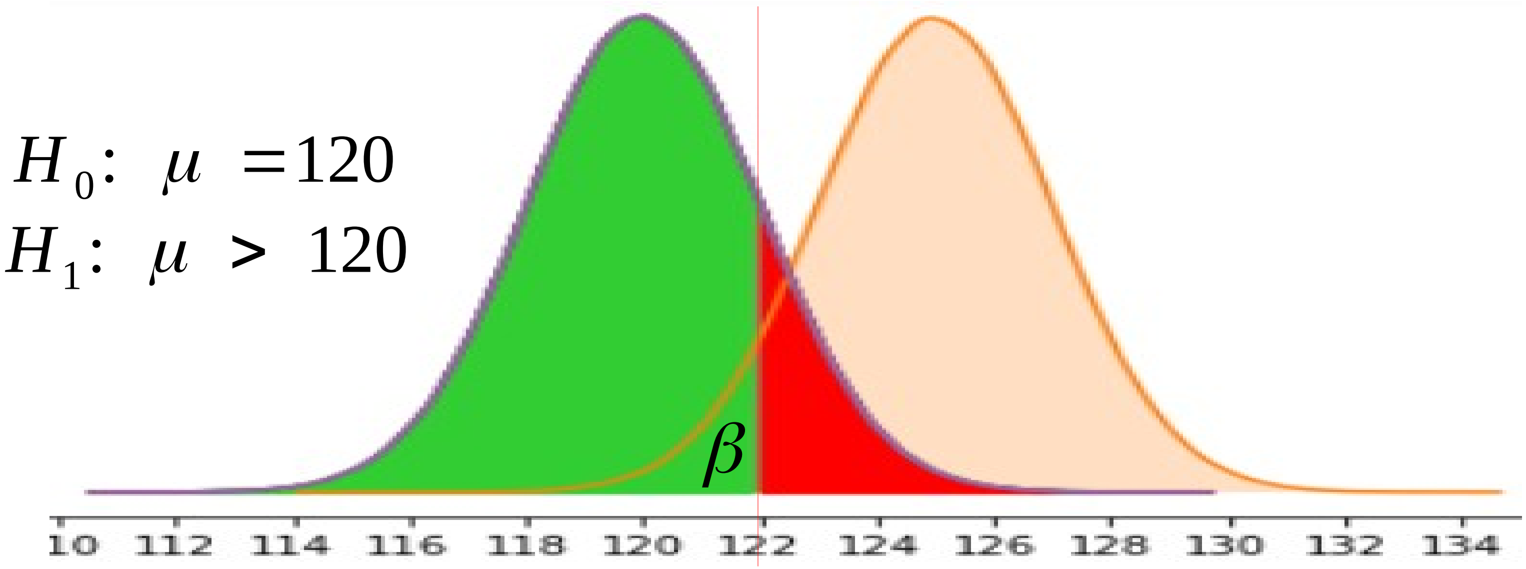
$$\frac{c-120}{2/\sqrt{3}}=1.645$$

$$c\approx121.9$$

Errors in Hypothesis Testing

Type II Error(β): Conditional probability of failing to reject Null Hypotesis

$$\mathit{Type\ II\ Error} = \beta = P\left(\mathit{Failed}\quad t\,0\quad \mathit{Reject\ Null\ Hypothesis} \mid H_0\ \ \mathit{is}\ \ \mathit{false}\right)$$



$$H_0:\ \mu = 120$$

$$H_1:\ \mu > 120$$

Why we should test Null Hypothesis?

The null hypothesis is based on evidence from past records , or that which is genereally assumed true

We know the distribution of our test variable under the null hypothesis

In our example mean 120 and standard deviation 2

$$H_0: \ \mu \ = 120$$
$$H_1: \ \mu \ > \ 120$$

The null hypothesis—that the two groups are drawn from the same distribution. If both sets are drawn from the same distribution, they’ll have the same expected value

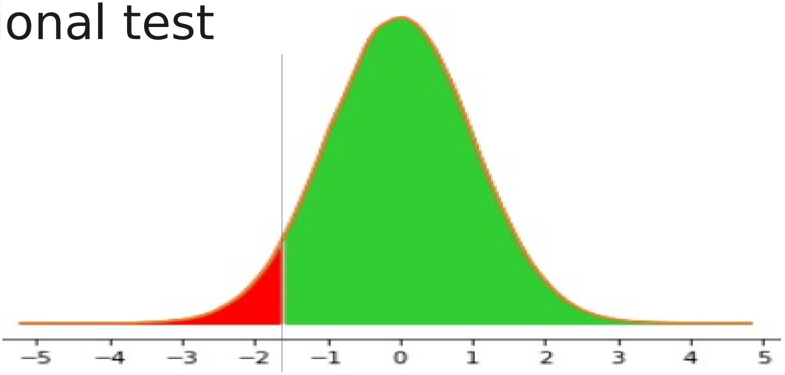
One- and two-tailed tests

types of hypothesis tests

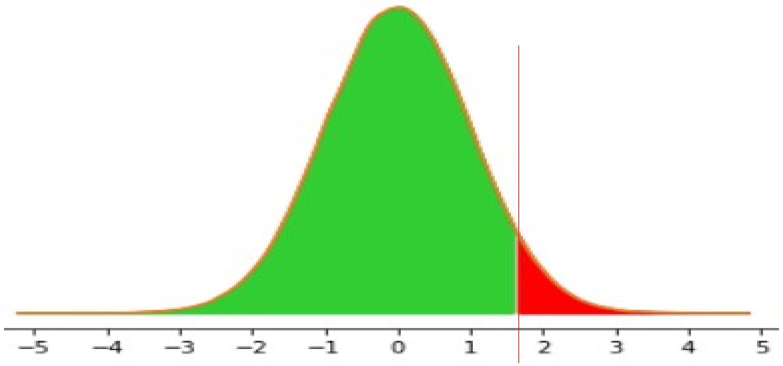
- One - tailed test
- Two - tailed test

A one-tailed test is a statistical test in which the rejection region will be on one side of a distribution

It is directional test



$Z = -1.645$



$Z = 1.645$

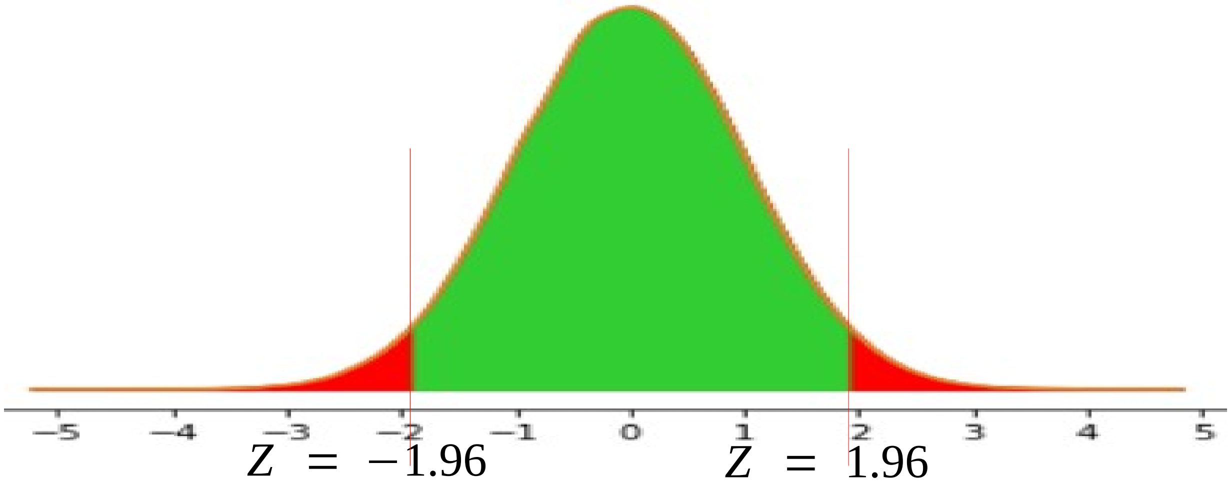
One- and two-tailed tests

types of hypothesis tests

- One - tailed test
- Two - tailed test

A Two-tailed test is a statistical test in which the rejection region will be on two sides of a distribution

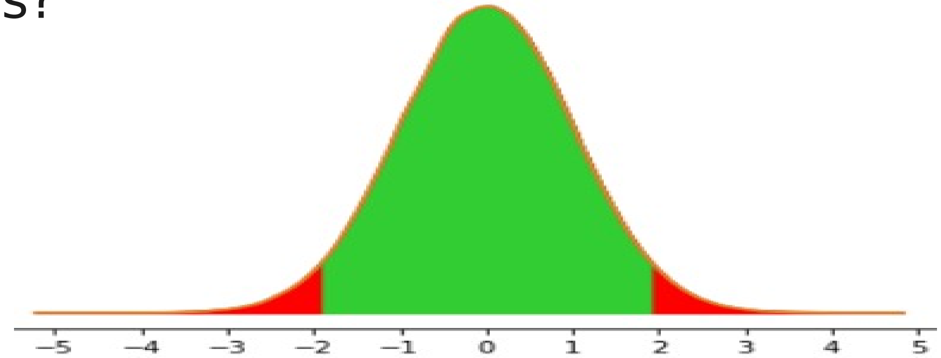
It is non directional test



two-tailed test

The school principal wants to test if it is true what teachers say – that high school juniors use the computer an average 3.2 hours a day. What are our null and alternative hypotheses?

$$H_0: \ \mu \ = \ 3.2$$
$$H_1: \ \mu \ \neq \ 3.2$$



Null hypothesis states that the population has a mean equal to 3.2 hours.

Alternative hypothesis states that the population has a mean that differs from 3.2 hours.

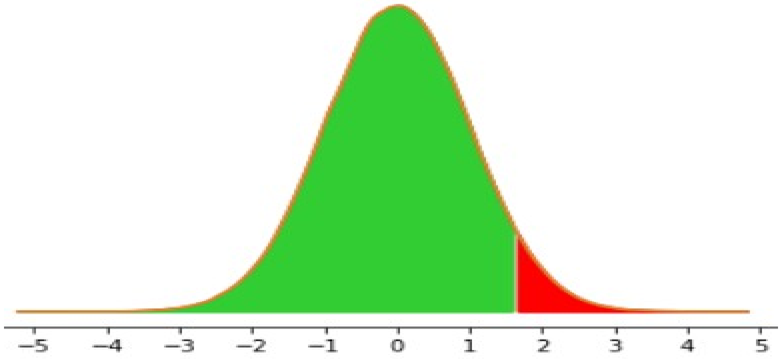
Right-Tailed Test

When the direction of the results is anticipated or we are only interested in one direction of the results.

A researcher claims that Indian men are, on average, more than 10 kgs heavier than Indian women, women average is 55 kgs. What is the null hypothesis, and what kind of test is this?

$$H_0:\ \mu\ \leq\ 65$$

$$H_1:\ \mu\ >\ 65$$



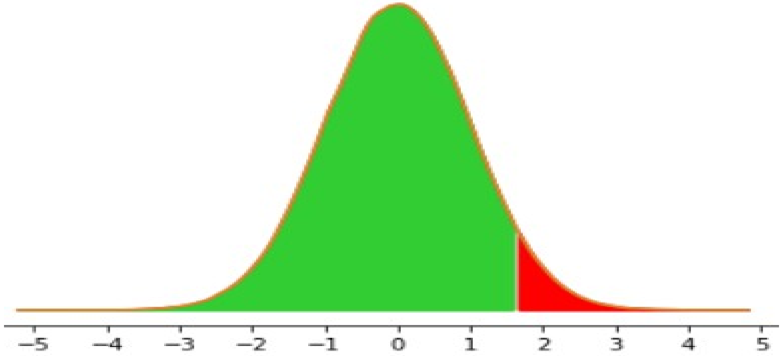
This is a right-tailed test, since rejection region will be in the right tail of the graph. Recognize that values above 65 would indicate that the null hypothesis be rejected.

Right-Tailed Test

A Deodorant brand claims their deodorant fragrance stays more than 24 hours

$$H_0 : \mu \leq 24$$

$$H_1 : \mu > 24$$



Left-Tailed Test

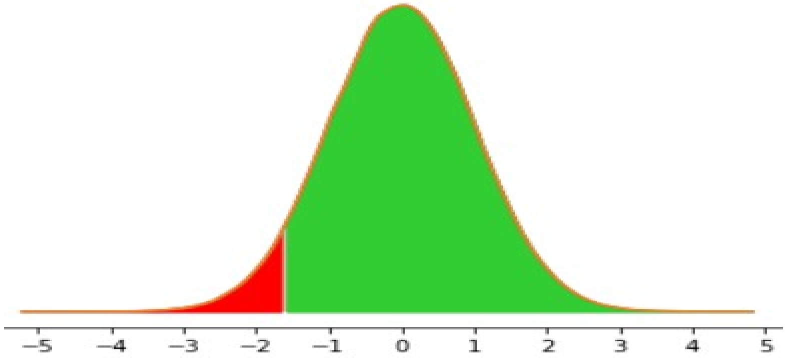
Suppose that in a particular geographic region in a reading test the mean and standard deviation of scores are 100 points, and 12 points, respectively.

In a particular school The mean score of 55 students is 96.

Is this mean score is significantly lower than the regional mean?

$$H_0: \ \mu \ \geq \ 100$$
$$H_1: \ \mu \ < \ 100$$

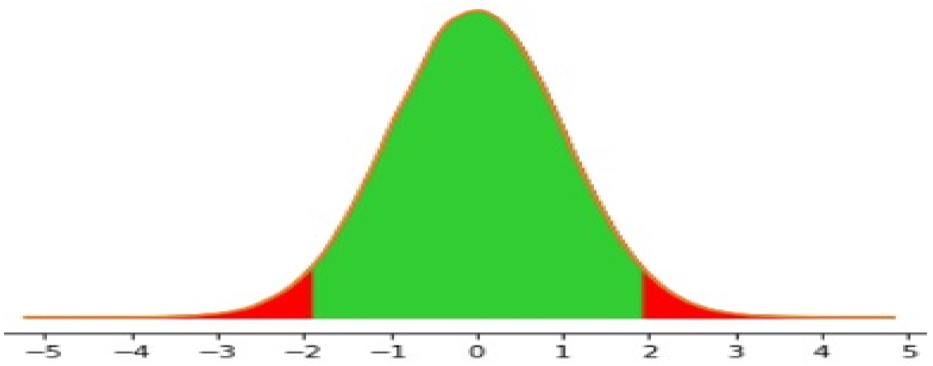
This is a left-tailed test



Two-Tailed Test

The average salary of Computer Science and Electronics stsudents is different

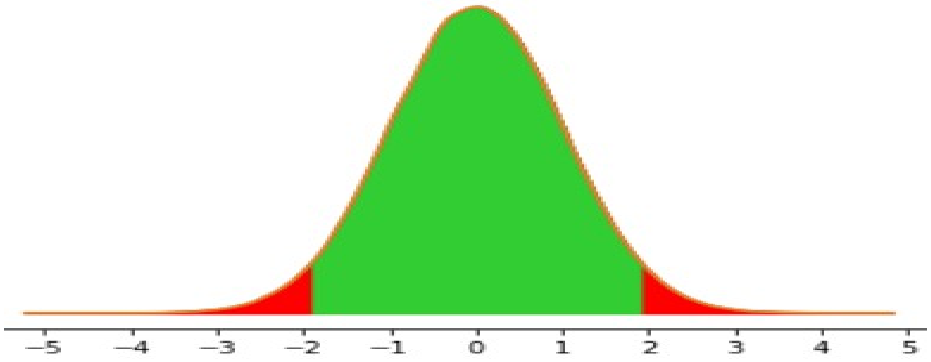
$$H_0 : \ \mu_{CSE} \ = \ \mu_{ECE}$$
$$H_1 : \ \mu_{CSE} \ \neq \ \mu_{ECE}$$



Two-Tailed Test

More men experience heart attacks than women

$$H_0: \; \mu_m = \mu_f$$
$$H_1: \; \mu_m \neq \mu_f$$



Z test

High P-Values: Your data are likely with a true null

Low P-Values: Your data are unlikely with a true null

Significance Level – It is the probability of making type I error and is denoted by α

Test Statistics measure how close the sample has come to the null hypothesis. This observation differs from a random sample to a sample. A test statistic results contain insights about the data that helps in making the decision of whether to reject the null hypothesis or not.

There are different probability models for different types of populations. Based on probability distribution different hypothesis tests are selected.

Z test

Sample data is like a mirror image for the population. So, the sample data must provide sufficient evidence to reject the null hypothesis and conclude that the effect exists in the population. If it is not able to do so then effect doesn’t exist in the population and thus we would fail to reject the null hypothesis.

Formulate the Null and Alternate Hypothesis

Based on data and probability distribution select the hypothesis test to be performed

Based on the business are and problem statement selects the level of significance if 0.05 (standard alpha) is not acceptable.

Calculate test statistics on the sample data collected

Calculate the p-value

Based on p-value draw insights to reject or fail to reject the null hypothesis.

Draw your business conclusion.

Popular Hypothesis Tests

Hypothesis test		Assumptions or notes
Z tests	One-sample Z test	(Normal population **or** n large) **and** σ *known*
	Two-sample Z test	Normal population **and** independent observations **and** $\sigma_1 \wedge \sigma_2$ *known*
T tests	One-sample t test	Normal population **and** σ *unknown*
	Paired t test	Normal population of differences **and** σ *unknown*
	Two-sample t test with equal variances	Normal populations **and** independent observations **and** $\sigma_1 = \sigma_2$ *unknown*
	Two-sample t test with unequal variances	Normal populations **and** independent observations **and** $\sigma_1 \neq \sigma_2$ *both unknown*
Chi-squared test	Chi-squared goodness fit test	Will be applied on categorical variables **and** All expected counts are atleast 5
F test	ANOVA	Normal Population

Z test

Z test

A z-test is a statistical test to determine whether two population means are different when the variances are known and the sample size is large.

One sample Z test

The One Sample Z Test compares a sample mean to a hypothesized population mean to determine whether the two means are significantly different

Two sample Z test

In many cases we would like to compare parameters of two different populations to check for any difference in parameter(mean) values

The purpose of the test is to determine whether the difference between these two populations is statistically significant.

Hypothesis Testing with Statistical Significance

The process for Hypothesis testing

1. Formulate Null Hypothesis and Alternate Hypothesis

2. Choose level of significance(α)

3. Determine appropriate test to use and find the test statistic

4. Determine if we need a 1 tailed or a 2 tailed Z-critical value and find the Z-critical value for desired confidence level

5. Compare test statistic with Z-critical value for desired confidence level

Or

Find the p value for test statistic compare with significance level(α)

One Sample Z test

One Sample Z test Assumptions

1. The population standard deviation(σ) is known

2. The population follow the normal probability distribution or The sample size is large(n>30)

One sample Z-test

Suppose that in a particular geographic region in a reading test the mean and standard deviation of scores are 100 points, and 12 points, respectively.

In a particular school The mean score of 55 students is 96.

We want to test whether this mean score is different from the regional mean?

that is, are the students in this school having same scores as the region students

or

either their scores lower than regional mean or their scores greater than regional mean?

One sample Z-test

One sample Z-tests can be performed

1. Here we know the population Standard Deviation(σ)
2. sample size is large(>30)

Here

$$Population\,Standard\;Deviation\left(\,\sigma\right)=12$$
$$SampleSize\left(n\right)=55$$
$$Sample\;Mean\left(\overline{x}\right)=96$$
$$Population\,Mean\left(\mu_0\right)=100$$

One sample Z-test

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0: \ \mu \ = \ 100\left(\mu_0\right)$$
$$H_1: \ \mu \ \neq \ 100\left(\mu_0\right)$$

Step 2: Choose level of significance(α)

α = 0.05

A value of α = 0.05 implies that the null hypothesis is rejected 5 % of the time when it is in fact true

One sample Z-test

Step 3: Determine appropriate test to use and find the test statistic

Calculate the z-statistic

$$Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

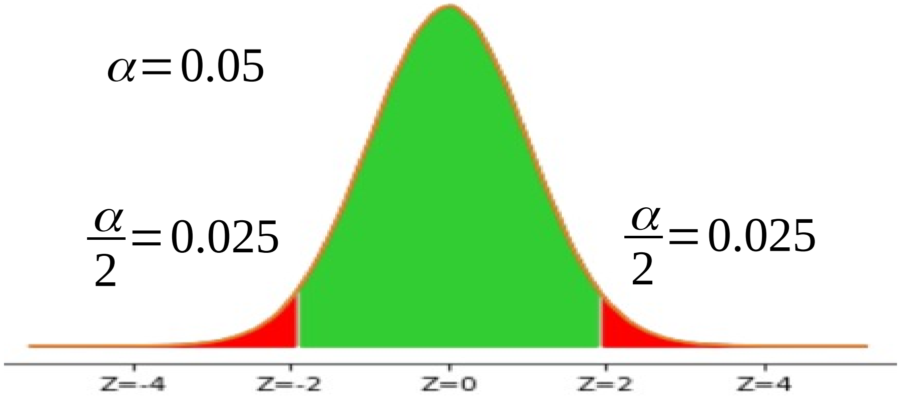
$$Z = \frac{96 - 100}{\frac{12}{\sqrt{55}}} = \frac{-4}{1.62} = -2.47$$

The sample mean score is 96, which is −2.47 standard error units from the population mean of 100

One sample Z-test Two-tailed

Step 4: Determine if we need a 1 tailed or a 2 tailed Z-critical value and find the Z-critical value for desired confidence level

$$H_0: \ \mu \ = \ 100\left(\mu_0\right)$$
$$H_1: \ \mu \ \neq \ 100\left(\mu_0\right)$$



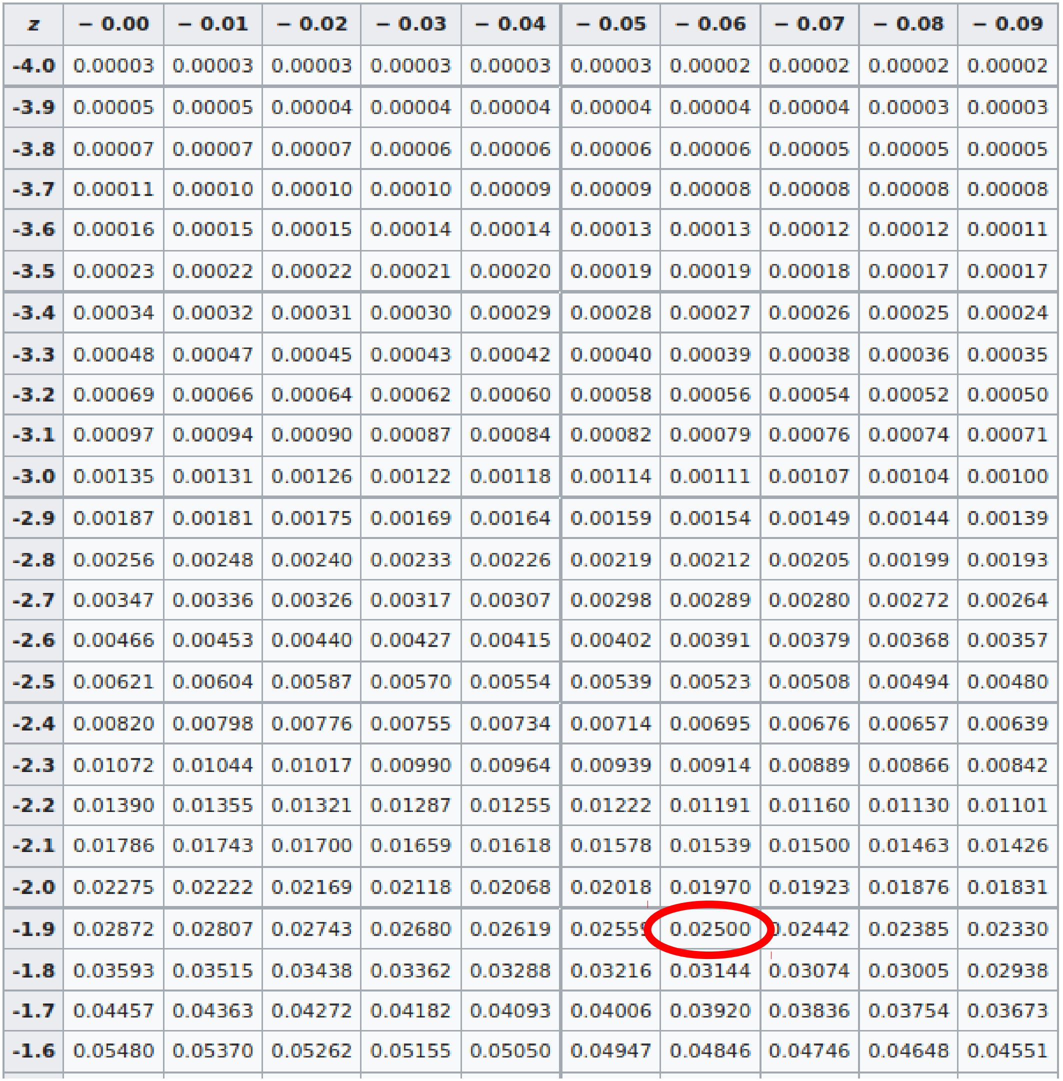
$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

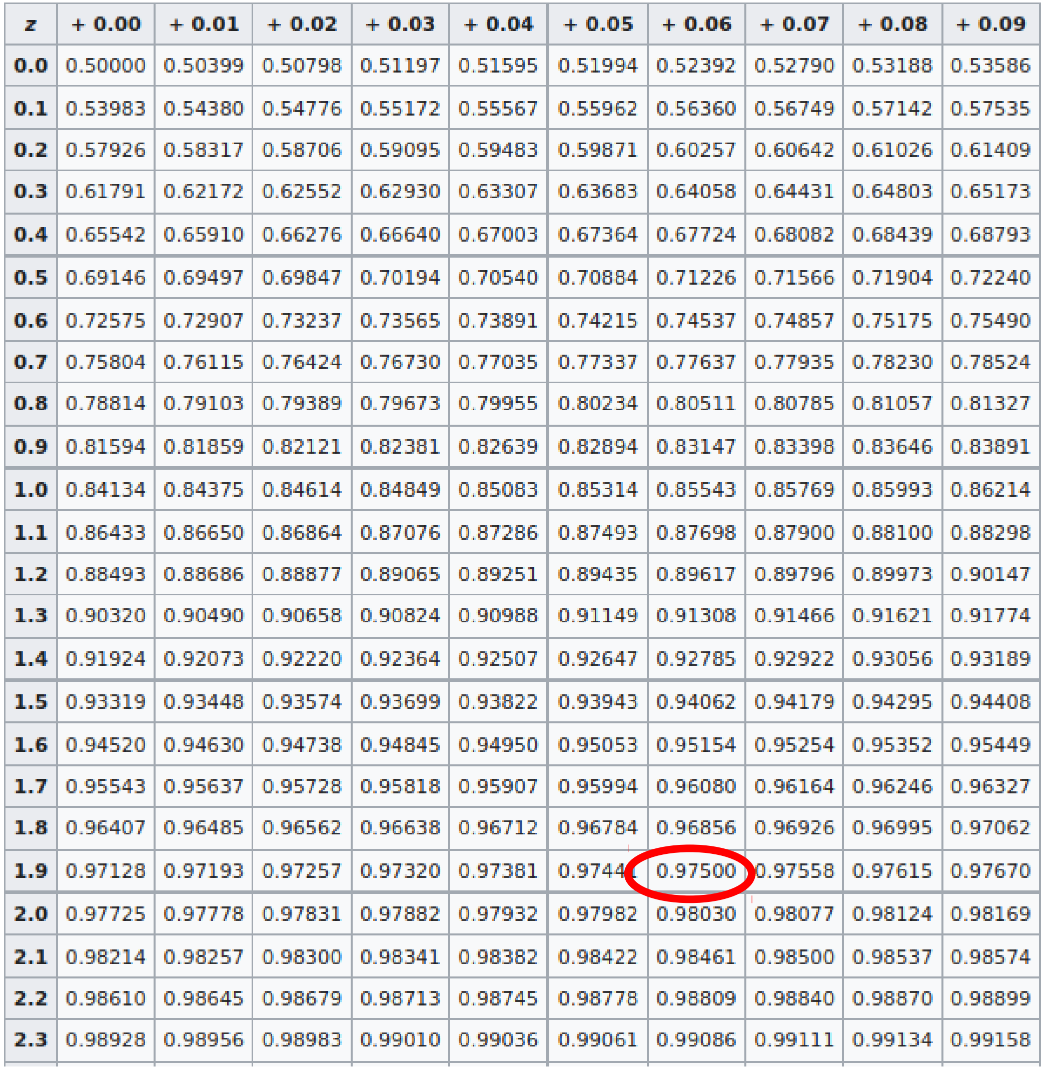
$\frac{\alpha}{2} = 0.025$

Z critical value for two tailed test for 95% confidence is +/-1.96

Cumilative probabilities - statistic is less than Z (i.e. between negative infinity and Z)

