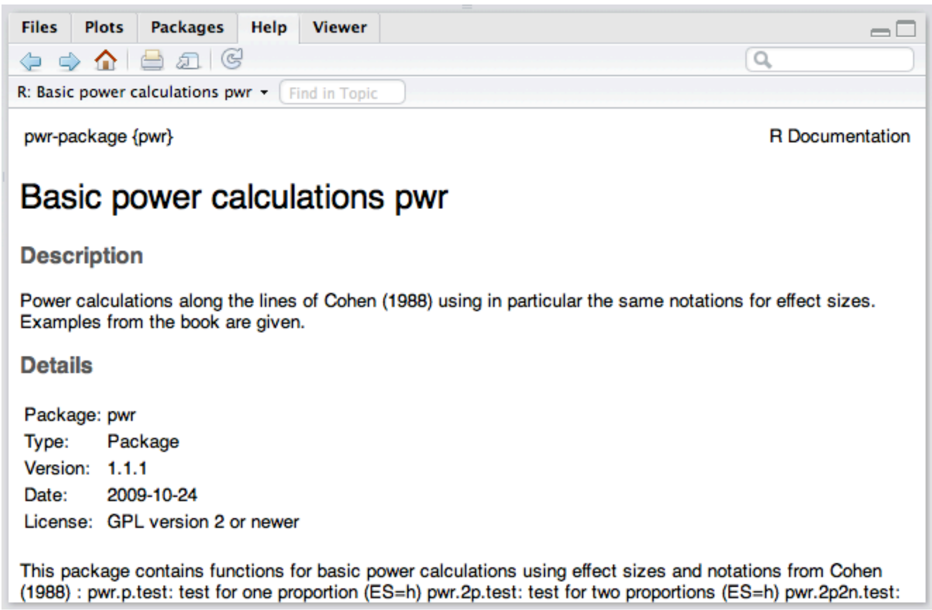
ADA Homework 4 Wenxin Liang wl2455

Question 1. Install package pwr library(pwr); help(pwr)

```
> library(pwr)
> help(pwr)
```



Files Plots Packages Help Viewer

R: Basic power calculations pwr

pwr-package {pwr} R Documentation

Basic power calculations pwr

Description

Power calculations along the lines of Cohen (1988) using in particular the same notations for effect sizes. Examples from the book are given.

Details

Package:	pwr
Type:	Package
Version:	1.1.1
Date:	2009-10-24
License:	GPL version 2 or newer

This package contains functions for basic power calculations using effect sizes and notations from Cohen (1988) : pwr.p.test: test for one proportion (ES=h) pwr.2p.test: test for two proportions (ES=h) pwr.2p2n.test:

Question 2.

Refer to Example 2. Suppose an investigator wishes to determine the sample size required to compare two anti-hypertension drugs under the following assumptions. It is known that the clinically meaningful detectable difference with respect to diastolic blood pressure is at least 4 mmHg; and from historical data the common standard deviation $\sigma = 10$ mmHg. Further, it is desired to use a two-sided test of level $\alpha = 0.05$ and power of 80%.

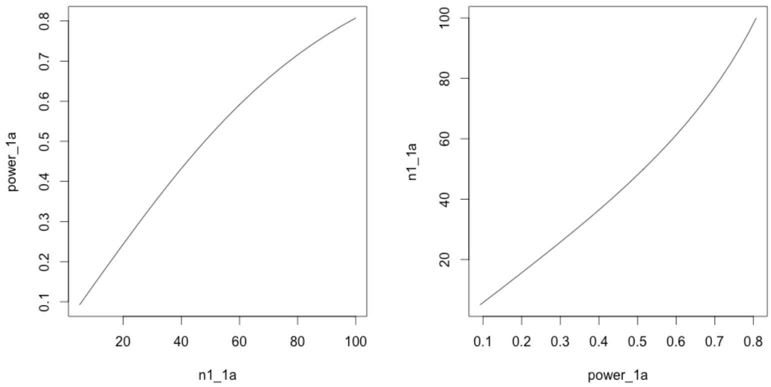
Write a function in R and use it.

a) To plot POWER as a function of SAMPLE SIZE (assuming equal n for the two groups) You may use POWER from 0% to 100%, in increments of 10%. Similarly, sample sizes in increments of 5 to 10.

First we need a function between POWER, which is 1-β and SAMPLE SIZE n.

