Normalita

Normální rozdělení a ověření normality dat

Opakování

- Jaké znáte typy dat?
  Uveďte příklady…

- Jak popisujeme data?

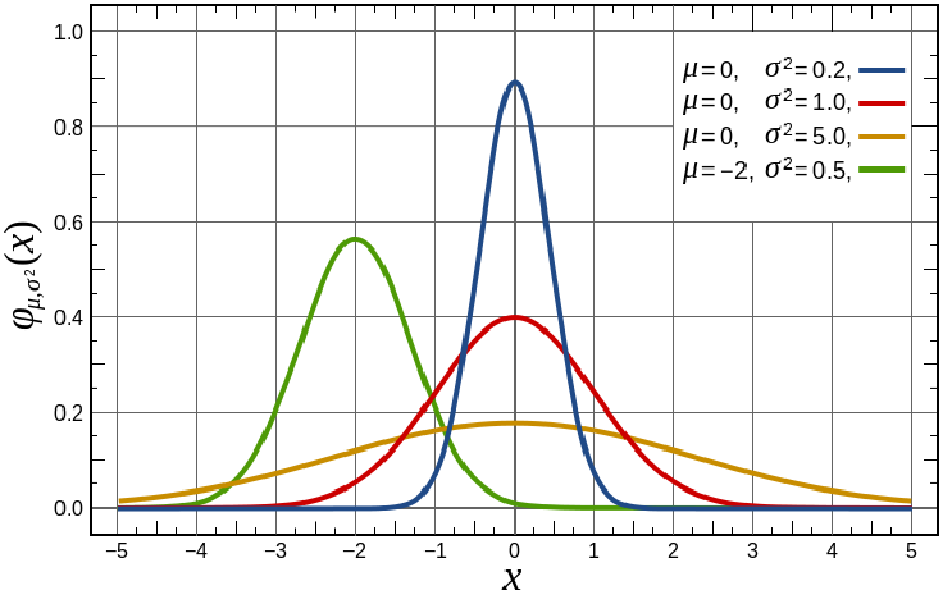
Normální rozdělení I

- Nejklasičtějším modelovým rozložením, od něhož je odvozena celá řada statistických analýz je tzv. normální rozložení, známé též jako **Gaussova křivka**.
- Popisuje rozdělení pravděpodobnosti spojité náhodné veličiny: např. výška v populaci, chyba měření…
- Je kompletně popsáno dvěma parametry:

  μ – střední hodnota

  σ^2 – rozptyl

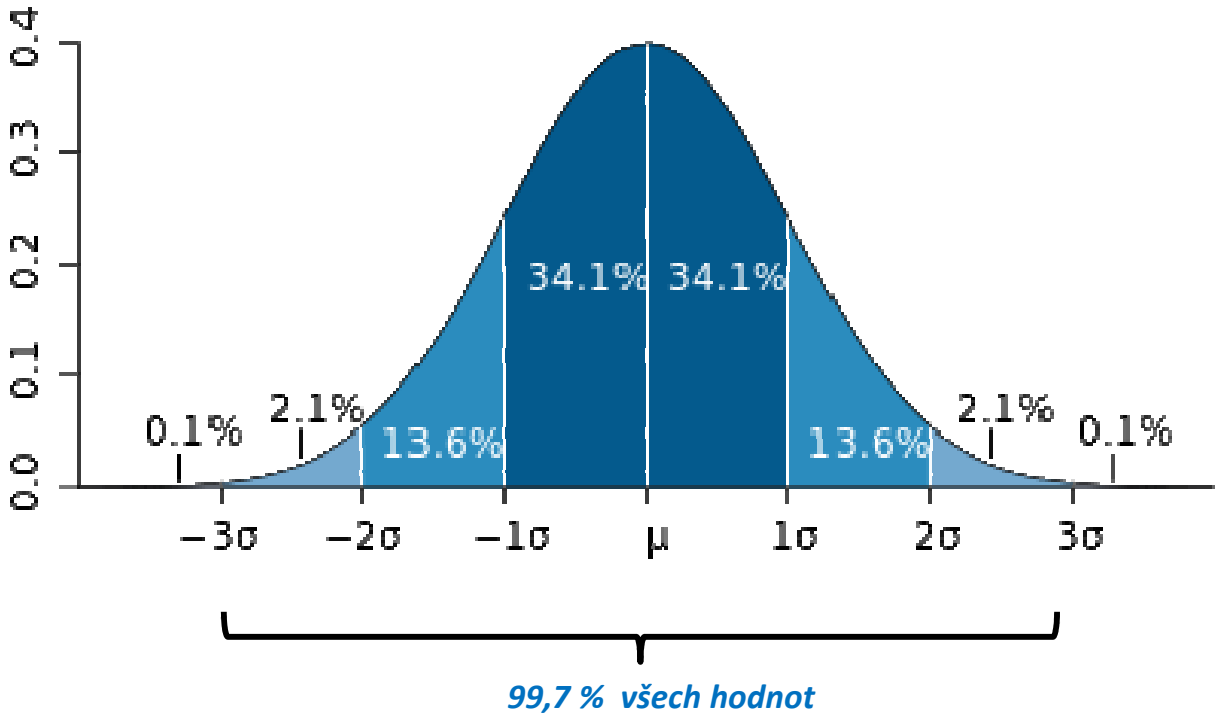
  Označení: **N(μ, σ^2)**



- Normalita je klíčovým předpokladem řady statistických metod
- Pro ověření normality existuje řada testů a grafických metod

Normální rozdělení II

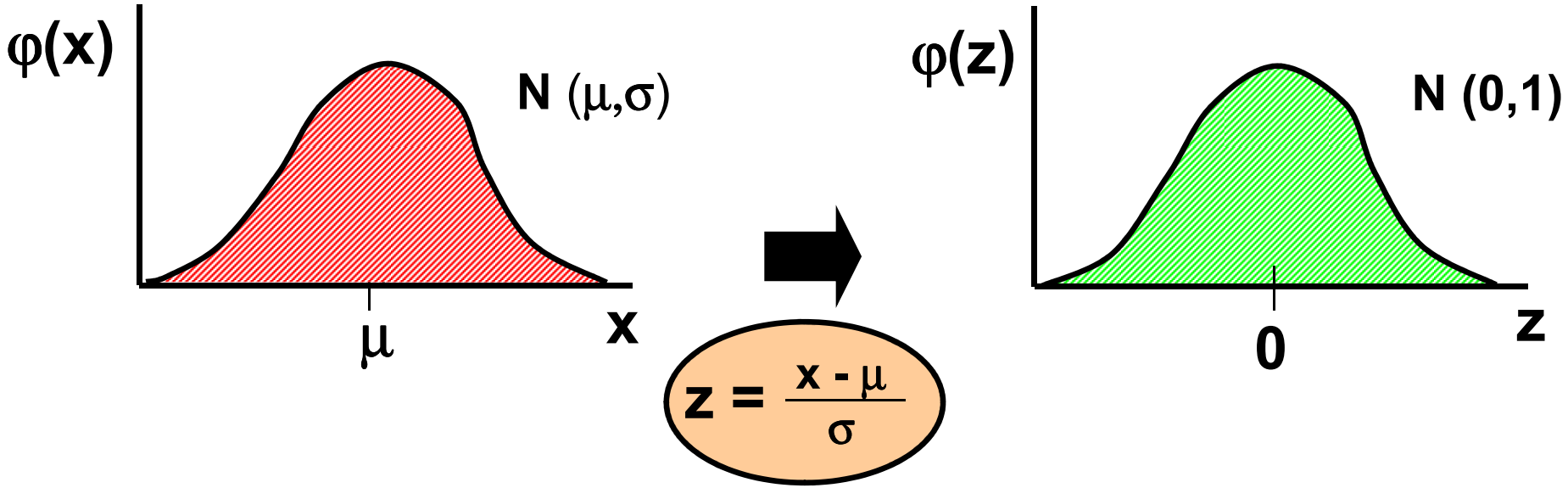
- V rozmezí μ ± 3σ by se mělo vyskytovat 99,7 % všech hodnot



- Použití: zhodnotíme tvar rozdělení a přítomnost odlehlých hodnot

Standardizované normální rozdělení

- Normální rozdělení se střední hodnotou nula a jednotkovým rozptylem



$$\varphi(x) = \frac{1}{\sigma.\sqrt{2\pi}}.e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Vzorec pro hustotu normálního rozdělení

