Homework 6

Jackson Dial

10/26/2021

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.4 v dplyr 1.0.7  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 2.0.1 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

## Question 1

### Part A

How much power would there be to detect the clinically relevant difference at the 0.05 significance level, if 20 individuals who spike in PF4 and 20 individuals who do not spike in PF4 are enrolled? Show some methodology for full credit.

MyPwrFun <- function(alpha, m1, m2, s1, s2, n1, n2){  
 z1a2 <- qnorm(1 - alpha / 2)  
 za2 <- qnorm(alpha / 2)  
 quant <- (m1-m2) / sqrt( (s1^2/n1) + (s2^2/n2) )  
 beta <- pnorm(z1a2 - quant) - pnorm(za2 - quant)  
 power <- 1- beta  
 return(power)  
}  
  
power1 <- MyPwrFun(alpha = .05, m1 = 1, m2 = .8, s1 = .32, s2 = .04, n1 = 20, n= 20)

The power for this scenario would be 0.79.

### Part B

Power is the probability of rejecting the null hypothesis when the alternative hypothesis is true. It is affected by variability, and as variability increases, power decreases.If we enroll more individuals in the population with more variability, that normalizes the variance of that group by a larger value, reducing variability, given that we have assumed normally distributed source populations.

### Part C

pwrdf <- data.frame()  
for (x in 0:40){  
 powers <- MyPwrFun(alpha = .05, m1 = 1, m2 = .8, s1= .32, s2 = .04, n1 = x, n2 = 40-x)  
 pwrdf <- rbind(pwrdf, c(x, powers))  
}

A sample split of 36-4 observations per group would yield us the highest power, with power being calculated as 0.9396 for that split.

## Question 2

Investigators hope to determine if a large dose of vitamin E will prevent cancer. The investigators will first conduct a feasibility study to determine the dosage of vitamin E. The investigators believe that 80% of participants on a high dose regimen will achieve adequate serum levels while 64% of those on a medium dose regimen will achieve adequate intake.

### Part A

How large of a sample is needed to have 90% power of a two-sided test of proportions at a 0.05 significance level of .05? Assume equal allocation to the two arms.

MyPwrFun2 <- function(p1, p2, alpha, p.bar, n1, n2){  
 z1a2 <- qnorm(alpha)  
 d <- abs(p1-p2)  
 quant <- (d - (z1a2 \* sqrt(p.bar \* (1 - p.bar) ( (1/n1) + (1/n2) ) ) ) ) / (sqrt(((p1 \* (1-p1) ) / n1) + (p2 \* (1-p2) / n2)))  
 pnorm(quant)  
}  
  
SampSizePropsFun <- function(alpha, k, beta, p1, p2){  
 z1a2 <- qnorm(1-alpha/2)  
 z1b <- qnorm(1-beta)  
 pstar <- (p1 + k \* p2) / (1 + k)  
 firstquantnum <- z1a2 \* sqrt( pstar \* (1- pstar) \* (1 + (1/k)))  
 secondquantnum <- z1b \* sqrt( p1 \* (1 - p1) + ((p2 \* (1-p2))/k))  
 numerator <- (firstquantnum + secondquantnum)^2  
 denom <- (p1 - p2)^2  
 n <- numerator / denom  
 n  
}  
#k = n2/n1 and is only used for unequal allocation of observations  
  
ans2a <- ceiling(SampSizePropsFun(alpha = .05, beta = .1, k = 1, p1 = .8, p2 = .64))  
ans2a

## [1] 164

A sample size of 328 is needed to have a power of 90% using equal allocation.

### Part B

Create a smooth plot of power as a function of sample size per group assuming equal allocation. Use a dense grid of sample sizes, i.e., nn <- 1:200. How much power does a sample size of 180 per group yield?