Based on the formula we have, $n = \frac{2(Z_{\alpha/2}+Z_\beta)^2\sigma^2}{\Delta^2}$, we have our function can be

$$POWER = 1\text{-}\beta = \Phi\left(\frac{\sqrt{n}|\Delta|}{|\sigma|\sqrt{2}} - Z_{\alpha/2}\right)$$

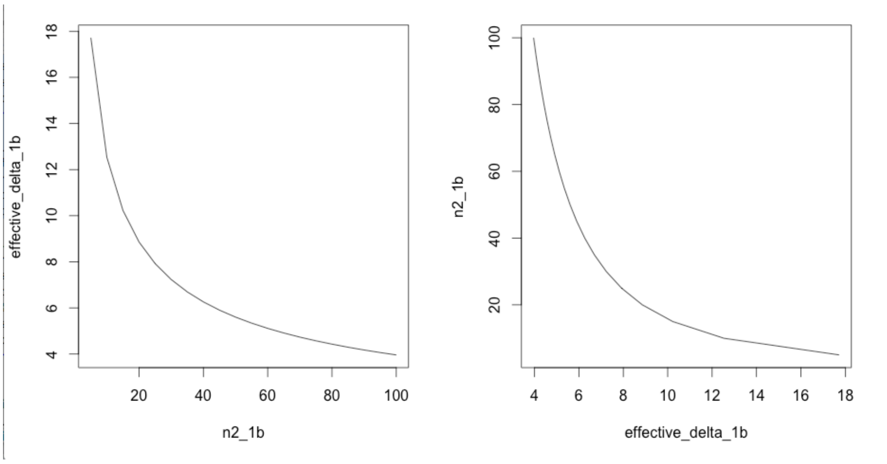


Based on the graphs for sample size from 5 to 100 we can conclude that when the sample size increase the power will increase and when the power increase the sample size required will increase. Also we notice that, as n equals to nearly 100 the increasing of sample size increase slowly. Therefore, there is a trade off since the larger the sample size the higher the cost will be.

b) To plot EFFECT SIZE as a function of SAMPLE SIZE (assuming equal n for the two groups)

First we need a function between EFFECT SIZE, which is μ_1-$\mu_2 = \Delta$ and SAMPLE SIZE n. Based on the formula we have, $n = \frac{2(Z\alpha/_2+Z_\beta)^2\sigma^2}{\Delta^2}$, we have our function can be

$$EFFECTIVE\ SIZE = \Delta = \frac{\sqrt{2}\left|\left(Z\alpha/_2 + Z_\beta\right)\sigma\right|}{\sqrt{n}}$$



Based on the graphs we know that for sample size from 5 to 100 when the sample size increase the effective size will decrease which means when there is a big difference when comparing the two means then there are less sample size we needed. Otherwise when there is a small difference when comparing the two means we need more sample size, the larger n, to make sure the difference can be tested out.

Question 3.

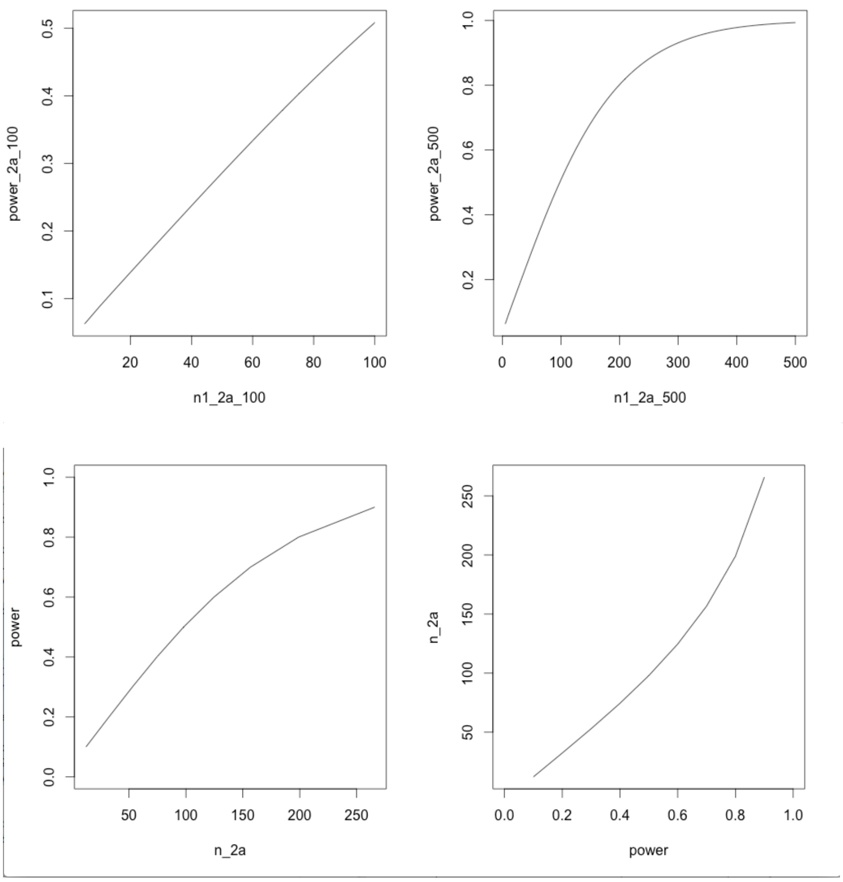
Refer to Example 3, and repeat Problem 2 using both normal approximation and arc sin transformation