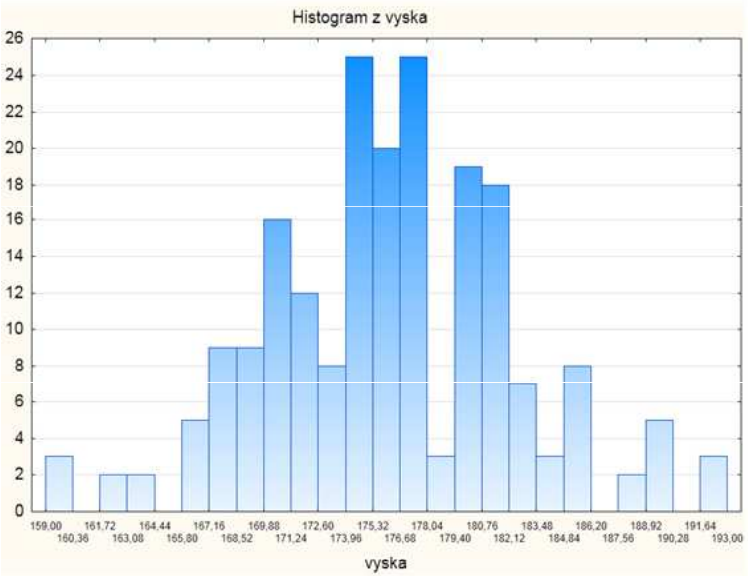
$$\varphi(z) = \frac{1}{\sqrt{2.\pi}}.e^{-\frac{z^2}{2}}$$

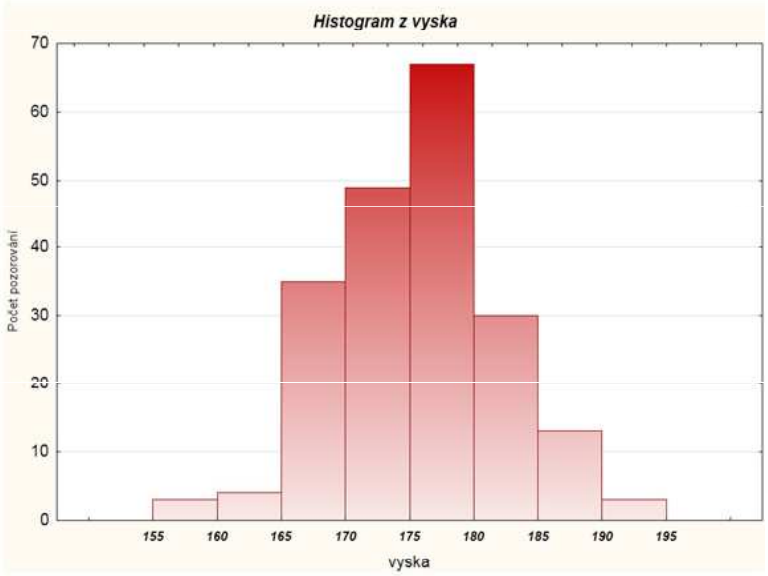
Vzorec pro hustotu standardizovaného normálního rozdělení

Tabelovaná podoba

Vizuální ověření normality

- Pro hodnocení tvaru rozložení lze využít histogram (nevýhoda: nutné určit „vhodný“ počet sloupců)





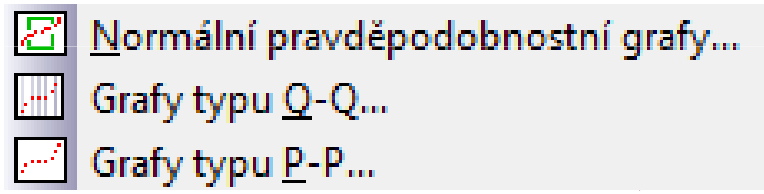
- Vhodnější jsou:
  1. Q-Q graf (kvantil-kvantilový graf)
  2. P-P graf (pravděpodobnostně-pravděpodobnostní graf)
  3. N-P graf (normální-pravděpodobnostní graf)

Opakování:
Co je kvantil?

