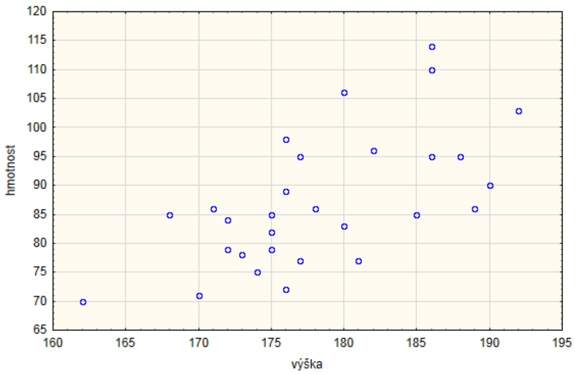
StatSoft

Novinky ve verzi 12 – Ortogonální proložení

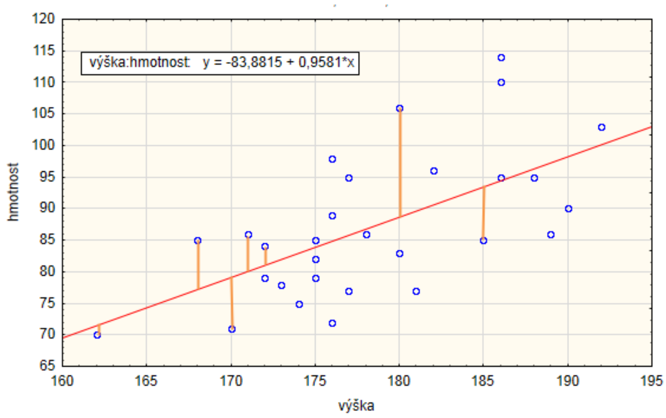
U 2D bodových grafů jsme se dočkali nového typu proložení, pojďme si jej ukázat.

Motivace

Na obrázku vpravo vidíte body, které představují výšku a hmotnost 30 jedinců.

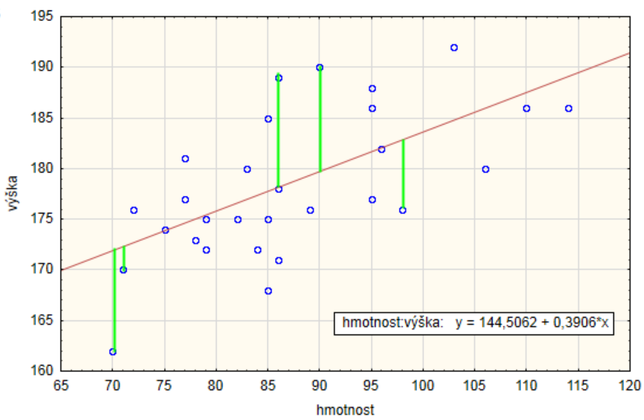


Pokud bychom chtěli vyšetřovat či modelovat jednu veličinu pomocí druhé, pravděpodobně bychom zkusili daty nejdříve proložit regresní přímku. Na dalším obrázku je vidět lineární závislost hmotnosti na výšce. Poněvadž jde o klasicky lineární regresní odhad, jedná se o metodu nejmenších čtverců. Ta se snaží najít přímku tak, aby součet kvadrátů vzdáleností bodů od přímky byl co nejmenší. Abychom si to představili, metoda vlastně sčítá kvadráty úseček naznačených na obrázku níže oranžově.

