

## Konstrukce křivky teoretického rozdělení

*cvičení ze statistických metod č. 4*

Zadání:

- 1) Sestrojte křivku normálního rozdělení průměrných ročních teplot vzduchu na stanici Praha, Klementinum za období 120 let od ..... do ..... (data viz. cvičení 2).
- 2) Použijte  $\chi^2$  test a Kolmogorovův – Smirnovův test (K-S test) ke zhodnocení, zda existuje statisticky významný rozdíl mezi empirickými a teoretickými četnostmi průměrných ročních teplot vzduchu.
- 3) Vhodnost použití normálního rozdělení otestujte dále pomocí tzv. pravděpodobnostního grafu.
- 4) Ujistěte se, že pro řadu sum celkové výšky sněhové pokrývky na Milešovce za období 1925-1993 je použití normálního rozdělení nevhodné a z nabídky programu STATISTIKA zvolte jiný vhodnější typ teoretického rozdělení. Jeho vhodnost opět dokumentujte výsledkem  $\chi^2$  testu, K-S testu a pravděpodobnostním grafem.
- 5) Výsledky zpracování pro oba statistické soubory prezentujte dvěma typy grafů podle dodaného vzoru.

**Podkladová data:**

Soubory Klementinum\_cv2.XLS a Mil\_snih.XLS ve složce E:\Statistika\Cviceni\_4

**Poznámky ke zpracování:**

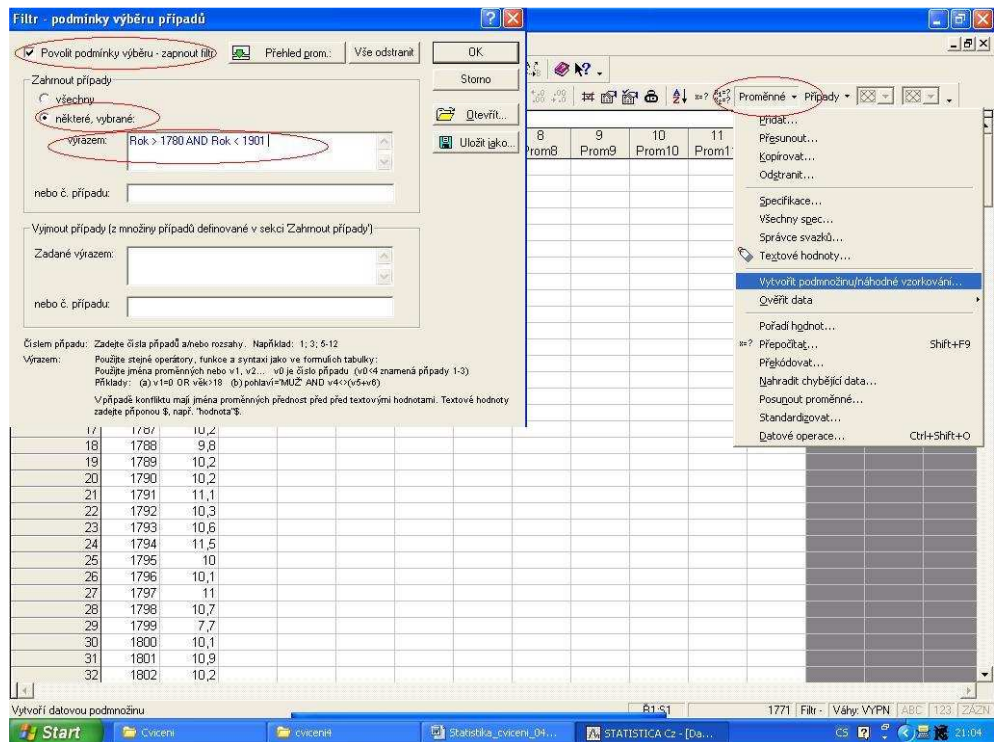
- 1) Spustíte program STATISTIKA
- 2) Proveďte import výše uvedeného souboru Klementinum.XLS: Soubor – Otevřít, zvolit Typ souboru \*.XLS a najít složku se jmenovaným souborem. Zatrhnout volbu „Importovat vybraný list do tabulky“. V dalším okénku zatrhněte „1. řádek jako názvy proměnných“,
