The Assisgnment one

Tie Ma

2024-03-02

#Tie Ma #student number: 101316917 #ECON 5027 #Assignment one

Q2-D

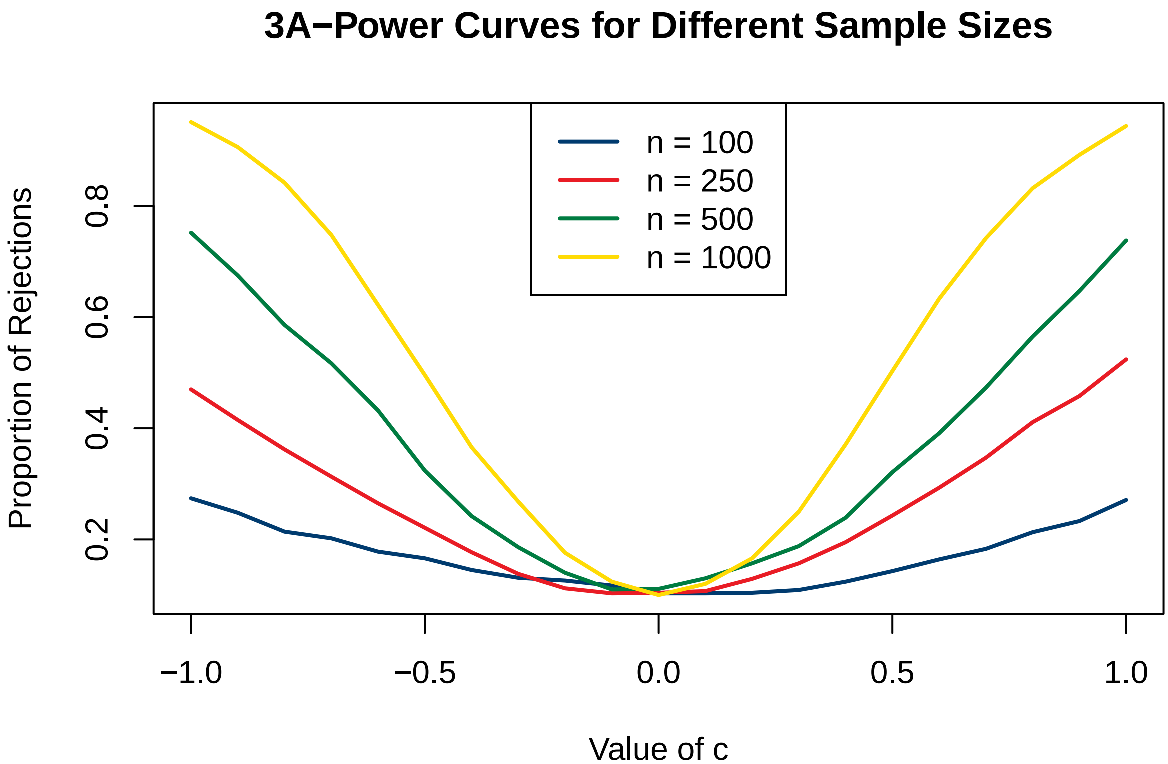
```
##                  beta_hat
## beta_one_hat    0.8037143
## beta_two_hat   -0.4528571
## beta_three_hat -0.3771429
```

```
##      E[Yi|Xi1=x1,Xi2=x2] Xi1 Xi2
## [1,]          0.42657143   0   1
## [2,]          0.04942857   0   2
## [3,]         -0.32771429   0   3
## [4,]         -0.02628571   1   1
## [5,]         -0.40342857   1   2
## [6,]         -0.78057143   1   3
```

