

# Data 606: Chapter-3 Homework

---

Student Name : Sachid Deshmukh

Date : 10/16/2018

- GitHub Location for rmd file
  - GitHub Location for pdf file
  - RPub's location of published file
- 

```
library(ggplot2)

NormalDistPlot = function(lrange = -5, hrange = 5, mean = 0, sd=1, lb=2, ub=3)
{
  num <- seq(lrange,hrange, length=100) * sd + mean
  dnum <- dnorm(num)

  plot(num, dnum, type="n",main="Normal Distribution", axes=FALSE, xlab="", ylab="")

  i <- num >= lb & num <= ub
  lines(num, dnum)
  polygon(c(lb,num[i],ub), c(0,dnum[i],0), col="blue")
  axis(1, at=seq(lrange, hrange, 1), pos=0)
}
```

1. 3.2 Area under curve

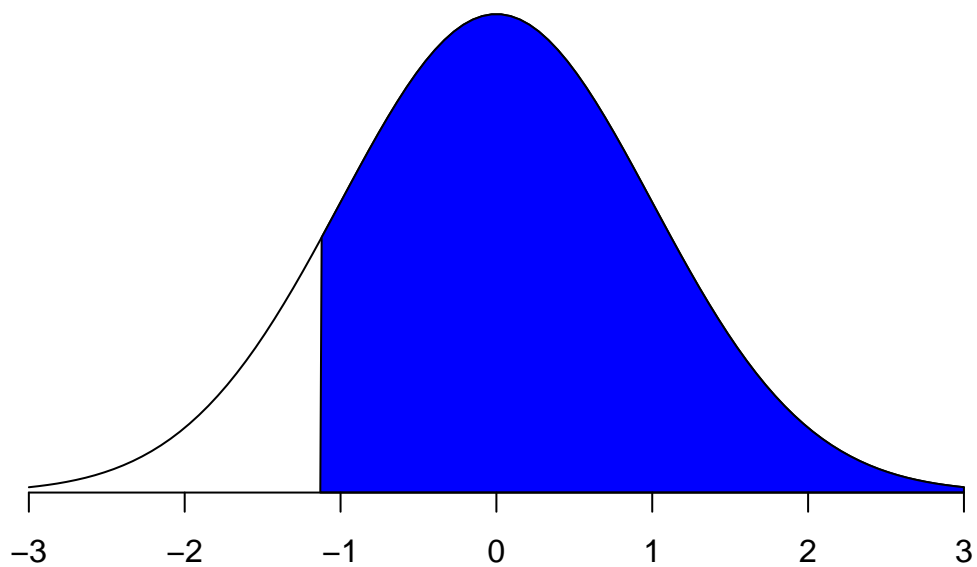
•

```
1 - pnorm(-1.13, mean = 0, sd = 1)
```

```
## [1] 0.8707619
```

```
NormalDistPlot(lrange = -3, hrange = 3, lb=-1.13, ub =3)
```

## Normal Distribution

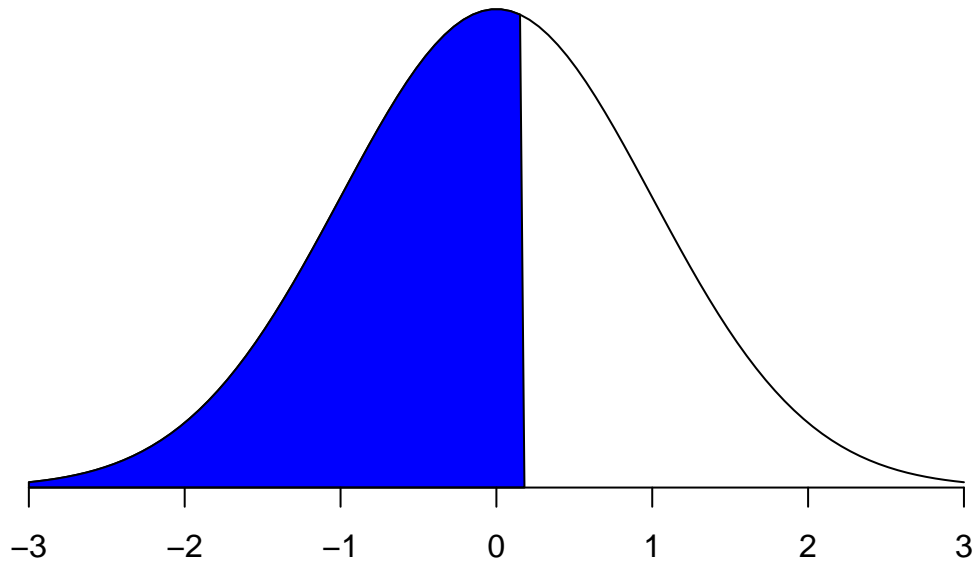


```
pnorm(0.18, mean = 0, sd = 1)
```

```
## [1] 0.5714237
```

