# Importing required libraries and dataset

Table of Hospitalisation details (first 5 rows):

Customer ID year month date children charges Hospital tier City tier \

0 Id2335 1992 Jul 9 0 563.84 1.0 2.0

1 Id2334 1992 Nov 30 0 570.62 1.0 0.0

2 Id2333 1993 Jun 30 0 600.00 1.0 0.0

3 Id2332 1992 Sep 13 0 604.54 2.0 2.0

4 Id2331 1998 Jul 27 0 637.26 2.0 2.0

State ID\_R1011 State ID\_R1012 State ID\_R1013 patient age

0 0.0 0.0 1.0 33

1 0.0 0.0 1.0 33

2 0.0 0.0 1.0 32

3 0.0 0.0 1.0 33

4 0.0 0.0 1.0 27

Table of Medical Examinations:

Customer ID BMI HBA1C NumberOfMajorSurgeries Heart Issues\_No \

0 Id1 47.410 7.47 1.0 1.0

1 Id2 30.360 5.77 1.0 1.0

2 Id3 34.485 11.87 0.0 0.0

3 Id4 38.095 6.05 1.0 1.0

4 Id5 35.530 5.45 1.0 1.0

Heart Issues\_yes smoker\_No smoker\_yes Any Transplants\_No \

0 0.0 0.0 1.0 1.0

1 0.0 0.0 1.0 1.0

2 1.0 0.0 1.0 1.0

3 0.0 0.0 1.0 1.0

4 0.0 0.0 1.0 1.0

Any Transplants\_yes Cancer history\_No Cancer history\_Yes

0 0.0 1.0 0.0

1 0.0 1.0 0.0

2 0.0 1.0 0.0

3 0.0 1.0 0.0

4 0.0 1.0 0.0

Table of Names:

Customer ID last\_name first\_name gender

0 Id1 Hawks Ms. Kelly f

1 Id2 Lehner Mr. Matthew D m

2 Id3 Lu Mr. Phil m

3 Id4 Osborne Ms. Kelsey f

4 Id5 Kadala Ms. Kristyn f

# Hypothesis Testing

## The average hospitalization costs for the three types of hospitals are not significantly different.

1. F-statistic: 495.4738392647216
2. P-value: 5.5985677688061205e-180
3. Since P-value > 0.05 (critical value), tyep of hospitals are not sigificantly different!

## 2. The average hospitalization costs for the three types of cities are not significantly different.

F-statistic: 1.5506960494320452

P-value: 0.21231948148820598

Since P-value > 0.05 (critical value), tyep of hospitals are not sigificantly different!

## 3. The average hospitalization costs for smokers is not significantly different from the average cost for nonsmokers.

F-statistic: 5499.05404908718

P-value: 0.0 10

Since P-value < 0.05 (critical value), smoker and non-smoker are sigificantly different!

## 4. Smoking and heart issues are independent.

Contingency Table:

Smoker 0.0 1.0

health issue

0.0 1110 297

1.0 735 191

P-value: 0.8194

Since P-value > 0.05 : health issue and being smoker are independet