MA677 A1

Ruxin Liu

Assignment Objective:

Reproduce example 3.11 on page 102 of Introduction of Probability.

Based on the question, we know that the ordinary aspirin can cure headaches for 60% of the time, which can be written in P(cure headache) = 0.6. The drug company claims that their new additive aspirin can be more effective, which means that P(cure headache) > 0.6. Therefore, in our hypothesis test the null hypothesis is that: $H_0 : P = 0.6$ and the alternative hypothesis is that: $H_1 : P > 0.6$.

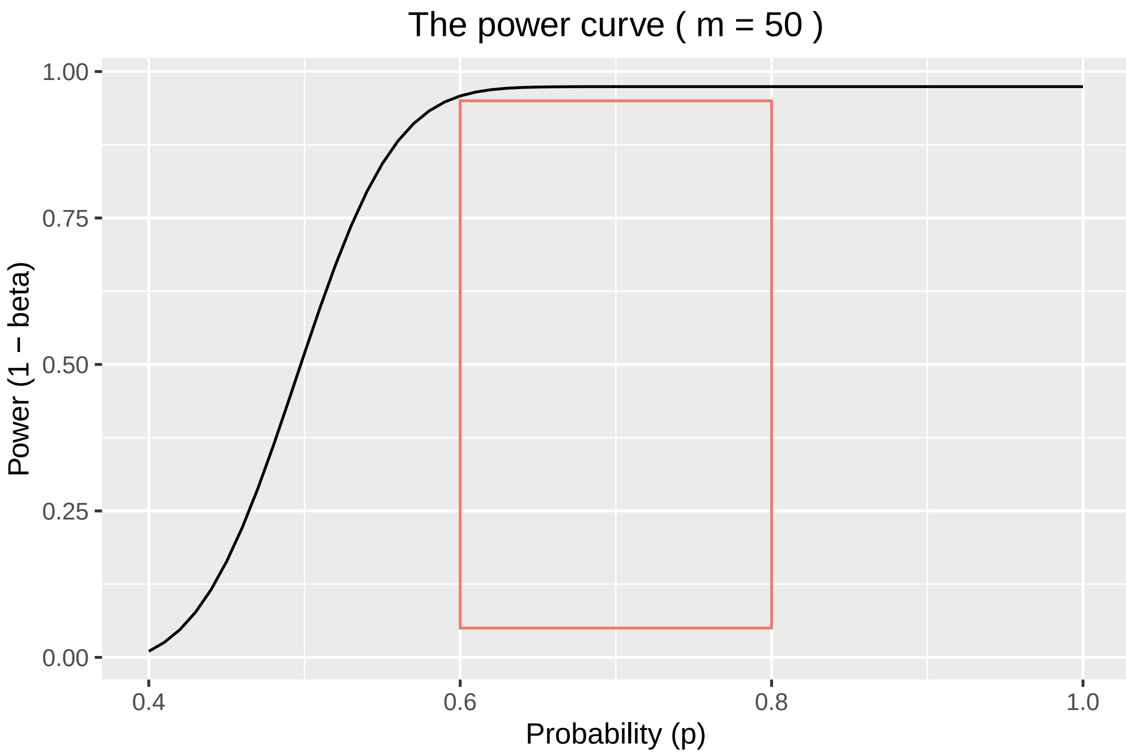
In this situation, we may encounter 2 types of errors: type 1 error and type 2 error. Type 1 error happens when we reject the null hypothesis when the null hypothesis is true and α is the probability of type 1 error. Type 2 error happens when we do not reject the null hypothesis when the alternative hypothesis is true and β is the probability of type 2 error. And we can calculate the power of the test by using the formula: $1 - \beta$. Here the goal is to find the critical value m that makes both probabilities (α and β) small.

