

# STATISTIKA A ZPRACOVÁNÍ DAT

## Anotace

V této aktivitě se žáci seznámí s výpočtem aritmetického průměru, modu a mediánu a tvorbou diagramů. Nabyté znalosti si vyzkouší jednak vlastním výpočtem, ale i konstrukcí formuláře pro sběr dat a jejich zpracováním v tabulkovém procesoru.

## Autor

Michal Heczko

Gymnázium a Jazyková škola s právem státní jazykové zkoušky Zlín

[heczko@gjszlin.cz](mailto:heczko@gjszlin.cz)

## Informace o výukovém materiálu

- Typ materiálu učebnice/cvičení
- Škola základní škola (2. stupeň)
- Ročník 9.
- Metody výklad, samostatná nebo skupinová práce

## Zařazení

- Vzdělávací obor: Matematika a její aplikace
- Tematický okruh: Závislosti, vztahy a práce s daty
- Očekávané výstupy:
  - M-9-2-01 žák vyhledává, vyhodnocuje a zpracovává data
  - M-9-2-02 žák porovnává soubory dat

## Mezipředmětové vztahy

- Vzdělávací obor: Informatika
- Tematický okruh: Data, informace a modelování
- Očekávané výstupy:
  - I-9-1-01 žák získá z dat informace, interpretuje data, odhaluje chyby v cizích interpretacích dat
  - I-9-1-03 žák vymezí problém a určí, jaké informace bude potřebovat k jeho řešení; situaci modeluje pomocí grafů, případně obdobných schémat; porovná svůj navržený model s jinými modely k řešení stejného problému a vybere vhodnější, svou volbu zdůvodní

## Časová dotace

- Výuka 45–90 minut
- Příprava 15–45 minut

## Pomůcky, hardware a software

- Učitel: Microsoft Excel (Office 365, 7/2022) či jiný tabulkový procesor  
Microsoft Forms nebo Google Forms
- Žák: Microsoft Excel (Office 365, 7/2022) či jiný tabulkový procesor  
Microsoft Forms nebo Google Forms

*Autorem materiálu a všech jeho částí, není-li uvedeno jinak, je Michal Heczko.*

*Koordinováno a recenzováno pracovní skupinou Národního kabinetu Matematika a její aplikace.*

*Dostupné z Metodického portálu [www.rvp.cz](http://www.rvp.cz), ISSN: 1802-4785. Provozuje Národní pedagogický institut České republiky.*

## Zdroje

- HERMAN, Jiří a kol. *Matematika pro nižší ročníky víceletých gymnázií – Úměrnosti (tercie)*. 1. vydání Praha: Nakladatelství Prometheus, 2000. ISBN 978-80-7196-480-3.
- HERMAN, Jiří a kol. *Matematika pro nižší ročníky víceletých gymnázií – Funkce (kvarta)*. 1. vydání Praha: Nakladatelství Prometheus, 2000. ISBN 978-80-7196-182-6.
- Formulář v MS Forms:  
<https://forms.microsoft.com/Pages/ShareFormPage.aspx?id=9ghLpK-rik6a5kdh35H5xkjJuBJkqthFpWPrDRketalUQ1IUOENDMUJYSUtENkdYSIVPU1NSSIBMMC4u&sharetoken=9BSjvzJMSNGouMRTFXZa>
- Tabulka se zdrojovými daty:  
<https://1drv.ms/x/s!AohVPCrNhfIOicEG487dtJkqWmfkLQ?e=YI3v9K>
- Výsledná tabulka: <https://1drv.ms/x/s!AohVPCrNhfIOicED9SSHUOVYoUwvA?e=ULs2rb>

## Potřebné vstupní znalosti a dovednosti

### Matematika

- Základní statistické pojmy
- Aritmetický průměr, modus, medián
- Diagramy

### Digitální dovednosti

- Základy ovládání tabulkového procesoru
- Práce se vzorci a funkcemi
- Práce s grafy

## Vzdělávací cíle

### Matematika

- Žák umí zpracovat tabulku četností a interpretovat tyto informace.
- Žák umí zpracovat statistická data a vypočítat základní statistické charakteristiky – aritmetický průměr, modus, medián
- Žák umí vytvořit diagramy k jednotlivým tabulkám a tyto informace vyhodnotit.

