Linear Regression Example

2025-09-26

Goal for this is to help you see how we code a linear model in R before we actually get to the lab section. First, we load up the data and any libraries we think we will need.

NOTE: most (all?) of you will need to install ggResidpanel. It's best to install it using the console (the command line down below) and not the markdown document itself. You only ever have to download a package once so keeping the code in the markdown is just asking for trouble/bugs/annoying output

```
library(ggplot2)
library(ggResidpanel)
```

```
##
## Attaching package: 'ggResidpanel'
```

```
## The following object is masked from 'package:datasets':
##
##     penguins
```

```
colleges <- read.csv("https://remiller1450.github.io/data/Colleges2019_Complete.csv")

head(colleges) #look at the first few rows of data
```

```
##   X                          Name        City State Enrollment Private
## 1 1 Abilene Christian University     Abilene    TX       3524 Private
## 2 3           Adelphi University Garden City    NY       5307 Private
## 3 4               Adrian College      Adrian    MI       1781 Private
## 4 5      AdventHealth University     Orlando    FL       1166 Private
## 5 8     Alabama A & M University      Normal    AL       4990  Public
## 6 9     Alabama State University  Montgomery    AL       3903  Public
##        Region Adm_Rate ACT_median ACT_Q1 ACT_Q3  Cost Net_Tuition
## 1  South West   0.5696         24     21     21 48046       16177
## 2    Mid East   0.7418         25     22     22 49008       24971
## 3 Great Lakes   0.6481         23     19     19 51626       14136
## 4  South East   0.8689         20     18     18 24338       15360
## 5  South East   0.8986         18     16     16 22489        7413
## 6  South East   0.9774         18     16     16 21476       10160
##   Avg_Fac_Salary PercentFemale PercentWhite PercentBlack PercentHispanic
## 1          69804     0.6118200       0.7946       0.0814          0.1635
## 2         111339     0.7211121       0.6669       0.1785          0.1292
## 3          72873     0.4221106       0.8861       0.0692          0.0318
## 4          69759     0.8251058       0.7622       0.1395          0.1338
## 5          63909     0.5640301       0.4684       0.4798          0.0379
## 6          69786     0.6134185       0.4269       0.5232          0.0409
##   PercentAsian FourYearComp_Males FourYearComp_Females Debt_median
```

```
## 1       0.0287          0.4115756            0.5283019       16000
## 2       0.0673          0.6114650            0.6998855       19500
## 3       0.0121          0.2320917            0.3319838       18468
## 4       0.0259          0.4761905            0.4132231       16646
## 5       0.0148          0.1471572            0.2313665       15000
## 6       0.0141          0.1282051            0.2679211       18950
##   Salary10yr_median
## 1             43000
## 2             58500
## 3             38600
## 4             56000
## 5             31000
## 6             27700
```

```
colnames(colleges) #and look at the column names
```

```
##  [1] "X"                    "Name"                 "City"
##  [4] "State"                "Enrollment"           "Private"
##  [7] "Region"               "Adm_Rate"             "ACT_median"
## [10] "ACT_Q1"               "ACT_Q3"               "Cost"
## [13] "Net_Tuition"          "Avg_Fac_Salary"       "PercentFemale"
## [16] "PercentWhite"         "PercentBlack"         "PercentHispanic"
## [19] "PercentAsian"         "FourYearComp_Males"   "FourYearComp_Females"
## [22] "Debt_median"          "Salary10yr_median"
```