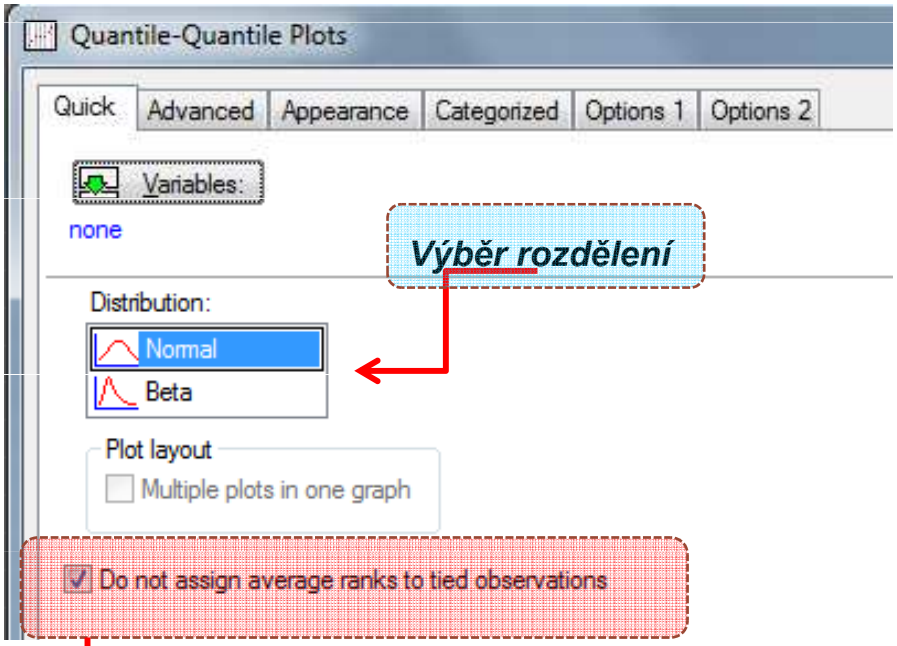
Řešení v softwaru Statistica I

• *V menu Graphs zvolíme 2D Graphs*

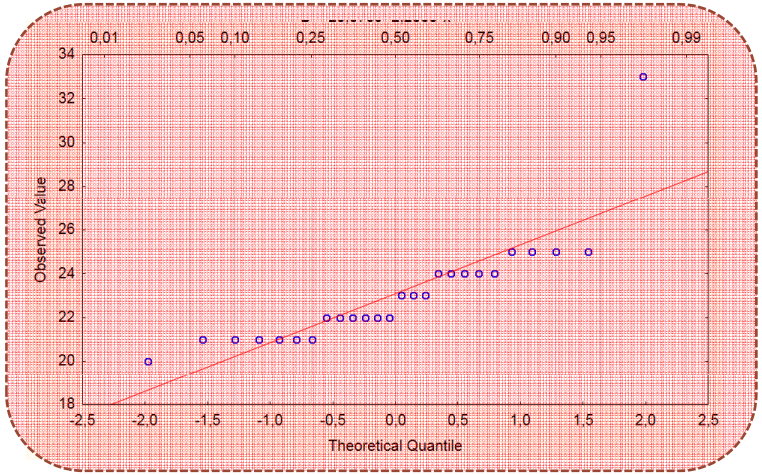
1



2

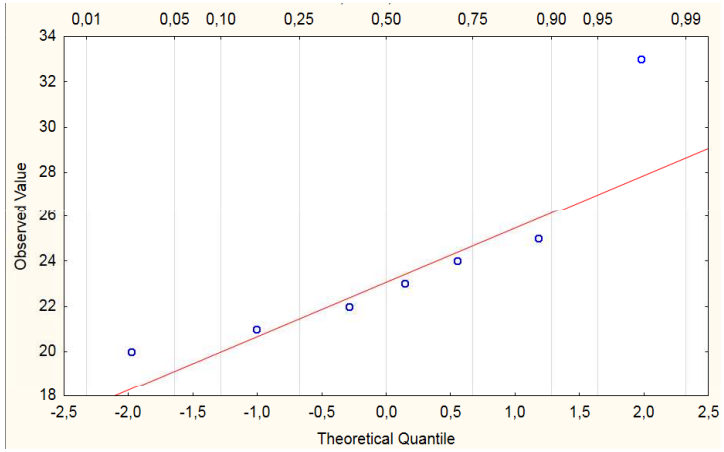


Výběr rozdělení

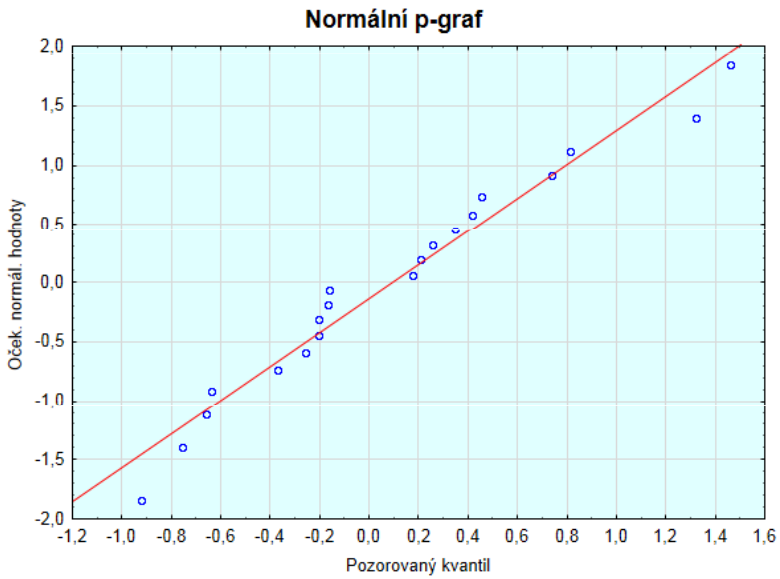


