

Konstrukce regresního modelu

Úkol č. 1

Zadání: Máte k dispozici soubor s průměrnými ročními teplotami vzduchu ze stanice Nová Ves za období 1961-1990 (Nova_Ves.XLS), ve kterém však chybějí údaje pro 4 roky. S využitím údajů ze stanice (viz. cvičení č. 3) doplňte chybějící hodnoty.

Základní kroky zpracování

- 1) Vytvořte soubor se dvěma sloupci, které obsahují průměrné roční teploty vzduchu z obou stanic.
- 2) Výpočtem korelačního koeficientu a vykreslením bodového grafu ověřte, zda existuje statisticky významný vztah mezi teplotami vzduchu na obou stanicích.
- 3) Sestavte lineární regresní model závislosti teplot vzduchu na obou stanicích. Vytvořte graf této závislosti a dále graf reziduálních hodnot. Regresní model doplňte pásy spolehlivosti na hladině významnosti 95%. Přehledně shrňte a **interpretujte** základní statistiky popisující tento vztah (korelační koeficient, koeficient determinace, střední chyba odhadu, výsledek testování vhodnosti modelu)
- 4) S využitím sestavené závislosti doplňte chybějící hodnoty.
- 5) Vykreslete obě řady hodnot do spojnicového grafu (osa x roky, osa y teplota vzduchu)

Úkol č. 2:

Zadání: Z údajů Českého statistického úřadu porovnávajících kraje ČR (soubor **Uzemni_srovnani.XLS**) si zvolte jeden ukazatel, který budete studovat a který bude představovat vysvětlovanou proměnnou (Y). K ní ve stejné tabulce vyberte jednu vysvětlující proměnnou (X), u které lze z logiky věci předpokládat, že na ní bude proměnná Y závislá. Sestavte lineární regresní model, který bude kvantifikovat vztah hodnot X a Y. Tento vztah zhodnoťte.

(Možné příklady řešení: závisí v krajích ČR

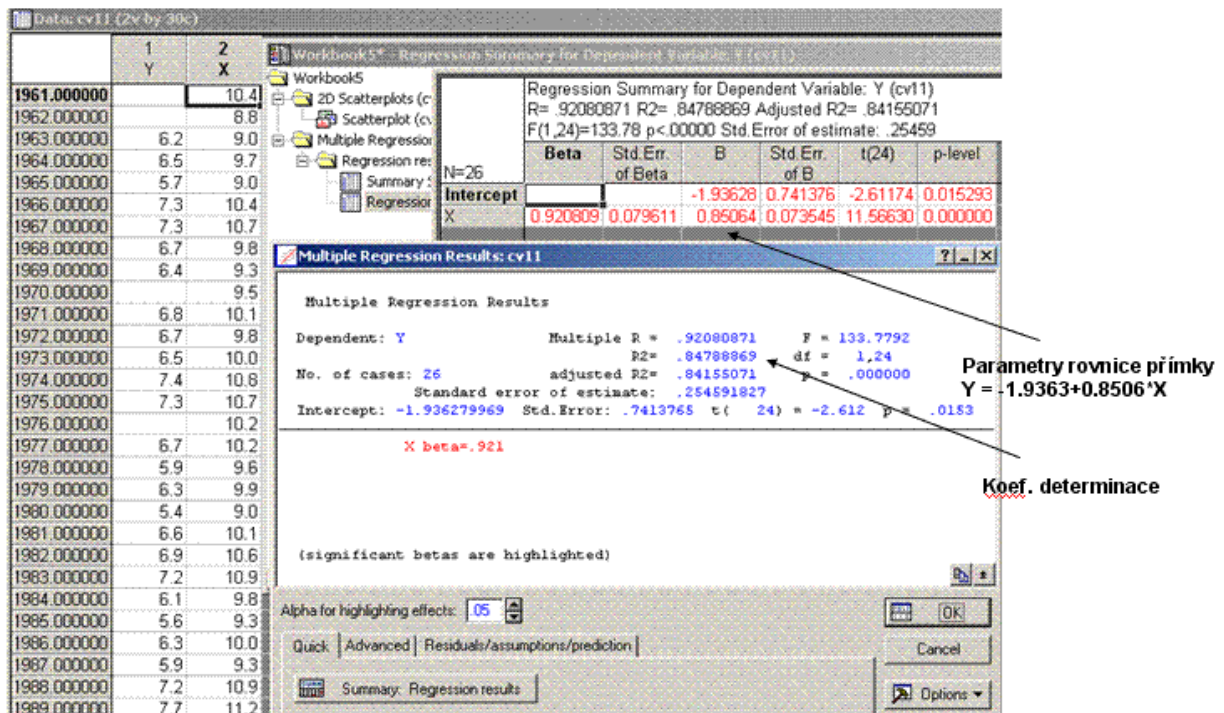
- počet zjištěných trestných činů na nezaměstnanosti?
- množství oxidu siřičitého na hrubém domácím produktu?
- množství oxidu siřičitého na počtu obyvatel?
- bytová výstavba na počtu obyvatel?
- atp.