Q2-E

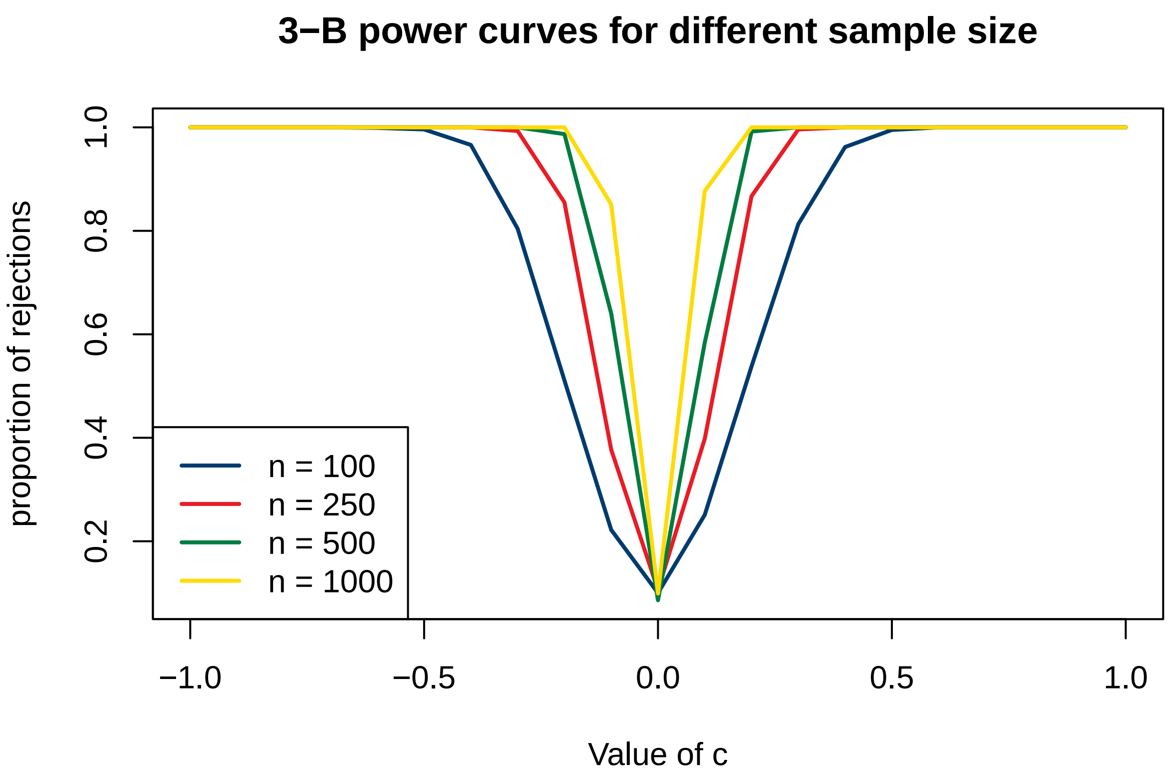
```
##      E[Yi|Xi1=x1,Xi2=x2]_QD E[Yi|Xi1=x1,Xi2=x2]_QA difference(QD -QA) Xi1 Xi2
## [1,]             0.42657143                   0.29         0.13657143   0   1
## [2,]             0.04942857                  -0.03         0.07942857   0   2
## [3,]            -0.32771429                  -0.18        -0.14771429   0   3
## [4,]            -0.02628571                   0.10        -0.12628571   1   1
## [5,]            -0.40342857                  -0.44         0.03657143   1   2
## [6,]            -0.78057143                  -0.96         0.17942857   1   3
```

This table displays the differences between the sample analogs of the conditional expectations we calculated manually in Question A and those estimated using OLS. The table suggests that there are differences between the two, but with a minor margin. Therefore, I think that model (0.1) is a plausible and correct representation.



Q3-A

This graph shows the proportions of rejections for values of c moving closer to the true value of $beta_2 = 0$ across four different sample sizes (n). We can observe that as n increases, the proportion of times that the null hypothesis is rejected also increases for different values of c. This may suggests that large number results in a lower chance of making a Type I error, as the law the big number function. The proportion rejected is not zero for any of the four values of n, according to the graph. This indicates that even when the null hypothesis is true, there is always a non-zero chance of rejection due to random sampling variation, which is a Type I error.



Q3-B

After removing β_1, where the true β_1 is equal to zero, we observed a significant drop in the proportion of Type I errors, which represent false positives. Its also show that while n increase, the type I errors has been droped, suggest the law of big number In the graph for part (a), the proportion of rejections exhibits a smooth decrease as the c-value approaches the true $\hat{\beta}_2$, indicating a gradual reduction in Type I errors. However, in comparison to part (b), Type I errors occur only when the c-value is close to 0.5 for $n = 100$, in contrast to starting at one in the graph for part (a) when $n = 100$. this happened because, including unnecessary predictors in a statistical model increases the risk of Type I and Type II errors due to an increase in model complexity.

Q3-C

```
##          Joint Prob (Not Reject 0.4 AND Reject 0.5) Prob Reject 0.4
## n = 100                                        0.071           0.243
## n = 250                                        0.061           0.421
## n = 500                                        0.030           0.666
## n = 1000                                       0.016           0.897
##          Prob Reject 0.5
## n = 100            0.100
## n = 250            0.108
## n = 500            0.096
## n = 1000           0.109
```

With the increase of the sample size n, the joint probability of failing to reject the null hypothesis for model (0.4) while rejecting it for model (0.5) has decreased. This trend suggests that as we have more data, our tests become more accurate in distinguishing between the correctly specified model and the one that might be misspecified. The increasing sample size leads to more precise estimates, reducing the likelihood of making incorrect inferences about the population parameters.

Q3-D I will using the example of Q3-A and A3-B to aruge with my boss that including an eredundant explanatory variable within the regression model will only increase the posibility of the type one error with a decreasing proference of the model.

