**ALY6010**

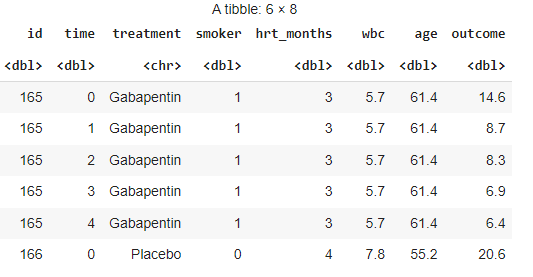
**R-PRACTICE MODULE 5**

**Name: Shreyansh Bhalodiya**

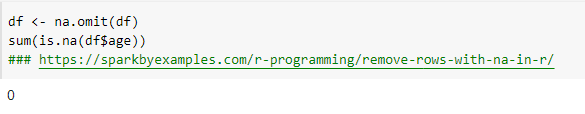
**NUID: 002664707**

**Instructor name: Mohsen Soltanifar**

**Dataset**

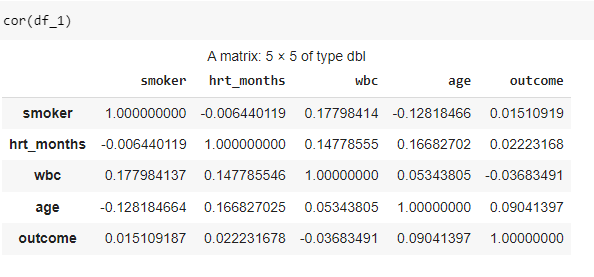
****

* We have a dataset with 8 columns. Looking at the columns we can see that it's about the treatment a person has gone through with different features like smoker, time, wbc, hrt\_months age and outcome.
* Cleaning data from null values.



**PART 1**

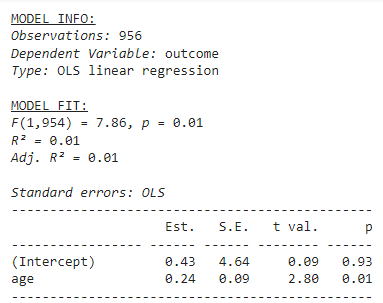
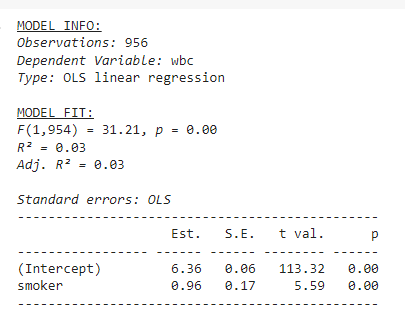
Let’s first check the correlation with numerical columns after cleaning data from null values. We have smoker, hrt\_months, wbc, age, outcome as major variables.



**ANALYSIS**

* We can clearly see that data has no major correlation between variables selected above. We do have positive and negative correlation between few variables for example smoker$hrt\_months, smoker$age and positive correlation between few variables like outcome$smoker, outcome$age and many more.
* We don’t have any strong positive or negative correlation between any pair from the selected variable.
* Let’s look at the regression analysis on the next page and see how it differs between pairs and groups of pairs. Do we get any strong R-value? Let's see.

**PART 2**



**ANALYSIS**

* We tested correlation between wbc and smoker and regression value between the same pair. We found that correlation is 0.117 which is a positive correlation but not a strong one. On the same lines we got an R-value as 0.03 which is again not strong hence we don’t have any strong relation between these two pairs.
* On the similar line if we check correlation between age and outcome is 0.09 which is again not a strong correlation and R-value we got is 0.01 which is again weak.
* We also tested the Regression fit for a group of variables like wbc, age and outcome with smoker as dependent variable but again the R-value we got is 0.05 which is not strong enough to hold the algorithm.

References

* N. (2022, July 20). *How to Remove Rows with NA in R*. Spark by {Examples}. <https://sparkbyexamples.com/r-programming/remove-rows-with-na-in-r/>
* How to check and create a regression line. https://cran.r-project.org/web/packages/jtools/vignettes/summ.html