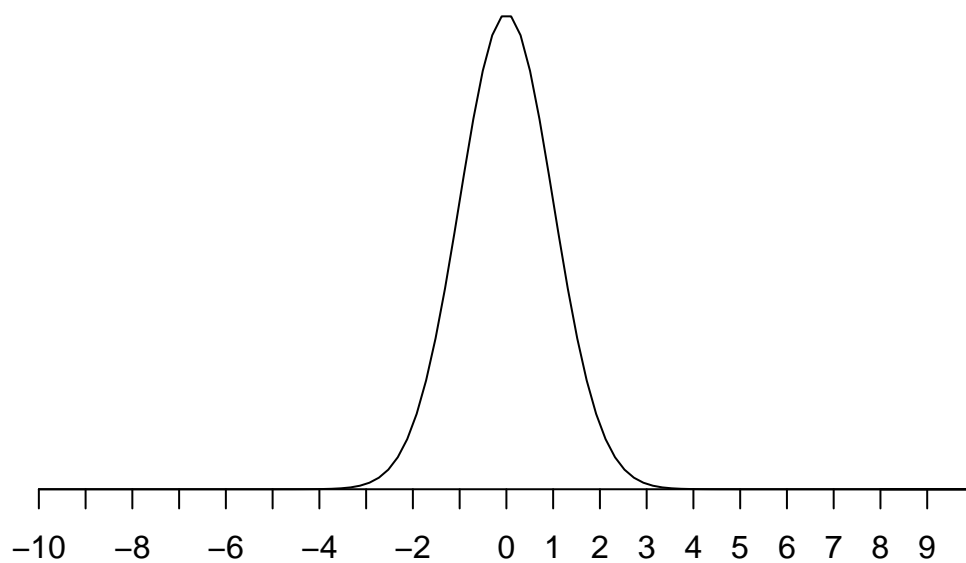
```
NormalDistPlot(lrange = -3, hrange = 3, lb=-3, ub =0.18)
```

## Normal Distribution



```
•  
1- pnorm(8, mean = 0, sd = 1)  
  
## [1] 6.661338e-16  
NormalDistPlot(lrange = -10, hrange = 10, lb=8, ub = 10)
```

