

# Assignment Exercise 16-18

Alex

10/10/2019

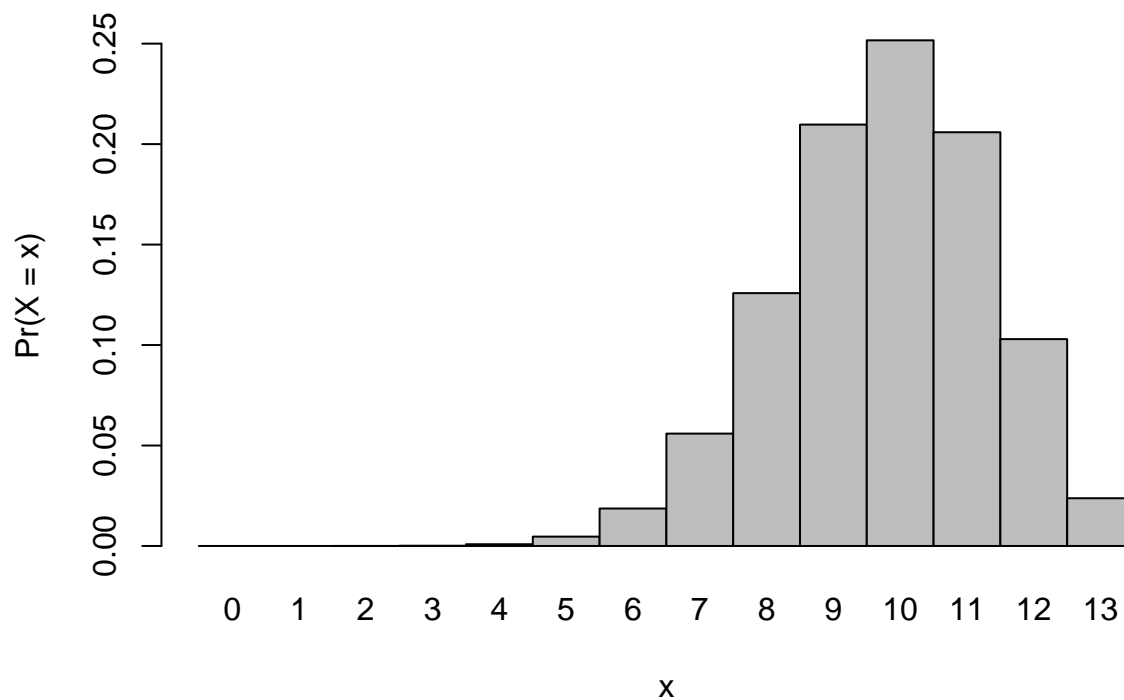
```
## Chapter 16 Exercise 16.1
```

```
# Question a
```

```
# 13 viewing platform, Probability = 75/100 or 0.75
```

```
Bdis<- dbinom(x=0:13,size=13,prob=0.75)
```

```
barplot(Bdis,names.arg=0:13,space=0,xlab="x",ylab="Pr(X = x)")
```



```
# Question b
```

```
#Probability to see birds at all times is Binomial dist
```

```
Bdis2<- dbinom(x=13,size=13,prob=0.75)
```

```
Bdis2
```

```
## [1] 0.02375726
```

```
#Question c
```

```
#Probability of more than 9 is Pr(X>9), q = 9
```

```
1-pbinom(q=9,size=13,prob=0.75)
```

```
## [1] 0.5842527
```

```
#Question d  
# Probability of  $8 \leq P \leq 11$   
# cumulative probability of d-function  
sum(dbinom(x=8:11,size=13,prob=0.75))
```

```
## [1] 0.793082
```

```
# probability using p-function q  
pbinom(q=11,size=13,prob=0.75)-pbinom(q=7,size=13,prob=0.75)
```

```
## [1] 0.793082
```

```
#Question e  
#Probability of cumulatively seeing less than 9  
pbinom(q=8, size=13,prob=0.75)
```

```
## [1] 0.2060381
```

```
#Question f  
#Probability of 10 random visits  
Visits <- rbinom(n= 10,size=13,prob=0.75)  
Visits
```

```
## [1] 10 9 10 11 11 8 11 12 6 10
```

```
#Question g  
MnP <- 13*0.75  
MnP
```

```
## [1] 9.75
```

```
stndD <- sqrt(MnP*0.25)  
stndD
```

```
## [1] 1.561249
```

```
## Exercise 16.2
```

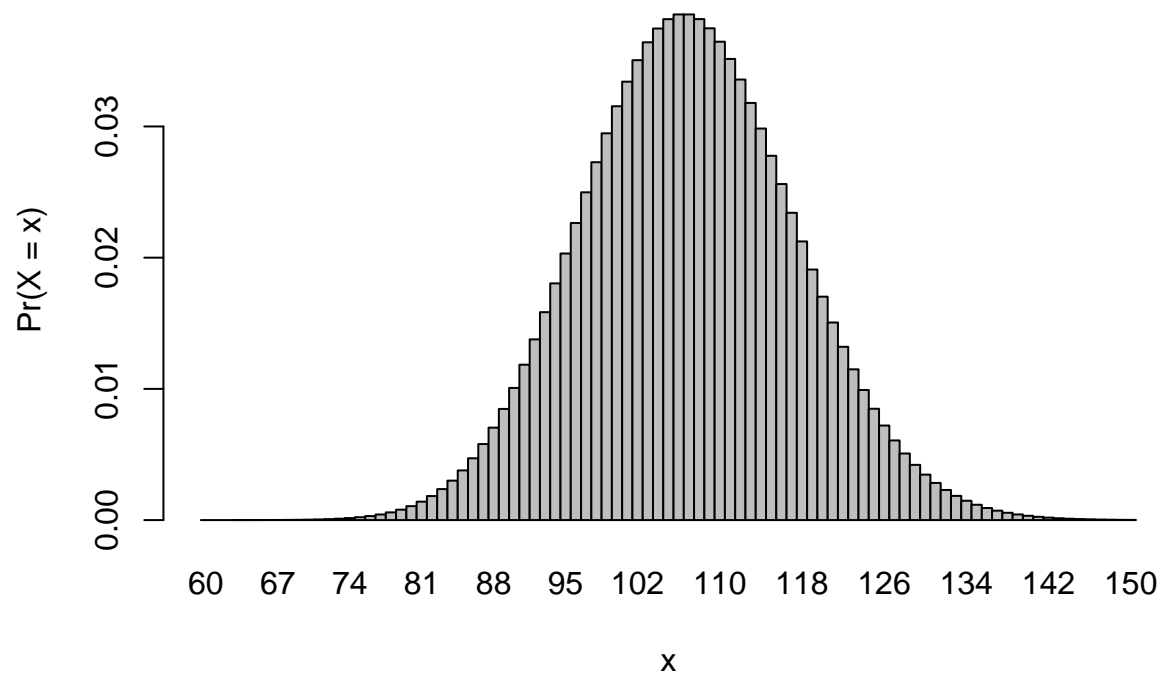
```
#Question a  
1-ppois(q=100,lambda=107)
```

```
## [1] 0.7319128
```

```
#Question b  
dpois(x=0,lambda=107)
```

