

4.5.2 Statistika

Předpoklady: 040501

Pedagogická poznámka: V hodině pracujeme s daty, které do tabulky vyplnili žáci v minulé hodině.

Pohlaví	Znám BM	Cesta do školy (minuty)	Kapesné	Známka chemie	Počet sourozenců	Oblíbená barva
muž	ano	30	700	2	1	černá
muž	ano	50	500	4	2	zelená
muž	ano	20	200	5	2	modrá
muž	ne	30	200	3	2	zelená
žena	ano	10	0	3	0	hnědá
žena	ne	10	150	2	2	červená
žena	ano	30	300	1	1	oranžová
žena	ano	30	700	2	2	modrá
žena	ano	10	200	2	1	černá
muž	ano	5	0	4	1	modrá
muž	ano	5	0	4	2	Fialová
muž	ano	5	1500	2	3	Modrá

Smyslem statistiky není hledat jména ke konkrétním řádkům, naopak hledání obecných charakteristik a předpovědí \Rightarrow u zjišťovaných údajů se neuvádí identifikátor dotazovaného a jde o takzvaná anonymizovaná data.

Př. 1: Tabulka, kterou vyplnili někteří v minulé hodině, se nazývá statistický soubor. Vypiš znaky, které jsme zjišťovali. Které z nich jsou kvalitativní? Které kvantitativní? Kolik má soubor statistických jednotek? Jaký je jeho rozsah?

Zjišťované znaky:

pohlaví, znalost BM, délka cesty do školy, výše kapesného, známka z chemie, počet sourozenců, oblíbená barva.

Kvantitativní znaky (vyjádřitelné v číslech, měřitelné): délka cesty do školy, výše kapesného, známka z chemie, počet sourozenců.

Kvalitativní znaky (nečíselné, o vlastnosti): pohlaví, znalost BM, oblíbená barva.

Soubor má 12 statistických jednotek, jeho rozsah je tedy 12.

Pedagogická poznámka: Neříkáme si před kontrolou, co znamená kvantitativní znak a kvalitativní znak. Pouze si připomínáme význam slov (kvalita – vlastnost, stav, jakost), kvantita (množství, počet).

Př. 2: Prohlédni si údaje v tabulce a odpověz na následující otázky.
a) Účastnilo se výzkumu více mužů nebo žen?
b) Jaká je nejvyšší a nejnižší částka měsíčního kapesného?
c) Kolika účastníkům trvá cesta do školy déle než 10 minut?
d) Kolik účastníků má nejraději modrou nebo zelenou barvu?

- e) Kolik procent účastníků má z chemie lepší známku než 4?
f) Jaká je pravděpodobnost, že náhodně vybraný účastník výzkumu zná BM?

a) Účastnilo se výzkumu více mužů nebo žen?

Výzkumu se účastnilo 7 mužů a 5 žen, tedy více mužů.

b) Jaká je nejvyšší a nejnižší částka měsíčního kapesného?

Nejnižší částkou kapesného je 0 Kč, nejvyšší 1500 Kč.

c) Kolika účastníkům trvá cesta do školy déle než 10 minut?

6 účastníkům trvá cesta do školy více než 10 minut.

d) Kolik účastníků má nejrady modrou nebo zelenou barvu?

Modrou nebo zelenou barvu má nejrady 6 účastníků (2 zelenou, 4 modrou).

e) Kolik procent účastníků má z chemie lepší známku než 4?

Lepší známku než 4 má 8 účastníků $\Rightarrow \frac{8}{12} = 0,667 \Rightarrow 66,7\%$ účastníků.

f) Jaká je pravděpodobnost, že náhodně vybraný účastník výzkumu zná BM?

BM neznají pouze 2 účastníci \Rightarrow zná ho 10 účastníků z 12 \Rightarrow pravděpodobnost, že ho zná

náhodně zvolený účastník je $\frac{10}{12} = 0,833 \Rightarrow 83,3\%$.

Př. 3: Naše tabulka má pouze 12 statistických jednotek, její rozsah je velmi malý, naprostá většina statistických výzkumů pracuje s podstatně většími počty účastníků (s počtem účastníků totiž při jejich správné volbě dochází k zpřesňování výsledků). Plnění většiny předchozích úkolů v tabulce, která by měla tisíce řádků, by bylo velmi náročné. Navrhni takové zpracování údajů o jednom znaku v tabulce (například údajů o známce z chemie), při kterém bychom mohli snadno odpovídat na otázky z předchozího příkladu i v případě, že by soubor obsahoval tisíce statistických jednotek (zpracování, které by usnadňovalo tvorbu odpovědí otázky, ale neztratilo by žádnou podstatnou informaci).

Při zodpovídání otázek v předchozím příkladu jsme potřebovali vědět, kolik účastníků dostalo jakou známku \Rightarrow pokud bychom na začátku spočetli počty lidí s jednotlivými známkami, mohli bychom odpovídat na otázky stejně rychle bez ohledu na počty účastníků.

Pro usnadnění zpracování souborů s velkým počtem účastníků můžeme sestavit tabulky, které pro každou možnou hodnotu znaku udávají, kolikrát nastal – tabulku četností.

