

# Učitel matematiky

---

František Mošna  
Medián a kvartily

*Učitel matematiky*, Vol. 30 (2022), No. 3, 149–159

Persistent URL: <http://dml.cz/dmlcz/151112>

## Terms of use:

© Jednota českých matematiků a fyziků, 2022

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ*:  
*The Czech Digital Mathematics Library* <http://dml.cz>

## MEDIÁN A KVARTILY

FRANTIŠEK MOŠNA

### Úvod

Popisná statistika sbírá a shromažďuje hromadná data, snaží se je utřídit, uspořádat, zpřístupnit a zprostředkovat o nich co nejvhodnější a nejprehlednější informace formou tabulek, přehledů a grafů. Mezi jednoduché nástroje, které nám mohou pomoci přiblížit vlastnosti souborů hromadných dat, patří čísla zvaná charakteristiky (někdy míry).

Charakteristiky rozlišujeme podle toho, co vyjadřují. Jeden typ charakteristik se snaží popsat a vyjádřit jakési centrální hodnoty, kolem kterých jsou data soustředěna – těm se říká charakteristiky polohy. Jiný typ charakteristik vyjadřuje, jak hodně či málo jsou okolo centrálních hodnot ty ostatní soustředěny či rozptýleny – ty nazýváme charakteristiky variability. Kromě toho existují charakteristiky popisující další vlastnosti souborů dat – šikmost, špičatost a podobně. Mezi charakteristikami se vyskytují také takové, jež jsou založeny na uspořádání dat podle velikosti. Této skupině někdy říkáme pořadové charakteristiky.

### Charakteristiky

K zavedení a výpočtu charakteristik používáme vzorce a postupy uvedené v každé učebnici statistiky, např. (Robová et al., 2013). Uvažujme nějaký  $n$ -členný soubor dat  $x_1, x_2, \dots, x_n$ , počet jeho prvků  $n$  nazýváme rozsahem souboru. Zcela běžnou charakteristikou polohy je aritmetický průměr  $\bar{x}$  (average, mean). Počítáme ho podle známého vzorce

$$\bar{x} = \frac{1}{n}(x_1 + \dots + x_n).$$

Mezi charakteristiky variability řadíme směrodatné odchylky. Směrodatnou odchylku  $\sigma$  (population standard deviation) užíváme většinou v případě, kdy uvažujeme celý základní soubor, a počítáme podle vzorce

$$\sigma = \sqrt{\frac{1}{n}((x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2)}.$$

Pro výběr ze základního souboru dat volíme zpravidla výběrovou směrodatnou odchylku  $s$  (sample standard deviation)<sup>1</sup>

$$s = \sqrt{\frac{1}{n-1}((x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2)}.$$

Jak jsme již zmínili, vedle průměru a dvou typů směrodatných odchylek užíváme ve statistice také tzv. pořadové charakteristiky. Nejznámější takovou charakteristikou polohy je patrně medián  $\tilde{x}$ . Běžně si představujeme, že jde o jakousi „prostřední“ hodnotu.

Uvažujme opět nějaký soubor dat  $x_1, x_2, \dots, x_n$  o rozsahu  $n$ . Seřadíme hodnoty podle velikosti a výsledný uspořádaný soubor budeme značit s indexy v závorkách  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ . Hodnotou  $x_{(k)}$  tedy označujeme (pro  $k$  přirozené)  $k$ -tý člen podle velikosti od nejmenšího do největšího. Bývá užitečné rozšířit zavedení hodnoty  $x_{(s)}$  také pro neceločíselná  $s > 1$  jako jakousi aproximaci. Každé reálné číslo  $s$  můžeme rozložit na celou část  $[s]$  a zlomkovou část  $\{s\}$ , platí  $s = [s] + \{s\}$ . Hodnotu  $x_{(s)}$  pak definujeme podle vzorce

$$x_{(s)} = x_{([s])} + \{s\} \cdot (x_{([s]+1)} - x_{([s])}).$$

Použijeme-li tuto definici, lze zavést medián jednoduše jako  $\tilde{x} = x_{(\frac{n+1}{2})}$ . Pokud je rozsah souboru  $n$  lichý, je medián roven přímo prostřední hodnotě, pokud je rozsah sudý, za medián bereme střed ze dvou prostředních hodnot.

