

Průměr

Průměr je ve statistice často používaná hodnota, která se počítá jako aritmetický průměr hodnot.

Co je to průměr

Průměrem se rozumí klasický aritmetický průměr sledovaných hodnot. Můžeme si pro příklad vzít tuto tabulku se mzdami v Kč:

Tomáš	10 000
Martin	15 000
Jiří	18 000
Miroslav	21 000
Jana	19 000
Lenka	25 000
Ondřej	15 000
Lukáš	45 000
Petra	29 000
Jan	85 000
Josef	15 000

Jsou to mzdy imaginárních 11 pracovníků naší imaginární firmy Matweb s.r.o. Jaká by byla průměrná mzda v této firmě? Vzali bychom všechny částky, sečetli je:

$$10000+15000+19000+21000+18000+25000+15000+45000+29000+85000+15000=297000$$

Celkem se na mzdách vyplatí 297 000 Kč měsíčně. Průměrnou mzdu získáme tak, že tuto částku vydělíme počtem zaměstnanců, tj. jedenácti: $297\,000 : 11 = 27\,000$. Průměrná mzda v naší firmě je 27 000 Kč. To není špatné, že?

Pokud tak máme soubor n hodnot, označme je x_1, x_2, \dots, x_n , tak průměrnou hodnotu získáme tak, že sečteme všechna x_i a výsledek vydělíme n . Průměrnou hodnotu obvykle značíme \bar{x} .

. Vzorec:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Případně stejný vzorec pomocí sumy:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Jak vypočítat průměr v Excelu

V českém Excelu slouží k výpočtu průměru funkce `průměr`, v anglickém pak funkce `average`. Jako parametr pak vložíme vybranou oblast buněk, ze které chceme průměr spočítat. Můžete se podívat na následující živou ukázkou z webového rozhraní Excelu, ve které je zobrazen soubor žáků spolu s počtem správných odpovědí z nějakého matematického testu.

Problémy s průměrem

Některé problémy, na které můžete s průměrnou hodnotou narazit:

- Průměrná hodnota neudává nejčastější ani nejpravděpodobnější hodnotu. Je to zkrátka jen průměrná hodnota. Pokud bychom počítali průměr známek ve škole, můžeme dojít k číslu 2,3, což ani není platná známka, kterou byste mohli ve škole dostat. Ale je to průměrná známka. Stejně tak průměrná mzda v naší firmě je 27 000, přitom tuto mzdu vlastně vůbec nikdo nemá.
- Průměrná hodnota nám může poskytnout špatný obraz o datech, pokud se v datech vyskytují nějaké velké extrémy. Průměrná mzda 27 000 Kč zní poměrně hezky, jenomže většina zaměstnanců na tuto mzdu zdaleka nedosáhne. Takto vysoká mzda je především díky jedinému zaměstnanci, Janovi, který bere 85 000 a strhává tak průměr o hodně výše. Kdybychom Jana vyhodili, byla by průměrná mzda pouze 21 200, což na první pohled vypadá jako méně zavádějící číslo.

Medián

Medián dělí nějaký soubor hodnot na dvě stejně velké části, přičemž platí, že nejméně 50 % hodnot je větších než medián a 50 % hodnot je větší než medián.

Co je to medián

U mediánu hraje roli, jestli je počet prvků populace sudý, nebo lichý. V případě, kdy je počet prvků populace lichý, získáme medián tak, že seřadíme všechny hodnoty od nejmenší po největší a hodnota, která je přesně uprostřed, bude mediánem.

Pro příklad si vezmeme soubor čísel 75, 4, 2, 3, 2, 5, 1. Je jich celkem 7, což je liché číslo. Čísla seřadíme, čímž získáme [posloupnost](#) 1, 2, 2, 3, 4, 5, 75. Mediánem, který označíme **Me**, bude prvek, který je uprostřed, což je číslo 3.

