Math Concepts

Anna Yeaton

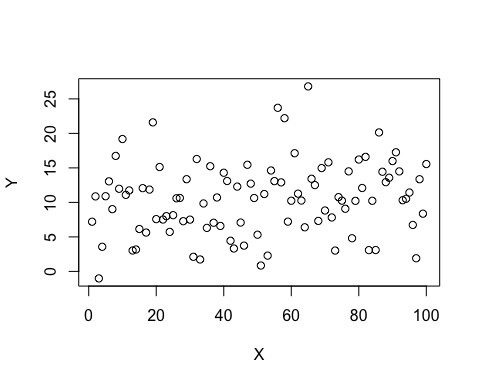
Fall 2018

# Math Concepts

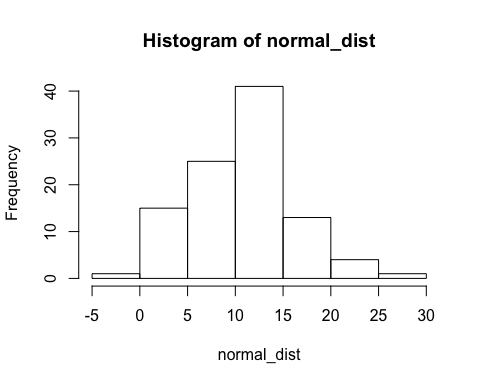
In this lab, we will go over math concepts that serve as a foundation for the unsupervised learning class. We will go over standard deviation, variance, covariance, eigen values, and eigen vectors.

## Distributions

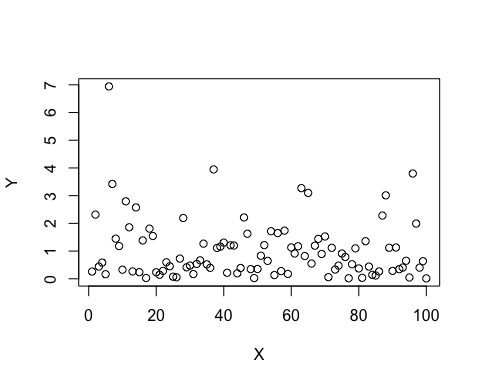
#Gaussian(normal) distribution  
normal\_dist <- rnorm(n = 100, mean = 10, sd = 5)  
plot(normal\_dist, ylab="Y", xlab = "X")



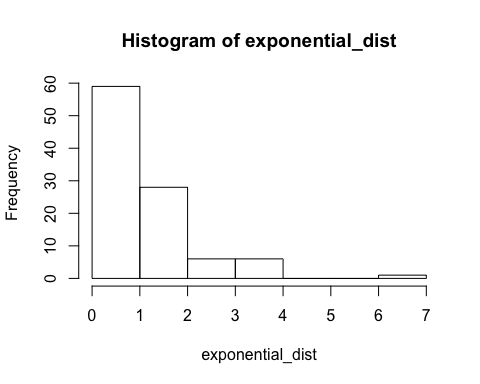
hist(normal\_dist)



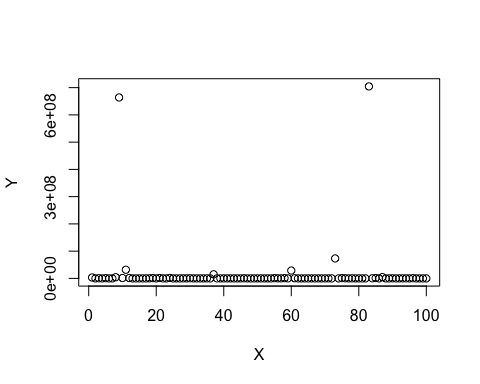
#exponential distribution  
exponential\_dist <- rexp(n=100, rate =1)  
plot(exponential\_dist, ylab="Y", xlab = "X")



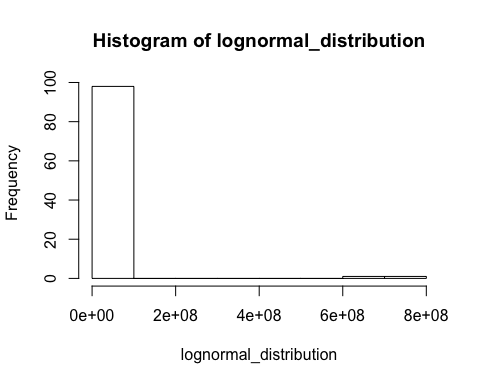
hist(exponential\_dist)



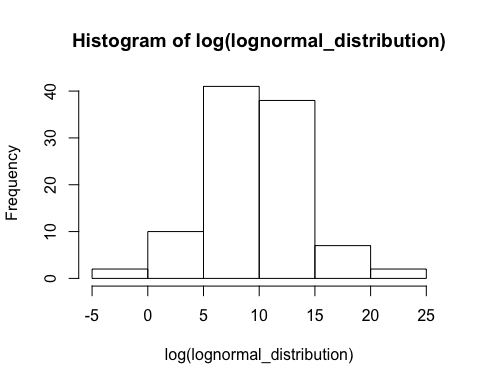
#lognormal distribution  
lognormal\_distribution <- rlnorm(100, meanlog = 10, sdlog = 5)  
plot(lognormal\_distribution, ylab="Y", xlab = "X")



hist(lognormal\_distribution)



hist(log(lognormal\_distribution))



## Variance

Variance is a measurememt of the spread of data.

#the spread of this set of values   
spread\_ex1 <- c(1,2,1,2,1,2,1,2,1,2)  
  
#is much different from the spread of this set of values  
spread\_ex2 <- c(2,5,8,10,1,35,67,32,89,100)

Calculate the variance of spread\_ex1 and spread\_ex2

var(spread\_ex1)

## [1] 0.2777778

var(spread\_ex2)

## [1] 1408.1

## Standard deviation

Standard deviation is also a measurememt of the spread of data. It is the square root of the variance.

Calculate the standard deviations of spread\_ex1 and spread\_ex2

sd(spread\_ex1)

## [1] 0.5270463

sd(spread\_ex2)

## [1] 37.52466

## Covariance

Covariance is a measure of the variance of two random variables with respect to each other. In other words, covariance is a measure of linear dependence between two random variables.

cov\_ex1 <- data.frame(spread\_ex1, spread\_ex2)

Calculate the covariance of cov\_ex1 between spread\_ex1 and spread\_ex2

cov(cov\_ex1$spread\_ex1, cov\_ex1$spread\_ex2)

## [1] 0.8333333

Calculate the covariance matrix of cov\_ex1. Did you see any of these values before?

cov(cov\_ex1)

## spread\_ex1 spread\_ex2  
## spread\_ex1 0.2777778 0.8333333  
## spread\_ex2 0.8333333 1408.1000000

## Eigen vectors and Eigen values

Eigen vectors and Eigen values come in pairs. An eigen vector is a direction, and the corresponding eigen value tells how much the data varies in that direction. You can only calculate eigen vectors and values for n x n matrices. For an n x n matrix, there will be n eigen values and eigen vectors. Eigen vectors are perpendicular to each other. The eigen vector with the highest eigen value is the principal component.

So how do you find an eigen vector? Why is it one direction and not another? The eigen vector is special in that when a linear transformation is applied to it, the direction of the vector does not change.

You can project the data onto an eigen vector space instead of on the X and Y axis. This is the visualization for PCA.

## Matrix operations

<http://home.cc.umanitoba.ca/~thomas/Courses/MatrixMultPractice.pdf>