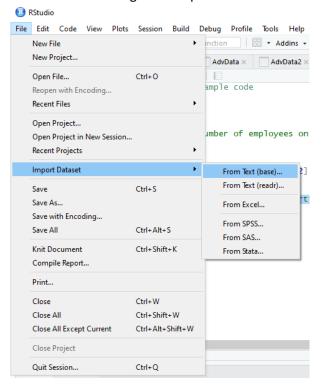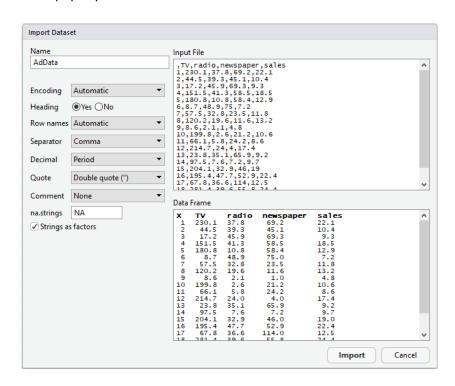# Computer lab 2 in MMS075, Jan 30, 2020

1. Download the advertising example in ISL from http://faculty.marshall.usc.edu/gareth-james/ISL/data.html and save it on the Desktop (i.e. in a computer-specific folder). Import the by using the menu in RStudio choosing File > Import Dataset > From Text (base)...

In this document, it will be assumed that the resulting dataset is called AdData – you can change the name to that (or another name of your preference) in the upper-left corner of the window that pops up:

2. Use "fix(AdData)" to open a simple editor within RStudio where you can make simple modifications of the data. Change the first letters of the column names to capital letters to test this function (and they will be referred this way henceforth, e.g. "Radio" instead of the original name "radio").

3. Construct the multiple linear regression model with Sales as response and TV, Radio and Newspaper as predictors, and call this model "AdModel". You do this by typing: "AdModel=lm(AdData$Sales~ AdData$TV+ AdData$Radio+ AdData$Newspaper)". Display a summary of the model using the "summary(AdData)" command. Now try typing "AdModel2=lm(AdData$Sales~ AdData$Newspaper+ AdData$TV+ AdData$Radio)" and display a summary of AdModel2. Do you see any difference?

4. It is not fun that we needed to type "AdData$" so many times to indicate every time that the variables belong to this data set. Try writing the command without that part, i.e. try typing: "AdModel=lm(Sales~TV+Radio+Newspaper)"
If you got an error message, type "attach(AdData)" and try the above command again. This time it should not give an error, because the attach command has made the column names of AdData directly available without having to specify that they are part of this data frame.

5. Answer some of the questions in the exercise from yesterday using the R software:
   a) Specify an approximate confidence interval for each coefficient.
   b) The management decides to spend 115000$ on TV advertisements and 40000$ on radio advertisements. How many sold units does the multiple linear regression model predict for these values?

   For this, you need to type "predict(AdModel,newdata)" where newdata represents the values where you want to make the prediction. Before using the "predict" command, you need to define newdata by typing "newdata=data.frame(TV=…, Radio=…, Newspaper=…)", where … is replaced by the values of predictors that you want to have the prediction for.

6. Just like in Lab 1, input the data underlying the example of completion time regressed on the number of employees (see the table below) and fit a simple linear model on these data.

Year	Employees in project	Completion time (days)
1	70	20
2	30	60
3	10	100
4	90	20

   Answer questions a) – b) below by using either the "predict" or the "predict.lm" function.

   a) Having 44 employees assigned to the project, what would be the expected completion time?
   b) How many employees need to be assigned to the project for a completion time of approximately one month?

Hint for part b): Define "newdata" to contain several different values to get a range of predictions. Example: "newdata=data.frame(Employees=1:100)" will contain all values from 1 to 100.

Hint 2: if you define the predictions to be part of the data frame "newdata", then they will be placed conveniently next to the values of the predictor and it is easy to get an overview of the predictor value – predicted response pairs. Example, assuming that the linear model defined above is called LinMod: "newdata$Predictions=predict(LinMod,newdata)".

- Does the scientific format of numbers disturb you? Type "options(scipen = 999)" to tell R to stop using that.
- Would you prefer fewer decimals displayed? Use the "round" function, for example like this: "round(predict(LinMod,newdata),2)"

7. Continue testing the code in Sections 2.3.1, 2.3.2 and 2.3.3 of the ISL book. These codes are also available online at http://faculty.marshall.usc.edu/gareth-james/ISL/code.html so that they can be copy-pasted into R, and reading the explanation in the book while copy-pasting ensures that you understand precisely what the commands do. Recall that the book can be downloaded for free at http://faculty.marshall.usc.edu/gareth-james/ISL/index.html.

8. If you would like to ask something or give feedback about the lecture, feel free to talk to me or enter your question/feedback at www.menti.com, using the code 22 00 02.