- 3) Z celkového souboru vyberte vašich 120 hodnot (proměnných): Ve druhém pruhu nástrojů (ikon) vpravo klikněte Proměnné – Vytvořit podmnožinu/náhodné vzorkování. V políčku Proměnné vyberte Rok i I-XII. Dále klikněte na Případy. Do políčka „Zahrnout případy konkrétní/vybrané výrazem“ zadejte podmínku, např.:  
Rok > 1780 AND Rok < 1901 Pozor – všechna ostatní políčka musejí být prázdná! (Tato podmínka vybere ze souboru 120 hodnot počínaje rokem 1781) – viz. obr. 1  
Pomocí správce jmen případů převedte proměnou „ROK“ na jména případů a tuto proměnou odstraňte.
- 4) Sestrojení křivky normálního rozdělení: Zvolte „Statistika – Prokládání rozdělení – Normální“. V dalším okně klikněte na Proměnná a zvolte I-XII a klikněte OK. Pro výpočet K-S testu je nutné zatrhnout příslušnou položku v záložce Možnosti (Kolmogorov – Smirnovův test – Ano (spojitý) – viz.obr.2. Dále zvolte „Graf pozorovaného a očekávaného rozdělení“
- 5) Upravte jednotlivé prvky grafu podle dodaného vzoru (poklepáním a jednotlivé prvky grafu – osy, popisky, čáry,... je lze změnit). Dále vyberte graf příkazy Zobrazit – Umístění grafu. Zkopírujte graf do schránky příkazem Úpravy – Kopírovat. Vložte graf do dokumentu. V programu WORD zadejte: Úpravy – Vložit jinak – obrázek.
- 6) Vytvoření pravděpodobnostního grafu. V programu STATISTIKA zvolte: Grafy – 2D grafy – Grafy typu P-P. V dalším okně klikněte na Proměnné a zvolte I-XII a klikněte OK. Jako typ rozdělení je vybráno Normální. Upravte jednotlivé prvky grafu (osy x, y