## Normal Distribution



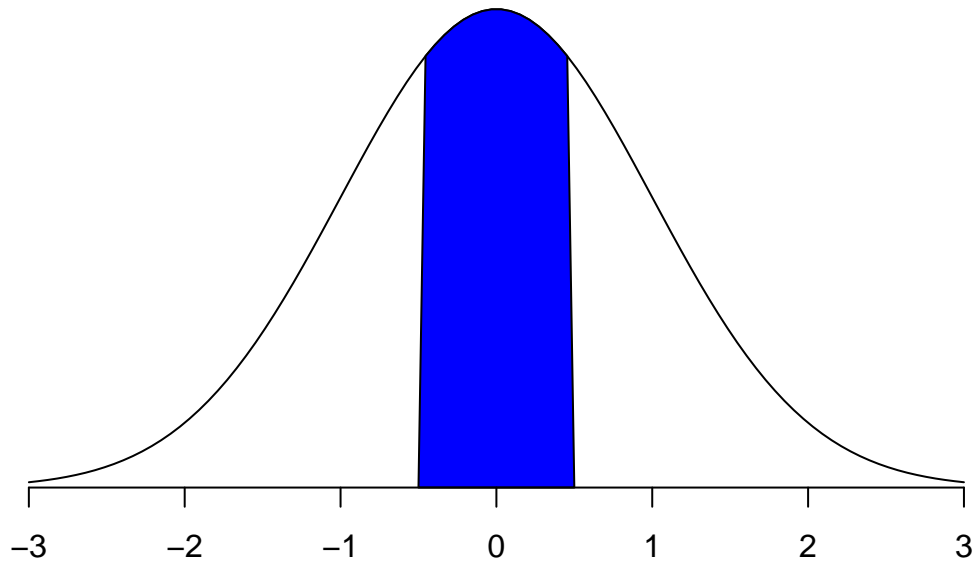
\*

```
pnorm(0.5, mean = 0, sd = 1) - pnorm(-0.5, mean = 0, sd = 1)
```

```
## [1] 0.3829249
```

```
NormalDistPlot(lrange = -3, hrange = 3, lb=-0.5, ub = 0.5)
```

## Normal Distribution



### 2. 3.4 Triathlon Times

- Men :  $N(\mu = 4313, \sigma = 583)$ , Women :  $N(\mu = 5261, \sigma = 807)$
- 

```
print(paste("Leo's Z Score = ", (4948-4313)/583))
```

```
## [1] "Leo's Z Score = 1.08919382504288"
```

```
print(paste("Mary's Z Score = ", (5513-5261)/807))
```

```
## [1] "Mary's Z Score = 0.312267657992565"
```

Leo is faster runner than Mary

- 

```
print(paste("Runners faster than Leo = ", 1- pnorm(4948, 4313, 583)))
```

```
## [1] "Runners faster than Leo = 0.13803421070203"
```

```
print(paste("Runners faster than Mary = ", 1-pnorm(5513,5261,807)))
```

```
## [1] "Runners faster than Mary = 0.3774185585735"
```

Only 13% runners are faster than Leo, however 37% runners are faster than Mary. Leo rank better than Mary in their respective groups

- 

```
print(paste("Leo is faster than", pnorm(4948, 4313, 583)*100, "%", "runners in his group"))
```

```
## [1] "Leo is faster than 86.196578929797 % runners in his group"
•
print(paste("Mary is faster than", pnorm(5513,5261,807)*100, "%", "runners in his group"))
```

```
## [1] "Mary is faster than 62.25814414265 % runners in his group"
```

According to central limit theorem, we can assume normal distribution for any sample size greater than 30. If the distributions of finishing times are not nearly normal, we can still assume data normality if the no of runners in each group is greater than 30. We can answer questions b-e by assuming data normality as per central limit theorem.

### 3. 3.18 Heights of female college students

```
heights = c(54,55,56,56,57,58,58,59,60,60,60,61,61,62,62,63,63,63,64,65,67,67,69,73)
```

