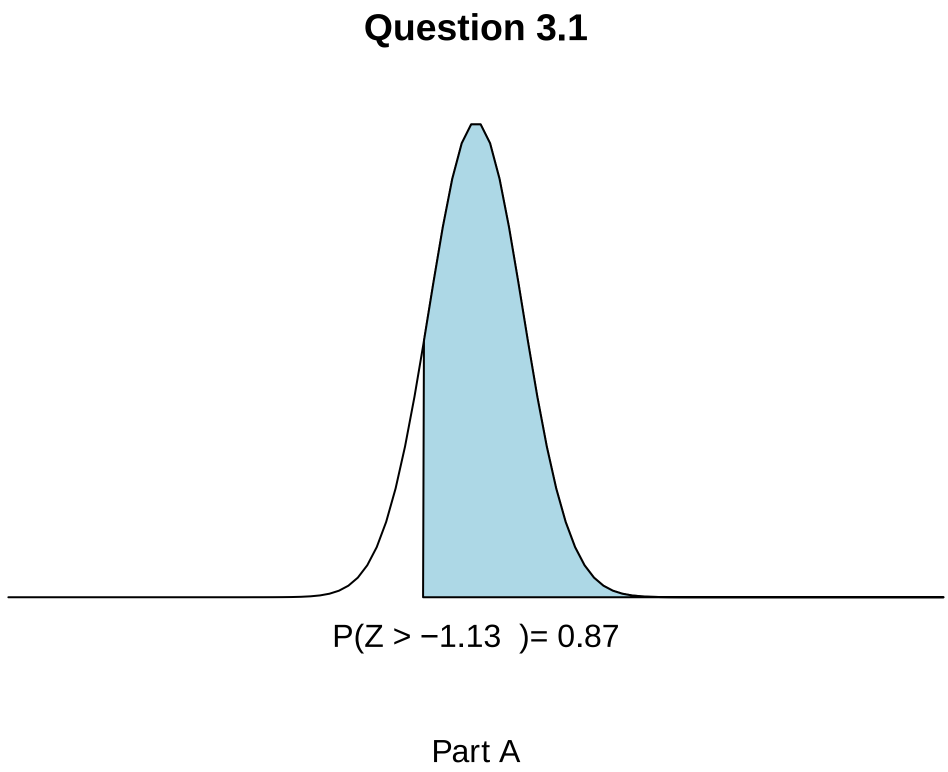
Chapter 3 HW

Munkhnaran Gankhuyag

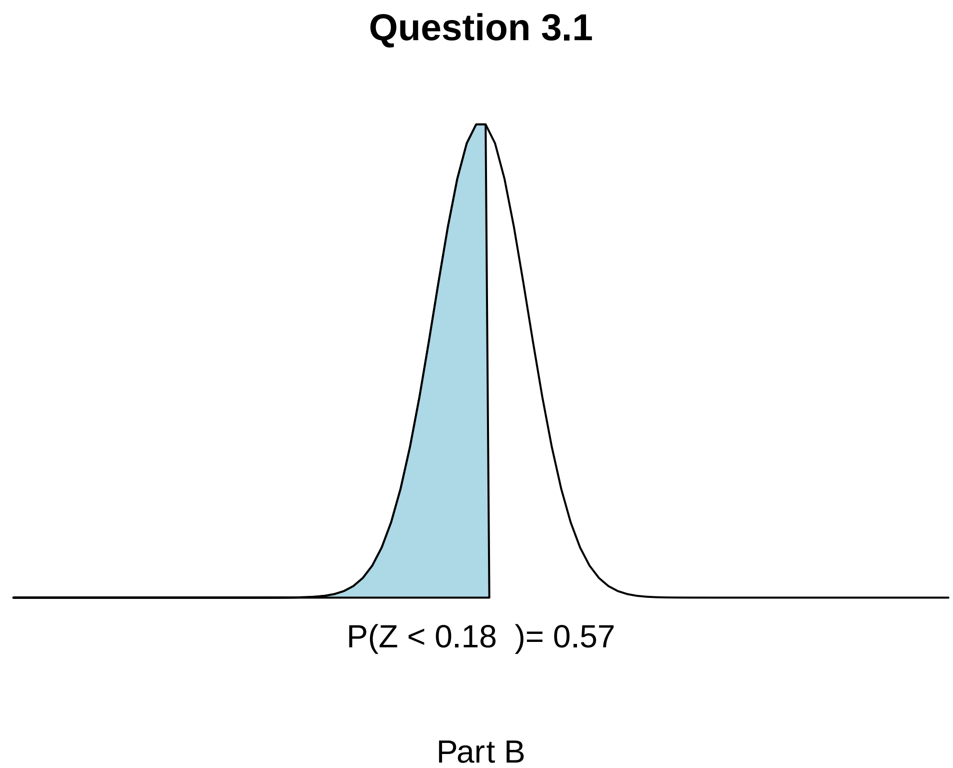
September 15, 2017

3.2 Area under the curve, Part II.

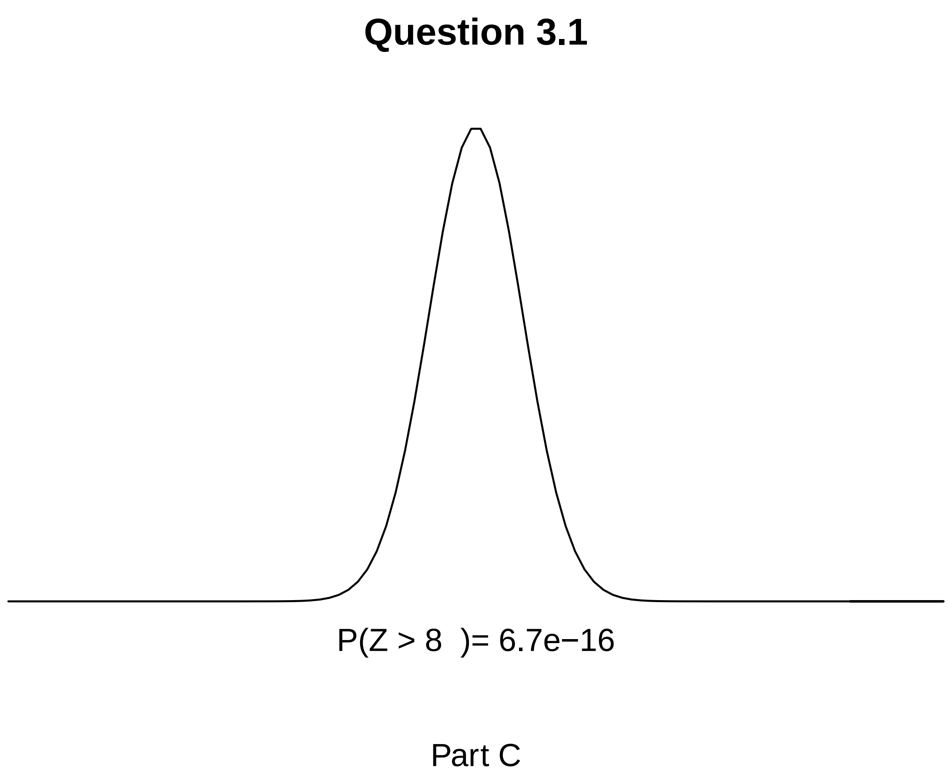
```
mean=0
sd=1
low= -10
high= 10
q1a = -1.13
q1b = .18
q1c = 8
q1d1 =  -.5
q1d2 =  .5

# create vector of x's: 100 observations around the given mean and sd
x <- seq(-10,10,length=100)*sd + mean
# find probabilities for these values given the distribution parameters
y <- dnorm(x,mean,sd)
# plot
plot(x, y, type="n", xlab="Part A", ylab="", main="Question 3.1", axes=FALSE)
# add lines at the upper and lower bounds
i <- x >= q1a
lines(x, y)
# create polygon for shading area of distribution
polygon(c(q1a,x[i],high), c(0,y[i],0), col="lightblue")
area <- pnorm(high, mean, sd) - pnorm(q1a, mean, sd)
# add title and axis labels
mtext(paste("P(Z >",q1a," )=",signif(area, digits=2)),1)
axis(1, at=seq(-10, 10, 1), pos=-1)
```