Now, we are going to reproduce the program PowerCurve that being mentioned in the book to find the critical value m. The sample size n is set to be 100 and p has a range from 0.4 to 1. Firstly, we can create a function to plot the power curve.

```
library(tidyverse)
power_curve <- function(m){
  p <- seq(from = 0.4, to = 1, by = 0.01)
  # Binomial distribution
  a <- dbinom(m, size = 100, prob = p)
  power <- cumsum(a)
  return(power)
}
```

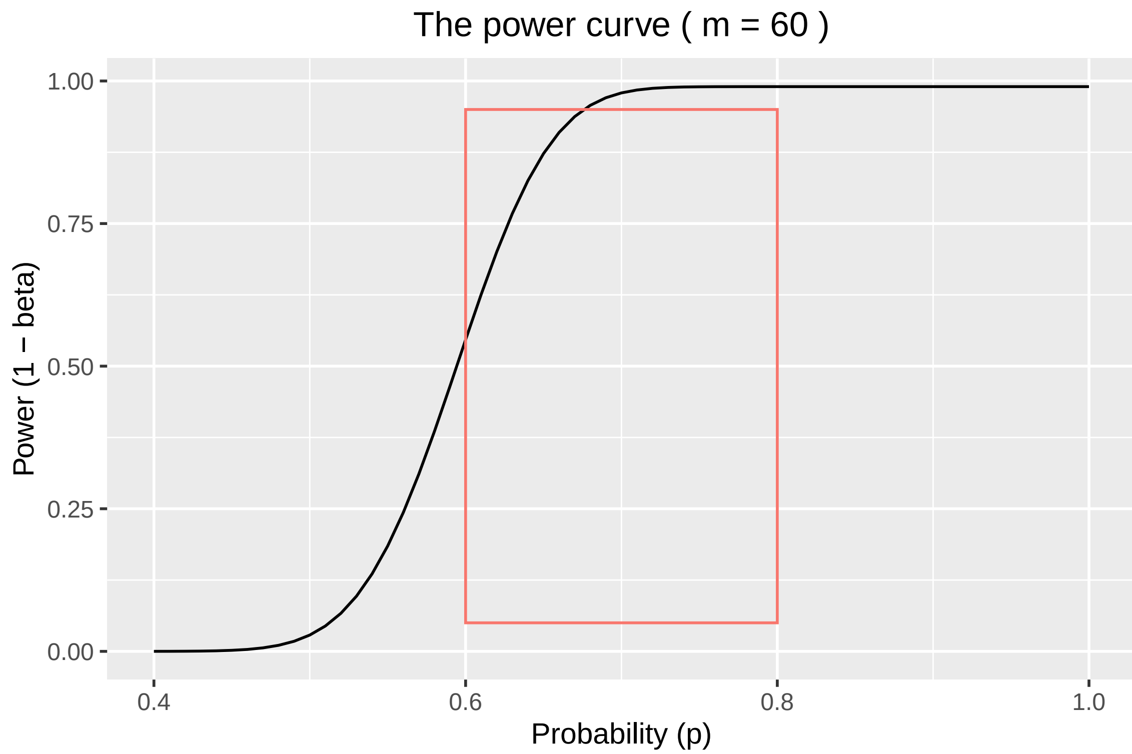
```
box <- function(m){
  x_axis <- seq(from = 0.4, to = 1, by = 0.01)
plot <- ggplot() +
  # Plot the power curve
  geom_line(aes(x_axis, power_curve(m))) +
  # Adding the box from 0.6 to 0.8, with bottom and top at heights 0.05 and 0.95
  geom_rect(aes(xmin = 0.6, xmax = 0.8, ymin = 0.05, ymax = 0.95, col = "red"),
            fill = NA) +
  xlab("Probability (p)") +
  ylab("Power (1 - beta)") +
  ggtitle(paste("The power curve ( m =", m, ")")) +
  theme(legend.position = "none") +
  theme(plot.title = element_text(hjust = 0.5))
return(plot)
}
```

```
box(50)
```