a. Based on normal approximation, to plot POWER as a function of SAMPLE SIZE

First we need a function between POWER, which is $1\text{-}\beta$ and SAMPLE SIZE n.

Based on the formula we have, $n = \frac{\left[Z_{\alpha/2}\sqrt{2\bar{p}\bar{q}}+Z_{\beta}\sqrt{p_1q_1+p_2q_2}\right]^2}{(p_2\text{-}p_1)^2}$ $where\ \bar{p} = \frac{p_1+q_1}{2}, \bar{q} = 1\text{-}\bar{p}$, we have our function can be

$$POWER = 1\text{-}\beta = \Phi\left(\frac{\sqrt{n}|p_2\text{-}p_1|\text{-}Z_{\alpha/2}\sqrt{2\bar{p}\bar{q}}}{\sqrt{p_1q_1+p_2q_2}}\right) where\ \bar{p} = \frac{p_1+q_1}{2}, \bar{q} = 1\text{-}\bar{p}$$



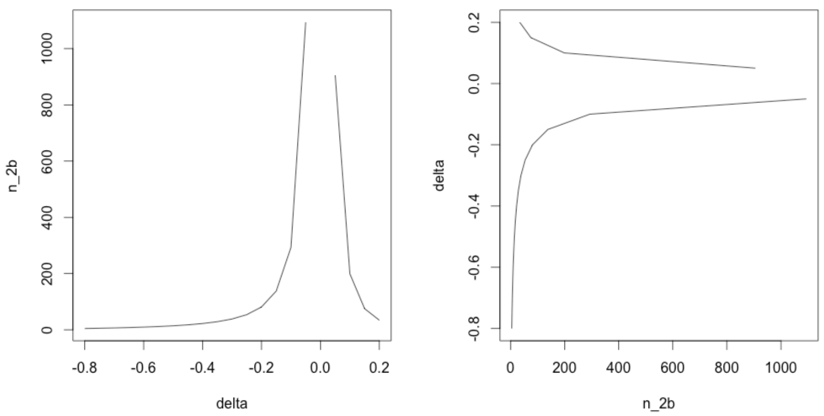
Based on the graphs, for the sample size from 5 to 100, for the sample size from 5 to 500 and for power from 0 to 1 we can conclude that when the sample size increase the power will increase and when the power increases the sample size required will increase.

b. Based on normal approximation, to plot EFFECT SIZE as a function of

SAMPLE SIZE

First we need a function between EFFECT SIZE, which is μ_1-$\mu_2 = \Delta$ and SAMPLE SIZE n. Based on the formula we have, $n = \frac{\left[Z\alpha_{/2}\sqrt{2\bar{p}\bar{q}}+Z_\beta\sqrt{p_1q_1+p_2q_2}\right]^2}{(p_2\text{-}p_1)^2}$ $where\ \bar{p} = \frac{p_1+q_1}{2}$, $\bar{q} = 1\text{-}\bar{p}$, we have our function can be

$$n = \frac{\left[Z\alpha_{/2}\sqrt{2\bar{p}\bar{q}} + Z_\beta\sqrt{p_1q_1 + p_2q_2}\right]^2}{(p_2\text{-}p_1)^2} where\ \bar{p} = \frac{p_1 + q_1}{2}, \bar{q} = 1\text{-}\bar{p}, p_2 = p_1 + \Delta$$



