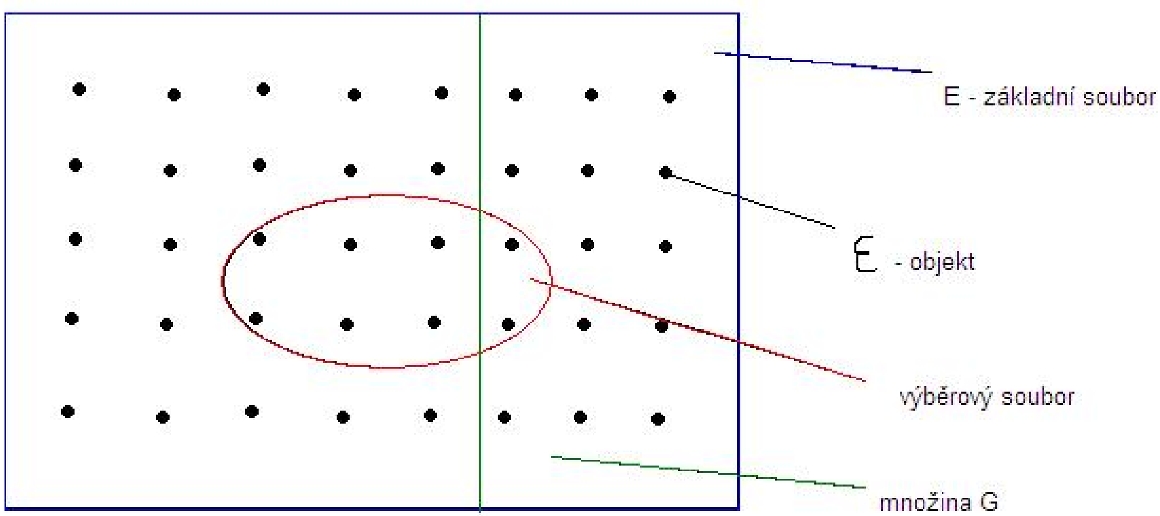
Základní, výběrový a datový soubor

Základním souborem rozumíme libovolnou neprázdnou množinu E. Prvky množiny E značíme ε a nazýváme je objekty. Libovolnou neprázdnou podmnožinu $\{\varepsilon_1, \ldots, \varepsilon_n\}$ základního souboru E nazýváme výběrový soubor rozsahu n. Je-li množina $G \subseteq E$, pak symbolem N(G) rozumíme absolutní četnost množiny G ve výběrovém souboru, tj. počet těch objektů množiny G, které patří do výběrového souboru. Relativní četnost množiny G ve výběrovém souboru zavedeme vztahem $p(G) = \frac{N(G)}{n}$.

Ilustrace



Příklad: Základním souborem E je množina všech ekonomicky zaměřených studentů 1. ročníku českých vysokých škol. Množina G_1 je tvořena těmi studenty, kteří uspěli v prvním zkušebním termínu z matematiky a množina G_2 obsahuje ty studenty, kteří uspěli v prvním zkušebním termínu z angličtiny. Ze základního souboru bylo náhodně vybráno 20 studentů, kteří tvoří výběrový soubor $\{\varepsilon_1, ..., \varepsilon_{20}\}$. Z těchto 20 studentů 12 uspělo v matematice, 15 v angličtině a 11 v obou předmětech. Zapište absolutní a relativní četnosti úspěšných matematiků, angličtinářů a oboustranně úspěšných studentů.

Řešení:

$$N(G_1) = 12, N(G_2) = 15, N(G_1 \cap G_2) = 11, n = 20, p(G_1) = \frac{12}{20} = 0{,}6, p(G_2) = \frac{15}{20} = 0{,}75,$$

$$p(G_1 \cap G_2) = \frac{11}{20} = 0{,}55$$

Vidíme, že úspěšných matematiků je 60%, angličtinářů 75% a oboustranně úspěšných studentů jen 55%.

Vlastnosti relativní četnosti: Relativní četnost má následujících 12 vlastností, které jsou obdobné vlastnostem procent.

- $p(\emptyset) = 0$
- $p(G) \geq 0$ (nezápornost)
- $p(G) \leq 1$
- $p(G_1 \cup G_2) + p(G_1 \cap G_2) = p(G_1) + p(G_2)$
- $1 + p(G_1 \cap G_2) \geq p(G_1) + p(G_2)$
- $p(G_1 \cup G_2) + 0 \leq p(G_1) + p(G_2)$ (subaditivita)
- $G_1 \cap G_2 = \emptyset \Rightarrow p(G_1 \cup G_2) = p(G_1) + p(G_2)$ (aditivita)
- $p(G_2 \setminus G_1) = p(G_2) - p(G_1 \cap G_2)$
- $G_1 \subseteq G_2 \Rightarrow p(G_2 \setminus G_1) = p(G_2) - p(G_1)$ (subtraktivita)
- $G_1 \subseteq G_2 \Rightarrow p(G_1) \leq p(G_2)$ (monotonie)
- $p(E) = 1$ (normovanost)
- $p(G) + p(\overline{G}) = 1$ (komplementarita)

