

### **Společné cvičení – ověřování normality dat:**

1. Načtete si do programu STATISTICA soubor *pacienti.sta*.
2. Vypište základní popisné statistiky pro proměnné *Leukocyty*, *Výška* a *Náklady za hospitalizaci*, pro celý soubor pacientů.

#### **Normální rozdělení – proměnná *Leukocyty*:**

3. Ověřte normalitu proměnné *Leukocyty* pomocí:
  - histogramu (*Nápověda: Graphs – Histogram*),
  - krabicového grafu (*Nápověda: Graphs – 2D – Box Plots*),
  - diagnostických grafů (Q-Q grafu, N-P grafu a P-P grafu) (*Nápověda: Graphs – 2D – Quantile-Quantile Plots / Normal Probability Plots / Probability-Probability Plots*),
  - Shapirova-Wilkova testu nebo Lilieforsovy modifikace Kolmogorovova-Smirnovova testu (*Nápověda: lze provést třemi způsoby: 1) v nastavení histogramu: záložka Advanced → Statistics: vybereme test, 2) v nastavení N-P grafu: záložka: Quick → Statistics: zaškrtneme test, 3) v menu Basic statistics → Frequency tables → záložka Normality → vybereme test a klikneme na Tests for Normality*).
4. Podívejte se, jak vypadají jednotlivé diagnostické grafy v případě normálního rozdělení.

#### **Normální rozdělení s odlehlou hodnotou – proměnná *Výška*:**

5. Ověřte normalitu proměnné *Výška* pomocí:
  - histogramu,
  - krabicového grafu,
  - diagnostických grafů (Q-Q grafu, N-P grafu a P-P grafu),
  - Shapirova-Wilkova testu / Lilieforsovy modifikace Kolmogorovova-Smirnovova testu.
6. Jak se projeví odlehlá hodnota v grafech?
7. Zkopírujte proměnnou *výška* (nebo vytvořte pomocí vzorce) do nové proměnné a vymažte v této nové proměnné odlehlou hodnotu (*nápověda: seřadte si data podle proměnné výška: karta Data → Sort → vložíme proměnnou výška*). Ověřte, zda se po vynechání odlehlé hodnoty data řídí normálním rozložením.

#### **Logaritmicko-normální rozdělení – proměnná *Náklady za hospitalizaci*:**

9. Vykreslete histogram proměnné *Náklady za hospitalizaci*. Proložte histogram nejdříve normálním rozložením, poté log-normálním rozložením.
10. Dále ověřte normalitu dat pomocí:
  - diagnostických grafů (Q-Q grafu, N-P grafu a P-P grafu),
  - Shapirova-Wilkova testu / Lilieforsovy modifikace Kolmogorovova-Smirnovova testu.
11. Jak se výsledky liší ve srovnání s daty, která se řídí normálním rozdělením?
12. Transformujte proměnnou *Náklady za hospitalizaci* pomocí přirozeného logaritmu do nové proměnné (*nápověda: Data → Transforms: LogNaklady=Log(v10)*).
13. Ověřte normalitu dat nové proměnné *LogNaklady* pomocí:
  - histogramu,
  - krabicového grafu,
  - diagnostických grafů (Q-Q grafu, N-P grafu a P-P grafu),
  - Shapirova-Wilkova testu / Lilieforsovy modifikace Kolmogorovova-Smirnovova testu.

14. Vypočtete geometrický a aritmetický průměr proměnné *Náklady za hospitalizaci*. Jaký průměr se bude více blížit mediánu? Podívejte se na histogram proměnné *Náklady za hospitalizaci* a svou odpověď odůvodněte. Vypočtete aritmetický průměr zlogaritmované proměnné *LogNáklady*, hodnotu porovnejte s předchozími průměry (*Nápověda: zpětnou transformací aritmetického průměru na zlogaritmovaných datech, pomocí funkce exp() v excelu*).