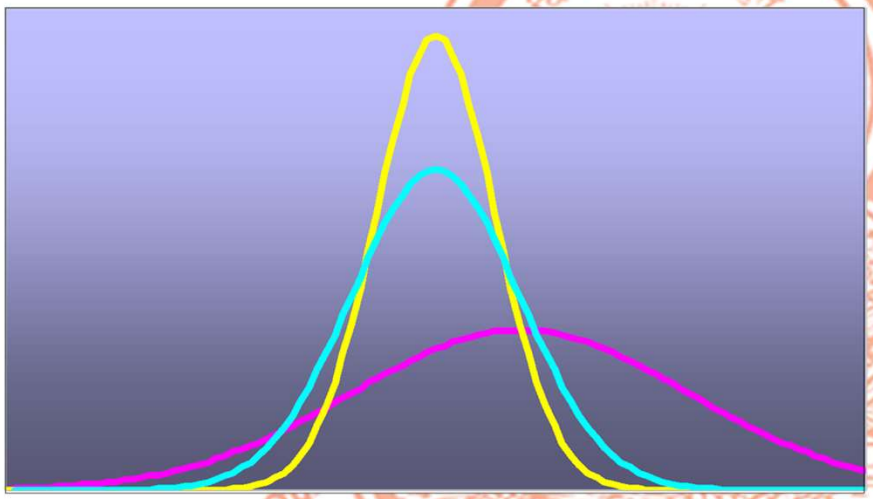
Course Name :Basic Statistics using GUI-R (RKWard)
Module : Z-Table (Practical Demonstration of Normal Distribution)
Week 3 Lecture : 1

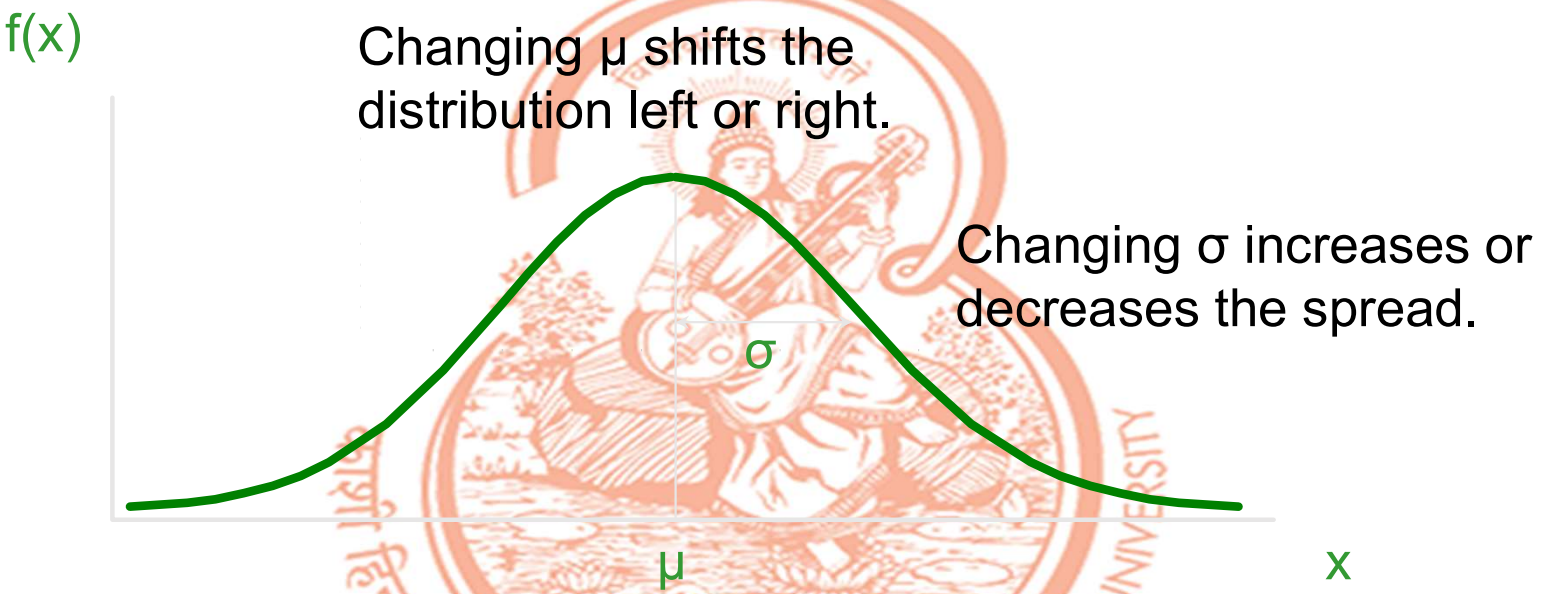
Harsh Pradhan, Assistant Professor,
Institute of Management Studies, BHU
https://bhu.ac.in/Site/FacultyProfile/1_5?FA000562