### Digitální kompetence

- Vyhodnocování a porovnávání souborů dat, prezentace a interpretace výsledků i za pomoci digitálních technologií

### Digitální dovednosti

- Zpracování statistických dat za pomoci tabulkového procesoru a jeho funkcí pro výpočet aritmetického průměru, modu a mediánu.
- Tvorba a vyhodnocování diagramů za pomoci tabulkového procesoru.

## Přínos využití digitálních technologií

Ukázka praktického využití ICT pro zpracování statistických dat většího rozsahu, včetně doplnění diagramů. Žáci si sami vyzkouší, jak anketu vytvořit a data následně zpracovat.

## Metodická poznámka

Materiál lze použít buď se vzorovými daty (ke stažení zde: <https://1drv.ms/x/s!AohVPCRnHfIOicEG487dtJkqWmfkLQ?e=YI3v9K>) nebo lze s žáky vytvořit vlastní data (buď za pomoci nástroje pro tvorbu formulářů, nebo z jiného průzkumu).

## Popis materiálu (vzdělávací aktivity)

### Základní statistické pojmy

**Statistický soubor** – skupina, kterou zkoumáme.

**Znak** – informace, kterou zkoumáme. Znaky lze rozdělit do dvou kategorií – **kvantitativní znak**, který lze vyjádřit číslem (kvantita = množství) – např. doba cesty do školy. Druhým typem znaku je **kvalitativní znak** (např. barva vlasů, druhý cizí jazyk, ...).

Jednotlivé prvky statistického souboru jsou **statistické jednotky** (např. jednotliví žáci, kterých se ptáme) a jejich počet nazýváme rozsah statistického souboru.

### Četnosti a diagramy

#### Tabulka četností

Hodnoty daného statistického znaku lze shrnout v tabulce četností. **Četnost** je počet výskytů dané hodnoty statistického znaku ve statistickém souboru. Jak vypadá tabulka četností, je ukázáno v následující úloze.

**Úloha 1:** Ve třídě byly známky z posledního testu z matematiky následující: 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 5, 5. Sestavte tabulku četností a relativních četností.

Řešení:

Data můžeme je shrnout v tabulce četností, kdy v prvním řádku tabulky budou uvedeny známky a ve druhém řádku jejich počet.

Známka	1	2	3	4	5
Počet žáků	5	12	8	3	2

Druhou možností je použití tzv. relativních četností, které vyjadřují podíl hodnoty na celkovém počtu hodnot. Tedy, pokud výše uvedená tabulka shrnuje známky 30 žáků a 5 z nich mělo 1, potom relativní četnost známky 1 je  $5 : 30 = 0,17$ .

Součet relativních četností bude 1.

Známka	1	2	3	4	5
Počet žáků	0,17	0,4	0,27	0,1	0,06

Relativní četnost nám umožňuje porovnávat dva soubory s různým rozsahem (s různým počtem hodnot). Lze ji vyjádřit i v procentech.

### Diagramy

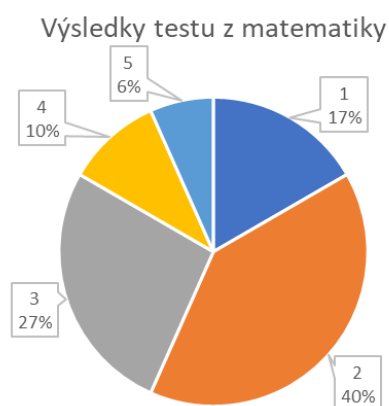
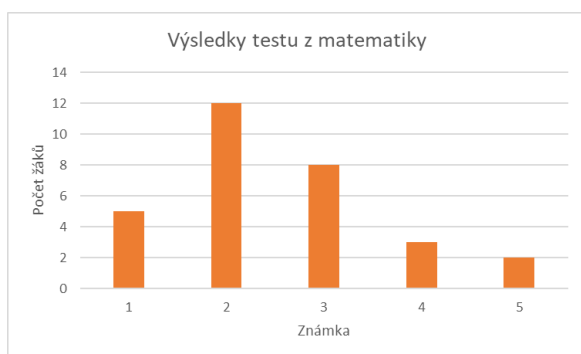
Druhou možností, jak shrnout a vizualizovat data jsou diagramy. Podle způsobu znázornění diagramů rozlišujeme sloupkové a kruhové diagramy.

