HW3_SOLUTION

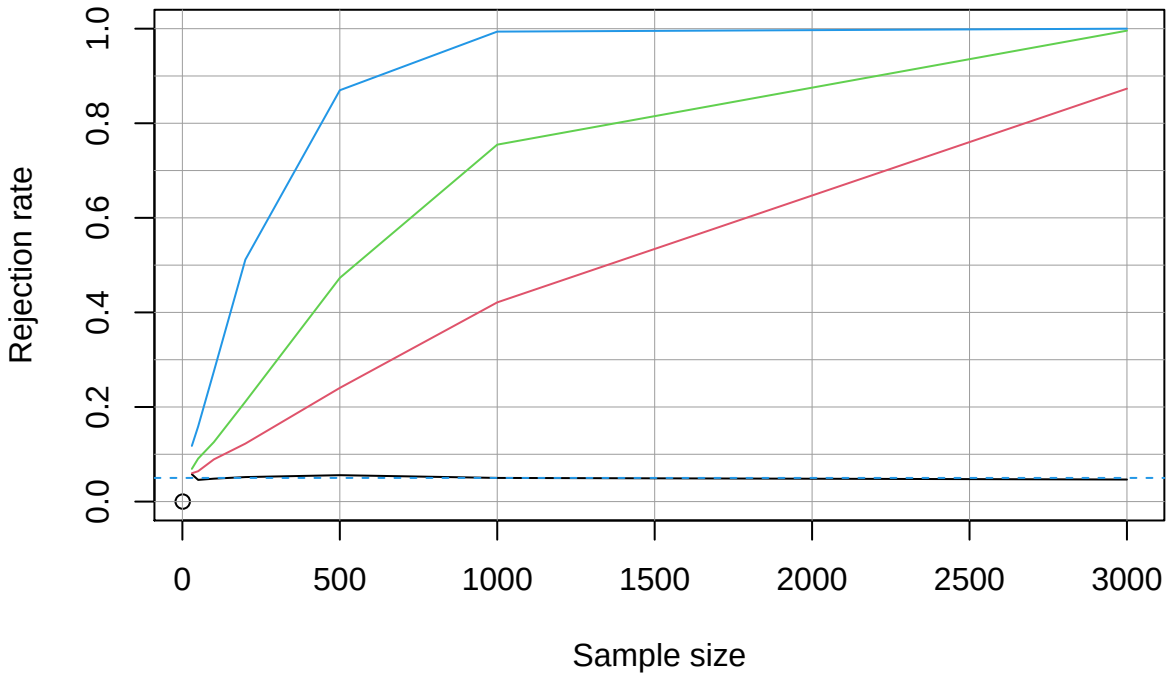
gustavoc@msu.edu

11/30/2020

Question 1

```
N=c(30,50,100,200,500,1000,3000)
EFFECTS=c(0,0.2,0.3,0.5)
nRep=5000
REJ_RATE=matrix(nrow=length(N),ncol=length(EFFECTS),0)
rownames(REJ_RATE)=N
colnames(REJ_RATE)=EFFECTS
for(i in 1:length(N)){
  n=N[i]
  for(j in 1:length(EFFECTS)){
    b2=EFFECTS[j]
    for(k in 1:nRep){
      M=rep(0:1,each=n/2) # male dummy variable

      BMI.M=rnorm(mean=27.4,sd=sqrt(16.7),n=n/2)
      BMI.F=rnorm(mean=26.5,sd=sqrt(30),n=n/2)
      BMI=ifelse(M==1,BMI.M,BMI.F)
      Z=BMI-mean(BMI)
      signal=120-3*M+b2*Z
      error=rnorm(n,sd=sqrt(300))
      SBP=signal+error
      pVal=summary(lm(SBP~M+BMI))$coef[3,4]
      REJ_RATE[i,j]=REJ_RATE[i,j]+(pVal<0.05)/nRep
    }
  }
}
plot(0,ylim=c(0,1),xlim=range(N),ylab='Rejection rate',xlab='Sample size')
abline(h=seq(from=0,to=1,by=.1),col=8,lwd=.5)
abline(v=seq(from=0,to=max(N),by=500),lwd=.5,col=8)
for(i in 1:ncol(REJ_RATE)){
  lines(x=N,y=REJ_RATE[,i],col=i)
}
abline(h=.05,col=4,lty=2)
```