Many Normal Distributions



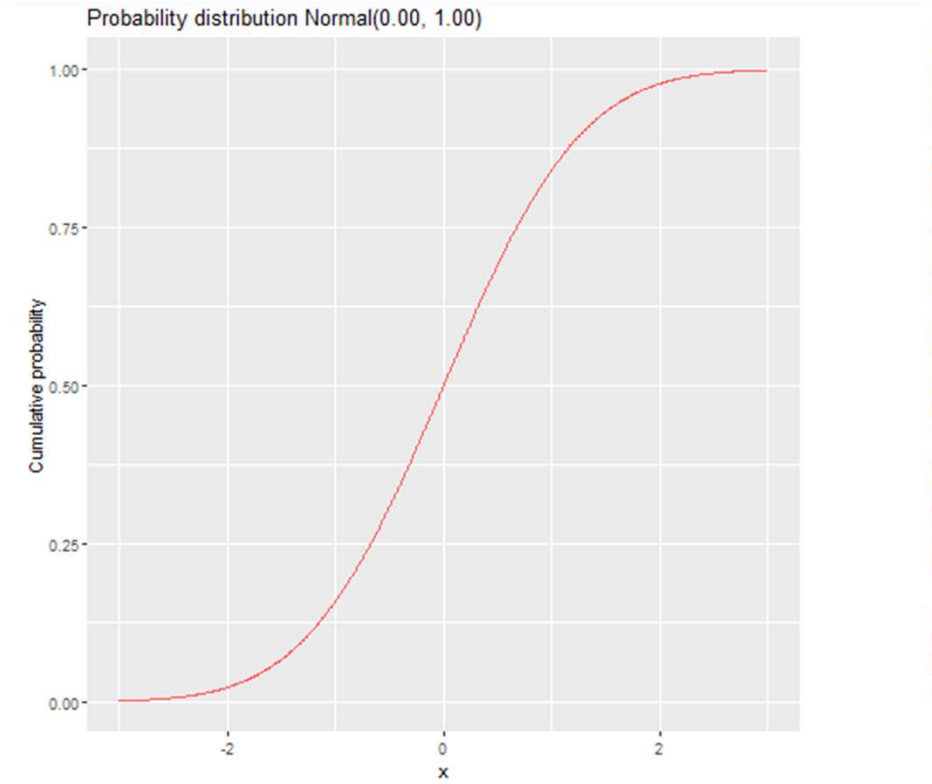
By varying the parameters μ and σ, we obtain different normal distributions

