College\_Sucess\_Analyse

Superuser

18 11 2021

# Data Description:

This data set, “College Success”, provides high school grades, SAT scores, and Grade Point Average of 224 university students.

Variables:

* id - Participant ID.
* gpa - Grade Point Average (GPA) after three semester in college.
* hsm - Average high-school grade in mathematics.
* hss - Average high-school grade in science.
* hse - Average high-school grade in English.
* satm - SAT score for mathematics.
* satv - SAT score for verbal knowledge.
* sex - Gender (labels not available)

This example Analysis demonstrates the use of linear regression. Specifically, we will examine which variables best predict GPA. First, we will fit a model predicting GPA by high school grades. Then, we will use a model that predicts GPA by SAT scores. Finally, we wil fit a model that uses both high shool grades and SAT scores to predict GPA.

# Data Import

As the data is given in form of a .csv file we have to import it using the read.csv() or readr::read\_csv() functions.

library(readr)  
data\_college\_success <-   
 read\_csv("College\_Success.csv")

## Rows: 224 Columns: 8

## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (8): id, gpa, hsm, hss, hse, satm, satv, sex

##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# References

Moore, D. S., McCabe, G. P., and Craig, B. A. (2012). Introduction to the Practice of Statistics (7th ed.). New York: Freeman.

Campbell, P. F. and McCabe, G. P. (1984). Predicting the success of freshmen in a computer science major. Communications of the ACM, 27: 1108–1113.