1. Mějme množinu naměřených bodů:
   $x_1 = 2$; $y_1 = 0,5$
   $x_2 = 3$; $y_2 = 15$
   $x_3 = 4$; $y_3 = 2$
   $x_4 = 6$; $y_4 = 6,5$

Body si zadejte do Exceové tabulky a udělejte si jejich graf.

- Je v této množině bodů hrubá chyba (outlier)? Pokud ano, jaká a proč.
- Pracujte s množinou bodů, ze které jste odstranili hrubé chyby.
- Vypočítejte směrnici (b1) a úsek (b0) lineární rovnice, kterou proložíte těmito body (použijte lineárni regresi).
- Vypočítejte korelační index R^2.

2. Doplňte následující tabulku:

	Název molekuly	**pKa**	**Náboj na atomu**		
			H	**O**	**C**
Tréningová sada	Carboxyacetic acid	2.85	0.48		
	Hydroxyethanoic acid	3.83		-0.694	
	Dipropylacetic acid	4.6			0.439
	n-Butanoic acid	4.82		-0.727	
	n-Dodecanoic acid	5.3	0.396		
Testovací sada	Almond acid	3.41		-0.706	
	Amber acid	4.21			0.524
	n-Capric acid	4.9			0.475

Poznámka: V tabulce jsou karboxylové kyseliny, náboje zjišťujeme na COOH skupině. O označuje kyslík, na kterém je vázán H. Struktury molekul získáme z PubChemu. Náboje počítejte pomocí programu ACC2 (https://acc2.ncbr.muni.cz/), použijte defaultní nastavení (= nahrajte molekulu a dejte “Compute charges”).

- Pro QSPR model: pka = p1*qH + p2 vytvořte v Excelu graf závislosti pKa na pH. Pro vytvoření modelu použijte jen tréningovou sadu.
- Pro tento model dopočítejte p1 a p2.
- Pomocí modelu predikujte pKa pro všechny molekuly. (Přidejte si do tabulky sloupec pka_p.)
- Vypočítejte relativní odchylku pro všechny body. (Přidejte si do tabulky sloupec pka_d.)
- Vypočítejte R2, RMSD a průměrnou relativní odchylku pro tréninkovou sadu.
- Vypočítejte Q2, RMSD a průměrnou relativní odchylku pro testovací sadu.

Domácí úkol:

Pro QSPR model: pka = pp1*qH + pp2*qO + pp3*qC + pp4 dopočítejte pp1, pp2, pp3 a pp4. (Pomocí např: http://home.ubalt.edu/ntsbarsh/business-stat/otherapplets/MultRgression.htm, http://www.wessa.net/rwasp_multipleregression.wasp).
Pomocí modelu predikujte pKa pro všechny molekuly. (Přidejte si do tabulky sloupec pka_p2.)
Vypočítejte relativní odchylku pro všechny body. (Přidejte si do tabulky sloupec pka_d2.)
Vypočítejte R2, RMSD a průměrnou relativní odchylku.
Vypočítejte Q2, RMSD a průměrnou relativní odchylku pro testovací sadu.
Bonus navíc (za 5% navíc):-))>
Dokážete najít další nábojový deskriptor, který by model zlepšil? Pokud ano, který to je. Ukažte, jak vypadá nový QSPR model s tímto descriptorem.