Pojem podmíněné relativní četnosti: Pokud se v daném základním souboru zajímáme o dvě podmnožiny, můžeme zavést pojem podmíněné relativní četnosti jedné podmnožiny v daném výběrovém souboru za předpokladu, že objekt pochází z druhé podmnožiny.
Nechť E je základní soubor, G_1, G_2 jeho podmnožiny, $\{\varepsilon_1, ..., \varepsilon_n\}$ výběrový soubor. Definujeme
podmíněnou relativní četnost množiny G_1 ve výběrovém souboru za předpokladu G_2:

$$p(G_1/G_2) = \frac{N(G_1 \cap G_2)}{N(G_2)} = \frac{p(G_1 \cap G_2)}{p(G_2)} \text{ a}$$

podmíněnou relativní četnost G_2 ve výběrovém souboru za předpokladu G_1:

$$p(G_2/G_1) = \frac{N(G_1 \cap G_2)}{N(G_1)} = \frac{p(G_1 \cap G_2)}{p(G_1)}.$$

Příklad: Pro údaje z příkladu o studentech vypočtěte podmíněnou relativní četnost úspěšných matematiků mezi úspěšnými angličtináři a podmíněnou relativní četnost úspěšných angličtinářů mezi úspěšnými matematiky.
(Připomínáme, že z 20 studentů 12 uspělo v matematice, 15 v angličtině a 11 v obou předmětech.)

Řešení:

$N(G_1) = 12, N(G_2) = 15, N(G_1 \cap G_2) = 11, n = 20,$

$p(G_1/G_2) = \dfrac{N(G_1 \cap G_2)}{N(G_2)} = \dfrac{11}{15} = 0{,}73$ (tzn., že 73% těch studentů, kteří byli úspěšní v angličtině, uspělo i v matematice)

$p(G_2/G_1) = \dfrac{N(G_1 \cap G_2)}{N(G_1)} = \dfrac{11}{12} = 0{,}92$ (tzn., že 92% těch studentů, kteří byli úspěšní v matematice, uspělo i v angličtině)

Pojem četnostní nezávislosti dvou množin: O četnostní nezávislosti dvou množin v daném výběrovém souboru hovoříme tehdy, když informace o původu objektu z jedné množiny nijak nemění šance, s nimiž soudíme na jeho původ i z druhé množiny.

V příkladě se studenty by množiny úspěšných matematiků a úspěšných angličtinářů byly četnostně nezávislé, pokud podíl úspěšných matematiků mezi úspěšnými angličtináři by byl stejný jako podíl úspěšných matematiků mezi všemi zkoušenými studenty a stejně tak podíl úspěšných angličtinářů mezi úspěšnými matematiky by byl stejný jako podíl úspěšných angličtinářů mezi všemi zkoušenými studenty, tj.

$$\frac{n(G_1 \cap G_2)}{n(G_2)} = \frac{n(G_1)}{n} \wedge \frac{n(G_1 \cap G_2)}{n(G_1)} = \frac{n(G_2)}{n}.$$

Po snadné úpravě dostaneme multiplikativní vztah

$$\frac{n(G_1 \cap G_2)}{n} = \frac{n(G_1)}{n} \cdot \frac{n(G_2)}{n}, \text{ tj. } p(G_1 \cap G_2) = p(G_1)\,p(G_2)$$

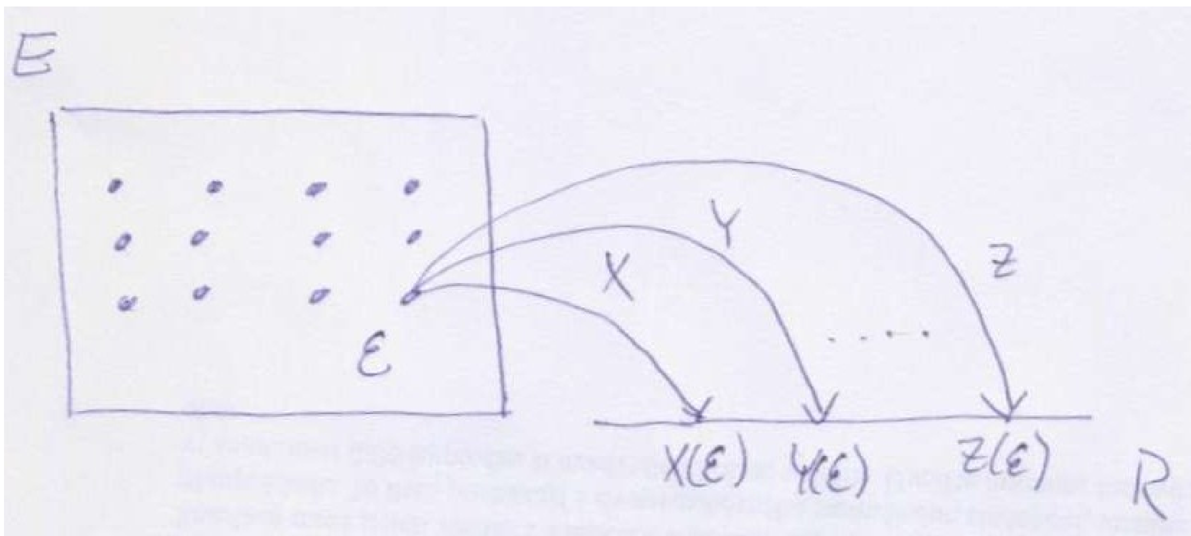
Řekneme tedy, že množiny G_1, G_2 jsou četnostně nezávislé v daném výběrovém souboru, jestliže $p(G_1 \cap G_2) = p(G_1)\,p(G_2)$.