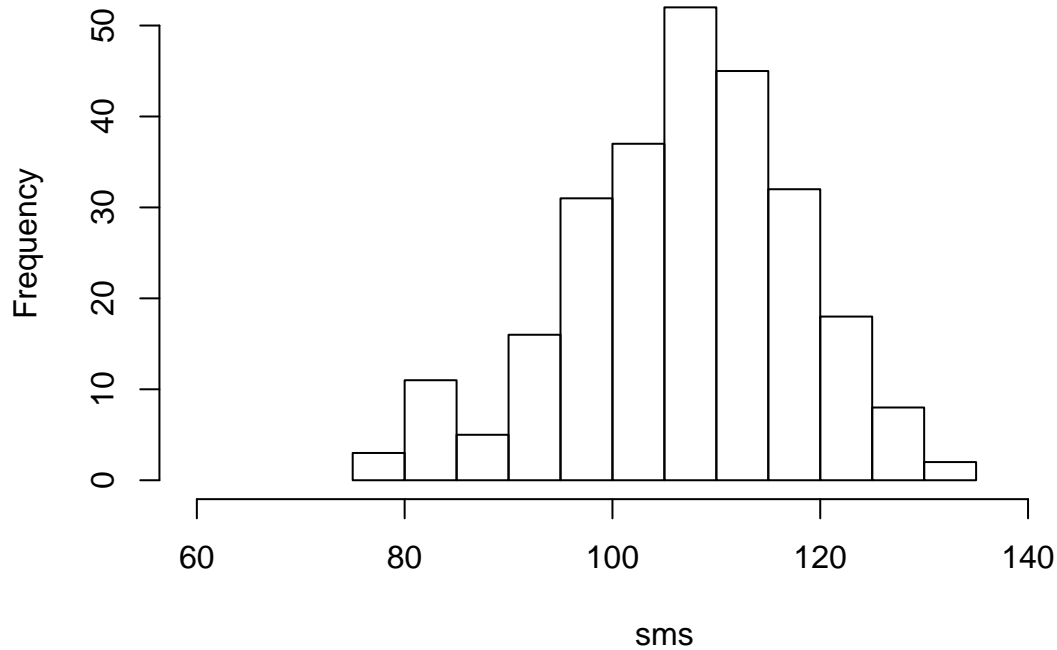
```
## [1] 3.39227e-47
```

```
#Question c
#plotting 60<=x<=150
barplot(dpois(x=60:150,lambda=107),names.arg=60:150,space=0,xlab="x",ylab="Pr(X = x)")
```



```
#Question d
sms <- rpois(n=260,lambda=107)
hist(sms,xlim=c(60,150))
```

## Histogram of sms



```
## Exercise 16.3
#min tree height 3, and max tree height 70
Treeh1 <- 3
Treeh2 <- 70
# Question a
punif(q=5.5,min=Treeh1,max=Treeh2)
```

```
## [1] 0.03731343
```

```
#Question b
qunif(p=1-0.15,min=Treeh1,max=Treeh2)
```

```
## [1] 59.95
```

```
#Question c
#mean of tree heights
mTree <- (Treeh1+Treeh2)/2
mTree
```

```
## [1] 36.5
```

```
# standard deviation
stdTrees <- sqrt((Treeh2-Treeh1)^2/12)
stdTrees
```

```
## [1] 19.34123
```

```
#Question d
```

```
punif(mTree+0.5*stdTrees,Treeh1,Treeh2) - punif(mTree-0.5*stdTrees,Treeh1,Treeh2)
```

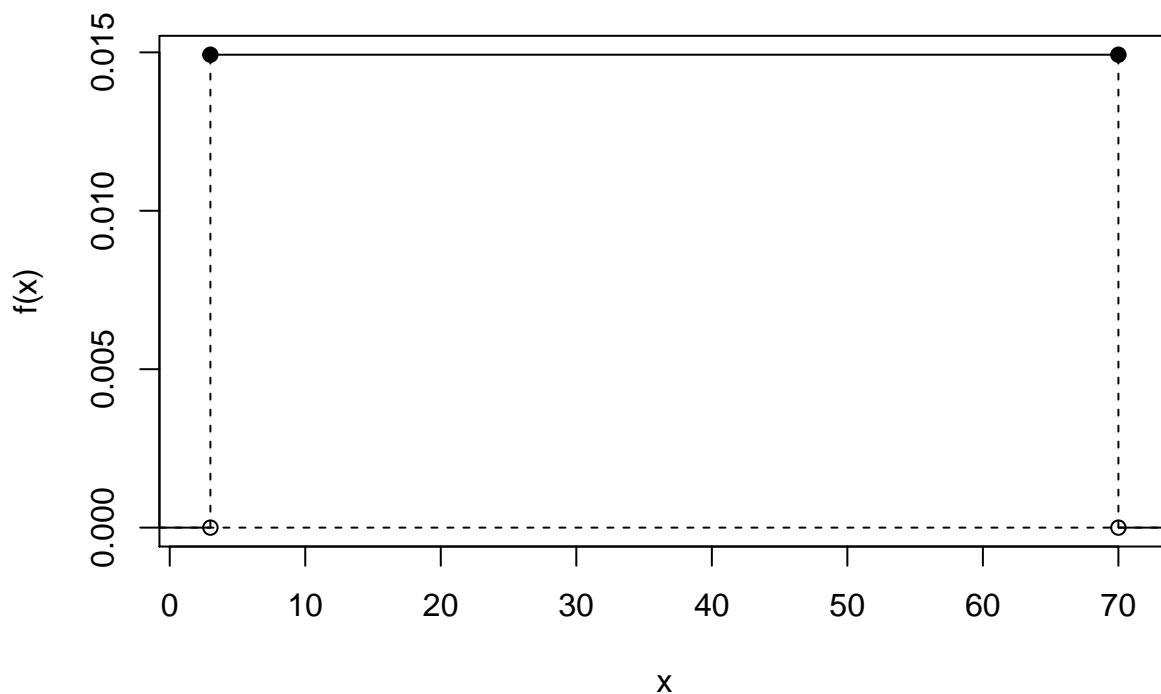
```
## [1] 0.2886751
```

```
#Question e
```

```
dens <- dunif(mTree,Treeh1,Treeh2)  
dens
```

```
## [1] 0.01492537
```

```
plot(c(Treeh1,Treeh2),rep(dens,2),type="o",pch=19,xlim=c(Treeh1-1,Treeh2+1),ylim=c(0,dens),ylab="f(x)",  
abline(h=0,lty=2)  
segments(c(Treeh1-5,Treeh2+5,Treeh1,Treeh2),rep(0,4),rep(c(Treeh1,Treeh2),2),rep(c(0,dens),each=2),lty=2),  
points(c(Treeh1,Treeh2),c(0,0))
```



```
#Question f
```

```
simTreeh1 <- runif(n=10,min=Treeh1,max=Treeh2)  
simTreeh1
```

```
## [1] 5.478355 16.165529 51.366896 58.569852 54.068964 15.547940 17.824444  
## [8] 40.347898 18.516190 26.891039
```

```
quan1 <- quantile(x=simTreeh1,prob=1-0.15)
quan1
```

```
##      85%
## 53.12324
```

```
simTreeh2 <- runif(n=1000,min=Treeh1,max=Treeh2)
quan2 <- quantile(x=simTreeh2,prob=1-0.15)
quan2
```

```
##      85%
## 59.44521
```

