Statistika je věda, která se snaží zkoumat reálná data a s pomocí [teorii pravděpodobnosti](http://www.matweb.cz/pravdepodobnost) se tato data snaží popisovat.

**Základní pojmy**

Ve statistice se pracuje s několika základními pojmy, které si zde popíšeme.

V prvé řadě jde o **statistický soubor**, což je konečná množina nějakých dat, která chceme zkoumat. Data mohou být obecná, může to být v zásadě cokoliv. Pokud chcete zkoumat průměrný plat v České republice, bude statistickým souborem množina všech lidí v České republice. Počet prvků ve statistickém souboru se nazývá rozsah souboru. Rozsah námi definovaného statistického souboru by tak byl roven počtu obyvatel České republiky.

Dále existuje pojem **statistická jednotka**, což je konkrétní prvek statistického souboru. V našem případě by tak statistická jednotka byl jeden konkrétní člověk.

Nakonec máme **statistický znak**, což je to, co chceme měřit. V našem příkladě by statistickým znakem byl právě plat. Statistický znak může být buď kvalitativní nebo kvantitativní. Kvantitativní (kvantita = množství, počet) znak je takový znak, který je vyjádřitelný čísly (například zrovna ten plat, výška, počet dětí, …), kvalitativní znak je pak vyjádřitelný slovně (barva, ano/ne, zaměstnání, …).

**Četnosti**

Četnost může být buď relativní nebo absolutní a udává, kolik hodnot daného znaku se vyskytuje ve statistickém souboru — buď absolutně, nebo relativně vzhledem k celkovému počtu prvků souboru.

Důležité je, že při počítání četnosti musíme aplikovat statistický znak, protože pokud bychom chtěli vypočítat absolutní četnost statistické jednotky, musíme nutně dojít buď k číslu nula, nebo jedna, protože statistický soubor je množina a [množina jako taková nepřipouští, aby v ní bylo více stejných prvků](http://www.matweb.cz/mnoziny).

Takže absolutní četnost hodnoty statistického znaku z udává počet výskytů znaku z ve statistickém souboru S. Příklad: mějme třídu o deseti žácích. Každý žák dostal na vysvědčení nějakou známku z matematiky, od jedné do pěti. Známky zaznamenává následující tabulka:

Poznámka: statistickým souborem pro tento příklad by bylo těch deset spolužáků, něco jako

V tabulce máme pro jednoduchost v prvním řádku číselný identifikátor studenta, v prvním řádku jsou tak statistické jednotky, tj. prvky statistického souboru. Ve druhém řádku máme hodnoty statistického znaku, tj hodnoty „výsledná známka z matematiky“ daného studenta.

Absolutní četnost znaku (známky na vysvědčení) z = 3 by tak byla rovna dvěma, pouze dva žáci dostali na vysvědčení trojku (jsou to „žáci“ 3 a 10). Absolutní četnost znaku z = 1 by byla tři („žáci“ 5, 6 a 9).

Relativní četnost udává, kolik procent hodnot znaku ze statistického souboru je rovno hodnotě z. Relativní četnost znaku z vypočteme takto:

kde za je absolutní četnost znaku z a |S| je rozsah statistického souboru, tj. počet prvků. Relativní četnost známky tři by tak byla:

Rozsah našeho souboru je deset, protože máme ve třídě deset žáků. Zápis v procentech získáme vynásobením stem, tedy dostali bychom 20 %. Relativní četnost jedničky by byla

**Aritmetický průměr**

Aritmetický průměr, nebo často též jen průměr, je průměr všech hodnot ve statistickém souboru. Tím slovem hodnota je myšleno po aplikaci statistického znaku. Průměr vypočteme tak, že sečteme všechny hodnoty a vydělíme je počtem hodnot v souboru. Takže zhruba takto:

Přidal jsem i vyjádření pomocí [sumy](http://www.matweb.cz/symboly), kdyby vám přišla čitelnější. Ale stačí ten předchozí výraz se zlomkem. Hodnoty x představují všechny hodnoty našeho souboru.

Příklad: vezmeme si data z předchozí tabulky a vypočteme si průměrnou známku na žáka.

V naší třídě je průměrná známka 2,4. Jak vidíte, aritmetický průměr nám může vrátit i hodnotu, která vlastně vůbec není platná — nemůžete dát známku 2,4.

Aritmetický průměr je také špatně použitelný v případě, kdy má část dat zásadně jinou hodnotu než zbytek dat. Takže pokud máme soubor s hodnotami 1, 3, 2, 5, 4, 2, 75, tak nám aritmetický průměr vyjde

Vidíme, že výsledná hodnota je daleko od všech hodnot v souboru. Je několikrát větší než prvních šest čísel a několikrát menší než poslední hodnota. To je problém, který může řešit například medián, viz dále. Aspoň už víte, proč dvě třetiny lidí nedosáhnou na průměrný plat — existuje malá skupina lidí, která má hodně nadstandardní platy, které zvyšují aritmetický průměr.

**Geometrický průměr**

Geometrický průměr se vypočte podobně jako aritmetický průměr, pouze se místo sčítání používá násobení a místo dělení [odmocnina](http://www.matweb.cz/odmocniny). Takže geometrický průměr vypočteme takto:

Geometrický průměr se může použít jako indikátor růstu. Pro příklad předpokládejme, že cena nějakého produktu vzrostla za rok o 10 %, další rok o 15 % a další rok o 5 %. Původní cena c tak byla po těch třech letech na hodnotě

Geometrický průměr těchto koeficientů by byl:

Co to znamená? Že kdyby se cena každý roz zvýšila právě o 1,0992419, tak by výsledná cena byla na stejné hodnotě:

**Modus a medián**

Modus znaku je hodnota, která má nejvyšší četnost, značí se Mod(x). Pokud se vrátíme k příkladu se známkami, pak modusem jsou hodnoty 1 a 2, protože se vyskytují nejčastěji — oba mají absolutní četnost 3.

Medián je pak prostřední hodnota, značíme Med(x). Pokud umíme uspořádat hodnoty do [neklesající](http://www.matweb.cz/funkce) [posloupnosti](http://www.matweb.cz/posloupnosti)

tak medián představuje hodnotu, která je uprostřed této posloupnosti. Prostředí hodnota se pak liší, pokud má posloupnost lichý, nebo sudý počet prvků. Pokud lichý, je medián prvek na pozici

Pokud má posloupnost sudý počet prvků, pak nemá prvek, který by byl zcela uprostřed (příklad: posloupnost 1, 2, 3, 4 prostě nemá prostřední prvek). Proto vezmeme průměr z dvou prostředních hodnot (průměr hodnot 2 a 3). Takže vzorec pro sudá čísla:

Vraťme se k příkladu, který ilustroval špatně použitý aritmetický průměr. Měli jsme soubor hodnot s=1, 3, 2, 5, 4, 2, 75. Modus by byl rovný dvěma, je to jediné číslo, které se opakuje. Pro výpočet mediánu si čísla seřadíme do posloupnosti:

Posloupnost má sedm prvků, prostředí prvek je tak a4 a ten je roven třem.