(V praxi jen zřídka dojde k tomu, že uvedený vztah platí přesně. Většinou je jen naznačena určitá tendence četnostní nezávislosti.)

Příklad: Pro údaje z příkladu o studentech zjistěte, zda úspěchy v matematice a angličtině jsou v daném výběrovém souboru četnostně nezávislé.
(Připomínáme, že oboustranně úspěšných studentů bylo 55%, úspěšných matematiků 60% a úspěšných angličtinářů 75%.)

Řešení:
$p(G_1 \cap G_2) = 0{,}55$, $p(G_1)p(G_2) = 0{,}6 \times 0{,}75 = 0{,}45$, tedy skutečná relativní četnost oboustranně úspěšných studentů je větší než by odpovídalo četnostní nezávislosti množin G_1, G_2 v daném výběrovém souboru. Znamená to, že úspěch v matematice se zpravidla sdružuje s úspěchem v angličtině a naopak.

Pojem skalárního a vektorového znaku: Vlastnosti objektů vyjadřujeme číselně pomocí znaků.
Nechť E je základní soubor. Funkce
$X: E \to R$, $Y: E \to R$, ..., $Z: E \to R$,
které každému objektu přiřazují číslo, se nazývají (skalární) znaky.
Uspořádaná p-tice (X, Y, ..., Z) se nazývá vektorový znak.



Označení: Nechť je dán výběrový soubor $\{\varepsilon_1, ..., \varepsilon_n\} \subseteq E$. Hodnoty znaků X, Y, ..., Z pro i-tý objekt označíme $x_i = X(\varepsilon_i)$, $y_i = Y(\varepsilon_i)$, ..., $z_i = Z(\varepsilon_i)$, $i = 1, ..., n$.

Pojem datového souboru:

Matice $\begin{pmatrix} x_1 & y_1 & \cdots & z_1 \\ x_2 & y_2 & \cdots & z_2 \\ \cdots & \cdots & \cdots & \cdots \\ x_m & y_n & \cdots & z_n \end{pmatrix}$ typu n × p se nazývá datový soubor. Její řádky odpovídají jednotlivým objektům, sloupce znakům.

Libovolný sloupec této matice nazýváme jednorozměrným datovým souborem. Jestliže uspořádáme hodnoty některého znaku (např. znaku X) v jednorozměrném datovém souboru vzestupně podle velikosti, dostaneme uspořádaný datový soubor $\begin{pmatrix} x_{(1)} \\ \vdots \\ x_{(n)} \end{pmatrix}$, kde $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$.

Vektor $\begin{pmatrix} x_{[1]} \\ \vdots \\ x_{[r]} \end{pmatrix}$, kde $x_{[1]} < \ldots < x_{[r]}$ jsou navzájem různé hodnoty znaku X, se nazývá vektor variant.

Příklad: Pro studenty z výběrového souboru uvedeného výše byly zjišťovány hodnoty znaků X – známka z matematiky v prvním zkušebním termínu, Y – známka z angličtiny v prvním zkušebním termínu, Z – pohlaví studenta (0 ... žena, 1 ... muž). Byl získán datový soubor

$$\begin{pmatrix} 2 & 2 & 0 \\ 1 & 3 & 1 \\ 4 & 3 & 1 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \\ 4 & 4 & 1 \\ 3 & 3 & 1 \\ 3 & 4 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 4 & 2 & 1 \\ 4 & 4 & 0 \\ 2 & 2 & 0 \\ 4 & 3 & 1 \\ 2 & 3 & 1 \\ 4 & 4 & 0 \\ 1 & 1 & 0 \\ 4 & 3 & 1 \\ 4 & 4 & 1 \\ 1 & 3 & 0 \end{pmatrix}$$

Utvořte jednorozměrný uspořádaný i neuspořádaný datový soubor pro známky z matematiky a vektory variant pro známky z matematiky.