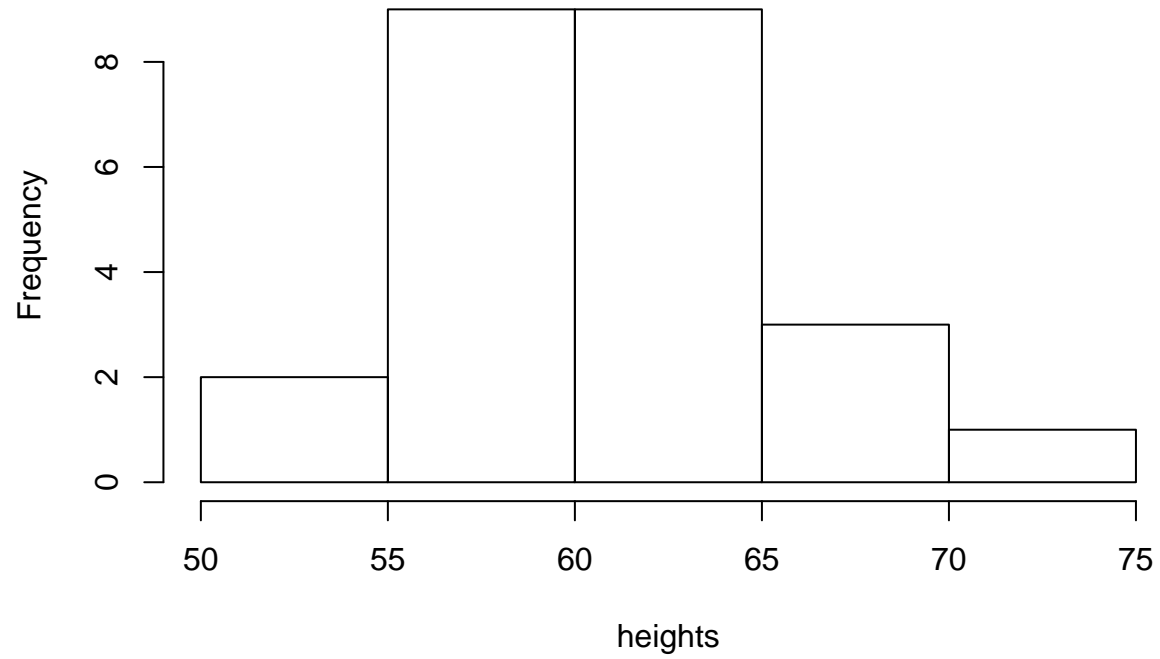
```
•
mean = 61.52
sd = 4.58
for(i in 1:3)
{
  min.range = mean - i * sd
  max.range = mean + i * sd
  heights.sel = heights[heights >= min.range & heights <= max.range]
  print(paste("Heights in", i, "sd = ", length(heights.sel)/length(heights)*100, "%"))
}
```

```
## [1] "Heights in 1 sd = 66.6666666666667 %"
## [1] "Heights in 2 sd = 95.8333333333333 %"
## [1] "Heights in 3 sd = 100 %"
```

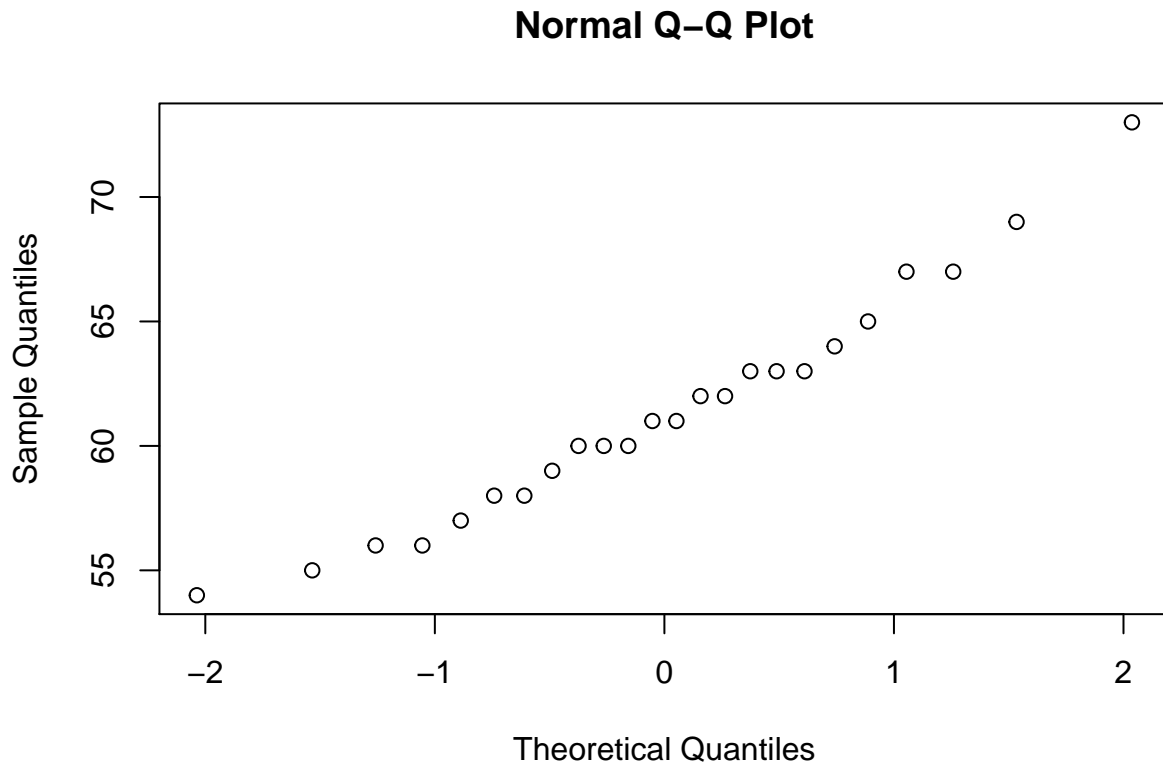
From the above analysis we can conclude that heights approximately follow 68-95-99.7 rule

```
•
hist(heights)
```