The Normal Distribution Shape

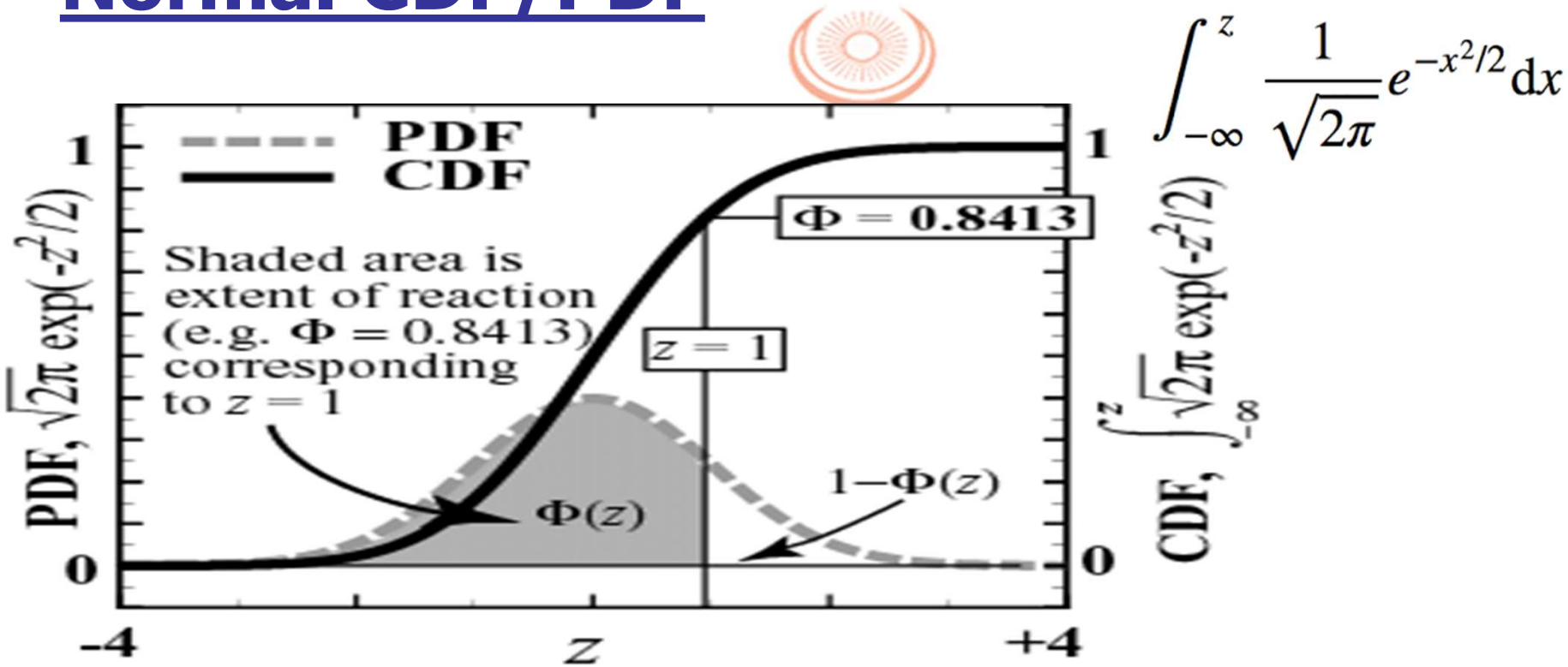


Given the mean μ and variance σ we define the normal distribution using the notation

$$X \sim N(\mu, \sigma^2)$$



Normal CDF/PDF



$$\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx$$

CDF

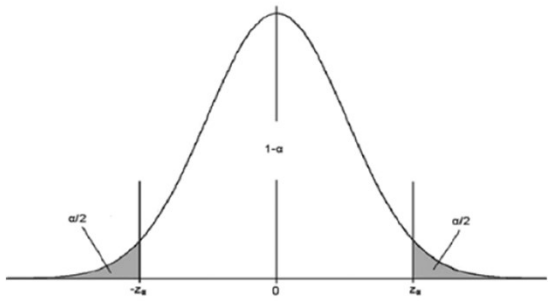
$$F(X,\mu,\sigma) = Pr(-\infty < x \le X) = \int_{-\infty}^{X} f(x,\mu,\sigma)dx$$

where

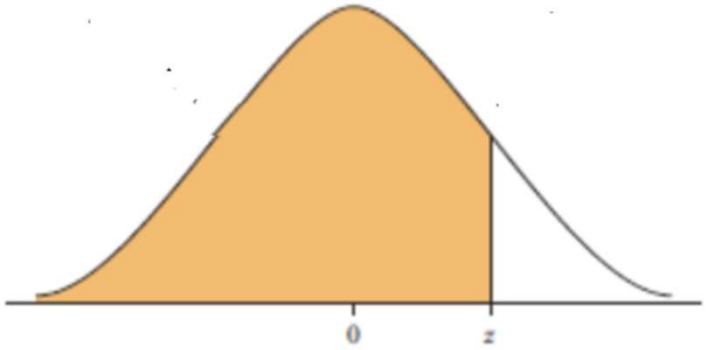
PDF

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

Different Z Table



R 06



R 12

Prob Value

R 07

Z Table based on CDF

Which one is this, among three ?

Z	Second Decimal Place in Z 0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.10	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.20	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.30	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.90	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.00	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.10	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.20	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
2.00	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
3.00	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.40	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.50	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998

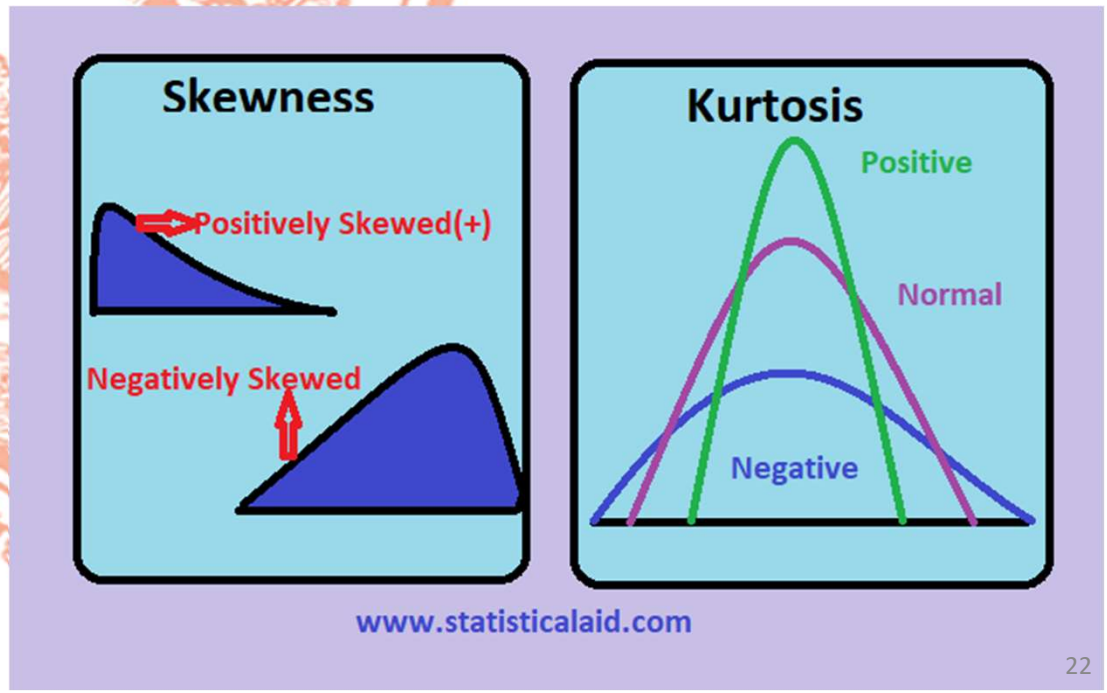
- Example : The grades of an examination are known to be normally distributed with a mean of 60 and standard deviation of 8.
- If the top 10% qualify for a scholarship, what is the minimum grade for this scholarship?
- What will be the percentage of students failing in the examination if the minimum grade for passing is 40?