Řešení:

$$\begin{pmatrix} 2 \\ 1 \\ 4 \\ 1 \\ 1 \\ 4 \\ 3 \\ 3 \\ 1 \\ 1 \\ 4 \\ 4 \\ 2 \\ 4 \\ 2 \\ 4 \\ 1 \\ 4 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ 3 \\ 3 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

Pojem jevu: Nechť $\{\varepsilon_1, \dots, \varepsilon_n\}$ je výběrový soubor, X, Y, ..., Z jsou znaky, B, B_1, B_p jsou číselné množiny.
Zápis $\{X \in B\}$ znamená jev „znak X nabyl hodnoty z množiny B" .
Zápis $\{X \in B_1 \wedge Y \in B_2 \wedge \dots \wedge Z \in B_p\}$ znamená jev „znak X nabyl hodnoty z množiny B_1 a současně znak Y nabyl hodnoty z množiny B_2 atd. až znak Z nabyl hodnoty z množiny B_p".
Symbol $N(X \in B)$ značí absolutní četnost jevu $\{X \in B\}$ ve výběrovém souboru, tj. počet těch objektů ve výběrovém souboru, pro něž $x_i \in B$.
Symbol $p(X \in B)$ znamená relativní četnost jevu $\{X \in B\}$ ve výběrovém souboru, tj. $p(X \in B) = \frac{N(X \in B)}{n}$.
Analogicky $N(X \in B_1 \wedge Y \in B_2 \wedge \dots \wedge Z \in B_p)$ resp. $p(X \in B_1 \wedge Y \in B_2 \wedge \dots \wedge Z \in B_p)$ znamená absolutní resp. relativní četnost jevu $\{X \in B_1 \wedge Y \in B_2 \wedge \dots \wedge Z \in B_p\}$ ve výběrovém souboru.

Příklad: Pro datový soubor s údaji o známkách najděte relativní četnost
a) matematických jedničkářů
b) úspěšných matematiků
c) oboustranně neúspěšných studentů.
Datový soubor má tvar:

$$\begin{pmatrix} 2 & 2 & 0 \\ 1 & 3 & 1 \\ 4 & 3 & 1 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \\ 4 & 4 & 1 \\ 3 & 3 & 1 \\ 3 & 4 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 4 & 2 & 1 \\ 4 & 4 & 0 \\ 2 & 2 & 0 \\ 4 & 3 & 1 \\ 2 & 3 & 1 \\ 4 & 4 & 0 \\ 1 & 1 & 0 \\ 4 & 3 & 1 \\ 4 & 4 & 1 \\ 1 & 3 & 0 \end{pmatrix}$$

Řešení:

ad a) $p(X = 1) = \frac{7}{20} = 0{,}35;$

ad b) $p(X \leq 3) = \frac{12}{20} = 0{,}60;$

ad c) $p(X = 4 \wedge Y = 4) = \frac{4}{20} = 0{,}20.$

Zjistili jsme, že jedničku z matematiky mělo 35% studentů, zkoušku z matematiky úspěšně složilo 60% studentů a oboustranně neúspěšných bylo 20% studentů.

Jednorozměrné bodové rozložení četností

Jestliže počet variant znaku X v jednorozměrném datovém souboru není příliš velký, pak přiřazujeme četnosti jednotlivým variantám a hovoříme o bodovém rozložení četností.

Nechť je dán jednorozměrný datový soubor $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, v němž znak X nabývá r variant.

Pro j = 1, ..., r definujeme:

$n_j = N(X = x_{[j]})$ – absolutní četnost varianty $x_{[j]}$ ve výběrovém souboru

$p_j = \frac{n_j}{n}$ – relativní četnost varianty $x_{[j]}$ ve výběrovém souboru

$N_j = N(X \le x_{[j]}) = n_1 + \ldots + n_j$ – absolutní kumulativní četnost prvních j variant ve výběrovém souboru

$F_j = \frac{N_j}{n} = p_1 + \ldots + p_j$ – relativní kumulativní četnost prvních j variant ve výběrovém souboru

Tabulka typu

$x_{[j]}$	n_j	p_j	N_j	F_j
$x_{[1]}$	n_1	p_1	N_1	F_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_{[r]}$	n_r	p_r	N_r	F_r

se nazývá variační řada (nebo též tabulka rozložení četností).

Příklad: Máme jednorozměrný datový soubor, který obsahuje údaje o známkách z matematiky (znak X) u 20 studentů.

$$\begin{pmatrix} 2 \\ 1 \\ 4 \\ 1 \\ 1 \\ 4 \\ 3 \\ 3 \\ 1 \\ 1 \\ 4 \\ 4 \\ 2 \\ 4 \\ 2 \\ 4 \\ 1 \\ 4 \\ 4 \\ 1 \end{pmatrix}$$

Sestavte tabulku rozložení četností.

Řešení:

$x_{[j]}$	n_j	p_j	N_j	F_j
1	7	7/20=0,35	7	7/20=0,35
2	3	3/20=0,15	10	10/20=0,50
3	2	2/20=0,10	12	12/20=0,60
4	8	8/20=0,40	20	20/20=1,00
$\sum$	20	1,00	-	-

Pojem četnostní funkce a empirické distribuční funkce

Pomocí relativních četností zavedeme četnostní funkci.

Funkce $p(x) = \begin{cases} p_j \text{ pro } x = x_{[j]},\ j = 1, \dots, r \\ 0 \text{ jinak} \end{cases}$ se nazývá četnostní funkce.

