# Simple Linear Regression in R

# (smoking and cancer)

This exercise uses file smoking\_cancer.csv.

## Tasks to be performed:

1. Read the data into R.
2. Form a H0.
3. Form a H1.
4. Explore your dataset.
5. View the data in a scatter plot.
6. Calculate correlation coefficient and comment on the results.
7. Use the linear model function in R to build a linear model.
8. Display the best fit line.
9. Summarise your model and comment on the results.

Note that dataset shows total number of people who smoked (second column) and total number of people who got cancer (third column) for different occupations (first column). Sometimes the number of people who got cancer is higher than number of people who smoked.

So, there is a chance that smoking is not linked to cancer. In addition to that, there is no record what type of cancer it was (we cannot assume it was only lungs cancer). Maybe they got cancer because of the dangerous substances they were dealing with at work?

And yet, we can apply statistical methods to this dataset.