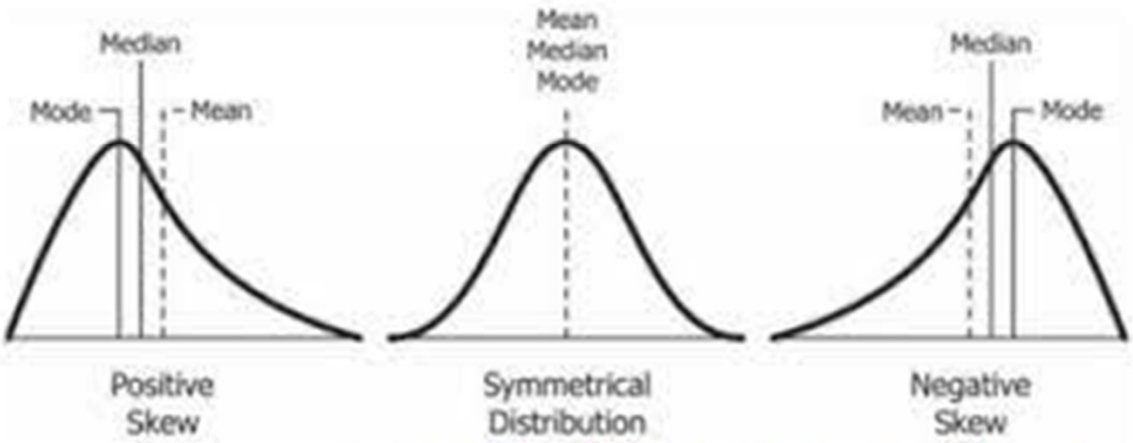
MEASURING DIVERGENCE FROM NORMALITY

- SKEWNESS
  - Skewness is the degree of asymmetry of the distribution
  - Properties of the skewed distribution
  - Positively skewed and negatively skewed
  - Measuring of skewness

- Lack of Symmetry Skewness
- Lack of Pointyness Kurtosis

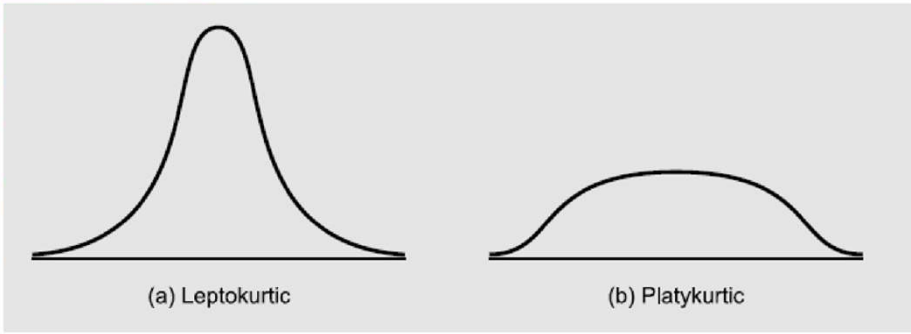
Andy Field, Jeremy Miles, Zoë Field - Discovering Statistics Using R -pg 20




Median
Mode
Mean
Mean
Median
Mode
Median
Mean
Mode
Positive
Skew
Symmetrical
Distribution
Negative
Skew

MEASURING DIVERGENCE FROM NORMALITY

- KURTOSIS
- Kurtosis refers to the flatness or peakedness of a frequency distribution as compared with the normal distribution
- Mesokurtic: same as normal distribution
- Leptokurtic: more peaked than the normal
- Platykurtic: flatter than the normal
- Negative Kurtosis , beta<3, broad, thin in the tails (has light tails) and tends to be flatter than normal, platy-kurtic
- Positive Kurtosis, Beta>3, thin, heavy tail data, lepto-kurtic



Kurtosis

- the observed values are concentrated more around the mode (a peaked curve) or away from the mode toward both tails of the frequency curve.
- average, variation, and skewness

