Regression

2022-09-21

Contents

In this document, we do basic descriptive and regression analysis to understand what variables best predict freshman GPA.

1. First install/ load the R packages we need

```
library(tidyverse)
library(broom)
library(modelsummary)

gpa.data <- read_csv("data/satgpa.csv")
attach(gpa.data)
```

Exploratory questions

How well do SAT scores correlate with freshman GPA?

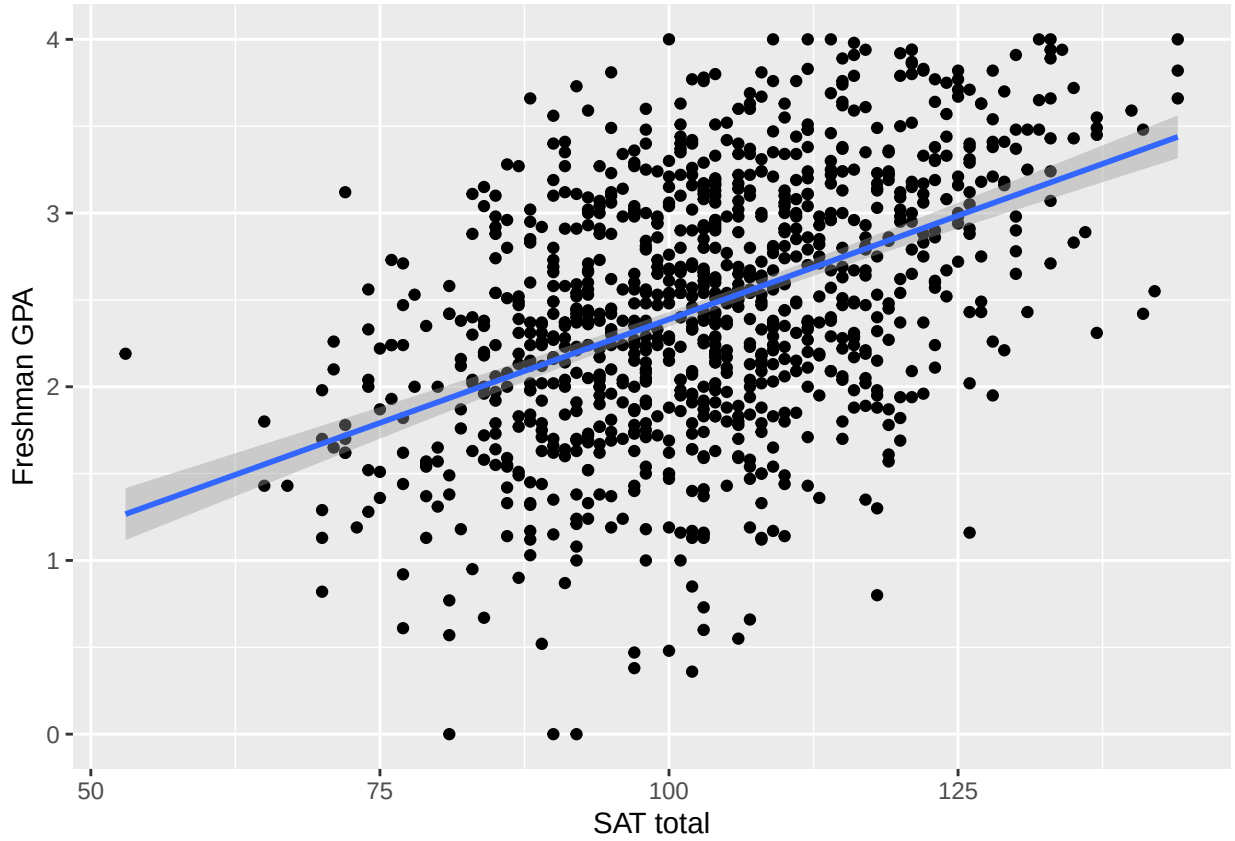
```
cor(gpa_fy, sat_total)
```

```
## [1] 0.460281
```

A correlation of -1 means perfect negative correlation. A correlation of 0 means, no correlation between the two. And a correlation of 1 means perfect positive correlation. The above result shows a positive correlation between SAT scores and freshman GPA.But it is not very strong, meaning close to 1.

```
ggplot(data = gpa.data, mapping = aes(x = sat_total, y = gpa_fy)) +
  geom_point() +
  geom_smooth(method = lm) +
  labs(y = "Freshman GPA", x = "SAT total")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



The above plot shows the positive correlation between SAT scores and freshman GPA.