

- 1) The data frame `buildings` contains data about buildings in the Bronx. Each row represents one building, and contains the building's address, number of floors, number of elevators in the building, and the average floor size in square feet. The first 5 rows are:

Address	Number_of_floors	Number_of_elevators	Floor_size
250 Bedford Park Blvd	3	7	1000
500 Kingsbridge Ave.	6	2	1200
300 Fordham Road	10	4	1500
2300 Grand Concourse	7	1	800
800 Sedgewick Ave.	6	3	2000

The following Python code creates a linear model from the data in `buildings`:

```
lm = smf.ols(formula = 'Number_of_elevators ~ Number_of_floors +  
                    Floor_size', data = buildings).fit()
```

- a) What is/are the dependent variable(s)?
- b) What is/are the independent variable(s)?
- c) Write Python code to compute the R-Squared value of this linear model. It is fine if you code computes more than the R-Squared value.

d) What is the best possible R-Squared value?

e) Suppose instead we had the linear model:

```
lm2 = smf.ols(formula = 'Number_of_elevators ~ Number_of_floors',  
              data = buildings).fit()
```

Write Python code to predict the number of elevators in a building with 5 floors.