• *V případě, že máme v datech několik stejných hodnot, je vhodné odškrtnout Neurčovat průměrnou pozici svázaných pozorování*

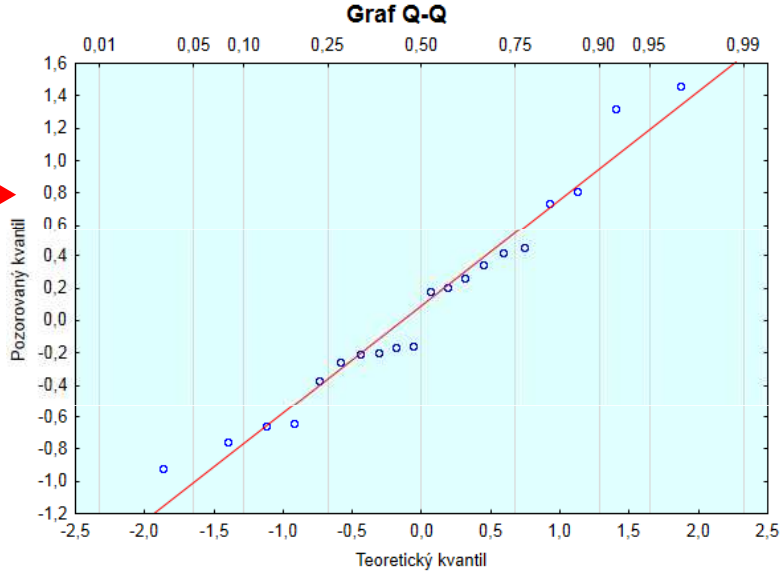
3



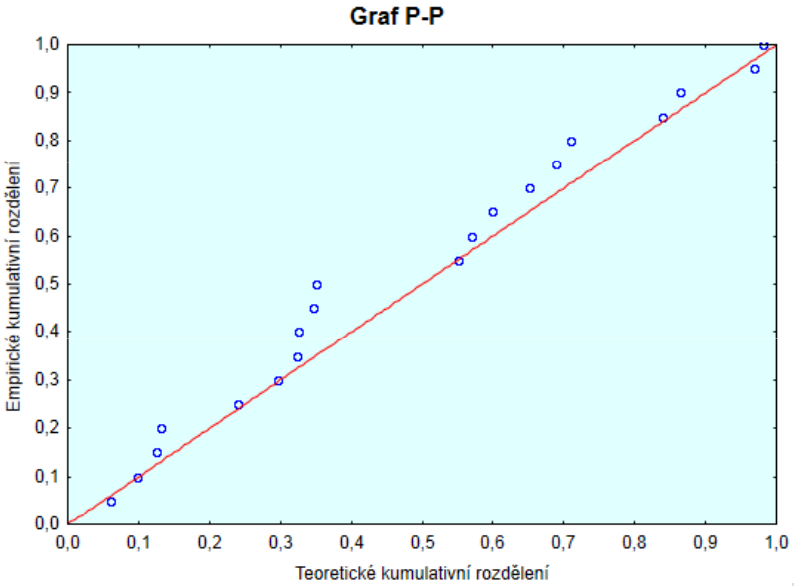
Rozdíl mezi N-P, Q-Q, P-P grafem



???