Medián rozděluje soubor dat na dvě části. Někdy bývá vhodné rozdělit uspořádaná data nejen na poloviny, ale na čtyři části.

<sup>1</sup>Dá se ověřit, že tato charakteristika má mezi odhady vlastnost nestrannosti, což přibližně znamená, že odhad nenadhodnocuje ani nepodhodnocuje odhadovaný parametr.

K tomu slouží tzv. kvartily, dolní  $\tilde{x}_{25}$  a horní  $\tilde{x}_{75}$ . Představujeme si opět, že jde o jakési „prostřední“ hodnoty dolní a horní části souboru dat. S využitím kvartilů lze zavést charakteristiku variability zvanou mezikvartilové rozpětí  $MKR = \tilde{x}_{75} - \tilde{x}_{25}$  (Robová et al., 2013, s. 175).<sup>2</sup>

Mnohdy stojíme před otázkou, který typ charakteristik je v dané situaci více a který méně vhodný. První skupina charakteristik – tedy aritmetický průměr a směrodatné odchylky – závisejí na všech hodnotách souboru dat, tedy i na tzv. odlehlých hodnotách – na těch, které se hodně odlišují od většiny z nich. Druhá skupina charakteristik – medián, kvartily, mezikvartilové rozpětí – je zavedena pomocí uspořádaných dat. Data, která se výrazně vymykají jakémusi centru, na takové charakteristiky nemají velký vliv. Říkáme, že tyto charakteristiky jsou stabilnější, robustnější, že je odlehlé hodnoty příliš neovlivní.

Za příklad nevhodného užití statistiky je často uváděno vyjádření výše mzdy jisté populace prostřednictvím aritmetického průměru (Swoboda, 1976, s. 45–47). Nevhodnost této charakteristiky spočívá v tom, že bývá zpravidla výrazně ovlivněna velmi vysokými příjmy několika málo jedinců. Ukazuje se, že v tomto případě je vhodnější užívat medián.

## Zavedení kvartilů

Způsoby přesného zavedení kvartilů bývají různé. Při jejich výpočtu pomocí kalkulaček či statistického softwaru často obdržíme odlišné výsledky. Účelem tohoto článku je především seznámit čtenáře se způsoby zavedení kvartilů a s jejich rozdíly. Uvedeme čtyři nejčastěji užívané metody výpočtu (budeme je označovat A, B, C a D).

První dvě zavedení dávají stejné výsledky pro sudý rozsah souboru  $n$ . V tomto případě je soubor rozdělen na dvě části, jež medián neobsahují, a kvartily pak zavádíme jako medián horní a dolní části. Pro lichý rozsah  $n$  se však metody od sebe liší. Zá-

---

<sup>2</sup>Někdy se zavádí též mezikvartilová odchylka jako polovina mezikvartilového rozpětí (Caldá & Dupač, 2000, s. 145).

kladní volba spočívá v tom, jestli medián zahrneme do obou částí souboru (do horní i dolní) nebo do žádné části souboru.

**A. metoda including:** Kvartily zavádíme jako mediány horní a dolní části dat, přičemž medián  $\tilde{x}$  zahrnujeme do obou částí.

**B. metoda excluding:** Kvartily zavádíme jako mediány horní a dolní části dat, přičemž medián  $\tilde{x}$  vylučujeme z obou částí. Tuto metodu používá například kalkulačka CASIO fx-85CEX.

Ukažme si tyto metody na jednoduchém příkladu uspořádaných dat 1, 2, 3, 4, 5. Rozsah souboru je  $n = 5$ , medián máme zaveden jako  $x_{(\frac{n+1}{2})} = x_{(3)} = 3$ . Nyní budeme počítat kvartily. Při užití metody including (A) zahrneme medián do obou částí, ta dolní je tedy 1, 2, 3 a ta horní 3, 4, 5. Dolní kvartil je tedy medián dolní části, což jest 2, horní kvartil dostaneme jako medián horní části, což jest 4.

