Start-Tech Academy

Multiple Linear Regression

F statistics

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -6.498625   5.264076  -1.235   0.2176
crime_rate               0.009710   0.348185   0.028   0.9778
resid_area              -0.040875   0.057585  -0.710   0.4782
air_qual               -15.897400   4.003793  -3.971 8.24e-05 ***
room_num                 4.019017   0.426606   9.421  < 2e-16 ***
age                     -0.005715   0.013606  -0.420   0.6747
teachers                 1.007001   0.122098   8.247 1.50e-15 ***
poor_prop               -0.577271   0.052695 -10.955  < 2e-16 ***
airportYES               1.131516   0.454266   2.491   0.0131 *
n_hos_beds               0.329221   0.152239   2.163   0.0311 *
n_hot_rooms              0.091868   0.082174   1.118   0.2641
waterbodyLake            0.264086   0.641963   0.411   0.6810
`waterbodyLake and River` -0.687556   0.714023  -0.963   0.3361
waterbodyRiver           -0.291319   0.546656  -0.533   0.5943
rainfall                 0.016119   0.017839   0.904   0.3667
avg_dist                -1.218640   0.188933  -6.450 2.68e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.925 on 490 degrees of freedom
Multiple R-squared:  0.7208,    Adjusted R-squared:  0.7123
F-statistic: 84.34 on 15 and 490 DF,  p-value: < 2.2e-16
```

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0 \qquad H_a : \text{ at least one } \beta_j \text{ is non-zero.}$$

Multiple Linear Regression

F statistics

Coin is biased if I get 5 consecutive heads in 5 tosses





Probability of Head	1/2	1/2	1/2	1/2	1/2

Probability of classifying a fair coin as a biased coin = (1/2)^5 = 0.03125

If 100 coins are tossed 5 times each ,
What is the probability of getting all heads in at least one of the coin

$$1 - \left(1 - \frac{1}{32}\right)^{100} \approx 95\%$$

Multiple Linear Regression

F statistics

For our model

	β	β	β	β	β
Probability of Wrongly classifying B as significant	5%	5%	5%	5%	5%

If number of variables is large, there is very high chance that one of the B is wrongly classified

$$F = \frac{(\mathrm{TSS} - \mathrm{RSS})/p}{\mathrm{RSS}/(n-p-1)},$$