- budou mít rozsah 0-1, úprava rozsahu os popsána v „úvodu do programu Statistica, viz. obr. 3) a vložte ho výše popsáním způsobem do dokumentu v programu WORD.
- 7) Proveďte import souboru Mil\_snih.XLS a to stejnými kroky jako v případě předchozího souboru (viz bod 2).
  - 8) Sestrojte křivku normálního rozdělení podle instrukcí v bodu 4. Z jakého důvodu je aplikace normálního rozdělení pro tento soubor nevhodná?
  - 9) Na liště ve spodní části obrazovky otevřete opět okno „Proložení spojitých rozdělení“. Otestujte jiná spojitá rozložení nabízená programem a jedno z nich vyberte. Vaši volbu ověřte sestrojením P-P grafu podle instrukcí v bodě 6.
  - 10) Oba grafy pro soubor Mil\_snih.XLS upravte a zkopírujte do dokumentu v programu WORD

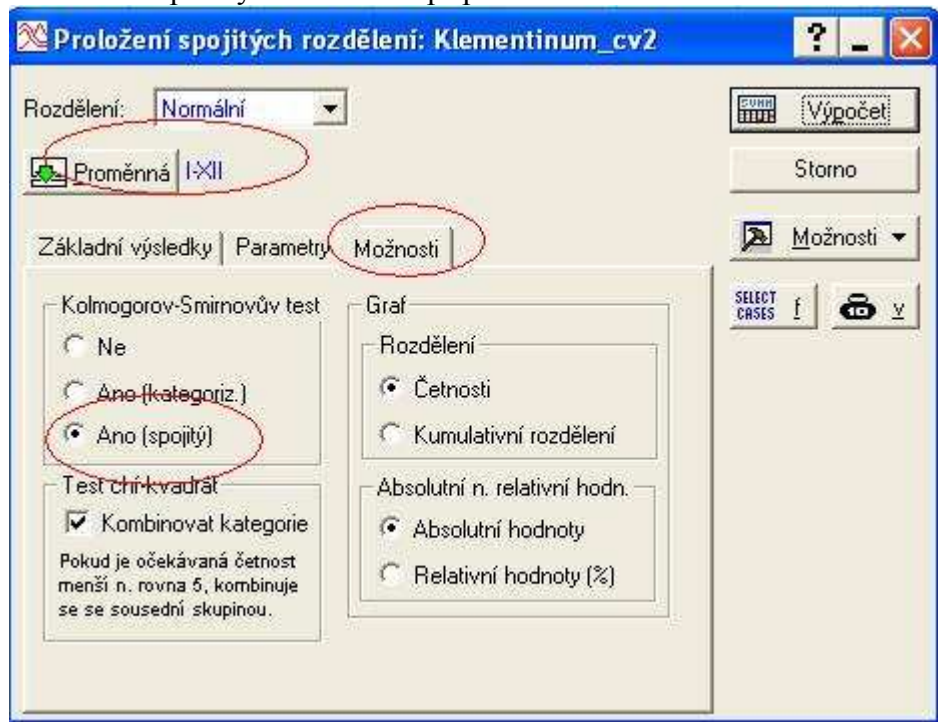
**Poznámka k interpretaci  $\chi^2$  testu:** Tímto testem se testuje shoda empirického a teoretického rozdělení. Testování je založeno na stanovení nulové hypotézy o existenci shody obou rozdělení. Pro interpretaci výsledků testu v programu STATISTIKA je vedle hodnoty testovacího kritéria rozhodující vypočtená hodnota hladiny významnosti  $p$ . Obecně používaná hladina významnosti testu je  $p=0,05$ . Vychází-li tedy  $p$ -hodnota podstatě větší než  $0,05$ , potom nemůžeme zamítnout nulovou hypotézu a můžeme prohlásit, že mezi empirickými a teoretickými četnostmi není statisticky významný rozdíl. Jinými slovy – použití normálního rozdělení je vhodné.

**Poznámka k interpretaci K-S testu:** Tento test lze použít pro testování významnosti shody teoretického a empirického rozložení i v případech, kdy nelze použít  $\chi^2$  testu. Hodnota testovacího kritéria u K-S testu však vychází pouze z jedné hodnoty – z maximálního rozdílu empirických kumulovaných četností. Proto není tak spolehlivým testem. Vhodnost či nevhodnost testovaného rozdělení posuzujte opět podle  $p$  hodnoty jako v případě  $\chi^2$  testu. Zkratka n.s. značí nevýznamný výsledek.

