Popisná štatistika.

Štatistika skúma javy na prípadoch v rozsiahlych súboroch a hľadá tie vlastnosti javov, ktoré sa prejavujú až pri veľkom výskyte prípadov a nie v jednotlivých prípadoch. Základným pojmom štatistiky je štatistický súbor.

Definícia: Štatistický súbor je ľubovoľná dobre definovaná množina štatistických jednotiek. Štatistický súbor môže byť určený zoznamom svojich prvkov (jednotiek) alebo môže byť definovaný pomocou nejakého pravidla (vlastnosti). V prípade nejakých pochybnosti musí existovať možnosť overiť, či daná jednotka patrí do štatistického súboru.

Štatistický súbor pracuje so štatistickými jednotkami, na ktorých sa meria jeden alebo viacero štatistických znakov. Meranie týchto štatistických znakov môže byť uskutočnené podľa viacerých kritérií. Najbežnejšie kritériá sú:

(1) Nominálne: predpokladá disjunktné kategórie, ktoré obsahujú všetky možné hodnoty merania. Je to analógia rozkladu náhodného javu z teórie pravdepodobnosti. Medzi jednotlivými hodnotami nie je žiaden vzťah, ani usporiadanie. (napr. farba očí pri výskume dedičnosti)
(2) Ordinálne: je to v podstate nominálne kritérium, v ktorom pribudlo usporiadanie jednotlivých hodnôt. Možným hodnotám môžeme priradiť indexy (resp. poradové čísla), pričom hodnota s menším indexom sa nachádza pred hodnotou s väčším indexom. (napr. najväčšie dosiahnuté vzdelanie, počet hviezdičiek priradených kategórii hotela, úroveň jazykovej zdatnosti) Tieto indexy vyjadrujú iba poradie týchto hodnôt a nie ich vzdialenosť!
(3) Intervalové: toto kritérium predpokladá jednoznačné číselné označenie jednotlivých možných hodnôt. Určuje to ich usporiadanie a súčasne sa predpokladá, že vzdialenosti medzi susednými hodnotami sú konštantné. Dôležité je, že umiestnenie nuly je iba dohoda – napr. pri meraní teploty, je nula umiestnená v závislosti od teplotnej stupnice.
(4) Pomerové: vyjadruje vzťah nejakej nameranej veličiny k nejakej dohodnutej jednotke, t.j. udáva jej násobok. Nula v tomto prípade udáva neexistenciu meranej vlastnosti. (Najčastejšie sem patria fyzikálne veličiny.)

Kvalitatívne znaky: sú štatistické znaky, ktoré sú merané podľa nominálnaho resp. ordinálneho kritéria.
Kvantitatívne znaky (spojité): sú štatistické znaky, ktoré sú merané podľa intervalového alebo pomerného kritéria.

Definícia: Predpokladajme, že sme pre n štatistických jednotiek namerali hodnoty

$$x_1, x_2, x_3, \dots , x_{n-1}, x_n$$

daného znaku. Potom tieto hodnoty nazývame súbor hodnôt daného znaku štatistického súboru. Ak tieto hodnoty navyše usporiadame (aspoň podľa ordinálneho kritéria) do neklesajúcej postupnosti čísel $\{x_{(1)}, x_{(2)}, \dots , x_{(n-1)}, x_{(n)}\}$, kde

$$x_{(1)} \le x_{(2)} \le \dots \le x_{(n-1)} \le x_{(n)},$$

potom tento súbor hodnôt nazývame usporiadaný súbor hodnôt. Indexy v zátvorkách vyjadrujú poradie jednotlivých nameraných hodnôt. Najmenšiu hodnotu označíme $x_{(1)}$ a najväčšiu hodnotu označíme symbolom $x_{(n)}$.

Prehľadná tabuľka

TYPY DÁT

<table>
<tr><th colspan="2">Typ dát</th><th>Stručný opis</th><th>Príklad</th></tr>
<tr><td rowspan="2">K
v
a
l
i
t
a
t</td><td>Nominálny</td><td>Len triedy. Dáta sa nedajú usporiadať podľa poradia</td><td>Rozdelenie voličov:
45 demokratov
80 republikánov   Len kategórie
90 nezávislých</td></tr>
<tr><td>Ordinálny</td><td>Triedy sú usporiadané, ale rozdiely medzi dátovými hodnotami sa nedajú určiť alebo nemajú zmysel.</td><td>Rozdelenie voličov:
45 nízky príjem   Usporiadanie je určené
80 stredný príjem   poradím
90 vysoký príjem   „ nízky, stredný, vysoký“.</td></tr>
<tr><td rowspan="2">K
v
a
n
t
i
t
a
t</td><td>Intervalový</td><td>Zmysluplné rozdiely medzi hodnotami sa dajú nájsť, ale neexistuje na stupnici začiatočný bod. Podiely nemajú zmysel.</td><td>Teplotu oceľových drôtov
45˚ F   90˚ F nie je
80˚ F   dvakrát teplejší
90˚ F   ako 45˚ F</td></tr>
<tr><td>Podielový</td><td>Rozdiely medzi hodnotami majú zmysel. Na stupnici existuje začiatočný bod. Podiely majú zmysel.</td><td>Dĺžky oceľových drôtov.
45 cm   90 cm je
80 cm   dvakrát dlhší
90 cm   ako 45 cm.</td></tr>
</table>

