Application of Normal Probability Distributions

RESAN S. LIPAR

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents.

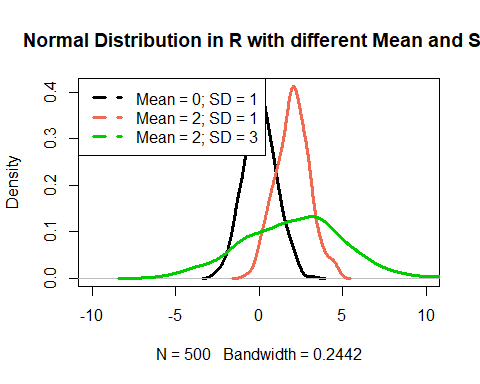
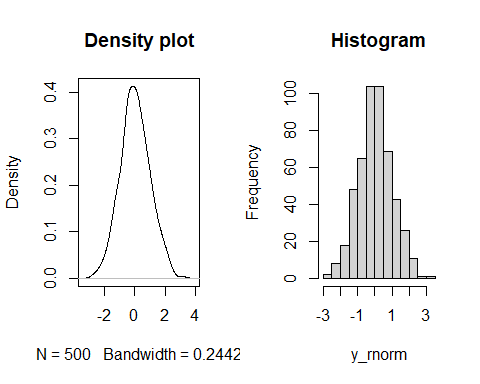
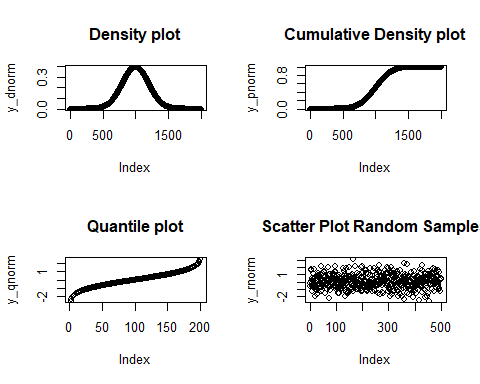
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

R codes are put inside what is called as code chunks. Each chunk should have a different name. To compile an R MArkdown file (RMD) just klick on the knit button and select the type of output that you want. All three outputs html. pdf and word format are not possible at the same time. Some RMD commandes work onyl for html output and not for pdf and word output. For example alignment of figure does not work for word output.

# HOW TO CREATE AN RMD FILE

Click File in main menu  
Click New File  
Click on R Markdown  
Type title  
Click on Word if you want Word output  
Click Ok

To run the sample Rmarkdown file  
Click on the Knit button  
Click on Knit to Word  
Wait until Word icon blink



## Normal Distribution

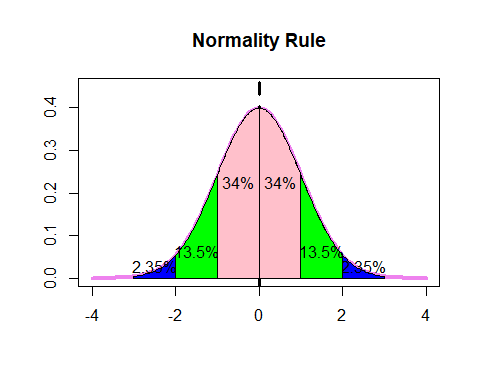
A type of continuous probability distribution for a real-valued random variable. It is a type of symmetric distribution where most of the observations cluster around the central peak and the probabilities for values further away from the mean taper off equally in both directions. It is represented using a bell-shaped density curve described by its mean and standard deviation. It is also known as the Gaussian distribution. It has got the following features: Symmetric bell shape

# VERY IMPORTANT PROPERTY: Total Area Under the Normal Curve is 1

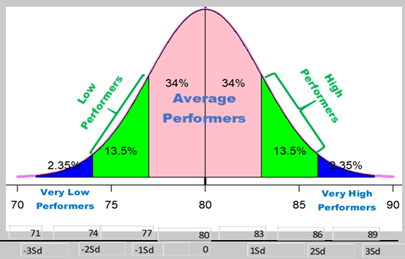
# PARTITIONING OF AREAS UNDER THE NORMAL CURVE

Mean and median are equal; both located at the center of the distribution  
- 68% of the data falls within 1 standard deviation of the mean  
- 95% of the data falls within 2 standard deviations of the mean  
- 99.7% percent of the data falls within 3 standard deviations of the mean

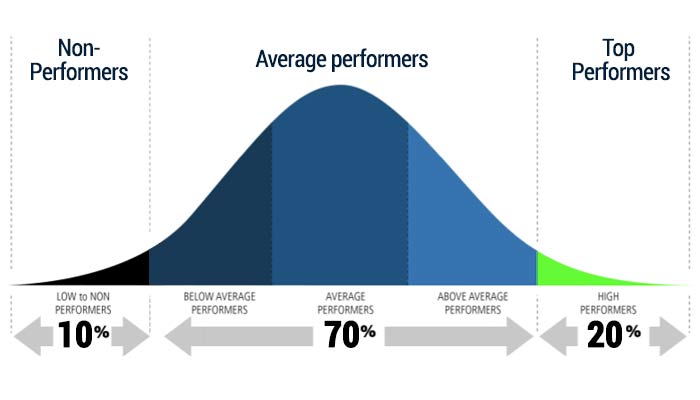
# Empirical Rule for Normal Distribution

This is the percentage of area covered under a standard Normal distribution 

and this is an example on how to load a ready image Normal curve of examination scores with mean = 80 and standard deviation 3 units away from the mean value.



and here is another illustration of the normality rule



# Example of Simulation Using Probability Distribution

## Notations for our Random Variable representing the raw data

Use capital letter like X to represent a normal random variable with mean and standard deviation as X ~ N(,).