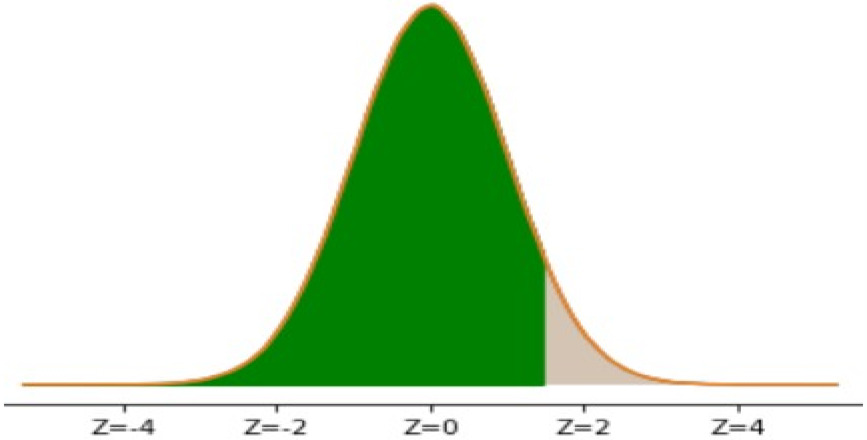


z	− 0.00	− 0.01	− 0.02	− 0.03	− 0.04	− 0.05	− 0.06	− 0.07	− 0.08	− 0.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551

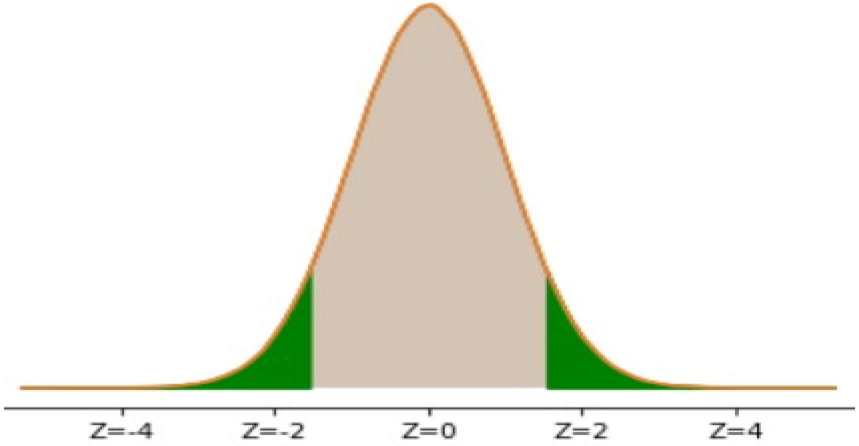


z	+ 0.00	+ 0.01	+ 0.02	+ 0.03	+ 0.04	+ 0.05	+ 0.06	+ 0.07	+ 0.08	+ 0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158

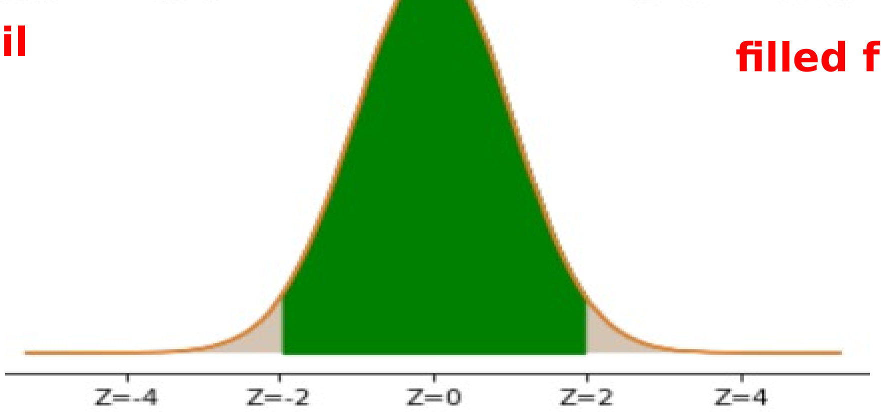
Z Tables



filled from left tail



filled from both sides



filled from the middle

Z critical Values

	Two-tailed	One-tailed
$Alpha(\alpha)$	$Z_{\alpha/2}$	Z_{α}
0.01	2.575	2.33
0.05	1.96	1.645
0.10	1.645	1.28

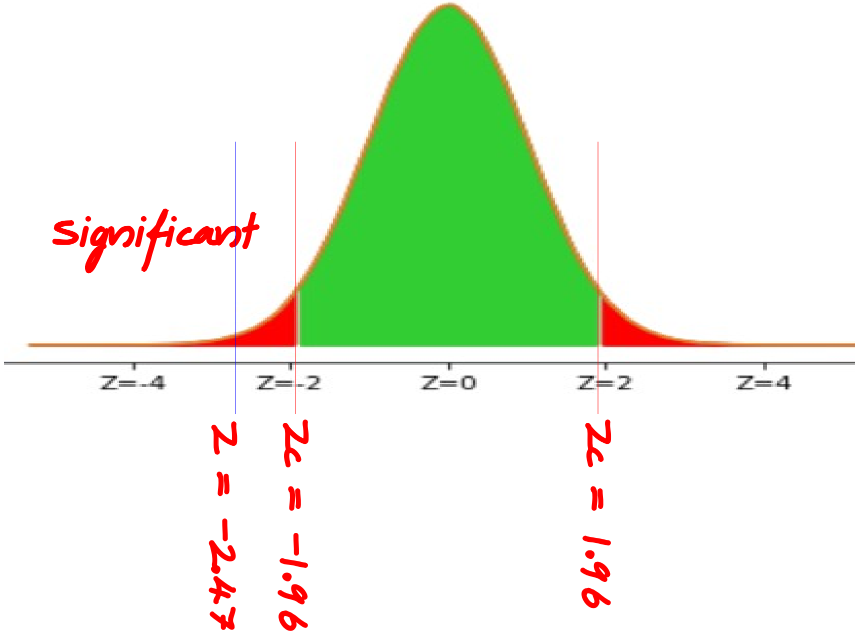
One sample Z-test Two-tailed

Step 5: Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

Z-statistic(-2.47) < Z critical value(-1.96)

Z statistic is falls in the rejection region

$$H_0: \ \mu \ = \ 100(\mu_0)$$
$$H_1: \ \mu \ \neq \ 100(\mu_0)$$



Result is statistically significant so we need to reject Null Hypothesis

One sample Z-test Two-tailed

Step 5: Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

The probability of observing a standard normal value below −2.47 is approximately is 0.0068(1 tailed p value)

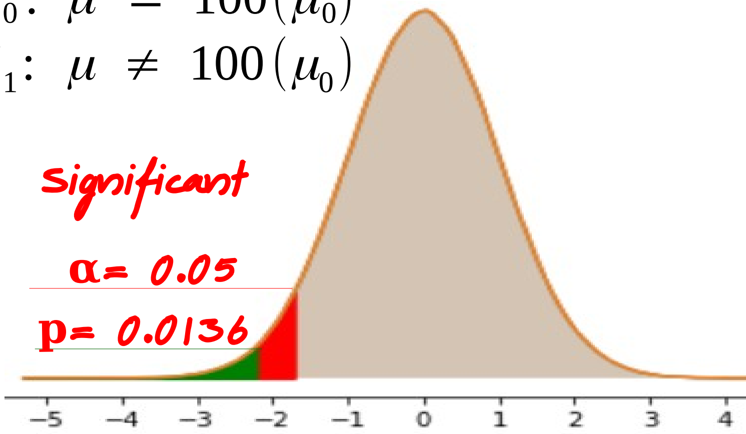
2 tailed p= 0.0136 = 2* 0.0068

Significance Level(α) = 0.05

Since p(0.0068)< α(0.05)

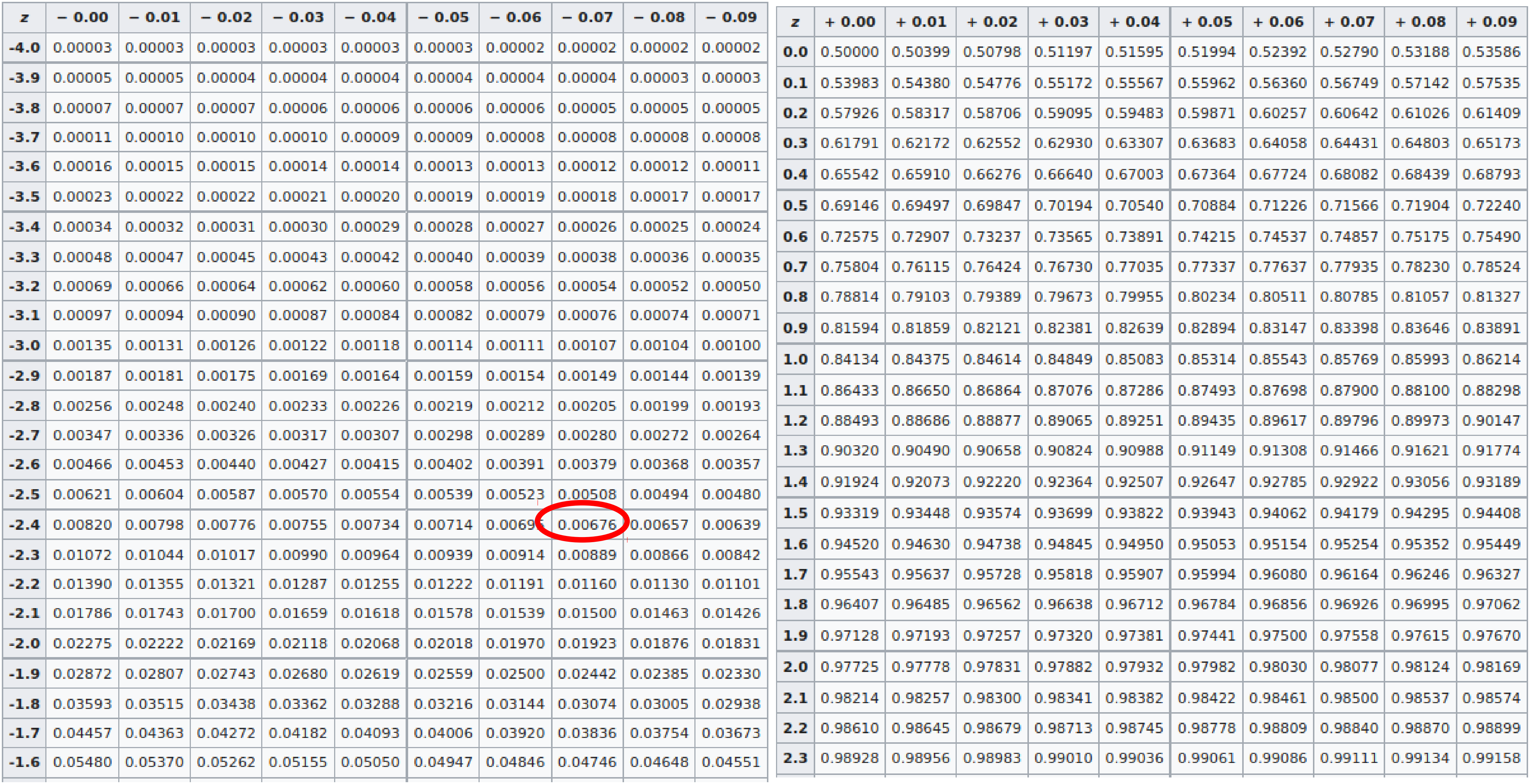
$$H_0: \ \mu \ = \ 100(\mu_0)$$
$$H_1: \ \mu \ \neq \ 100(\mu_0)$$

Means result is significant so we need to reject Null Hypothesis



The school students have different mean test score from region mean score

Cumilative probabilities - statistic is less than Z (i.e. between negative infinity and Z)



z	− 0.00	− 0.01	− 0.02	− 0.03	− 0.04	− 0.05	− 0.06	− 0.07	− 0.08	− 0.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551

z	+ 0.00	+ 0.01	+ 0.02	+ 0.03	+ 0.04	+ 0.05	+ 0.06	+ 0.07	+ 0.08	+ 0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158

One sample Z-test Two-tailed with Confidence Intervals

$$n = 55, \overline{X} = 96, \sigma = 12, \alpha = 0.05$$

$$H_0:\ \mu\ =\ 100\left(\mu_0\right)$$
$$H_1:\ \mu\ \neq\ 100\left(\mu_0\right)$$

$$CI = \overline{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$CI = 96 \pm 1.96 \frac{12}{\sqrt{55}}$$
$$CI = 96 \pm 1.96 * 1.62$$
$$CI = 96 \pm 3.1752$$
$$CI = \left[92.8248, 99.1752\right]$$

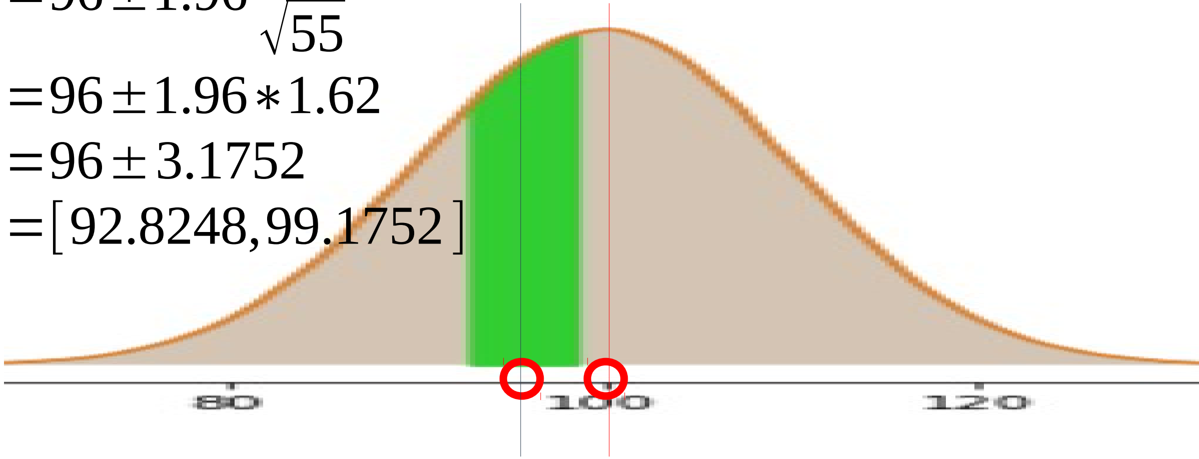
One sample Z-test Two-tailed with Confidence Intervals

$$n=55, \overline{X}=96, \sigma=12, \alpha=0.05$$

$$H_0: \ \mu \ = \ 100\left(\mu_0\right)$$
$$H_1: \ \mu \ \neq \ 100\left(\mu_0\right)$$

$$CI=\overline{X}\pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$CI=96\pm 1.96\frac{12}{\sqrt{55}}$$
$$CI=96\pm 1.96*1.62$$
$$CI=96\pm 3.1752$$
$$CI=\left[92.8248, 99.1752\right]$$



100 is not in the confidence interval [92.82,99.17] range

there is statistical difference so null hypothesis is not true

One sample Z-test Two-tailed

A study claims that Indian smartphone users spend three and half hours a day on their smartphones. Try to verify this with following data

$$Population\, Standard\ Deviation(\sigma)=2\, hours$$

$$SampleSize\,(n)=121$$

$$Sample\ Mean(\overline{x})=3.7\, hours$$

$$Significance\ Level(\alpha)=0.01$$

One sample Z-test Two-tailed

Here

$$
\begin{aligned}
&Population\,Standard\ Deviation(\sigma)=2\,hours\\
&SampleSize(n)=121\\
&Sample\ Mean(\bar{x})=3.7\,hours\\
&Population\ Mean(\mu_0)=3.5\,hours\\
&Significance\ Level(\alpha)=0.01
\end{aligned}
$$

One sample Z-test Two-tailed

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0: \ \mu \ = \ 3.5(\mu_0)$$
$$H_1: \ \mu \ \neq \ 3.5(\mu_0)$$

Step 2: Choose level of significance(α)

α = 0.01

A value of α = 0.01 implies that the null hypothesis is rejected 1 % of the time when it is in fact true

One sample Z-test Two-tailed

Step 3: Determine appropriate test to use and find the test statistic

One sample Z-tests can be performed

1. Here we know the population Standard Deviation(σ)

2. sample size is large(>30)

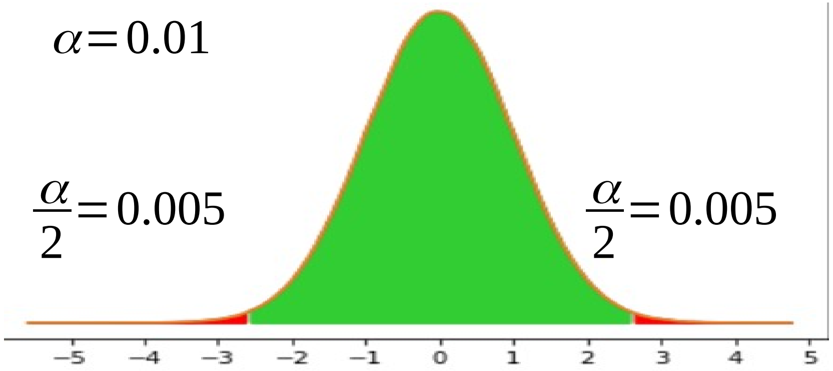
Calculate the z-statistic $Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ $Z = \frac{3.7 - 3.5}{\frac{2}{\sqrt{121}}} = \frac{0.2}{0.18} = 1.1$

The sample mean is 3.7, which is 1.1 standard error units from the population mean of 3.5

One sample Z-test Two-tailed

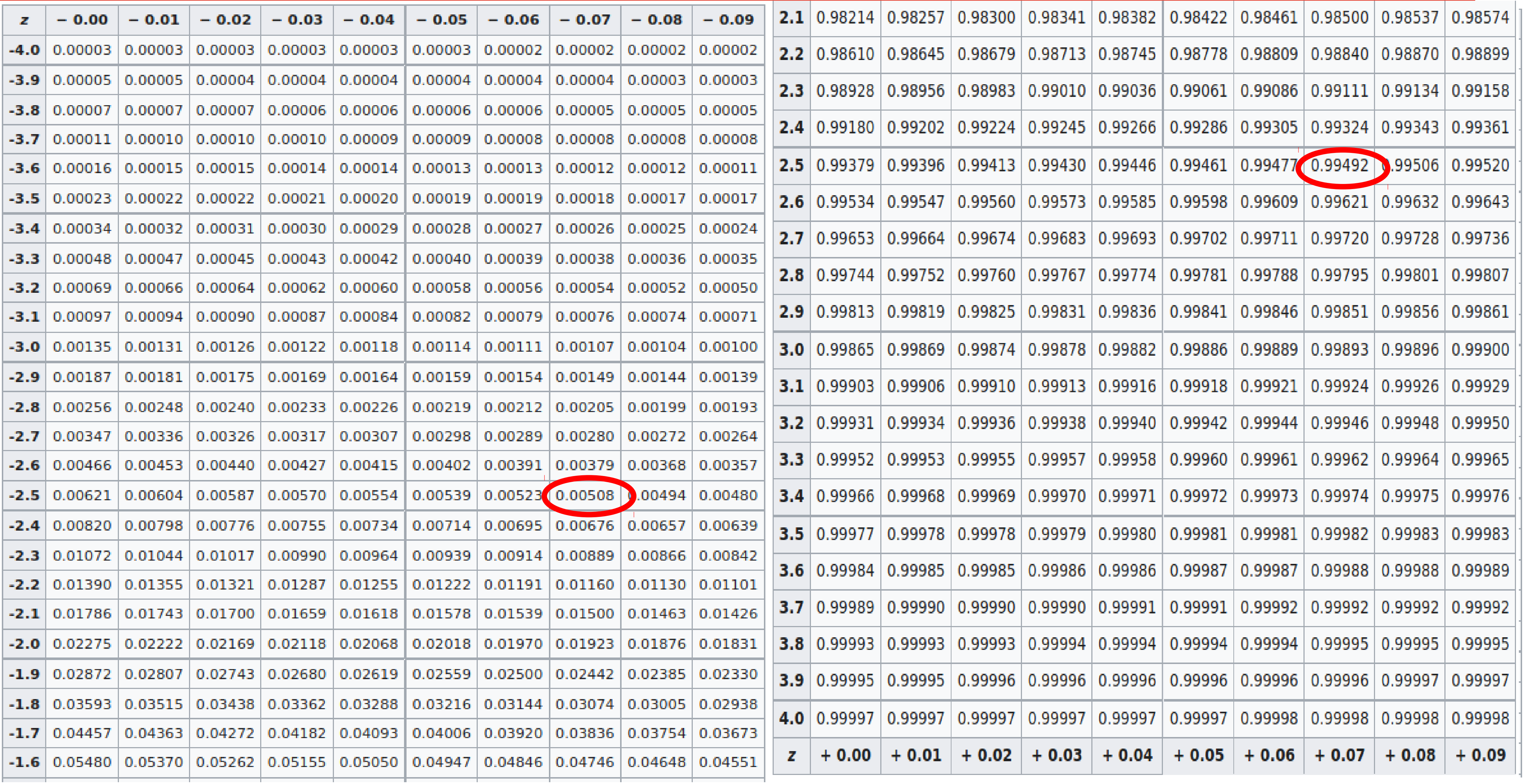
Step 4: Determine if we need a 1 tailed or a 2 tailed Z-critical value and find the Z-critical value for desired confidence level

$$H_0: \ \mu \ = \ 3.5\left(\mu_0\right)$$
$$H_1: \ \mu \ \neq \ 3.5\left(\mu_0\right)$$



Z critical value for two tailed test for 99% confidence is +/-2.575

Cumilative probabilities - statistic is less than Z (i.e. between negative infinity and Z)



z	− 0.00	− 0.01	− 0.02	− 0.03	− 0.04	− 0.05	− 0.06	− 0.07	− 0.08	− 0.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551

2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998
z	+ 0.00	+ 0.01	+ 0.02	+ 0.03	+ 0.04	+ 0.05	+ 0.06	+ 0.07	+ 0.08	+ 0.09

Z critical Values

	Two-tailed	One-tailed
$Alpha(\alpha)$	$Z_{\alpha/2}$	Z_{α}
0.01	2.575	2.33
0.05	1.96	1.645
0.10	1.645	1.28

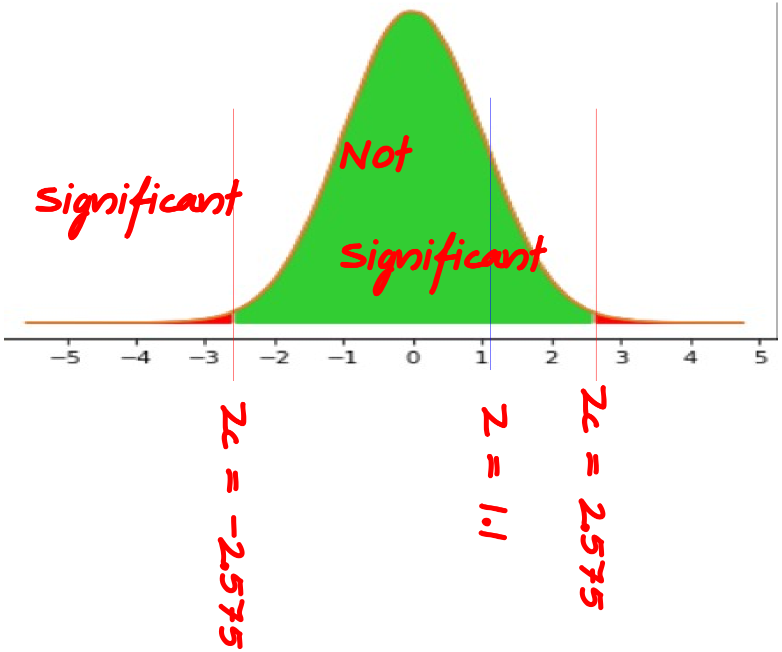
One sample Z-test Two-tailed

Step 5: Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

Z critical value(-2.575)< Z-statistic(1.1) < Z critical value(2.575)

Z statistic do not falls in the rejection region

$$H_0:\ \mu\ =\ 3.5(\mu_0)$$
$$H_1:\ \mu\ \neq\ 3.5(\mu_0)$$



Result is not statistically significant so we can’t reject Null Hypothesis

One sample Z-test Two-tailed

Step 5: Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

The probability of observing a standard normal value above 1.1 is approximately is 0.0135(1 tailed p value)

Significance Level(α) = 0.01

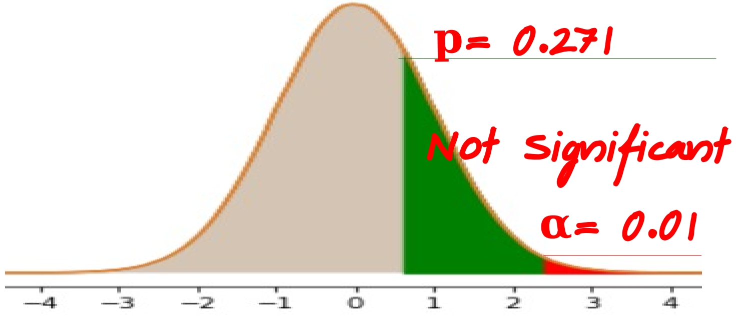
2 tailed p= 0.271 = 2*0.135

$$H_0: \ \mu \ = \ 3.5(\mu_0)$$

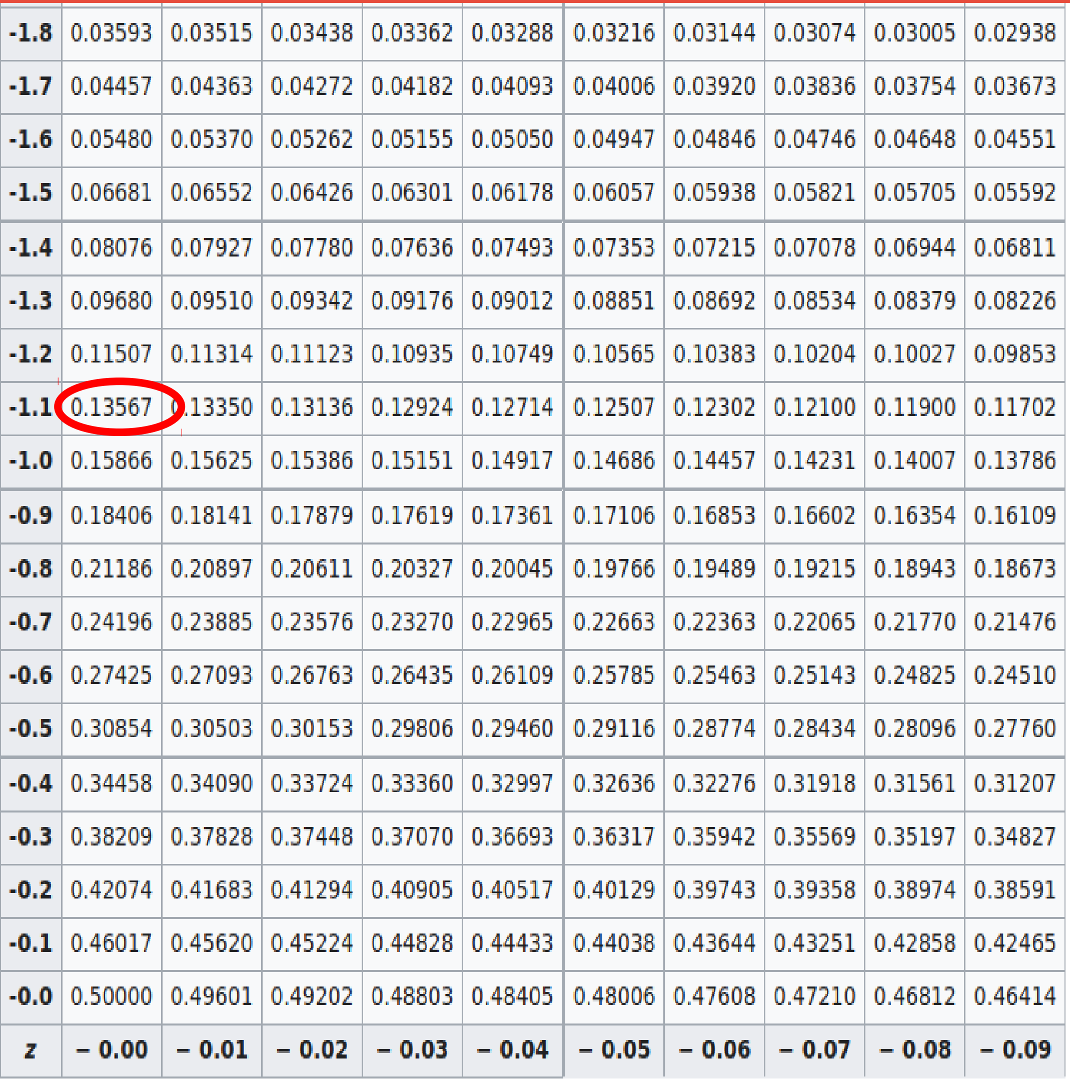
$$H_1: \ \mu \ \neq \ 3.5(\mu_0)$$

Since $p(0.27) \nless 0.01(\alpha)$

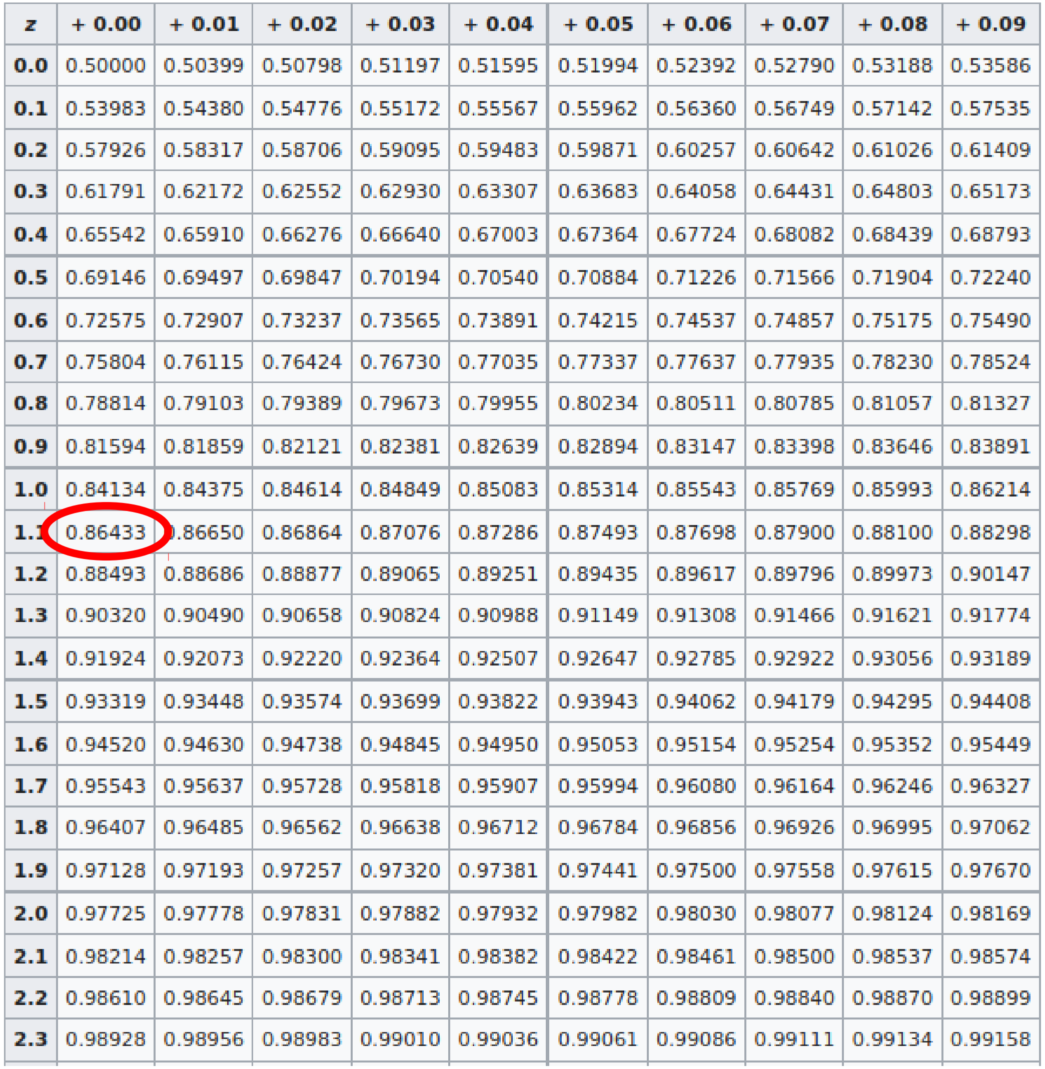
Means result is not significant so can't reject Null Hypothesis