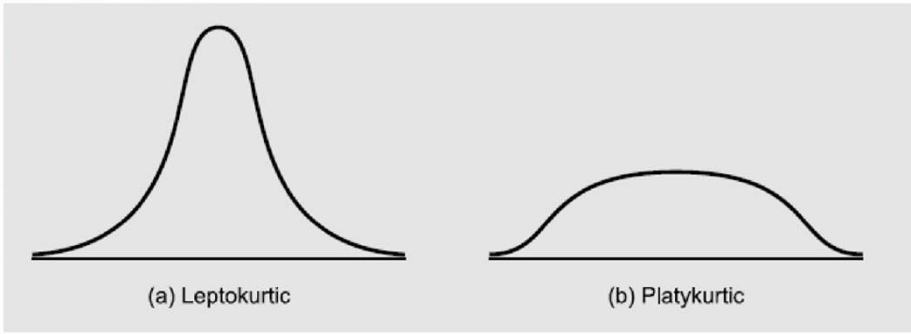
- Rcmdr→ Summaries→ numerical summary→ statistic
- install.packages("pastecs");library(pastecs); stat.desc(x,norm=TRUE,basic = TRUE)
- install.packages("psych");library(psych);describe(x)
- install.packages("e1071");library(e1071); skewness(x);kurtosis(x)
- plot(density(x)) ; plot density(x)
- More in detail --DeCarlo (1997).

Course Name :Basic Statistics using GUI-R (RKWard)
Module : Z-Table (Practical Demonstration of Normal Distribution)
Week 3 Lecture : 2

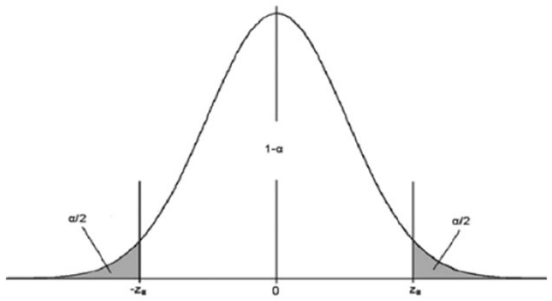
Harsh Pradhan, Assistant Professor,
Institute of Management Studies, BHU
https://bhu.ac.in/Site/FacultyProfile/1_5?FA000562

MEASURING DIVERGENCE FROM NORMALITY

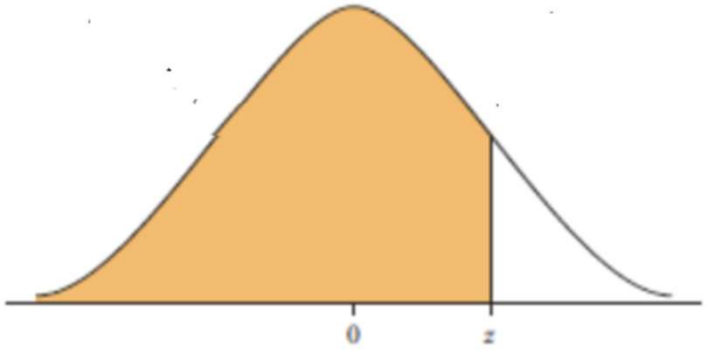
- KURTOSIS
- Kurtosis refers to the flatness or peakedness of a frequency distribution as compared with the normal distribution
- Mesokurtic: same as normal distribution
- Leptokurtic: more peaked than the normal
- Platykurtic: flatter than the normal
- Negative Kurtosis , beta<3, broad, thin in the tails (has light tails) and tends to be flatter than normal, platy-kurtic
- Positive Kurtosis, Beta>3, thin, heavy tail data, lepto-kurtic



Different Z Table



R 06



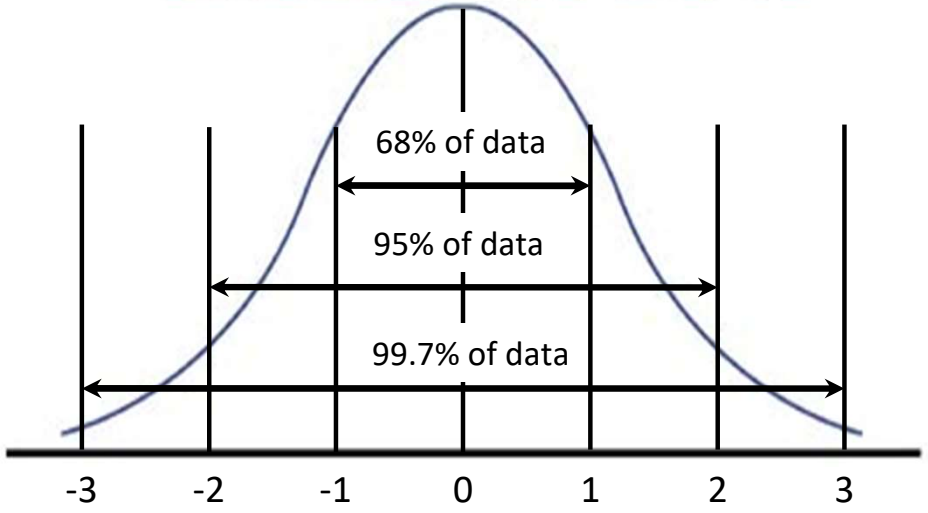
R 12

Prob Value

R 07

Standard normal distribution

- Mean of zero and variance of one
- Notation: N(0,1)
- Map between deviation and probability