Príklad 1. Máme súbor 155 osobných automobilov, u ktorých sledujeme tieto znaky: *spotreba, počet valcov, výkon motora, zrýchlenie, rok výroby, hmotnosť, pôvod, výrobca, model***.** Klasifikujte uvedené znaky podľa štyroch typov aj podľa dvoch typov.

Riešenie. Výsledok je zobrazený v nasledujúcej tabuľke:

KVALITATÍVNY ÚDAJ		KVANTITATÍVNY ÚDAJ	
nominálny typ	ordinálny typ	intervalový typ	podielový typ
pôvod výrobca model	počet valcov rok výroby		spotreba výkon motora zrýchlenie hmotnosť

<u>**Triedením**</u> rozumieme rozdelenie jednotiek súboru do takých skupín (tried), aby čo najlepšie vynikli charakteristické vlastnosti skúmaných javov. Výsledkom triedenia býva tabuľka početnosti.

<u>**Tabuľka početnosti:**</u> Predpokladajme, že máme veľký súbor ($n \geq 30$) a namerané hodnoty podľa kvalitatívneho kritéria. Nech sa znaky často opakujú a nech

$$a_1, a_2, \dots , a_m \quad \text{pre} \quad m < n$$

sú všetky rôzne hodnoty štatistického súboru. V prípade ordinálneho kritéria sú usporiadané

$$a_1 \leq a_2 \leq \dots \leq a_m.$$

Nech $n_1, n_2, \dots , n_m$ sú zistené <u>(absolútne) početnosti</u> týchto hodnôt. Potom musí platiť, že:

$$n = \sum_{i=1}^{m} n_i \; .$$

Takýto typ tabuľky početnosti sa najčastejšie používa pri kvalitatívnych kritériách a pri súboroch, ktorých znaky nadobúdajú len celočíselné hodnoty.
V prípade kvantitatívnych kritérií sa uvedený postup trochu modifikuje. Meraný znak môže nadobúdať veľa hodnôt, ale často sa neopakujú, ale vyskytujú sa tam s malými odchýlkami. Potom umelo zmenšíme počet rôznych hodnôt tak, že obor hodnôt rozdelíme na disjunktné intervaly:

$$(-\infty, t_0) \cup \langle t_1, t_2) \cup \langle t_2, t_3) \cup \dots \cup \cup \langle t_m, \infty)$$

resp.

$$(-\infty, t_0\rangle \cup (t_1, t_2\rangle \cup (t_2, t_3\rangle \cup \dots \cup \cup (t_m, \infty).$$

Všetky hodnoty z j-teho intervalu $(t_{j-1}, t_j\rangle$ stotožníme s hodnotou

$$a_j = (t_{j-1} + t_j)/2 \quad \text{pre } j = 1. ,,, . m–1.$$

Hodnota

$$a_0 = t_0 - (t_1 - t_0)/2$$

a

$$a_m = t_{m-1} + (t_{m-2} - t_{m-1})/2.$$

Najčastejšie sa zvolia deliace body tak, aby intervaly mali rovnakú dĺžku (krajné intervaly môžu tvoriť výnimku) a dĺžka intervalu sa zvykne označovať

$$h = tj - t_{j-1} \text{ pre } j = 1, \dots , m - 1.$$

Potom určíme počty n_j pre $j = 0, 1, 2, \dots , m - 1, m$ (takzvané <u>triedne početnosti</u>) hodnôt x_i, ktoré patria do jednotlivých intervalov (takzvané <u>triedy</u>) a dostávame tabuľku početnosti.

Tabuľku početnosti môžeme doplniť o ďalšie údaje:

– <u>relatívnu početnosť</u> $f_j = n_j / n$, pre $j = 0, 1, .. , m$
– <u>kumulatívnu početnosť</u> $N_j = \sum_{i=0}^{j} n_i$, pre $j = 0, 1, \dots , m$
– <u>relatívnu kumulatívnu početnosť</u> $F_j = N_j / n$, pre $j = 0, 1, \dots , m$

<u>Grafy (diagramy):</u>

– polygón početností
– histogram
– kruhový (koláčový)
– prstencový (pre viacero radov)

Optimálny počet tried: $m = 1 + 3,322 \cdot \log n$

Konštrukcia tabuľky početností:

Vytváranie tabuľky početností je dobrá metóda na opis veľkej množiny dát. Pôvodné dáta sa zatrieďujú (grupujú) do tried (kategórií) a zisťujú sa početnosti jednotlivých tried, čím sa vytvorí rozdelenie početností.