Cumilative probabilities - statistic is less than Z (i.e. between negative infinity and Z)



-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
-1.5	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
-1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
-1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08692	0.08534	0.08379	0.08226
-1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09853
-1.1	0.13567	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
-1.0	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
-0.9	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
-0.8	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
-0.7	0.24196	0.23885	0.23576	0.23270	0.22965	0.22663	0.22363	0.22065	0.21770	0.21476
-0.6	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.24510
-0.5	0.30854	0.30503	0.30153	0.29806	0.29460	0.29116	0.28774	0.28434	0.28096	0.27760
-0.4	0.34458	0.34090	0.33724	0.33360	0.32997	0.32636	0.32276	0.31918	0.31561	0.31207
-0.3	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
-0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
-0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
-0.0	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.47210	0.46812	0.46414
z	**− 0.00**	**− 0.01**	**− 0.02**	**− 0.03**	**− 0.04**	**− 0.05**	**− 0.06**	**− 0.07**	**− 0.08**	**− 0.09**



z	**+ 0.00**	**+ 0.01**	**+ 0.02**	**+ 0.03**	**+ 0.04**	**+ 0.05**	**+ 0.06**	**+ 0.07**	**+ 0.08**	**+ 0.09**
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158

Two-tailed Z-test with Confidence Intervals

Confidance Interval:

$$CI = \overline{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

99% Confidence Interval:

$$CI = \overline{X} \pm 2.576 \frac{\sigma}{\sqrt{n}}$$

$$H_0 : \ \mu \ = \ 3.5(\mu_0)$$

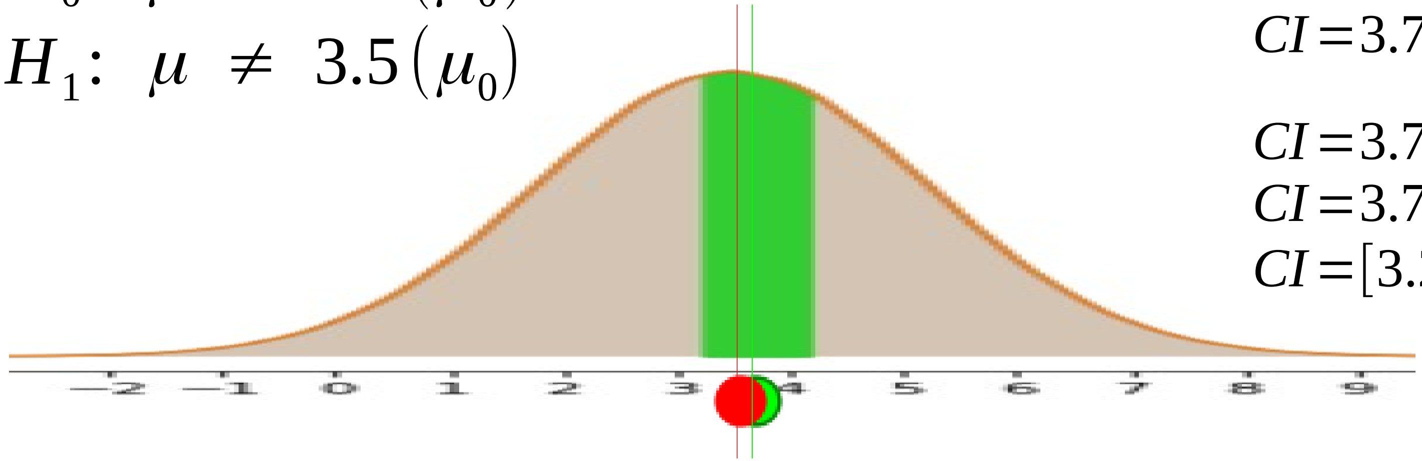
$$H_1 : \ \mu \ \neq \ 3.5(\mu_0)$$

$$CI = 3.7 \pm 2.575 \frac{2}{\sqrt{121}}$$

$$CI = 3.7 \pm 2.575 * 0.18$$

$$CI = 3.7 \pm 0.4635$$

$$CI = [3.23 \qquad 4.16]$$



Population Mean 3.5 in the confidence linterval range so Null Hypothesis can’t be rejected

Z critical Values

	Two-tailed	One-tailed
$Alpha(\alpha)$	$Z_{\alpha/2}$	Z_{α}
0.01	2.575	2.33
0.05	1.96	1.645
0.10	1.645	1.28

One Sample Z test

Suppose that in a particular geographic region in a reading test the mean and standard deviation of scores are 100 points, and 12 points, respectively.

In a particular school the mean score of 55 students is 96.

Can you determine with at least 95% confidence, are the students in this school comparable to the region students, or are their scores surprisingly low?

One sample Z-test

Here

$$Population\,Standard\ Deviation(\sigma)=12$$
$$SampleSize\,(n)=55$$
$$Sample\ Mean(\overline{x})=96$$
$$Population\ Mean(\mu_0)=100$$
$$Significance\ Level(\alpha)=0.05$$

One sample Z-test Left-tailed

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0: \ \mu \ \geq \ 100(\mu_0)$$
$$H_1: \ \mu \ < \ 100(\mu_0)$$

Step 2: Choose level of significance(α)

α = 0.05

A value of α = 0.05 implies that the null hypothesis is rejected 5 % of the time when it is in fact true

One sample Z-test Left-tailed

Step 3: Determine appropriate test to use and find the test statistic

One sample Z-tests can be performed

1. Here we know the population Standard Deviation(σ)

2. sample size is large(>30)

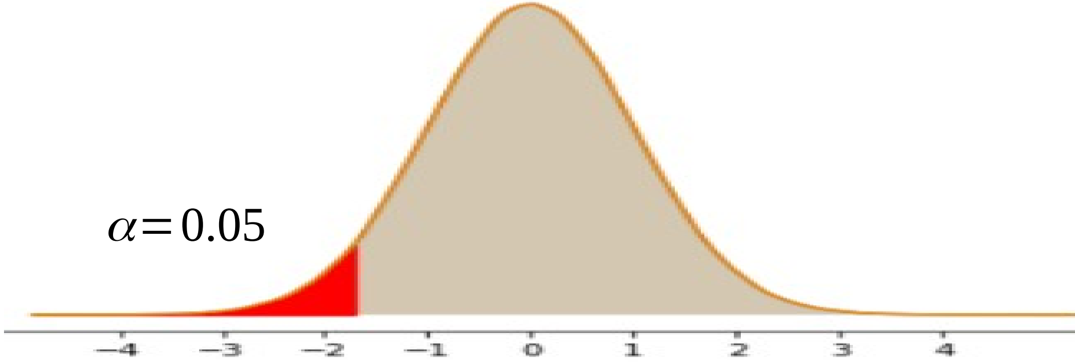
Calculate the z-statistic $Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ $Z = \frac{96 - 100}{\frac{12}{\sqrt{55}}} = \frac{-4}{1.62} = -2.47$

The class mean score is 96, which is −2.47 standard error units from the population mean of 100

One Sample Z test - Left tailed

Step 4: Determine if we need a 1 tailed or a 2 tailed Z-critical value and find the Z-critical value for desired confidence level

$$H_0: \ \mu \ \geq \ 100(\mu_0)$$
$$H_1: \ \mu \ < \ 100(\mu_0)$$



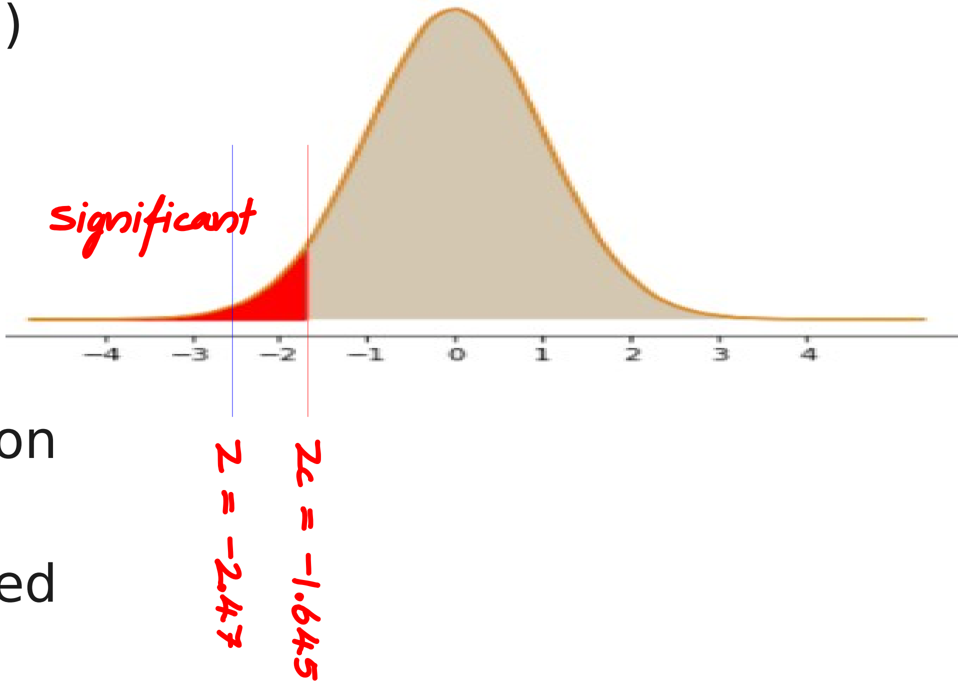
Z critical value for two tailed test for 95% confidence is -1.645

One Sample Z test – Left tailed

Step 5: Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

Z-statistic(-2.47) < Z critical value(-1.645)

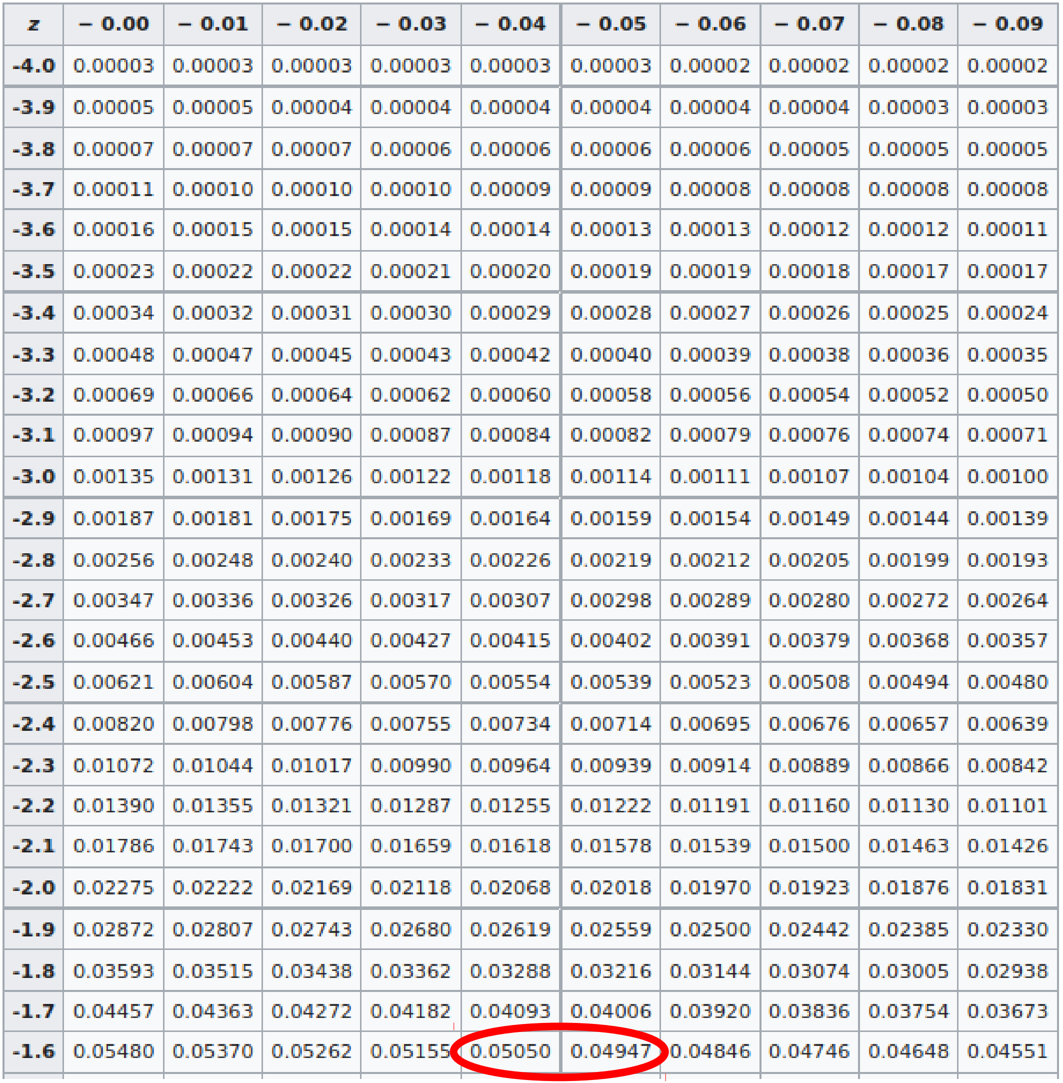
$$H_0 : \ \mu \ \geq \ 100(\mu_0)$$
$$H_1 : \ \mu \ < \ 100(\mu_0)$$



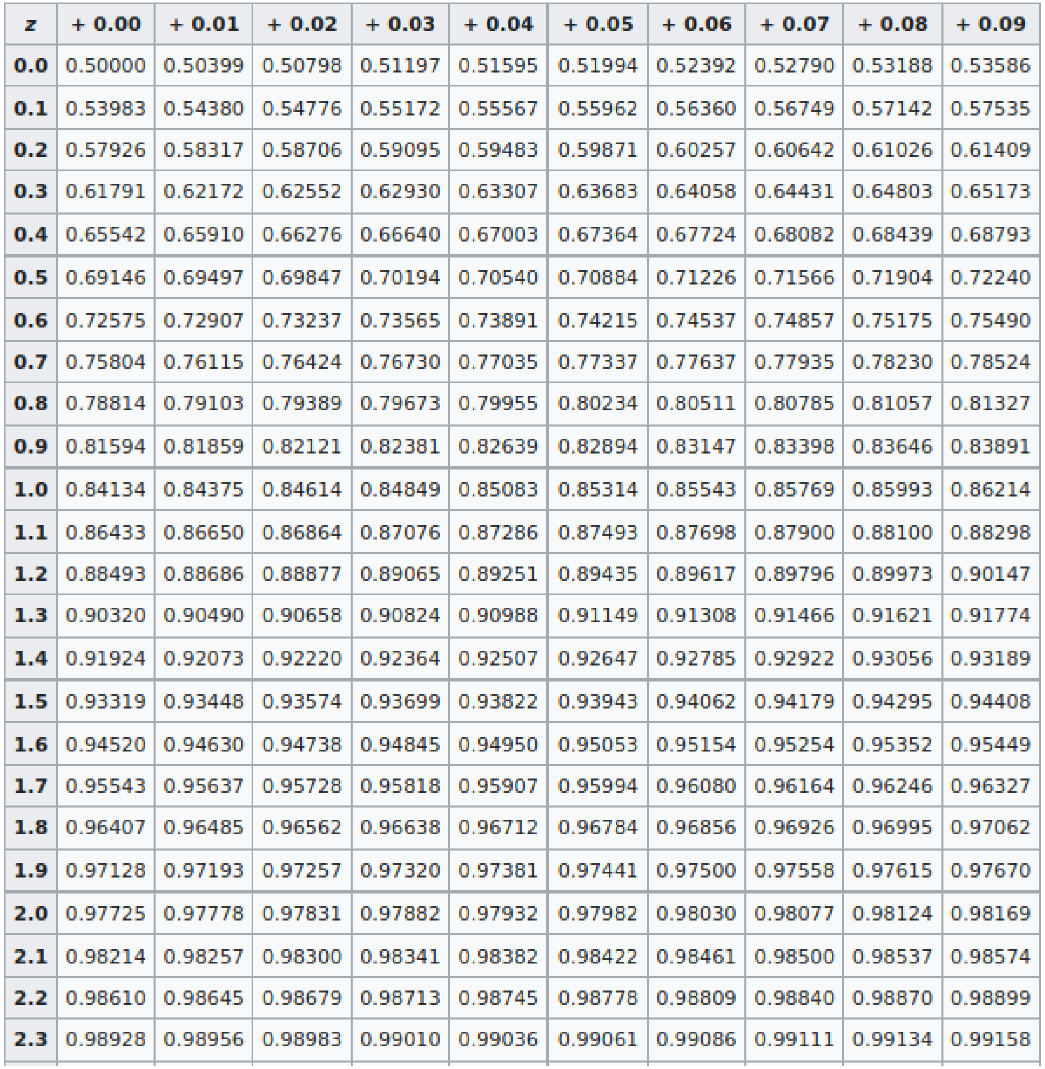
Z statistic is falls in the rejection region means it is very unlikely

Result is statistically significant so we need to reject Null Hypothesis

Cumilative probabilities - statistic is less than Z (i.e. between negative infinity and Z)



z	− 0.00	− 0.01	− 0.02	− 0.03	− 0.04	− 0.05	− 0.06	− 0.07	− 0.08	− 0.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551



z	+ 0.00	+ 0.01	+ 0.02	+ 0.03	+ 0.04	+ 0.05	+ 0.06	+ 0.07	+ 0.08	+ 0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158

One Sample Z test - Left tailed

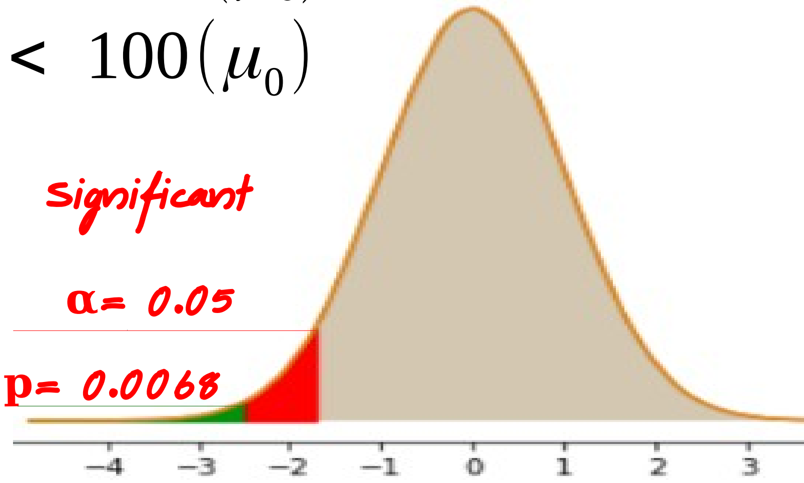
Step 5: Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

The probability of observing a standard normal value below −2.47 is approximately is 0.0068.

$$H_0: \ \mu \ \geq \ 100(\mu_0)$$

$$H_1: \ \mu \ < \ 100(\mu_0)$$

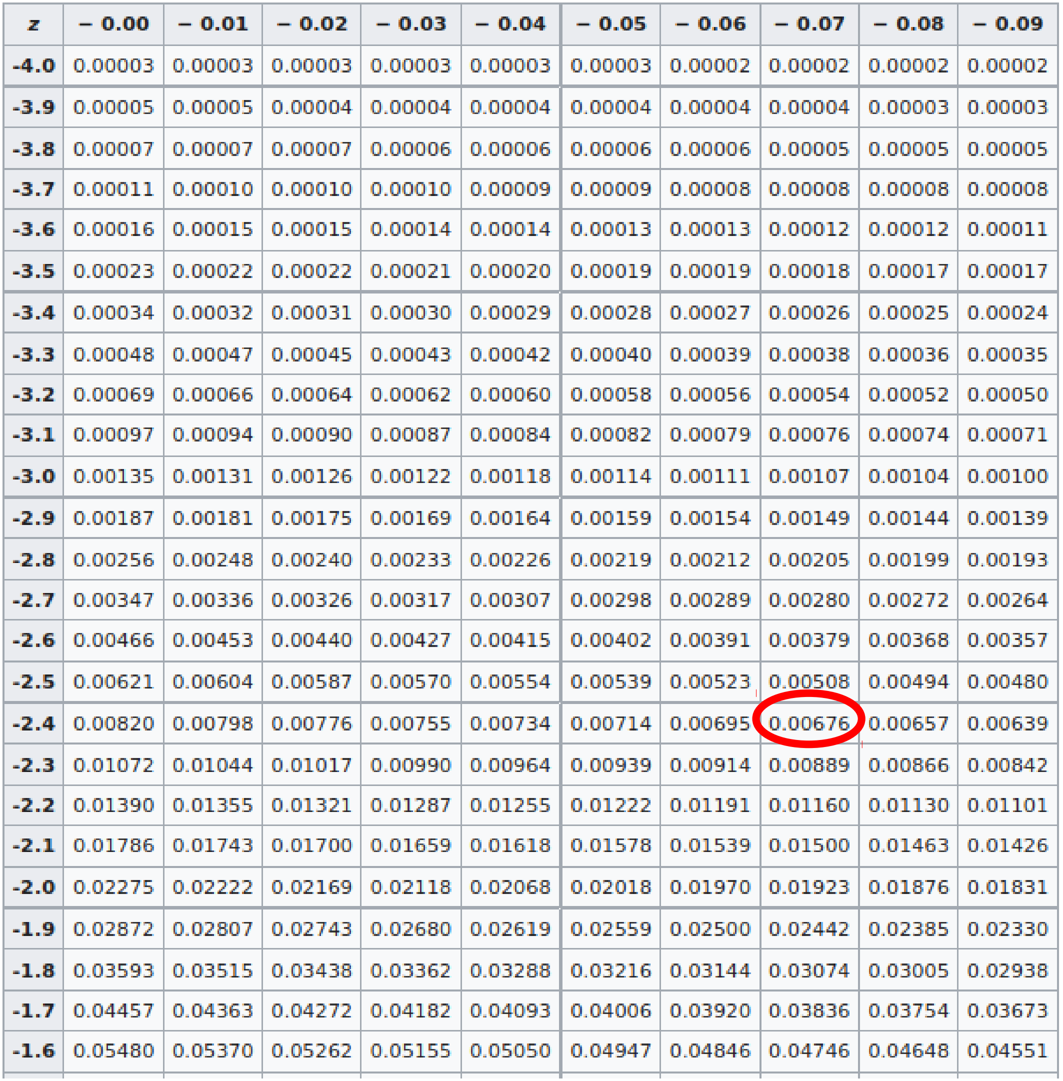
Since p(0.0068)< α(0.05)



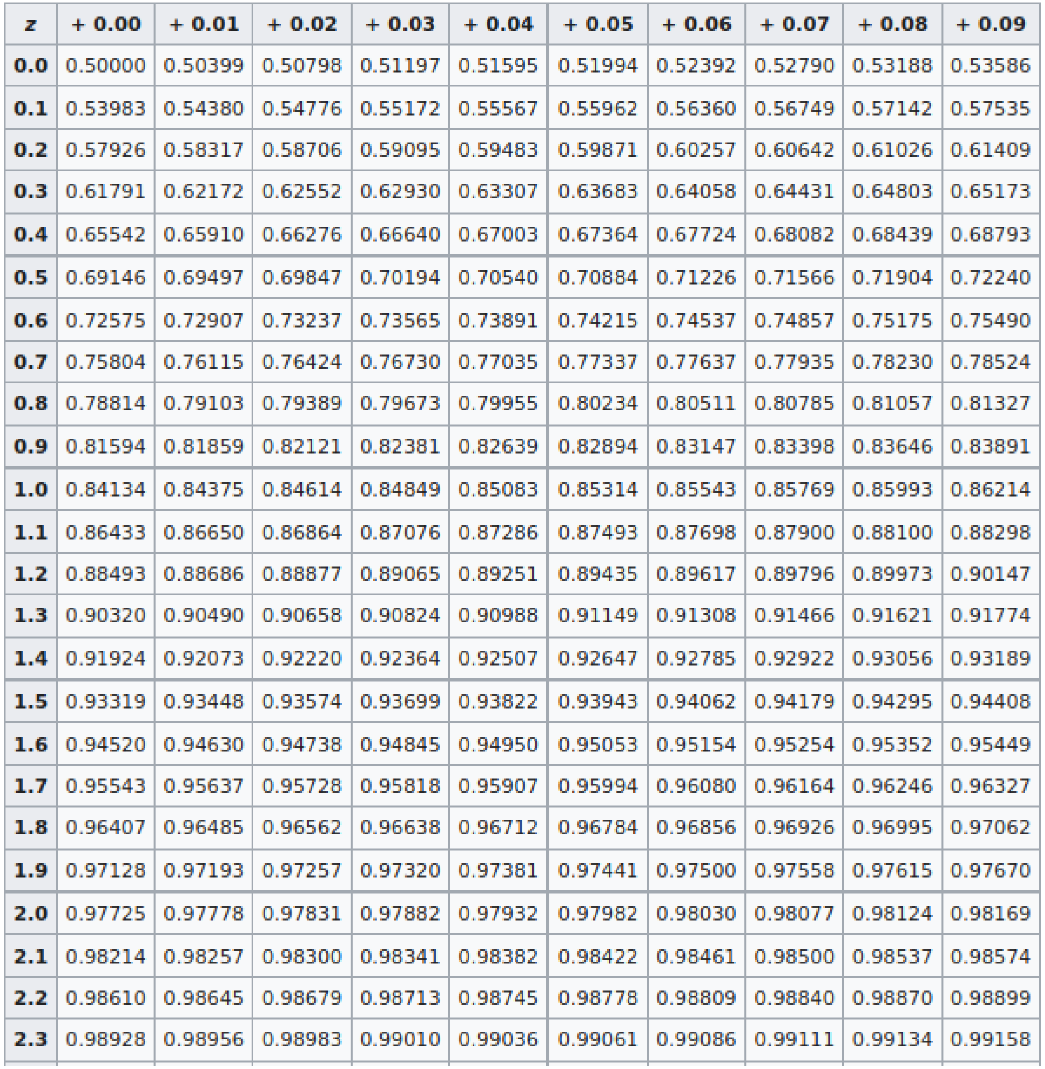
Means result is significant so we need to reject Null Hypothesis