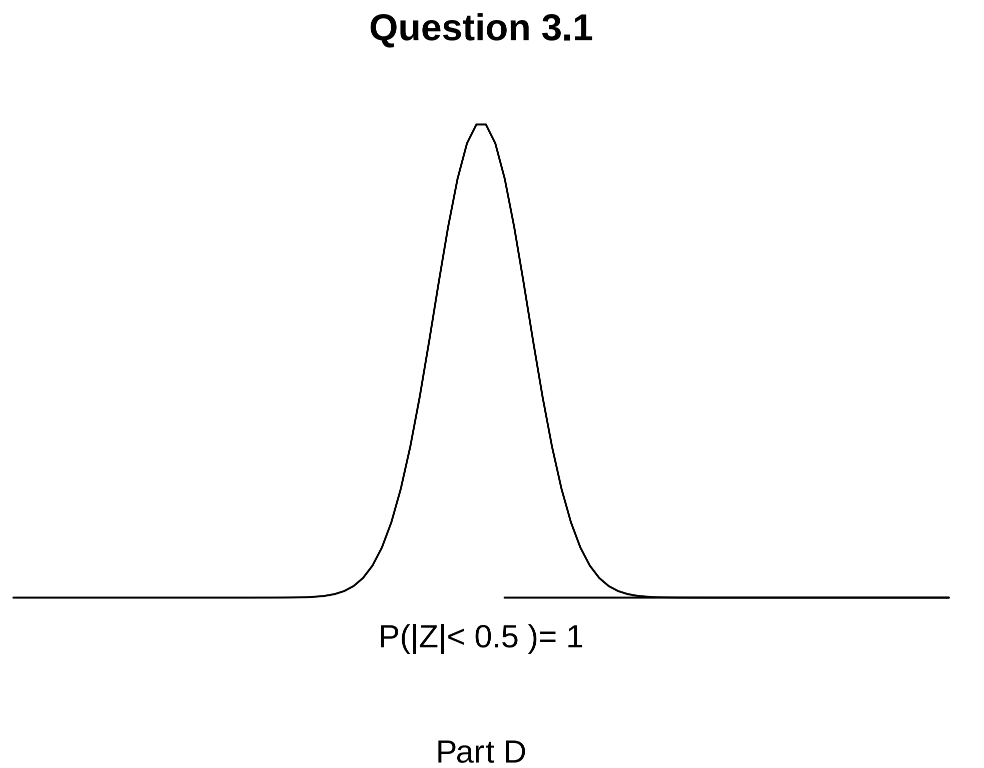

```
plot(x, y, type="n", xlab="Part B", ylab="", main="Question 3.1", axes=FALSE)
i2 <- x < q1b
lines(x, y)
polygon(c(q1b,x[i2],q1b), c(0,y[i2],0), col="lightblue")
area <- pnorm(q1b, mean, sd) -pnorm(low, mean, sd)
mtext(paste("P(Z <",q1b," )=",signif(area, digits=2)),1)
axis(1, at=seq(-10, 10, 1), pos=-1)
```




```r
plot(x, y, type="n", xlab="Part C", ylab="", main="Question 3.1", axes=FALSE)
i3 <- x > q1c
lines(x, y)
polygon(c(q1c,x[i3],q1c), c(0,y[i3],0), col="lightblue")
area <- pnorm(high, mean, sd) - pnorm(q1c, mean, sd)
mtext(paste("P(Z >",q1c," )=",signif(area, digits=2)),1)
axis(1, at=seq(-10, 10, 1), pos=-1)
```




```
plot(x, y, type="n", xlab="Part D", ylab="", main="Question 3.1", axes=FALSE)
# add lines at the upper and lower bounds
i4 <- x >= q1d2 & x <= q1d1
lines(x, y)
# create polygon for shading area of distribution
polygon(c(q1d2,x[i4],high), c(0,y[i4],0), col="lightblue")
area <- pnorm(high, mean, sd) - pnorm(low, q1d1, sd)
# add title and axis labels
mtext(paste("P(|Z|<",q1d2,")=",signif(area, digits=3)),1)
axis(1, at=seq(40, 160, 20), pos=0)
```