Obecný vzorec pro populaci X o velikosti $|X| = N$, jejíž prvky jsou x_i a jsou seřazené od nejmenšího po největší, by tak byl:

$$Me(X) = x_{(N+1)/2}$$

Pokud bychom to zkusili pro náš soubor hodnot $X = [1, 2, 2, 3, 4, 5, 75]$, tak bychom získali:

$$Me(X) = x_{(N+1)/2} = x_{(7+1)/2} = x_4 = 3$$

Pro sudý počet prvků musíme provést drobnou úpravu, protože sudá posloupnost hodnot nemá „prvek uprostřed“. Když z naší populace ubereme jednu dvojku, získáme populaci: $Y = [1, 2, 3, 4, 5, 75]$ a ta nemá žádný prostřední prvek. Obě čísla 3 a 4 jsou „stejně uprostřed“. V tomto případě se to řeší tak, že se vezmou oba prvky, které jsou takto uprostřed a mediánem bude jejich aritmetický průměr. Mediánem hodnot Y by tak byla hodnota $(3 + 4)/2 = 3,5$.

Obecný vzorec pro hodnoty X o velikosti N se seřazenými prvky x_i by vypadal takto:

$$Me(X) = \frac{x_{N/2} + x_{(N/2)+1}}{2}$$

Pokud takto vypočítáme medián hodnot $Y = [1, 2, 3, 4, 5, 75]$, kde prvky označíme y_i , získáme:

$$Me(Y) = (y_{N/2} + y_{(N/2)+1}) / 2 = (y_{6/2} + y_{(6/2)+1}) / 2 = (y_3 + y_4) / 2 = 3 + 4 / 2 = 3,5$$

Výhody mediánu

Medián, na rozdíl od [průměru](#), netrpí problémy, pokud jsou v souboru hodnot nějaké extrémny. Ve výše zmíněném příkladu se souborem $X = [1, 2, 2, 3, 4, 5, 75]$ bychom získali průměr $92/7 \approx 13.1428$, což je taková trochu divná hodnota, když šest ze sedmi prvků je daleko menších. Mediánem je pak rozumné číslo tři.

Medián lze dále použít pro všechny hodnoty, které lze nějak [seřadit](#). Průměr můžeme udělat jen u hodnot, které dokážeme sečíst a následně vydělit. Můžeme si vzít například velikost oblečení: tam bývá nějaká posloupnost typu XS, S, M, L, XL, XXL. Těžko bychom z takových hodnot udělali průměr, protože není jasně dané, jak sečíst XS + L nebo XL + L a není ani jasně, jak bychom to nakonec vydělili. Nicméně pokud nějaká populace koupila oblečení S, S, M, L, L, L, XXL, tak můžeme říci, že mediánem je velikost L. Problém může nastat ve chvíli, kdy máme sudý počet velikostí, protože pak nemůžeme udělat aritmetický průměr. V tomto případě obvykle prostě zvolíme jedno ze dvou prostředních čísel jako medián.

Jak vypočítat Medián v Excelu

V českém i anglickém Excelu se používá funkce `median`, která bere jako parametr vybranou oblast buněk. Můžete se podívat na následující živou ukázkou z webového rozhraní Excelu, ve které je zobrazen soubor žáků spolu s počtem správných odpovědí z nějakého matematického testu.

Modus

Modus soubor hodnot představuje hodnotu, která se v daném souboru vyskytuje nejčastěji.

Co je to modus

Pokud máme soubor hodnot $X = [2, 3, 1, 8, 8, 4, 8, 6, 2, 3]$, tak modus je hodnota, která je v tomto souboru zastoupena nejčastěji. Nejčastěji se v souboru nachází číslo 8, všechna ostatní čísla jsou tam méně často. Kdybychom do souboru přidali ještě jedno číslo 2, pak bychom měli dva mody: číslo 8 a číslo 2, protože obě čísla by se nacházela v souboru třikrát.

Výhodou modu je, že ho můžeme použít na jakýkoliv soubor prvků – prvky ani nemusí jít seřadit. Pokud tak máme soubor prvků $Y = [\text{bílá}, \text{červená}, \text{červená}, \text{žlutá}, \text{modrá}]$, tak modem bude červená barva, protože je tam nejvícekrát.

Jak vypočítat modus v Excelu

V českém i anglickém Excelu se používá funkce `mode.sngl`, která bere jako parametr vybranou oblast buněk. Tato funkce vrátí jeden modus, pokud jich očekáváte více, použijte funkci `mode.mult`. Můžete se podívat na následující živou ukázkou z webového rozhraní Excelu, ve které je zobrazen soubor žáků spolu s počtem správných odpovědí z nějakého matematického testu.