The school students have different mean test score from region mean score

Cumilative probabilities - statistic is less than Z (i.e. between negative infinity and Z)



z	− 0.00	− 0.01	− 0.02	− 0.03	− 0.04	− 0.05	− 0.06	− 0.07	− 0.08	− 0.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551



z	+ 0.00	+ 0.01	+ 0.02	+ 0.03	+ 0.04	+ 0.05	+ 0.06	+ 0.07	+ 0.08	+ 0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158

One Sample Z test - Left tailed with Confidence Intervals

Confidance Interval:

$$CI = \overline{X} \pm Z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

95% Confidence Interval for left tailed test:

$$CI = \overline{X} \pm 1.645 \frac{\sigma}{\sqrt{n}}$$

$$CI = 96 \pm 1.645 \frac{12}{\sqrt{55}}$$

$$CI = 96 \pm 1.645 * 1.62$$

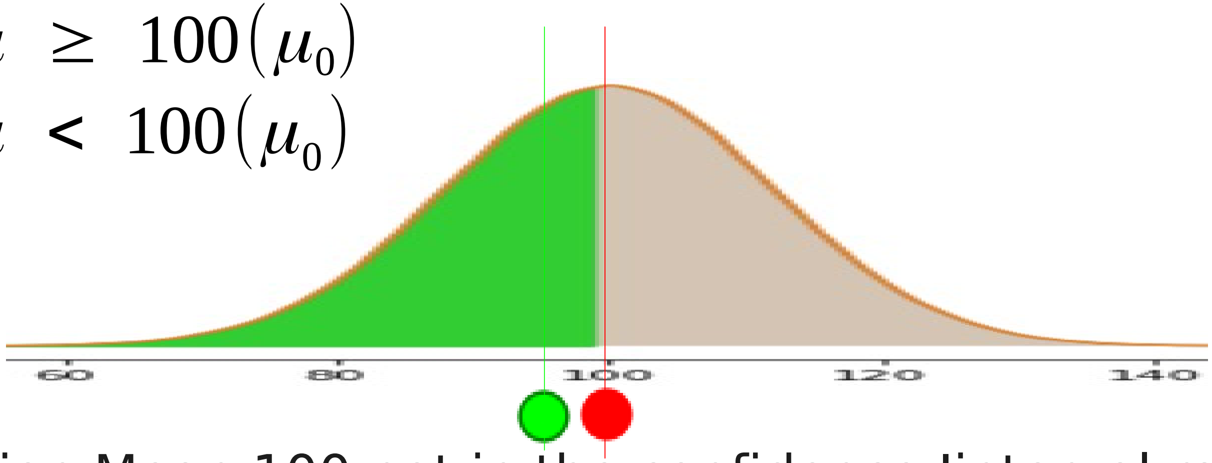
$$CI = 96 \pm 2.66$$

$$CI = [93.34 \quad 98.7]$$

$$CI = [-\infty \quad 98.7]$$

$$H_0: \ \mu \ \geq \ 100(\mu_0)$$

$$H_1: \ \mu \ < \ 100(\mu_0)$$



Population Mean 100 not in the confidence linterval range so

Null Hypothesis should be rejected

Z critical Values

	Two-tailed	One-tailed
$Alpha(\alpha)$	$Z_{\alpha/2}$	Z_{α}
0.01	2.575	2.33
0.05	1.96	1.645
0.10	1.645	1.28

One Sample Z test - Left tailed

A company claims that their oats contain at least 15.5g of protein per 100 grams. You try to verify with fallowing sample data

$$Population\,Standard\ Deviation\left(\sigma\right)=2\,g$$
$$SampleSize\left(n\right)=49$$
$$Sample\ Mean\left(\overline{x}\right)=13.75\,g$$
$$Significance\ Level\left(\alpha\right)=0.05$$

One Sample Z test - Left tailed

Here

$$Population\,Standard\;Deviation(\sigma)=2\,g$$
$$SampleSize\,(n)=49$$
$$Sample\;Mean(\bar{x})=13.75\,g$$
$$Population\;Mean(\mu_0)=15.5\,g$$
$$Significance\;Level(\alpha)=0.05$$

One sample Z-test Left-tailed

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0 : \ \mu \ \geq \ 15.5(\mu_0)$$
$$H_1 : \ \mu \ < \ 15.5(\mu_0)$$

Step 2: Choose level of significance(α)

α = 0.05

A value of α = 0.05 implies that the null hypothesis is rejected 5 % of the time when it is in fact true

One Sample Z test - Left tailed

Step 3: Determine appropriate test to use and find the test statistic

One sample Z-tests can be performed

1. Here we know the population Standard Deviation(σ)

2. sample size is large(>30)

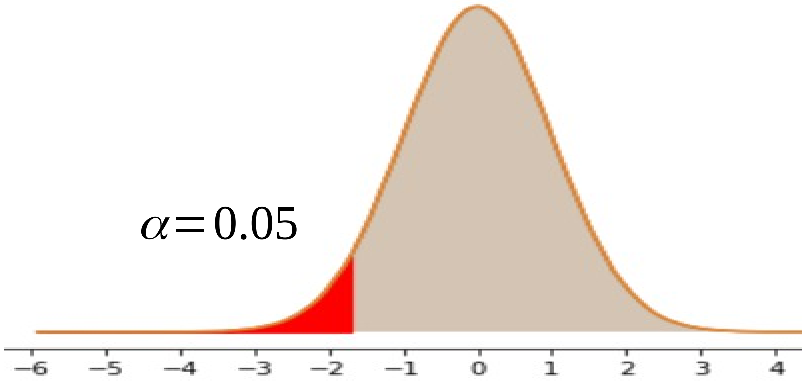
Calculate the z-statistic $Z=\frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ $Z=\frac{13.75-15.5}{\frac{2}{\sqrt{49}}}=\frac{-1.75}{0.28571}=-6.125$

The sample mean is 13.75, which is −6.125 standard error units from the population mean of 15.5

One Sample Z test - Left tailed

Step 4: Determine if we need a 1 tailed or a 2 tailed Z-critical value and find the Z-critical value for desired confidence level

$$H_0 : \ \mu \ \geq \ 15.5(\mu_0)$$
$$H_1 : \ \mu \ < \ 15.5(\mu_0)$$



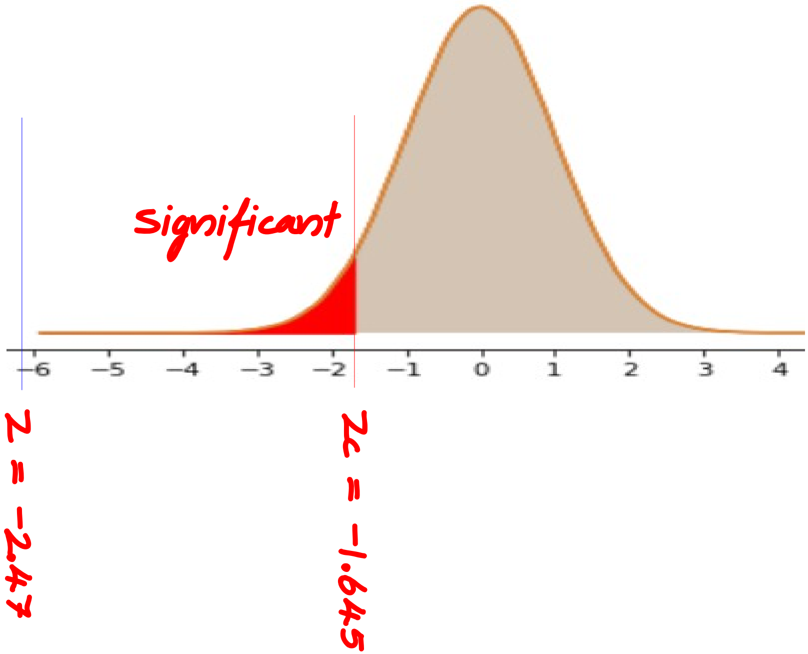
Z critical value for two tailed test for 95% confidence is -1.645

One Sample Z test – Left tailed

Step 5: **Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)**

Z-statistic(-6.125) < Z critical value(-1.645)

$$H_0: \ \mu \ \geq \ 15.5(\mu_0)$$
$$H_1: \ \mu \ < \ 15.5(\mu_0)$$



Z statistic is falls in the rejection region means it is very unlikely

Result is statistically significant so we need to reject Null Hypothesis

One Sample Z test - Left tailed

Step 5: **Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)**

The probability of observing a standard normal value below −6.125 is less than 0.00001.

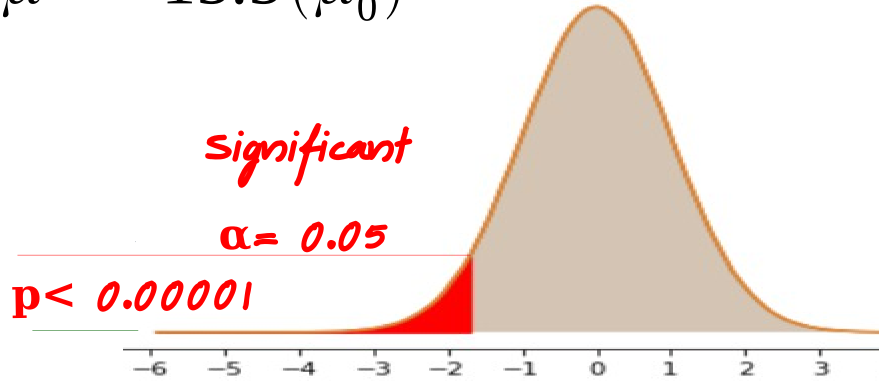
$$H_0 :\ \mu \ \geq\ 15.5(\mu_0)$$

$$H_1 :\ \mu \ <\ 15.5(\mu_0)$$

Since p(0.00001)< α(0.05)

Means result is significant so we need to reject Null Hypothesis

The Z-test tells us that the company oats have an unusually low protien whent it comared to population mean 15.5g



Cumilative probabilities - statistic is less than Z (i.e. between negative infinity and Z)

z	− 0.00	− 0.01	− 0.02	− 0.03	− 0.04	− 0.05	− 0.06	− 0.07	− 0.08	− 0.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551

z	+ 0.00	+ 0.01	+ 0.02	+ 0.03	+ 0.04	+ 0.05	+ 0.06	+ 0.07	+ 0.08	+ 0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158

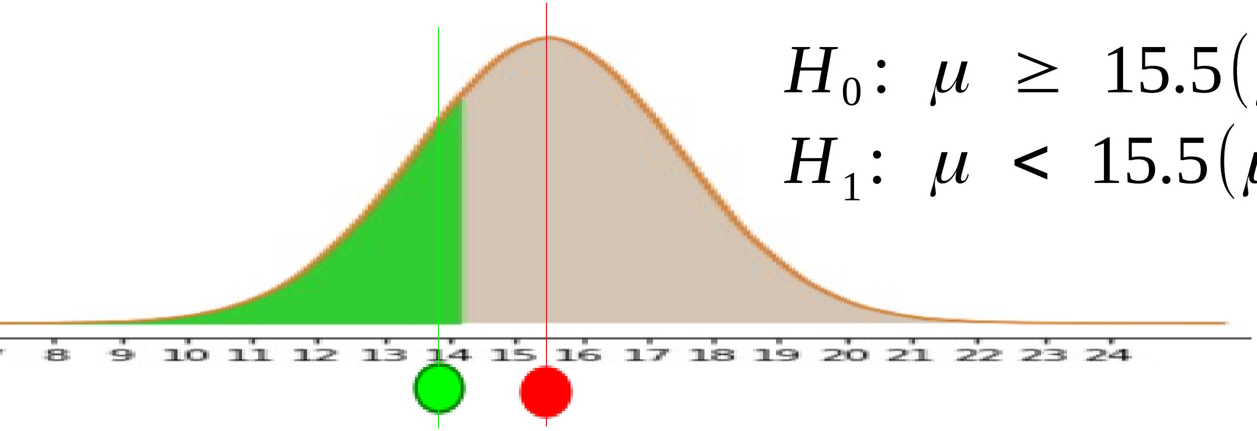
One Sample Z test - Left tailed with Confidence Intervals

Confidance Interval:

$$CI = \overline{X} \pm Z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

95% Confidence Interval for left tailed test:

$$CI = \overline{X} \pm 1.645 \frac{\sigma}{\sqrt{n}}$$



$$H_0: \ \mu \ \geq \ 15.5(\mu_0)$$
$$H_1: \ \mu \ < \ 15.5(\mu_0)$$

$$CI = 13.75 \pm 1.645 \frac{2}{\sqrt{49}}$$
$$CI = 13.75 \pm 1.645 * 0.286$$
$$CI = 13.75 \pm 0.47$$
$$CI = [13.3 \quad 14.2]$$
$$CI = [-\infty \quad 14.2]$$

Population Mean 15.5 not in the confidence linterval range so

Null Hypothesis should be rejected

Z critical Values

	Two-tailed	One-tailed
$Alpha(\alpha)$	$Z_{\alpha/2}$	Z_{α}
0.01	2.575	2.33
0.05	1.96	1.645
0.10	1.645	1.28

One Sample Z test - Right tailed

A wine manufacturer claims that his wine has at most 19.8 ppm(parts per million) impurities in barrel of wine. You try to verify this with following data.

$$Sample\ Standard\ Deviation\left(s\right)=3$$
$$SampleSize\left(n\right)=144$$
$$Sample\ Mean\left(\overline{x}\right)=20$$
$$Significance\ Level\left(\alpha\right)=0.10$$

One Sample Z test - Right tailed

Here

$$Sample\,Standard\,Deviation\left(s\right)=3$$
$$SampleSize\left(n\right)=144$$
$$Sample\,Mean\left(\overline{x}\right)=20$$
$$Population\,Mean\left(\mu_0\right)=19.8$$
$$Significance\,Level\left(\alpha\right)=0.10$$

One Sample Z test - Right tailed

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0:\ \mu\ \leq\ 19.8(\mu_0)$$

$$H_1:\ \mu\ >\ 19.8(\mu_0)$$

Step 2: Choose level of significance(α)

α = 0.1

A value of α = 0.1 implies that the null hypothesis is rejected 10 % of the time when it is in fact true

One Sample Z test - Right tailed

Step 3: Determine appropriate test to use and find the test statistic

One sample Z-tests can be performed

1. Here we know the population Standard Deviation(σ)

2. sample size is large(>30)

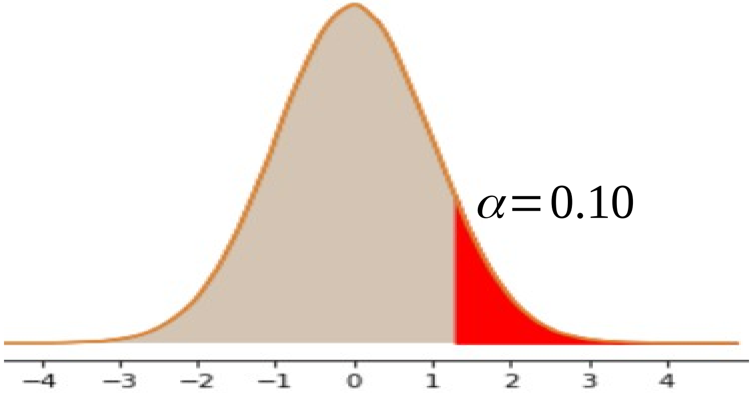
Calculate the z-statistic $$Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$ $$Z = \frac{20 - 19.8}{\frac{3}{\sqrt{144}}} = \frac{0.199}{0.25} = 0.8$$

The sample mean is 20, which is 0.8 standard error units from the population mean of 19.8

One Sample Z test - Right tailed

Step 4: Determine if we need a 1 tailed or a 2 tailed Z-critical value and find the Z-critical value for desired confidence level

$$H_0: \ \mu \ \le \ 19.8\left(\mu_0\right)$$
$$H_1: \ \mu \ > \ 19.8\left(\mu_0\right)$$



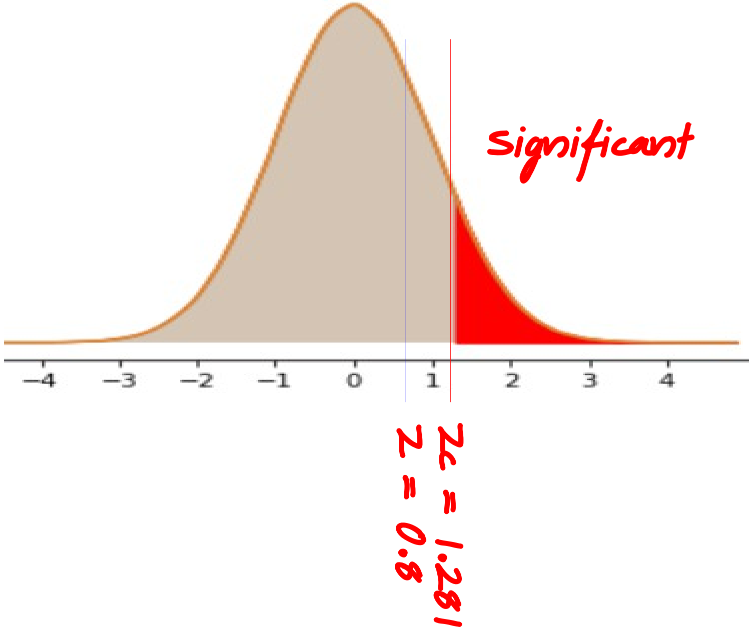
Z critical value for two tailed test for 90% confidence is 1.281

One Sample Z test - Right tailed

Step 5: Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

Z-statistic(0.8) < Z critical value(1.281)

$$H_0:\ \mu \ \le\ 19.8(\mu_0)$$
$$H_1:\ \mu \ >\ 19.8(\mu_0)$$



Z statistic do not falls in the rejection region means it is very likely in null hypothesis

Result is not statistically significant so we can’t reject Null Hypothesis

One Sample Z test - Right tailed

Step 5: Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

The probability of observing a standard normal value above 0.8 is 0.212.

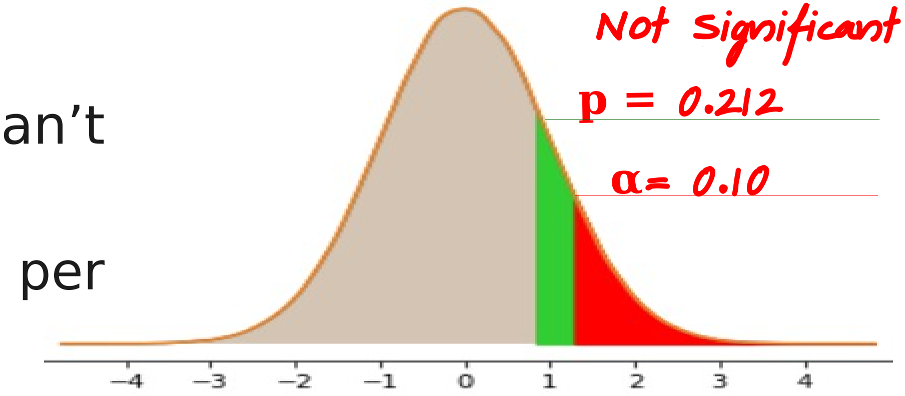
$$H_0: \ \mu \ \leq \ 19.8(\mu_0)$$

$$H_1: \ \mu \ > \ 19.8(\mu_0)$$

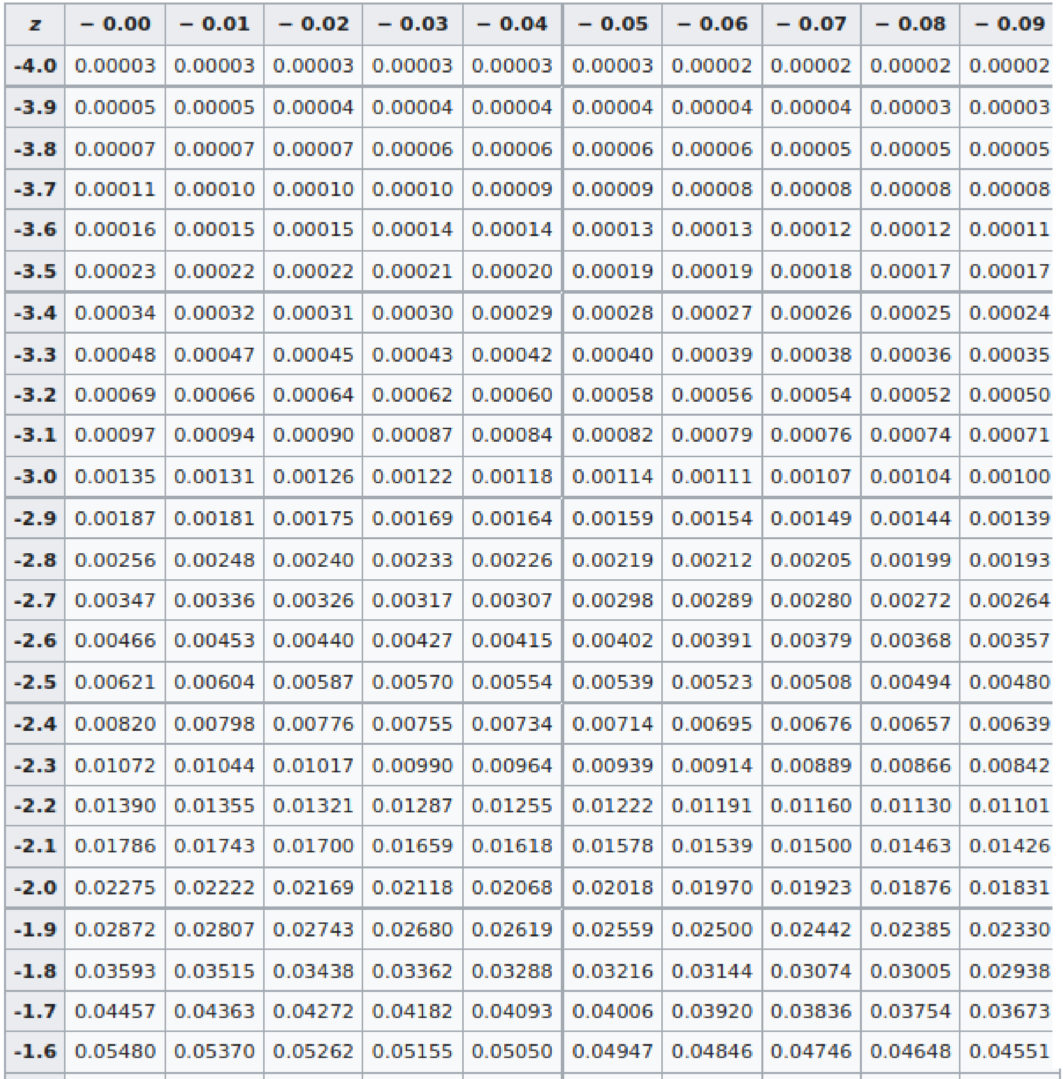
Since p(0.212)> α(0.10)

Means result is not significant so we can’t reject Null Hypothesis

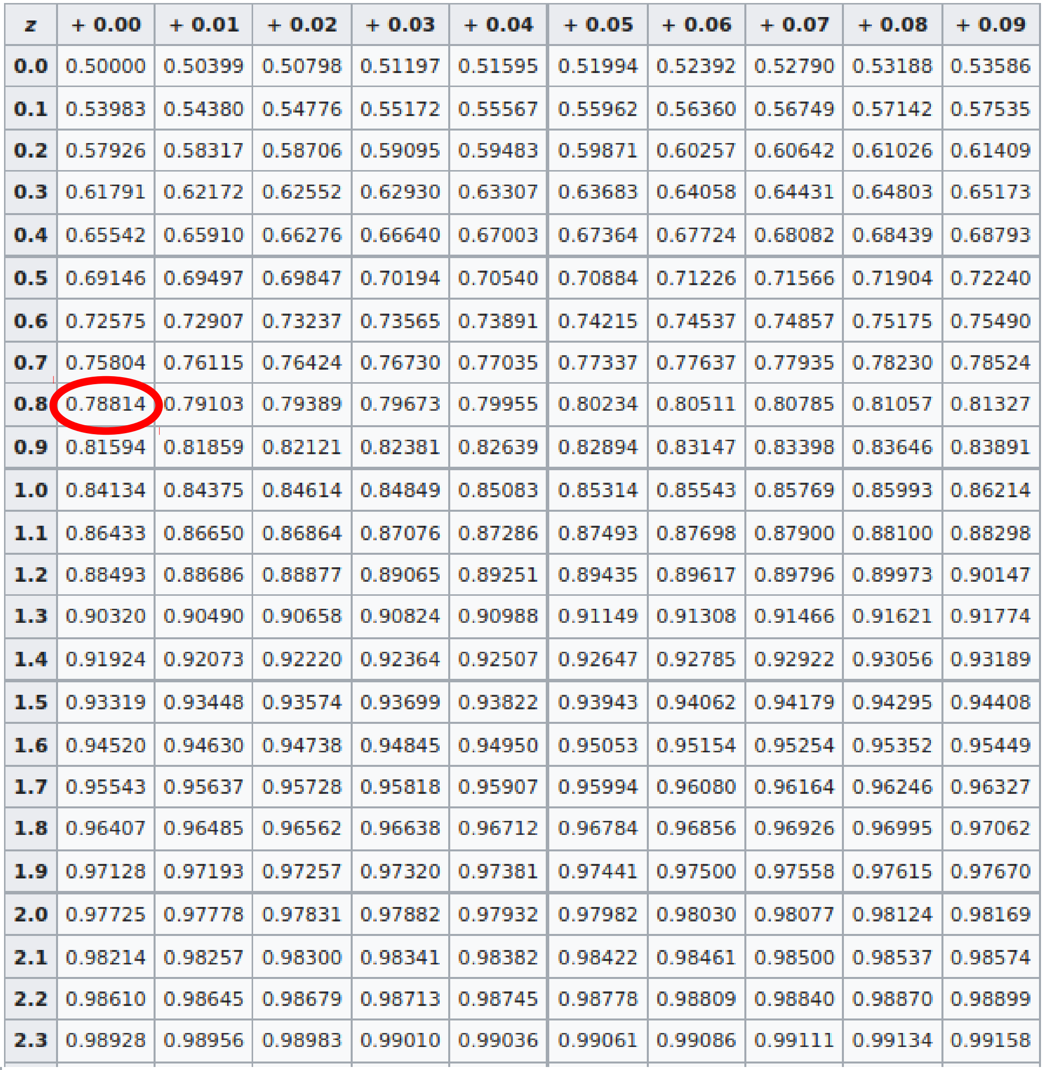
The wine has at most 19.8 ppm(parts per million) impurities in barrel of wine



Cumilative probabilities - statistic is less than Z (i.e. between negative infinity and Z)



z	− 0.00	− 0.01	− 0.02	− 0.03	− 0.04	− 0.05	− 0.06	− 0.07	− 0.08	− 0.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551



z	+ 0.00	+ 0.01	+ 0.02	+ 0.03	+ 0.04	+ 0.05	+ 0.06	+ 0.07	+ 0.08	+ 0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158

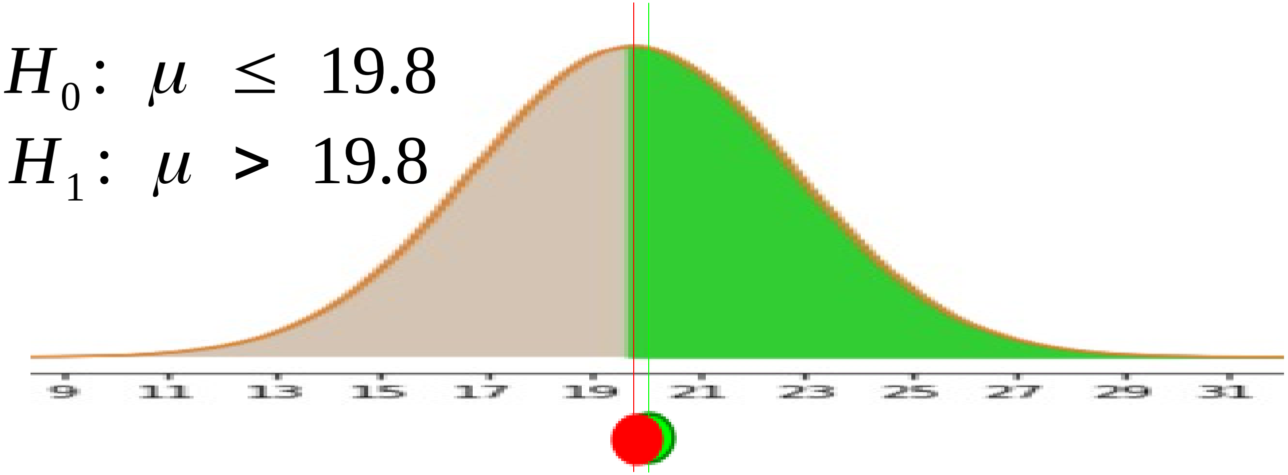
One Sample Z test - Right tailed with Confidence Intervals

Confidance Interval:

$$CI = \overline{X} \pm Z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

90% Confidence Interval for right tailed test:

$$CI = \overline{X} \pm 1.2816 \frac{\sigma}{\sqrt{n}}$$



$$H_0: \ \mu \ \leq \ 19.8$$

$$H_1: \ \mu \ > \ 19.8$$

$$CI = 20 \pm 1.2816 \frac{3}{\sqrt{144}}$$

$$CI = 20 \pm 1.2816 * 0.25$$

$$CI = 20 \pm 0.3204$$

$$CI = [19.679 \quad 20.3204]$$

$$CI = [19.679 \quad \infty]$$

Population Mean 19.8 in the confidence linterval range so

Null Hypothesis can’t be rejected

Z critical Values

	Two-tailed	One-tailed
$Alpha(\alpha)$	$Z_{\alpha/2}$	Z_{α}
0.01	2.575	2.33
0.05	1.96	1.645
0.10	1.645	1.28

Two Sample Z test

Two Sample Z-test

In many cases we would like to compare parameters of two different populations to check for any difference in the parameter vallues such as mean.

a two-sample test is a test performed on the data of two random samples, each independently obtained from a different population.

