Statistical Analysis Data Collection Project 1

Benjamin Goldstone, Dina Maltser, Eleanor Raff, Faith Fezenko

2/3/23

Question of Interest

As college students ourselves, we recognize how expensive college can be, and also how much the cost of college tuition varies across different 4-year colleges. Therefore, we wanted to compare which factors influence the tuition cost of 4-year undergraduate degrees across the United States.

Variables

Response Variable:

Tuition: Cost (in U.S. dollars) of tuition at 4-year undergraduate programs at colleges across the United States. For public colleges, the out-of-state tuition price was used.

Explanatory Variables:

Public/Private: Whether the college was subsidized by a particular state or privately owned.

Undergraduate Enrollment: The number of students currently pursuing an undergraduate degree at each college.

Acceptance Rate: The proportion of incoming first year students who applied to a particular college that were accepted, often looked at as a measure of academic prestige.

Number of Majors: the number of four-year degree offerings at a particular college.

Data Collection Method

We began by searching for a list of all 4-year colleges in the United States (the list used was from 4icu.org). Then, we followed the numbers assigned to each college in the online list, which was sorted to some extent by academic prestige, and entered those numbers (1-1762) into a random number generator. We ran the random number generator to randomly select 50 numbers, then matched each number with the college it represented on the list. When the random number generator selected a college that only offered master's degree programs, we discarded that result and randomly selected another number. Each group member was then assigned one variable to collect results on for the 50 colleges selected.

```
library(readr)
StatsProject1 <- read_csv(
  "StatsProjectCollegeData.csv",
  col_types = cols(
    `Name of College` = col_character(),
    `Tuition price (out of state)` = col_integer(),
    `Public/Private` = col_character(),
    `School Size` = col_integer(),
```

```
    `Acceptance Rate` = col_double(),
    `Number of Majors` = col_integer()
  )
)
```

```
## Warning: The following named parsers don't match the column names: Name of
## College, Tuition price (out of state), Public/Private, School Size, Acceptance
## Rate, Number of Majors
```

```
head(StatsProject1)
```

```
## # A tibble: 6 x 6
##   NameOfCollege             TuitionPrice PublicPrivate SchoolSize AcceptanceRate
##   <chr>                            <dbl> <chr>              <dbl>          <dbl>
## 1 Ohio Wesleyan University~        48832 Private             1339           74
## 2 Houghton College (1093)          16446 Private              871           79.1
## 3 Lee University (1039)            20500 Private             3927           82.9
## 4 Samford University (503)         35360 Private             3607           83.9
## 5 Old Dominion University ~        30960 Public             18678           96.2
## 6 Allen University (1663)          13340 Private              656           56.6
## # ... with 1 more variable: NumberOfMajors <dbl>
```

```
nrow(StatsProject1)
```

```
## [1] 50
```