Při užití metody excluding (B) medián nezahrneme ani do jedné z částí, ta dolní je tedy 1, 2 a ta horní 4, 5. Dolní kvartil je tedy medián dolní části, což jest 1,5 a horní kvartil dostaneme jako medián horní části, což jest 4,5.

Na dalším jednoduchém příkladu uspořádaných dat 1, 2, 3, 4, 5, 6 se můžeme přesvědčit, že obě metody (A a B) dávají stejný výsledek pro sudý rozsah souboru  $n = 6$ . Medián máme zaveden jako  $x_{(\frac{n+1}{2})} = x_{(3,5)} = 3,5$ . Medián nepatří ani do dolní části 1, 2, 3, ani do horní 4, 5, 6. Dolní kvartil je tedy medián dolní části, což jest 2, horní kvartil dostaneme jako medián horní části, což jest 5.

Medián a kvartily se používají také při vytváření tzv. krabíkových grafů neboli box-plotů (Mošna, 2015). Metoda including (A) se shoduje se zavedením tzv. hinges (panty, závěsy), kterých John Tukey (1977) původně užíval pro konstrukci těchto diagramů. Při tomto postupu vypočítáme nejprve pozici mediánu  $M = \frac{n+1}{2}$  a pozici dolního kvartilu  $H = \frac{\lfloor M \rfloor + 1}{2}$ . Potom zavádíme  $\tilde{x} = x_{(M)}$ ,  $\tilde{x}_{25} = x_{(H)}$ ,  $\tilde{x}_{75} = x_{(n+1-H)}$ .

Ukažme si tento postup opět na příkladu uspořádaných dat 1, 2, 3, 4, 5. Rozsah souboru je  $n = 5$ , pozice mediánu je tedy  $M = \frac{5+1}{2} = 3$  a pozice dolního kvartilu  $H = \frac{3+1}{2} = 2$ . Proto  $\tilde{x} = x_{(3)} = 3$ ,  $\tilde{x}_{25} = x_{(2)} = 2$  a  $\tilde{x}_{75} = x_{(5+1-2)} = x_{(4)} = 4$ .

Podobně postupujeme u druhého uvedeného příkladu uspořádaných dat 1, 2, 3, 4, 5, 6. Rozsah souboru je  $n = 6$ , pozice mediánu je tedy  $M = \frac{6+1}{2} = 3,5$  a pozice dolního kvartilu  $H = \frac{[3,5]+1}{2} = 2$ . Proto  $\tilde{x} = x_{(3,5)} = 3,5$ ,  $\tilde{x}_{25} = x_{(2)} = 2$  a  $\tilde{x}_{75} = x_{(6+1-2)} = x_{(5)} = 5$ .

Další dvě metody C a D jsou založeny na pojmu kvantil. Proto si nejprve přiblížíme způsoby zavedení a výpočtu kvantilu a ty pak využijeme k popisu metod C a D.

## Kvantily a jejich užití při zavedení kvartilů

Soubory dat můžeme rozdělit i v jiných poměrech než na poloviny nebo na čtvrtiny. Obecně se pro  $p$  mezi 0 a 100 zavádí tzv.  $p$ -kvantil, který rozděluje soubor dat na část  $p$  procent hodnot menších nebo rovných  $p$ -kvantilu a  $100 - p$  procent hodnot větších nebo rovných  $p$ -kvantilu. Představíme si dvě metody výpočtu.

- 1) **metoda including:** Hledaný  $p$ -kvantil představuje hodnotu<sup>3</sup>  $x_{(s)}$ , kde  $s = \frac{p}{100} \cdot (n - 1) + 1$ .
- 2) **metoda excluding:** Hledaný  $p$ -kvantil představuje hodnotu  $x_{(s)}$ , kde  $s = \frac{p}{100} \cdot (n + 1)$ .

Odtud odvozujeme výpočet kvartilů.

- C. metoda kvantil including:** Dolní (resp. horní) kvartil je 25% (resp 75%) kvantil získaný metodou including. Takto zavádí kvartily a kvantily např. Robová et al. (2013, s. 172). Tuto metodu užívá funkce programu MS Excel quartile nebo quartile.inc, příkaz quantile v R s defaultním nastavením nebo s vlastností type=7.

