***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 dependence 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 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 bootstrap bias-correction and 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. Compare the confidence intervals obtained with respect to their length and the location of the estimated mean in these intervals.
4. Use linear model y=ax+b where y is Price and x is SqFt. Construct confidence bands using 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()*)

## Submission procedure

If you are neither speaker or opponent, you should just send-in the solutions via It’s learning. **Note that there is a deadline for submission!** If you are *speaker*, you must also do the following:

* Put your report and PowerPoint presentation to the folder *Lab 4* which is in the folder *Seminars*. Specify Active : “Set time span”. Specify activation time as **30 april 15:00. Make sure that the time settings are specified!**

If you are *opponent*, you may find the report and PPT-file for revision available immediately after **30 april 15:00** in the folder*Seminars->Lab 4*