GS example 3.11

Zhaosheng-Xie

2021/2/10

Introduction

In the final paragraph of the example on page 102, the authors write, "A few experiments have shown us that ... ". They report the results but don't show you their experiments. Similarly, they show you Figure 3.7, but not how it was produced. MY assignment is to fill in these gaps. Produce an explanation of the example showing how the authors reached their conclusion that the critical value should be between 69 and 73 people cured. Replicate and explain Figure 3.7.

Explanation of the example

First we have type 1 error, which is to reject the null hypothesis when in fact it is true. Also, we have type 2 error, which is to accept the null hypothesis when it is false. $\alpha(p) = \sum_{m \leq k \leq n} b(n, p, k)$ represents the probability of type 1. $\beta(p) = 1 - \alpha(p)$ represents the probability of type 2.

In this case, we want to choose a m to make $\alpha(p)$ smaller. As m increases above the most probable value np = .6n, $\alpha(0.6)$, being the upper tail of a binomial distribution, approaches 0. In this way increasing m makes type 1 error less likely. We also want a small type 2 error which is $1 - \alpha(p)$. But increasing m makes type 2 error more likely.

In conclusion, we want to find a m which makes both these probabilities small, preferably less than 0.05.

```
# Probability of type 1 error when p=0.6
n <- 100
m <- rep(0,41)
p.k <- rep(0,41)
a_6 <- rep(0,41)
for (i in 1:41) {
  m[i]=i+59
  p.k[i] <- dbinom(x = i+59,size = n,prob = 0.6)
}
for (j in 1:41) {
  a_6[j] <- sum(p.k[j:41])
}
a <- data.frame(m,p.k,a_6)
view(a)
```

I set m from 60 to 100. p.k is b(n,p,k) and a_6 is $\alpha(0.6)$

```
# Probability of type 2 error when p=0.8
n <- 100
m <- rep(0,41)
p.k <- rep(0,41)
b_8 <- rep(0,41)
for (i in 1:41) {
  m[i]=i+59
  p.k[i] <- dbinom(x = i+59,size = n,prob = 0.8)
```

```
}
for (j in 1:41) {
  b_8[j] <- 1-sum(p.k[j:41])
}
b <- data.frame(m,p.k,b_8)
view(b)
```

I set m from 60 to 100. p.k is b(n,p,k) and b_8 is $\beta(0.8)$

```
# Range of m
library(sqldf)
```

```
## Warning: package 'sqldf' was built under R version 4.0.3
```

```
## Loading required package: gsubfn
```

```
## Warning: package 'gsubfn' was built under R version 4.0.3
```

```
## Loading required package: proto
```

```
## Warning: package 'proto' was built under R version 4.0.3
```

```
## Loading required package: RSQLite
```

```
## Warning: package 'RSQLite' was built under R version 4.0.3
```

```
lower <- min(sqldf('SELECT m FROM a WHERE a_6 <= 0.05'))
upper <- max(sqldf('SELECT m FROM b WHERE b_8 <= 0.05'))
cat("In this case, the range of m is from",lower,"to",upper,".")
```

```
## In this case, the range of m is from 69 to 73 .
```

Replicate and explain Figure 3.7

```
n = 100
m = c(69,73)
p = seq(0.4, 1, 1/n)
# Prepare data when m is 69
P69 <- data.frame(p, curve='P_69',Power=cumsum(dbinom(69,n,p)))
# Prepare data when m is 73
P73 <- data.frame(p, curve='P_73',Power=cumsum(dbinom(73,n,p)))
# Rbind 2 data set
P <- rbind(P69,P73)
```

```
ggplot(P)+
  geom_rect(aes(xmin = 0.6, xmax = 0.8, ymin = 0.05, ymax = 0.95))+
  geom_line(aes(p, Power, color = curve))+
  labs(x="p",y="Power",caption= "Figure 3.7: The power curve.")+
  theme(plot.caption = element_text(hjust=0.5,size=15))+
  theme(panel.grid=element_blank(),panel.border=element_blank(),axis.line=element_line(colour="black"))
```
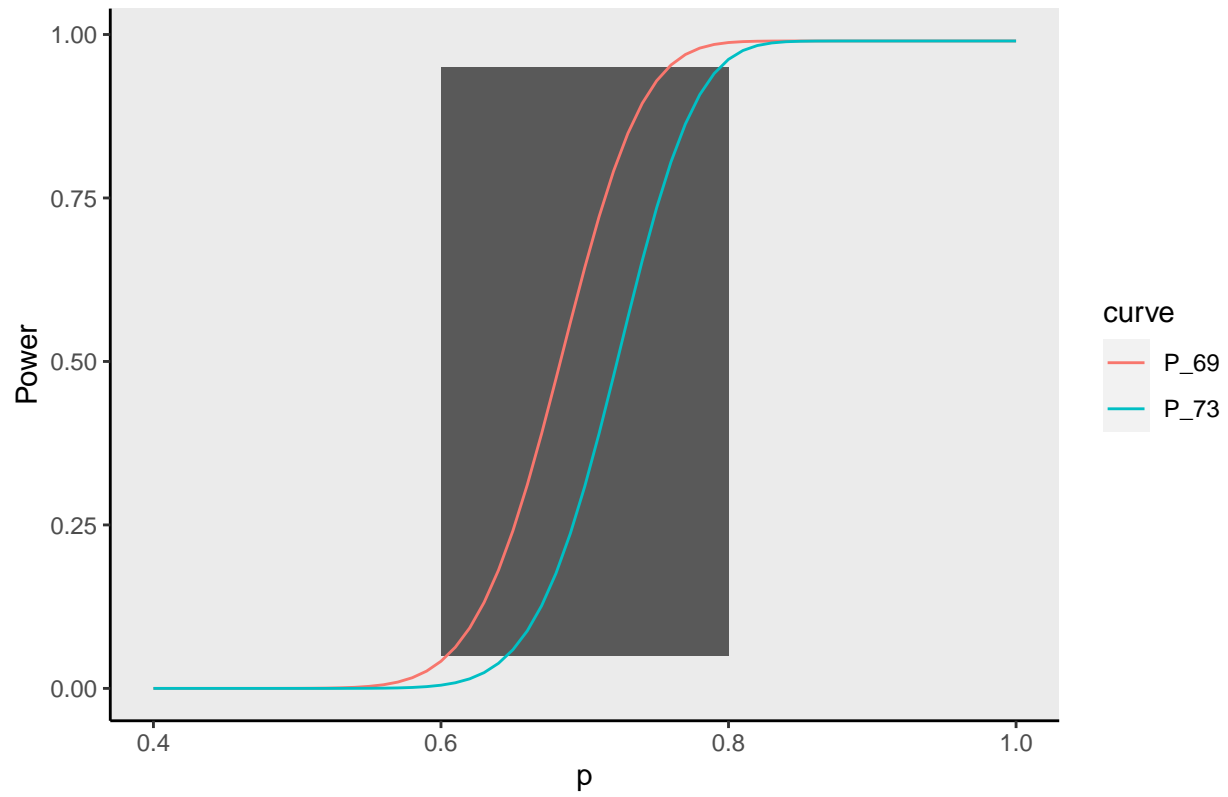
Figure 3.7: The power curve.