*# In all, both estimates seem to be centered on the 'true' value from (b), but those based on simulation*

*## Exercise 16.4*

*#Question a*

```
m <- 17
stdv <- 4.5
# i
1-pnorm(20,m,stdv)
```

```
## [1] 0.2524925
```

*# ii*

```
pnorm(10,m,stdv)-pnorm(5,m,stdv)
```

```
## [1] 0.05607653
```

*# iii*

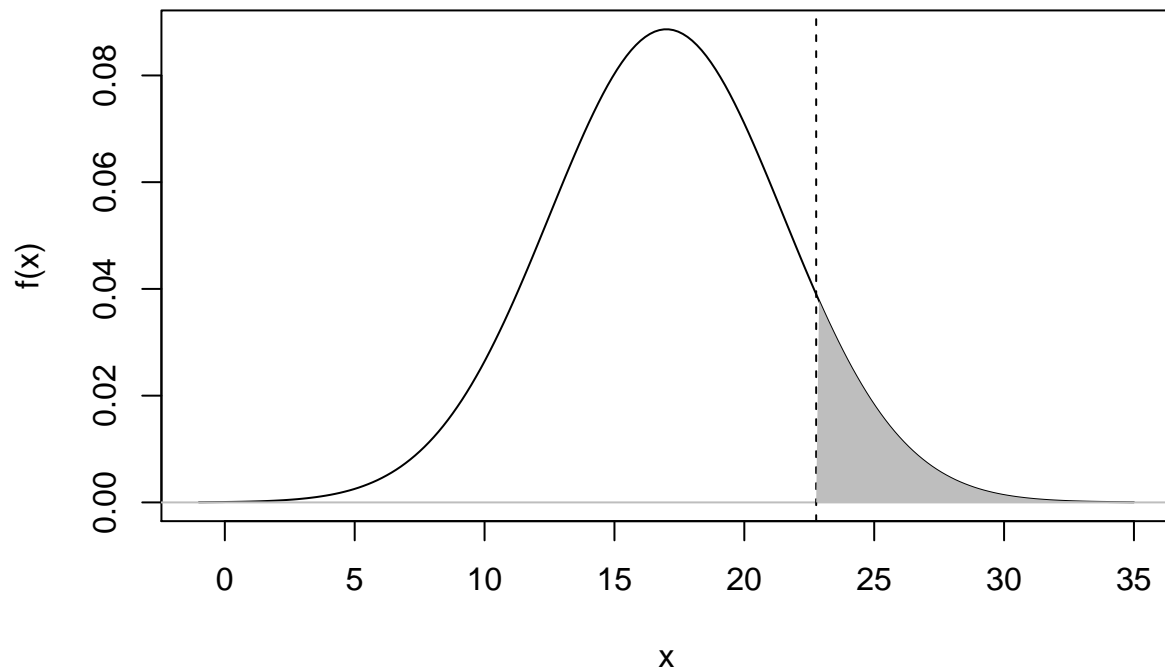
```
slw10 <- qnorm(1-0.1,m,stdv)
slw10
```

```
## [1] 22.76698
```

*# iv*

```
xvals <- seq(m-4*stdv,m+4*stdv,length=200)
f <- dnorm(xvals,m,stdv)
xsub <- xvals[xvals>=slw10]
fsub <- f[xvals>=slw10]
plot(xvals,f,type="l",main="N(17,4.5) distribution",xlab="x",ylab="f(x)")
abline(h=0,col="gray")
abline(v=slw10,lty=2)
polygon(rbind(c(slw10,0),cbind(xsub,fsub),c(max(xvals),0)),border=NA,col="gray")
```

## N(17,4.5) distribution



```
# v
rnorm(10,m,stdv)
```

```
## [1] 18.84411 18.39804 21.35697 22.36821 21.58974 15.68306 18.13435
## [8] 10.39036 14.17182 14.92251
```

```
# Question b
m <- 10
stdv <- sqrt(2)
# i
pnorm(11,m,stdv)-pnorm(9.5,m,stdv)
```

```
## [1] 0.3984131
```

```
# ii
standValue9.5 <- (9.5-m)/stdv
standValue9.5
```

```
## [1] -0.3535534
```

```
standValue11 <- (11-m)/stdv
standValue11
```

```
## [1] 0.7071068
```

```
pnorm(standValue11)-pnorm(standValue9.5)
```

```
## [1] 0.3984131
```

```
# iii  
shortest2.5 <- qnorm(0.025,m,stdv)  
shortest2.5
```

```
## [1] 7.228192
```

```
# iv  
(shortest2.5-m)/stdv
```

```
## [1] -1.959964
```

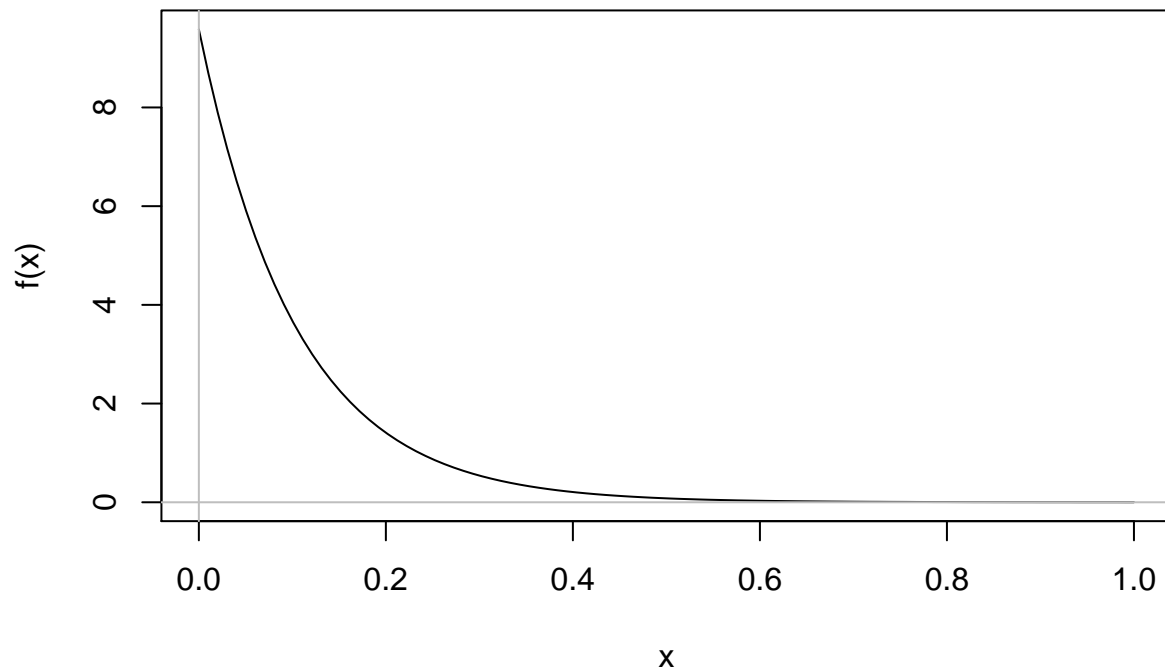
```
## Exercise 16.5  
#Question a  
# i  
lambda.d <- 3500/365.25  
lambda.d
```

```
## [1] 9.582478
```

```
# ii  
xvals <- seq(0,1,length=100)  
plot(xvals,dexp(xvals,lambda.d),type="l",xlab="x",ylab="f(x)",main="EXP(0.89) distribution")  
abline(h=0,col="gray")  
abline(v=0,col="gray")
```