**Sloupkový diagram** obsahuje na vodorovné ose jednotlivé hodnoty a na svislé ose jejich četnosti (případně relativní četnosti).

**Kruhový diagram** znázorňuje jednotlivé hodnoty pomocí kruhových výsečí. Velikost středového úhlu této výseče odpovídá relativní četnosti dané hodnoty.

Pokud bychom vizualizovali výše uvedenou tabulku se známkami, odpovídal by známce 1 úhel 60 stupňů (celý kruh představuje 360 stupňů):

$$\frac{5}{30} \cdot 360^\circ = 60^\circ$$



## Základní statistické charakteristiky

### Aritmetický průměr

Aritmetický průměr daného znaku (někdy označováno jako „průměrná hodnota“ nám soubor hodnot umožňuje nahradit jednou hodnotou, která tento soubor charakterizuje. Zjednodušeně lze říct, že se jedná o součet hodnot vydělený jejich počtem. Při výpočtu však můžeme vyjít i z tabulky četností, jak je vidět v následující tabulce shrnující výsledky písemné práce z matematiky.

**Úloha 2:** Vypočítejte průměrnou známku z testu z matematiky, pokud jsou známky dány tabulkou četností.

Řešení:

Známka	1	2	3	4	5
Počet žáků	5	12	8	3	2

Výpočet je potom následující:

$$x = \frac{5 \cdot 1 + 12 \cdot 2 + 8 \cdot 3 + 3 \cdot 4 + 2 \cdot 5}{30} = \frac{75}{30} = 2,5$$

### Modus

Modus je nejčastější hodnota ve statistickém souboru. Pokud se v tomto souboru vyskytuje více hodnot s touto četností, jsou jako modus označeny všechny tyto hodnoty. V případě, že by všechny hodnoty měly četnost jedna, tak lze jako modus označit všechny hodnoty statistického souboru.

U výše uvedené tabulky (na které byl vysvětlen výpočet aritmetického průměru) je modus 2 (protože známka 2 se vyskytuje 12x, tedy nejčastěji).

*Autorem materiálu a všech jeho částí, není-li uvedeno jinak, je Michal Heczeko.*

*Koordinováno a recenzováno pracovní skupinou Národního kabinetu Matematika a její aplikace.*

*Dostupné z Metodického portálu [www.rvp.cz](http://www.rvp.cz), ISSN: 1802-4785. Provozuje Národní pedagogický institut České republiky.*

## Medián

Medián určuje prostřední hodnotu souboru hodnot – dělí hodnoty tak, že polovina hodnot je menší, než medián a polovina hodnot je větší než medián. Výhodou mediánu je, že hodnotu mediánu tolik neovlivňují extrémní hodnoty, jako v případě aritmetického průměru.

Výpočet mediánu můžeme rozdělit do dvou kategorií:

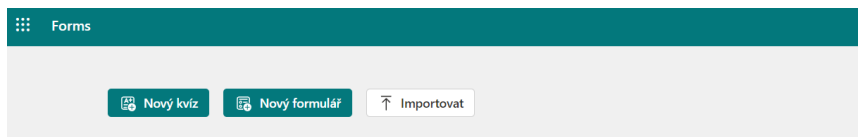
- Pokud máme lichý počet hodnot, je mediánem hodnota, která je uprostřed seřazených dat. Např. pokud máme známky 1 1 2 **3** 3 4 4, mediánem je hodnota 3, která je přesně uprostřed.
- V případě, že máme lichý počet hodnot, najdeme uprostřed hodnoty dvě, v tomto případě je mediánem jejich aritmetický průměr. Např. pokud máme známky 1 1 2 **2** **3** 3 4 4, mediánem je hodnota  $(2 + 3) : 2 = 2,5$