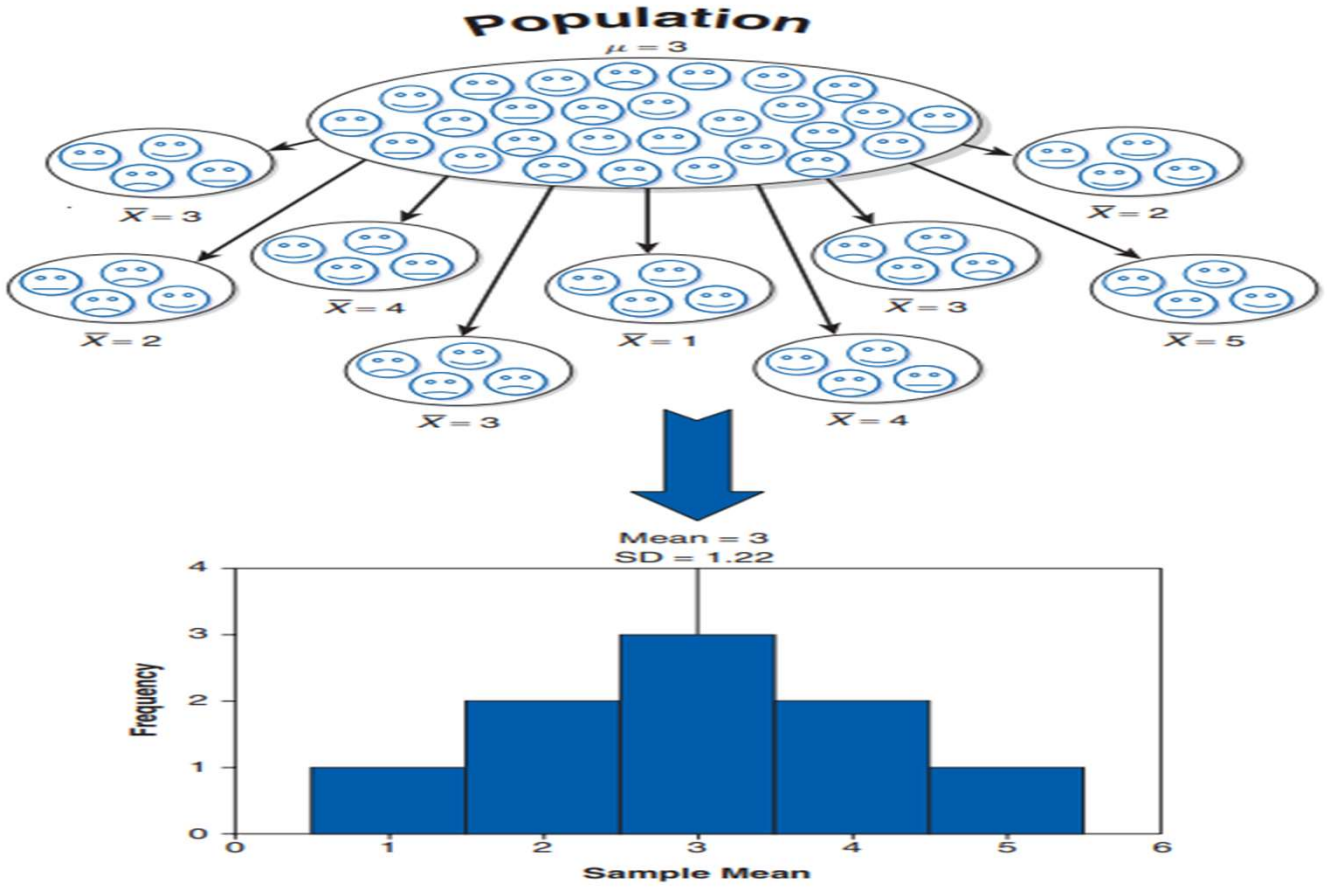
Course Name :Basic Statistics using GUI-R (RKWard)
Module : Introduction to Inferential Statistics
Week 3 Lecture : 3

Harsh Pradhan, Assistant Professor,
Institute of Management Studies, BHU
https://bhu.ac.in/Site/FacultyProfile/1_5?FA000562