The purpose of the test is to determine whether the difference between these two populations is statistically significant.

Examples:

What is the effect of taking a drug vs. a placebo in controlling pain?

Do interest rates rise more quickly when wages increase or stay the same?

Who is more likely to vote in an election, Democrats or Republicans?

Each of these involves comparison of two samples.

Two Sample Z test Assumptions

1. The population standard deviations σ_1 and σ_2 are known

2. The population follow the normal probability distribution or The sample size is large(n>30)

3. Observations should be independent

Z statistic for Two Sample Z-test

Assume that μ_1 and μ_2 are population means. Our intrest is to check a hypothesis on difference between μ_1 and μ_2 that is $\left(\mu_1-\mu_2\right)$.

If $\overline{X}_1$ and $\overline{X}_2$ are the estimated mean values from two samples drawn from two populations, the statistic $\left(\overline{X}_1-\overline{X}_2\right)$ follows a standard normal distribution with mean $\left(\mu_1-\mu_2\right)$ and standard deviation $\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

where n_1 and n_2 are sample sizes of two samples. The corresponding Z statistic is given by

$$Z=\frac{\left(\overline{X}_1-\overline{X}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

Two Sample Z-test

Test the hypothesis that the population means are equal for the two samples.

The first sample is miles per gallon for U.S. cars and the second sample is miles per gallon for Japanese cars.

the summary statistics for each sample are shown below

sample	Sample size	Mean	Population Standard deviation
Sample1(US)	249	20.14458	6.41470
Sample2(Japan)	79	30.48101	6.10771

Two Sample Z-test - Two tailed

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0 : \mu_1 = \mu_2 \qquad H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 \neq \mu_2 \qquad H_1 : \mu_1 - \mu_2 \neq 0$$

Step 2: Choose level of significance(α)

α = 0.05

A value of α = 0.05 implies that the null hypothesis is rejected 5 % of the time when it is in fact true

Two Sample Z-test - Two tailed

Step 3: Determine appropriate test to use and find the test statistic

Two sample Z-tests can be performed

1. Here we know the population Standard Deviations (σ$_1$ and σ$_2$)

2. Sample size is large(>30)

3. Observations are independent

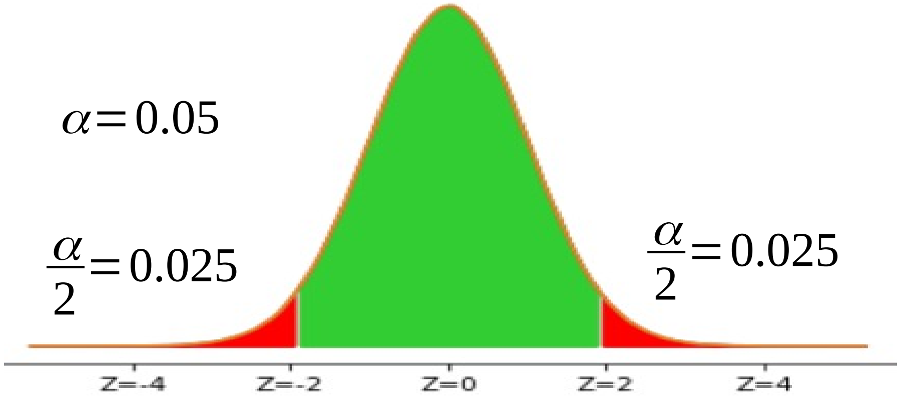
$$Z-statistic=\frac{\left(\overline{X}_1-\overline{X}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{\left(20.14458-30.48101\right)-\left(0\right)}{\sqrt{\frac{6.41470^2}{249}+\frac{6.10771^2}{79}}}=\frac{-10.33643}{0.798}=-12.95$$

The difference of sample means $\left(\overline{X}_1-\overline{X}_2\right)$ is -12.95 standard error units from the expected population mean difference $\left(\mu_1-\mu_2\right)$

Two Sample Z-test - Two tailed

Step 4: Determine if we need a 1 tailed or a 2 tailed Z-critical value and find the Z-critical value for desired confidence level

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



Z critical value for two tailed test for 95% confidence is 1.96

Two sample Z-test Two-tailed

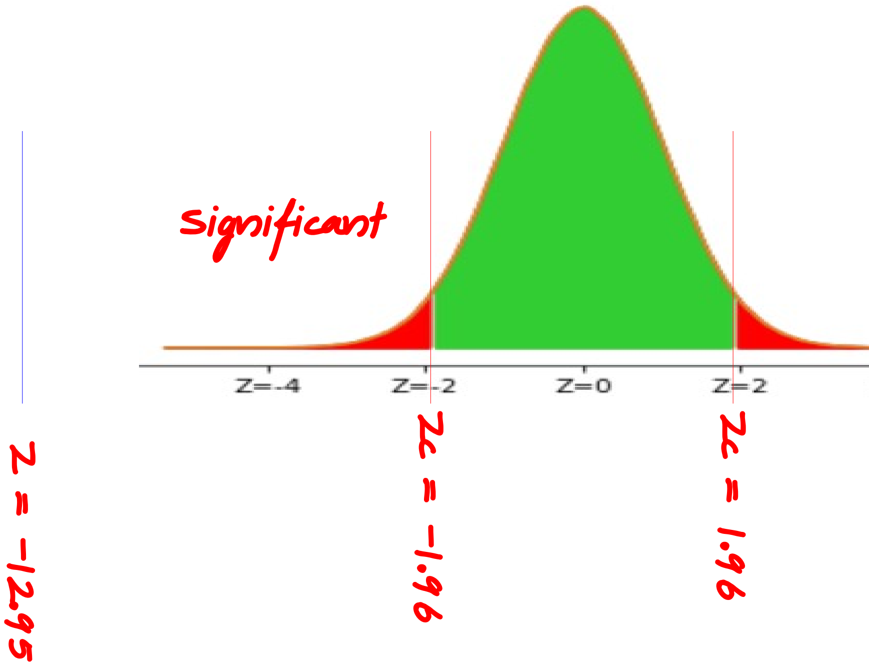
Step 5: Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

Z-statistic(-12.95) < Z critical value(-1.96)

Z statistic is falls in the rejection region

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Result is statistically significant so we need to reject Null Hypothesis



Two sample Z-test Two-tailed

Step 5: Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

The probability of observing a standard normal value below −12.95 is < 0.00001 (2 tailed p value)

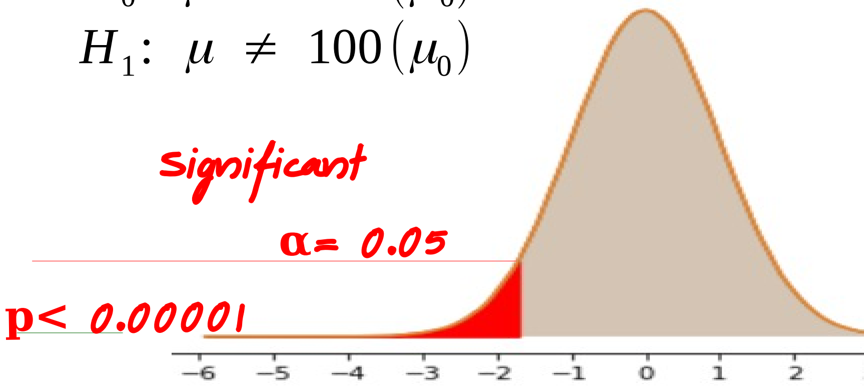
Significance Level(α) = 0.05

Since p(0.00001)< α(0.05)

$$H_0:\ \mu\ =\ 100\left(\mu_0\right)$$
$$H_1:\ \mu\ \neq\ 100\left(\mu_0\right)$$

Means result is significant so we need to reject Null Hypothesis

The two population means are different at the 0.05 significance level

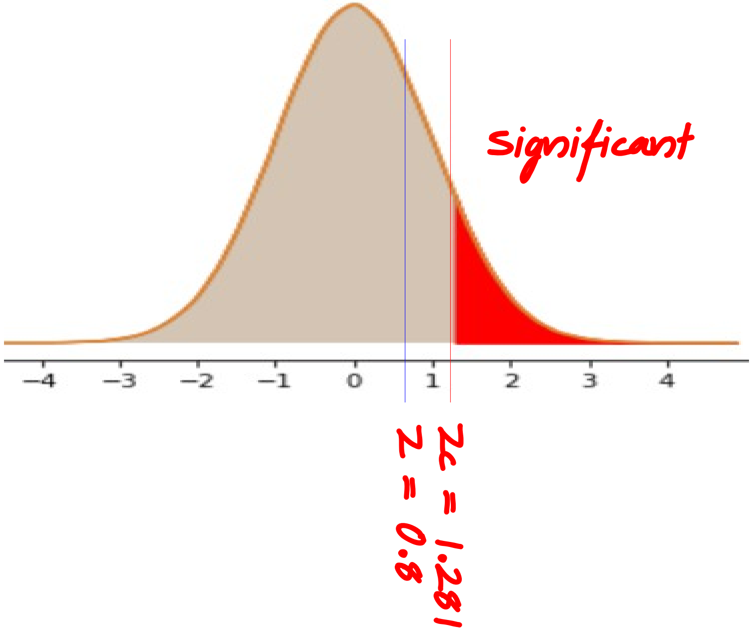


Two Sample Z-test - Two tailed

Step 5: **Compare test statistic with Z-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)**

Z-statistic(0.8) < Z critical value(1.281)

$$H_0: \ \mu \ \leq \ 19.8(\mu_0)$$
$$H_1: \ \mu \ > \ 19.8(\mu_0)$$



Z statistic do not falls in the rejection region means it is very likely in null hypothesis

Result is not statistically significant so we can’t reject Null Hypothesis

Two Sample Z-test

Marketing specialization students earns atleast 5000 more per month than operations management students test this with following sample data

specialization	Sample size	Mean salary	Population Standard deviation
Marketing	120	67500	7200
Operations	45	58950	4600

Two Sample Z-test

The dataset "Normal Body Temperature, Gender, and Heart Rate" contains 130 observations of body temperature, along with the gender of each individual and his or her heart rate. In the dataset, the first column gives body temperature and the second column gives the value "1" (male) or "2" (female) to describe the gender of each subject. Using the MINITAB "DESCRIBE" command with the "BY" subcommand to separate the two genders provides the following information:

Is there a significant difference between the mean body temperatures for men and women?

specialization	Sample size	Mean salary	Population Standard deviation
Marketing	120	67500	7200
Operations	45	58950	4600

Two Sample Z-test

We have $n_1 = 120, n_2 = 45, \overline{X}_1 = 67500, \overline{X}_2 = 58{,}950, \mu_1 = 7200$ and $\mu_2 = 4600$

$$H_0 : \mu_1 - \mu_2 \le 5000$$
$$H_1 : \mu_1 - \mu_2 > 5000$$

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(67500 - 58950) - (5000)}{\sqrt{\frac{7200^2}{120} + \frac{4600^2}{45}}} = \frac{3550}{949.85} = 3.7374$$

Z critical value for 95% confidence level is 1.645

Since Z statistic value is higher than Z-critical value we reject null hypothesis

Two Sample Z-test

Typing Speed on PC, typing speed of men and women are not equal

	Sample size	Sample Mean	Population Standard deviation
Men	50	65 wpm	10 wpm
Women	60	68 wpm	14 words per min

test at $\alpha = 0.01$

Two Sample Z-test

We have $n_1=50, n_2=60, \overline{X}_1=65, \overline{X}_2=68, \mu_1=10$ and $\mu_2=14$

$$H_0:\mu_1-\mu_2=0$$
$$H_1:\mu_1-\mu_2\neq 0$$

$$Z=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{(65-68)-(0)}{\sqrt{\frac{10^2}{50}+\frac{14^2}{60}}}=\frac{-3}{\sqrt{\frac{100}{50}+\frac{196}{60}}}=\frac{-3}{2.29}=-1.31$$

Z critical value for 99% confidence level is 2.575

Since Z statistic value is in acceptance region we can’t reject null hypothesis

Student’s t distribution

Student's t-distribution

Student's t-distribution is any member of a family of continuous probability distributions.

That arises when estimating the mean of a normally distributed population.

In situations where the sample size is small and the population standard deviation is unknown.

It was developed by William Sealy Gosset under the pseudonym Student.

If we take a sample of n observations from a normal distribution, then the t-distribution with ν = n-1 degrees of freedom can be defined as the distribution of the location of the sample mean relative to the true mean, divided by the sample standard deviation, after multiplying by the standardizing term $\sqrt{n}$.

Degrees of freedom

Suppose the average of 5 numbers is 3

If given a equation $\frac{10+X}{5}=3$

Here X must be 5, X do not have any freedom to vary.

$$\frac{X_1+X_2+X_3+X_4+X_5}{5}=3$$

X and Mean are dependent on each other so after estimating mean we have only 4 independent choices for our sample of size 5.

Degrees of freedom

The number of degrees of freedom is the number of values in the final calculation of a statistic that are free to vary.

In general, the degrees of freedom of an estimate of a parameter are equal to the number of independent scores that go into the estimate minus the number of parameters used as intermediate steps in the estimation of the parameter itself

If there are n observations in sample and k parameters are estimated from the sample then degree of freedom is (n-k).

Degrees of freedom

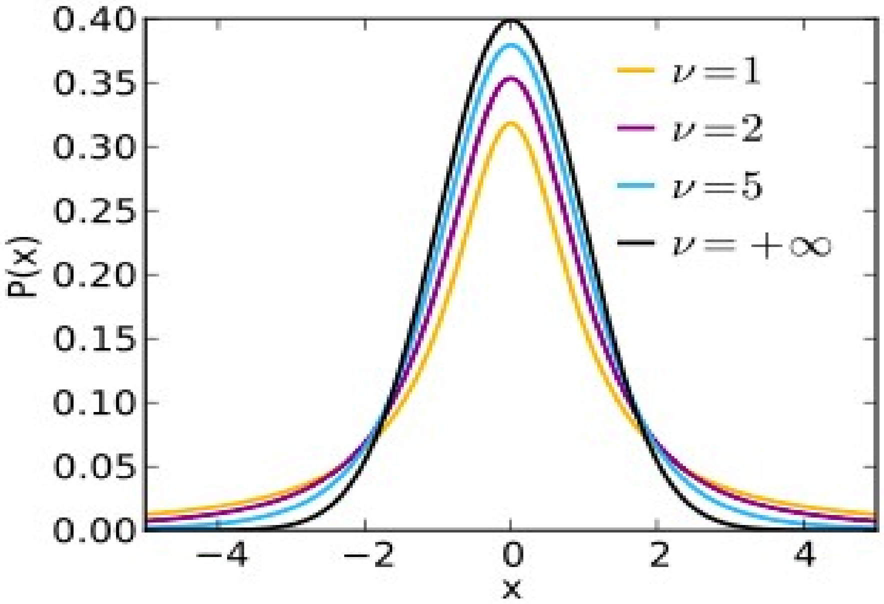
Most of the time the sample variance has N − 1 degrees of freedom, since it is computed from n random scores minus the only 1 parameter estimated as intermediate step, which is the sample mean.

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

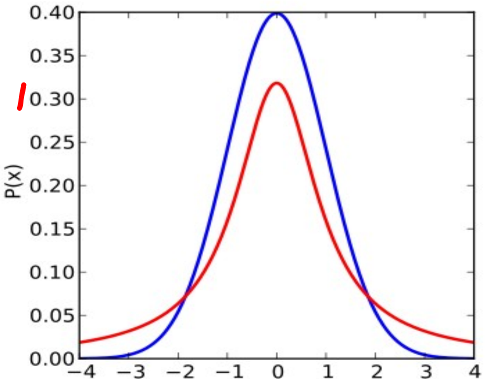
Here $\overline{X}$ is estimated from sample thus the degrees of freedom is (n-1)

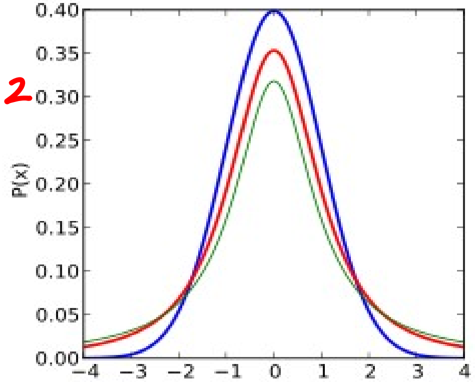
Student's t-distribution

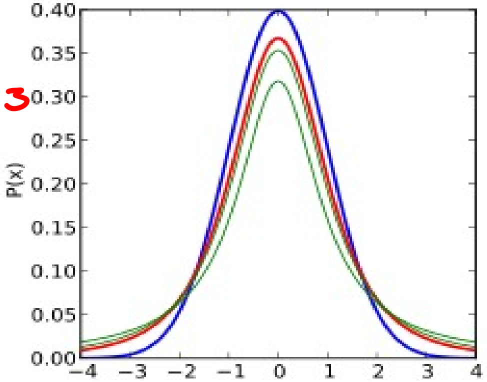
The t-distribution is symmetric and bell-shaped, like the normal distribution, but has heavier tails, meaning that it is more prone to producing values that fall far from its mean.

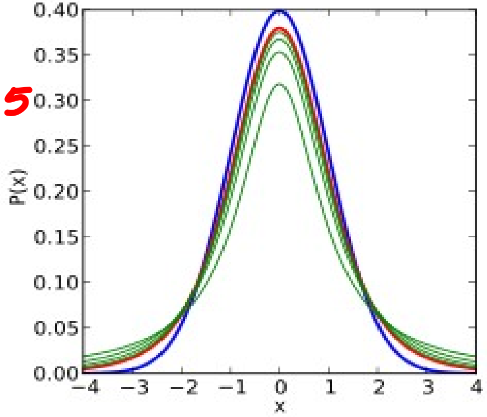


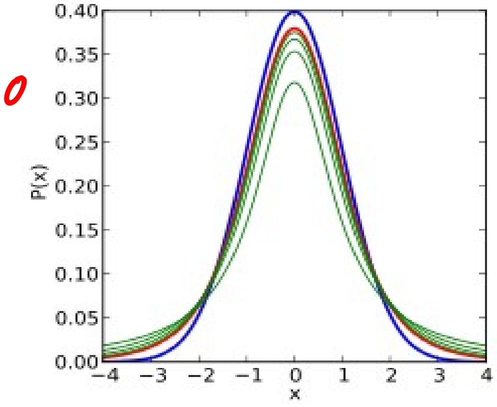
Student's t-distribution

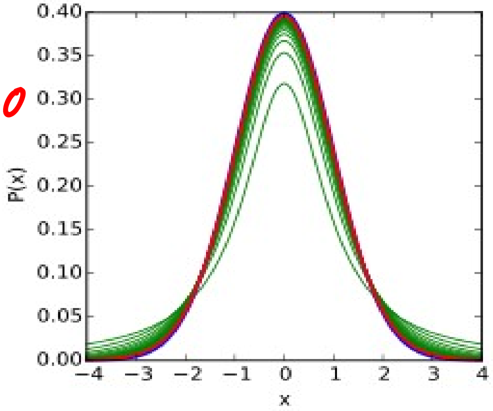

df = 1


df = 2


df = 3


df = 5


df = 10


df = 30

Student's t-distribution

Let X $_1$, … , X $_n$ be independently and identically drawn from the distribution N(μ,σ^2) i.e. this is a sample of size n from a normally distributed population with expected mean value μ and variance σ^2.

Let $\overline{X}=\frac{\sum_{i=1}^{n} X_i}{n}$ be the sample mean $S^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$ sample variance.

Then the random variable $\frac{\overline{X}-\mu}{\sigma\ /\sqrt{n}}$ has standard normal distribution and the

Random variable $\frac{\overline{X}-\mu}{S\ /\sqrt{n}}$ has a Student's t-distribution with n-1 degrees of freedom.

where S has been substitute for σ

t test

Popular Hypothesis Tests

Hypothesis test		Assumptions or notes
Z tests	One-sample Z test	(Normal population **or** n large) **and** $\sigma \quad known$
	Two-sample Z test	Normal population **and** independent observations **and** $\sigma_1 \wedge \sigma_2 \quad known$
T tests	One-sample t test	(Normal population **or** n large) **and** $\sigma \quad unknown$
	Paired t test	(Normal population of differences **or** n large) **and** $\sigma \quad unknown$
	Two-sample t test with equal variances	(Normal populations **or** $n_1 + n_2 > 40$) **and** independent observations **and** $\sigma_1 = \sigma_2 \ \ unknown$
	Two-sample t test with unequal variances	(Normal populations **or** $n_1 + n_2 > 40$) **and** independent observations **and** $\sigma_1 \neq \sigma_2 \ \ both \ \ unknown$
Chi-squared test	Chi-squared goodness fit test	Will be applied on categorical variables **and** All expected counts are atleast 5
F test	ANOVA	Normal Population

One Sample t test

One Sample t test Assumptions

1. The population variance(σ^2) is unknown

2. The population follow the normal probability distribution or the sample size is large

One-sample t-test

In testing the null hypothesis that the population mean is equal to a specified value μ_0, one uses the statistic

$$t = \frac{\overline{X} - \mu_0}{s \;\; / \sqrt{n}}$$

where $\overline{X}$ is the sample mean, s is the sample standard deviation and n is the sample size. The degrees of freedom used in this test are n − 1.

One Sample t test

A manufacturer want to test a hypothesis to see if the diameter of their M30 bolts are being manufactured properly at α = 0.05.

He collected a sample of 6 and their values(in mm) are 30.02, 29.99, 30.11, 29.97, 30.01, 29.99.

One Sample t test

A manufacturer want to test a hypothesis to see if the diameter of their M30 bolts are being manufactured properly at α = 0.05.

He collected a sample of 5 and their values(in mm) are 30.02, 29.99, 30.11, 29.97, 30.01, 29.99.

Here $\overline{X}$ = 30.015, μ_0 = 30mm, s = 0.049 and n = 6

One Sample t test - Two tailed

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0: \ \mu \ = \ 30\left(\mu_0\right)$$
$$H_1: \ \mu \ \neq \ 30\left(\mu_0\right)$$

Step 2: Choose level of significance(α)

α = 0.05

A value of α = 0.05 implies that the null hypothesis is rejected 5 % of the time when it is in fact true

One Sample t test - Two tailed

Step 3: Determine appropriate test to use and find the test statistic

1. Here we do not know the population Standard Deviation(σ)
2. sample size is not large(<30) so Assume population is normally distributed

One sample t-tests can be performed

Calculate the t-statistic

$$t = \frac{\overline{X} - \mu_0}{s \ / \sqrt{n}}$$

$$t = \frac{30.015 - 30}{0.05 \ / \sqrt{6}} = \frac{0.015}{0.02} = 0.75$$

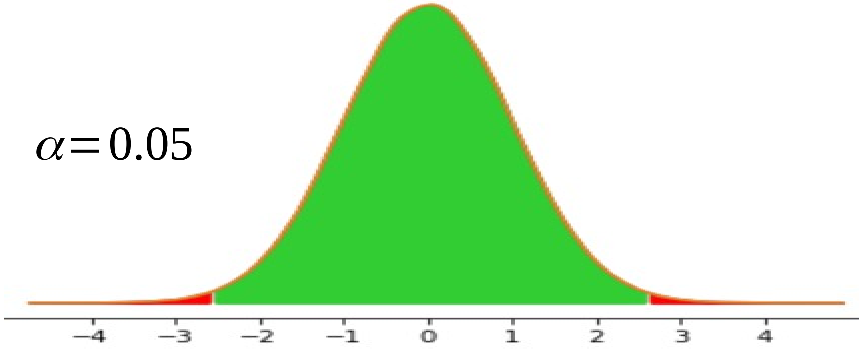
The sample mean score is 30.015, which is 0.75 standard error units from the population mean of 30

One Sample t test - Two tailed

Step 4: Determine if we need a 1 tailed or a 2 tailed t-critical value and find the t-critical value for desired confidence level

$$H_0: \ \mu \ = \ 30(\mu_0)$$
$$H_1: \ \mu \ \neq \ 30(\mu_0)$$

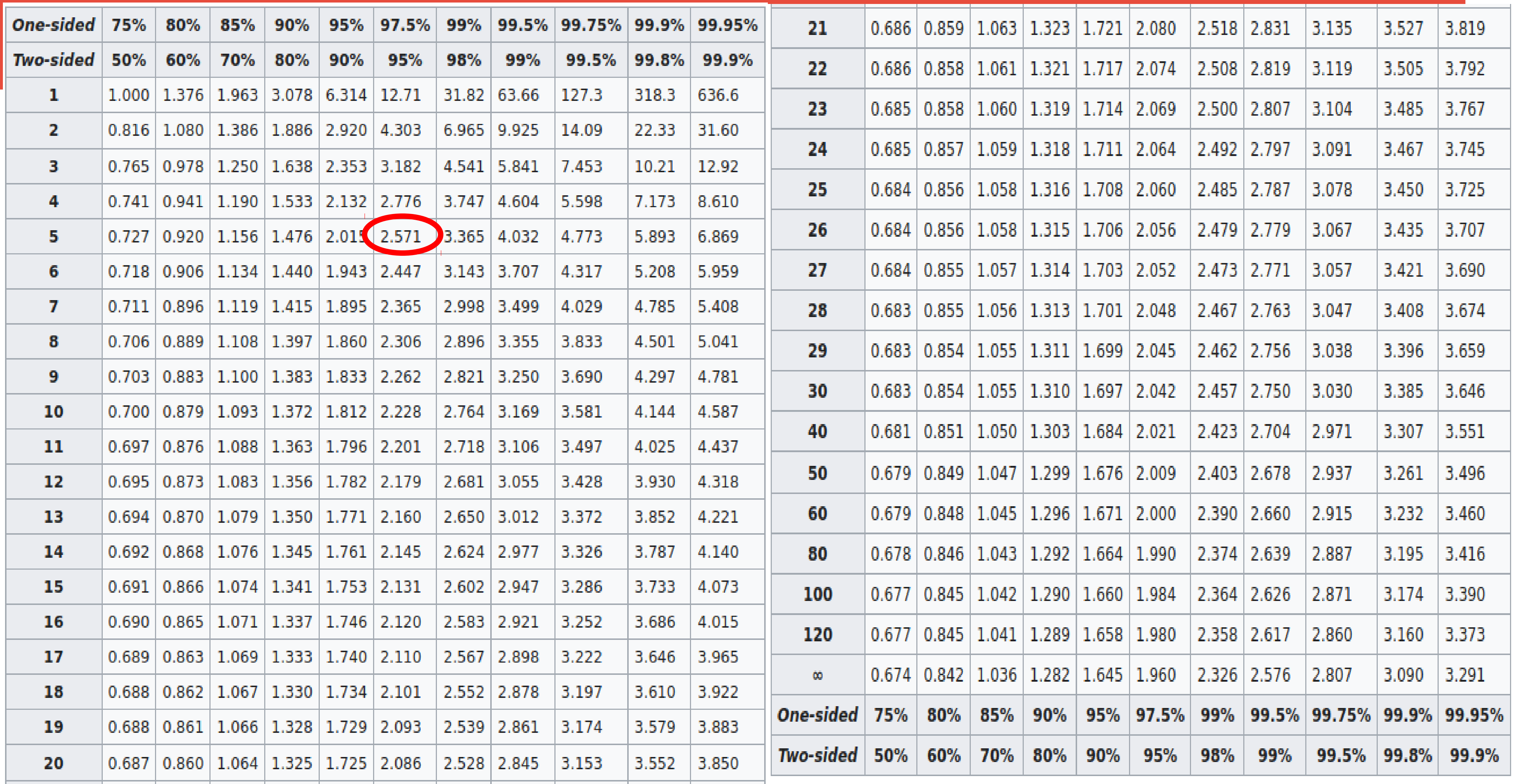
$\alpha = 0.05$



t critical value for two tailed test for 95% confidence at d.f = 5

is +/- 2.571

t-table with selected degrees of freedom for one-sided or two-sided critical values



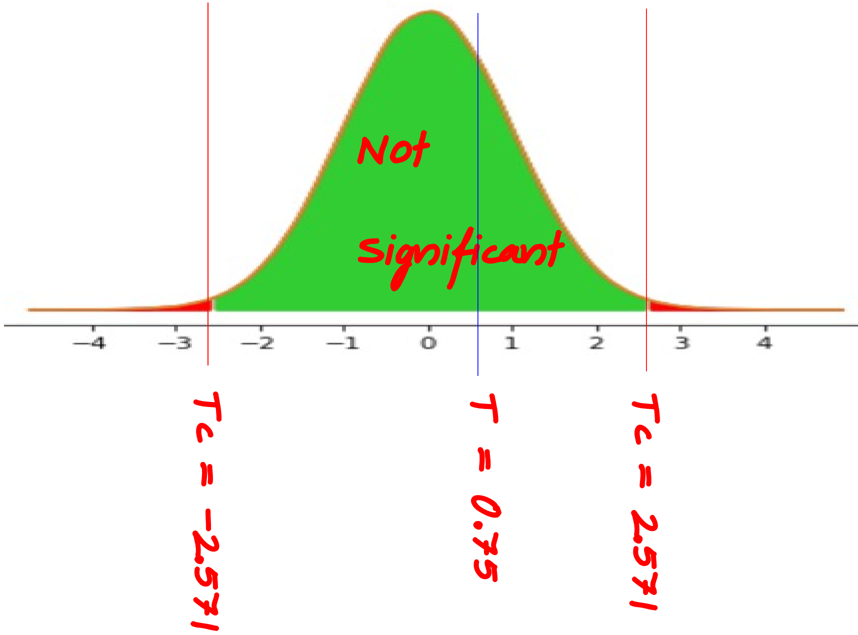
One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%

One Sample t test - Two tailed

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

t_c(-2.571)<t-statistic(0.75) < t_c(2.571)

$$H_0: \ \mu \ = \ 30(\mu_0)$$
$$H_1: \ \mu \ \neq \ 30(\mu_0)$$



t statistic do not falls in the rejection region means it is very likely in null hypothesis

Result is not statistically significant so we can’t reject Null Hypothesis

One Sample t test - Two tailed

Step 5: **Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)**

The probability of observing a standard normal value above 0.75 is 0.24.