3.4 Triathlon times, Part I.

```
score <- matrix(c(160,151,7,157,153,7.67),ncol = 3, byrow = TRUE)
colnames(score)<- c("score", "mean", "sd")
rownames(score) <-c("Verbal","Quantitative")
score <- as.table(score)
score
```

```
##               score   mean     sd
## Verbal       160.00 151.00   7.00
## Quantitative 157.00 153.00   7.67
```

```
Verbal<- pnorm(160, mean = 151, sd=7)
Quantiative <- pnorm(157, mean =153.7, sd =7.67)
Verbal
```

```
## [1] 0.9007286
```

```
Quantiative
```

```
## [1] 0.6664923
```

The z scores tell me that Sophia is in the 90th percentile for the verbal reasoning section and in the 67th percentile for the quantitative section. Overall she did better on the verabal section. about 10% of test taker scored better on the verbal section. Comparing raw scores may conclude to incorrect assumptions because the difficulty is unknown so we compare the standard deviations. If the distrubtions of the scores on these exams are not nearly normal, then yes my answers would change. For example if only GRE test experts decided to take the exam with Sophia that day, that would increase the mean, and we would assume that Sophia did poorly on the GRE.

3.4 Triathlon times, Part I.

a)

```
triathlon <- matrix(c(4948,4313,583,5313,5261,807),ncol = 3, byrow = TRUE)
colnames(triathlon)<- c("score", "mean", "sd")
rownames(triathlon) <-c("Leo","Mary")
triathlon <- as.table(triathlon)
triathlon
```

```
##      score mean   sd
## Leo   4948 4313  583
## Mary  5313 5261  807
```

```
Leoz <-((4948-4313)/583)
MaryZ <-((5313-5261)/807)
Leoz #Leo's Z score
```

```
## [1] 1.089194
```

```
MaryZ #Mary's Z Score
```

```
## [1] 0.06443618
```

```
pnorm(4948,4313,583)
```

```
## [1] 0.8619658
```

```
pnorm(5313,5261,807)
```

```
## [1] 0.5256885
```

b)The z score tells me that Leo did better in his race. c)Leo ranked better in his group because he is in the 86th percentile while Mary is in the 53rd percentile in her group. d)Leo finished faster than 86.2% of his group. e)Mary finished faster than 52.6% of her group. f)If the distribution of finishing times are not normal, I expect my answers to change. For example if Leo's group had all professional triathlon racers, his z score and percentile rank would drop.

3.18 Heights of female college students.

```
Heights <- c(54,55,56,56,57,58,58,59,60,60,60,61,61,62,62,63,63,63,64,65,65,67,67,69,73)
mean(Heights)
```

```
## [1] 61.52
```

```
sd(Heights)
```

```
## [1] 4.583667
```

```
mean(Heights)-sd(Heights)
```

```
## [1] 56.93633
```

```r
mean(Heights)+sd(Heights)
```

```
## [1] 66.10367
```

```r
mean(Heights)-(sd(Heights)*2)
```

```
## [1] 52.35267
```

```r
mean(Heights)+(sd(Heights)*2)
```

```
## [1] 70.68733
```

```r
mean(Heights)-(sd(Heights)*3)
```

```
## [1] 47.769
```

```r
mean(Heights)+(sd(Heights)*3)
```

```
## [1] 75.271
```

```r
Rule1 <- Heights > 56.9 & Heights < 66.9
Rule2 <- Heights > 52.4 & Heights < 70.7
Rule3 <- Heights > 47.8 & Heights < 75.3
length(which(Rule1==TRUE))/25
```

```
## [1] 0.68
```

```r
length(which(Rule2==TRUE))/25
```