- ***Pouze výměna os***
- ***Znázorněn pozorovaný a teoretický kvantil***



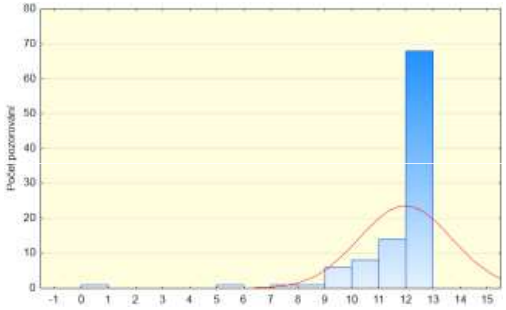
- ***Vykresleno kumulativní rozdělení***

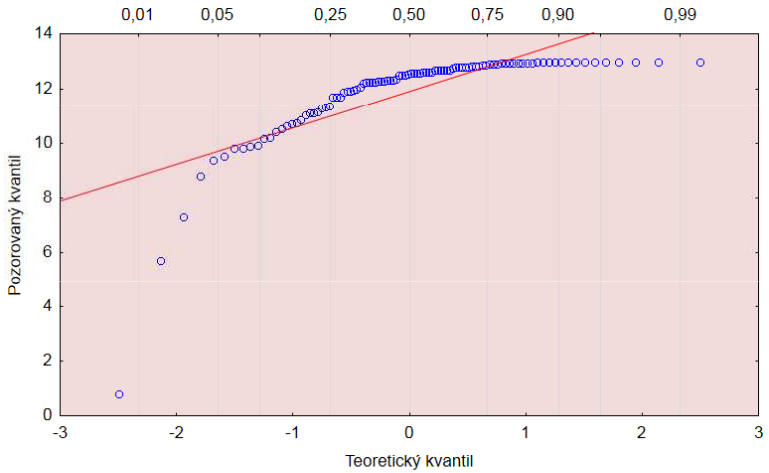
PAMATUJ:

Pocházejí-li data z normálního rozložení, pak body budou ležet okolo přímky

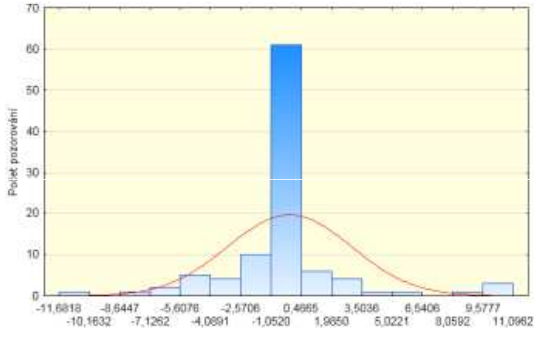
Vizualizace I.

