

Homework 4

Jacob Thielemier

30 April 2024

Question 1

- The Student Performance Dataset is a dataset designed to examine the factors influencing academic student performance. The dataset consists of 10,000 student records with the outcome measure being a Performance Index rating of 0-100. The covariates measured are: Hours Studied, Previous Scores, Extracurricular Activities, Sleep Hours, and Sample Question Papers Practiced.

Question 2

- Question 1: How does the number of Hours Studied influence the students performance?
- Question 2: What is the interaction of Hours Studied with Hours Slept and how does this influence the students performance?

Question 3

- Alhazmi, E., & Sheneamer, A. (2023). Early Predicting of Students Performance in Higher Education. IEEE Access, 11. <https://doi.org/10.1109/ACCESS.2023.3250702>
- O., O., & P., C. (2017). Predicting Students' Academic Performances – A Learning Analytics Approach using Multiple Linear Regression. International Journal of Computer Applications, 157(4). <https://doi.org/10.5120/ijca2017912671>
- Yang, S. J. H., Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Ogata, H., & Lin, A. J. Q. (2018). Predicting students' academic performance using multiple linear regression and principal component analysis. Journal of Information Processing, 26. <https://doi.org/10.2197/ipsjjip.26.170>

Question 4

- The covariates Extracurricular Activities and Sample Question Papers Practiced are going to be included as potential confounders. Since we are physically limited in the number of hours in a day, we should be able to see the impact of the confounders on Hours Studied and how that influences the overall Students Performance.

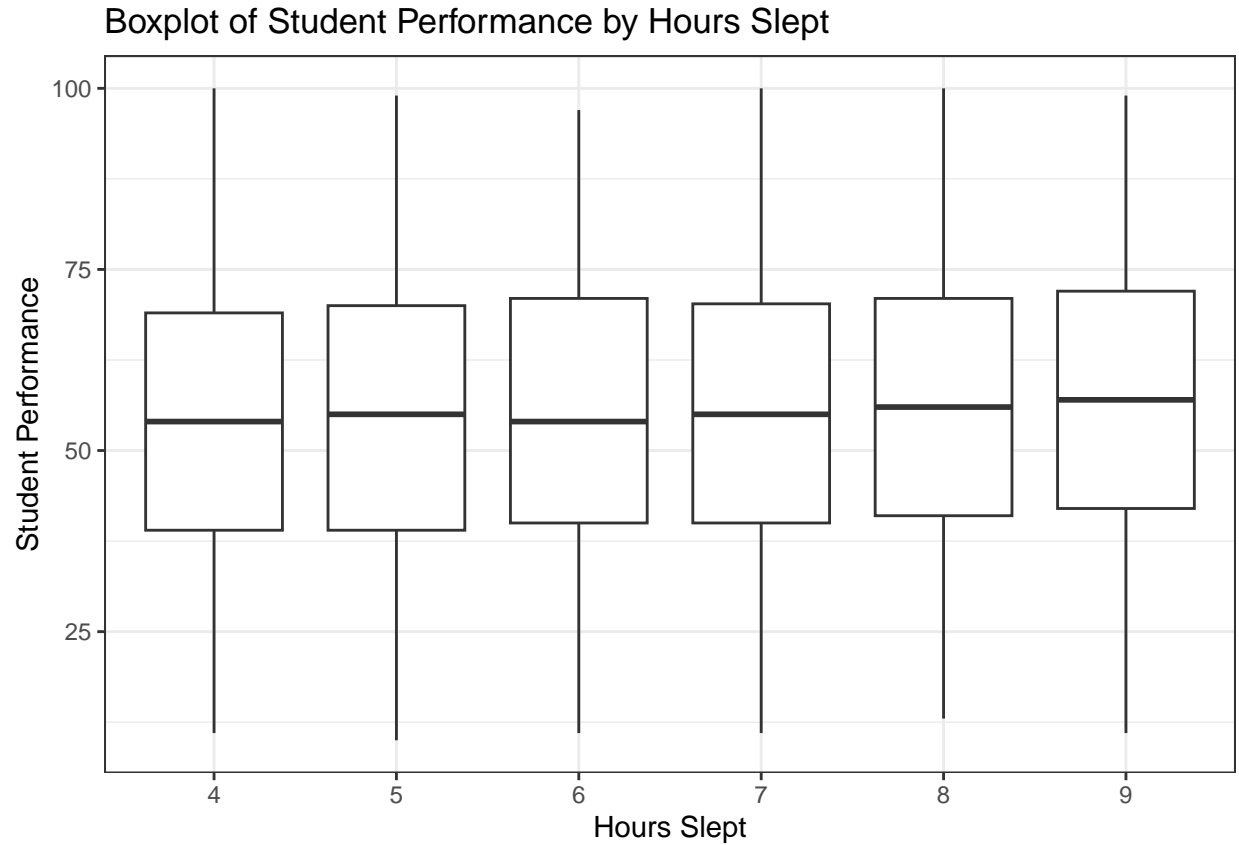
Question 5

- The dataset was collected from 10,000 highschool senior students. The students were from multiple school districts throughout the country. We do not know the sex, race, or geographic location of the students, but we do know they are all the same age.