2 tailed p= 0.48 = 2*0.24

$$H_0: \ \mu \ = \ 30(\mu_0)$$

$$H_1: \ \mu \ \neq \ 30(\mu_0)$$

Since p(0.48)> α(0.05)

Not significant

Means result is not significant so we can’t reject Null Hypothesis

The t-test tells us that the bolts are manufactured properly

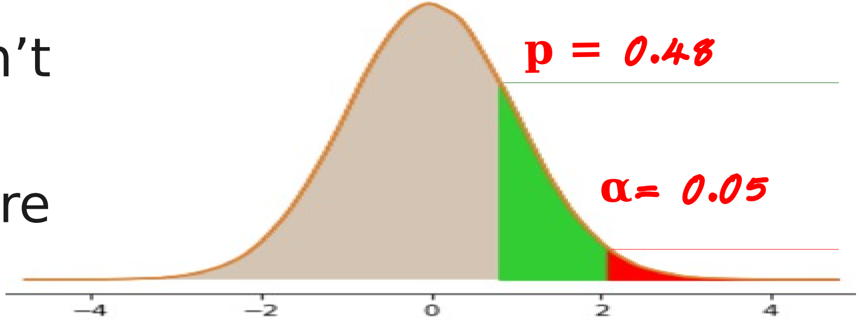
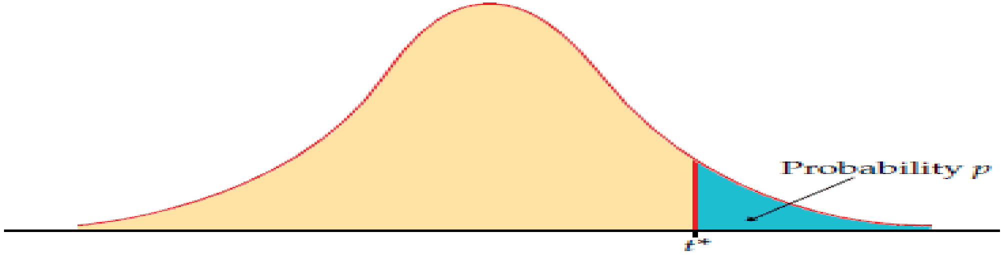


Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.



t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

One Sample t test - Two tailed with Confidence Intervals

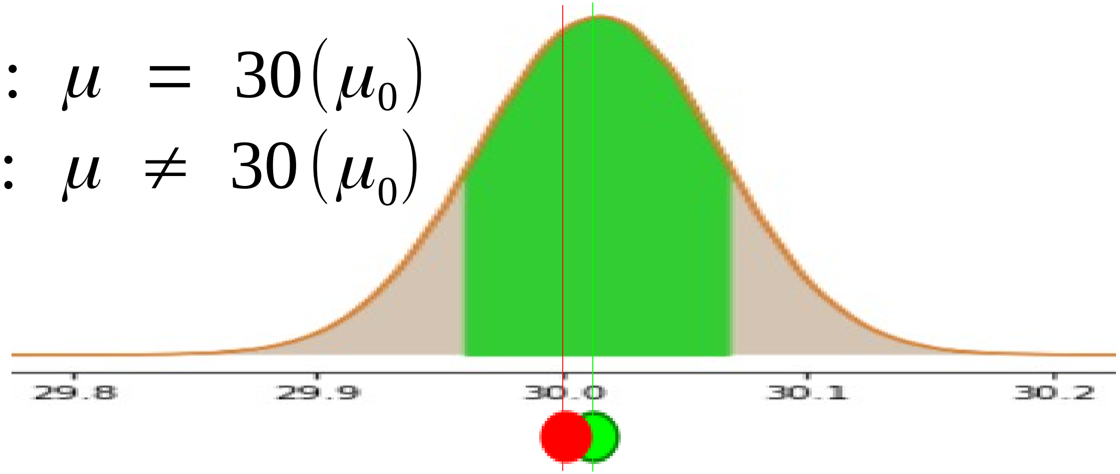
Confidance Interval:

$$CI = \overline{X} \pm t_{(\frac{\alpha}{2}, n-1)} \frac{s}{\sqrt{n}}$$

95 %, two-sided confidence interval:

$$CI = \overline{X} \pm 2.571 \frac{s}{\sqrt{n}}$$

$$H_0 : \ \mu \ = \ 30 \left(\mu_0\right)$$
$$H_1 : \ \mu \ \neq \ 30 \left(\mu_0\right)$$



$$CI = 30.015 \pm 2.571 \frac{0.049}{\sqrt{6}}$$
$$CI = 30.015 \pm 2.571 * 0.02$$
$$CI = 30.015 \pm 0.05142$$
$$CI = [29.96358 \qquad 30.06642]$$

Population Mean 30 in the confidence linterval range so

Null Hypothesis can’t be rejected

One Sample t test

Perform one sample t test for following summary data to test the null hypothesis that the population mean is less than or equal to 5.

$$Sample\ Standard\ Deviation\left(s\right)=0.022789$$
$$SampleSize\left(n\right)=195$$
$$Sample\ Mean\left(\overline{x}\right)=9.26146$$
$$Significance\ Level\left(\alpha\right)=0.05$$

One Sample t test

Given

$$Sample\ Standard\ Deviation\left(s\right)=0.022789$$
$$SampleSize\left(n\right)=195$$
$$Sample\ Mean\left(\overline{x}\right)=9.26146$$
$$Population\ Mean\left(\mu_0\right)=5$$
$$Significance\ Level\left(\alpha\right)=0.05$$

One Sample t test - Right tailed

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0: \ \mu \ \leq \ 5(\mu_0)$$
$$H_1: \ \mu \ > \ 5(\mu_0)$$

Step 2: Choose level of significance(α)

α = 0.05

A value of α = 0.05 implies that the null hypothesis is rejected 5 % of the time when it is in fact true

One Sample t test - Right tailed

Step 3: Determine appropriate test to use and find the test statistic

One sample t-tests can be performed

1. Here we do not know the population Standard Deviation(σ)

2. sample size is large(>30) **assume population normally distributed**

Calculate t-statistic

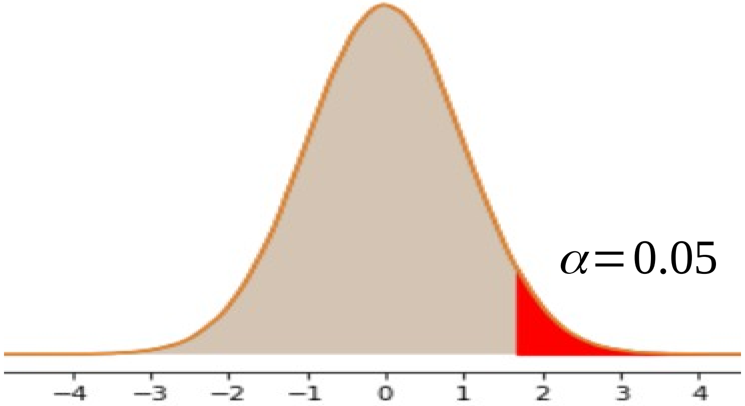
$$t = \frac{\overline{X} - \mu_0}{s \;\; / \sqrt{n}} \qquad t = \frac{9.261460 - 5}{0.022789 \;\; / \sqrt{195}} = \frac{4.26146}{0.00163} = 2614.392$$

The sample mean score 9.26146, which is 2614.39 standard error units from the population mean of 5

One Sample t test - Right tailed

Step 4: Determine if we need a 1 tailed or a 2 tailed t-critical value and find the t-critical value for desired confidence level

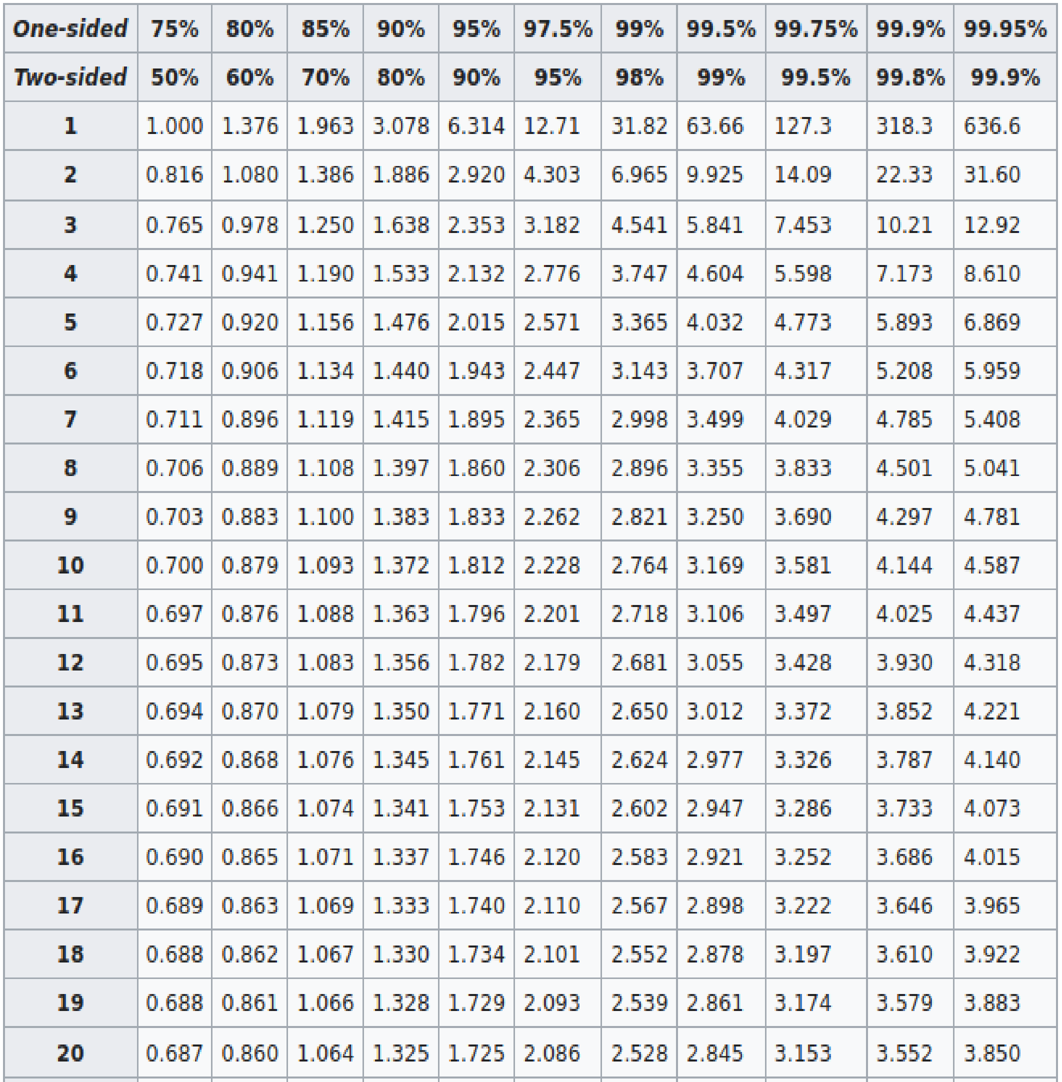
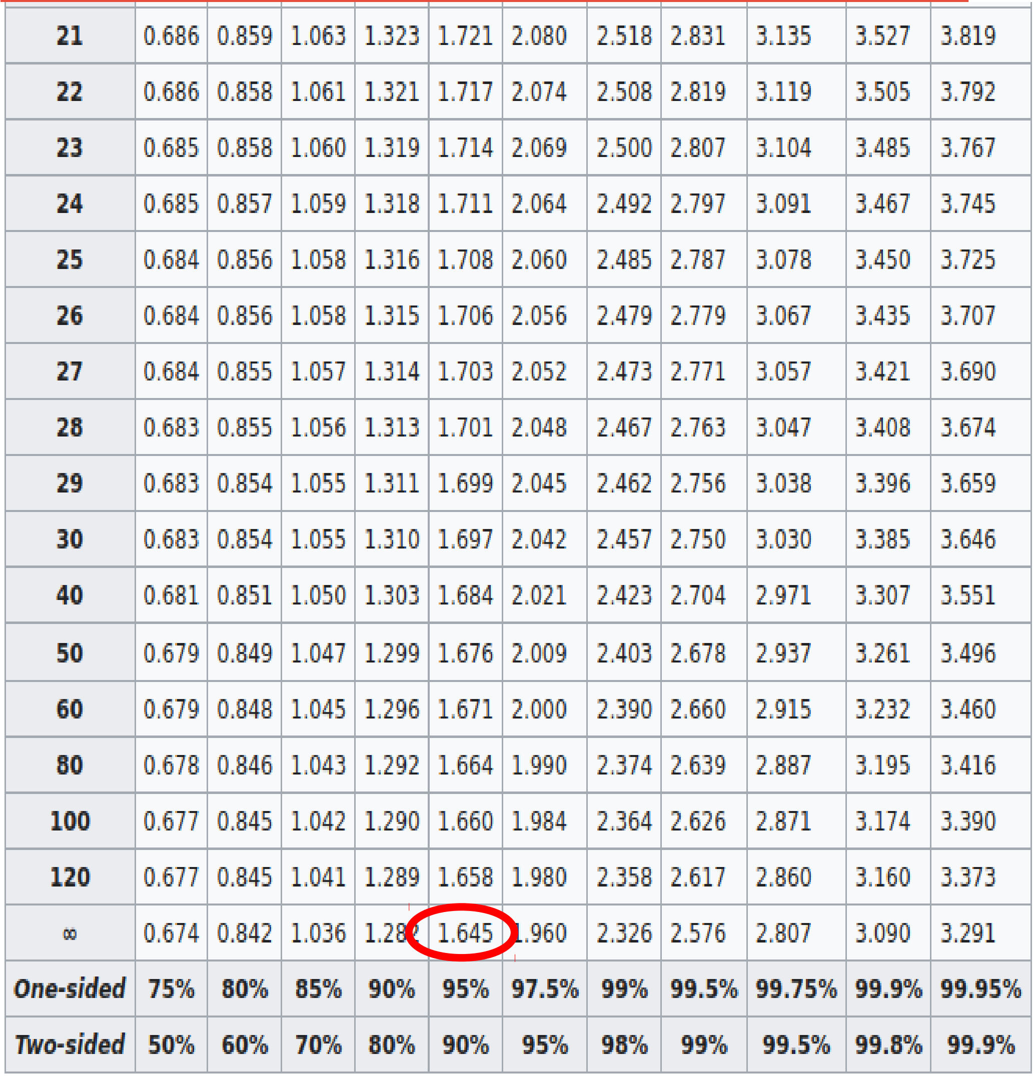
$$H_0: \ \mu \ \leq \ 5(\mu_0)$$
$$H_1: \ \mu \ > \ 5(\mu_0)$$



t critical value for one tailed test for 95% confidence at d.f = 194

Is 1.645

t-table with selected degrees of freedom for one-sided or two-sided critical values

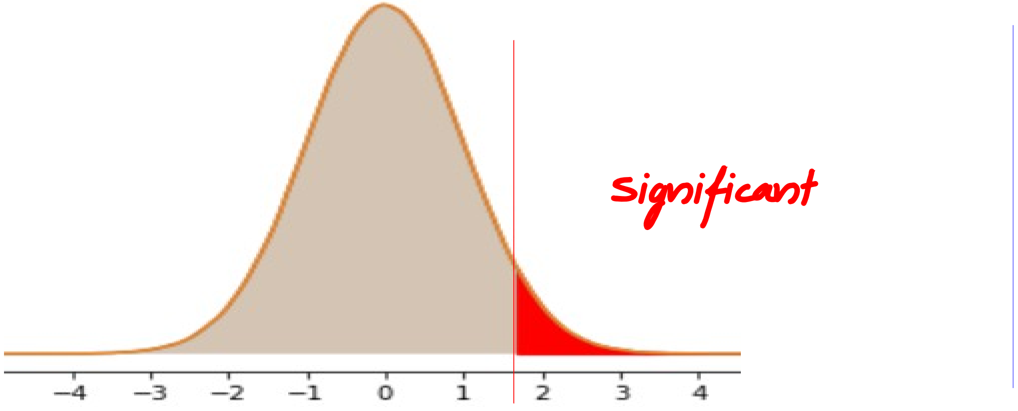
One-sided	**75%**	**80%**	**85%**	**90%**	**95%**	**97.5%**	**99%**	**99.5%**	**99.75%**	**99.9%**	**99.95%**
Two-sided	**50%**	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.5%**	**99.8%**	**99.9%**
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
One-sided	**75%**	**80%**	**85%**	**90%**	**95%**	**97.5%**	**99%**	**99.5%**	**99.75%**	**99.9%**	**99.95%**
Two-sided	**50%**	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.5%**	**99.8%**	**99.9%**

One Sample t test - Two tailed

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

t-statistic(2614.392) > t_c(1.646)

$$H_0: \ \mu \ \leq \ 5(\mu_0)$$
$$H_1: \ \mu \ > \ 5(\mu_0)$$



t statistic falls in the rejection region means it is very unlikely in null hypothesis

Result is statistically significant so we reject Null Hypothesis

One Sample t test - Two tailed

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

The probability of observing a standard normal value above 2614.392 is less than 0.00001

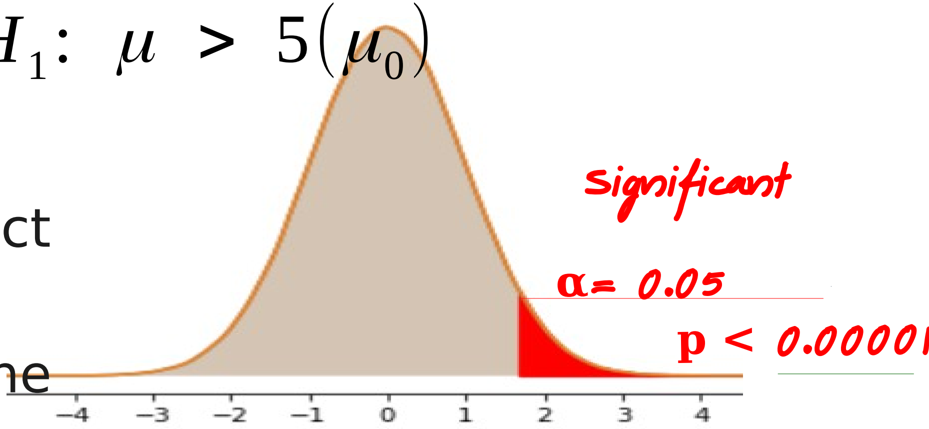
2 tailed p < 0.00001

$$H_0: \ \mu \ \leq \ 5(\mu_0)$$

$$H_1: \ \mu \ > \ 5(\mu_0)$$

Since p(0.00001)< α(0.05)

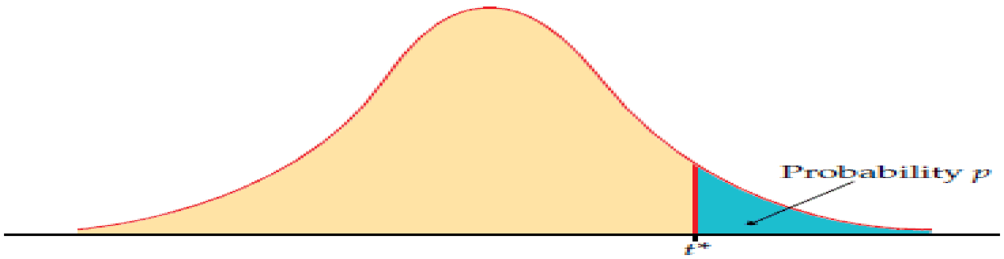
Means result is significant so we can reject Null Hypothesis



The t-test tells us that indicating that the population mean greater 5

at the 0.05 level of significance.

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.



t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

One Sample t test - Right tailed with Confidence Intervals

Confidance Interval:

$$CI = \overline{X} \pm t_{(\alpha, n-1)} \frac{s}{\sqrt{n}}$$

95 %, one-sided confidence interval:

$$CI = \overline{X} \pm 1.645 \frac{s}{\sqrt{n}}$$

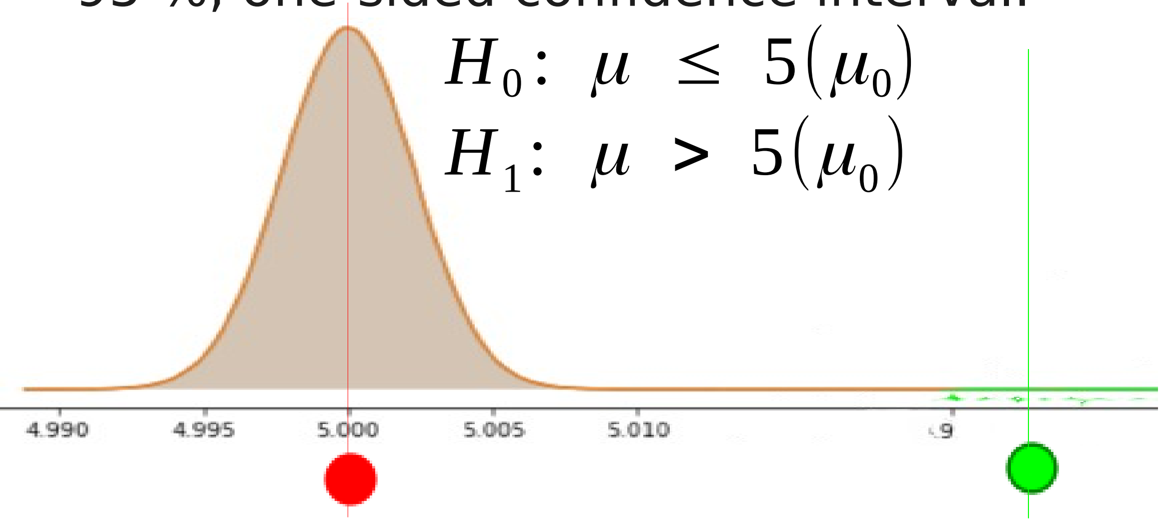
$$CI = 9.261460 \pm 1.645 \frac{0.022789}{\sqrt{195}}$$

$$CI = 9.261460 \pm 1.645 * 0.00163$$

$$CI = 9.261460 \pm 0.00268$$

$$CI = [9.258778 \quad 9.264141]$$

$$CI = [9.258778 \quad \infty]$$



$$H_0: \mu \le 5(\mu_0)$$

$$H_1: \mu > 5(\mu_0)$$

Population Mean 5 is not in the confidence linterval range so

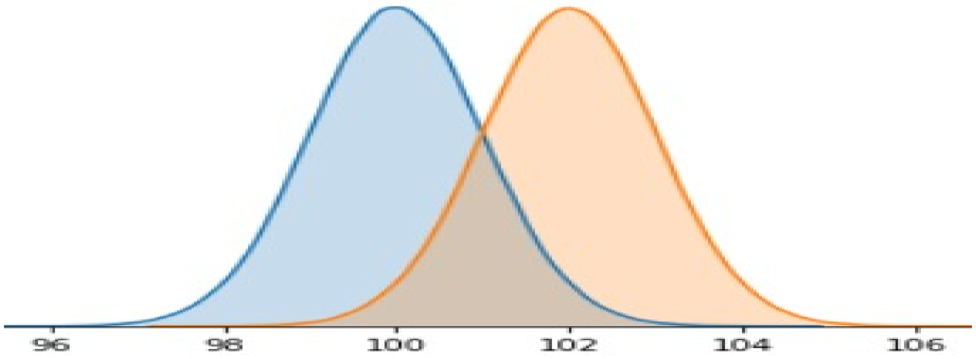
Null Hypothesis can be rejected

Two Sample t test

Two-sample t-test

To compare parameters of two different populations to check for any difference in parameter(mean) values based on sample data

The purpose of the test is to determine whether the difference between these two populations is statistically significant.



Two-sample t-test

Two-sample t-tests for a difference in mean involve independent samples (unpaired samples) or paired samples

1. Independent (unpaired) samples

Two separate sets of independent and identically distributed samples are obtained, one from each of the two populations being compared.

randomly assign 50 subjects to the treatment group and 50 subjects to the control group when evaluating the effect of a medical treatment.

2. Dependent (paired) samples

sample of matched pairs of similar units, or one group of units that has been tested twice (a "repeated measures" t-test).

subjects are tested prior to a treatment, say for high blood pressure, and the same subjects are tested again after treatment with a blood-pressure-lowering medication

Independent two-sample t-test

This test is used when the samples are Independent

If we want to estimate difference in two population means when standard deviations of the populations are unknown

Two types of Independent two-sample t-test

1. Two sample t test with equal variance

This test is used only when it can be assumed that the two distributions have the same variance.

2. Two sample t test with unequal variance

This test, also known as Welch's t-test, is used only when the two population variances are not assumed to be equal.

Two Sample t test with equal variance

Two-sample t-test with equal varience Assumptions

1. The population variances(σ_1^2 & σ_2^2) are unknown

2. The population follow the normal distribution **or** (n_1+n_2 >40)

3. The two samples are Independent

4. The two samples variences are equal($\sigma_1^2 = \sigma_2^{2)}$

Equal variance two-sample t-test t statistic

The sampling distribution of the difference in estimated means $(\overline{X}_1 - \overline{X}_2)$ follows t distribution with (n_1 + n_2 – 2) degrees of freedom with mean $(\mu_1 - \mu_2)$

And standard deviation $\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

Where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ is an estimator of the pooled variance

of two samples. n_1 , n_2 are sample sizes for sample1 and sample2 respectively

Equal variance two-sample t-test t statistic

Statistic $t = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

Here

$\overline{X}_1 = \textit{is the average of Sample 1}$

$\overline{X}_2 = \textit{is the average of Sample 2}$

$s_1 = \textit{is the sample standard deviation of Sample 1}$

$s_2 = \textit{is the sample standard deviation of Sample 2}$

$n_1 = \textit{is samplesize of Sample 1}$

$n_2 = \textit{is samplesize of Sample 2}$

Two-sample t-test Example

Let A_1 denote a set obtained by drawing a random sample of six measurements:

A_1={30.02,29.99,30.11,29.97,30.01,29.99}

and let A_2 denote a second set obtained similarly:

A_2={29.89,29.93,29.72,29.98,30.02,29.98}

These could be, for example, the weights of screws that were chosen out of a bucket.

Test the hypothesis that the means of the populations from which the two samples were taken are equal.