---

<sup>3</sup>Někdy se u této metody uvádí hodnota  $x_{(s)}$ , kde  $s = \frac{p}{100} \cdot (n - 1)$  s tím, že uspořádaný soubor dat číslujeme od nuly do  $n - 1$  (Mošna, 2017, s. 63).

**D. metoda kvantil excluding:** Dolní (resp. horní) kvartil je 25% (resp 75%) kvantil získaný metodou excluding. Tento způsob užívá funkce programu MS Excel quartile.exc, program SPSS nebo příkaz quantile v R s vlastností type=6.

Ukažme si výpočet kvartilů na příkladu uspořádaných dat 1, 2, 3, 4, 5, 6. Rozsah souboru je  $n = 6$  a medián  $\tilde{x} = 3,5$ .

Postupujme nejprve metodou kvantilu including (C). Pozice 25% kvantilu je  $s = 0,25 \cdot 5 + 1 = 2,25$ , a dolní kvartil je tedy  $x_{(2,25)} = 2,25$ . Podobně pozice 75% kvantilu je  $s = 0,75 \cdot 5 + 1 = 4,75$  a horní kvartil je  $x_{(4,75)} = 4,75$ .

Nyní obrátíme pozornost k výpočtu kvartilů pomocí metody kvantilu excluding (D). Pozice 25% kvantilu je  $s = 0,25 \cdot 7 = 1,75$  a dolní kvartil je  $x_{(1,75)} = 1,75$ . Podobně pozice 75% kvantilu je  $s = 0,75 \cdot 7 = 5,25$  a horní kvartil je  $x_{(5,25)} = 5,25$ .

## Ukázka výpočtu

Výpočty mediánu a kvartilů si ukážeme na dvou příkladech. Inspirovali jsme se v učebnici od Robové a kol. (2013, s. 162–173) a využíváme dat zde uvedených.

**Příklad 1.** U dvanácti přístrojů určitého typu byla zjištěna doba od uvedení do provozu do první poruchy. Získané hodnoty jsou vyjádřeny v celých dnech:

$$59, 72, 64, 68, 2, 38, 68, 42, 61, 57, 84, 66.$$

Vypočítejte medián a kvartily jednotlivými metodami.

**Řešení.** Při užití každé z metod výpočtu mediánu a kvartilů je vhodné si data uspořádat podle velikosti:

$$2, 38, 42, 57, 59, 61, 64, 66, 68, 68, 72, 84.$$

Rozsah souboru je  $n = 12$ , medián je  $\tilde{x} = x_{\left(\frac{n+1}{2}\right)} = x_{(6,5)} = 61 + 0,5 \cdot (64 - 61) = 62,5$ .

Medián rozděluje data na dvě části a protože rozsah souboru je sudý, medián nepatří ani do dolní části 2, 36, 42, 57, 59, 61, ani do horní části 64, 66, 68, 68, 72, 84. Proto se metody A a B neliší

a dávají nám stejný výsledek. Dolní kvartil počítáme jako střed čísel 42 a 57, což je 49,5. Horní kvartil vyjde 68. Metodu B užívá např. kalkulačka CASIO fx-85CEX a program Statistica (obr. 1).

Proměnná	Popisné statistiky (Tabulka dat1)				
	Průměr	Medián	Dolní kvartil	Horní kvartil	sm.odch.
<b>příklad1</b>	56,75000	62,50000	49,50000	68,00000	21,22659

Obr. 1: Charakteristiky, metoda B – příklad 1 (program Statistica)

