

Základy štatistiky

Základy štatistiky

- Úvod
- Základné pojmy
- Popisná štatistika
 - Triedenie
 - Tabuľky rozdelenia početností
 - Grafické znázornenie
- Charakteristiky štatistického znaku
 - charakteristiky polohy (**priemer, modus, medián**)
 - charakteristiky variability (**rozptyl, smerodajná odchyľka**)

Úvod

- **Historický základ štatistiky**
 - opis obyvateľstva, územia, hospodárskeho a politického stavu štátu
- **Lat. slovo „status“= štát; stav**
- **Vývoj**
 - popis zemepisného a hospodárskeho stavu štátu → popisná štatistika
 - matematická štatistika – využitie poznatkov z matematiky
- **Súčasnosť**
 - skúmanie stavu a vývoja hromadných javov
- **Hromadný jav**
 - prírodný alebo spoločenský jav, ktorý sa vyskytuje vo veľkom počte nositeľov tohto javu
- **Štatistika**
 1. praktická činnosť
 2. štatistické údaje
 3. vedná disciplína

Základné údaje

Základné pojmy

- **Štatistický súbor**
 - konečná neprázdna množina prvkov, na ktorých pozorujeme určitý znak, vlastnosť;
rozsah – počet prvkov štatistického súboru;
 - základný súbor;
 - výberový súbor;
- **Štatistická jednotka**
 - základný prvok štatistického súboru, ktorý je predmetom štatistického skúmania;
- **Štatistický znak (premenná)**
 - vlastnosť štatistickej jednotky, merateľná alebo opisateľná;
 - vymedzenie: vecné, časové, priestorové;

Základné pojmy

- **Štatistické znaky**

- **kvalitatívne**

- **nominálne** (názvové) – ich úrovne popisujeme slovne (národnosť, farba očí, krvná skupina, ...)
 - dichotomický (alternatívny) – 2 úrovne (áno/nie, muž/žena, ...)
 - **ordinálne** – ich hodnoty vieme zoradiť do poradia (prospech, veľkosť odevu S, M, L, XL, kvalita vody ...)

- **kvantitatívne** (kardinálne)

- ich hodnoty sú reálne čísla
 - o 2 hodnotách viem povedať, či sú rovnaké/rozdielne, môžem ich usporiadať do poradia a určiť o koľko je jedna väčšia ako druhá (rozloha, počet obyvateľov, vek, príjem, ...)
 - **diskrétne** (počet detí, počet bodov v teste, ...)
 - **spojité** (výška, hmotnosť, teplota, vek, ...)

Popisná štatistika

- **Triedenie dát**

- usporiadanie štatistických jednotiek do skupín/tried na základe hodnôt štatistického znaku;
- výsledok: rozdelenie početností;
- zásady:
 - úplnosť – každá nameraná hodnota je zatriedená;
 - jednoznačnosť – každá štat. jednotka je zaradená práve do jednej triedy/intervalu;
- typy triedenia
 - jednoduché triedenie;
 - intervalové/skupinové triedenie;

Popisná štatistika

- **Jednoduché triedenie**

Príklad: Žiaci 6.A dostali z písomnej práce z matematiky nasledovné známky:

3, 2, 1, 1, 2, 2, 3, 4, 4, 1, 2, 3, 3, 3, 5, 4, 2, 2, 2, 3, 3, 3, 3, 1, 3. Zostavte tabuľku rozdelenia početností.

- štatistický súbor: žiaci 6. A, ktorí písali písomku z matematiky
- štatistická jednotka: žiak
- štatistický znak (X): známka z matematiky, nadobúda 5 úrovní: 1, 2, 3, 4, 5
- rozsah súboru $n = 25$

Základné údaje

- Jednoduché triedenie:
tabuľka rozdelenia početností

Úrovne znaku x_i	Absolútna početnosť n_i	Kumulatívna početnosť	Relatívna početnosť n_i / n	Relatívna početnosť v % $(n_i / n) \cdot 100$
1	4	4	$4/25 = 0,16$	16
2	7	$4+7 = 11$	$7/25 = 0,28$	28
3	10	$11+10 = 21$	$10/25 = 0,4$	40
4	3	$21+3 = 24$	$3/25 = 0,12$	12
5	1	$24+1 = 25$	$1/25 = 0,04$	4

25

1

100

Popisná štatistika

- **Intervalové/skupinové triedenie**
 - štatistický znak má veľký počet úrovní, príp. máme veľký rozsah štatistického súboru
 - počet intervalov k môžeme zvoliť sami alebo použiť pravidlo $k \cong \sqrt{n}$;
 - šírka intervalu h : $h = \frac{x_{max} - x_{min}}{k}$;
 - každému intervalu priradíme hodnotu znaku rovnú stredu intervalu

Popisná štatistika

- **Intervalové/skupinové triedenie**

Príklad: Pri meraní výšky 14 žiakov sme získali hodnoty v cm: 158, 161, 168, 163, 174, 179, 176, 172, 178, 168, 179, 174, 173, 179. Zostavte tabuľku rozdelenia početností pre namerané výšky žiakov.