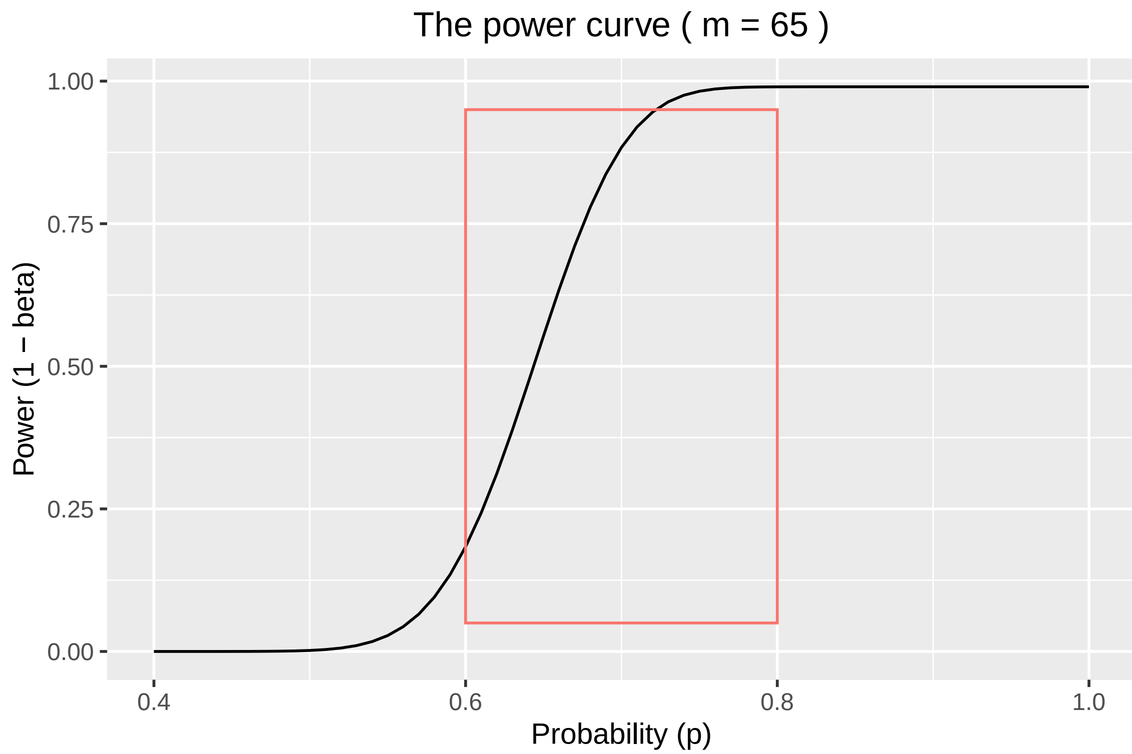


As mentioned in the book, when the critical value m increases, type 1 error is less likely to happen; while when the critical value m decreases, type 2 error is less likely to happen, we would like both error probabilities less than 5%. From the plot, the boundary of type 1 error probability is the upper line of the box (power = 0.95) and the boundary of type 2 error probability is the bottom line (power = 0.05) of the box. Therefore, we are looking for the value for m that produces the power curve entering the box from the bottom and leaving from the top. Now, we will try different values of m.