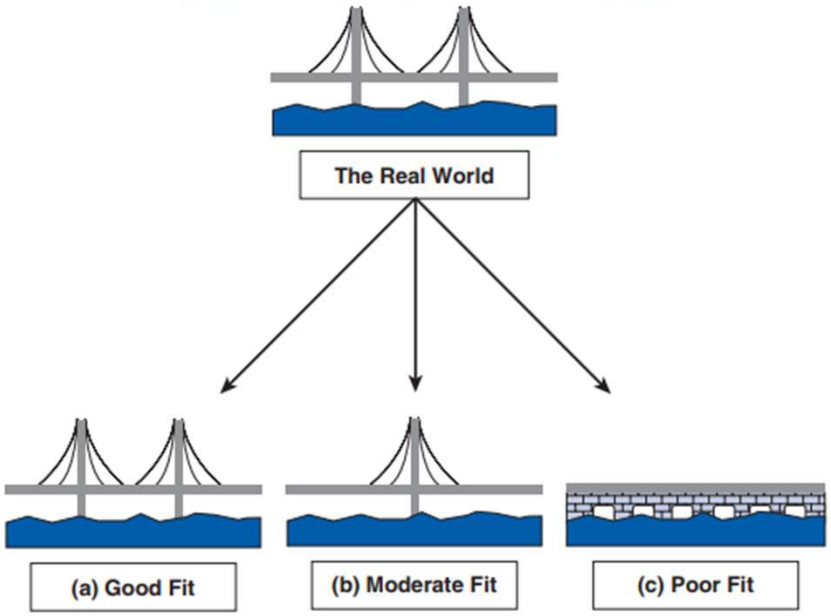
Agenda

- Sample and Population
- Inferential Statistics
- Model

Population and Sample



Model



Model Fit

Excel File

Trend Line in Excel

if file name is my.csv.data, and Y= JP_01, X=JP_02

Commands in R console

```
remotes::install_github("dustinfife/flexplot")

flexplot::flexplot(JP_01~JP_02,data=my.csv.data,method="loess",raw.data = FALSE,se = FALSE)
flexplot::flexplot(JP_01~JP_02,data=my.csv.data,method="polynomial",raw.data = FALSE,se = FALSE)
flexplot::flexplot(JP_01~JP_02,data=my.csv.data,method="lm",raw.data = FALSE,se = FALSE)

?flexplot::flexplot
```

LOESS (Locally Estimated Scatterplot Smoothing)

Documentation of Flexplot Package

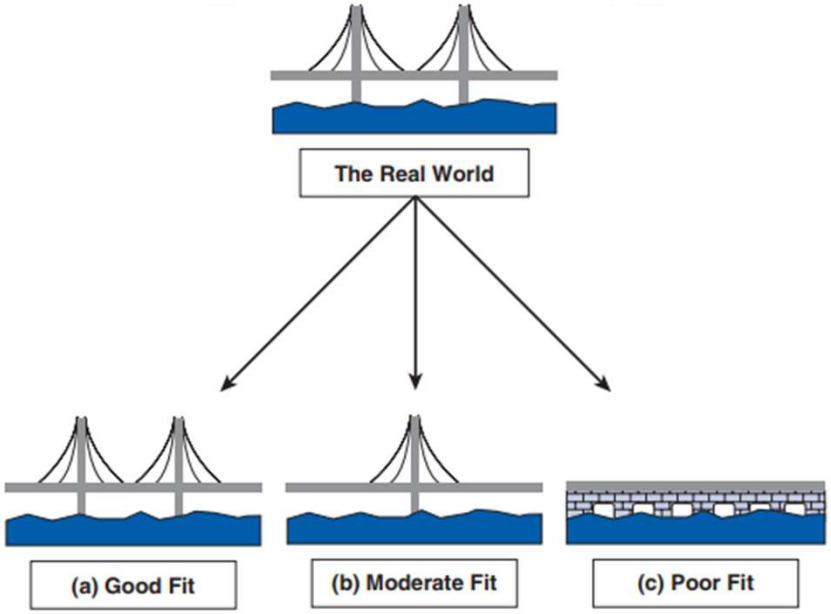
Course Name :Basic Statistics using GUI-R (RKWard)

Module : Model Fit

Week 3 Lecture : 4

Harsh Pradhan, Assistant Professor,
Institute of Management Studies, BHU
https://bhu.ac.in/Site/FacultyProfile/1_5?FA000562

Model



Model Fit

Excel File

Trend Line in Excel

if file name is my.csv.data, and Y= JP_01, X=JP_02

Commands in R console

```
remotes::install_github("dustinfife/flexplot")

flexplot::flexplot(JP_01~JP_02,data=my.csv.data,method="loess",raw.data = FALSE,se = FALSE)
flexplot::flexplot(JP_01~JP_02,data=my.csv.data,method="polynomial",raw.data = FALSE,se = FALSE)
flexplot::flexplot(JP_01~JP_02,data=my.csv.data,method="lm",raw.data = FALSE,se = FALSE)

?flexplot::flexplot
```

