***Preliminaries***

The solutions to the assignments involving R coding should be reported in details, all R code that you write should be **included** into the report. Also, all appropriate pictures or diagrams should be included.

The data you are going to continue analyzing is the database of home prices in Albuquerque , 1993.

The variables present are Price; SqFt – the area of a house; FEATS – number of features such as dishwasher, refrigerator and so on; Taxes – annual taxes paid for the house. Explore the file ***prices1.xls***

***Assignment 1: Model selection***

1. Import necessary information to R
2. Plot three graphs showing respective dependences of the Price on SqFt, FEATS, and Taxes. If you would need to choose the best explanatory variable among these three, but only one, which one would you choose and why?
3. Perform model selection by using leave-one-out Cross-Validation among three univariate linear models where Price is the outcome and SqFt, FEATS, and Taxes are predictors. Which predictor was selected? Compare the prediction-sum-of-squares (PRESS) for these models and make necessary conclusions(**Hint**: for cross-validation, use function *cv.glm()*)
4. Make the same analysis as in step 3 but with leave-K-out Cross-Validationfor different K.
5. Plot the dependence of CV on the number K for each predictor used. Conclusions?

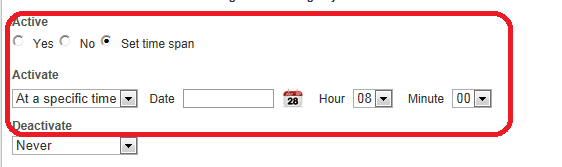
***Assignment 2: Bootstrap***

Use the same data set as in assignment 1.

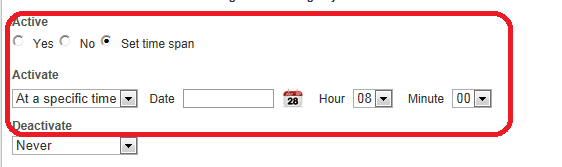
1. Plot the histogram of Price. Does it remind any conventional distribution? Compute the mean price.
2. Estimate the distribution of the mean price of the house using bootstrap. Determine the bootstrap bias-correction and the variance of the mean price. Compute the 95% confidence interval for the mean price using bootstrap percentile, bootstrap BCa, and first-order normal approximation (**Hint**: use *boot(),boot.ci(),plot.boot(),print.bootci()*)
3. Estimate the variance of the mean price using the jackknife and compare it with the bootstrap estimate
4. Compare the confidence intervals obtained with respect to their length and the location of the estimated mean in these intervals.
5. Consider linear model y=ax+b where y is Price and x is SqFt. Compute a confidence band using the bootstrap percentile. Plot the band obtained together with original data and fitted regression line. Comment on the shape of confidence bounds. Any other comments? (**Hint**: use *envelope(),plot(), points(), order()*)

***Submission procedure***

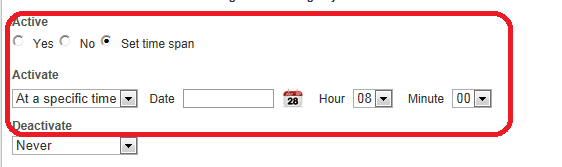
**If you are speaker,**

* Create a PowerPoint presentation describing your lab solutions. Submit your report and the PowerPoint file via “Submit answer” function of assignment ***Lab 5*** which is in the folder **Labs** before the deadline.
* Put your report also in the folder ***Lab 5 for uploading*** which is in the folder Labs and specify activation time (Specify Active : “Set time span”) to be the same as the deadline for the lab. **IMPORTANT: Do not forget to specify the activation time!**
* 
* Put your PowerPoint presentation in the folder ***Lab 5 for uploading*** which is in the folder Labs and specify activation time (Specify Active : “Set time span”) to be the same as the deadline for the lab. **IMPORTANT: Do not forget to specify the activation time!**
* Check folder ***Lab 5 for uploading***soon after deadline for the lab***,*** and find there the report of the classmate that you need to correct (see Couse Information🡪 Schedule for Labs and Seminars)
* Put your critical comments to the report of your classmate and submit the commented report via assignment ***Lab 5 student comments.***There is a deadline for this assignment!
* Afterwards, put the commented report into folder ***Lab 5 revised for uploading***

**If you are opponent,**

* Submit your report via “Submit answer” function of assignment ***Lab 5*** which is in the folder Labs before deadline.
* Put your report also in the folder ***Lab 5 for uploading*** which is in the folder Labs and specify activation time (Specify Active : “Set time span”) to be the same as the deadline for the lab. **IMPORTANT: Do not forget to specify the activation time!**
* 
* Check folder ***Lab 5 for uploading***soon after deadline for the lab***,*** and find there the report of the speaker (see Couse Information🡪 Schedule for Labs and Seminars) and his/her PowerPoint presentation.
* Put your critical comments to the report of the speaker. Create a Word or a PDF file called **Questions** containing at least 4 critical comments/ questions to the PowerPoint file of the speaker (primarily) and his/her lab report .
* Put the commented report into folder ***Lab 5 revised for uploading* at latest 24 hours after the deadline specified in “*Lab 5”* assignment at It’s learning** .
* Submit the commented report and the file **Questions** via assignment ***Lab 5 student comments.***There is a deadline for this assignment!

**If you are neither speaker nor opponent,**

* Submit your report via “Submit answer” function of assignment ***Lab 5*** which is in the folder Labs before deadline
* Put your report also in the folder ***Lab 5 for uploading*** which is in the folder Labs and specify activation time (Specify Active : “Set time span”) to be the same as the deadline for the lab. **IMPORTANT: Do not forget to specify the activation time!**
* 
* Check folder ***Lab 5 for uploading***soon after deadline for the lab***,*** and find there the report of the classmate that you need to correct (see Couse Information🡪 Schedule for Labs and Seminars)
* Put your critical comments to the report of your classmate and submit the commented report via assignment ***Lab 5 student comments.***There is a deadline for this assignment!
* Afterwards, put the commented report into folder ***Lab 5 revised for uploading***