Proces konštruovania tabuľky početností sa skladá z týchto krokov:

1. Určenie počtu tried *m*, ktoré bude obsahovať tabuľka početností. Ak nevieme určiť počet tried, použijeme napr. vzorec
$$m = 1 + 3,322 \cdot \log n$$
2. Nájdeme najmenšiu $x_{\min}$ a najväčšiu $x_{\max}$ hodnotu zo súboru hodnôt.
3. Nájdeme šírku triedy h delením rozdielu medzi najväčšou a najmenšou hodnotou (nazývanou *rozpätie* = $x_{\max} - x_{\min}$) počtom tried. Výsledky zaokrúhľujte nahor na vhodné číslo tak, aby každá hodnota zo súboru hodnôt bola obsiahnutá v tabuľke početností.
$$h = \frac{x_{\max} - x_{\min}}{m}$$
4. Zvolíme ako štartovací bod najmenšiu hodnotu alebo hodnotu trocha menšiu ako najmenšia hodnota. Štartovací bod je *dolná hranica prvej triedy* t_0.
5. Pripočítajte šírku triedy h k štartovaciemu bodu t_0, aby ste dostali *dolnú hranicu druhej triedy* t_1. Pripočítajte šírku triedy h k dolnej hranici druhej triedy t_1, aby ste dostali *dolnú hranicu tretej triedy* t_2. atď.
6. Dajte všetky dolné hranice do jedného stĺpca a horné hranice do vedľajšieho stĺpca, ako je uvedené v nasledujúcej tabuľke.
7. Zaznamenajte každú hodnotu do riadku príslušnej triedy.
8. Spočítajte počty záznamov prislúchajúcich jednotlivým triedam. Dostanete tak *početnosti jednotlivých tried* n_i. Tieto zaznamenajte v ďalšom stĺpci.
9. Vypočítajte *relatívne početnosti*, *kumulatívne početnosti*, *kumulatívne relatívne početnosti* a zapíšte ich do ďalších troch stĺpcov. Takto sme vytvorili celú tabuľku početností.

Tabuľka 1 Schéma rozdelenia početnosti:

Obmena znaku x_i	Absolútna početnosť n_i	Relatívna početnosť f_i	Kumulatívna abs. početnosť N_i	Kumulatívna rel. početnosť F_i
x_1 x_2 $\vdots$ x_k	n_1 n_2 $\vdots$ n_k	f_1 f_2 $\vdots$ f_k	$N_1 = n_1$ $N_2 = n_1 + n_2$ $\vdots$ $N_k = \sum_{i=1}^{k} n_i = n$	$F_1 = f_1$ $F_2 = f_1 + f_2$ $\vdots$ $F_k = \sum_{i=1}^{k} f_i = 1$
súčet	$\sum_{i=1}^{k} n_i = n$	$\sum_{i=1}^{k} f_i = 1$		

Tabuľka 2 *Schéma* intervalového rozdelenia početnosti:

Trieda interval	Dolná medza	Horná medza	Triedny znak	Absolútna početnosť	Relatívna početnosť	Kumulatívna početnosť	Relat. kum. početnosť
0.		t_0	a_0	n_0	$f_0 = \frac{n_0}{n}$	n_0	f_0
1.	t_0	t_1	a_1	n_1	$f_1 = \frac{n_1}{n}$	$n_0 + n_1$	$f_0 + f_1$
2.	t_1	t_2	a_2	n_2	$f_2 = \frac{n_2}{n}$	$n_0 + n_1 + n_2$	$f_0 + f_1 + f_2$
$\vdots$							
m.	t_{m-1}	t_m	a_m	n_m	$f_k = \frac{n_m}{n}$	$n = \sum_{i=0}^{m} n_i$	$\sum_{i=0}^{m} f_i = 1$
súčet				$\sum_{i=0}^{m} n_i = n$	$\sum_{i=0}^{m} f_i = 1$		

V tabuľke platia nasledujúce vzťahy:

$t_0 < x_{\min}$, $\quad t_k > x_{\max}$, $\quad a_i = \frac{t_{i-1} + t_i}{2}$ je stred *i*-teho intervalu (*i*-tej triedy)

Na určenie počtu tried *m* môžeme použiť niektoré z pravidiel:

(1) $m = \frac{R}{h}$, kde $h = t_i - t_{i-1}$ je šírka intervalu, ktorú môžeme určiť zo vzťahu

$$h = 0{,}08R \quad \text{alebo} \quad h < \frac{R}{12} < 2h$$

Ak tento podiel nie je celé číslo, berieme najbližšie vyššie celé číslo.

(2) podľa Sturgersovho pravidla:
$m = 1 + 3{,}322 \log n$, kde *m* zaokrúhľujeme nahor.

