1. Mějme množinu naměřených bodů:

x1 = 2; y1 = 0,5

x2 = 3; y2 = 15

x3 = 4; y3 = 2

x4 = 6; y4 = 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 R2.

1. Mějme následující tabulku:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Název molekuly** | **pKa** | **Náboj na atomu (q)** | | |
| **H** | **O** | **C** |
| Tréninková sada | 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 |
| Testova- cí sada | 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 |

* Pro QSPR model: pka = p1\***qH** + p2 vytvořte v Excelu graf závislosti pKa na qH. Pro vytvoření modelu použijte jen tréninkovou 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 a průměrnou relativní odchylku pro tréninkovou sadu.

**Domácí úkol:**

* Vypočítejte Q2 a průměrnou relativní odchylku pro testovací sadu.
* 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 a průměrnou relativní odchylku.
* Vypočítejte Q2 a průměrnou relativní odchylku pro testovací sadu.