Two-sample t-test with equal variance

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0 : \mu_1 = \mu_2 \qquad H_0 : \mu_1 - \mu_2 = 0$$

or

$$H_1 : \mu_1 \neq \mu_2 \qquad H_1 : \mu_1 - \mu_2 \neq 0$$

Step 2: Choose level of significance(α)

α = 0.05

A value of α = 0.05 implies that the null hypothesis is rejected 5 % of the time when it is in fact true

Two-sample t-test with equal variance

Step 3: Determine appropriate test to use and find the test statistic

1. Here we do not know the population Standard Deviations
2. The two samples are Independent
3. sample size is not large(n_1+n_2 < 40) so Let us assume population is normally distributed
4. Let assume population variances are equal

Two sample t-tests can be performed

Two-sample t-test with equal variance

From given data $\overline{X}_1 = 30.015, \ \overline{X}_2 = 29.92, \ s_1 = 0.05, \ s_2 = 0.11$

$n_1 = 6, n_2 = 6, \ \overline{X}_1 - \overline{X}_2 = 0.095$

Then $s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(6-1)0.05^2 + (6-1)0.11^2}{6+6-2}} \approx 0.08544$

And $d.f = n_1 + n_2 - 2 = 6 + 6 - 2 = 10$

Two-sample t-test with equal variance

t statistic

$$t=\frac{\left(\overline{X}_1-\overline{X}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$t=\frac{\left(30.015-29.92\right)-\left(0\right)}{\sqrt{0.0073\left(\frac{1}{6}+\frac{1}{6}\right)}}$$

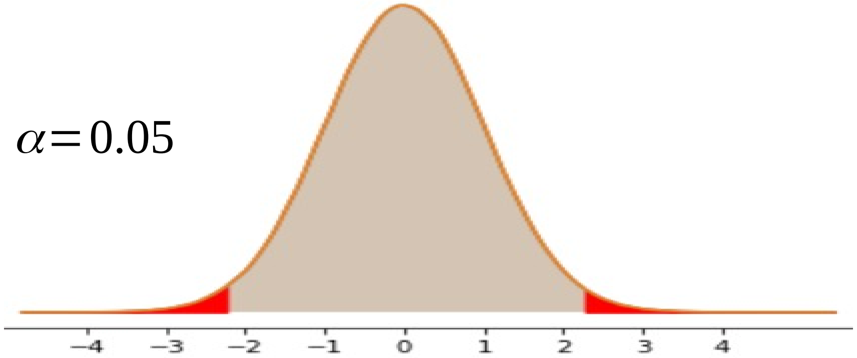
$$t\approx 1.938$$

Two-sample t-test with equal variance

Step 4: Determine if we need a 1 tailed or a 2 tailed t-critical value and find the t-critical value for desired confidence level

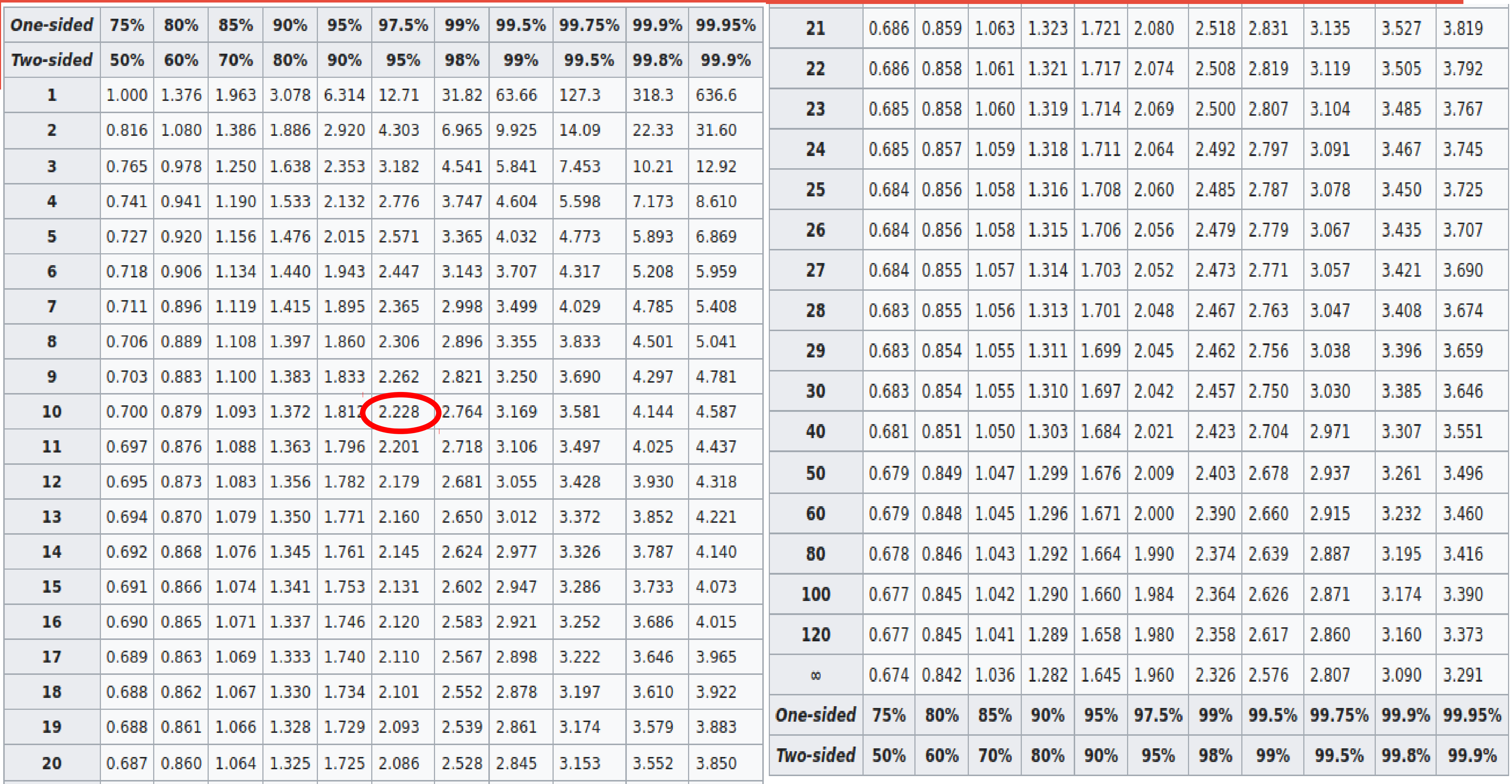
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

$\alpha = 0.05$



The critical value of t for $\alpha = 0.05$ and d.f = 10 for two tailed test is +/- 2.228

t-table with selected degrees of freedom for one-sided or two-sided critical values



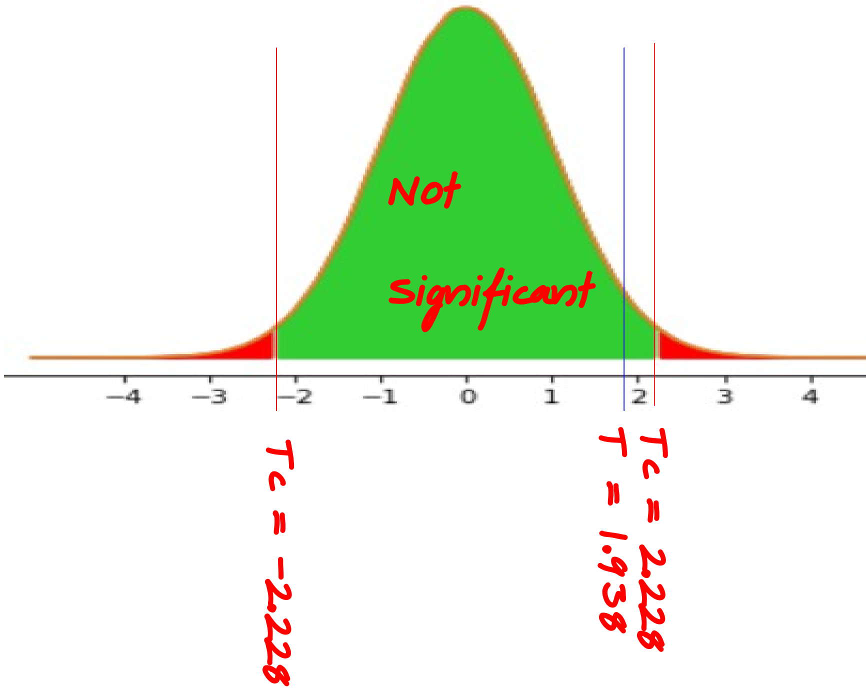
One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%

Two-sample t-test with equal variance

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

$t_c(-2.228) < \text{t-statistic}(1.938) < t_c(2.228)$

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



t statistic do not falls in the rejection region means it is very likely in null hypothesis

Result is not statistically significant so we can’t reject Null Hypothesis

Two-sample t-test with equal variance

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

The probability of observing a standard normal value above 1.938 is 0.04.

2 tailed p= 0.08 = 2*0.04

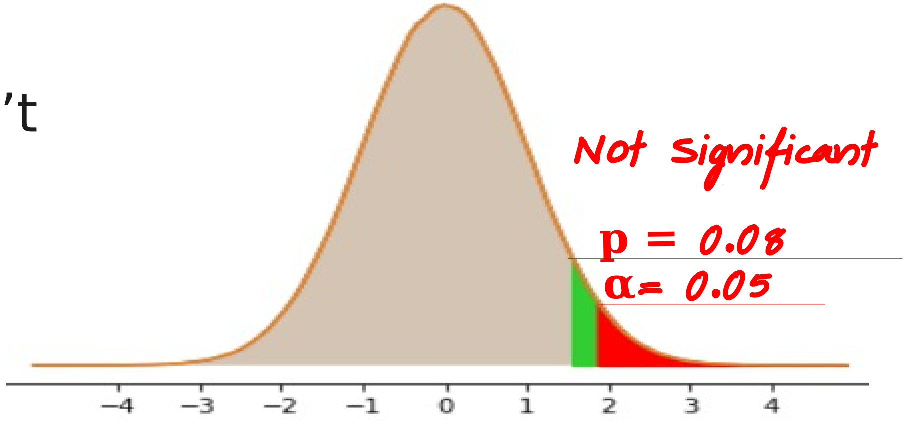
$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Since p(0.08) > α(0.05)

Means result is not significant so we can't reject Null Hypothesis

The population means are equal

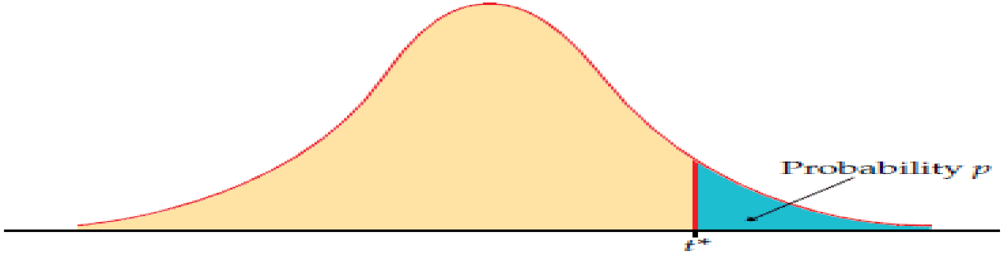


Not significant

p = 0.08

α= 0.05

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.



t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Two-sample t-test with equal variance

The first sample is miles per gallon for U.S. cars and the second sample is miles per gallon for Japanese cars.

the summary statistics for each sample are shown below

sample	Sample size	Mean	Sample Standard deviation
Sample1(US)	249	20.14458	6.41470
Sample2(Japan)	79	30.48101	6.10771

Test the hypothesis that US cars have significantly lower fuel economy than the Japan cars at 95% confidence level.

Two-sample t-test with equal variance

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0 : \mu_1 \geq \mu_2$$
$$H_1 : \mu_1 < \mu_2$$

or

$$H_0 : \mu_1 - \mu_2 \geq 0$$
$$H_1 : \mu_1 - \mu_2 < 0$$

Step 2: Choose level of significance(α)

α = 0.05

A value of α = 0.05 implies that the null hypothesis is rejected 5 % of the time when it is in fact true

Two-sample t-test with equal variance

Step 3: Determine appropriate test to use and find the test statistic

1. Here we do not know the population Standard Deviations
2. The two samples are Independent
3. Sample size is large(n_1+n_2 > 40)
4. Let assume population variances are equal

Two sample t-tests can be performed

Two-sample t-test with equal variance

Given data $\overline{X}_1 = 20.14458, \ \overline{X}_2 = 30.48101, \ s_1 = 6.41470, \ s_2 = 6.10771$

$n_1 = 249, n_2 = 79, \ \overline{X}_1 - \overline{X}_2 = -10.33643$

Then $s_p = \sqrt{\dfrac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}} = \sqrt{\dfrac{(249 - 1) 6.41470^2 + (79 - 1) 6.10771^2}{249 + 79 - 2}} \approx 6.34260$

And $d.f = n_1 + n_2 - 2 = 249 + 79 - 2 = 326$

Two-sample t-test with equal variance

t statistic

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$t = \frac{(20.14458 - 30.48101) - (0)}{\sqrt{40.228\left(\frac{1}{249} + \frac{1}{79}\right)}}$$

$$t = -\frac{10.33643}{0.819}$$

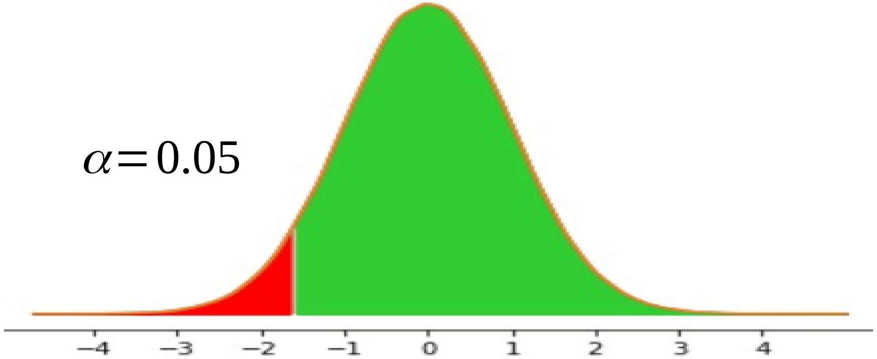
$$t \approx -12.62059$$

Two-sample t-test with equal variance

Step 4: Determine if we need a 1 tailed or a 2 tailed t-critical value and find the t-critical value for desired confidence level

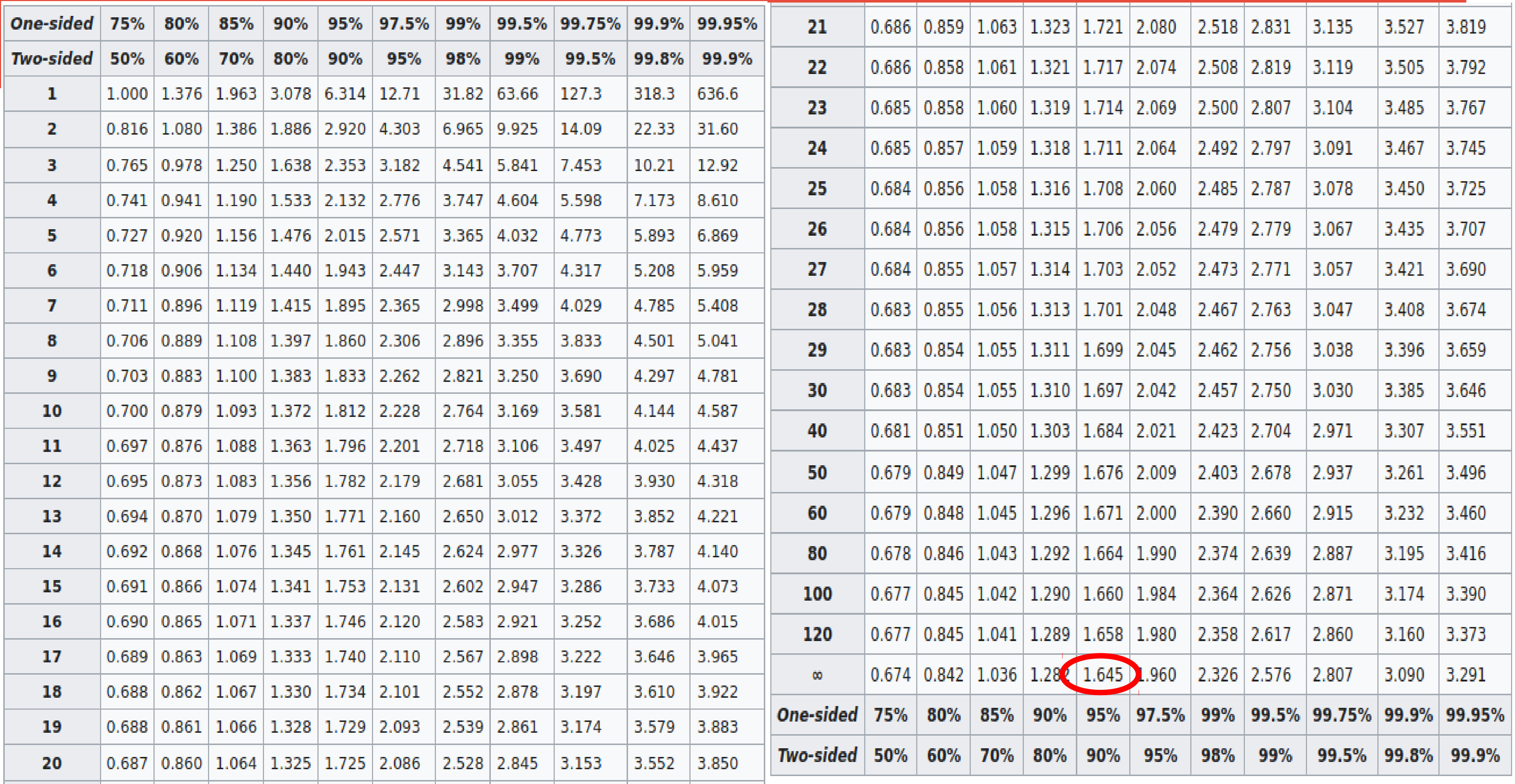
$$H_0 : \mu_1 \geq \mu_2$$
$$H_1 : \mu_1 < \mu_2$$

$\alpha = 0.05$



The critical value of t for $\alpha = 0.05$ and d.f = 326 for lower tail is -1.645

t-table with selected degrees of freedom for one-sided or two-sided critical values



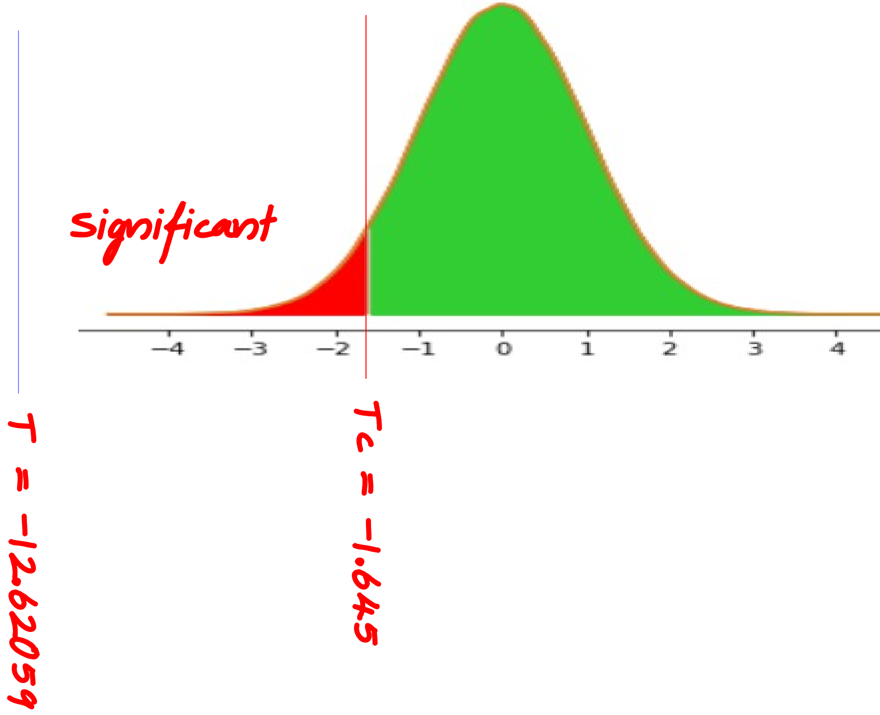
One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	**50%**	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.5%**	**99.8%**	**99.9%**
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
One-sided	**75%**	**80%**	**85%**	**90%**	**95%**	**97.5%**	**99%**	**99.5%**	**99.75%**	**99.9%**	**99.95%**
Two-sided	**50%**	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.5%**	**99.8%**	**99.9%**

Two-sample t-test with equal variance

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

t-statistic(-12.62059) < t_c(-1.645)

$$H_0 : \mu_1 \geq \mu_2$$
$$H_1 : \mu_1 < \mu_2$$



t statistic falls in the rejection region means it is very unlikely in null hypothesis

Result is statistically significant so we can reject Null Hypothesis at 95% confidence

Two-sample t-test with equal variance

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

The probability of observing a standard normal value below -12.62059 is <0.00001

$$H_0: \mu_1 \geq \mu_2$$
$$H_1: \mu_1 < \mu_2$$

Since p(0.00001) < α(0.05)

Means result is significant so we can reject Null Hypothesis

The t-test tells us that us cars have low fuel economy

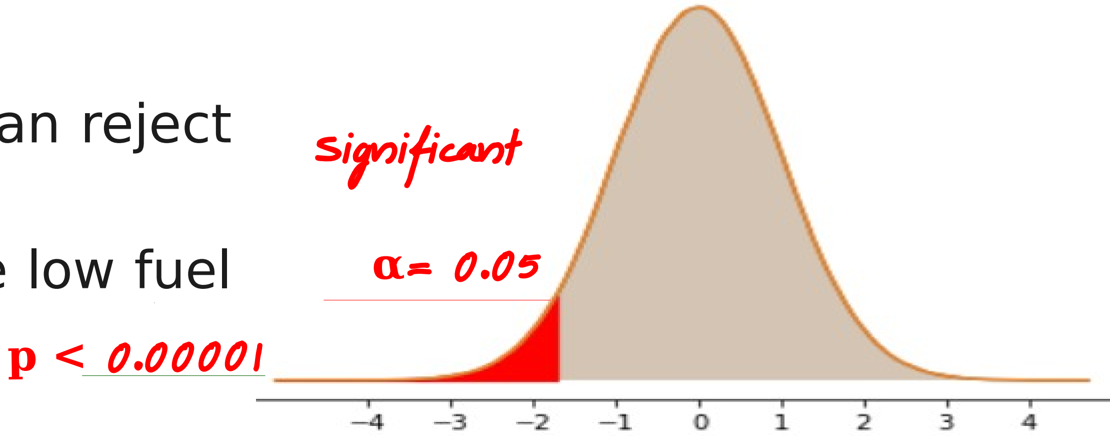
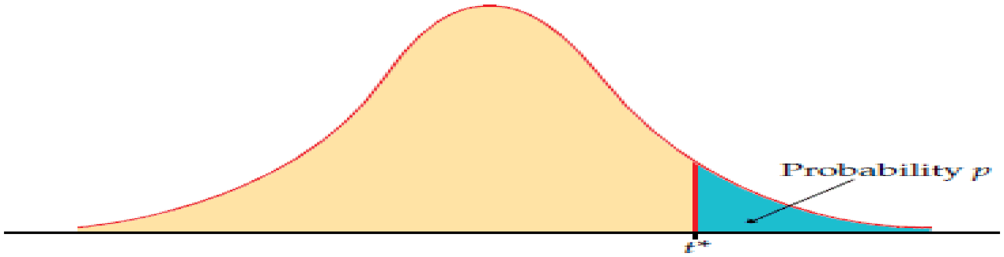


Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.



t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Two Sample t test with unequal variances (Welch’s t test)

Two-sample t-test with unequal varience Assumptions

1. The population variances(σ_1^2 & σ_2^2) are unknown

2. The population follow the normal distribution **or** ($n_1+n_2 >40$)

3. The two samples are Independent

4. The two samples variences are not equal $\sigma_1^2 \neq \sigma_2^2$

Unequal variance two-sample t-test t statistic

The sampling distribution of the difference in estimated means $(\overline{X}_1 - \overline{X}_2)$

follows t distribution with mean $(\mu_1 - \mu_2)$ and standard deviation $s_u = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

And the degree of freedom is given by $d.f = \left\lfloor \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}} \right\rfloor$

Unequal variance two-sample t-test t statistic

Statistic $$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Here

$\overline{X}_1 \;=\;$ *is the average of Sample* 1

$\overline{X}_2 \;=\;$ *is the average of Sample* 2

$s_1 \;=\;$ *is the sample standard deviation of Sample* 1

$s_2 \;=\;$ *is the sample standard deviation of Sample* 2

$n_1 \;=\;$ *is samplesize of Sample* 1

$n_2 \;=\;$ *is samplesize of Sample* 2

Two-sample t-test Example

Let A_1 denote a set obtained by drawing a random sample of six measurements:

$A_1=\{30.02,29.99,30.11,29.97,30.01,29.99\}$

and let A_2 denote a second set obtained similarly:

$A_2=\{29.89,29.93,29.72,29.98,30.02,29.98\}$

These could be, for example, the weights of screws that were chosen out of a bucket.

Test the hypothesis that the means of the populations from which the two samples were taken are equal.

Two-sample t-test with unequal variance

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0 : \mu_1 = \mu_2 \qquad H_0 : \mu_1 - \mu_2 = 0$$

or

$$H_1 : \mu_1 \neq \mu_2 \qquad H_1 : \mu_1 - \mu_2 \neq 0$$

Step 2: Choose level of significance(α)

α = 0.05

A value of α = 0.05 implies that the null hypothesis is rejected 5 % of the time when it is in fact true

Two-sample t-test with unequal variance

Step 3: Determine appropriate test to use and find the test statistic

1. Here we do not know the population Standard Deviations
2. The two samples are Independent
3. sample size is not large(n_1+n_2 < 40) so Let us assume population is normally distributed
4. Let assume population variances are not equal

Two sample t-test with unequal varianes can be performed

Two-sample t-test with unequal variance

From given data $\overline{X}_1 = 30.015, \ \overline{X}_2 = 29.92, \ s_1 = 0.05, \ s_2 = 0.11$

$n_1 = 6, n_2 = 6, \ \overline{X}_1 - \overline{X}_2 = 0.095$

Then $$s_u = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{0.05^2}{6} + \frac{0.11^2}{6}} \approx 0.04849$$

And $$d.f = \left\lfloor \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}} \right\rfloor = \left\lfloor \frac{\left(\frac{0.05^2}{6} + \frac{0.11^2}{6}\right)^2}{\frac{\left(0.05^2/6\right)^2}{6 - 1} + \frac{\left(0.11^2/6\right)^2}{6 - 1}} \right\rfloor \approx \lfloor 7.031 \rfloor = 7$$

Two-sample t-test with unequal variance

t statistic

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = \frac{(0.095) - (0)}{0.04849}$$

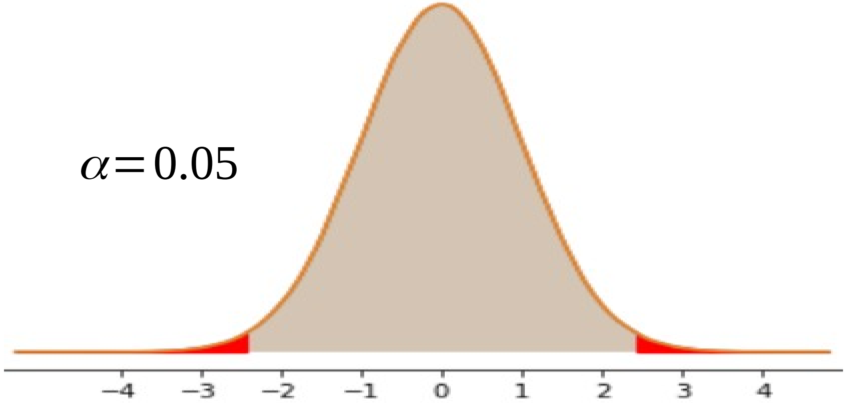
$$t \approx 1.959$$

Two-sample t-test with unequal variance

Step 4: Determine if we need a 1 tailed or a 2 tailed t-critical value and find the t-critical value for desired confidence level

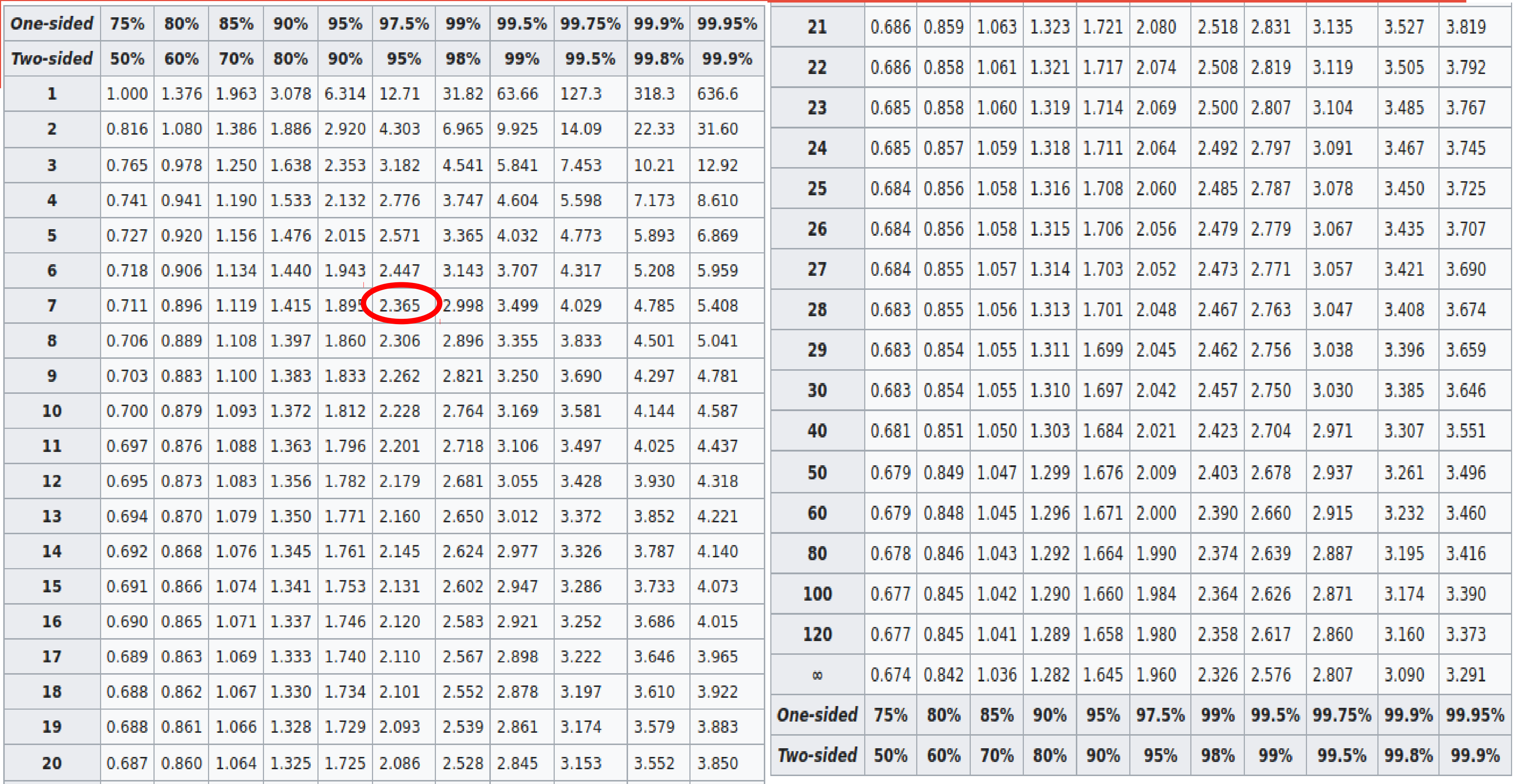
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

$\alpha = 0.05$



The critical value of t for $\alpha = 0.05$ and d.f = 7 for two tailed test is +/- 2.365

t-table with selected degrees of freedom for one-sided or two-sided critical values



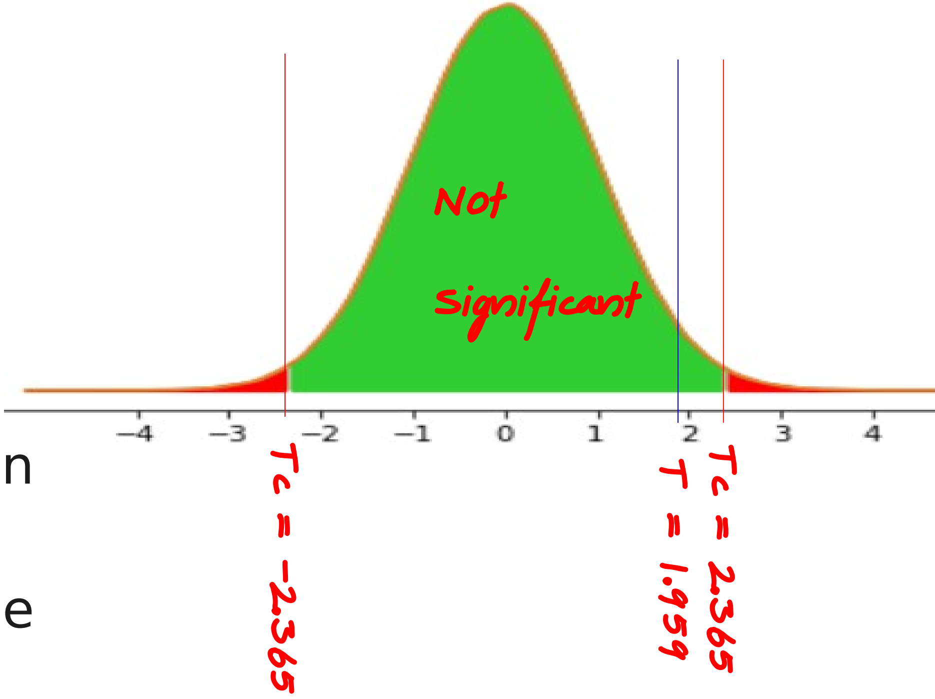
One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%

Two-sample t-test with unequal variance

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

$t_c(-2.365) < \text{t-statistic}(1.959) < t_c(2.365)$

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



t statistic do not falls in the rejection region means it is very likely in null hypothesis

Result is not statistically significant so we can’t reject Null Hypothesis

Two-sample t-test with unequal variance

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

The probability of observing a standard normal value above 1.959 is 0.045.