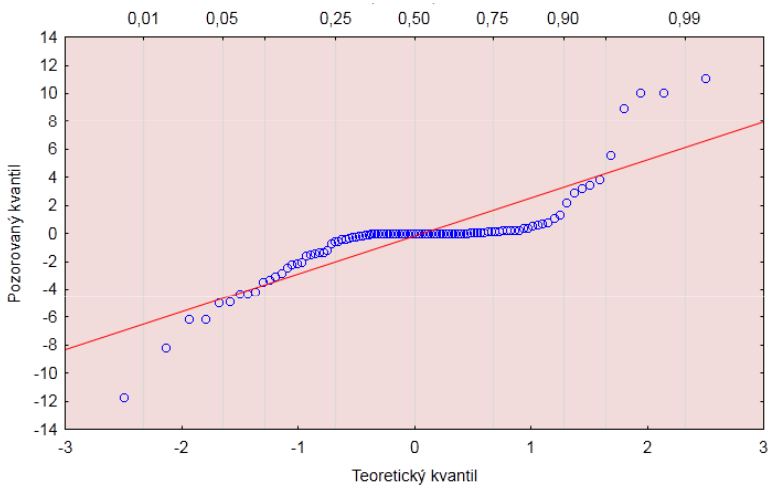
Zešikmená (nesymetrická) data





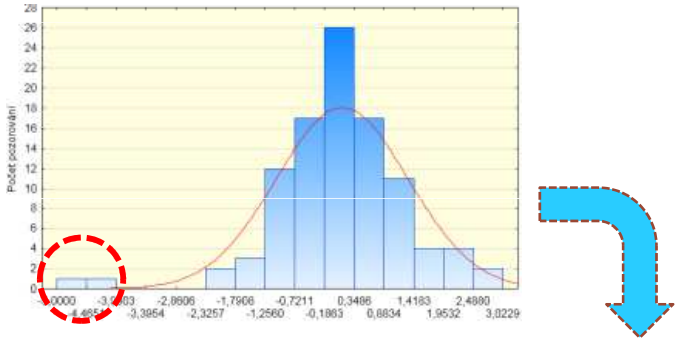
Špičatost

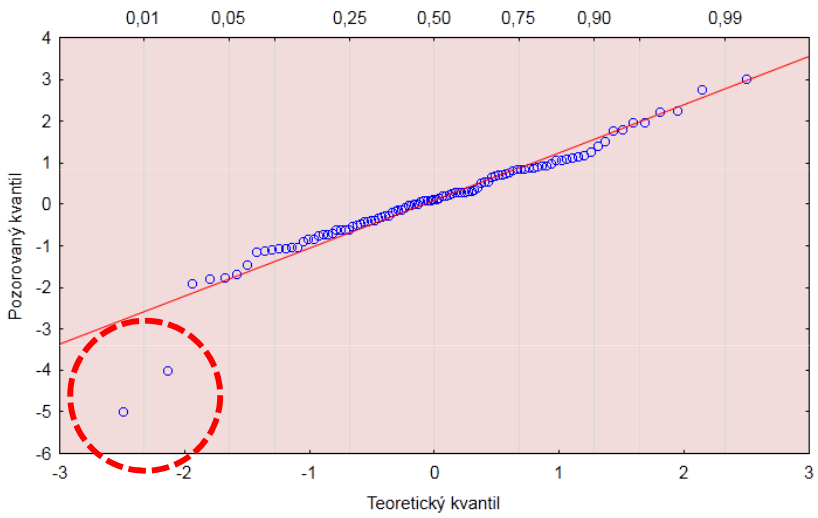
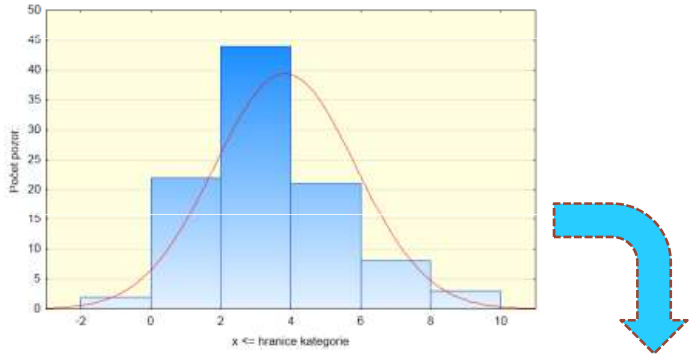
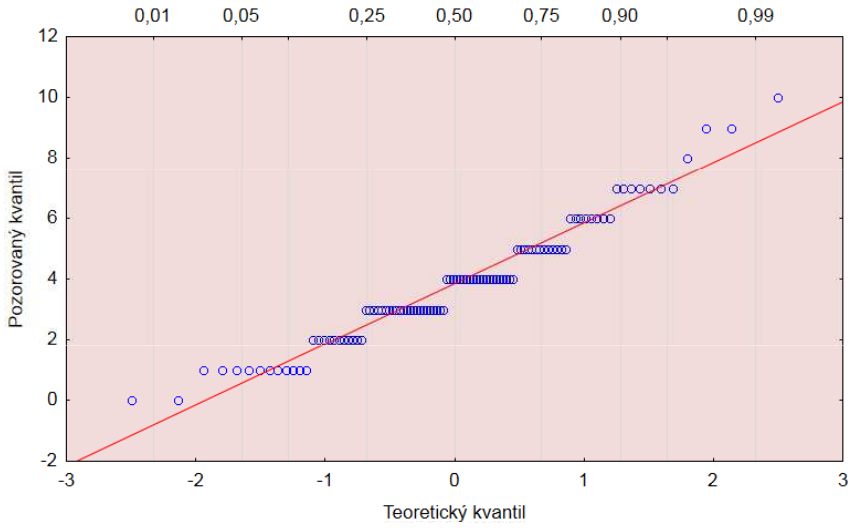




Vizualizace II.

Odlehlé hodnoty





Diskrétní rozdělení