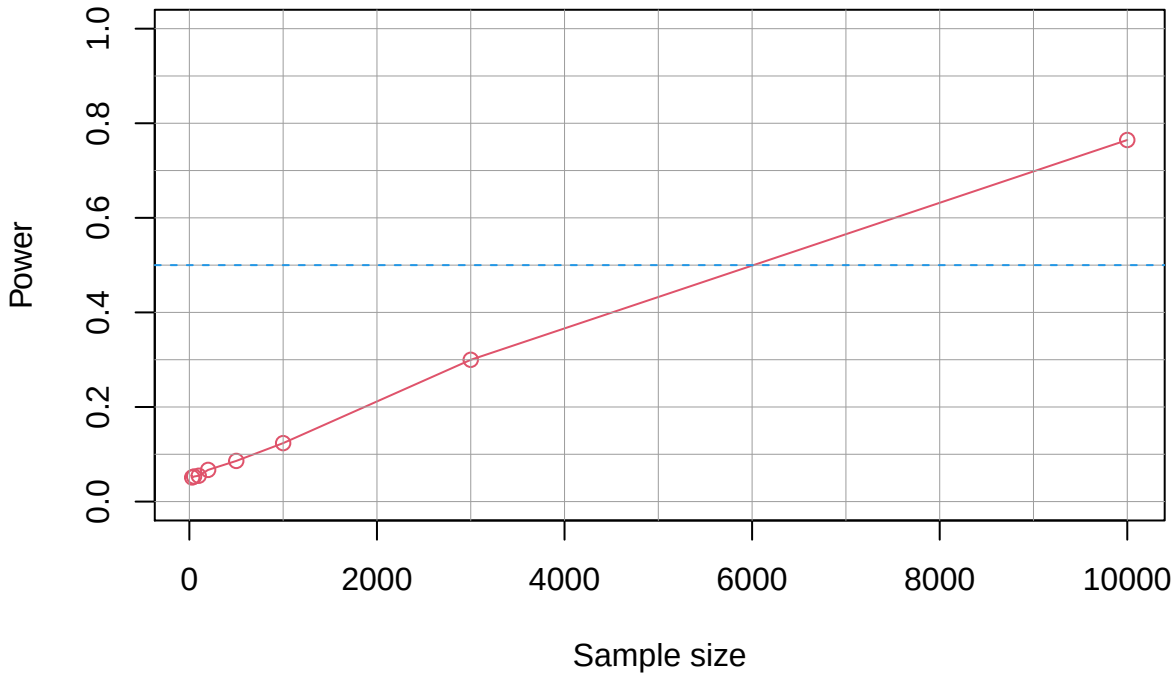
Answer: We need a sample size of at least 1,500 to reach a power of 0.8 with an effect size. (for nicer plots using `ggplot` see code in the Appendix.)

Question 2

```
N=c(30,50,100,200,500,1000,3000,10000)
nRep=5000
REJ_RATE2=rep(0,length(N))
bM=0.4
bF=0.2
for(i in 1:length(N)){
  n=N[i]
  for(k in 1:nRep){
    M=rep(0:1,each=n/2) # male dummy variable

    BMI.M=rnorm(mean=27.4,sd=sqrt(16.7),n=n/2)
    BMI.F=rnorm(mean=26.5,sd=sqrt(30),n=n/2)
    BMI=ifelse(M==1,BMI.M,BMI.F)
    Z=BMI-mean(BMI)

    signal=120-3*M+bM*M*Z+bF*(1-M)*Z
    error=rnorm(n,sd=sqrt(300))
    SBP=signal+error
    pVal=summary(lm(SBP~M+BMI+BMI*M))$coef[4,4]
    REJ_RATE2[i]=REJ_RATE2[i]+(pVal<0.05)/nRep
  }
}
plot(REJ_RATE2~N,ylab='Power', xlab='Sample size',ylim=c(0,1),type='o',col=2)
abline(h=seq(from=0,to=1,by=.1),col=8,lwd=.5)
abline(v=seq(from=0,to=max(N),by=1000),lwd=.5,col=8)
abline(h=.5,col=4,lty=2)
```



Answer: Yes, the power analysis suggests that a power of 50% can be achieved with N~6,000.

Appendix: Plots using ggplot

The ggplot R-package can be used to produce nice-looking, rather complex, plots. The following code illustrates how to produce a power plot like the one in Question 1 using `ggplot`.

To plot your data, you first need to put the results in a stacked format.