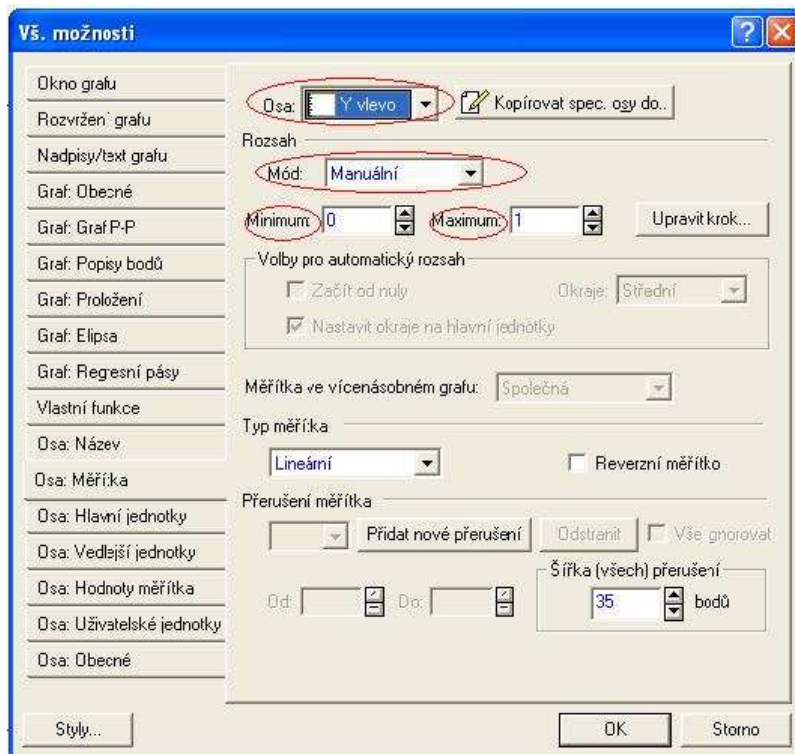
**Poznámka k interpretaci pravděpodobnostního grafu (P-P grafu):** V tzv. P-P grafu jsou vyneseny na ose  $y$  pravděpodobnosti výskytu empirických (původních, měřených) kumulativních hodnot studovaného znaku, na ose  $x$  potom pravděpodobnosti teoretických (vypočtených) kumulativních hodnot znaku. Pokud lze body grafu proložit přímkou, potom můžeme tvrdit, že uvažované teoretické rozdělení dobře aproximuje hodnoty studovaného souboru. Pokud přímkou netvoří, zvolené teoretické rozdělení není vhodné a je třeba hledat jiné.



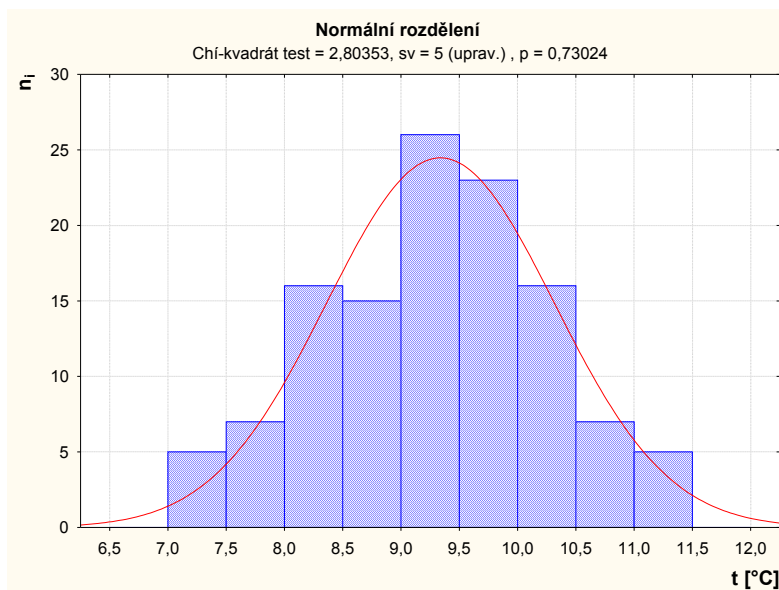
Obr. 1. Filtr pro výběr množství případů