– n_i je absolútna početnosť *i*-tej triedy a

– $f_i = \frac{n_i}{n}$ je relatívna početnosť *i*-tej triedy

– $N_i = \sum_{j=1}^{i} n_j$ je kumulatívna početnosť *i*-tej triedy

.

-- $F_i = \frac{N_i}{n}$ je kumulatívna relatívna početnosť *i*-tej triedy

.

Kumulatívne rozdelenie početnosti nám umožňuje zaradiť individuálnu hodnotu znaku do celkového rozdelenia početnosti daného znaku.
Potom pomocou tabuľky početností môžeme konštruovať *histogram*, *polygon početností*, *polygon relatívnych početností*, *polygon kumulatívnych početností* a *polygon kumulatívnych relatívnych početností*. Všetky tieto grafy nám svojim spôsobom ukazujú rozdelenie početností nameraných dát.

Charakteristiky polohy štatistického súboru: (Miery polohy)

Charakteristiky polohy nám udávajú hodnotu okolo ktorej sa jednotlivé pozorovania zhromažďujú.

Priemer: (výberový priemer) Vyžaduje aspoň intervalové kritérium a rovnako závisí na všetkých hodnotách štatistického znaku.

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{n} \cdot \sum_{i=1}^{m} n_i \cdot x_i$$

Pre aritmetický priemer platí:

$$\overline{(a+bx)} = a + b\bar{x}$$, kde *a*, *b* sú ľubovoľné reálne čísla.

Geometrický priemer: má zmysel len, ak sú všetky hodnoty kladné. Je analógiou aritmetického priemeru, lebo jeho logaritmus je aritmetickým priemerom logaritmov. Geometrický priemer nie je invariantný voči lineárnej transformácii. Používa sa v prípadoch, kde ide o násobenie. Najčastejšie vo finančníctve: úrokovanie, inflácia a pod.

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n}$$

Harmonický priemer: má zmysel len, ak sú všetky hodnoty kladné. Tiež nie je invariantný voči lineárnej transformácii. Je definovaný vzťahom:

$$\bar{x}_H = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i} \right)^{-1}.$$

Ak sú všetky hodnoty znaku kladné, tak platia nasledujúce nerovnosti:

$$\bar{x}_H \le \bar{x}_G \le \bar{x} \,.$$

Medián: (výberový medián) Je definovaný pomocou usporiadaného súboru hodnôt $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n-1)} \le x_{(n)}$ nasledujúcim vzťahom:

$$\tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & \text{pre } n \text{ nepárne} \\ \frac{1}{2} \cdot \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) & \text{pre } n \text{ párne} \end{cases}$$

Medián je hodnota, ktorá nám rozdelí usporiadané hodnoty na rovnaké súbory, čo do počtu hodnôt. Z toho vyplýva, že nezáleží na konkrétnych hodnotách na začiatku a konci súboru hodnôt, len kde sa v tomto usporiadaní nachádzajú. Pre medián platí rovnaká vlastnosť, ako pre priemer:

$$(\widetilde{a+bx}) = a + b\tilde{x}$$, kde *a*, *b* sú ľubovoľné reálne čísla.

Takto definovaný medián má pre zmysel už pre ordinálne kritérium, ak n je nepárne. Ak n je párne, tak potrebujeme aspoň intervalové kritérium. Ak by sme pre párne n definovali medián ako akúkoľvek hodnotu spĺňajúcu nerovnosť:

$$x_{(\frac{n}{2})} \leq \tilde{x} \leq x_{(\frac{n}{2}+1)},$$

tak by sme stratili jednoznačnosť v definícii, ale mohli by sme ju použiť aj pre ordinálne kritérium.
Zovšeobecnený medián je takzvaný <u>p-tý kvantil</u> (<u>percentil</u>), ktorý neoddeľuje polovicu najmenších hodnôt od ostatných, ale oddeľuje p-ty diel hodnôt. Nech $0 < p < 1$, potom definujeme (výberový) p-ty <u>kvantil</u> (<u>percentil</u>) vzťahom:

$$x_p = \begin{cases} x_{([np]+1)} & \text{pre } np \neq [np] \\ \frac{1}{2} \cdot \left(x_{(np)} + x_{(np+1)} \right) & \text{pre } np = [np] \end{cases}$$

Ak np nie je celé číslo, tak z usporiadaného súboru hodnôt sa vyberie hodnota s najbližším väčším indexom. Ak np je celé číslo, tak sa použije priemer z daných dvoch hodnôt (v prípade ordinálneho kritéria sa vezme ľubovoľná hodnota medzi nimi). Z definície percentilu dostávame, že platí:

$$\tilde{x} = x_{0,5} .$$

Pri grafických znázorneniach sa často používa <u>horný</u> a <u>dolný</u> <u>kvartil</u>:

$$Q_1 = x_{0,25} \quad \text{a} \quad Q_3 = x_{0,75} .$$

Ak máme k dispozícii len tabuľku početností hodnôt daného znaku, tak na výpočet mediánu využijeme nasledujúci vzťah:

$$\tilde{x} = a_e + h \cdot \frac{\frac{n+1}{2} - N_{i-1}}{n_i},$$

kde a_e je dolná hranica intervalu triedy v ktorej sa nachádza medián, h je dĺžka intervalu triedy, n je rozsah súboru, n_i je početnosť triedy , v ktorej sa nachádza medián a N_{i-1} je kumulatívna početnosť triedy, ktorá predchádza triedu obsahujúcu medián.