- Štatistický súbor: 14 žiaci ($n=14$ je rozsah súboru)
Štatistický prvok: žiak
Štatistický znak (X): výška žiakov
- počet intervalov $k \cong 4$
- šírka intervalu $h = (179-158)/4 \cong 5$

Popisná štatistika

- Intervalové triedenie:
tabuľka rozdelenia početností

Interval	Stred x_i	Absolútna početnosť n_i	Kumulatívna početnosť	Relatívna početnosť n_i / n	Relatívna početnosť v % $(n_i / n) \cdot 100$
<155,160)	157,5	1	1	$1/14 = 0,07$	7
<160,165)	162,5	2	3	$2/14 = 0,14$	14
<165,170)	167,5	2	5	$2/14 = 0,14$	14
<170,175)	172,5	4	9	$4/14 = 0,29$	29
<175,180)	177,5	5	14	$5/14 = 0,36$	36

14

1

100

Popisná štatistika

- **Grafické znázornenie**

- slúži pre rýchlu a názornú prezentáciu štatistických údajov

- najčastejšie používané grafy:

- Histogram

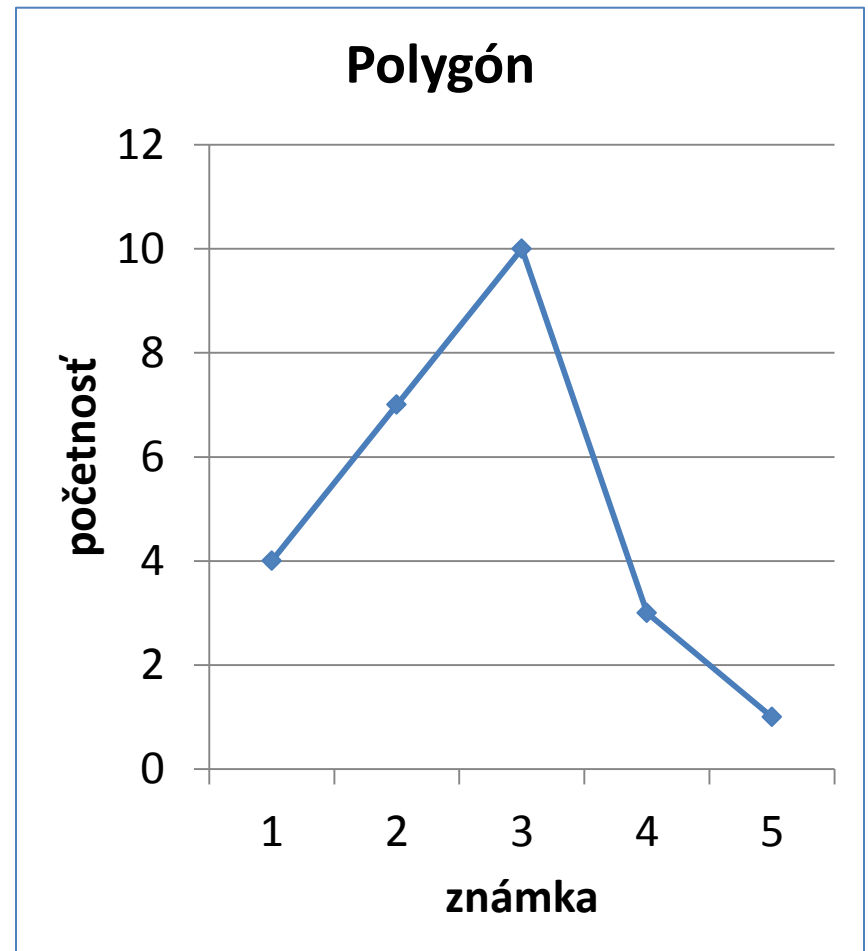
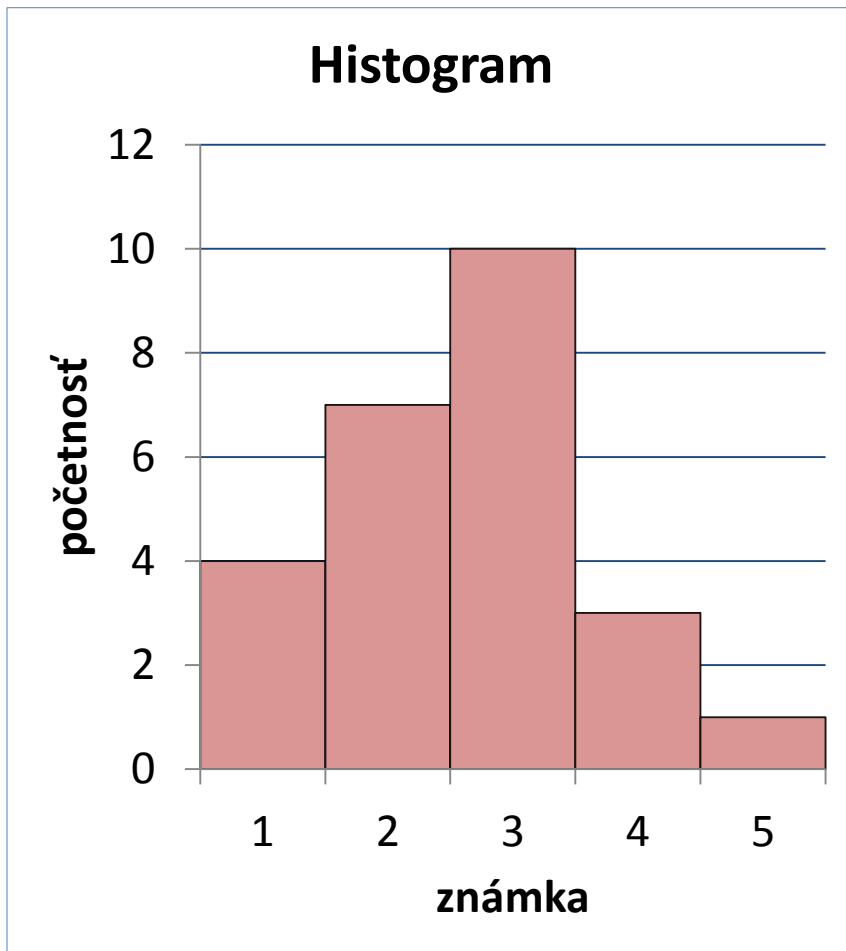
- stĺpcový graf, používaný hlavne pri intervalovom triedení