Základní kroky zpracování

- 1) Z tabulky vhodně zvolte závisle a nezávisle proměnnou. Uložte do XLS souboru se dvěma sloupci obsahujícími hodnoty závisle a nezávisle proměnné.
- 2) Sestavte graf korelačního pole jako první přiblížení možné závislosti mezi studovanými charakteristikami.
- 3) Graf stručně charakterizujte, rozhodněte, zda budete dále pracovat se všemi jednotkami (kraji) resp. je nutné ze zpracování některý vyloučit a zdůvodněte proč (Praha ?)
- 4) Sestavte lineární regresní model. Vytvořte graf této závislosti, graf reziduálních hodnot. Regresní model doplňte pásy spolehlivosti na hladině významnosti 95%. Přehledně shrňte a **interpretujte** základní statistiky popisující tento vztah (korelační koeficient, koeficient determinace, střední chyba odhadu, výsledek testování vhodnosti modelu)

Jednotlivé kroky práce s regresním modelem pro úkol č. 1

1. Statistika – Vícerozměrná regrese (zadat závisle a nezávisle proměnnou – Základní výsledky – Výpočet: Výsledek regrese



2. Detailní výsledky – ANOVA (Celk. vhodnost modelu)

Analysis of Variance; DV: Y (cv11)

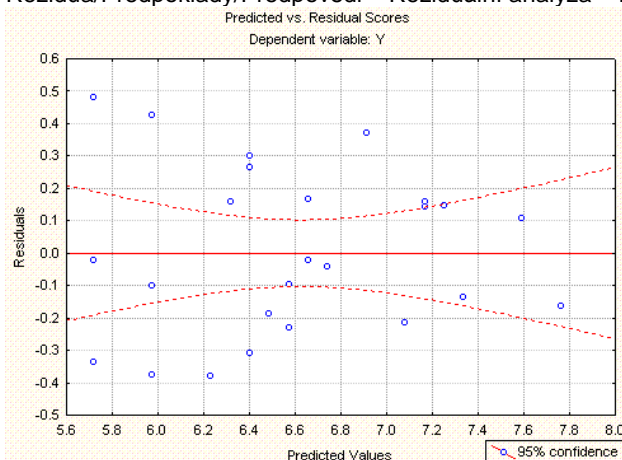
Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	8.67117	1	8.671166	133.7792	0.000000
Residual	1.55561	24	0.064817		
Total	10.22677				

ANOVA: Testujeme, zda se variabilita vysvětlená modelem významně liší od variability nevysvětlené (reziduální), H₀: neliší se.

Interpretace výsledku: P-hodnota (pravděpodobnost příslušející vypočtené hodnotě testovacího kritéria (F=133.77) je menší než alfa = 0,05 – tedy zamítáme nulovou hypotézu H₀. MODEL JE VHODNÝ