Četnostní funkce je

nezáporná ($\forall x \in R$: $p(x) \geq 0$)

a normovaná ($\sum_{x=-\infty}^{\infty} p(x) = 1$).

Pomocí kumulativních relativních četností zavedeme empirickou distribuční funkci.

Funkce $F(x) = \begin{cases} 0 \text{ pro } x < x_{[1]} \\ F_j \text{ pro } x_{[j]} \leq x < x_{[j+1]},\ j = 1, \dots, r-1 \\ 1 \text{ pro } x \geq x_{[r]} \end{cases}$ se nazývá empirická distribuční funkce.

Empirická distribuční funkce je

neklesající ($\forall x_1, x_2 \in R$, $x_1 < x_2$: $F(x_1) \leq F(x_2)$),

zprava spojitá ($\forall x_0 \in R$ libovolné, ale pevně dané: $\lim_{x \to x_{0+}} F(x) = F(x_0)$)

a normovaná ($\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$).

Platí $\forall x \in R : F(x) = \sum_{t \leq x} p(t)$.

Příklad: Pro známky z matematiky nakreslete graf četnostní funkce a empirické distribuční funkce.

Řešení:

Variační řada

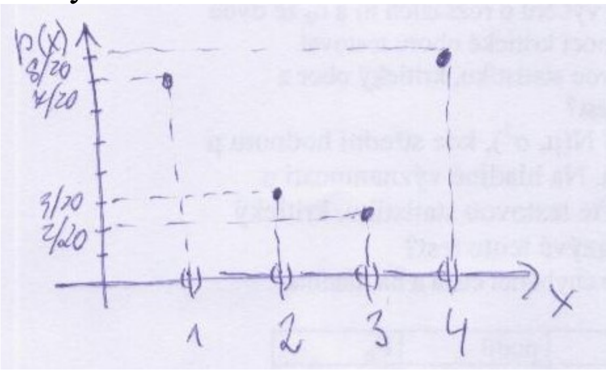
$x_{[j]}$	n_j	p_j	N_j	F_j
1	7	7/20=0,35	7	7/20=0,35
2	3	3/20=0,15	10	10/20=0,50
3	2	2/20=0,10	12	12/20=0,60
4	8	8/20=0,40	20	20/20=1,00
∑	20	1,00	-	-

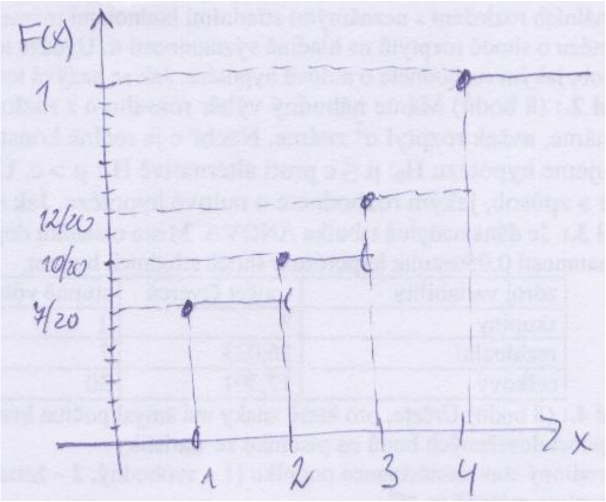
Vzorce

$$p(x) = \begin{cases} p_j \text{ pro } x = x_{[j]},\ j = 1, \ldots, r \\ 0 \text{ jinak} \end{cases}$$

$$F(x) = \begin{cases} 0 \text{ pro } x < x_{[1]} \\ F_j \text{ pro } x_{[j]} \le x < x_{[j+1]},\ j = 1, \ldots, r-1 \\ 1 \text{ pro } x \ge x_{[r]} \end{cases}$$

