677 HW1

Zixuan Liu

2/10/2021

311 Assignment: Due – Wednesday, February 10 In Chapter 2 of G&S, Example 3.11 uses the binomial distribution to demonstrate the logic of hypothesis testing with a simple example. The intuitive simplicity of the example and the availability of tools in R for calculations and plotting distributions encourages you to explore hypothesis testing by reproducing the results presented in the book. 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. Your 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.

Find critical value

Type I and Type II error

Type I error is the error of rejecting the null hypothesis, which is when it is actually true. suppose the sample size is 100, then the proportion $\hat{p}$ such that Z statistics should be to 1.64

$$z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$$

plug in the number here

$$1.64 = \frac{\hat{p} - 0.6}{\sqrt{\frac{0.6(1-0.6)}{100}}}$$

So we get

```
1.64*sqrt(0.6*0.4)/10+0.6
```

```
## [1] 0.6803433
```

Type II error is the error of not rejecting a null where the alternative hypothesis is the true state of nature. For this question, it should be

$$-1.64 = \frac{\hat{p} - 0.8}{\sqrt{\frac{0.8(1-0.8)}{100}}}$$

So we get

```
-1.64*sqrt(0.8*0.2)/10+0.8
```

```
## [1] 0.7344
```

The minimum value of m should be 69 and largest value of m should be 73.

Reproducing the graph in Figure 3.7

The type-I and type-II error need to be smaller than 0.05 in order to be significant. When m = 69, the probability which is p should be 0.6 to reject null hypothesis. When m = 73, the probability which is p should be 0.8 to reject null hypothesis.

```r
#install.packages("tidyverse")
library(tidyverse)
```

```
## Warning in as.POSIXlt.POSIXct(Sys.time()): unknown timezone 'zone/tz/2020d.
## 1.0/zoneinfo/America/Chicago'
```

```
## -- Attaching packages -------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.2.1      v purrr   0.3.3
## v tibble  2.1.3      v dplyr   0.8.4
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
```

```
## Warning: package 'readr' was built under R version 3.4.4
```

```
## Warning: package 'stringr' was built under R version 3.4.4
```

```
## Warning: package 'forcats' was built under R version 3.4.4
```

```
## -- Conflicts ----------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
#plot
curve1 <- function(x){
  sum(dbinom(69:100,100,x))
}
curve2 <- function(x){
  sum(dbinom(73:100,100,x))
}
curve1<-Vectorize(curve1)
curve2<-Vectorize(curve2)
ggplot(data.frame(x=c(0.4,1)), aes(x=x)) + xlab("P") + ylab("Probability of rejecting null hypothesis")
  stat_function(fun=curve1)+
  stat_function(fun=curve2)+
  geom_rect(aes(xmin = 0.6, xmax = 0.8, ymin = 0.05, ymax = 0.95),
            fill = "transparent", color = "black", size = 0.5)+
  scale_x_continuous(breaks = seq(0.4,1,0.1), minor_breaks = NULL)+
  scale_y_continuous(breaks = seq(0,1,0.1), minor_breaks = NULL)
```