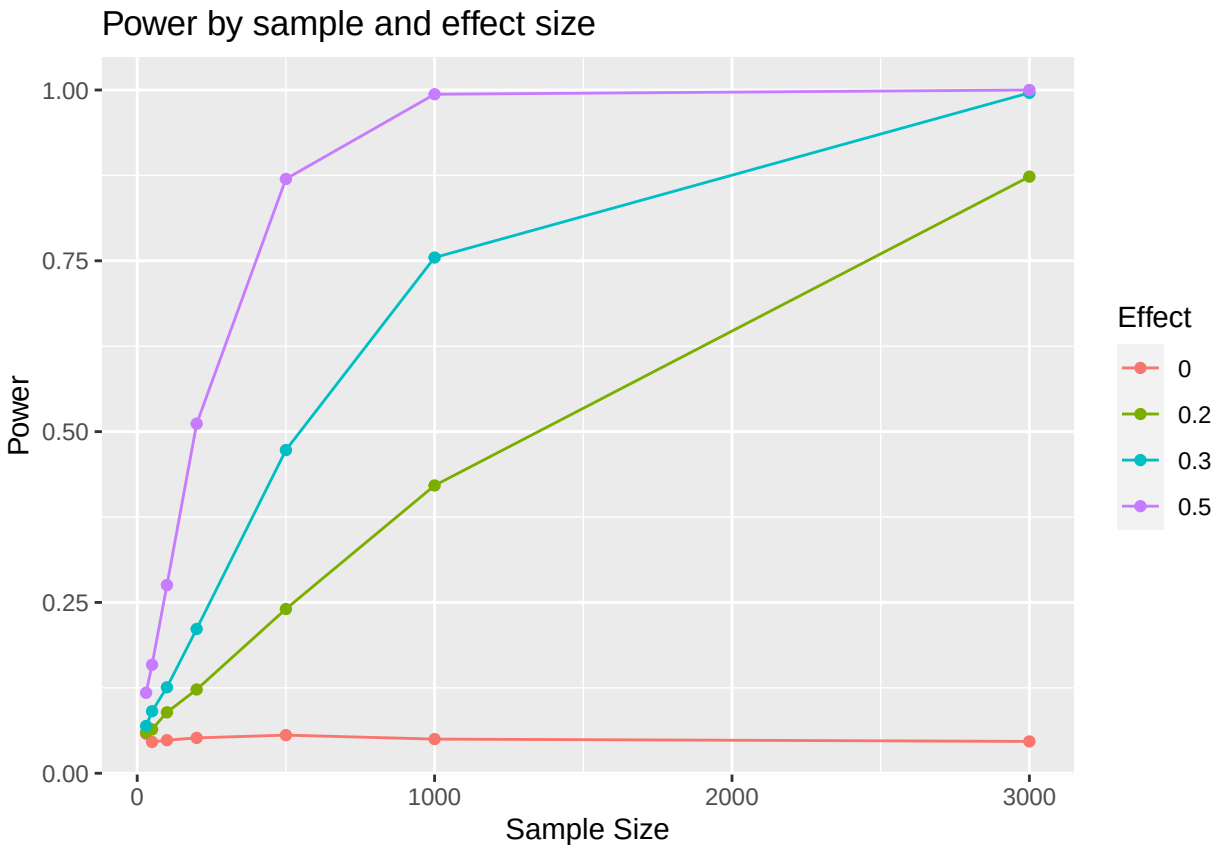
Preparing the data

```
library(ggplot2)
DATA=data.frame(N=rep(rownames(REJ_RATE),ncol(REJ_RATE)),Effect=rep(colnames(REJ_RATE),each=nrow(REJ_RA
head(DATA)
```

```
##      N Effect  Power
## 1   30      0 0.0578
## 2   50      0 0.0458
## 3  100      0 0.0484
## 4  200      0 0.0518
## 5  500      0 0.0558
## 6 1000      0 0.0500
```

```
DATA$N=as.integer(DATA$N)

p=ggplot(DATA,aes(x=N,y=Power)) + # creates a plot
geom_line(aes(colour=Effect)) + # use + to add/modify features of a plot, in this case defining lines
geom_point(aes(colour=Effect)) + # use + to add/modify features of a plot, in this case adding points
labs( x="Sample Size",y="Power",title='Power by sample and effect size') # adding labels
plot(p)
```


Power by sample and effect size
1.00
0.75
0.50
0.25
0.00
Power
0
1000
2000
3000
Sample Size
Effect
0
0.2
0.3
0.5