Spring 2022 **Lealand Morin**

University of Central Florida
College of Business

QMB 6912
Capstone Project in Business Analytics

Problem Set #3

A data engineer for several dealerships has gathered relevant and appropriate information, and organized a dataset concerning 9,861 sales involving a trade-in of a truck at nine dealerships (some having more than one location), around one-half of which were sold at retail, while the other one-half were sold at auction. These data are contained in the file `UsedTrucks.dat`, which is available on the course Webpage under Module 3. Each vehicle in the dataset is a row, while the columns correspond to the variables whose names and definitions are the following:

Variable	Definition
`type`	sale type (an integer); `Auction=1` versus `Retail=0`
`pauc`	price when sold at auction (an integer)
`pret`	price when sold retial (an integer)
`mileage`	odometer mileage (an integer)
`make`	make of vehicle (an integer)
`year`	model year of vehicle (an integer)
`damage`	an index of damage to vehicle; `1` little damage, `10` a lot (an integer)
`dealer`	dealer id (an integer)
`ror`	rate-of-return (a real number)
`cost`	net amount given to trade-in (an integer)

Download the file `UsedTrucks.dat`; load the data described above into R; calculate summary statistics for these data; finally, present these statistics in a LaTeX table. (You may want to use an R package, such as `xtable`, to automate the production of such a table from a data frame in R.)

Some of the variables above warrant some description. The first set of variables describe the characteristics of the vehicles. The `damage` feature variable was constructed by workers at the various dealership, and is really just an ordinal ranking, at best. The `dealer` variable is an integer between `1` and `9`. The `make` variable is an integer between `1` and `9`, where `1` and `2` denote Ford, while `3` denotes Chevrolet, `4` Dodge, `5` General Motors, `6` Toyota, `7` Nissan, and `8` Subaru, and `9` others. The `cost` variable is how much the dealer gave the buyer of a new vehicle for the truck given in trade, plus any taxes and fees that might be associated with the truck, rounded to the nearest integer.

The variables listed above also include the dependent variables. The `ror` variable is the natural logarithm of the ratio either the price garnered at auction `pauc` or the price earned from retail to the variable `cost`; it is a real number. Suppose a particular vehicle `cost` \$1,000 net in trade, and the vehicle sold \$1,100, then the rate-of-return `ror` would be $\log(1.1) = 0.0953101\ldots$, or about 9.5 percent.

Next, analyze the data in subsets, according to sale type calculating the summary statistics for each subset and presenting these statistics in a LaTeX table. Then continue analyzing the data according to other categorizations that you might find relevant for the analysis of this dataset. Present these statistics in LaTeX tables as well.

Finally, assemble your results into a LaTeX script building on the example in the folder named `assignment_03` in the course repository. Augment this with a shell script that runs your script for your analysis, generates the LaTeX tables and builds the `pdf` document. Submit your work by either uploading a zip file to Webcourses containing all the relevant files and folders or by pushing the files to a GitHub repository, in which case you need only submit the url to your repository. In either case, your software package should run from a single shell script and include a `README` file that describes the instructions for users.

Due Date: Thursday, 10 February 2022, before the beginning of class.