

1. Mějme množinu naměřených bodů:

$$x_1 = 2; y_1 = 0,5$$

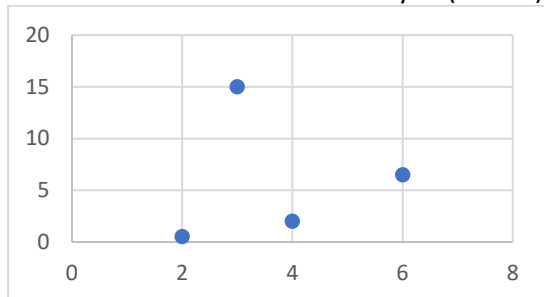
$$x_2 = 3; y_2 = 15$$

$$x_3 = 4; y_3 = 2$$

$$x_4 = 6; y_4 = 6,5$$

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

Je v této množině bodů hrubá chyba (outlier)? Pokud ano, jaká a proč. $x_2 = 3; y_2 = 15$, viz graf:



- Pracujte s množinou bodů, ze které jste odstranili hrubé chyby.
- Vypočítejte směrnici (b_1) a úsek (b_0) lineární rovnice, kterou proložíte těmito body (použijte lineární regresi).
 $b_1 = 1.5, b_0 = 3$
- Vypočítejte korelační index R^2 .
 $R^2 = 0.9231$

2. Doplněte následující tabulku:

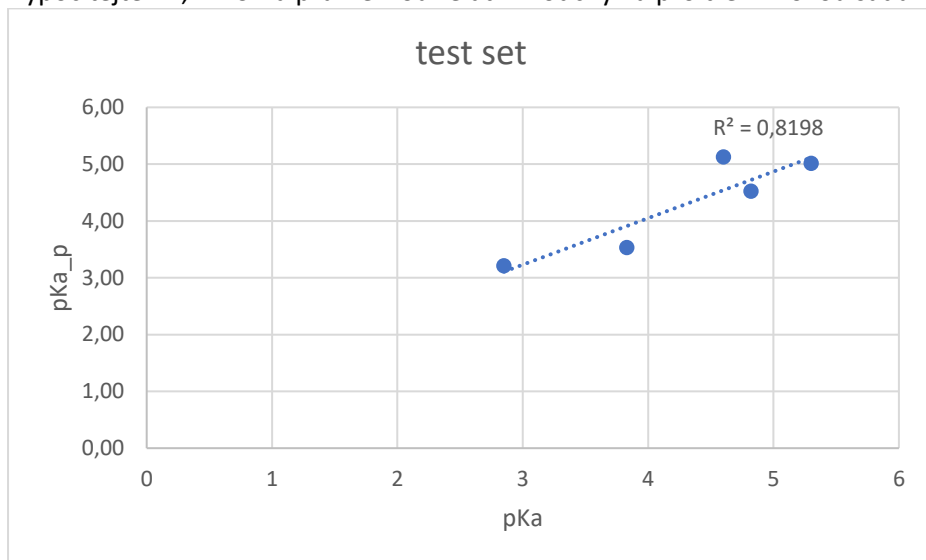
	Název molekuly	pKa	Náboj na atomu	Pka_p	pKa_d
			H		
Tréningová sada	Carboxyacetic acid	2.85	0.48	3.21	0.36
	Hydroxyethanoic acid	3.83	0.4649	3.53	0.30
	Dipropylacetic acid	4.6	0.3907	5.12	0.52
	n-Butanoic acid	4.82	0.4187	4.52	0.30
	n-Dodecanoic acid	5.3	0.396	5.01	0.29
Testovací sada	Almond acid	3.41	0.4371	4.13	0.72
	Amber acid	4.21	0.4628	3.58	0.63
	n-Capric acid	4.9	0.3991	4.94	0.04

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").

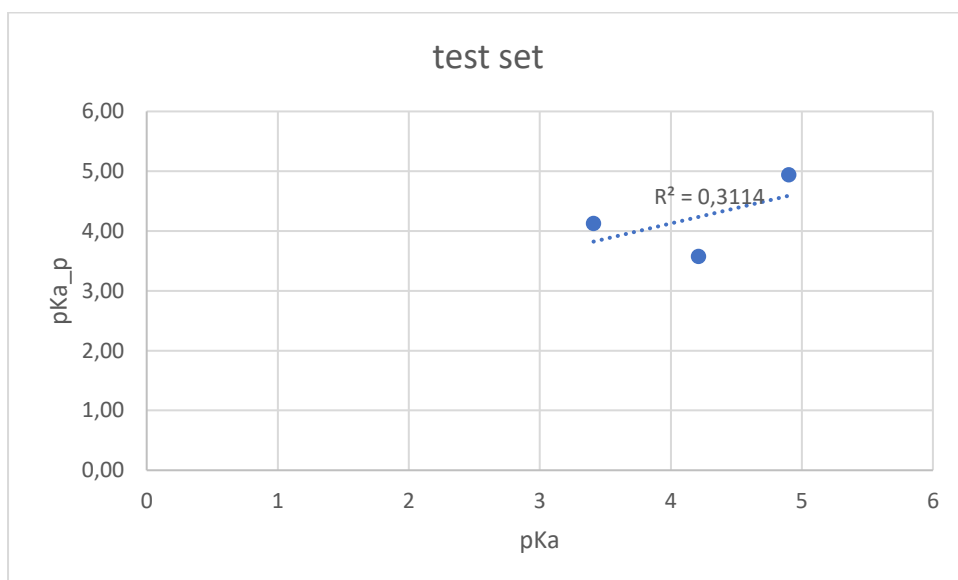
- Pro QSPR model: $pka = p_1 \cdot qH + p_2$ vytvořte v Excelu graf závislosti pKa na pH. Pro vytvoření modelu použijte jen tréningovou sadu.
- Pro tento model dopočítejte p_1 a p_2 . $p_1 = -21.473, p_2 = 13.514$
- Pomocí modelu predikujte pKa pro všechny molekuly. (Přidejte si do tabulky sloupec pka_p.)
- Vypočítejte relativní odchylku pro všechny body. (Přidejte si do tabulky sloupec pka_d.)

- Vypočítejte R2, RMSD a průměrnou relativní odchylku pro tréninkovou sadu. $R^2 = 0.8198$



RMSD = 0.163

- Vypočítejte Q2, RMSD a průměrnou relativní odchylku pro testovací sadu.



$Q^2 = 0.3114$

RMSD = 0.3196