Question 6

- I added a boxplot to give us a better visual of the data. We can interpret the median Student Performance better at each of the Hours Slept.





Question 7

Part (a)

Table 1: Main Effect Model

Term	Estimate	Std..Error	t.value	P.value	Lower.C.I.	Upper.C.I.
Intercept	131.351	0.452	290.911	< 0.001	130.466	132.236
Ten Year CHD	13.582	0.923	14.718	< 0.001	11.773	15.391
Male	-2.467	0.669	-3.687	< 0.001	-3.779	-1.155

Part (b)

Table 2: Interaction Effect Model

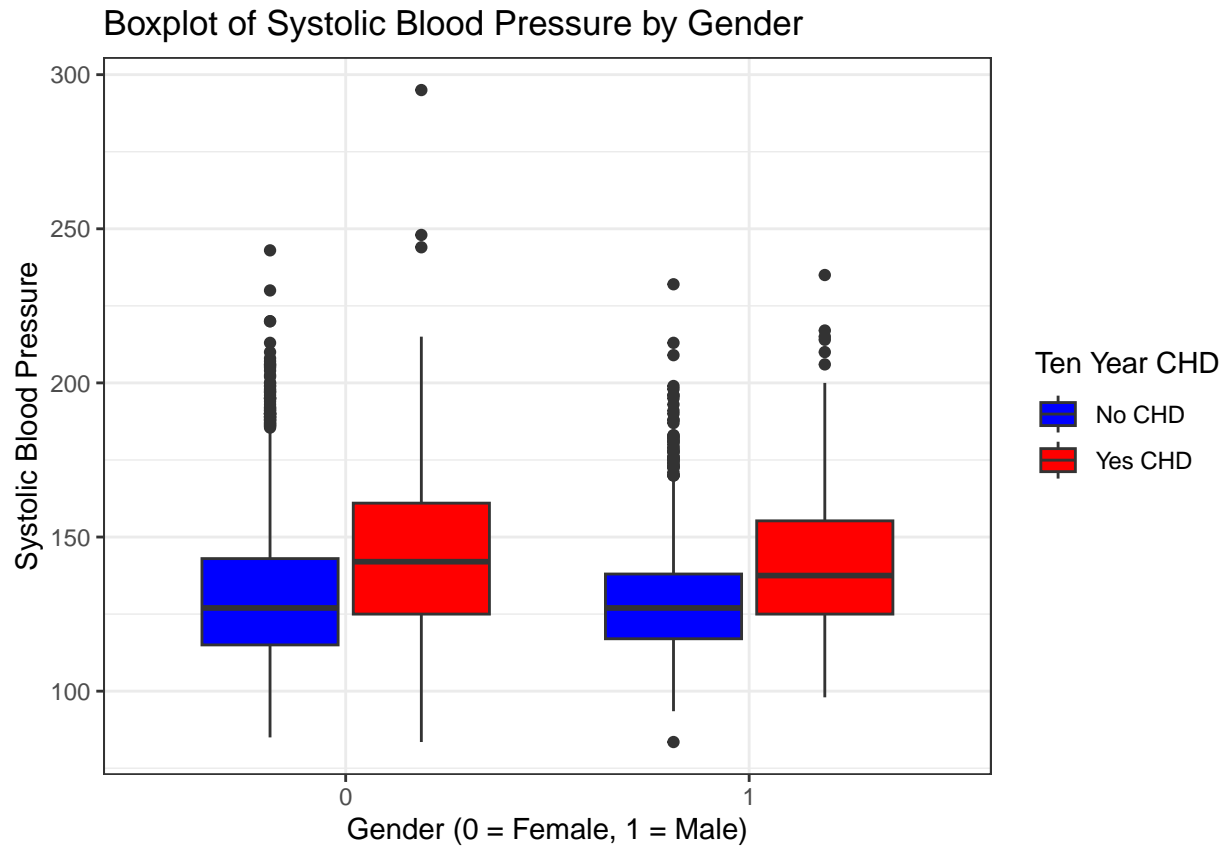
Term	Estimate	Std..Error	t.value	P.value	Lower.C.I.	Upper.C.I.
Intercept	131.090	0.466	281.035	< 0.001	130.176	132.005
Ten Year CHD	15.677	1.323	11.853	< 0.001	13.084	18.270
Male	-1.833	0.728	-2.518	0.012	-3.260	-0.406
Ten Year CHD:Male	-4.079	1.845	-2.210	0.027	-7.697	-0.461

Part (c)

- We can see the change between the two model results in the P-value of Male. There is a slight increase in P-value in the interaction effect model. We also see that the confidence interval is smaller male. The confidence interval for Ten Year CHD also shifted higher.