<u>Modus:</u> Je najčastejšie sa vyskytujúca hodnota v súbore hodnôt. Má zmysel hlavne v prípadoch, v ktorých je počet m skutočne sa vyskytujúcich rôznych hodnôt podstatne menší, ako je rozsah štatistického súboru n. Modus nemusí byť určený jednoznačne, ale môžeme ho použiť pri všetkých kritériách (pri nominálnom kritériu sa nedá hovoriť o charakteristike polohy).
Ak máme k dispozícii len tabuľku početností hodnôt daného znaku, tak na výpočet mediánu využijeme nasledujúci vzťah:

$$\hat{x} = a_o + h \cdot \frac{d_1}{d_1 + d_2},$$

kde a_o je dolná hranica intervalu, v ktorom sa nachádza modus, h je dĺžka triedneho intervalu, n je rozsah súboru, $d_1 = n_i - n_{i-1}$ (i je trieda do ktorej patrí modus) a $d_1 = n_{i+1} - n_i$ (i je trieda do ktorej patrí modus).

Charakteristiky variability štatistického súboru: (Miery variability)

Miery variability charakterizujú veľkosť variability hodnôt okolo nejakej ich miery polohy alebo veľkosť ich vzájomnej rozdielnosti. Základnou požiadavkou by mala byť invariantnosť voči posunutiu, lebo pripočítaním rovnakej konštanty ku všetkým hodnotám sa variabilita nesmie zmeniť.

Rozptyl: (výberový rozptyl) je definovaný vzťahom:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \overline{x}^2 = s_x^2 .$$

Niekedy sa namiesto výrazu ($1/n$) používa výraz ($1/(n - 1)$), ale to popíšeme až neskôr pri výberovom rozptyle. Ak je štatistický súbor popísaný len tabuľkou početnosti, tak rozptyl vypočítame podľa vzťahu:

$$s^2 = \frac{1}{n}\sum_{i=1}^{m}n_i \cdot (a_i - \overline{x})^2 ,$$

tento vzťah sa ešte zvykne upraviť tak, že sa odčíta od daného výrazu hodnota $h^2/12$, kde h je dĺžka triednych rovnako širokých intervalov (takzvaná Scheppardova korekcia).

Smerodajná odchýlka: Je to odmocnina z výberového rozptylu a označíme ju s alebo s_x, ak chceme zdôrazniť, ku ktorej premennej prislúcha. Smerodajná odchýlka rovnako ako aj rozptyl a priemer závisí na všetkých pozorovaniach.

Rozpätie: je to rozdiel maximálnej a minimálnej hodnoty. Rozpätie je závislé len od najmenšej a najväčšej hodnoty súboru.

$$R = x_{(n)} - x_{(1)}$$

Kvartilové rozpätie: je definované ako rozdiel dvoch kvartilov:

$$R_Q = Q_3 - Q_1.$$

Hodnota $R_Q/2$ sa tiež nazýva kvartilová odchýlka.

Priemerná odchýlka: je to priemerná vzdialenosť jednotlivých hodnôt x_i na reálnej osi od mediánu (niekedy od priemeru):

$$d = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \tilde{x}\right| .$$

Všetky uvedené miery variability predpokladajú aspoň intervalové kritérium.

Veta: Funkcia $S(t) = \frac{1}{n}\sum_{i=1}^{n}(x_i - t)^2$ nadobúda svoju minimálnu hodnotu v bode $t = \bar{x}$.

Veta: Funkcia $T(t) = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - t\right|$ nadobúda svoju minimálnu hodnotu v bode $t = \tilde{x}$.

<u>Charakteristiky šikmosti a ostrosti štatistického súboru:</u> (Miery šikmosti a ostrosti)

<u>Výberový koeficient šikmosti:</u> tento koeficient je definovaný vzťahom:

$$\gamma_1 = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}.$$

Ľahko vidieť súvislosť s momentovými charakteristikami, ktoré sme uviedli v teórii pravdepodobnosti pri popise charakteristík náhodnej veličiny. Stačí zaviesť náhodnú veličinu X tak, že každej hodnote x_i pridelíme rovnakú pravdepodobnosť $1/n$ (prípadne každej hodnote a_j pravdepodobnosť n_j/n). Potom stredná hodnota EX je totožná s aritmetickým priemerom a rozptyl var$X$ je totožný s rozptylom s^2 a dostaneme, že koeficient šikmosti náhodnej veličiny X je ekvivalentný s naším vzorcom.