2 tailed p= 0.09 = 2*0.045

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Since p(0.09) > α(0.05)

Means result is not significant so we can’t reject Null Hypothesis

The t-test tells us that population means are equal

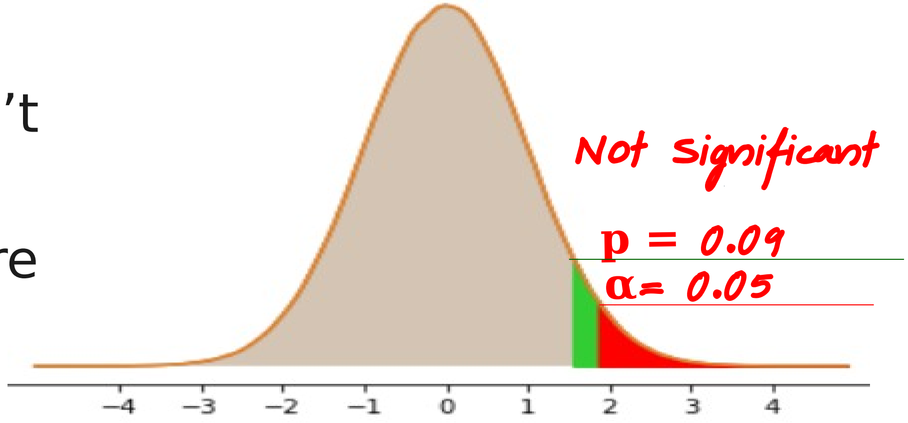
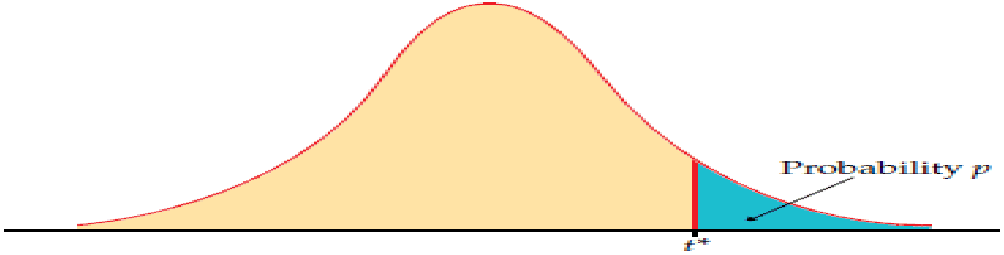


Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.



t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Two-sample t-test with unequal variance

The first sample is miles per gallon for U.S. cars and the second sample is miles per gallon for Japanese cars.

the summary statistics for each sample are shown below

sample	Sample size	Mean	Sample Standard deviation
Sample1(US)	249	20.14458	6.41470
Sample2(Japan)	79	30.48101	6.10771

Test the hypothesis that US cars have significantly lower fuel eonomy than the Japan cars at 95% confidence level.

Two-sample t-test with unequal variance

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0 : \mu_1 \geq \mu_2$$
$$H_1 : \mu_1 < \mu_2$$

or

$$H_0 : \mu_1 - \mu_2 \geq 0$$
$$H_1 : \mu_1 - \mu_2 < 0$$

Step 2: Choose level of significance(α)

α = 0.05

A value of α = 0.05 implies that the null hypothesis is rejected 5 % of the time when it is in fact true

Two-sample t-test with unequal variance

Step 3: Determine appropriate test to use and find the test statistic

1. Here we do not know the population Standard Deviations
2. The two samples are Independent
3. Sample size is large($n_1+n_2 > 40$)
4. Let assume population variances are not equal

Two sample t-test with unequal varianes can be performed

Two-sample t-test with unequal variance

Given data $\overline{X}_1 = 20.14458,\ \overline{X}_2 = 30.48101,\ s_1 = 6.41470,\ s_2 = 6.10771$

$n_1 = 249, n_2 = 79,\ \overline{X}_1 - \overline{X}_2 = -10.33643$

$$s_u = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{6.61470^2}{249} + \frac{6.10771^2}{79}} = \sqrt{\frac{41.1484}{249} + \frac{37.3041}{79}} = \sqrt{0.1652 + 0.4722} \approx 0.7983$$

$$d.f = \left\lfloor \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \right\rfloor = \left\lfloor \frac{\left(\frac{6.61470^2}{249} + \frac{6.10771^2}{79}\right)^2}{\frac{(6.61470^2/249)^2}{249 - 1} + \frac{(6.10771^2/79)^2}{79 - 1}} \right\rfloor = \left\lfloor \frac{0.6373}{0.00011 + 0.00286} \right\rfloor$$

$$d.f. \approx \lfloor 214.58 \rfloor = 214$$

Two-sample t-test with unequal variance

t statistic

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

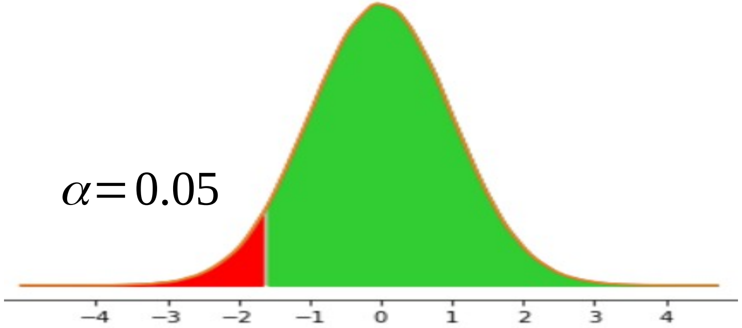
$$t = \frac{(-10.33643) - (0)}{0.7983}$$

$$t \approx -12.94$$

Two-sample t-test with uequal variance

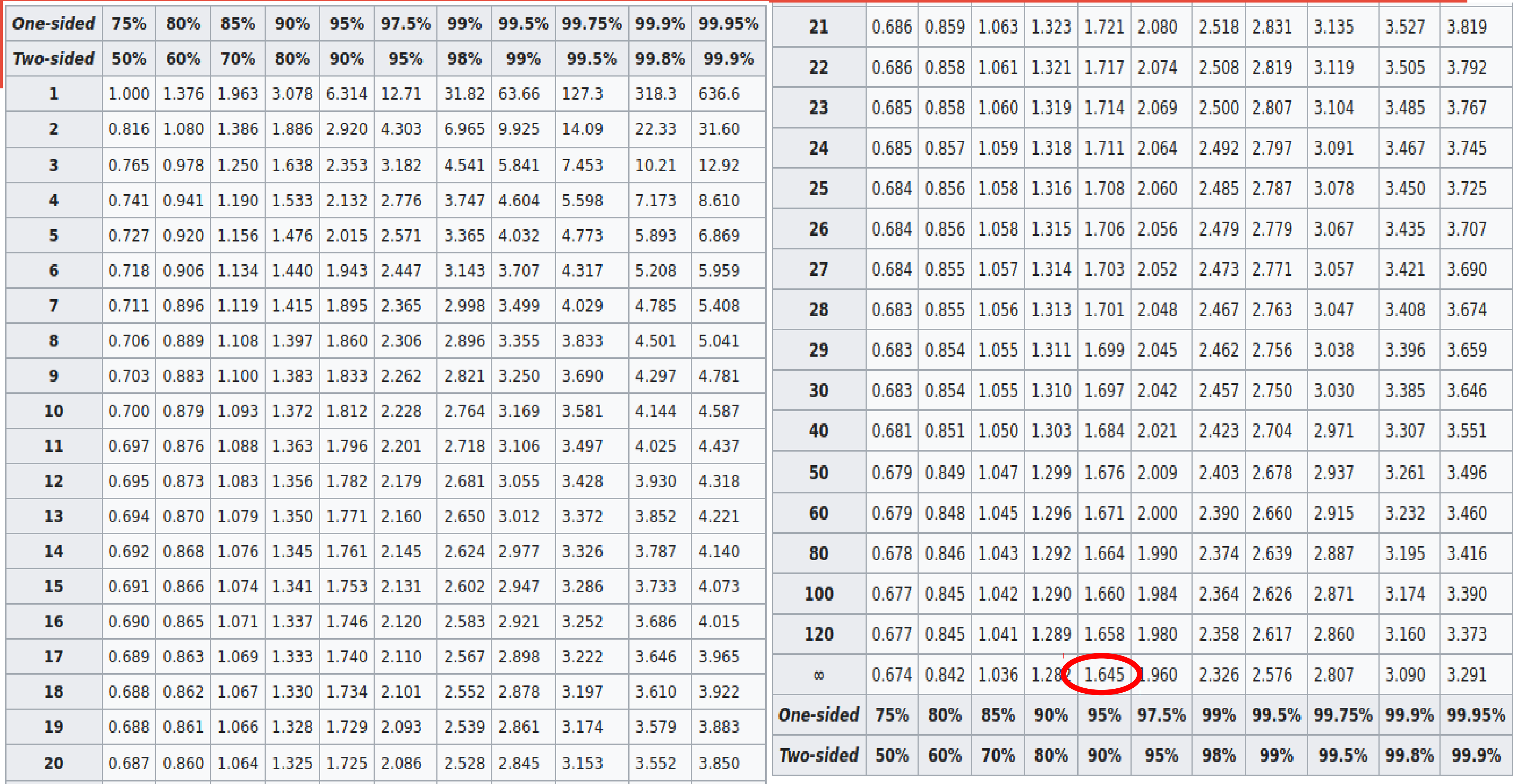
Step 4: Determine if we need a 1 tailed or a 2 tailed t-critical value and find the t-critical value for desired confidence level

$$H_0: \mu_1 \geq \mu_2$$
$$H_1: \mu_1 < \mu_2$$



The critical value of t for $\alpha = 0.05$ and d.f = 214 for lower tail is -1.645

t-table with selected degrees of freedom for one-sided or two-sided critical values



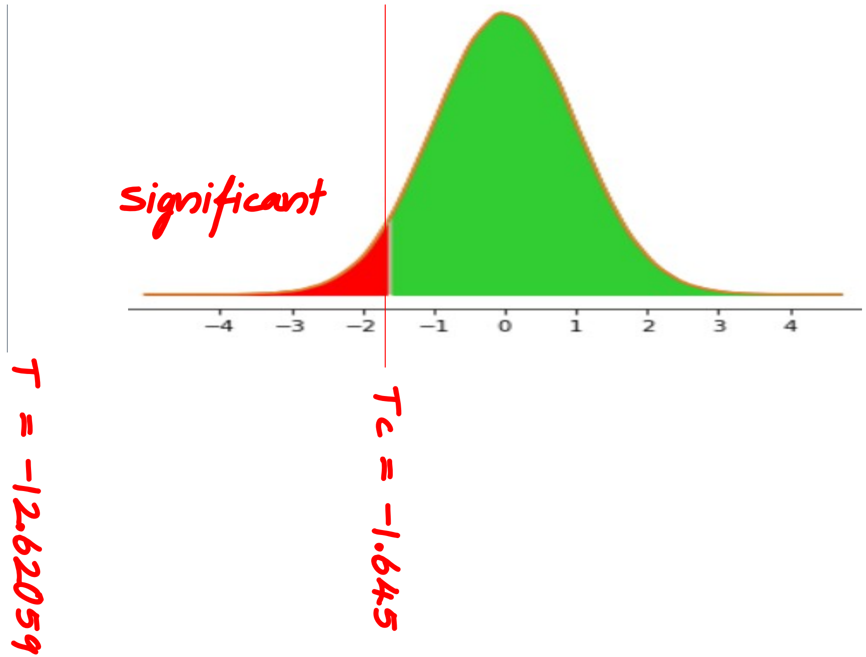
One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	**50%**	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.5%**	**99.8%**	**99.9%**
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
One-sided	**75%**	**80%**	**85%**	**90%**	**95%**	**97.5%**	**99%**	**99.5%**	**99.75%**	**99.9%**	**99.95%**
Two-sided	**50%**	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.5%**	**99.8%**	**99.9%**

Two-sample t-test with unequal variance

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

t-statistic(-12.94) < t_c(-1.645)

$$H_0 : \mu_1 \geq \mu_2$$
$$H_1 : \mu_1 < \mu_2$$



t statistic falls in the rejection region means it is very unlikely in null hypothesis

Result is statistically significant so we can reject Null Hypothesis at 95% confidence

Two-sample t-test with unequal variance

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

The probability of observing a standard normal value below -12.94 is <0.00001

$$H_0 : \mu_1 \geq \mu_2$$
$$H_1 : \mu_1 < \mu_2$$

Since p(0.00001) < α(0.05)

Means result is significant so we can reject Null Hypothesis

The t-test tells us that us cars have low fuel economy

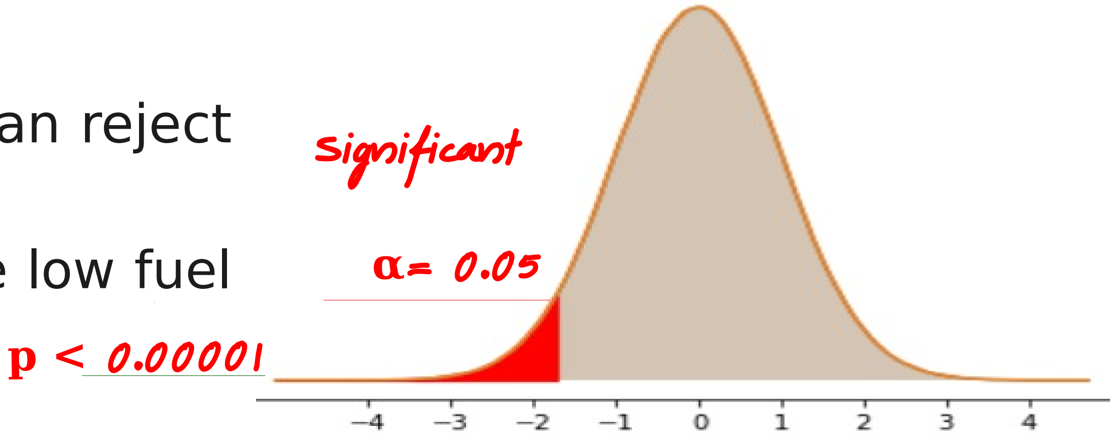
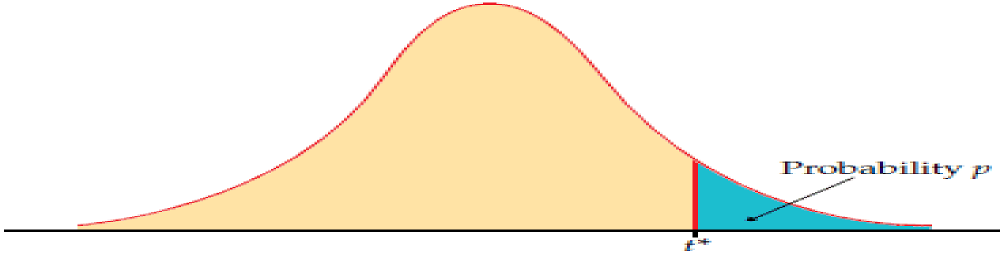


Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.



t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Paired Sample t test

Dependent t-test for paired samples

This test is used when the samples are dependent

That is, when there is only one sample that has been tested twice (repeated measures) or when there are two samples that have been matched or "paired".

1. Math scores befor attending course after completion of course

2. Between-pairs of persons matched into meaningful groups (for instance drawn from the same family or age group.

Example of repeated measures

S.No	Name	Test1	Test2
1	Mike	35%	67%
2	Melnie	50%	46%
3	Melissa	90%	86%
4	Mitchell	78%	91%

Example of matched pairs

Pair	Name	Age	Test
1	Jhon	35	250
1	Jane	36	340
2	Jimmy	22	460
2	Jessy	21	200

Paired sample t – test examples

1. Sugar levels before and after treatment.
2. Heartrate before drinking a cup of coffee and after cup of coffee
3. Alcohol consumtion before and after marriage
4. Math scores befor attending course and after completion of course.
5. Test scores based on age groups.

Paired sample t – test t statistic

t statistic is calculated as $t = \frac{\overline{X}_d - \mu_d}{s_d \;/ \sqrt{n}}$

Statistic t follows a t distribution with degrees of freedom with (n-1)

Where $\overline{X}_d \;=\; the\ mean\ difference\ between\ pairs$

$s_d \;=\; the\ standard\ deviation\ difference\ between\ pairs$

$\mu_d \;=\; the\ hypothesized\ mean\ difference$

Here Assumtion is $\overline{X}_d$ follows a normal distribution

Paired sample t – test Example

Math scores befor attending course and after completion of course

S.No	Name	Test1	Test2
1	Mike	35%	67%
2	Melnie	50%	46%
3	Melissa	90%	86%
4	Mitchell	78%	91%

Conduct appropriate hypothesis test to check weather the average math score is more after completion of course(i.e $\mu_d > 0$) at 95% confidence.

Paired sample t – test Example

From the given data

$$\begin{array}{ll} \overline{X_1} = 63.25 & s_1 = 25.21 \\ \overline{X_2} = 72.50 & s_2 = 20.47 \\ \overline{X_d} = 9.25 & s_d = 4.74 \end{array}$$

Paired sample t – test Example

Step 1: Formulate Null Hypothesis and Alternate Hypothesis

$$H_0 : \mu_d \leq 0$$
$$H_1 : \mu_d > 0$$

Step 2: Choose level of significance(α)

α = 0.05

A value of α = 0.05 implies that the null hypothesis is rejected 5 % of the time when it is in fact true

Paired sample t – test Example

Step 3: Determine appropriate test to use and find the test statistic

1. Here we do not know the population Standard Deviation
2. sample size is not large(<30) so Assume population difference is normally distributed

Paired t-tests can be performed

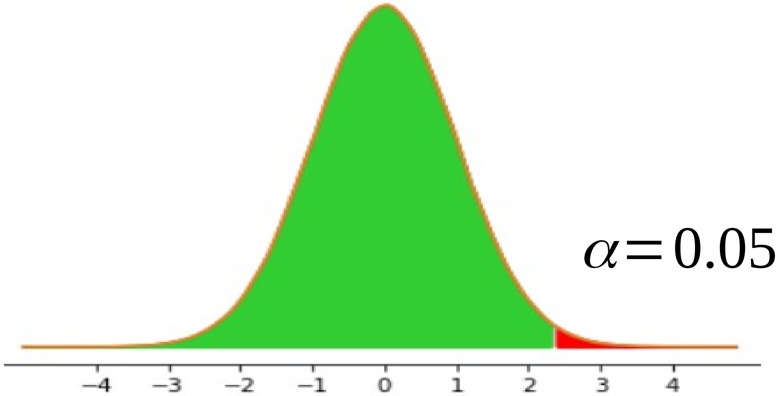
Calculate the t-statistic

$$Statistic \quad t = \frac{\overline{X}_d - \mu_d}{s_d \; / \sqrt{n}} = \frac{9.25 - 0}{4.74 / \sqrt{4}} = \frac{9.25}{2.37} \approx 3.9$$

Paired sample t – test Example

Step 4: Determine if we need a 1 tailed or a 2 tailed t-critical value and find the t-critical value for desired confidence level

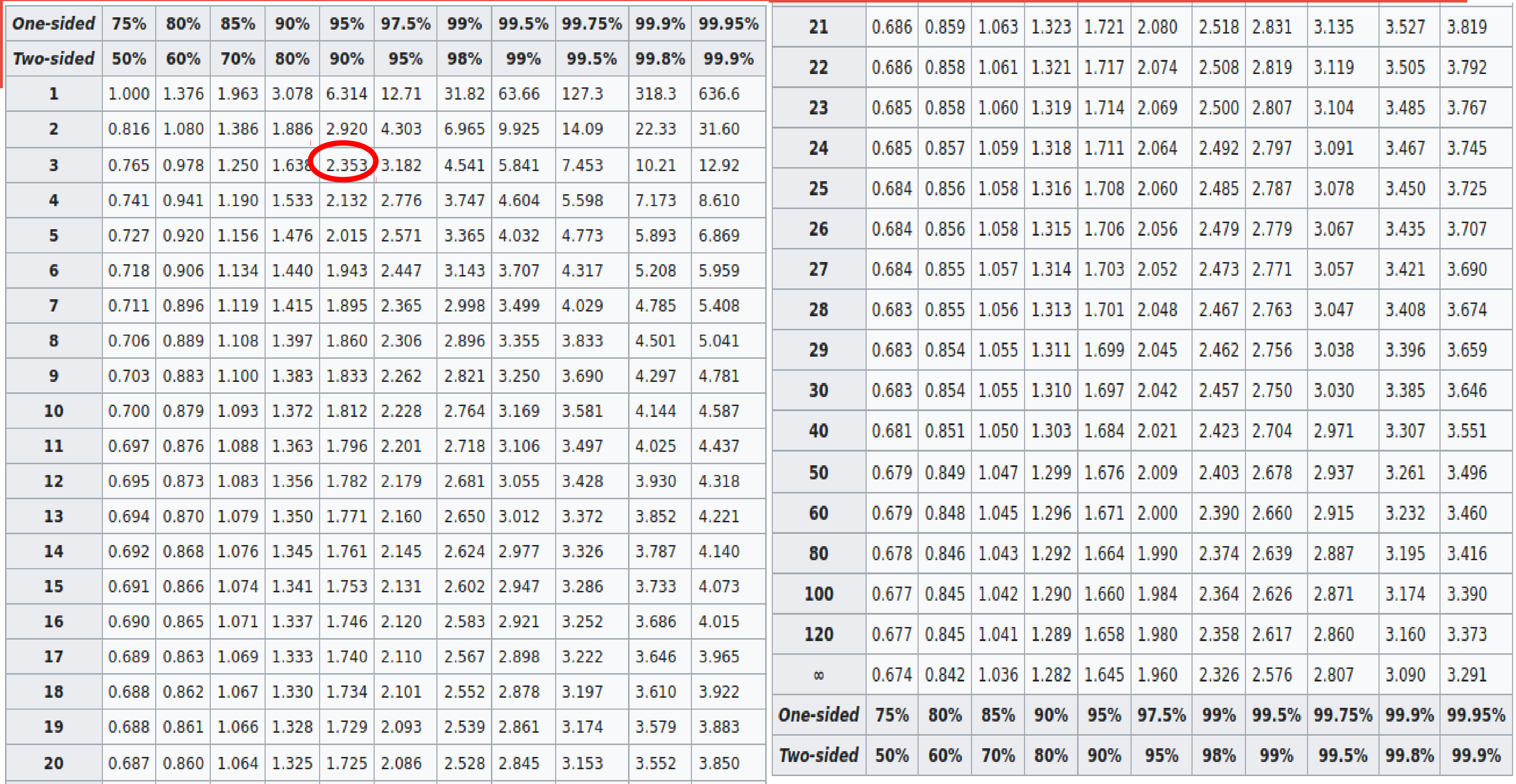
$$H_0 : \mu_d \le 0$$
$$H_1 : \mu_d > 0$$



$\alpha = 0.05$

The critical value of t for $\alpha = 0.05$ and $d.f = 4 - 1 = 3$ for right tailed test is 2.353

t-table with selected degrees of freedom for one-sided or two-sided critical values



One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291
One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%

Paired sample t – test Example

Step 5: Compare test statistic with t-critical value for desired confidence level (Or) Find the p value for test statistic compare with significance level(α)

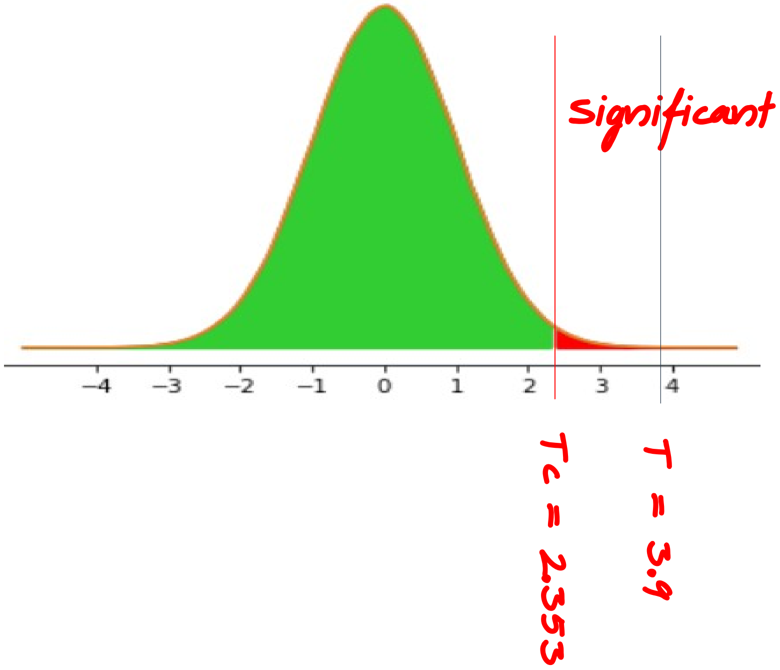
T-statistic(3.9) > t_c(2.353)

$$H_0 : \mu_d \leq 0$$
$$H_1 : \mu_d > 0$$

t statistic falls in the rejection region means it is very unlikely in null hypothesis

Result is statistically significant so we can reject Null Hypothesis at 95% confidence

Means after completion of course average math score is more


significant
Tc = 2.353
T = 3.9