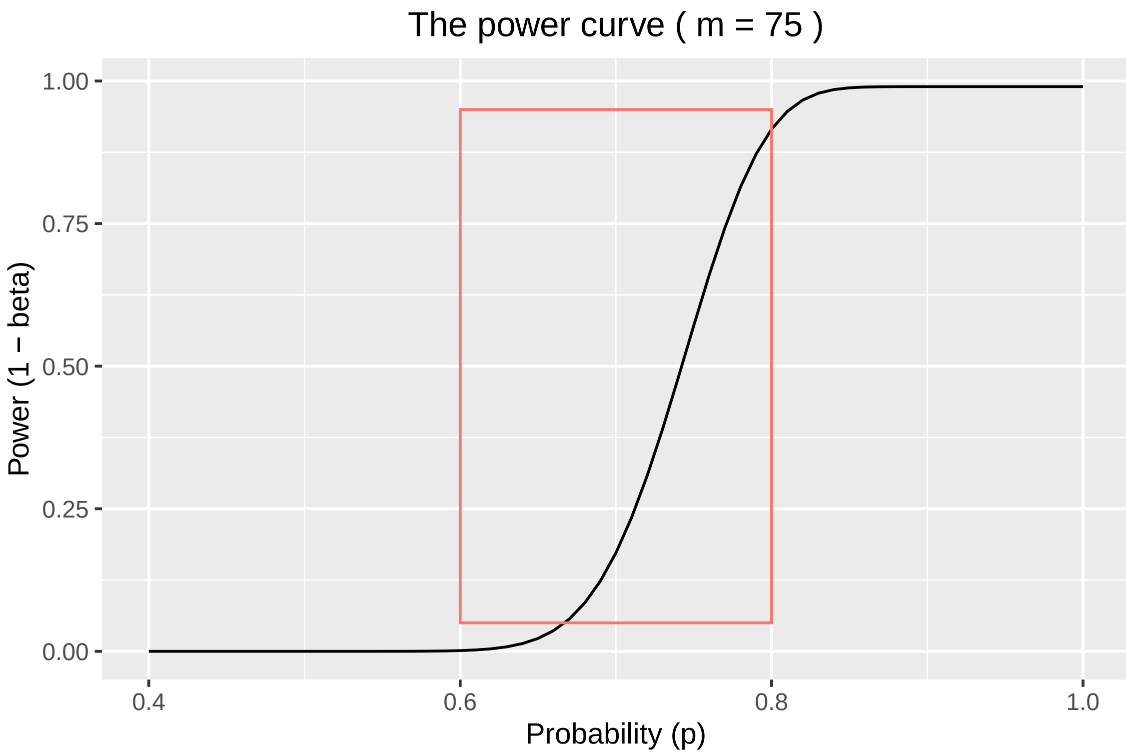
```
box(60)
```


The power curve (m = 60)
1.00
0.75
0.50
0.25
0.00
Power (1 − beta)
0.4
0.6
0.8
1.0
Probability (p)