## EXP(0.89) distribution



```
# iii  
pexp(q=0.5/24,rate=lambda.d)
```

```
## [1] 0.1809703
```

```
# iv  
qexp(p=1-0.1,rate=lambda.d)*24
```

```
## [1] 5.766989
```

```
# Question b  
# i  
pexp(q=5,1/11)
```

```
## [1] 0.3652636
```

```
# ii  
pexp(q=6,1/9)
```

```
## [1] 0.4865829
```

```
# iii
1-pexp(q=15,1/11)
```

```
## [1] 0.2557292
```

```
1-pexp(q=15,1/9)
```

```
## [1] 0.1888756
```

## Exercise 17.1

```
stndM <- 41.1
#Question a
StndE <- 11.3/sqrt(6)
StandE
```

```
## [1] 4.613206
```

```
# Question b
pnorm(55,mean=stndM,sd=StandE)-pnorm(45,mean=stndM,sd=StandE)
```

```
## [1] 0.197651
```

```
#Question c
pnorm(32.5,stndM,StandE)
```

```
## [1] 0.03114587
```

```
#Question d
SurveyTeens<- 140*0.35
SurveyTeens
```

```
## [1] 49
```

```
SamplingD<-140*(1-0.35)
SamplingD
```

```
## [1] 91
```

```
# Both are greater than 5 so, meaning using the normal distribution to represent the sampling distribut
#Question e
1-pnorm(0.4,mean=0.35,sd=sqrt(0.35*0.65/140))
```

```
## [1] 0.1074235
```

```
#Question f  
qnorm(0.9,0.35,sqrt(0.35*0.65/140))
```

```
## [1] 0.401661
```

```
# Upper limit - tail above this value has probability 0.1  
qnorm(0.1,0.35,sqrt(0.35*0.65/140))
```

```
## [1] 0.298339
```

```
# Lower limit - tail below this value has probability 0.1. Together, these two limits therefore mark of  
#Question g  
# Even though raw data are not normal, sample size is large (n>30 by rule-of-thumb). Standard deviation  
stdE <- 34.51/sqrt(63)  
stdE
```

```
## [1] 4.347851
```

```
#Question h  
# i  
1-pt((40-37.8)/stdE,df=62)
```

```
## [1] 0.3073266
```

```
# ii  
pt((30-37.8)/stdE,df=62)
```

```
## [1] 0.03884552
```

```
# iii  
pt((40-37.8)/stdE,df=62)-0.5
```

```
## [1] 0.1926734
```

```
## Exercise 17.2
```

```
sprintM <- 14.22  
stdD <- 2.9
```

```
# Question a  
#90% confidence Interval 0.95  
sprintM+c(-1,1)*qnorm(0.95)*stdD/sqrt(34)
```

```
## [1] 13.40194 15.03806
```

```
# Question b  
sprintM+c(-1,1)*qt(0.95,df=33)*stdD/sqrt(34)
```

```
## [1] 13.37831 15.06169
```

```
# Question c
Lhand <- 37/400
Lhand
```

```
## [1] 0.0925
```

```
Lhand+c(-1,1)*qnorm(0.995)*sqrt(Lhand*(1-Lhand)/400)
```

```
## [1] 0.05518519 0.12981481
```

```
# Question d
Lhand <- (37+11)/400
Lhand
```

```
## [1] 0.12
```

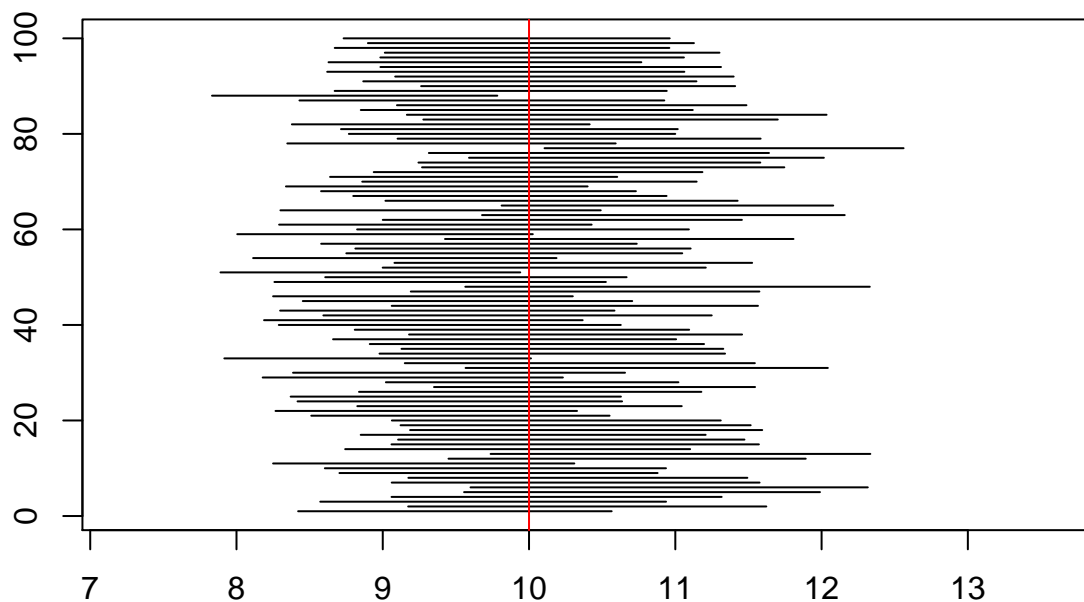
```
Lhand+c(-1,1)*qnorm(0.995)*sqrt(Lhand*(1-Lhand)/400)
```

```
## [1] 0.07814773 0.16185227
```

```
# Question e
CImat <- matrix(NA,nrow=5000,ncol=3)
Size <- 300
lambda.e <- 0.1
M <- 1/lambda.e
for(i in 1:5000){
  samples <- rexp(Size,rate=lambda.e)
  sampleci <- mean(samples)+c(-1,1)*qt(0.975,Size-1)*sd(samples)/sqrt(Size)
  CImat[i,1:2] <- sampleci
  CImat[i,3] <- M>=sampleci[1] && M<=sampleci[2]
}
mean(CImat[,3])
```

```
## [1] 0.9504
```

```
# Question f
plot(CImat[1,1:2],c(1,1),xlim=range(CImat[,1:2]),ylim=c(1,100),type="l",xlab="",ylab="")
for(i in 2:100){
  lines(CImat[i,1:2],c(i,i))
}
abline(v=M,col=2)
```