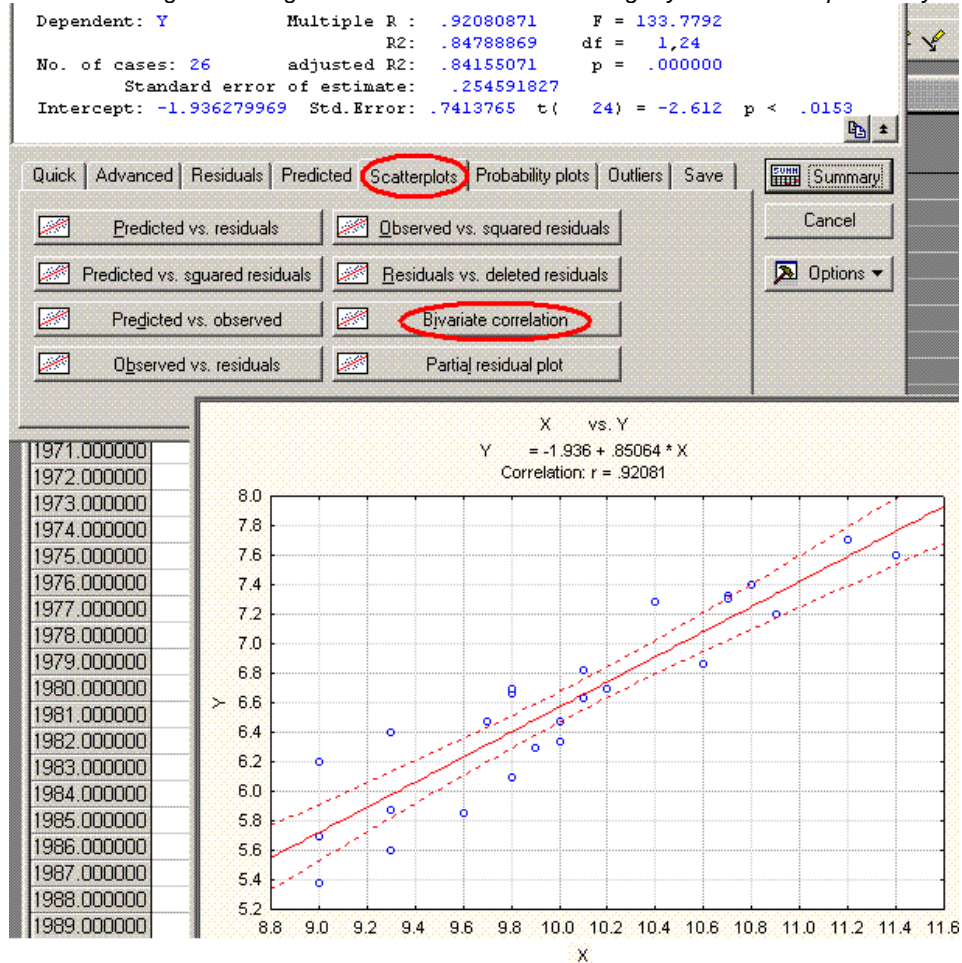
3. Analýza reziduálních hodnot

Rezidua/Předpoklady/Předpovědi – Reziduální analýza – Bodové grafy – předpovědi vs. rezidua



Interpretace: Pokud v tomto grafu hodnoty reziduí náhodně kolísají kolem nuly, použitý lineární model je vhodný. Pokud by rezidua např. postupně rostla či klesala resp. obecně by netvořila náhodný shluk bodů v grafu – potom by to ukazovalo na nutnost použití jiného regresního modelu než rovnice přímky.

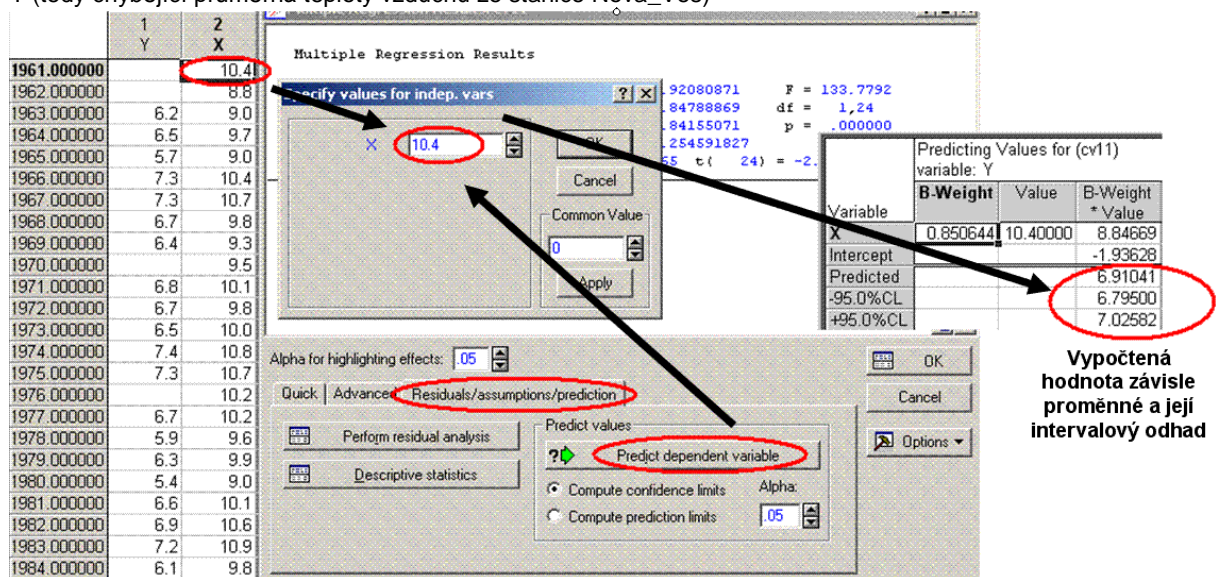
4. Konstrukce grafu lin. regresního modelu: OK – Bodové grafy – Korelace 2 proměnných



5. Výpočet chybějících měření s využitím regresního modelu

Rezidua/Předpoklady/Předpovědi – Předpovědi závisle proměnné

Pro zadanou hodnotu X (nezávisle proměnná) se podle sestaveného modelu vypočte hodnota závisle proměnné Y (tedy chybějící průměrná teploty vzduchu ze stanice Nova_Ves)



Postupně se zadávají hodnoty nezávisle proměnné X a pomocí modelu se vypočítají hodnoty závisle proměnné Y včetně hodnot intervalu spolehlivosti.