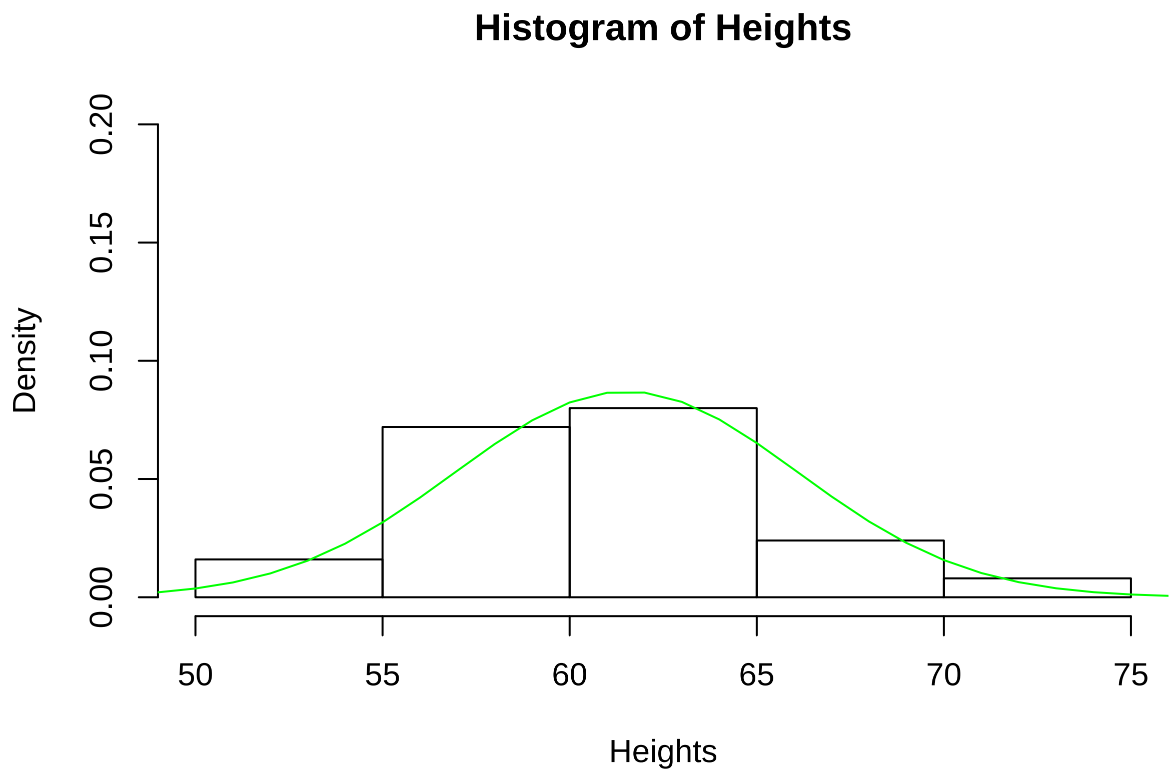
```
## [1] 0.96
```

```r
length(which(Rule3==TRUE))/25
```

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

a) Based on this information, it follows the 68-95-99.7 rule.

```r
hist(Heights, probability = TRUE, ylim = c(0, 0.20))

x <- 1:100
y <- dnorm(x = x, mean = mean(Heights), sd = sd(Heights))
lines(x = x, y = y, col = "green")
```



b) It looks like the both of the data does follow a normal distribution. I would use the first one to visual where the mean is and I would use the second one to find the range of the distribution.

3.22 Defective rate.

```
.98^9 *.02
```

```
## [1] 0.01667496
```

```
.98^100
```

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

```
1/.02
```

```
## [1] 50
```

```
sqrt(2*98)
```

```
## [1] 14
```

```
1/.05
```

```
## [1] 20
```

```
sqrt(5*95)
```

```
## [1] 21.79449
```

a) 1.67% that the first one defective is the 10th stansistor.

b) 13.3 % that the machine produces no defective transistors in a batch of 100. c)50 units is expected to be produced before the first defect and the standard deviation is 14. d)20 units is expected to be produced before the first defect and the standard deviation is 21.79. e)increasing the probablity of defect increases the standard deviation and increases the time before first defect.

3.38 Male children.

```
outcomes <-c("boy","girl")
child <- sample(outcomes, size = 3,replace = TRUE,prob =c(.51,.49))
table(child)
```

```
## child
##  boy girl
##    1    2
```

```
(.51*.49*.51) + (.51*.49*.51) + (.51*.49*.51)
```

```
## [1] 0.382347
```

probabily of having 2 boys out of 3 children is 38.23%. If we try to calculate 3 boys out of 8 children, the odds would decrease because since the chances of having a boy is higher and the couple wants only 3 of 8 to be a boy.

3.42 Serving in volleyball.

```
.15^7*.85^2 *.85
```

```
## [1] 1.04929e-06
```

a) .000104929%
b) .15%
c) Since each serve is an independent event, her chances are still 15% of success.