

# Statistika

## Vybrané statistické funkce

### Četnosti výskytu hodnot v statistickém souboru

### Korelace

Tento text je jen úvodem do problematiky. Jeho cílem je ukázat možnost užití tabulkového kalkulátoru Excel pro vybrané statistické funkce bez podrobného popisu příslušných funkcí.

## Vybrané statistické funkce

Následující statistické funkce jsou rozděleny podle stupně užitých mocnin v jejich vzorcích.

### Střední hodnoty

- **Modus** je dán hodnotou, která se ve statistickém souboru nejčastěji vyskytuje. Je to tedy hodnota s největší četností výskytu v souboru.
  - V Excelu se užívá pro modus funkce **MODE(číslo1; číslo2...)**, přičemž číslo lze nahradit polem hodnot.
- **Medián** je prostřední hodnota v uspořádaném statistickém souboru.
  - V Excelu se užívá pro medián funkce **MEDIAN(číslo1; číslo2...)**, přičemž číslo lze nahradit polem hodnot.
- **Průměr** (aritmetický prostý) vznikne součtem hodnot statického souboru dělených jejich počtem.
  - V Excelu se užívá pro průměr funkce **PRŮMĚR(číslo1; číslo2...)**, přičemž číslo lze nahradit polem hodnot.

Výše uvedené funkce popisují střední hodnotu. **Modus** je nejčastěji se vyskytující hodnota v množině hodnot, **medián** je prostřední hodnota a **průměr** je průměrná hodnota. Žádné jednotlivé měření nějaké střední hodnoty neposkytuje úplný obraz o datech. Předpokládáme, že polovina dat je soustředěna kolem jedné nízké hodnoty a druhá polovina kolem dvou vysokých hodnot. Funkce PRŮMĚR i MEDIAN vrátí asi hodnoty z relativně prázdného středu a funkce MODE vrátí dominantní nízkou hodnotu.

### Funkce popisující odchylky jednotlivých hodnot od aritmetického průměru

- **Rozptyl** se vypočítá je součet čtverců (druhých mocnin) rozdílů hodnot a aritmetického průměru dělený počtem hodnot.
  - V Excelu se užívá pro rozptyl funkce **VAR(číslo1; číslo2...)**, přičemž číslo lze nahradit polem hodnot.
- **Směrodatná odchylka** je druhá odmocnina rozptylu.
  - V Excelu se užívá pro směrodatnou odchylku funkce **SMODCH(číslo1; číslo2...)**, přičemž číslo lze nahradit polem hodnot.

Výše uvedené funkce popisují, jak se jednotlivé hodnoty liší od střední hodnoty. Jsou to statistické funkce, které pracují s druhou mocninou.

### Funkce popisující šikmost statistického souboru

- **Šikmost** označuje stupeň asymetričnosti rozdělení veličiny kolem střední hodnoty. Kladné zešikmení označuje rozdělení s asymetrickou stranou, která se vychyluje směrem k více kladným hodnotám. Záporné zešikmení označuje rozdělení s asymetrickou stranou, která se vychyluje směrem k více záporným hodnotám. Při výpočtu šikmosti se užívá třetí mocniny.
  - V Excelu se užívá funkce **SKEW(číslo1; číslo2;...)**

### Funkce popisující špičatost statistického souboru

- **Špičatost** určuje relativní strmost nebo plochost rozdělení v porovnání s normálním rozdělením. Kladná špičatost znamená, že rozdělení je poměrně strmé. Záporná špičatost znamená, že rozdělení je poměrně ploché. Při výpočtu špičatosti se užívá čtvrté mocniny.
- V Excelu se užívá funkce **KURT**(číslo1;číslo2;...).

### Četnosti výskytu hodnot v statistickém souboru

Funkce ČETNOSTI vypočte počet výskytů hodnot v rámci oblasti hodnot a vrátí vertikální matici čísel. Funkce ČETNOSTI se například používá k sečtení počtu výsledků testu, které spadají do určitých rozmezí. Jelikož funkce ČETNOSTI vrací matici, musí být zadána jako maticový vzorec.

#### Syntaxe

**ČETNOSTI**(data;hodnoty)

- ◆ **Data** je matice nebo odkaz na množinu hodnot, jejichž četnosti chcete vypočítat. Pokud argument data neobsahuje žádné hodnoty, vrátí funkce ČETNOSTI nulovou matici.
- ◆ **Hodnoty** je matice intervalů (nebo odkaz na ně), do kterých chcete seskupit hodnoty uvedené v argumentu data. Pokud argument hodnoty neobsahuje žádné hodnoty, vrátí funkce ČETNOSTI počet prvků v argumentu data.

Počet prvků ve vrácené matici je o jeden větší než počet prvků argumentu hodnoty. Prvek, který vrácená matice obsahuje navíc, vrací počet všech hodnot převyšujících nejvyšší interval. Při počítání například tří oblastí hodnot (intervalů) zadaných do tří buněk je k získání výsledků nutné zadat funkci ČETNOSTI do čtyř polí. V buňce navíc se zobrazí počet hodnot v argumentu data větších, než hodnota třetího intervalu.

### Korelace

Pomocí korelace lze usuzovat na souvislost dvou jevů na základě dvou souborů hodnot. V Excelu lze užít:

- analytického nástroje korelace
- funkce CORREL

#### Literatura:

[1] Kontextová nápověda programu EXCEL 2000