**Histogram of heights**



```
qqnorm(heights)
```



Above plots suggests that distribution is normal

4. 3.22 Defective Rate

•

```
dbinom(1, size=10, prob=0.02)
```

```
## [1] 0.1667496
```

•

```
dbinom(0, size=100, prob=0.02)
```

```
## [1] 0.1326196
```

•

```
print(paste("With 0.02 defective probability, on average we need to produce", 1/0.02, "transistors before first with defective"))
```

```
## [1] "With 0.02 defective probability, on average we need to produce 50 transistors before first with defective"
```

```
print(paste("With 0.02 defective probability and 50 sample size, standard deviation of binomial distribution is", sqrt(50*0.02*0.98)))
```

```
## [1] "With 0.02 defective probability and 50 sample size, standard deviation of binomial distribution is 0.99"
```

•

```
print(paste("With 0.05 defective probability, on average we need to produce", 1/0.05, "transistors before first with defective"))
```

```
## [1] "With 0.05 defective probability, on average we need to produce 20 transistors before first with defective"
```



```
print(paste("With 0.05 defective probability and 20 sample size, standard deviation of binomial distrib
```

```
## [1] "With 0.05 defective probability and 20 sample size, standard deviation of binomial distribution
```

•

When the probability increases, mean and standard deviation decreases

5. 3.38 Male Children

•

```
print(paste("Probability of having two boys = ", dbinom(2, size=3, prob=0.51)))
```

```
## [1] "Probability of having two boys = 0.382347"
```

•

```
child.order = c("BBG", "BGB", "GBB")
```

```
p.boy = 0.51
```

```
p.girl = 1-0.51
```

```
p.BBG = 0.51 * 0.51 * (1-0.51)
```

```
p.BGB = 0.51 * (1-0.51) * 0.51
```

```
p.GBB = (1-0.51) * 0.51 * 0.51
```

```
p.twoboys = p.BBG + p.BGB + p.GBB
```

```
print(paste("Probability of having two boys = ", p.twoboys))
```

```
## [1] "Probability of having two boys = 0.382347"
```

Above analysis confirms that answers from a and b matches

•

For calculating 3 boys out of 8 kids, we need to evaluate 3 out of 8 combinations. That is total 56 scenarios. This makes method b more time consuming compared to method 1

6. 3.42 Serving in Volleyball

•

```
print(paste("Probability of making 3 successful serves from 10 attempts = ", dbinom(3, size=10, prob=0.15)))
```

```
## [1] "Probability of making 3 successful serves from 10 attempts = 0.129833720753906"
```

•

Probability that 10th serve will be successful is 0.15%. All the serves are independent of each other so probability of successful serve remains the same which is 0.15

•

Scenario 1 is about getting probability of three successful serves in 10 attempts. Here we are looking for probability for total three successful serves. Scenario 2 is about getting 10th serve successful after we had two successful serves in 9 attempts. Scenario 2 focuses more on individual 10th serve and probability of that being successful is 0.15. This makes a difference in scenario 1 and 2 even though they look similar