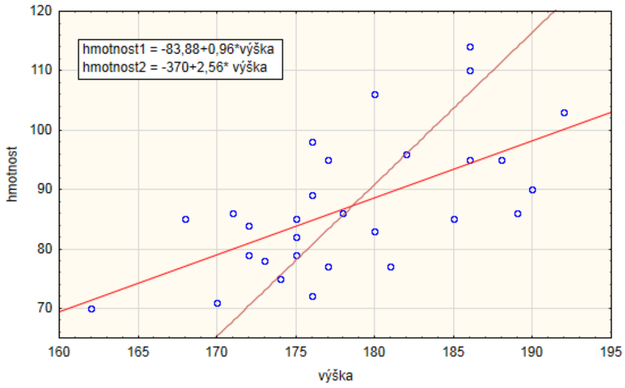


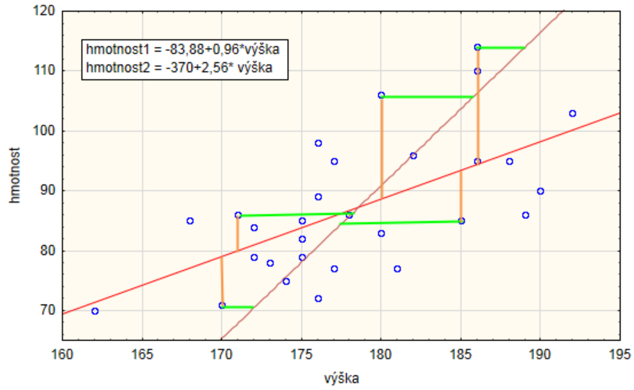
Toto je klasický nejběžněji používaný přístup.

Představme si nyní, že bychom chtěli z nějakého důvodu naopak vyšetřovat závislost hmotnosti na výšce. Pokud použijeme stejný postup jako před chvílí, dostaneme následující proložení a rovnici regresní přímky, zelené úsečky značí residua modelu, jejichž součet kvadrátů se snaží metoda minimalizovat.



Na první pohled se může zdát, že jsme na obou grafech získali stejné proložení. Nenechme se ale mýlit, stejné není! Pro porovnání přímek a důkaz, že opravdu stejné není, zkusme nakreslit obě do jednoho grafu (to znamená překlopit osy druhého grafu):

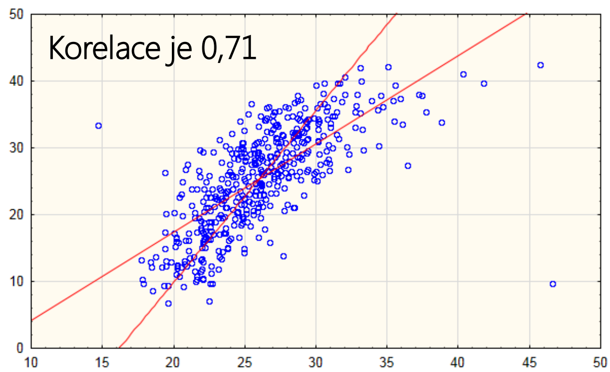


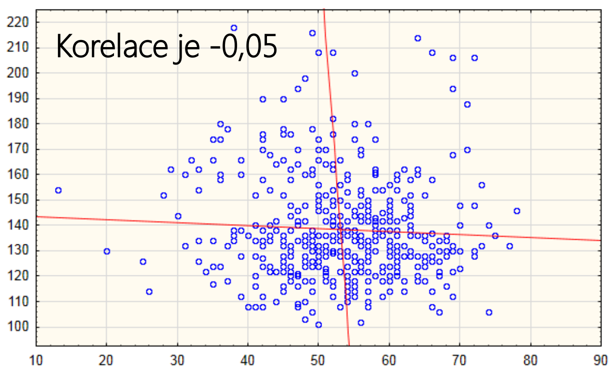


Přímky tedy opravdu stejné nejsou, není se také čemu divit, když se vlastně řeší úplně jiné úlohy – v jedné se minimalizuje součet kvadrátů horizontálních a ve druhé vertikálních vzdáleností od přímky.

Kdy jsou přímky stejné a jak to souvisí s korelačním koeficientem?