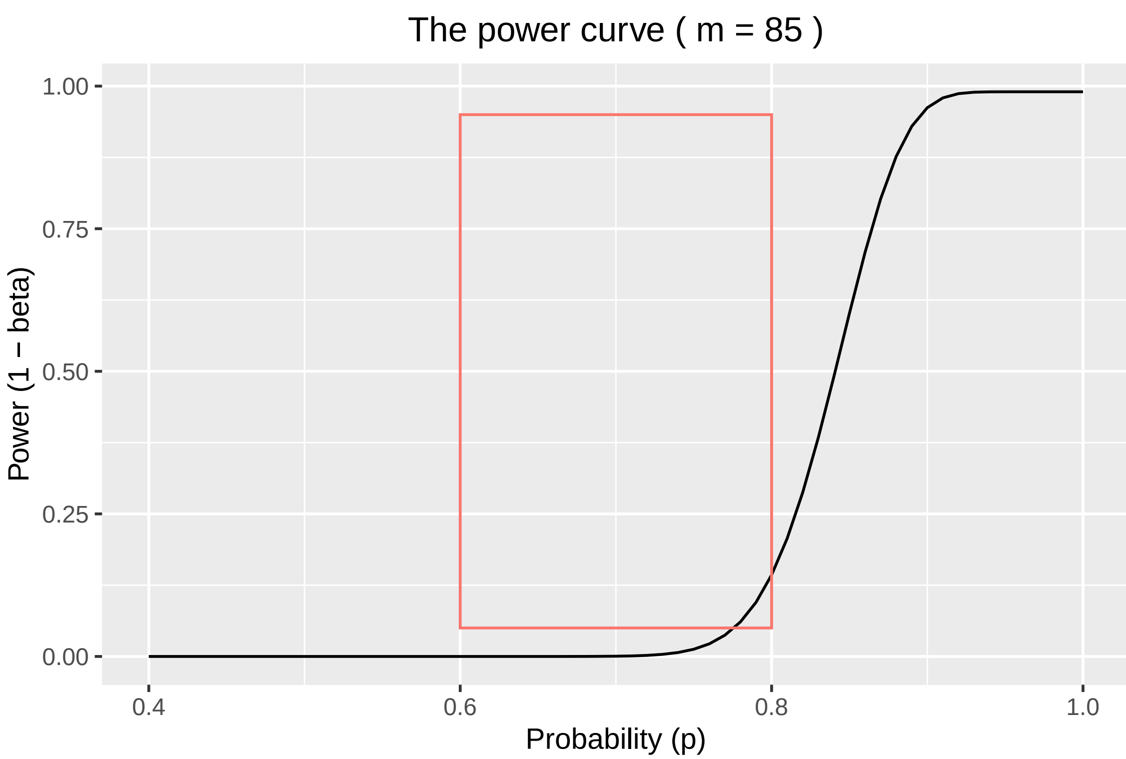
```
box(65)
```


The power curve (m = 65)
1.00
0.75
0.50
0.25
0.00
Power (1 − beta)
0.4
0.6
0.8
1.0
Probability (p)

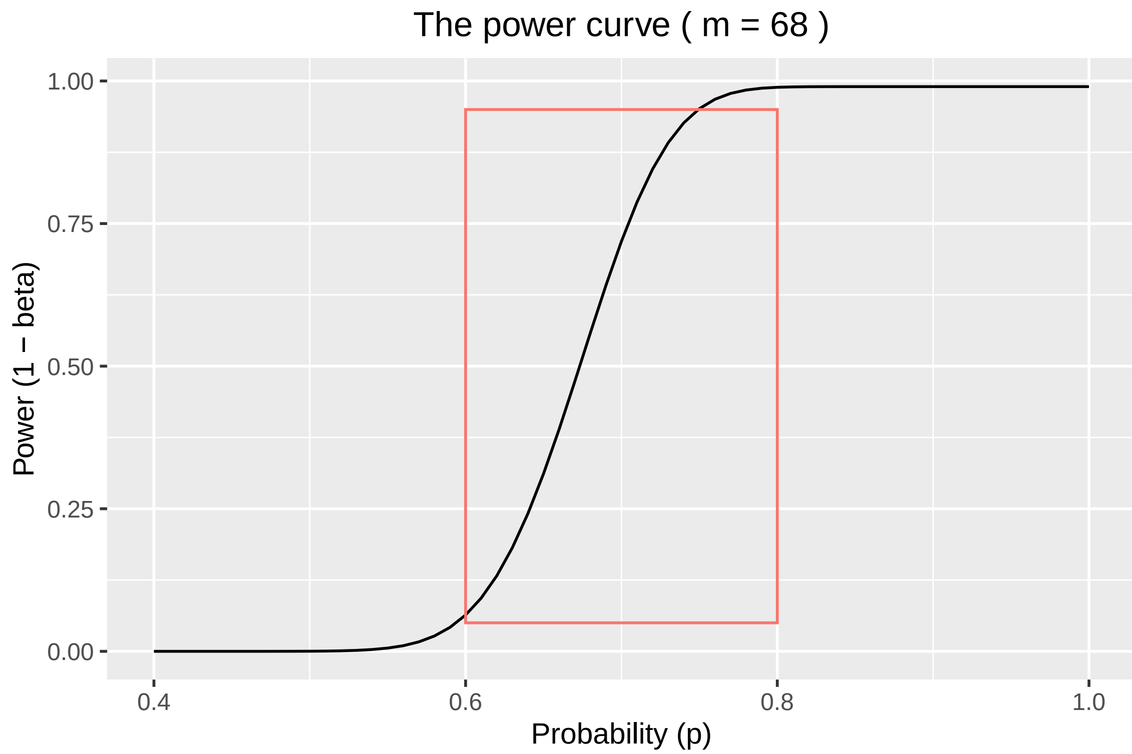
```
box(75)
```


The power curve (m = 75)
1.00
0.75
0.50
0.25
0.00
Power (1 − beta)
0.4
0.6
0.8
1.0
Probability (p)