Př. 4: Urči četnost:

- a) výsledku muž u znaku Pohlaví,
b) výsledku 200 Kč u znaku Kapesné,
c) výsledku 25 min u znaku Cesta do školy [min].

a) výsledku muž u znaku Pohlaví

Anketu vyplnilo 7 mužů \Rightarrow četnost výsledku muž u znaku Pohlaví je 7.

b) výsledku 200 Kč u znaku Kapesné

Kapesné 200 Kč dostávají 3 účastníci \Rightarrow četnost výsledku 200 Kč u znaku Kapesné je 3.

c) výsledku 25 min u znaku Cesta do školy [min].

Hodnotu 25 minut nikdo u délky cesty do školy neuvedl \Rightarrow četnost výsledku 25 minut u znaku Cesta do školy [min] je 0.

Př. 5: Sestav tabulku četností pro znak:

- a) pohlaví, b) známka z chemie, c) oblíbená barva.
Co platí pro četnosti libovolného znaku?

a) pohlaví

Pohlaví	muž	žena
Četnost	7	5

b) známka z chemie

Známka z chemie	1	2	3	4	5
Četnost	1	5	2	3	1

c) oblíbená barva

Barva	černá	zelená	modrá	hnědá	červená	oranžová	fialová
Četnost	2	2	4	1	1	1	1

Ve všech případech se součet všech četností znaku rovná 12 (rozsah souboru).

Součet všech četností libovolného znaku se rovná počtu statistických jednotek.

Př. 6: V jaké situaci neumožňují četnosti rychlé porovnávání? Navrhni řešení.

Pokud mají dva soubory rozdílný rozsah, nemůžeme porovnávat hodnoty znaků podle četnosti (1 člověk s jedničkou z chemie z 12 účastníků není to samé jako 1 člověk s jedničkou z chemie z 30 účastníků).

Řešením by bylo spočítat poměr četnosti a rozsahu souboru.

Poměr četnosti a rozsahu souboru se nazývá relativní četnost.

Př. 7: Dopln tabulku četností známky z chemie o řádku s relativními četnostmi. Sestav tabulku četností a relativních četností pro počet sourozenců.
Co platí pro relativní četnosti libovolného znaku?

Známka z chemie

Známka z chemie	1	2	3	4	5
Četnost	1	5	2	3	1
Relativní četnost	0,0833	0,417	0,137	0,25	0,0833

Počet sourozenců

Počet sourozenců	0	1	2	3
Četnost	1	4	6	1

Relativní četnost	0,0833	0,333	0,50	0,0833
-------------------	--------	-------	------	--------

Součet relativních četností je vždy 1.

Součet všech relativních četností libovolného znaku se rovná 1.

Relativní četnosti se často znázorňují pomocí koláčových grafů, na kterých plocha jednotlivých částí kruhu ve tvaru výseče odpovídá relativním četnostem.

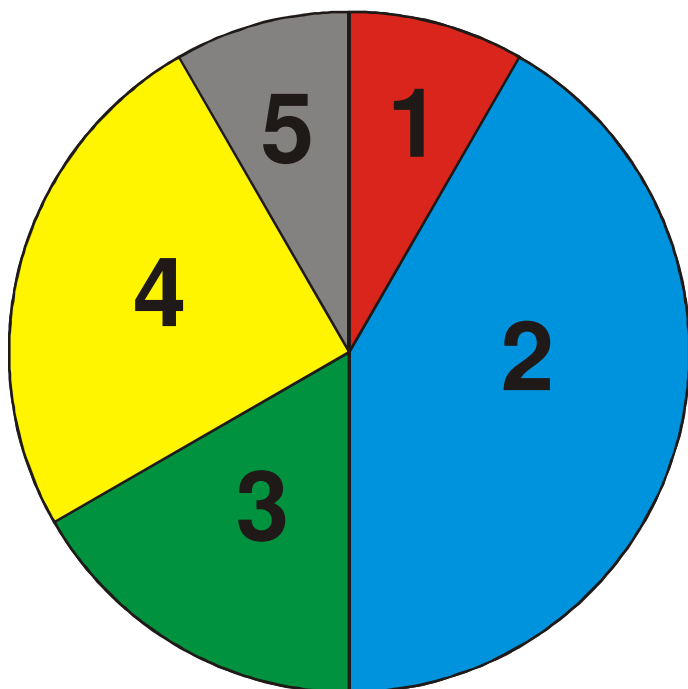
Př. 8: Narýsuj koláčový graf pro známky z chemie.

Pro každou hodnotu musíme vypočítat středový úhel výseče.

Celkem 12 známek ... 360°
 1 známka ... $360 : 12 = 30$

Do tabulky můžeme doplnit další řádek s hodnotami úhlů.

Známka z chemie	1	2	3	4	5
Četnost	1	5	2	3	1
Relativní četnost	0,0833	0,417	0,137	0,25	0,0833
Středový úhel výseče	30	150	60	90	30



Shrnutí: Statistické informace můžeme úsporně vyjadřovat pomocí četnosti nebo relativní četnosti.