V dokumentu je jak průměr, tak medián a modus. Obsahy buněk můžete přímo zde na webu měnit a průměr se okamžitě přepočítá, takže si můžete vyzkoušet, co se stane s průměrem a s ostatními hodnotami, když změníte několik buněk. Rozsah buněk lze buď naklikat nebo lze zapsat pomocí syntaxe `B2:B21`, což znamená „vyber všechny buňky od buňky B2 po buňky B21“.

STATISTIKA – PRACOVNÍ LIST – pro žáka

ARITMETICKÝ PRŮMĚR, MODUS, MEDIÁN

- 1) Pro statistické účely byl u 90 osob zjištěn počet rodinných příslušníků a byla získána tabulka rozdělení četností. Doplňte tuto tabulku o relativní četnosti ve tvaru des. čísla a v %. Určete z tabulky modus a medián.

Počet rod. příslušníků	1	2	3	4	5	6	7
Četnost	3	9	24	33	18	2	1
Rel. četnost							
Rel. četnost (%)							

-
- 2) Pro 304 studentů prvního ročníku vysoké školy byla sestavena tabulka četností n_j jejich věku x_j . Určete relativní četnost v %, aritmetický průměr, modus, medián.

x_j	18	19	20	21	22	23
n_j	102	84	58	26	19	15
$v_j(\%)$						

- 3)** Je dán prvotní statistický soubor, který udává délky novorozenců v daném časovém období (délky v cm): 48, 47, 53, 48, 50, 48, 50, 49, 52, 56, 53, 50, 51, 50, 51, 53, 50, 49, 50, 49.
- a) Sestavte tabulku absolutní a relativní četnosti (%)
 - b) Vypočtěte aritmetický průměr
 - c) Určete modus a medián

-
- 4)** Tělesné výšky žáků jedné třídy jsou vyhodnoceny v tabulce. Vypočtěte nejmenší možnou a největší možnou průměrnou výšku žáků této třídy zaokrouhlenou na centimetry.

Výška (cm)	145-154	155-159	160-164	165-169	170-174	175-179	180-189
Počet žáků	3	3	5	5	8	6	5

- 5) V prodejně pánské obuvi zaznamenali velikosti prodaných párů během jednoho dne s tímto výsledkem: 41, 41, 41, 42, 42, 41, 39, 41, 37, 41, 45, 41, 42, 38, 40, 39, 38, 41, 41, 38, 42, 39, 44, 43, 43, 44, 39, 39, 43, 43, 40, 42, 43, 41, 41, 43, 40, 40, 40, 42, 42, 42, 41, 40, 42. Určete rozsah souboru, vypočtete absolutní a relativní četnosti i v % a ověřte, zda jejich součet činí 100%. Určete modus a medián.

STATISTIKA – PRACOVNÍ LIST – pro učitele

ARITMETICKÝ PRŮMĚR, MODUS, MEDIÁN

- 1) Pro statistické účely byl u 90 osob zjištěn počet rodinných příslušníků a byla získána tabulka rozdělení četností. Doplňte tuto tabulku o relativní četnosti ve tvaru des. čísla a v %. Určete z tabulky modus a medián.

Počet rod. příslušníků	1	2	3	4	5	6	7
Četnost	3	9	24	33	18	2	1
Rel. četnost	0,0333	0,1	0,2667	0,3667	0,2	0,0222	0,0111
Rel. četnost (%)	3,33	10	26,67	36,67	20	2,22	1,11

Mod(x) = 4 příslušníci

Med(x) = 4 příslušníci

-
- 2) Pro 304 studentů prvního ročníku vysoké školy byla sestavena tabulka četností n_j jejich věku x_j . Určete relativní četnost v %, aritmetický průměr, modus, medián.

x_j	18	19	20	21	22	23
n_j	102	84	58	26	19	15
$v_j(\%)$	33,55	27,63	19,08	8,55	6,25	4,93

$$\bar{x} = \frac{102 \cdot 18 + 84 \cdot 19 + 58 \cdot 20 + 26 \cdot 21 + 19 \cdot 22 + 15 \cdot 23}{304}$$

$$\bar{x} = \underline{\underline{19,4\text{let}}}$$

Mod(x) = 18 let

Med(x) = 19 let

3) Je dán prvotní statistický soubor, který udává délky novorozenců v daném časovém období (délky v cm): 48, 47, 53, 48, 50, 48, 50, 49, 52, 56, 53, 50, 51, 50, 51, 53, 50, 49, 50, 49.