## Zpracování dat na počítači

### Příprava dat

V první části budeme potřebovat data, která budeme zpracovávat. Jsou dvě možnosti. Použití vzorového souboru (odkaz v úvodu) nebo vlastní příprava dat, kdy si žáci připraví otázky a sesbírají data např. pomocí nástroje Microsoft Forms nebo Google Forms. V tomto případě je nutné připravit dotazník (viz obrázek) a následně sesbírat odpovědi a získat tabulku s daty. U tohoto nástroje je možné také ukázat, jak jej využít pro automatické zpracování dat.

Pro vytvoření nové ankety otevřeme internetový prohlížeč a zadáme adresu <http://forms.microsoft.com>, kde se přihlásíme školním účtem a klikneme na tlačítko „Nový formulář“.

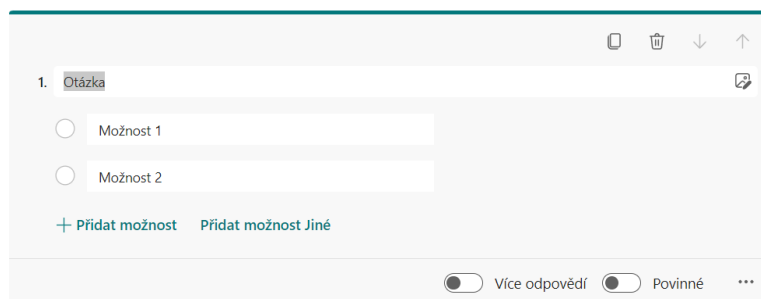


Následně již můžeme vytvářet jednotlivé otázky. Vhodnější je použít otázky s výběrem možností, ale je možno vložit i otevřenou otázku, ta však bude náročnější na vyhodnocení.

Před tvorbou otázek nejprve přejmenujeme formulář (kliknutím na text „Formulář bez názvu“ a přepsáním).

Pod názvem formuláře klikneme na tlačítko „Přidat nový“, kde následně vybereme požadovaný typ otázky (buď volba, nebo text).

V případě volby zadáme otázku a následně jednotlivé možnosti (případně kliknutím na „Přidat možnost“ můžeme přidat další odpovědi).



V případě textové odpovědi lze na otázku zadávat libovolnou odpověď, můžeme však nastavit omezení (např. aby respondent mohl zadat pouze číselnou odpověď) kliknutím na tři tečky v pravém dolním rohu otázky a výběrem „Omezení“.

Následně anketu nasdílíme respondentům. Sdílení provedeme kliknutím na tlačítko „Shromážďovat odpovědi“, kde najdeme odkaz pro sdílení. Před kopírováním odkazu je však vhodné vybrat možnost „Odpovědět může kdokoliv“, aby se respondent nemusel přihlašovat pro vyplnění anket.

Výsledná anketa potom může vypadat následovně:

Otázky

Odpovědi

## Anketa

1. Jak dlouho mi trvá dojíždění do školy (zaokrouhlete na desítky minut): \*

Hodnotou musí být číslo.