Obraťme pozornost k metodě C. Připomeňme, že rozsah souboru je  $n = 12$ , pozice 25% kvantilu je tedy  $s = 0,25 \cdot 11 + 1 = 3,75$ . Dolní kvartil je potom  $x_{(3,75)} = 42 + 0,75 \cdot (57 - 42) = 53,25$ . Pozice 75% kvantilu je  $s = 0,75 \cdot 11 + 1 = 9,25$  a horní kvartil je vzhledem ke shodě 9. a 10. hodnoty  $x_{(9,75)} = 68$ . Metodu C používá program R s defaultním nastavením nebo s volbou `type=7` (obr. 2) a funkce programu MS Excel `quartile.inc` nebo `quartile`.

```
> priklad1=c(59,72,64,68,2,38,68,42,61,57,84,66)
> quantile(priklad1); quantile(priklad1, type=7)
 0%   25%   50%   75%  100%
2.00 53.25 62.50 68.00 84.00
 0%   25%   50%   75%  100%
2.00 53.25 62.50 68.00 84.00
```

Obr. 2: Kvartily, metoda C – příklad 1 (program R)

Užijeme-li metodu D, pozice 25% kvantilu je  $s = 0,25 \cdot 13 = 3,25$ . Dolní kvartil je tedy  $x_{(3,25)} = 42 + 0,25 \cdot (57 - 42) = 45,75$ . Podobně pozice 75% kvantilu je  $s = 0,75 \cdot 13 = 9,75$  a horní kvartil je tedy  $x_{(9,75)} = 68 + 0,75 \cdot (68 - 68) = 68$ . Metodu D používá program R s volbou `type=6` (obr. 3), program SPSS (obr. 4) a funkce programu MS Excel `quartile.exc`.

```
> priklad1=c(59,72,64,68,2,38,68,42,61,57,84,66)
> quantile(priklad1, type=6)
 0%   25%   50%   75%  100%
2.00 45.75 62.50 68.00 84.00
```

Obr. 3: Kvartily, metoda D – příklad 1 (program R)



**Statistics**

příklad1

N	Valid	12
	Missing	0
Median		62,5000
Percentiles	25	45,7500
	50	62,5000
	75	68,0000

Obr. 4: Kvartily, metoda D – příklad 1 (program SPSS)

Kromě mediánu a kvartilů uvedme navíc také průměr a směrodatné odchyly  $\bar{x} = 56,750$ ,  $\sigma_x = 20,323$  a  $s_x = 21,227$ .

**Příklad 2.** V příkladu 1 se mezi dvanácti přístroji vyskytl jeden – ten pátý – od jiného dodavatele. Proto příslušnou hodnotu 2 vypustíme. Jak se změní medián a kvartily?

**Řešení.** Vyjdeme opět z uspořádaného souboru:

38, 42, 57, 59, 61, 64, 66, 68, 68, 72, 84.

Medián je prostřední hodnota 64. Metody A a B pro výpočet kvartilu se ale mohou lišit, protože rozsah souboru  $n = 11$  je tentokrát lichý. Medián nám rozděluje soubor dat na dvě části. Při výpočtu kvartilů metodou A ho zahrneme do obou částí. Dolní kvartil počítáme jako prostředek dolní části 36, 42, 57, 59, 60, 64, což je průměr hodnot 57 a 59, tedy 58. Prostředek horní části 64, 66, 68, 68, 72, 84 nám dá horní kvartil 68.

Užijeme-li metodu B, nezahrneme medián 64 ani do jedné z částí. Dolní kvartil počítáme jako prostředek dolní části 36, 42, 57, 59, 60, což je hodnota 57. Podobně prostředek horní části 66, 68, 68, 72, 84 dá horní kvartil 68. Opět si uvědomíme, že metodu B užívá např. kalkulačka CASIO fx-85CEX a program Statistica (obr. 5).

Proměnná	Popisné statistiky (Tabulka dat1)				
	Průměr	Medián	Dolní kvartil	Horní kvartil	sm.odch.
<b>příklad2</b>	61,727271	64,000000	57,000000	68,000000	12,98531

Obr. 5: Charakteristiky, metoda B – příklad 2 (program Statistica)

Rozsah souboru je  $n = 11$ , podle metody C obdržíme pozici 25% kvantilu rovnou číslu  $s = 0,25 \cdot 10 + 1 = 3,5$  a dolní kvartil je  $x_{(3,5)} = 57 + 0,5 \cdot (59 - 57) = 58$ . Pozice 75% kvantilu je  $s = 0,75 \cdot 10 + 1 = 8,5$  a horní kvartil je 68.

