

Lineární regrese

Cíl

Ověřit, jak lze použít dvourozměrný soubor dat tak, aby bylo možné určit hodnotu jedné z proměnných pro zadanou hodnotu druhé proměnné.

Vysvětlení úlohy

U mnoha dvourozměrných znázornění se vyskytuje vztah mezi dvěma množinami dat. Pokud například srovnáme známky studentů z testu z fyziky se známkami z testu z matematiky, bude zřejmé, že student s dobrými známkami ve fyzice má často dobré známky také v matematice. Mezi těmito známkami je patrně pozitivní vzájemný vztah (**pozitivní korelace**). Pokud známky každého studenta znázorníme do grafu, můžeme použít přímkou, která je nejvhodnější pro znázornění rozptylu známek, abychom odhadli hodnotu jedné známky druhou známou známkou.



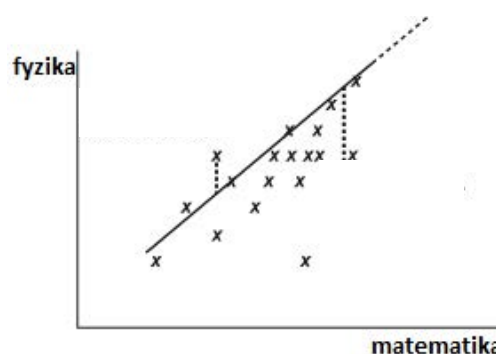
Pokud známky znázorníme, vznikne nám **korelační diagram**.

Přímka je graf lineární regrese. Pro výpočet její rovnice můžeme použít kalkulačku.

Lineární regresní přímka má funkční předpis:

$$y = a + b x.$$

Přímku v korelačním diagramu umístíme tak, že součet čtverců odchylek vůči rovnici přímky bude co nejmenší.



Tento postup je znám jako **metoda nejmenších čtverců**.

Tohoto postupu využívá také kalkulačka.

Pokud zadáme soubor dvourozměrných dat, můžeme použít kalkulačku pro výpočet hodnot **a** a **b**, které určí rovnici regresní přímky. Tak můžeme odhadnout hodnoty proměnné **y**, pokud známe hodnotu proměnné **x**.

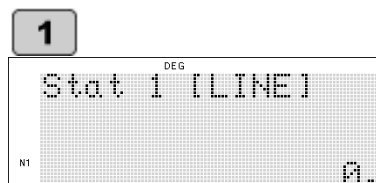
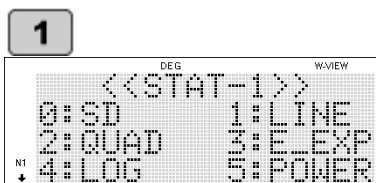
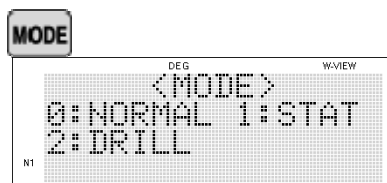
Podle charakteru rozložení bodů v korelačním diagramu můžeme určit, zda je mezi proměnnými přímá-pozitivní korelační závislost (obě proměnné stoupají) nebo nepřímá-negativní korelační závislost (jedna proměnná stoupá, druhá klesá). Míru lineární závislosti mezi proměnnými (korelace) pak vyjadřuje korelační koeficient, který může nabývat hodnot od -1 až po +1. Kalkulačka EL-W531 umí vypočítat korelační koeficient:

r pro množství dvoudimenzionálních dat.
Hodnotu **r** nastavíme třetí funkcí tlačítka dělení:





Využití kalkulačky

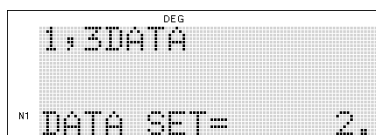
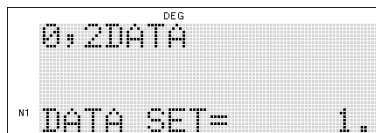
Na kalkulačce nastavíme režim statistika.



Kalkulačka je nyní nastavena na dvoudimenzionální zadávání hodnot.

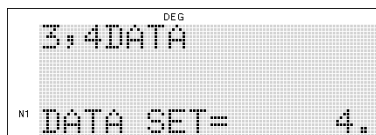
Každý pár p, q zadáme následovně:

p  klávesou
DATA
q  klávesou

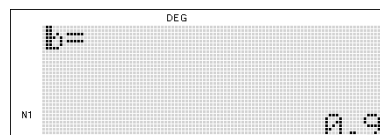
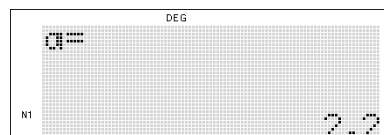
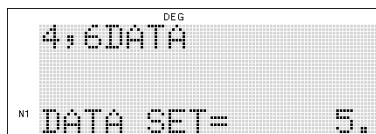
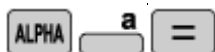


Nyní zadáme dvoudimenzionální hodnoty z následující tabulky do kalkulačky. Zobrazení na kalkulačce je vidět vpravo.

p	0	1	2	3	4
q	2	3	5	4	6



Po zadání hodnot můžeme pomocí kalkulačky nalézt hodnoty pro **a** a **b** v rovnici regresního stupně (??) pro $y = a + b x$. Pomocí klávesy Alpha můžeme zobrazit hodnoty pro **a** a **b**:



Vzorec pro lineární regresi je tedy $y = 2,2 + 0,9 x$.

Odhadované hodnoty y můžeme vypočítat dosazením daných hodnot x do rovnice.