# Analyze the report of Swedish Motor Insurance

*Business Analytic Foundation with R Tools- Question*

# Question

The data gives the details of third party motor insurance claims in Sweden for the year 1977. In Sweden, all motor insurance companies apply identical risk arguments to classify customers, and thus their portfolios and their claims statistics can be combined. The data were compiled by a Swedish Committee on the Analysis of Risk Premium in Motor Insurance. The Committee was asked to look into the problem of analyzing the real influence on the claims of the risk arguments and to compare this structure with the actual tariff.

The insurance dataset holds 7 variables and the description of these variables are given below:

| Variable | Description |
| --- | --- |
| Kilometers | Kilometers travelled per year |
| | 1: < 1000 |
| | 2: 1000-15000 |
| | 3: 15000-20000 |
| | 4: 20000-25000 |
| | 5: > 25000 |
| | |
| Zone | Geographical zone |
| | 1: Stockholm, Göteborg, and Malmö with surroundings |
| | 2: Other large cities with surroundings |
| | 3: Smaller cities with surroundings in southern Sweden |
| | 4: Rural areas in southern Sweden |
| | 5: Smaller cities with surroundings in northern Sweden |
| | 6: Rural areas in northern Sweden |
| | 7: Gotland |

Bonus          No claims bonus; equal to the number of years, plus one, since the
               last claim

Make           1-8 represents eight different common car models. All other models
               are combined in class 9.

Insured        Number of insured in policy-years

Claims         Number of claims

Payment        Total value of payments in Skr (Swedish Krona)

---

To complete this project, you will require a strong understanding of the following concepts:

- Lesson 3: Correlation, summarizing, subset summary, and data visualization
- Lesson 4: Linear Regression

After understanding the data, you need to help the committee with the following by the use of the R tool:

- The committee is interested to know each field of the data collected through descriptive analysis to gain basic insights into the data set and to prepare for further analysis.
- The total value of payment by an insurance company is an important factor to be monitored. So the committee has decided to find whether this payment is related to number of claims and the number of insured policy years. They also want to visualize the results for better understanding.

- The committee wants to figure out the reasons for insurance payment increase and decrease. So they have decided to find whether distance, location, bonus, make, and insured amount or claims are affecting the payment or all or some of these are affecting it.
- The insurance company is planning to establish a new branch office, so they are interested to find at what location, kilometer, and bonus level their insured amount, claims, and payment get increased. (Hint: Aggregate Dataset)
- The committee wants to understand what affects their claim rates so as to decide the right premiums for a certain set of situations. Hence, they need to find whether the insured amount, zone, kilometer, bonus, or make affects the claim rates and to what extent.

The data can be downloaded from the URL mentioned below (Swedish motor insurance):

http://instruction.bus.wisc.edu/jfrees/jfreesbooks/Regression%20Modeling/BookWebDec2010/data.html

The total time provided to complete this task is 4 hours.