

Faktoriály a kombinační čísla

6. kapitola. Několik otázek z matematické statistiky

In: Jiří Sedláček (author): Faktoriály a kombinační čísla. (Czech).
Praha: Mladá fronta, 1985. pp. 72–81.

Persistent URL: <http://dml.cz/dmlcz/404118>

Terms of use:

© Jiří Sedláček, 1964

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>

NĚKOLIK OTÁZEK Z MATEMATICKÉ STATISTIKY

Představme si, že máme vyšetřit výšku a váhu patnáctiletých chlapců v naší republice. Předmětem zkoumání je tedy soubor jedinců, který obsahuje dosti značný počet prvků. Nebylo by proto hospodárné, kdybychom k získání výsledku chtěli skutečně změřit a zvážit všechny chlapce tohoto věku. Nebylo by to možné třeba ani z toho důvodu, že by tento experiment trval příliš dlouho, a my máme předepsáno, abychom výsledný údaj získali v době co nejkratší. Jak budeme postupovat? Místo celého velmi početného základního souboru si vezmeme jen určitý výběr, tj. skupinu patnáctiletých chlapců, která není příliš početná a jež je vybrána podle určitých zásad z celého souboru zkoumaných jedinců. Výšku, resp. váhu, pak měříme jen u tohoto výběru. Zmíněné zásady spočívají zejména v tom, že žádáme, aby výběr byl náhodný a co možná nejreprezentativnější, tj. aby v určitém smyslu vzhledem ke zkoumanému znaku reprezentoval celý základní soubor.

Měli jsme tedy základní soubor, ve kterém se měl měřit některý kvantitativní znak (výška, váha), a přešli jsme k menšímu výběru. Jak se liší třeba aritmetický průměr v základním souboru od aritmetického průměru ve výběru a jaký je vztah mezi těmito čísly? Tomuto zkoumání věnujeme další úvahu. Přitom odhlédneme od konkrétního významu kvantitativního znaku a budeme pracovat prostě s čísly.

Nechť je tedy dán základní soubor, který má N prvků. Na každém z nich je naměřen kvantitativní znak x_i , takže dostáváme skupinu N čísel $x_1, x_2, x_3, \dots, x_N$. Kromě toho uvažujeme výběr, který má r prvků ($1 \leq r \leq N$). Takových výběrů existuje celkem $\binom{N}{r}$. Číslům N , resp. r , říkáme někdy též *rozsah* základního souboru, resp. výběru. Označme \bar{x} aritmetický průměr v základním souboru; je tedy

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N}. \quad (1)$$

Dále označme $\bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(s)}$ po řadě všechny aritmetické průměry ve výběrech rozsahu r ; je tedy $s = \binom{N}{r}$. Jak velký je aritmetický průměr z čísel $\bar{x}^{(i)}$? Odpověď najdeme v dalším příkladě.

Příklad 37. Určete aritmetický průměr z výběrových aritmetických průměrů $\bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(s)}$.

Řešení. Předmětem naší úvahy je výraz

$$\frac{\bar{x}^{(1)} + \bar{x}^{(2)} + \dots + \bar{x}^{(s)}}{s}.$$

Rozšíříme-li tento zlomek číslem r , vychází

$$\frac{r\bar{x}^{(1)} + r\bar{x}^{(2)} + \dots + r\bar{x}^{(s)}}{rs}. \quad (2)$$

Každé z čísel $r\bar{x}^{(i)}$ si můžeme představit jako součet některých čísel x_j , přičemž sčítanců je vždy r . V čitateli zlomku (2) se tedy vyskytují všechna čísla x_j — případně vícekrát. Kolikrát je tu číslo x_1 ? Musíme určit, kolikrát se číslo x_1 vyskytuje ve výběrech rozsahu r ze základního

souboru rozsahu N . Každý výběr obsahující prvek x_1 dostaneme tak, že k prvku x_1 přidáme libovolnou skupinu o $r - 1$ prvcích; je tedy výběrů obsahujících prvek x_1 právě tolik, kolik kombinací $(r - 1)$ -ní třídy lze vytvořit z $N - 1$ prvků, tj. $\binom{N-1}{r-1}$. Číslo x_1 je tedy v čitateli zlomku (2) právě $\binom{N-1}{r-1}$ - krát a totéž platí ovšem i o každém z dalších čísel x_2, x_3, \dots, x_N .