## Notations for our standard Random Variable

Use capital letter Z to represent a standard normal random variable with mean 0 and sd = 1 as Z ~ N(0,1).

Step1: Generate a set of random samples of size 1000 with a mean score of 70 and standard deviation 10.

Step2: Obtain descriptive statistics of the simulated dataset

Step3: Plot the histogram and of the random samples

Step4: Finding probability using pnorm() command in R

Step5: Finding some quantiles for a given probability value

# answers

Step1: Display first 100 values

## [1] 64.39524 67.69823 85.58708 70.70508 71.29288 87.15065 74.60916 57.34939  
## [9] 63.13147 65.54338 82.24082 73.59814 74.00771 71.10683 64.44159 87.86913  
## [17] 74.97850 50.33383 77.01356 65.27209 59.32176 67.82025 59.73996 62.71109  
## [25] 63.74961 53.13307 78.37787 71.53373 58.61863 82.53815 74.26464 67.04929  
## [33] 78.95126 78.78133 78.21581 76.88640 75.53918 69.38088 66.94037 66.19529  
## [41] 63.05293 67.92083 57.34604 91.68956 82.07962 58.76891 65.97115 65.33345  
## [49] 77.79965 69.16631 72.53319 69.71453 69.57130 83.68602 67.74229 85.16471  
## [57] 54.51247 75.84614 71.23854 72.15942 73.79639 64.97677 66.66793 59.81425  
## [65] 59.28209 73.03529 74.48210 70.53004 79.22267 90.50085 65.08969 46.90831  
## [73] 80.05739 62.90799 63.11991 80.25571 67.15227 57.79282 71.81303 68.61109  
## [81] 70.05764 73.85280 66.29340 76.44377 67.79513 73.31782 80.96839 74.35181  
## [89] 66.74068 81.48808 79.93504 75.48397 72.38732 63.72094 83.60652 63.99740  
## [97] 91.87333 85.32611 67.64300 59.73579

Step2: Compute dewscriptive statistics

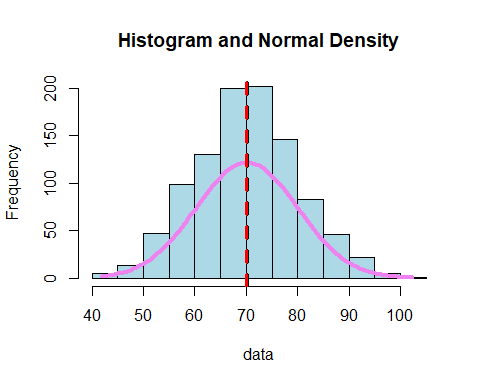
## [1] "Mean = 70.1612786593489"

## [1] "Median = 70.0920963876889"

## [1] "Skewness = 0.0652939143523381"

## [1] "Kurtosis = 2.92574655726915"

## [1] "Variance = 98.3458926215703"

Step3: Display histogram and Normal density 

Step4: Find the probability that the normal random variable  
a) is less than 70, is less than 150  
b) between 60 and 90  
c) more than 95  
D) what is that X value where 85% of the random samples are below it?

# Step4 Solution: Prob(X < 70) (Answer should be 0.5 Since half of the curve is below x = 70)

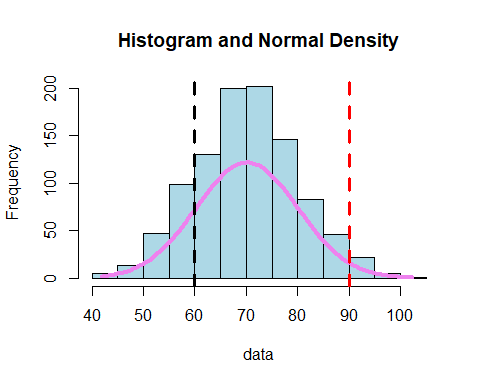
## [1] 0.5

# Step4 Solution: Prob(X < 150) (Answer should be 1 since the whole curve is found on the left side of X = 150)

## [1] 1

# Step4 Solution: Prob(60 < X < 90)

## [1] 0.8185946



# That area between the two lines x = 60, x = 90 and below the normal curve is equal to 0.8185946

# Step4 Solution: Prob( X > 95)

## [1] 0.004661188

# Step4 Solution: Solve for the X value with a desired percentage below it

## [1] 80.36433

# SECOND EXAMPLE ABOUT NORMAL CURVES THIS TIME WITH TRANSFPORMATION

Among students taking a standardized exam, scores are normally distributed with a mean of 550 and standard deviation 100. What proportion of the students score above 700?

The transformation formula from a normal random variable X to a standard normal random variable is given by

Z = (X - mean)/sd

At X = 700, mean = 550 and sd = 100, we have Z = (700 - 550)/100 = 150/100 = 1.5

Therefore P( X > 700) = P(Z > 1.5) = 1 - P( Z < 1.5) = 1 - 0.9331928 = 0.0668072 which represents the area to the right side of Z = 1.5.

As a summary, an application of the Normal probability distribution is called as a

## Z test.