Part (d)

- We can interpret from the box plot by comparing the means of the boxes for the sex and yes/no CHD. We can see that the median for Yes CHD is different between males and females. This indicates that there is an interaction effect since the medians are different. We can also see that females have a few more outliers based on yes/no CHD indicating a potential interaction effect.



Question 8

The linear model is given by:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 (x_{i1} * x_{i2}) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

Here, x_{i2} is a binary covariate. We want to calculate the variance of the estimator for the difference in means of two observations, where observation A has $x_{A2} = 1$ and $x_{A1} = 1$, and observation B has $x_{B2} = 0$ and $x_{B1} = 1$.

$$E[A|x_{a2} = 1] = \beta_0 + \beta_1 x_{a1} + \beta_2 = \beta_3 x_{a1}$$

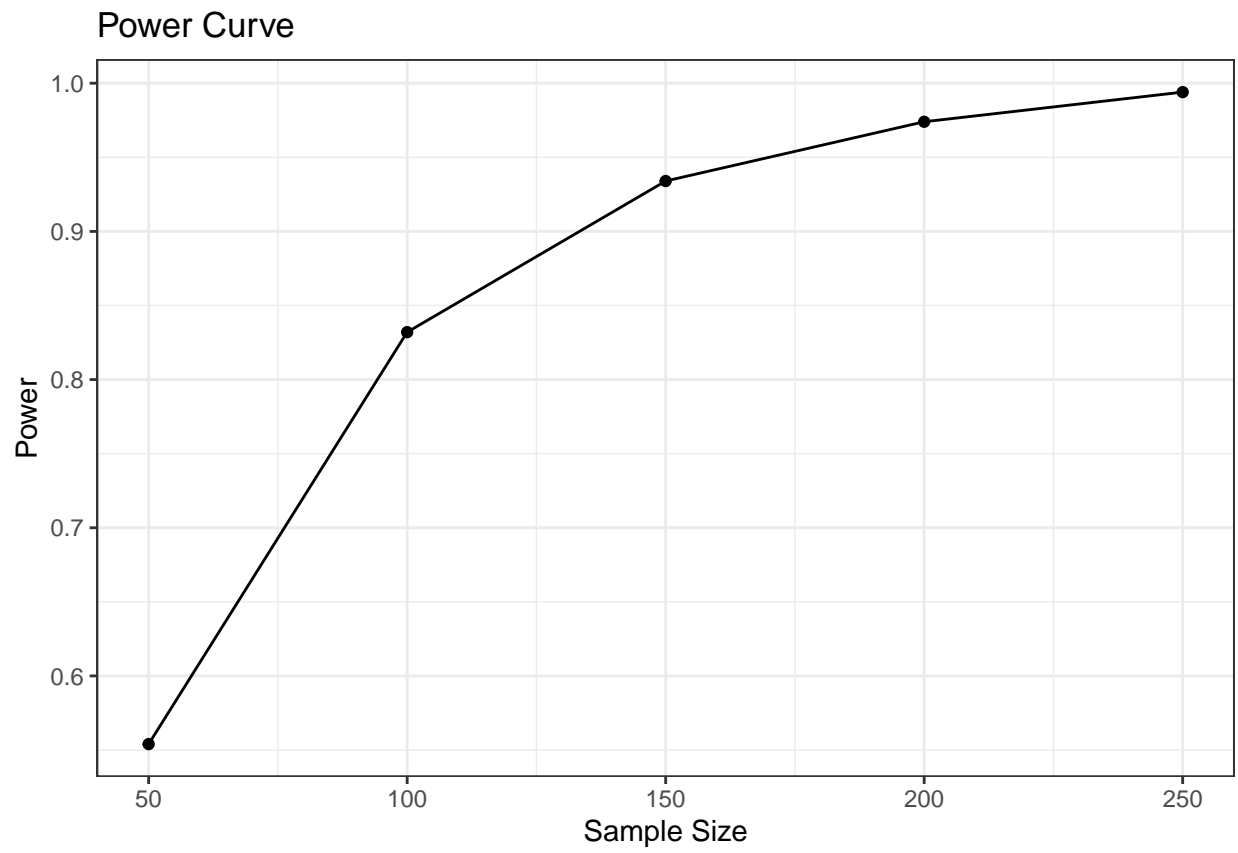
$$E[B|x_{b2} = 0] = \beta_0 + \beta_1 x_{b1}$$

$$E[A - B] = \hat{\beta}_1 + \hat{\beta}_2 + \beta_3 \hat{x}_{a1}$$

This follows the distribution of

$$\hat{\beta} \sim N(\beta, \sigma^2(X^t X)^{-1})$$

Question 9



Question 10