Čitatele zlomku (2) tedy můžeme vyjádřit ve tvaru

$$\binom{N-1}{r-1} \cdot (x_1 + x_2 + \dots + x_N).$$

Podle (1) má proto čítec tvar

$$\begin{aligned} \binom{N-1}{r-1} \cdot N \cdot \bar{x} &= \frac{(N-1)!}{(r-1)!(N-r)!} \cdot N \cdot \bar{x} = \\ &= \frac{N!}{(r-1)!(N-r)!} \cdot \bar{x}. \end{aligned}$$

Jmenovatel zlomku (2) je

$$r \cdot s = r \cdot \binom{N}{r} = r \cdot \frac{N!}{r!(N-r)!} = \frac{N!}{(r-1)!(N-r)!}.$$

Dělením tedy dostáváme podíl \bar{x} .

Odpověď. Aritmetický průměr ze všech výběrových aritmetických průměrů se rovná aritmetickému průměru v základním souboru.

Další důležitou charakteristikou, která se vyskytuje v matematické statistice, je tzv. *směrodatná odchylka* σ . Jsou-li dána čísla x_1, x_2, \dots, x_N , pak σ je definováno vztahem

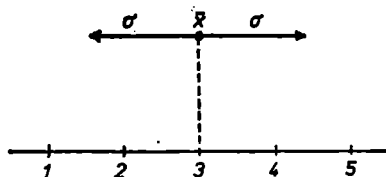
$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}}$$

Směrodatná odchylka σ je mírou, jak se hodnoty x_i „roztýkají“ nebo „kupí“ kolem aritmetického průměru \bar{x} ; to je vidět z toho, že σ závisí na odchylkách $x_i - \bar{x}$. Jsou-li odchylky malé, je σ malé, jsou-li odchylky velké, je σ velké.

Uvedeme si numerický příklad. Čísla 1, 2, 3, 4, 5 mají aritmetický průměr 3, takže

$$\sigma = \sqrt{\frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{5}} = \sqrt{2}$$

Bývá zvykem znázorňovat daná čísla na ose číselné a připojit tam též směrodatnou odchylku. Pro náš případ je toto znázornění vidět na obr. 5.



Obr. 5

Vraťme se ještě k výběrovým aritmetickým průměrům a studujme otázku, jak se tato čísla $\bar{x}^{(i)}$ „kupí“ kolem aritmetického průměru \bar{x} . Srovnáme tedy směrodatnou odchylku všech těchto výběrových aritmetických průměrů se směrodatnou odchylkou čísel x_i v základním souboru. Tomuto srovnání je věnován další příklad.

Příklad 38. Znáte-li rozsah základního souboru a rozsah výběrů (čísla N a r) a směrodatnou odchylku σ v zá-

kladním souboru, vypočtete směrodatnou odchylku všech výběrových aritmetických průměrů $\bar{x}^{(i)}$.

Řešení. Případ $r = 1$ je triviální, neboť pak je hledaná směrodatná odchylka rovna přímo číslu σ . V dalším textu tedy uvažujme jen případ $r \geq 2$. Použijme výsledku z předcházejícího příkladu, podle něhož čísla $\bar{x}^{(i)}$ mají aritmetický průměr \bar{x} . Jejich směrodatnou odchylku σ vypočteme tedy podle vzorce

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

Budeme zatím pracovat jen se zlomkem pod odmocninou, který rozšíříme číslem r^2 . Dostáváme tak zlomek

$$\frac{(r\bar{x}^{(1)} - r\bar{x})^2 + (r\bar{x}^{(2)} - r\bar{x})^2 + \dots + (r\bar{x}^{(s)} - r\bar{x})^2}{r^2 s}. \quad (3)$$

