STA 325 Case Study

Load libraries and data

```
## # A tibble: 89 x 7
##       St    Re      Fr R_moment_1 R_moment_2 R_moment_3 R_moment_4
##    <dbl> <dbl>   <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
##  1  0.1    224   0.052   0.00216     0.130      14.4       1586.
##  2  3      224   0.052   0.00379     0.470      69.9      10404
##  3  0.7    224 Inf       0.00291     0.0435      0.822      15.6
##  4  0.05    90 Inf       0.0635      0.0907      0.467       3.27
##  5  0.7    398 Inf       0.000369    0.00622     0.126       2.57
##  6  2       90   0.3     0.148       2.01       36.2        672.
##  7  0.2     90 Inf       0.0813      0.324       3.04       33.0
##  8  3      224 Inf       0.00575     0.120       2.75       63.2
##  9  0.9    224 Inf       0.00302     0.0452      0.845      15.8
## 10  0.6    398   0.052   0.000314    0.00447     0.0821      1.51
## # ... with 79 more rows
```

```
##                     St          Re  Fr R_moment_1 R_moment_2 R_moment_3
## St          1.00000000 -0.03169871 NaN  0.2147681  0.1479257  0.1647465
## Re         -0.03169871  1.00000000 NaN -0.7747206 -0.3932344 -0.3844289
## Fr                 NaN         NaN   1        NaN        NaN        NaN
## R_moment_1  0.21476813 -0.77472058 NaN  1.0000000  0.6298829  0.6217326
## R_moment_2  0.14792571 -0.39323445 NaN  0.6298829  1.0000000  0.9984335
## R_moment_3  0.16474648 -0.38442895 NaN  0.6217326  0.9984335  1.0000000
## R_moment_4  0.18004537 -0.37741773 NaN  0.6150484  0.9946671  0.9988414
##            R_moment_4
## St          0.1800454
## Re         -0.3774177
## Fr                NaN
## R_moment_1  0.6150484
## R_moment_2  0.9946671
## R_moment_3  0.9988414
## R_moment_4  1.0000000
```

```
## # A tibble: 23 x 3
##       St    Re      Fr
##    <dbl> <dbl>   <dbl>
##  1  0.05   398   0.052
##  2  0.2    398   0.052
##  3  0.7    398   0.052
##  4  1      398   0.052
##  5  0.1    398 Inf
##  6  0.6    398 Inf
##  7  1      398 Inf
##  8  1.5    398 Inf
```

```
## 9   3       398  Inf
## 10  3       224   0.3
## # ... with 13 more rows
```

Exploratory Data Analysis

```
# We transform the 'Fr' variable using the sigmoid function so that this variable
# will be within a finite range.

train1 <- train %>%
  mutate(Fr_sigmoid = 1 / ( 1 + exp(-Fr))) %>%
  subset(select = c(1:3, 8, 4:7))

train1
```

```
## # A tibble: 89 x 8
##        St    Re      Fr Fr_sigmoid R_moment_1 R_moment_2 R_moment_3 R_moment_4
##     <dbl> <dbl>   <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
##  1  0.1     224   0.052      0.513   0.00216     0.130      14.4       1586.
##  2  3       224   0.052      0.513   0.00379     0.470      69.9      10404
##  3  0.7     224 Inf          1       0.00291     0.0435      0.822      15.6
##  4  0.05     90 Inf          1       0.0635      0.0907      0.467       3.27
##  5  0.7     398 Inf          1       0.000369    0.00622     0.126       2.57
##  6  2        90   0.3        0.574   0.148       2.01       36.2       672.
##  7  0.2      90 Inf          1       0.0813      0.324       3.04       33.0
##  8  3       224 Inf          1       0.00575     0.120       2.75       63.2
##  9  0.9     224 Inf          1       0.00302     0.0452      0.845      15.8
## 10  0.6     398   0.052      0.513   0.000314    0.00447     0.0821      1.51
## # ... with 79 more rows
```

```
cor(train1)
```

```
##                     St          Re  Fr  Fr_sigmoid R_moment_1 R_moment_2
## St          1.00000000 -0.03169871 NaN -0.04734175  0.2147681  0.1479257
## Re         -0.03169871  1.00000000 NaN  0.11152749 -0.7747206 -0.3932344
## Fr                 NaN         NaN   1         NaN        NaN        NaN
## Fr_sigmoid -0.04734175  0.11152749 NaN  1.00000000 -0.1364384 -0.2896720
## R_moment_1  0.21476813 -0.77472058 NaN -0.13643841  1.0000000  0.6298829
## R_moment_2  0.14792571 -0.39323445 NaN -0.28967203  0.6298829  1.0000000
## R_moment_3  0.16474648 -0.38442895 NaN -0.28369640  0.6217326  0.9984335
## R_moment_4  0.18004537 -0.37741773 NaN -0.27852083  0.6150484  0.9946671
##            R_moment_3 R_moment_4
## St          0.1647465  0.1800454
## Re         -0.3844289 -0.3774177
## Fr                NaN        NaN
## Fr_sigmoid -0.2836964 -0.2785208
## R_moment_1  0.6217326  0.6150484
## R_moment_2  0.9984335  0.9946671
## R_moment_3  1.0000000  0.9988414
## R_moment_4  0.9988414  1.0000000
```

```r
test1 <- test %>%
  mutate(Fr_sigmoid = 1 / ( 1 + exp(-Fr)))

test1
```

```
## # A tibble: 23 x 4
##       St    Re      Fr Fr_sigmoid
##    <dbl> <dbl>   <dbl>      <dbl>
##  1  0.05   398   0.052      0.513
##  2  0.2    398   0.052      0.513
##  3  0.7    398   0.052      0.513
##  4  1      398   0.052      0.513
##  5  0.1    398 Inf          1
##  6  0.6    398 Inf          1
##  7  1      398 Inf          1
##  8  1.5    398 Inf          1
##  9  3      398 Inf          1
## 10  3      224   0.3        0.574
## # ... with 13 more rows
```

```r
# R_moment_2 is almost perfectly correlated with R_moment_3 and R_moment 4.
```

We will try to create these 4 models:

- **Response**: R_moment_1 & **Predictors (Main Effects)**: St, Re, Fr_sigmoid

We will attempt to use a combination of subset selection, polynomial, transformation, and interaction variables.

- **Response**: R_moment_2 & **Predictors (Main Effects)**: St, Re, Fr_sigmoid, R_moment_1

We will attempt to use a combination of subset selection, polynomial, transformation, and interaction variables. We will also include R_moment_1 since it is has significant positive relationship with R_moment_2 (~0.63).

- **Response**: R_moment_3 & **Predictors (Main Effects)**: R_moment_2

We know that R_moment_2 is almost perfectly correlated (>0.99) with R_moment_3, so only using one predictor variable is enough. We try to avoid overfitting by using only R_moment_2 as our only predictor to predict R_moment_3. We will attempt to use polynomial and transformation variables.

- **Response**: R_moment_4 & **Predictors (Main Effects)**: R_moment_2, R_moment_3

Same reasoning - R_moment_2 and R_moment_3 are almost perfectly correlated with R_moment_4. We will only use these 2 predictors and will attempt to use both transformation and interaction variables (since R_moment_2 and R_moment_3 are also highly correlated to each other).

Predictive models

Apply to test data