<u>Kvantilový koeficient šikmosti:</u> Nech $0 < p < 0{,}5$. Potom kvantilový koeficient šikmosti je definovaný vzťahom:

$$\frac{(x_{1-p} - \tilde{x}) - (\tilde{x} - x_p)}{x_{1-p} - x_p}.$$

Špeciálne ak položíme $p = 0{,}25$, tak dostaneme <u>kvartilový koeficient šikmosti</u>:

$$\frac{(Q_3 - \tilde{x}) - (\tilde{x} - Q_1)}{Q_3 - Q_1}.$$

<u>Výberový koeficient ostrosti:</u> je definovaný vzťahom:

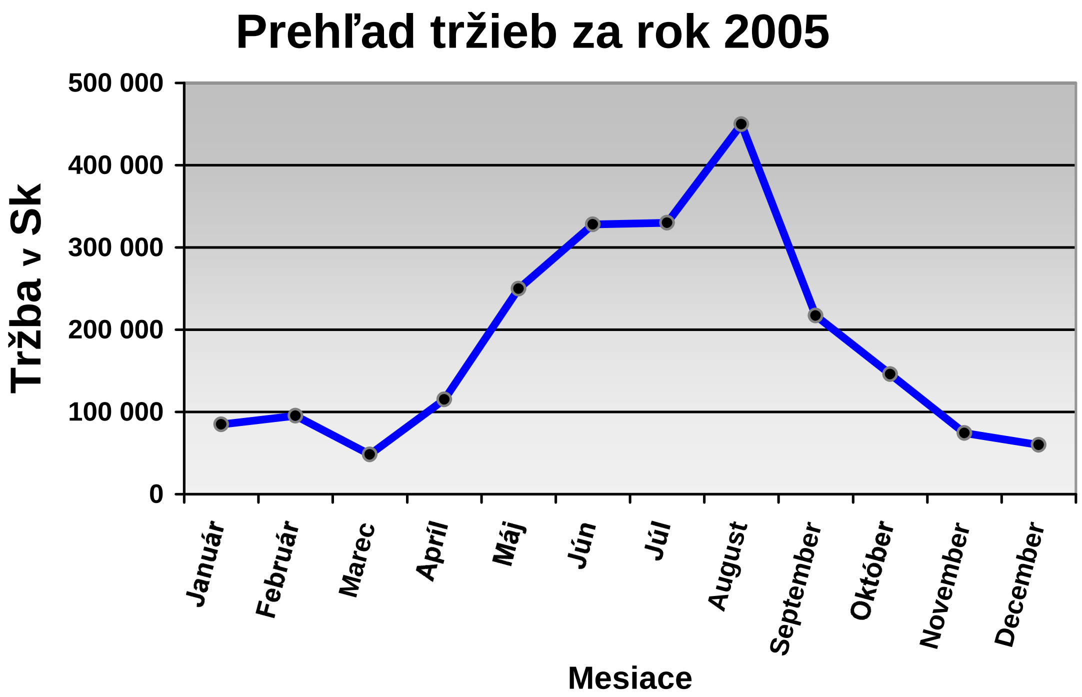
$$\gamma_2 = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3.$$

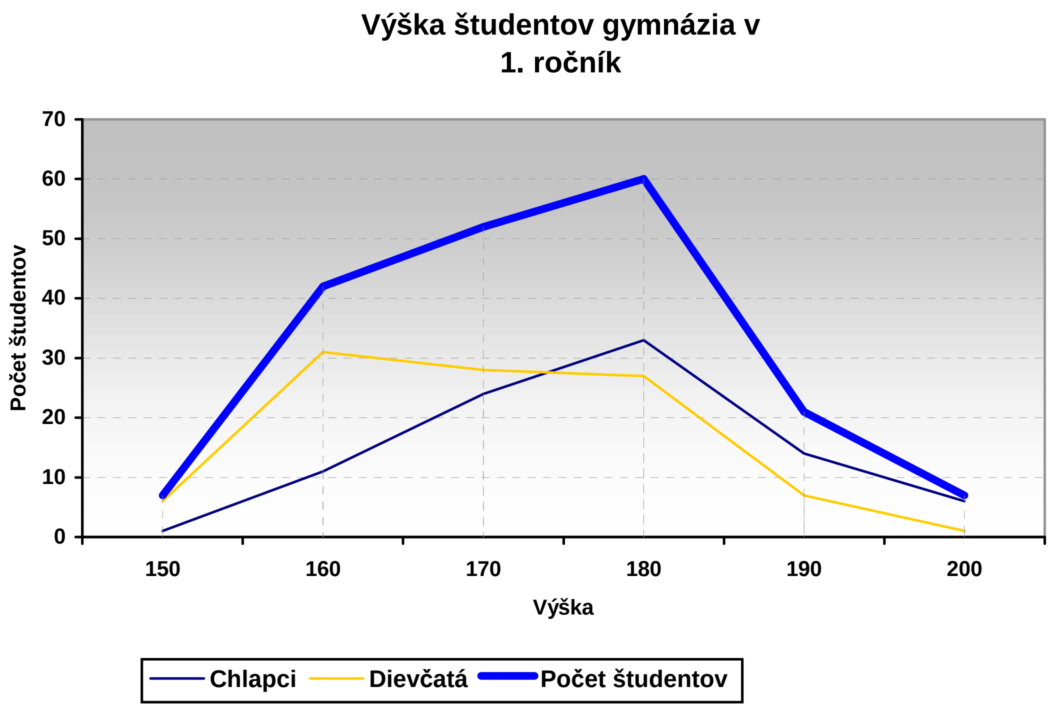
<u>Kvantilový koeficient ostrosti:</u> je definovaný pre $0 < p < 0{,}5$ vzťahom:

$$\frac{x_{(n)} - x_{(1)}}{x_{1-p} - x_p}.$$

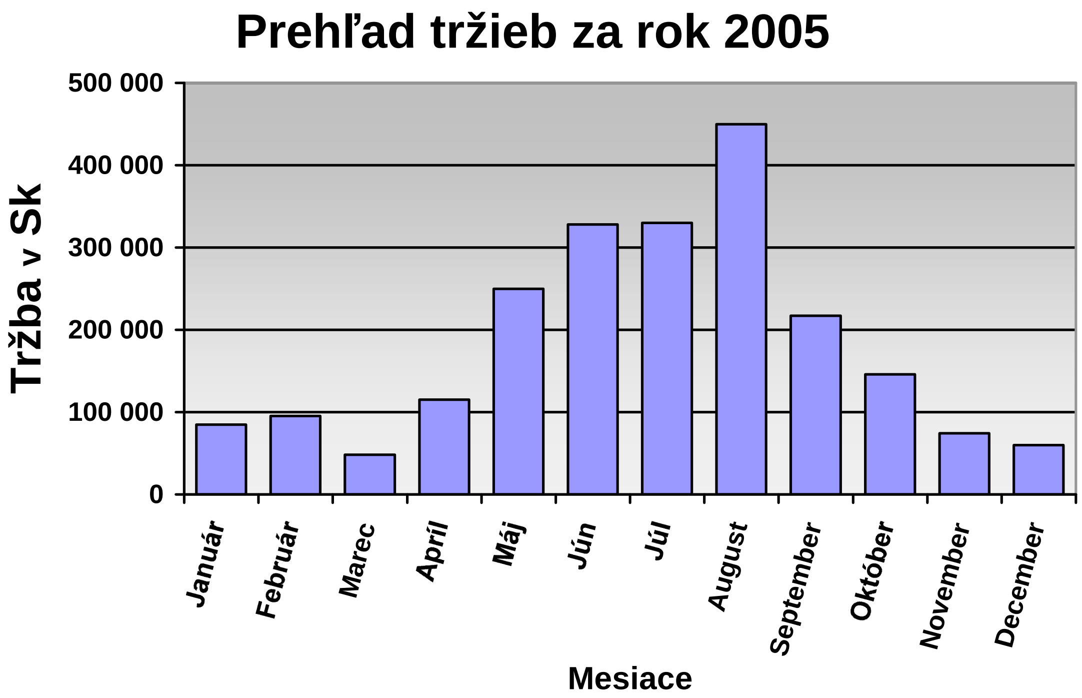
Grafy (diagramy): Ukážeme základné typy diagramov a grafov.

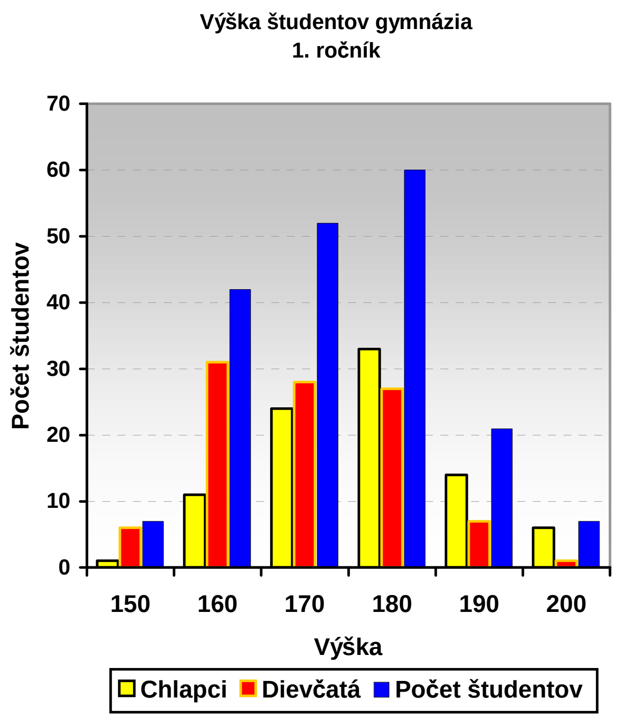
(1) Polygón resp. polygón početností:

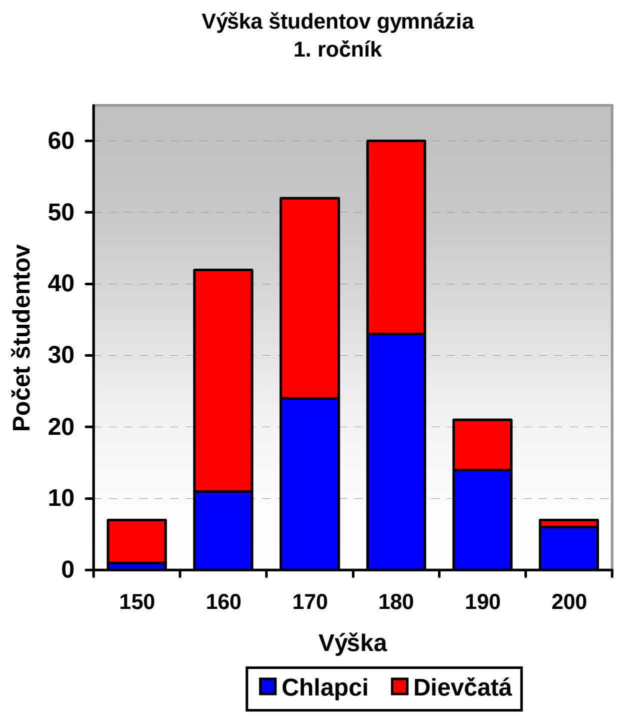




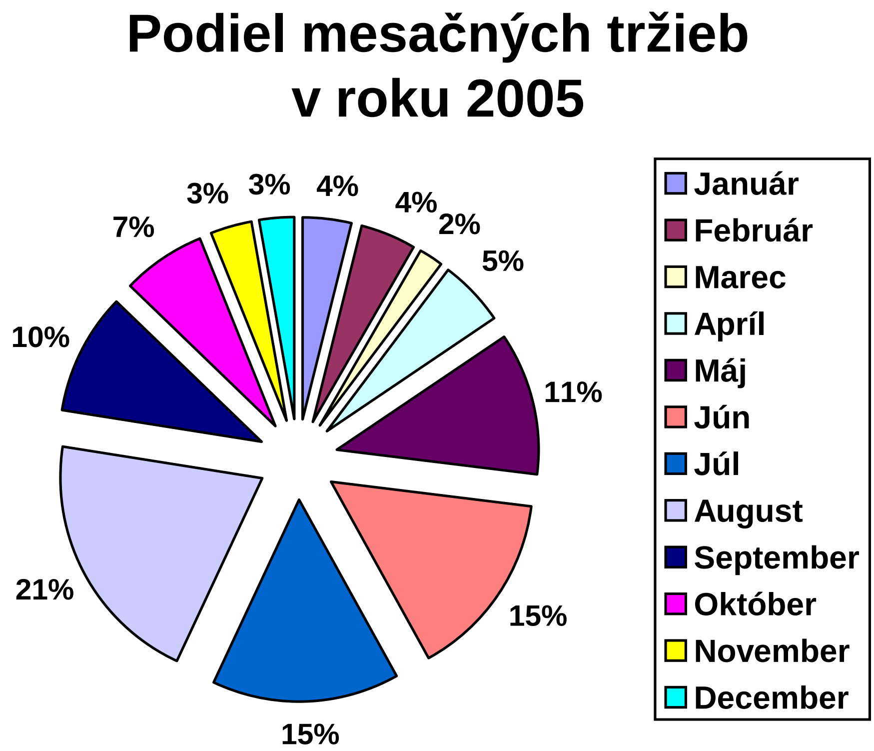
(2) Histogram:


Prehľad tržieb za rok 2005
Tržba v Sk
500 000
400 000
300 000
200 000
100 000
0
Január
Február
Marec
Apríl
Máj
Jún
Júl
August
September
Október
November
December
Mesiace


Výška študentov gymnázia
1. ročník
Počet študentov
70
60
50
40
30
20
10
0
150
160
170
180
190
200
Výška
Chlapci
Dievčatá
Počet študentov


Výška študentov gymnázia
1. ročník
Počet študentov
60
50
40
30
20
10
0
150
160
170
180
190
200
Výška
Chlapci
Dievčatá

(3) <u>Kruhový (koláčový):</u>



(4) <u>Prstencový (pre viacero radov):</u>