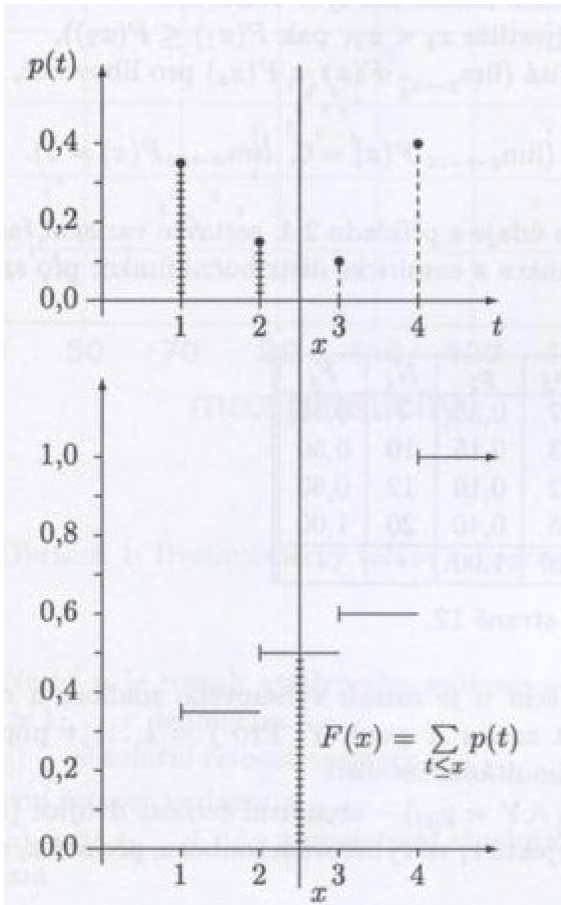
Grafy





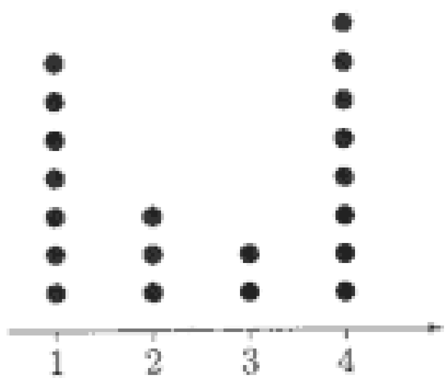
Vztah mezi četnostní funkcí a empirickou distribuční funkcí

$$\forall x \in R : F(x) = \sum_{t \le x} p(t)$$

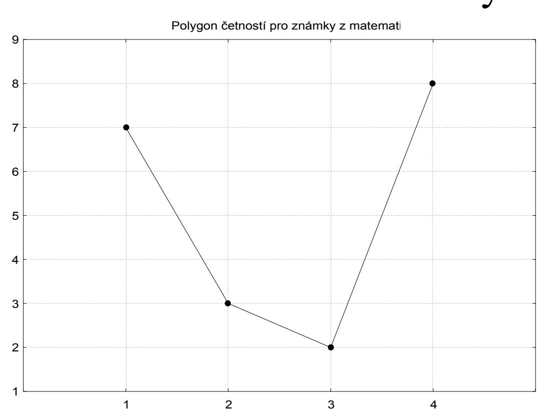


Grafické znázornění jednorozměrnho bodového rozložení četností

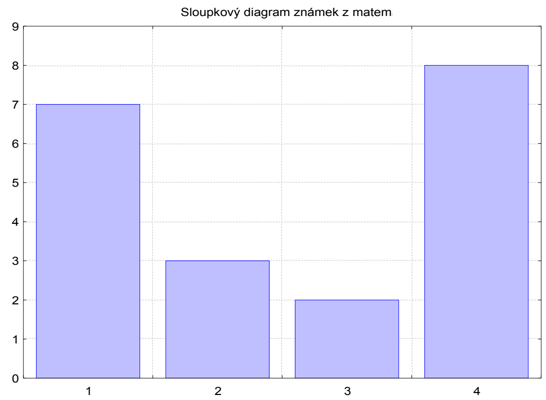
Tečkový diagram: na číselné ose vyznačíme jednotlivé varianty znaku X a nad každou variantu nakreslíme tolik teček, jaká je její absolutní četnost.



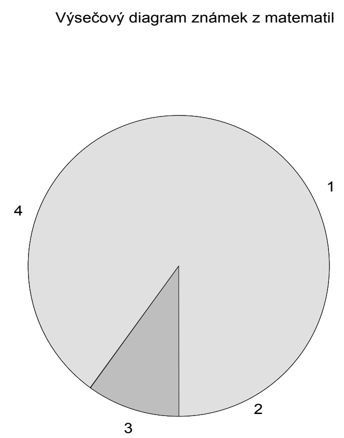
Polygon četností: je lomená čára spojující body, jejichž x-ová souřadnice je varianta znaku X a y-ová souřadnice je absolutní či relativní četnost této varianty.



Sloupkový diagram: je soustava na sebe nenavazujících obdélníků, kde střed základny je varianta znaku X a výška je absolutní či relativní četnost této varianty.



Výsečový graf: je kruh rozdělený na výseče, jejichž vnější obvod odpovídá absolutním četnostem variant znaku X.



Dvourozměrné bodové rozložení četností

Nechť je dán dvourozměrný datový soubor $\begin{pmatrix} x_1 & y_1 \\ \dots & \dots \\ x_n & y_n \end{pmatrix}$, kde znak X má r variant a znak Y má s variant. Pak definujeme:

$n_{jk} = N(X = x_{[j]} \wedge Y = y_{[k]})$ – simultánní absolutní četnost dvojice $(x_{[j]}, y_{[k]})$ ve výběrovém souboru

$p_{jk} = \frac{n_{jk}}{n}$ – simultánní relativní četnost dvojice $(x_{[j]}, y_{[k]})$ ve výběrovém souboru

$n_{j.} = N(X = x_{[j]}) = n_{j1} + \dots + n_{js}$ – marginální absolutní četnost varianty $x_{[j]}$

$p_{j.} = \frac{n_{j.}}{n} = p_{j1} + \dots + p_{js}$ – marginální relativní četnost varianty $x_{[j]}$

$n_{.k} = N(Y = y_{[k]}) = n_{1k} + \dots + n_{rk}$ – marginální absolutní četnost varianty $y_{[k]}$

$p_{.k} = \frac{n_{.k}}{n} = p_{1k} + \dots + p_{rk}$ – marginální relativní četnost varianty $y_{[k]}$

Simultánní četností zapisujeme do kontingenční tabulky.