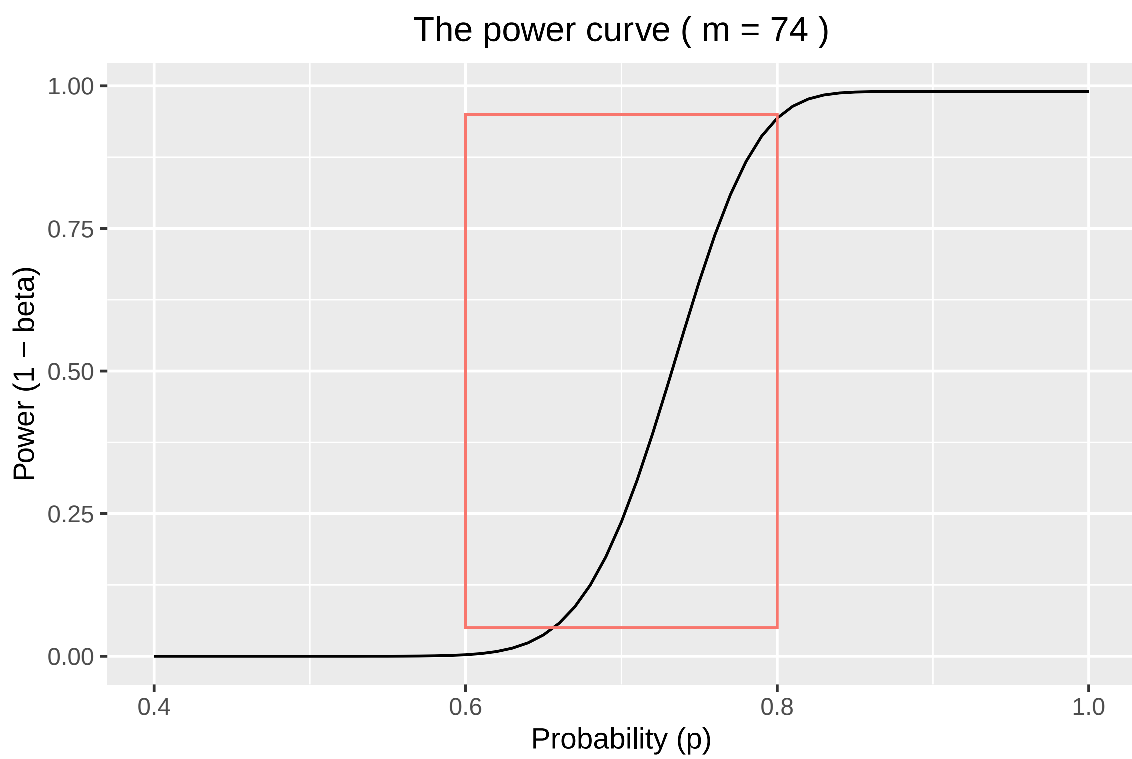
```
box(85)
```


The power curve (m = 85)
1.00
0.75
0.50
0.25
0.00
Power (1 − beta)
0.4
0.6
0.8
1.0
Probability (p)

```
box(68)
```




```
box(74)
```



As doing experiments with different m values, we can confirm with the book that m = 69 is the smallest value for m that obstructs a type 1 error, while m = 73 is the largest which obstructs a type 2. The plot below is showing how the power curve looks like when the critical value is between 69 and 73, which is the same as Figure 3.7 from the book shows.

```r
x_axis <- seq(from = 0.4, to = 1, by = 0.01)
ggplot() +
  geom_line(aes(x_axis, power_curve(69))) +
  geom_line(aes(x_axis, power_curve(73))) +
  geom_rect(aes(xmin = 0.6, xmax = 0.8, ymin = 0.05, ymax = 0.95, col = "red"),
            fill = NA) +
  xlab("Probability (p)") +
  ylab("Power (1 - beta)") +
  ggtitle("The power curve") +
  theme(legend.position = "none") +
  theme(plot.title = element_text(hjust = 0.5))
```