Součin $r\bar{x}^{(i)}$ si zase můžeme představit jako součet některých čísel x_i , takže rozdíl $r\bar{x}^{(i)} - r\bar{x}$ lze převést na tvar

$$(x_a - \bar{x}) + (x_b - \bar{x}) + \dots + (x_t - \bar{x}),$$

kteřý se týká jednoho výběru. Čitatele zlomku (3) lze tedy vyjádřit jako součet, v němž jsou jednak sčítanci tvaru $(x_a - \bar{x})^2$, jednak sčítanci tvaru $2(x_a - \bar{x})(x_b - \bar{x})$ (při $a \neq b$). Kolikrát je zde sčítanec $(x_1 - \bar{x})^2$? Stejná úvaha jako v předcházejícím příkladě nás vede k výsledku $\binom{N-1}{r-1}$. Také každý ze sčítanců

$$(x_2 - \bar{x})^2, (x_3 - \bar{x})^2, \dots, (x_N - \bar{x})^2$$

má v čitateli zlomku (3) stejnou násobnost.

Kolikrát se v čitateli zlomku (3) vyskytuje sčítanec tvaru $2(x_1 - \bar{x})(x_2 - \bar{x})$? Jsme tu vedeni ke kombinač-

nímu číslu $\binom{N-2}{r-2}$, které se ovšem týká i dalších sčítanců uvedeného tvaru.

Čitatel zlomku (3) je tedy součtem dvou výrazů; první má tvar

$$A = \binom{N-1}{r-1} \cdot [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2],$$

druhý má tvar

$$B = \binom{N-2}{r-2} \cdot [2(x_1 - \bar{x})(x_2 - \bar{x}) + \dots + 2(x_{N-1} - \bar{x})(x_N - \bar{x})].$$

Víme, že platí

$$\binom{N-1}{r-1} = \binom{N-2}{r-2} + \binom{N-2}{r-1},$$

takže v součtu $A + B$ si můžeme nejprve všimnout jen těch sčítanců, jež mají u sebe jako koeficient kombinační číslo $\binom{N-2}{r-2}$. Je vidět, že součet těchto sčítanců lze vyjádřit jako

$$[(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_N - \bar{x})]^2,$$

což podle vzorce (1) je zřejmě rovno nule. Součet $A + B$ v čitateli zlomku (3) se tedy převádí na tvar

$$\binom{N-2}{r-1} \cdot [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2].$$

Podle definice směrodatné odchylky lze tento výsledek vyjádřit jako

$$\binom{N-2}{r-1} \cdot N\sigma^2 = \frac{(N-2)!}{(r-1)!(N-r-1)!} \cdot N\sigma^2.$$

Upravujme ještě jmenovatele zlomku (3). Máme

$$r^2 s = r^2 \binom{N}{r} = r^2 \frac{N!}{r!(N-r)!} = r \frac{N!}{(r-1)!(N-r)!}.$$

Zlomek (3) se tedy rovná

$$\frac{N-r}{r(N-1)} \sigma^2.$$

Odpověď. Směrodatnou odchylku všech výběrových aritmetických průměrů určuje vzorec

$$\bar{\sigma} = \sigma \sqrt{\frac{N-r}{r(N-1)}}.$$

Výsledek, ke kterému jsme právě dospěli, se zdá na první pohled trochu složitý, avšak pro praktické případy jej můžeme ještě zjednodušit. Obvykle bývá totiž číslo N velmi veliké a číslo r poměrně malé. V takovém případě můžeme ještě upravovat zlomek, který se vyskytuje ve výsledném vzorci pro $\bar{\sigma}$. Platí

$$\frac{N-r}{r(N-1)} = \frac{1 - \frac{r}{N}}{r - \frac{r}{N}},$$

přičemž zlomek $\frac{r}{N}$ je „poměrně malý“. V praktických příkladech můžeme dokonce předpokládat, že se rovná nule; tím dostáváme přibližnou rovnost

$$\frac{N-r}{r(N-1)} \doteq \frac{1}{r}$$

a pro směrodatnou odchylku přibližný vzorec

$$\bar{\sigma} \doteq \frac{\sigma}{\sqrt{r}}.$$

Zamysleme se ještě nad tím, jaký význam má výsledek dosažený v příkladě 38. Zjistili jsme, že se výběrové aritmetické průměry „hromadí“ kolem průměru \bar{x} mnohem více než původně uvažovaná čísla x_1, x_2, \dots, x_N . Směrodatná odchylka nás poučuje o tom, jak jsou čísla na číselné ose roztroušena. Bereme-li z velkého základního souboru např. výběry rozsahu 25, pak směrodatná odchylka $\bar{\sigma}$ je zhruba pětinou směrodatné odchylky σ . Abychom si mohli lépe představit, jak se výběrové průměry hromadí kolem \bar{x} , vypočteme si ještě jeden numerický příklad. V něm jsme zvolili jen malá čísla, protože při větších rozsazích a větších číslech x_i velmi rychle vzrůstají technické potíže s numerickým výpočtem.