- Sestavte tabulku absolutní a relativní četnosti (%)
- Vypočtěte aritmetický průměr
- Určete modus a medián

a)

Délka (cm)	47	48	49	50	51	52	53	56
Absol. četnost	1	3	3	6	2	1	3	1
Rel. četnost (%)	5	15	15	30	10	5	15	5

b)
$$\bar{x} = \frac{47 + 3 \cdot 48 + 3 \cdot 49 + 6 \cdot 50 + 2 \cdot 51 + 52 + 3 \cdot 53 + 56}{20}$$

$$\bar{x} = \underline{\underline{50,35cm}}$$

c) $\text{mod}(x) = 50 \text{ cm}$

$\text{med}(x) = 50 \text{ cm}$

4) Tělesné výšky žáků jedné třídy jsou vyhodnoceny v tabulce. Vypočtěte nejmenší možnou a největší možnou průměrnou výšku žáků této třídy zaokrouhlenou na centimetry.

Výška (cm)	145-154	155-159	160-164	165-169	170-174	175-179	180-189
Počet žáků	3	3	5	5	8	6	5

Nejmenší možná průměrná výška:

$$\bar{x}_1 = \frac{3 \cdot 145 + 3 \cdot 155 + 5 \cdot 160 + 5 \cdot 165 + 8 \cdot 170 + 6 \cdot 175 + 5 \cdot 180}{35}$$

$$\bar{x}_1 = \underline{\underline{167cm}}$$

Největší možná průměrná výška:

$$\bar{x}_1 = \frac{3 \cdot 154 + 3 \cdot 159 + 5 \cdot 164 + 5 \cdot 169 + 8 \cdot 174 + 6 \cdot 179 + 5 \cdot 189}{35}$$

$$\bar{x}_1 = \underline{\underline{172cm}}$$

- 5) V prodejně pánské obuvi zaznamenali velikosti prodaných párů během jednoho dne s tímto výsledkem: 41, 41, 41, 42, 42, 41, 39, 41, 37, 41, 45, 41, 42, 38, 40, 39, 38, 41, 41, 38, 42, 39, 44, 43, 43, 44, 39, 39, 43, 43, 40, 42, 43, 41, 41, 43, 40, 40, 40, 42, 42, 42, 41, 40, 42. Určete rozsah souboru, vypočtete absolutní a relativní četnosti i v % a ověřte, zda jejich součet činí 100%. Určete modus a medián.

Velikost obuvi	37	38	39	40	41	42	43	44	45	
Absol. četnost	1	3	5	6	12	9	6	2	1	
Rel. četnost (%)	2,22	6,67	11,1	13,3	26,7	20	13,3	4,44	2,22	⇒ <u>99,95%</u>

Mod(x) = velikost č. 41

Med(x) = velikost č. 41

ZDROJE:

POLÁK, Josef: *Středoškolská matematika v úlohách II*. 1. vydání. Praha: Prometheus, 1999. 626 s. ISBN 80-7196-166-3.

CALDA,E; DUPAČ,V: *Kombinatorika, pravděpodobnost, statistika*. Dotisk 2. vydání. Praha: Prometheus, 1996. 163 s. ISBN 80-85849-10-0.

ZHOUF, Jaroslav a kol: *Sbírka testových úloh k maturitě z matematiky*. 1. vydání. Praha: Prometheus, 2002. 279 s. ISBN 80-7196-249-X

JIRÁSEK, F, kol.: *Sbírka úloh z matematiky pro SOŠ a pro studijní obory SOU, 2. část*. Dotisk 3. upraveného vydání. Praha: Prometheus, 2000. 479 s. ISBN 80-7196-012-8