## Exercise 18.1

```
#Question a
# H0: mu = 3.5; HA: mu != 3.5 (two-sided test)
Hstat <- (3.97-3.5)/(2.21/sqrt(73))
Hstat
```

```
## [1] 1.817051
```

```
pt(-Hstat,df=72)+(1-pt(Hstat,df=72))
```

```
## [1] 0.07337077
```

```
# p-value is around 0.073, this is > than alpha=0.05, therefore insufficient evidence to reject the nul
# Question b
# H0: mu = 4.3; HA: mu > 4.3 (one-sided test)
data("quakes")
Result<- t.test(quakes$mag,mu=4.3,alternative="greater",conf.level=0.99)
Result
```

```
##
## One Sample t-test
```

```
##
## data:  quakes$mag
## t = 25.155, df = 999, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 4.3
## 99 percent confidence interval:
##  4.590722      Inf
## sample estimates:
## mean of x
##    4.6204

# p-value very small; strong evidence to reject the null. There is evidence to suggest that the true me
#Question c
MQ<-mean(quakes$mag)+c(-1,1)*qt(0.995,df=999)*sd(quakes$mag)/sqrt(1000)
MQ

## [1] 4.587529 4.653271

## Exercise 18.2
# Question a
library("MASS")
?anorexia

## starting httpd help server ... done

Resul<- t.test(anorexia[,3],anorexia[,2],alternative="greater",paired=TRUE)
Resul

##
## Paired t-test
##
## data:  anorexia[, 3] and anorexia[, 2]
## t = 2.9376, df = 71, p-value = 0.002229
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  1.195825      Inf
## sample estimates:
## mean of the differences
##                2.763889

# From the followmg p-value ~0.0023. Less than 0.05; some evidence to reject H0. No evidence to suggest
# Question b
# conducting three separte hypothesis
H1 <- t.test(anorexia[anorexia$Treat=="Cont",3],anorexia[anorexia$Treat=="Cont",2],alternative="greater")
H1

##
## Paired t-test
##
## data:  anorexia[anorexia$Treat == "Cont", 3] and anorexia[anorexia$Treat == "Cont", 2]
## t = -0.28723, df = 25, p-value = 0.6118
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
```

```
## -3.126168      Inf
## sample estimates:
## mean of the differences
## -0.45
```

```
H2<- t.test(anorexia[anorexia$Treat=="CBT",3],anorexia[anorexia$Treat=="CBT",2],alternative="greater",p
H2
```

```
##
## Paired t-test
##
## data: anorexia[anorexia$Treat == "CBT", 3] and anorexia[anorexia$Treat == "CBT", 2]
## t = 2.2156, df = 28, p-value = 0.01751
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 0.6981979      Inf
## sample estimates:
## mean of the differences
## 3.006897
```

```
H3<- t.test(anorexia[anorexia$Treat=="FT",3],anorexia[anorexia$Treat=="FT",2],alternative="greater",pai
H3
```

```
##
## Paired t-test
##
## data: anorexia[anorexia$Treat == "FT", 3] and anorexia[anorexia$Treat == "FT", 2]
## t = 4.1849, df = 16, p-value = 0.0003501
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 4.233975      Inf
## sample estimates:
## mean of the differences
## 7.264706
```

```
# From the results of H1, H2 and H3 there is no statistical evidence to reject the claim that there is
# Question c
```

```
data("PlantGrowth")
?PlantGrowth
controlGroup <- PlantGrowth$weight[PlantGrowth$group=="ctrl"]
controlGroup
```

```
## [1] 4.17 5.58 5.18 6.11 4.50 4.61 5.17 4.53 5.33 5.14
```

```
treatedGroup <- PlantGrowth$weight[PlantGrowth$group!="ctrl"]
treatedGroup
```

```
## [1] 4.81 4.17 4.41 3.59 5.87 3.83 6.03 4.89 4.32 4.69 6.31 5.12 5.54 5.50
## [15] 5.37 5.29 4.92 6.15 5.80 5.26
```

```
# H0: mu_control - mu_treated = 0; HA: mu_control - mu_treated < 0
max(c(sd(controlGroup),sd(treatedGroup)))/min(c(sd(controlGroup),sd(treatedGroup)))
```

```
## [1] 1.315153
```

```
# Ratio of (large sd) / (small sd) is less than 2 so use pooled variance according to rule-of-thumb.
# Question d
ResG <- t.test(x=controlGroup,y=treatedGroup,alternative="less",var.equal=TRUE)
ResG
```

```
##
## Two Sample t-test
##
## data: controlGroup and treatedGroup
## t = -0.22272, df = 28, p-value = 0.4127
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 0.4082382
## sample estimates:
## mean of x mean of y
##      5.0320      5.0935
```

```
# Large p-value ~0.41. There is no evidence to reject H0. There is insufficient evidence to conclude th
```

```
# Question e
```

```
# A wrapper function
```

```
test <- function(x,y,paired=FALSE,var.equal=FALSE,...){
  if(!paired){
    stdX <- sd(x)
    stdY <- sd(y)
    big <- max(c(stdX,stdY))
    small <- min(c(sdX,sdY))
    var.equal <- (big/small)<2
  }
  return(t.test(x=x,y=y,paired=paired,var.equal=var.equal,...))
}
```

```
# Question f
```

```
# Snack Packet Example 1
```

```
snacks <- c(87.7,80.01,77.28,78.76,81.52,74.2,80.71,79.5,77.87,81.94,80.7,82.32,
            75.78,80.19,83.91,79.4,77.52,77.62,81.4,74.89,82.95,73.59,77.92,77.18,
            79.83,81.23,79.28,78.44,79.01,80.47,76.23,78.89,77.14,69.94,78.54,79.7,
            82.45,77.29,75.52,77.21,75.99,81.94,80.41,77.7)