- Polygón

- spojnicový graf (vznikne spojením bodov, ktorých súradnice sú stredy intervalov a ich početností)

Popisná štatistika

- Grafické znázornenie



Charakteristiky štatistického znaku (číselné hodnoty, ktoré reprezentujú celý súbor)

- **Charakteristiky polohy (stredné hodnoty)**

– poskytujú predstavu o celom sledovanom štatistickom súbore vo forme jednej alebo viacerých charakteristík

- **aritmetický priemer**

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_k}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

v prípade tabuľky rozdelenia početnosti:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n} = \frac{1}{n} \sum_{i=1}^k x_i n_i$$

- **modus** – najčastejšie sa vyskytujúca hodnota

- **medián** – prostredná hodnota znaku v usporiadanej neklesajúcej postupnosti hodnôt $x_1 \leq x_2 \leq \dots \leq x_n$

Charakteristiky štatistického znaku

(číselné hodnoty, ktoré reprezentujú celý súbor)

- **Charakteristiky variability**

– charakterizujú mieru rozptýlenia hodnôt znaku okolo ich strednej hodnoty

- **rozptyl (s^2)**

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

v prípade tabuľky rozdelenia početnosti:

$$s^2 = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 n_i$$

- **smerodajná odchýlka (s)**

$$s = \sqrt{s^2}$$

Charakteristiky štatistického znaku

Výpočet v programe MS Excel:

aritmetický priemer

=AVERAGE(rozsah dát)

modus

=MODE(rozsah dát)

medián

=MEDIAN(rozsah dát)

rozptyl

=VAR(rozsah dát);

=VAR.S(rozsah dát)

smerodajná odchýlka

=SQRT(rozptyl)

Ponuka:

Údaje → Data Analysis → Descriptive Statistics