Kontingenční tabulka simultánních absolutních četností má tvar:

x \ y (n_{jk})	$y_{[1]}$	...	$y_{[s]}$	$n_{j.}$
$x_{[1]}$	n_{11}	...	n_{1s}	$n_{1.}$
$\vdots$	...	...	...	...
$x_{[r]}$	n_{r1}	...	n_{rs}	$n_{r.}$
$n_{.k}$	$n_{.1}$	...	$n_{.s}$	n

Příklad: Máme datový soubor, který obsahuje údaje o známkách z matematiky (znak X), z angličtiny (znak Y) a pohlaví studenta (znak Z, 0 – žena, 1 – muž) u 20 studentů:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
X	2	1	4	1	1	4	3	3	1	1	4	4	2	4	2	4	1	4	4	1
Y	2	3	3	1	2	4	3	4	1	1	2	4	2	3	3	4	1	3	4	3
Z	0	1	1	0	1	1	1	0	0	0	1	0	0	1	1	0	0	1	1	0

Vytvořte kontingenční tabulku simultánních absolutních a relativních četností pro známky z matematiky a angličtiny.

Řešení:

Kontingenční tabulka simultánních absolutních četností

x \ y, n_{jk}	1	2	3	4	$n_{j\cdot}$
1	4	1	2	0	7
2	0	2	1	0	3
3	0	0	1	1	2
4	0	1	3	4	8
$n_{\cdot k}$	4	4	7	5	$n = 20$

Kontingenční tabulka simultánních relativních četností

x \ y, p_{jk}	1	2	3	4	$p_{j\cdot}$
1	0,20	0,05	0,10	0,00	0,35
2	0,00	0,10	0,05	0,00	0,15
3	0,00	0,00	0,05	0,05	0,10
4	0,00	0,05	0,15	0,20	0,40
$p_{\cdot k}$	0,20	0,20	0,35	0,25	1,00

Pojem simultánní a marginální četnostní funkce

Pomocí simultánních relativních četností zavedeme simultánní četnostní funkci:

Funkce

$$p(x, y) = \begin{cases} p_{jk} \text{ pro } x = x_{[j]},\, y = y_{[k]},\, j = 1, \ldots, r,\, k = 1, \ldots, s \\ 0 \text{ jinak} \end{cases}$$

se nazývá simultánní četnostní funkce.

Pomocí marginálních relativních četností zavedeme marginální četnostní funkce pro znaky X a Y. Odlišíme je indexem takto:

$$p_1(x) = \begin{cases} p_{j.} \text{ pro } x = x_{[j]},\, j = 1, \ldots, r \\ 0 \text{ jinak} \end{cases},$$

$$p_2(y) = \begin{cases} p_{.k} \text{ pro } y = y_{[k]},\, k = 1, \ldots, s \\ 0 \text{ jinak} \end{cases}.$$

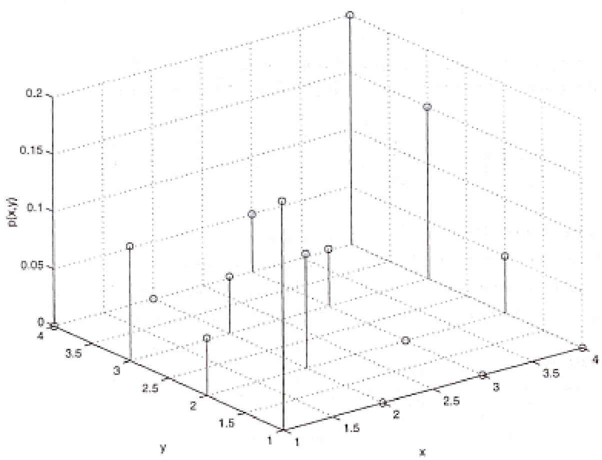
Mezi simultánní četnostní funkcí a marginálními četnostními funkcemi platí vztahy:

$$p_1(x) = \sum_{y=-\infty}^{\infty} p(x, y), \quad p_2(y) = \sum_{x=-\infty}^{\infty} p(x, y).$$

Příklad: Sestrojte graf simultánní četnostní funkce pro známky z matematiky a angličtiny.

Řešení: Vyjdeme z kontingenční tabulky simultánních relativních četností.

x \ y, p_{jk}	1	2	3	4	$p_{j\cdot}$
1	0,20	0,05	0,10	0,00	0,35
2	0,00	0,10	0,05	0,00	0,15
3	0,00	0,00	0,05	0,05	0,10
4	0,00	0,05	0,15	0,20	0,40
$p_{\cdot k}$	0,20	0,20	0,35	0,25	1,00



Pojem četnostní nezávislosti znaků v daném výběrovém souboru

Řekneme, že znaky X, Y jsou v daném výběrovém souboru četnostně nezávislé, právě když pro všechna j = 1, ..., r a všechna k = 1, ..., s platí multiplikativní vztah:

$p_{jk} = p_{j.}\, p_{.k}$ neboli pro $\forall (x, y) \in R^2$: $p(x, y) = p_1(x)\, p_2(y)$.