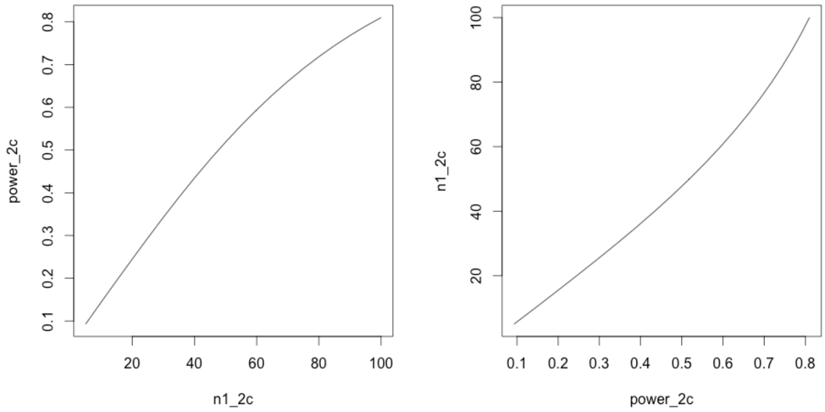
Based on the graphs we know that since $p_2 = p_1 + \Delta$ and the denominator of the formula $n = \frac{\left[Z\alpha_{/2}\sqrt{2\bar{p}\bar{q}}+Z_\beta\sqrt{p_1q_1+p_2q_2}\right]^2}{(p_2\text{-}p_1)^2}$ which is $(p_2\text{-}p_1)^2$ is the square of the delta then delta <0 as delta increase, the denominator will decrease then the sample size needed will increase. When delta equals to zero, the denominator will be zero then the sample size needed is infinity then there is a gap between delta<0 and delta>0. When delta >0 the denominator will increase then the sample size needed will decrease.

c. Based on arc sin transformation, to plot POWER as a function of SAMPLE SIZE

First we need a function between POWER, which is 1-β and SAMPLE SIZE n. Based on the formula we have, $n = \frac{(Z\alpha_{/2}+Z_\beta)^2}{\Delta^2}$, we have our function can be

$$POWER = 1\text{-}\beta = \Phi\left(\sqrt{n}|\Delta|\text{-}Z\alpha_{/2}\right)\ where\ \Delta = f(p_1)\text{-}f(p_2)$$

$$= 2arcsin\sqrt{p_1}\text{-}2arcsin\sqrt{p_2}$$

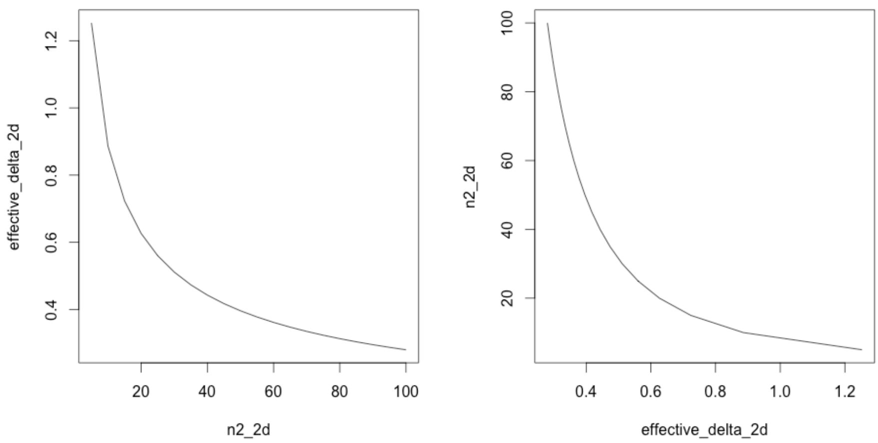


Based on the graphs we know that the when the sample size increase the power will increase and when the power increase the sample size required will increase.

d. Based on arc sin transformation, to plot EFFECT SIZE as a function of SAMPLE SIZE

First we need a function between EFFECT SIZE, which is μ_1-$\mu_2 = \Delta$ and SAMPLE SIZE n. Based on the formula we have, $n = \frac{(Z\alpha_{/2}+Z_\beta)^2}{\Delta^2}$, we have our function can be

$$EFFECTIVE\ SIZE = \Delta = \frac{\left|Z\alpha_{/2} + Z_\beta\right|}{\sqrt{n}}$$



Based on the graphs we know that the when the sample size increase the effective size will decrease which means when there is a big difference when comparing the two means then there are less sample size we needed vice versa.