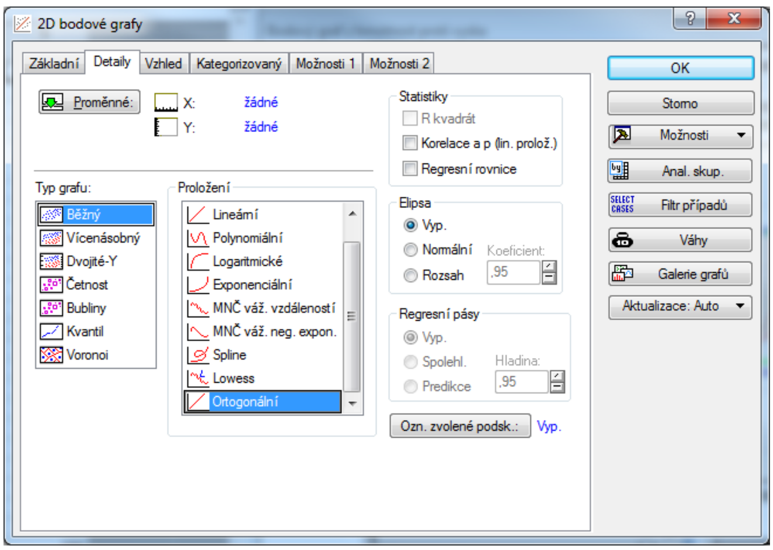
Dovolte nám malou odbočku. Pearsonův korelační koeficient je známá míra lineární závislosti. Věděli jste, že jedním ze zajímavých důsledků velikosti tohoto koeficientu je i vztah mezi proloženými přímkami? Pokud je korelační koeficient velký (v absolutní hodnotě), pak úhel mezi přímkami (jedna značí proložení Y na základě X a druhá X na základě Y) je tím bližší 0, čím je korelace silnější. Naopak úhel se blíží pravému s klesající velikostí korelačního koeficientu. Pokud je korelační koeficient roven 1 či -1, pak přímky splývají.





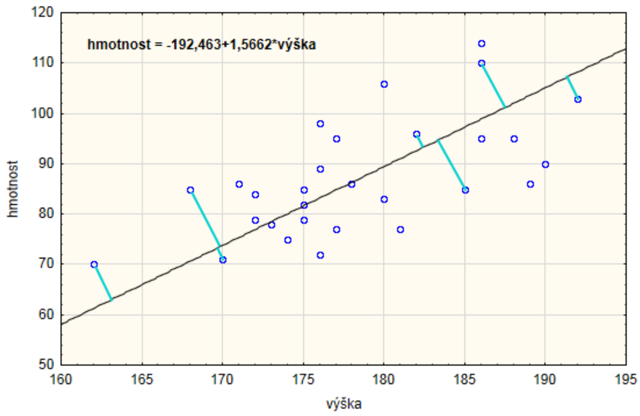
Ortogonální proložení

A teď již přejděme k novému proložení. Najdete jej v 2D bodových grafech v záložce *Detaily* pod názvem *Ortogonální* proložení:

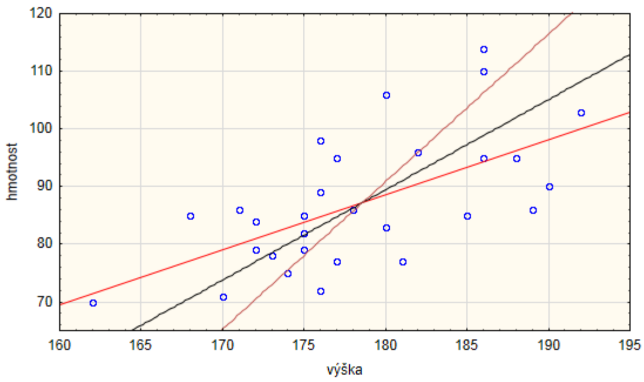


Tento typ proložení také prokládá daty přímku a vznikne minimalizací čtverců vzdáleností bodů od přímky, nicméně vzdálenosti jsou počítány jako délka kolmice k odhadnuté přímce (viz následující obrázek). Důsledkem toho je to, že ať už odhadujeme

X na základě Y nebo Y na základě X, odhadnuté přímky budou splývat.



Porovnání s klasickými proloženími, která jsme si ukazovali na začátku:



Metoda, která toto ortogonální proložení počítá, se jmenuje Total Least Squares a používá se místo klasické regrese v případě, kdy máme náhodné chyby jak v odezvě (tedy proměnné, kterou vyšetřujeme), tak v regresorech – jinak řečeno, kdy připouštíme chybu měření i v nezávislých proměnných. V klasickém lineárním regresním modelu se totiž předpokládá, že nezávislé proměnné jsou přesné a náhodná chyba je jen v závislé proměnné.