Q3-E

The significance of the main variable M_i might change after we remove a less important variable W_i. This can happen for a few reasons that we should carefully think about it. First, W_i could be a control variable. Removing it might make M_i seem more important than it really is. Also, just because W_i doesn't seem important at first doesn't mean it has no value in explaining M_i. It's possible that W_i and the outcome have a non-linear relationship that our model i doesn't pick up. To better understand how variables like M_i and W_i relate, we might want to try other methods like best subset selection or use lasso or ridge regression to decide if a variable should be included.

Q4-A

Q4-A-I

Use the law of iterated expectations

$$\begin{aligned} E[\epsilon_i x_i] &= E[E[\epsilon_i x_i | x_i]] \\ &= E[E[\epsilon_i | x_i] \cdot x_i] \\ &= E[\delta \cdot x_i] \end{aligned}$$

If $\delta = 0$, then:

$$\begin{aligned} E[\delta \cdot x_i] &= E[0 \cdot x_i] \\ &= 0 \end{aligned}$$

$\therefore$ if $E[\epsilon_i x_i] = 0$, then $E[\delta \cdot x_i] = 0$. For the exception to be zero. $\therefore$ we can concluded that

$$E[\epsilon_i x_i] = 0 \quad \text{if and only if} \quad \delta = 0$$

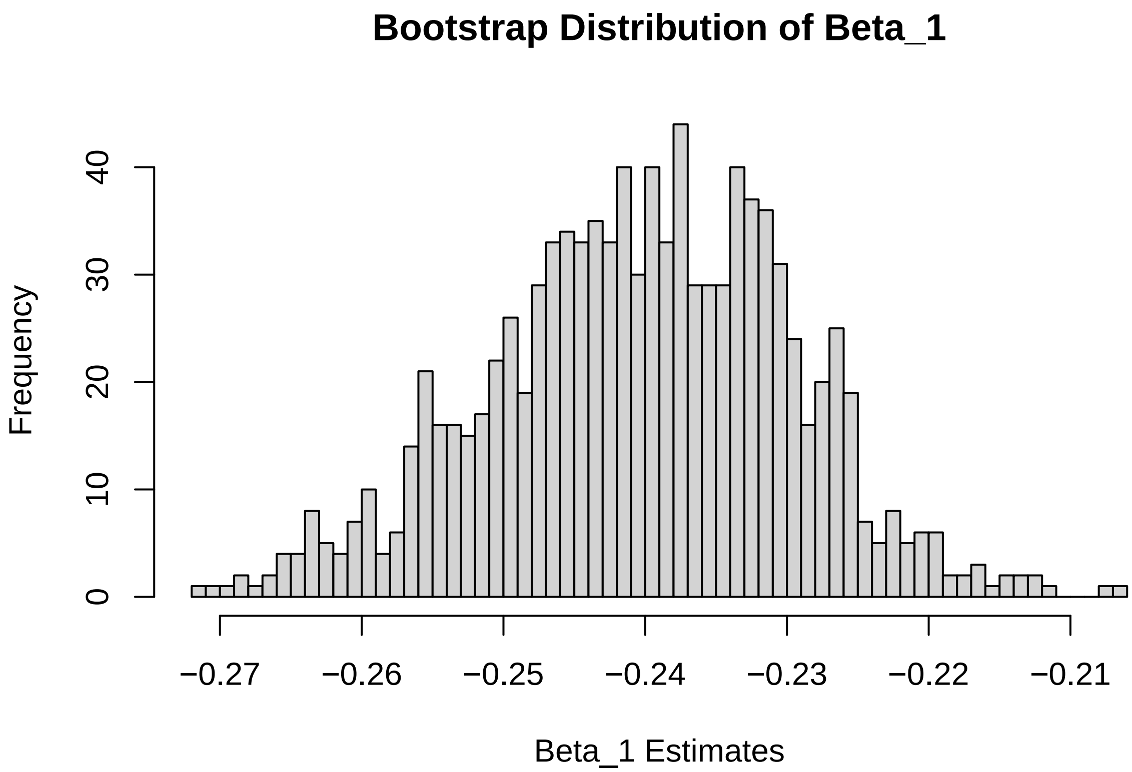
Q4-A-ii We start with the law of iterated expectations.

If $\delta \neq 0$, then:

$$\begin{aligned} E[\epsilon_i x_i] &= E_x[E[\epsilon_i x_i | x_i]] \\ &= E[E[\epsilon_i | x_i] \cdot x_i] \\ &= E[\delta \cdot x_i] \end{aligned}$$

Therefore, $E[\epsilon_i | x_i]$ is not zero if δ is not zero, menans that:

$$E[\epsilon_i | x_i] \neq 0$$



4-a

Q-4-B

```
##                        2.5%      97.5%
## Bootstrap CI   -0.2625680 -0.2204349
## Theoretical CI -0.2606526 -0.2190094
```

The output suggests that the confidence intervals are close between the Bootstrap and theoretical methods, meaning there is a robust estimate for beta_one in the linear regression model. Furthermore, it means that the true value is likely within the shared interval.