The R code attached

```
#Question 1 part a
par.n_1a <- function (n,alpha,delta,sigma){
```

```
    #Calculate Z_alpha/2
    z1_1a <-qnorm(1-alpha/2)
    #Calculate the power
    z2_1a <-sqrt(n)*abs(delta)/sqrt(2)/abs(sigma)-z1_1a
    ans_1a <- pnorm(z2_1a)
    return(ans_1a)
}

#Since sample sizes in increments of 5 to 10
n1_1a <- seq(5,100,by=5)
power_1a <- par.n_1a(n1_1a,0.05,4,10)

par(mfrow=c(1,2))
plot(n1_1a,power_1a,type="l")
plot(power_1a,n1_1a,type="l")

#Question 1 part b
par.n_1b <- function(n,alpha,beta,sigma){
    z1_1b <-qnorm(1-alpha/2)
    z2_1b <- qnorm(1-beta)
    ans_1b <- sqrt(2)*abs((z1_1b+z2_1b)*sigma)/sqrt(n)
    return(ans_1b)
}
n2_1b <- seq(5,100,by=5)
effective_delta_1b <- par.n_1b(n2_1b,0.05,0.2,10)
par(mfrow=c(1,2))
plot(n2_1b,effective_delta_1b,type="l")
plot(effective_delta_1b,n2_1b,type="l")

#Question 2 part a

par.n_2a_1 <- function (power,alpha,p1,p2){
    z1_2a <-qnorm(1-alpha/2)
    p_bar <- (p1+p2)/2
    #Calculate the power

n_1=(1/(p2-p1)^2)*(qnorm(1-alpha/2)*sqrt(2*p_bar*(1-p_bar))+qnorm(power
)*sqrt(p1*(1-p1)+p2*(1-p2)))^2
    return(n_1)
}
power=seq(0,1,by=0.1)
n_2a <- par.n_2a_1(power,0.05,0.8,0.9)
par(mfrow=c(1,2))
plot(n_2a,power,type="l")
```

```
plot(power,n_2a,type="l")


#Question 2 part b
par.n_2b <- function(delta,alpha,beta,p1){
   z1_2b    <- qnorm(1-alpha/2)
   z2_2b    <- qnorm(1-beta)
   p2         <- p1+delta
   p_bar    <- (p1+p2)/2
   ans_n                                                                                    <-
(z1_2b*sqrt(2*p_bar*(1-p_bar))+z2_2b*sqrt(p1*(1-p1)+p2*(1-p2)))^2/((p2-p1)
^2)
   return(ans_n)
}
delta <- seq(-0.8,0.2,by=0.05)
n_2b <- par.n_2b(delta,0.05,0.2,0.8)
par(mfrow=c(1,2))
plot(delta,n_2b,type="l")
plot(n_2b,delta,type="l")

#Question 2 part c
par.n_2c <- function (n,alpha,p1,p2){
   z1_2c <-qnorm(1-alpha/2)
   #Calculate the power
   delta=abs(2*asin(sqrt(p1))-2*asin(sqrt(p2)))
   z2_2c <-sqrt(n)*abs(delta)-z1_2c
   ans_2c <- pnorm(z2_2c)
   return(ans_2c)
}

#Since sample sizes in increments of 5 to 10
n1_2c <- seq(5,100,by=5)
power_2c <- par.n_2c(n1_2c,0.05,0.8,0.9)

par(mfrow=c(1,2))
plot(n1_2c,power_2c,type="l")
plot(power_2c,n1_2c,type="l")

#Question 2 part d
par.n_2d <- function(n,alpha,beta,p1,p2){
   z1_2d <-qnorm(1-alpha/2)
   z2_2d <- qnorm(1-beta)
   ans_2d <- abs(z1_2d+z2_2d)/abs(sqrt(n))
   return(ans_2d)
```

```
}
n2_2d <- seq(5,100,by=5)
effective_delta_2d <- par.n_2d(n2_2d,0.05,0.2,0.8,0.9)
par(mfrow=c(1,2))
plot(n2_2d,effective_delta_2d,type="l")
plot(effective_delta_2d,n2_2d,type="l")
```