snacks2 <- c(80.22,79.73,81.1,78.76,82.03,81.66,80.97,81.32,80.12,78.98,79.21,
            81.48,79.86,81.06,77.96,80.73,80.34,80.01,81.82,79.3,79.08,79.47,
            78.98,80.87,82.24,77.22,80.03,79.2,80.95,79.17,81)
SnackResult<- t.test(x=snacks2,y=snacks,alternative="greater",conf.level=0.9)
SnackResult
```

```
##
## Welch Two Sample t-test
##
## data: snacks2 and snacks
## t = 2.4455, df = 60.091, p-value = 0.008706
```



```
## alternative hypothesis: true difference in means is greater than 0
## 90 percent confidence interval:
## 0.5859714      Inf
## sample estimates:
## mean of x mean of y
## 80.15710 78.91068
```

```
# IQ scores between men and women Example 2
men <- c(102,87,101,96,107,101,91,85,108,67,85,82)
women <- c(73,81,111,109,143,95,92,120,93,89,119,79,90,126,62,92,77,106,105,111)
TestResult<- t.test(x=men,y=women,alternative="two.sided",conf.level=0.95)
TestResult
```

```
##
## Welch Two Sample t-test
##
## data: men and women
## t = -1.0571, df = 29.98, p-value = 0.2989
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -17.543272 5.576605
## sample estimates:
## mean of x mean of y
## 92.66667 98.65000
```

```
# Drug Efficacy to reduce heart rate Example 3
rate.before <- c(52,66,89,87,89,72,66,65,49,62,70,52,75,63,65,61)
rate.after <- c(51,66,71,73,70,68,60,51,40,57,65,53,64,56,60,59)
t.test(x=rate.after,y=rate.before,alternative="less",paired=TRUE,conf.level=0.95)
```

```
##
## Paired t-test
##
## data: rate.after and rate.before
## t = -4.8011, df = 15, p-value = 0.0001167
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.721833
## sample estimates:
## mean of the differences
##      -7.4375
```

```
## Exercise 18.3
```

```
# Question a
# H0: p=0.9; HA: p<0.9
Num <- 89
Result<- Num*0.9
Result
```

```
## [1] 80.1
```

```
Result2<-Num*0.1
Result2
```

```
## [1] 8.9
```

```
# Both np and n(p-1) > 5 so OK to continue with normal distribution according to the rule-of-thumb.
# Question b
pstat <- 71/Num
pstat
```

```
## [1] 0.7977528
```

```
R <- (pstat-0.9)/(sqrt(0.9*0.1/Num))
R
```

```
## [1] -3.215327
```

```
pnorm(R)
```

```
## [1] 0.0006514802
```

```
prop.test(x=71,n=Num,p=0.9,alternative="less",conf.level=0.9,correct=FALSE)
```

```
##
## 1-sample proportions test without continuity correction
##
## data: 71 out of Num, null probability 0.9
## X-squared = 10.338, df = 1, p-value = 0.0006515
## alternative hypothesis: true p is less than 0.9
## 90 percent confidence interval:
## 0.0000000 0.8466949
## sample estimates:
## p
## 0.7977528
```

```
# p-value very small; less than 0.1. There is evidence to reject H0 and conclude the true proportion of
# Question c
pstat+c(-1,1)*qnorm(0.95)*sqrt(pstat*(1-pstat)/Num)
```

```
## [1] 0.7277190 0.8677866
```

```
# Question d
x1 <- 97
Num1 <- 445
pstat1 <- x1/Num1
pstat1
```

```
## [1] 0.2179775
```

```
x2 <- 90
Num2 <- 419
pstat2 <- x2/Num2
pstat2
```

```
## [1] 0.2147971
```

```
p.star <- (x1+x2)/(Num1+Num2)
p.star
```

```
## [1] 0.2164352
```

```
Z <- (pstat2-pstat1)/sqrt(p.star*(1-p.star)*(1/Num1+1/Num2))
Z
```

```
## [1] -0.1134513
```

```
Result<- 2*pnorm(Z)
Result
```

```
## [1] 0.9096728
```

```
# p-value is very large; much greater than 0.05. No evidence to reject H0. Retain H0 and conclude there
# Question e
(pstat2-pstat1)+c(-1,1)*qnorm(0.975)*sqrt(p.star*(1-p.star)*(1/Num1+1/Num2))
```

```
## [1] -0.05812427 0.05176349
```

```
# We are 95% confident that the true difference in the proportion of support between State 2 and State
# Question f
Z.test <- function(p1,n1,p2=NULL,n2=NULL,p0,alternative="two.sided",conf.level=0.95){
  if(is.null(p2)||is.null(n2)){
    cat("One-sample Z-test.\n")
    if(p1*n1<=5||n1*(1-p1)<=5){
      warning("Normal distribution may not be valid; np or n(1-p) <= 5 detected.")
    }
    Z <- (p1-p0)/sqrt(p0*(1-p0)/n1)
    CI <- (p1)+c(-1,1)*qnorm(conf.level+(1-conf.level)/2)*sqrt(p0*(1-p0)/n1)
  } else {
    cat("Two-sample Z-test.\n")
    if(p1*n1<=5||n1*(1-p1)<=5||p2*n2<=5||n2*(1-p2)<=5){
      warning("Normal distribution may not be valid; np or n(1-p) <= 5 detected.")
    }
    p.star <- (p1*n1+p2*n2)/(n1+n2)
    Z <- (p1-p2-p0)/sqrt(p.star*(1-p.star)*(1/n1+1/n2))
    CI <- sort((p1-p2)+c(-1,1)*qnorm(conf.level+(1-conf.level)/2)*sqrt(p.star*(1-p.star)*(1/n1+1/n2)))
  }

  P <- pnorm(Z)
  if(alternative=="greater"){
```

```

    P <- 1-P
  } else if(alternative=="two.sided"){
    if(Z<0){
      P <- 2*P
    } else {
      P <- 2*(1-P)
    }
  }
  return(list(Z=Z,P=P,CI=CI))
}
# Question g
# Sick People Example 1
sick <- c(0,0,1,1,0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,1,1,1,0,0,0,1)
Z.test(p1=mean(sick),n1=length(sick),p0=0.2,alternative="two.sided",conf.level=0.95)

