Application 4

For this application you will create a data set of NBA players, their statistics, and their salaries. At the end, please submit your do file [your last name\_app4\_dofile] and answers [your last name\_app4.docx] to the Application 4 folder.

Step 1: collect data consisting of player stats from 2008-2009 season through the 2019-20 season. This can be accessed at <https://www.basketball-reference.com/>. We want season total statistics.

1. Create a single STATA data set with statistics for these NBA seasons using a loop.

Step 2: Collect salary data from the 2019-2020 season from the basketball-reference.com.

1. Create a single STATA data set with player salaries for 2019-20

Step 3: merge two data sets together.

1. Using 2018-2019 statistics data, merge it with 2019-20 salary data set.
2. After the merge, type the command: tab \_merge. Explain the three values of these variable.
3. Explain what data you will keep, what you will drop, and if there are any issues that you see with the merge.

Step 4: Data Cleaning

1. Create the following variables and fill in the table:

|  |  |  |
| --- | --- | --- |
|  | Mean | Std. Dev |
| Point per game |  |  |
| Rebounds per game |  |  |
| Assists per game |  |  |
| Steals per game |  |  |
| Turnovers per game |  |  |
| 2-pt Field Goal % |  |  |
| 3-pt Field Goal% |  |  |
| Free Throw% |  |  |
| Salary |  |  |

1. Estimate the following models and create a table with the regression outputs
2. What statistical factors affects player salary? Explain.

BONUS PART

1. Create a data set that has all NBA players included in Basketball reference with variables that include position, height, weight, and where they went to college.
2. This may require you to go outside of basketball reference, but create a data set with salary from as many years as you can find.

<https://rstudio-pubs-static.s3.amazonaws.com/369734_20ddb138d1af4b488b3a978ee055ec24.html>