LOESS (Locally Estimated Scatterplot Smoothing)

Documentation of Flexplot Package

Course Name :Basic Statistics using GUI-R (RKWard)
Module : Hypothesis and Error
Week 3 Lecture : 5

Harsh Pradhan, Assistant Professor,
Institute of Management Studies, BHU
https://bhu.ac.in/Site/FacultyProfile/1_5?FA000562

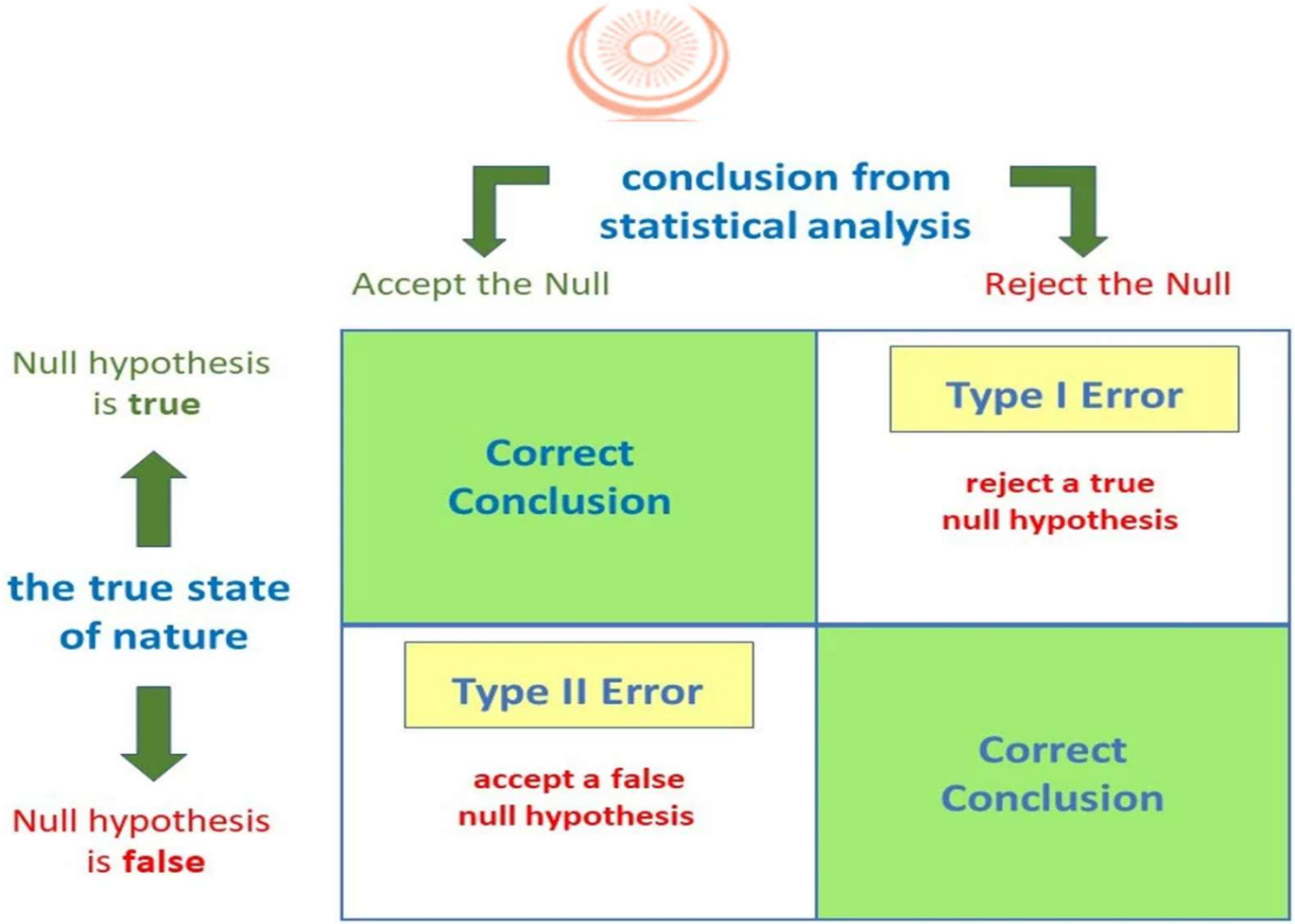
Hypothesis

A statistical hypothesis is a claim or assertion either about the value of a single parameter, about the values of several parameters, or about the form of an entire probability distribution.

Null Hypothesis

Alternative Hypothesis

Error


conclusion from statistical analysis
Accept the Null
Reject the Null
Null hypothesis is true
the true state of nature
Null hypothesis is false
Correct Conclusion
Type I Error
reject a true null hypothesis
Type II Error
accept a false null hypothesis
Correct Conclusion