2. Počet sourozenců: \*

0

1

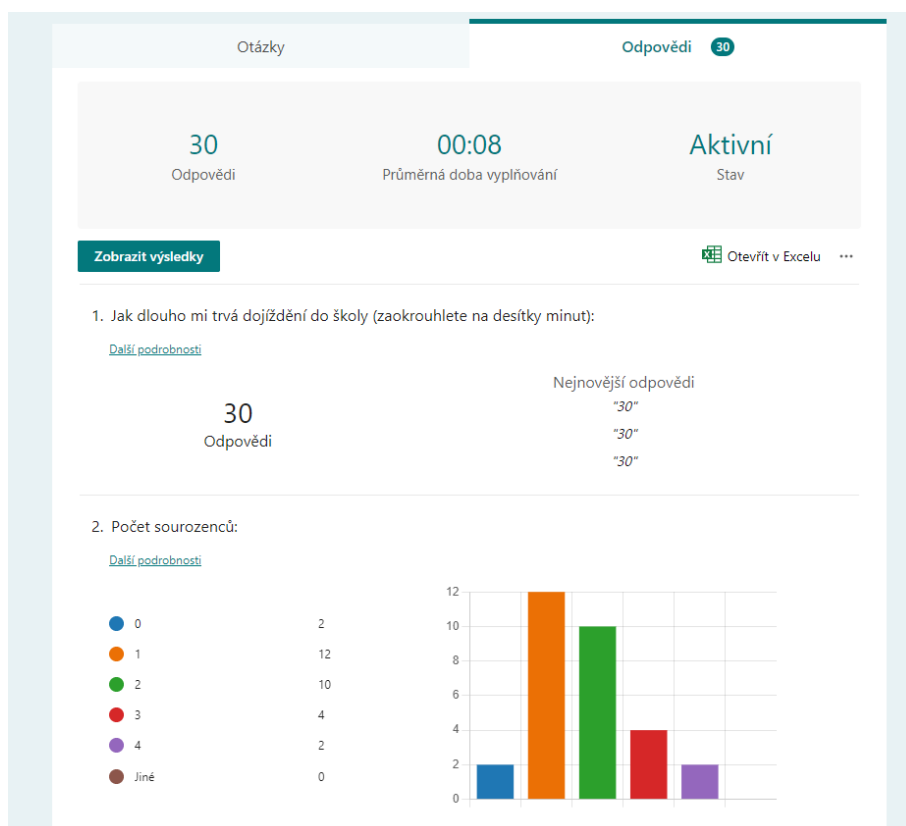
2

3

4

Jiné

Následně je potřeba nasbírat data, po nasbírání dat lze ukázat vyhodnocení ankety (záložka odpovědi) a stažení tabulky do počítače.



Pro stažení tabulky ve formátu MS Excel klikneme na tlačítko „Otevřít v Excelu“ v horní části stránky s výsledky. Vygeneruje se nám soubor XLSX, který můžeme otevřít v aplikaci MS Excel nebo nahrát do nástroje MS Excel Online nebo Google Spreadsheet.

ID	Počáteční čas	Čas dokončení	E-mail	Jméno	Jak dlouho mi trvá	Počet sourozenců	Věk otce	Věk matky
1	7.26.22 11:24:05	7.26.22 11:24:15	anonymous		20	2	35	32
2	7.26.22 11:24:21	7.26.22 11:24:29	anonymous		30	1	30	30
3	7.26.22 11:25:08	7.26.22 11:25:18	anonymous		30	1	40	42
4	7.26.22 11:25:20	7.26.22 11:25:30	anonymous		10	0	38	35
5	7.26.22 11:25:31	7.26.22 11:25:38	anonymous		50	3	35	36
6	7.26.22 11:25:39	7.26.22 11:25:49	anonymous		50	4	45	40
7	7.26.22 11:25:50	7.26.22 11:26:12	anonymous		40	0	30	29
8	7.26.22 11:26:14	7.26.22 11:26:22	anonymous		40	3	38	38
9	7.26.22 11:26:23	7.26.22 11:26:32	anonymous		30	2	35	33
10	7.26.22 11:26:33	7.26.22 11:26:39	anonymous		45	2	38	42
11	7.26.22 11:27:04	7.26.22 11:27:11	anonymous		40	2	40	40
12	7.26.22 11:27:12	7.26.22 11:27:19	anonymous		30	4	50	45
13	7.26.22 11:27:20	7.26.22 11:27:26	anonymous		50	1	35	35
14	7.26.22 11:27:26	7.26.22 11:27:33	anonymous		40	3	35	33
15	7.26.22 11:27:36	7.26.22 11:27:44	anonymous		40	1	39	41
16	7.26.22 11:27:45	7.26.22 11:27:51	anonymous		30	2	35	35

