1. We have a set of measured points:
   $x_1 = 2; y_1 = 0,5$
   $x_2 = 3; y_2 = 15$
   $x_3 = 4; y_3 = 2$
   $x_4 = 6; y_4 = 6,5$

Enter the points into the Excel (or other tool) table and make a graph of them.

- Is there an outlier (gross error) in this set of points? If yes, which point is it?
- Remove the outlier and continue with the point set without outlier.
- Calculate the slope (b1) and the intercept (b0) of the linear equation that you fit through these points (use linear regression).
- Calculate the Pearson squared correlation coefficient R^2.

2. We have the following table:

	Molecule name	**pKa**	**Charge on the atom (q)**		
			H	**O**	**C**
Training set	Carboxyacetic acid	2.85	0.48	-0.6845	0.5834
	Hydroxyethanoic acid	3.83	0.4649	-0.694	0.5179
	Dipropylacetic acid	4.6	0.3907	-0.7486	0.439
	n-Butanoic acid	4.82	0.4187	-0.727	0.4915
	n-Dodecanoic acid	5.3	0.396	-0.7433	0.473
Test set	Almond acid	3.41	0.4371	-0.706	0.464
	Amber acid	4.21	0.4628	-0.6924	0.524
	n-Capric acid	4.9	0.3991	-0.7408	0.475

- For a QSPR model: pka = p1***qH** + p2, create a graph of dependency between pKa and qH (use Excel or other tool). Use only a training set for creating of the model.
- Compute p1 and p2 for this model.
- Use the model for prediction of pKa for all the molecules. (Add a column pka_p into the table.)
- Compute R2 for a training set.
- Compute Q2 for a test set.

Domácí úkol:

- For QSPR model: pka = pp1*qH + pp2*qO + pp3*qC + pp4 compute pp1, pp2, pp3 and pp4. (Apply for example: http://home.ubalt.edu/ntsbarsh/business-stat/otherapplets/MultRgression.htm, http://www.wessa.net/rwasp_multipleregression.wasp).
- Predict pKa for all molecules using the model. (Add a column pka_p2 into the table.)
- Compute R2 for a training set.
- Compute Q2 for a test set.