**Linear regression: Predict the amount of insurance claim given the number of claims**

**Data Source:** Auto Insurance in Sweden

X = number of claims, Y = total payment for all the claims in thousands of Swedish Kronor for geographical zones in Sweden

Reference: Swedish Committee on Analysis of Risk Premium in Motor Insurance

## Read the data and import all necessary libraries

You can download the data set from the following link:

<https://s3.amazonaws.com/acadgildsite/wordpress_images/datasets/slr06/slr06.xls>

1. Collect X and Y
2. **The scikit-learn approach:**
3. Import libraries and tools
4. Cannot use Rank 1 matrix in scikit learn
5. Creating Model
6. Fitting training data
7. Y Prediction
8. Calculating RMSE and Score

**Multiple Linear Regression**

We will use a student score dataset in this case study. In this particular dataset, we have math, reading and writing exam scores of 1000 students. We will try to predict the score of a writing exam from math and reading scores. Thus, we have 2 features (input variables). Let us first start by importing the dataset.

## Read the data and import all necessary libraries

You can download the data set from the following link:

<https://drive.google.com/drive/u/0/folders/192X4XJbfiRkiLSTvKYSxYMrjm5u1dBYs>

1. get scores to an array.
2. Plot the scores as scatter plot
3. Generate our X, Y and β.
4. X and Y
5. **Model Initialization**
6. **Data Fitting**
7. **Y Prediction**

1. **Model Evaluation**