## Zpracování dat v tabulkovém procesoru

Nyní máme získána statistická data, která můžeme zpracovat. Nejprve si připravíme tabulky četností. Před tím přidáme sloupce pro zaokrouhlení věku otce a matky na desítky (klikneme pravým tlačítkem na záhlaví sloupce napravo od věku otce a vybereme „Přidat sloupec“) – pro zjednodušení tabulky četností. Zaokrouhlení provedeme funkcí Zaokrouhlit s následujícím zápisem: =ZAOKROUHLIT (H2; -1)

Jak je vidět, funkce zaokrouhlit má dva parametry, první je adresa buňky s hodnotou a druhým je počet míst. Hodnota „-1“ zaokrouhlí číslo na desítky (naopak 1 by zaokrouhlilo na desetiny). Po zapsání vzorce do první buňky vyplníme celý sloupec (klikneme na tečku v pravém dolním rohu buňky a tažením vyplníme do dalších buněk).

Věk otce:	Otec	Věk matky:
35		32
30		
40		
38		35
35		36
45		40
30		30

Stejný postup potom následně zopakujeme pro věk matky.

Nyní již máme tabulku upravenou a můžeme vyrobit tabulky četností. Tlačítkem „+“ v dolní části okna přidáme nový list, na kterém připravíme jednotlivé tabulky.



V následujícím příkladu je ukázána tabulka pro počet sourozenců.



Pro výpočet jednotlivých četností využijeme funkci `COUNTIF (oblast; hodnota)`, tedy v našem případě `=COUNTIF (Sheet1!$G$2:$G$31;B2)`. Prvním parametrem ve funkci je rozsah buněk, ve kterých hledáme (ve tvaru `JménoListu!PrvníBuňka:PosledníBuňka`). Je však možné tyto buňky i vybrat myší. Znak dolaru je použit, aby se adresa neměnila při vyplňování do dalších buněk tažením myší. Druhým parametrem je buňka s hodnotou, kterou hledáme (tj. hodnota z prvního řádku, jejíž četnost počítáme).

	A	B	C	D	E	F
1	Počet sourozenců	0	1	2	3	4
2	Četnost	2	12	10	4	2
3						

Následně můžeme vypočítat aritmetický průměr, modus a medián. Zde využijeme vzorce `PRŮMĚR ()` (v případě anglické verze tabulkového procesoru `AVERAGE ()`), `MODE ()` a `MEDIAN ()`. Do závorky následně vybereme sloupec z původní tabulky, tedy u průměru bude vzorec v buňce `=PRŮMĚR (Sheet1!G2:G31)`. Stejný zápis potom bude použit i u zbylých dvou funkcí – tedy modus a medián.

V některých verzích MS Excelu se při výběru buněk nahrazuje rozsah hodnot názvem sloupce tabulky. Takto vytvořený vzorec nemusíme přepisovat, ale můžeme ponechat v této podobě. Vzorec pro průměr v tomto případě může vypadat například takto: `=PRŮMĚR (Table1 [Počet sourozenců: ])`

Zbývá doplnit již jen diagram. V tabulce četností vybereme řádek s četnostmi a na záložce „Vložení“ vybereme vhodný kruhový nebo sloupcový diagram. U diagramu bude dále potřeba doplnit nadpis a přidat popis hodnot. Popis hodnot přidáme tak, že klikneme pravým tlačítkem na graf, zvolíme vybrat data a upravíme popisky vodorovné osy, kde kliknutím na tlačítko „Upravit“ vybereme první řádek tabulky četností.

Vytvoření zbylých tabulek a grafů je ukázáno v příloženém výsledném souboru.

## Zkušenosti s použitím materiálu

V případě tvorby vlastních dat (např. metodou formulářového výzkumu) je aktivita časově náročnější na realizaci, avšak bližší žákům v tom, že pracují s reálnými daty.

## Další možnosti využití materiálu

Zatím žádné