Příklad: Ověřte, zda v našem datovém souboru jsou známky z matematiky a angličtiny četnostně nezávislé.

Řešení: Vyjdeme z kontingenční tabulky relativních četností.

y		1	2	3	4	$p_{j\cdot}$
x	p_{jk}					
1		0,20	0,05	0,10	0,00	0,35
2		0,00	0,10	0,05	0,00	0,15
3		0,00	0,00	0,05	0,05	0,10
4		0,00	0,05	0,15	0,20	0,40
$p_{\cdot k}$		0,20	0,20	0,35	0,25	1,00

Známky z matematiky a angličtiny nejsou četnostně nezávislé, protože už pro j = 1, k = 1 je multiplikativní vztah porušen:

$p_{11} = 0{,}20$, $p_{1.} = 0{,}35$, $p_{.1} = 0{,}20$, tudíž $0{,}20 \neq 0{,}35.0{,}20$

Pojem řádkově a sloupcově podmíněných relativních četností

Sloupcově podmíněná relativní četnost varianty $x_{[j]}$ za předpokladu $y_{[k]}$:

$$p_{j(k)} = \frac{n_{jk}}{n_{.k}}$$

Řádkově podmíněná relativní četnost varianty $y_{[k]}$ za předpokladu $x_{[j]}$:

$$p_{(j)k} = \frac{n_{jk}}{n_{j.}}$$

Příklad: Pro datový soubor známek z matematiky a angličtiny sestavte kontingenční tabulku sloupcově a poté řádkově podmíněných relativních četností.

Řešení:

Nejprve se budeme zabývat sloupcově podmíněnými relativními četnostmi. Použijeme vzorec $p_{j(k)} = \frac{n_{jk}}{n_{.k}}$.

Vyjdeme z kontingenční tabulky simultánních absolutních četností.

x \ y (n_{jk})	1	2	3	4	$n_{j.}$
1	4	1	2	0	7
2	0	2	1	0	3
3	0	0	1	1	2
4	0	1	3	4	8
$n_{.k}$	4	4	7	5	$n = 20$

x \ y ($p_{j(k)}$)	1	2	3	4
1	1,00	0,25	0,29	0,00
2	0,00	0,50	0,14	0,00
3	0,00	0,00	0,14	0,20
4	0,00	0,25	0,43	0,80
$\sum$	1,00	1,00	1,00	1,00

Interpretujeme např. třetí sloupec: z těch studentů, kteří měli trojku z angličtiny, mělo 2/7 = 29% jedničku z matematiky, 1/7 = 14% dvojku z matematiky, 1/7 = 14% trojku z matematiky a 3/7 = 43% čtyřku z matematiky.

Dále se budeme zabývat řádkově podmíněnými relativními četnostmi.

Použijeme vzorec $p_{(j)k} = \frac{n_{jk}}{n_{j.}}$.

Opět nám poslouží kontingenční tabulka absolutních četností.

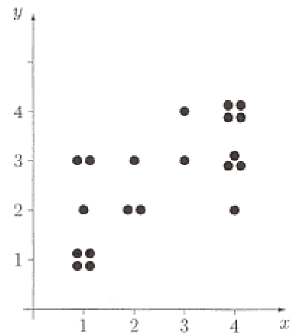
x \ y (n_{jk})	1	2	3	4	$n_{j.}$
1	4	1	2	0	7
2	0	2	1	0	3
3	0	0	1	1	2
4	0	1	3	4	8
$n_{.k}$	4	4	7	5	$n = 20$

x \ y ($p_{(j)k}$)	1	2	3	4	$\sum$
1	0,57	0,14	0,29	0,00	1,00
2	0,00	0,67	0,33	0,00	1,00
3	0,00	0,00	0,50	0,50	1,00
4	0,00	0,12	0,38	0,50	1,00

Interpretujeme např. první řádek: z těch studentů, kteří měli jedničku z matematiky, mělo 4/7 = 57% jedničku z angličtiny, 1/7 = 14% dvojku z angličtiny a 2/7 = 29% trojku z angličtiny.

Dvourozměrný tečkový diagram

Dvourozměrné rozložení četností lze znázornit pomocí dvourozměrného tečkového diagramu. Na vodorovnou osu vyneseme varianty znaku X, na svislou varianty znaku Y a do příslušných průsečíků nakreslíme tolik teček, jaká je absolutní četnost dané dvojice. V našem příkladě se studenty dostaneme tento diagram:



Dvourozměrný tečkový diagram svědčí o nepříliš výrazné tendenci k podobné klasifikaci v obou předmětech.
Zcela odlišný vzhled má diagram pro muže a pro ženy:

Pro muže	Pro ženy