

## Topic 2: Exercise 1

Daniel Alonso

November 25th, 2020

### Importing libraries

```
library(dplyr)
```

### Importing data as described by exercise

```
d <- read.csv("../datasets/Colleges.csv")
```

### Replacing binary variable Private with 1 and 0

```
d$Private <- ifelse(d$Private == "Yes", 1, 0)
```

### Selecting columns

```
data <- d %>% dplyr::select('Private', 'Apps', 'Accept', 'Enroll', 'F.Undergrad')
```

### Calculating covariances

```
cov_matrix <- cov(data)
cov_matrix
```

```
##           Private      Apps      Accept      Enroll  F.Undergrad
## Private      0.1986559    -745.3552    -519.2042    -235.1942    -1330.764
## Apps        -745.3552439  14978459.5301   8949859.8119   3045255.9876   15289702.474
## Accept      -519.2042169   8949859.8119   6007959.6988   2076267.7627   10393582.435
## Enroll      -235.1942393   3045255.9876   2076267.7627   863368.3923    4347529.884
## F.Undergrad -1330.7637175  15289702.4742  10393582.4355  4347529.8841   23526579.326
```

### Calculating correlations

```
corr_matrix <- cov2cor(cov_matrix)
corr_matrix
```

##	Private	Apps	Accept	Enroll	F.Undergrad
## Private	1.0000000	-0.4320947	-0.4752520	-0.5679078	-0.6155605
## Apps	-0.4320947	1.0000000	0.9434506	0.8468221	0.8144906
## Accept	-0.4752520	0.9434506	1.0000000	0.9116367	0.8742233
## Enroll	-0.5679078	0.8468221	0.9116367	1.0000000	0.9646397
## F.Undergrad	-0.6155605	0.8144906	0.8742233	0.9646397	1.0000000

### What information does the sample covariance provide?

We know that because the Private variable (binary variable) has only 2 possible values, its covariance with other variables is always going to be relatively small.

What information does the sample correlation provide?

Scatter plot of our quantitative variables and the Private binary variable

```
plot(d$Private, d$Apps)
```