Příklad 39. V základním souboru $\{1, 2, 3, 4, 5\}$ sestrojte všechny výběry rozsahu 3. Vypočtete výběrové aritmetické průměry a jejich směrodatnou odchylku. Znázorněte výsledek graficky.

Řešení. Víme už (viz str. 75), že $\bar{x} = 3$, $\sigma = \sqrt{2} \doteq 1,41$. Další výpočet můžeme upravit do tabulky (viz str. 80).

Sečteme čísla v posledním sloupci a dostáváme výsledek $3\frac{1}{3}$. Protože je celkem 10 výběrů, dělíme

$$3\frac{1}{3} : 10 = \frac{1}{3}$$

a pak odmocníme. Je tedy

$$\bar{\sigma} = \sqrt{\frac{1}{3}} = \frac{1}{3}\sqrt{3} \doteq 0,58.$$

výběr	$\bar{x}^{(4)}$	$\bar{x}^{(4)} - \bar{x}$	$(\bar{x}^{(4)} - \bar{x})^2$
1, 2, 3	2	-1	1
1, 2, 4	$2\frac{1}{3}$	$-\frac{2}{3}$	$\frac{4}{9}$
1, 2, 5	$2\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{9}$
1, 3, 4	$2\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{9}$
1, 3, 5	3	0	0
1, 4, 5	$3\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{9}$
2, 3, 4,	3	0	0
2, 3, 5	$3\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{9}$
2, 4, 5	$3\frac{2}{3}$	$\frac{2}{3}$	$\frac{4}{9}$
3, 4, 5	4	1	1

To je ve shodě se vzorcem pro $\bar{\sigma}$, který jsme odvodili výše. Podle něho je totiž

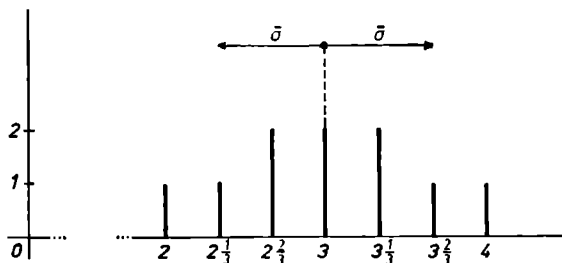
$$\bar{\sigma} = \sqrt{2} \cdot \sqrt{\frac{5-3}{3 \cdot (5-1)}} = \sqrt{\frac{1}{3}}.$$

Přibližný vzorec pro $\bar{\sigma}$ zde ovšem nemůžeme uplatnit, neboť je

$$\frac{r}{N} = \frac{3}{5} = 0,6,$$

což nelze považovat za číslo „skoro rovné“ nule.

Ještě ke grafickému znázornění. Protože se zde některé výběrové průměry $\bar{x}^{(i)}$ opakují, použijeme ke znázornění tzv. *sloupkového diagramu*. Na ose číselné u čísla $\bar{x}^{(i)}$ je připojen sloupek, jehož délka uvádí, kolikrát se toto číslo v úvaze vyskytuje. Výsledek je vidět na obr. 6, kde



Obr. 6

jsme také znázornili směrodatnou odchylku $\bar{\sigma}$. Srovnajte tento sloupkový diagram s obr. 5, který se týká základního souboru.*)

Úlohy

37. Graficky znázorněte přibližný vzorec

$$\bar{\sigma} \doteq \frac{\sigma}{\sqrt{r}};$$

přítom volte $\sigma = 10$, na vodorovnou osu nanášejte r a na svislou $\bar{\sigma}$.

*) Příklady, které jsme zde řešili, patří do oddílu matematické statistiky, nazývaného *teorie výběrových šetření* (z konečných souborů).