```

```
## One-sample Z-test.
```

```

## $Z
## [1] 1.021324
##
## $P
## [1] 0.3071008
##
## $CI
## [1] 0.1302796 0.4214445

```

```

# Psychology and Geography studentsExample 2
Psych.x1 <- 180
Psych.n1 <- 233
p.hat1 <- Psych.x1/Psych.n1
Geo.x2 <- 175
Geo.n2 <- 197
p.hat2 <- Geo.x2/Geo.n2
Result<-Z.test(p.hat2,Geo.n2,p.hat1,Psych.n1,p0=0,alternative="greater",conf.level=0.95)

```

```
## Two-sample Z-test.
```

```
Result
```

```

## $Z
## [1] 3.152693
##
## $P
## [1] 0.0008088606
##
## $CI
## [1] 0.04380676 0.18777861

```

```

# ...or you could flip the order of differencing and use alternative="less"
# Question h
WarningMess<- Z.test(p1=0.11,n1=10,p0=0.1)

```

```
## One-sample Z-test.
```

```
## Warning in Z.test(p1 = 0.11, n1 = 10, p0 = 0.1): Normal distribution may
## not be valid; np or n(1-p) <= 5 detected.
```

```
WarningMess
```

```
## $Z
## [1] 0.1054093
##
## $P
## [1] 0.9160511
##
## $CI
## [1] -0.07593851 0.29593851
```

```
## Exercise 18.4
```

```
# Question a
```

```
#H0: No relationship between hair and eye color; HA: There is a relationship.
```

```
data("HairEyeColor")
```

```
?HairEyeColor
```

```
TestofInd<- chisq.test(x=HairEyeColor[,1]+HairEyeColor[,2])
```

```
TestofInd
```

```
##
```

```
## Pearson's Chi-squared test
```

```
##
```

```
## data: HairEyeColor[, , 1] + HairEyeColor[, , 2]
```

```
## X-squared = 138.29, df = 9, p-value < 2.2e-16
```

```
# Very small P-value. Very strong evidence against the null. Reject H0 and conclude there does appear to be a relationship.
```

```
# Question b
```

```
#H0: p1=p2=p3=1/3; HA: H0 is incorrect
```

```
library("car")
```

```
## Loading required package: carData
```

```
data("Duncan")
```

```
?Duncan
```

```
jobtype <- Duncan$type
```

```
jobtype
```

```
## [1] prof prof prof prof prof prof prof prof prof wc prof prof prof prof prof
## [15] prof wc prof prof prof prof wc wc wc wc bc bc bc bc bc
## [29] bc bc bc bc bc bc bc bc bc bc bc bc bc bc bc
## [43] bc bc bc
## Levels: bc prof wc
```

```
jobtype.tab <- table(jobtype)
```

```
jobtype.tab
```

```
## jobtype
##   bc prof   wc
##   21  18    6
```

```
chisq.test(jobtype.tab)
```

```
##
## Chi-squared test for given probabilities
##
## data:  jobtype.tab
## X-squared = 8.4, df = 2, p-value = 0.015
```

```
# i With a significance level of 0.05 and a p-value of 0.015, there is weak evidence to reject H0 and w
# ii With a significance level of 0.01 and a P-value of 0.015, there is no evidence to reject H0 and we
```

```
## Exercise 18.5
```

```
# Question a
```

```
typeI.mean <- function(mu0,sigma,n,alpha,test="two.sided",ITERATIONS=10000){
  tstats <- rep(NA,ITERATIONS)
  for(i in 1:ITERATIONS){
    temporary.sample <- rnorm(n=n,mean=mu0,sd=sigma)
    temporary.mean <- mean(temporary.sample)
    temporary.sd <- sd(temporary.sample)
    tstats[i] <- (temporary.mean-mu0)/(temporary.sd/sqrt(n))
  }
  pvals <- pt(tstats,df=n-1)
  if(test=="less"){
    return(mean(pvals<alpha))
  } else if(test=="greater"){
    return(mean((1-pvals)<alpha))
  } else if(test=="two.sided"){
    result <- pvals
    result[tstats>0] <- 1-pvals[tstats>0]
    return(mean(result<alpha/2))
  } else {
    stop("'test' argument not recognised")
  }
}
# i
typeI.mean(mu0=0,sigma=1,n=40,alpha=0.05,test="less")
```

```
## [1] 0.0461
```

```
typeI.mean(mu0=0,sigma=1,n=40,alpha=0.05,test="greater")
```

```
## [1] 0.0475
```

```
typeI.mean(mu0=0,sigma=1,n=40,alpha=0.05,test="two.sided")
```

```
## [1] 0.0449
```

```
# ii
typeI.mean(mu0=-4,sigma=0.3,n=60,alpha=0.01,test="less")
```

```
## [1] 0.0101
```

```
typeI.mean(mu0=-4,sigma=0.3,n=60,alpha=0.01,test="greater")
```

```
## [1] 0.0093
```

```
typeI.mean(mu0=-4,sigma=0.3,n=60,alpha=0.01,test="two.sided")
```

```
## [1] 0.0098
```

```
# Question b
typeII.mean <- function(mu0,muA,sigma,n,alpha,test="two.sided",ITERATIONS=10000){
  tstats <- rep(NA,ITERATIONS)
  for(i in 1:ITERATIONS){
    temporary.sample <- rnorm(n=n,mean=muA,sd=sigma)
    temporary.mean <- mean(temporary.sample)
    temporary.sd <- sd(temporary.sample)
    tstats[i] <- (temporary.mean-mu0)/(temporary.sd/sqrt(n))
  }
  pvals <- pt(tstats,df=n-1)
  if(test=="less"){
    return(mean(pvals>=alpha))
  } else if(test=="greater"){
    return(mean((1-pvals)>=alpha))
  } else if(test=="two.sided"){
    result <- pvals
    result[tstats>0] <- 1-pvals[tstats>0]
    return(mean(result>=alpha/2))
  } else {
    stop("'test' argument not recognised")
  }
}
# i
typeII.mean(mu0=-3.2,muA=-3.3,sigma=0.1,n=25,alpha=0.05,test="two.sided")
```

```
## [1] 0.003
```

```
# ii
typeII.mean(mu0=8994,muA=5600,sigma=3888,n=9,alpha=0.01,test="less")
```

```
## [1] 0.5622
```

```
# iii
typeII.mean(mu0=0.44,muA=0.4,sigma=2.4,n=68,alpha=0.05,test="greater")
```

```
## [1] 0.9645
```

```
## Exercise 18.6
```

```
# Question a
```

```
power.mean <- function(nvec,...){  
  nlen <- length(nvec)  
  result <- rep(NA,nlen)  
  pbar <- txtProgressBar(min=0,max=nlen,style=3)  
  for(i in 1:nlen){  
    result[i] <- 1-typeII.mean(n=nvec[i],...)  
    setTxtProgressBar(pbar,i)  
  }  
  close(pbar)  
  return(result)  
}
```

```
# i
```

```
power.mean(nvec=50,mu0=10,muA=10.5,sigma=0.9,alpha=0.01,test="two.sided")
```

```
##  
|  
|                                     | 0%  
|  
|=====| 100%
```

```
## [1] 0.8816
```

```
# ii
```

```
power.mean(nvec=44,mu0=80,muA=78.5,sigma=3.1,alpha=0.05,test="less") #Yes, seems statistically powerful
```

```
##  
|  
|                                     | 0%  
|  
|=====| 100%
```

```
## [1] 0.9346
```

```
power.mean(nvec=44,mu0=80,muA=78.5,sigma=3.1,alpha=0.01,test="less") #No, power appears less than 80%,
```

```
##  
|  
|                                     | 0%  
|  
|=====| 100%
```

```
## [1] 0.7835
```

```
# Question b
```

```
sample.sizes <- 5:100  
pow <- power.mean(nvec=sample.sizes,mu0=80,muA=78.5,sigma=3.1,alpha=0.05,test="less")
```