Jak jsme uvedli, tyto výsledky obdržíme v programu R s defaultním nastavením nebo s volbou `type=7` (obr. 6) a pomocí funkcí programu MS Excel `quartile.inc` nebo `quartile`.

```
> priklad2=c(59,72,64,68,38,68,42,61,57,84,66)
> quantile(priklad2); quantile(priklad2, type=7)
0% 25% 50% 75% 100%
38 58 64 68 84
0% 25% 50% 75% 100%
38 58 64 68 84
```

Obr. 6: Kvartily, metoda C – příklad 2 (program R)

Podle metody D je pozice 25% kvantilu je  $s = 0,25 \cdot 12 = 3$  a dolní kvartil je  $x_{(3)} = 57$ . Pozice 75% kvantilu je  $s = 0,75 \cdot 12 = 9$  a horní kvartil je 68.

Tyto výsledky opět obdržíme v programu R s volbou `type=6` (obr. 7), v programu SPSS (obr. 8) a pomocí funkce programu MS Excel `quartile.exc`.

```
> priklad2=c(59,72,64,68,38,68,42,61,57,84,66)
> quantile(priklad2, type=6)
0% 25% 50% 75% 100%
38 57 64 68 84
```

Obr. 7: Kvartily, metoda D – příklad 2 (program R)

**Statistics**

příklad2

N	Valid	11
	Missing	1
Median		64,0000
Percentiles	25	57,0000
	50	64,0000
	75	68,0000

Obr. 8: Kvartily, metoda D – příklad 2 (program SPSS)

Při vynechání hodnoty 2 se průměr a směrodatné odchytky změni na  $\bar{x} = 61,727$ ,  $\sigma_x = 12,381$  a  $s_x = 12,985$ .

## Závěr

Ukázali jsme si čtyři způsoby výpočtu kvartilů. Uvedené informace mohou být užitečné zejména pro tu část učitelů matematiky, kteří učí statistiku a pro zpracování dat užívají kalkulačky, Excel nebo statistické programy jako R, SPSS či Statistica. Právě ti si totiž často potřebují ujasnit rozdíly mezi výsledky získanými některými z těchto nástrojů a následně je zdůvodnit studentům. Je však třeba mít na paměti, že výčet možností výpočtu kvartilů těmito čtyřmi uvedenými metodami (A, B, C a D) zdaleka nekončí. Užívá se celá řada dalších způsobů výpočtu kvartilů (Langford, 2006) a kvantilů (Hyndman & Fun, 1996).

## Literatura

- [1] Calda, E., & Dupač, V. (2000). *Matematika pro gymnázia: Kombinatorika, pravděpodobnost, statistika*. Prometheus.
- [2] Hyndman, R. J., & Fun, Y. (1996). Sample quantiles in statistical packages. *American Statistician*, 50(4), 361–365. <https://doi.org/10.1080/00031305.1996.10473566>

- [3] Langford, E. (2006). Quartiles in elementary statistics. *Journal of Statistics Education*, 14(3), 1–26. <https://doi.org/10.1080/10691898.2006.11910589>
- [4] Mošna, F. (2015). Box-ploty v MS Excelu a jejich využití ve výuce. *Učitel matematiky*, 23(4), 206–214.
- [5] Mošna, F. (2017). *Základní statistické metody*. PedF UK.
- [6] Robová, J., Hála, M., & Calda, E. (2013). *Komplexní čísla, kombinatorika, pravděpodobnost a statistika*. Prometheus.
- [7] Swoboda, H. (1976). *Moderní statistika*. Svoboda.
- [8] Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

## Abstract

The paper deals with measures of statistical data, especially with medians, and quartiles. There exist many ways how to introduce lower and upper quartiles. The text introduces the reader to four of them – method including, excluding and methods using quartiles. This contribution can serve to orientation among various results which are supplied by calculators, Excel and statistical programmes.

*František Mošna*  
*Pedagogická fakulta*  
*Univerzita Karlova*  
*Magdalény Rettigové 4*  
*116 39 Praha 1*  
*e-mail: mosna@tf.czu.cz*