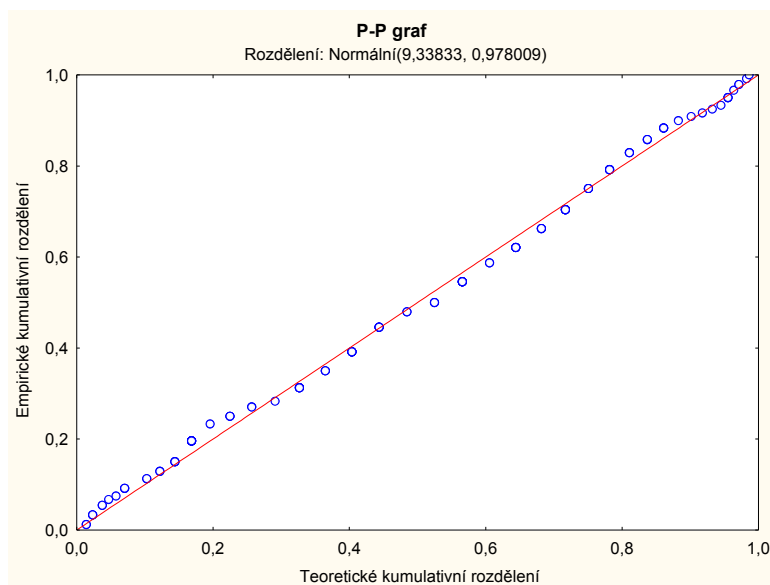
Obr. 2. Proložení spojitých rozdělení



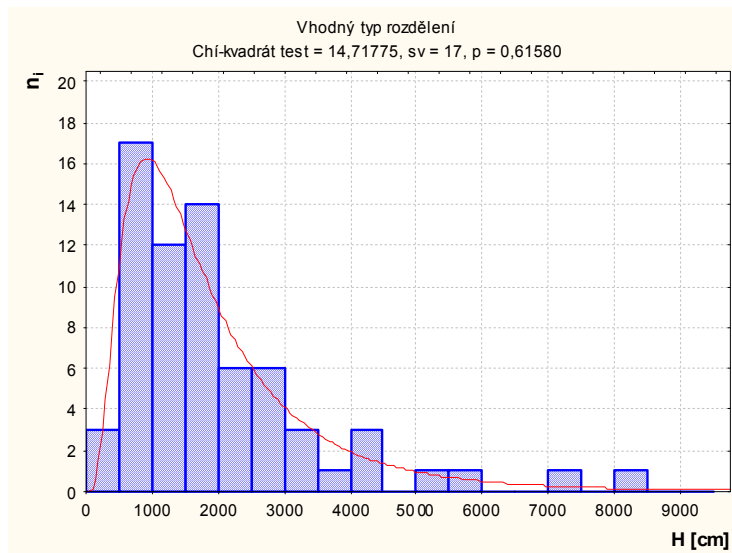
Obr. 3. Úprava rozmezí os x,y u P-P grafu



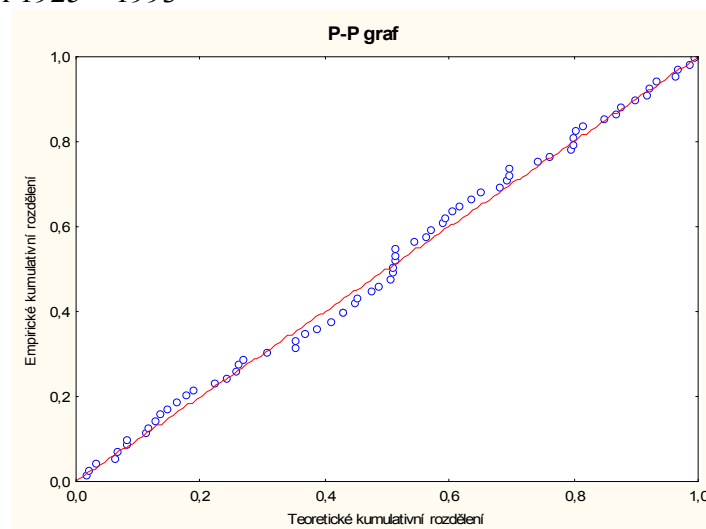
Obr. 4 Křivka normálního rozdělení průměrných ročních teplot vzduchu v Praze, Klementinu v období 1781-1900



Obr. 5 Pravděpodobnostní graf (P-P graf) normálního rozdělení průměrných ročních teplot vzduchu v Praze, Klementinu v období 1781-1900



Obr. 6 Křivka vhodného teoretického rozdělení celkové výšky sněhové pokrývky na Milešovce v období 1925 – 1993



Obr. 7 Pravděpodobnostní graf (P-P graf) vhodného teoretického rozdělení celkové výšky sněhové pokrývky na Milešovce v období 1925 – 1993