##

|       |  |     |
|-------|--|-----|
|       |  | 0%  |
|       |  |     |
| =     |  | 1%  |
|       |  |     |
| =     |  | 2%  |
|       |  |     |
| ==    |  | 3%  |
|       |  |     |
| ===   |  | 4%  |
|       |  |     |
| ===   |  | 5%  |
|       |  |     |
| ====  |  | 6%  |
|       |  |     |
| ===== |  | 7%  |
|       |  |     |
| ===== |  | 8%  |
|       |  |     |
| ===== |  | 9%  |
|       |  |     |
| ===== |  | 10% |
|       |  |     |
| ===== |  | 11% |
|       |  |     |
| ===== |  | 12% |
|       |  |     |
| ===== |  | 14% |
|       |  |     |
| ===== |  | 15% |
|       |  |     |
| ===== |  | 16% |
|       |  |     |
| ===== |  | 17% |
|       |  |     |
| ===== |  | 18% |
|       |  |     |
| ===== |  | 19% |
|       |  |     |
| ===== |  | 20% |
|       |  |     |
| ===== |  | 21% |
|       |  |     |
| ===== |  | 22% |
|       |  |     |
| ===== |  | 23% |
|       |  |     |
| ===== |  | 24% |
|       |  |     |
| ===== |  | 25% |
|       |  |     |
| ===== |  | 26% |
|       |  |     |

|       |     |
|-------|-----|
| ===== | 27% |
| ===== | 28% |
| ===== | 29% |
| ===== | 30% |
| ===== | 31% |
| ===== | 32% |
| ===== | 33% |
| ===== | 34% |
| ===== | 35% |
| ===== | 36% |
| ===== | 38% |
| ===== | 39% |
| ===== | 40% |
| ===== | 41% |
| ===== | 42% |
| ===== | 43% |
| ===== | 44% |
| ===== | 45% |
| ===== | 46% |
| ===== | 47% |
| ===== | 48% |
| ===== | 49% |
| ===== | 50% |
| ===== | 51% |
| ===== | 52% |
| ===== | 53% |
| ===== | 54% |

|       |     |
|-------|-----|
| ===== | 55% |
| ===== | 56% |
| ===== | 57% |
| ===== | 58% |
| ===== | 59% |
| ===== | 60% |
| ===== | 61% |
| ===== | 62% |
| ===== | 64% |
| ===== | 65% |
| ===== | 66% |
| ===== | 67% |
| ===== | 68% |
| ===== | 69% |
| ===== | 70% |
| ===== | 71% |
| ===== | 72% |
| ===== | 73% |
| ===== | 74% |
| ===== | 75% |
| ===== | 76% |
| ===== | 77% |
| ===== | 78% |
| ===== | 79% |
| ===== | 80% |
| ===== | 81% |
| ===== | 82% |

|       |      |
|-------|------|
| ===== | 83%  |
| ===== | 84%  |
| ===== | 85%  |
| ===== | 86%  |
| ===== | 88%  |
| ===== | 89%  |
| ===== | 90%  |
| ===== | 91%  |
| ===== | 92%  |
| ===== | 93%  |
| ===== | 94%  |
| ===== | 95%  |
| ===== | 96%  |
| ===== | 97%  |
| ===== | 98%  |
| ===== | 99%  |
| ===== | 100% |

```
minimum.n <- sample.sizes[min(which(pow>=0.8))]  
minimum.n
```

```
## [1] 29
```

```
pow2 <- power.mean(nvec=sample.sizes,mu0=80,muA=78.5,sigma=3.1,alpha=0.01,test="less")
```

```
##  
|  
|  
|= | 0%  
|  
|= | 1%  
|  
|= | 2%  
|  
|= | 3%  
|  
|= | 4%  
|
```

|       |     |
|-------|-----|
| ===   | 5%  |
| ===== | 6%  |
| ===== | 7%  |
| ===== | 8%  |
| ===== | 9%  |
| ===== | 10% |
| ===== | 11% |
| ===== | 12% |
| ===== | 14% |
| ===== | 15% |
| ===== | 16% |
| ===== | 17% |
| ===== | 18% |
| ===== | 19% |
| ===== | 20% |
| ===== | 21% |
| ===== | 22% |
| ===== | 23% |
| ===== | 24% |
| ===== | 25% |
| ===== | 26% |
| ===== | 27% |
| ===== | 28% |
| ===== | 29% |
| ===== | 30% |
| ===== | 31% |
| ===== | 32% |

|       |     |
|-------|-----|
| ===== | 33% |
| ===== | 34% |
| ===== | 35% |
| ===== | 36% |
| ===== | 38% |
| ===== | 39% |
| ===== | 40% |
| ===== | 41% |
| ===== | 42% |
| ===== | 43% |
| ===== | 44% |
| ===== | 45% |
| ===== | 46% |
| ===== | 47% |
| ===== | 48% |
| ===== | 49% |
| ===== | 50% |
| ===== | 51% |
| ===== | 52% |
| ===== | 53% |
| ===== | 54% |
| ===== | 55% |
| ===== | 56% |
| ===== | 57% |
| ===== | 58% |
| ===== | 59% |
| ===== | 60% |

|       |     |
|-------|-----|
| ===== | 61% |
| ===== | 62% |
| ===== | 64% |
| ===== | 65% |
| ===== | 66% |
| ===== | 67% |
| ===== | 68% |
| ===== | 69% |
| ===== | 70% |
| ===== | 71% |
| ===== | 72% |
| ===== | 73% |
| ===== | 74% |
| ===== | 75% |
| ===== | 76% |
| ===== | 77% |
| ===== | 78% |
| ===== | 79% |
| ===== | 80% |
| ===== | 81% |
| ===== | 82% |
| ===== | 83% |
| ===== | 84% |
| ===== | 85% |
| ===== | 86% |
| ===== | 88% |
| ===== | 89% |

|       |      |
|-------|------|
| ===== | 90%  |
| ===== | 91%  |
| ===== | 92%  |
| ===== | 93%  |
| ===== | 94%  |
| ===== | 95%  |
| ===== | 96%  |
| ===== | 97%  |
| ===== | 98%  |
| ===== | 99%  |
| ===== | 100% |

```
minimum.n2 <- sample.sizes[min(which(pow2>=0.8))]  
minimum.n2
```

```
## [1] 46
```

```
plot(sample.sizes,pow,xlab="sample size n",ylab="simulated power")  
points(sample.sizes,pow2,col="grey")  
abline(h=0.8,lty=2)  
abline(v=c(minimum.n,minimum.n2),lty=3,col=c("black","grey"))  
legend("bottomright",legend=c("alpha=0.05","alpha=0.01"),col=c("black","grey"),pch=1)
```



