

Popisná štatistika

Matematická štatistika sa zaoberá s náhodnými pokusmi; zbieraním údajov; z nich vytvára štatistické súbory. Potom tie súbory štatistické analyzuje, spracuje. A na konci vyvodí nejaké výsledky. Preto štatistickú činnosť môžeme rozdeliť do niekoľkých častí:

- príprava štatistického zisťovania
- získavanie údajov
- spracovanie štatistického súboru
- vyhodnotenie
- aplikácia výsledkov

štatistický súbor – konečná množina M obsahujúca štatistické jednotky, ktoré majú určité vlastnosti spoločné a niektoré odlišné

pr. trieda 1.A školy SPŠ stavebná OW v Lučenci v školskom roku 2019/2020 k dátumu 1.5.2020
→ vymenovali sme spoločné vlastnosti, tým sme jednoznačne vymedzili štatistický súbor

štatistická jednotka – prvok štatistického súboru ktorý je predmetom štatistického pozorovania
pr. sú to žiaci tej triedy

základný štatistický súbor – súhrn tých štatistických jednotiek, ktoré prichádzajú do úvahy pri štatistickom zisťovaní

štatistické zisťovanie môže byť:

úplné zisťovanie – zisťovanie na všetkých štatistických jednotkách základného súboru (práce, nákladné, možno až neuskutočniteľné)

výberové zisťovanie – je zamerané na časť jednotiek základného súboru (náhodným výberom)

rozsah štatistického súboru – počet štatistických jednotiek (prvkov) v súbore: n

pr. $n = 24$ (v 1.A je teraz 24 žiakov)

štatistický znak – vlastnosť ktorá je predmetom skúmania: x, y, z

pr. výška, váha, pohlavie, farba očí

a, **kvalitatívny štatistický znak** – nedá sa vyjadriť (merať) číslom: pohlavie, farba očí

b, **kvantitatívny štatistický znak** – nadobúda číselné hodnoty: výška, váha

hodnota štatistického znaku – hodnota štatistického znaku x štatistickej jednotky: $x_1; x_2; x_3; \dots; x_n$

pr. x – pohlavie: $x_1 =$ dievča; $x_2 =$ chlapec; $x_3 =$ chlapec; \dots ; $x_{24} =$ dievča

absolútna početnosť – počet, koľkokrát sa v štatistickom súbore vyskytuje hodnota x_i : n_i

$n_1; n_2; n_3; \dots; n_r$

platí: $r \leq n$;

$$n_1 + n_2 + n_3 + \dots + n_r = n$$

pr. $n_1 = 6$; $n_2 = 18$

relatívna početnosť – podiel absolútnej početnosti a rozsahu štatistického súboru: $\frac{n_i}{n}$

Charakteristiky úrovne (polohy)

Charakteristiky úrovne určia „strednú“ hodnotu štatistického znaku, ktorou by sme mohli charakterizovať celý štatistický súbor. Sem patria rôzne priemery, modus a medián. Všetky tie charakteristické hodnoty padajú medzi najmenšiu a najväčšiu hodnotu štatistického znaku.

1, **aritmetický priemer** – súčet čísel vydáme s počtom čísel

$$\bar{x}_A = \frac{a_1 + a_2 + a_3 + \dots + a_n}{n} = \frac{1}{n} \sum_{i=1}^n a_i$$

2, **geometrický priemer** – zo súčinu čísel toľkú odmocninu počítame, koľko ich bolo

$$\bar{x}_G = \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \cdot \dots \cdot a_n} = \sqrt[n]{\prod_{i=1}^n a_i}$$

3, **harmonický priemer** – počet čísel vydělíme so súčtom prevrátaných hodnôt čísel

$$\bar{x}_H = \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} + \dots + \frac{1}{a_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{a_i}}$$

4, **kvadratický priemer** – z aritmetického priemeru druhých mocnín počítame druhú odmocninu

$$\bar{x}_Q = \sqrt{\frac{a_1^2 + a_2^2 + a_3^2 + \dots + a_n^2}{n}} = \sqrt{\frac{1}{n} \sum_{i=1}^n a_i^2}$$

5, **modus (modálna hodnota)** – najčastejšia hodnota štatistického znaku, ak existuje (ak sú dve alebo viac s rovnakou absolútnou početnosťou, potom neexistuje): \hat{x}

6, **medián (stredná hodnota)** – stredná hodnota po usporiadaní podľa veľkosti hodnoty štatistického znaku (pri nepárnom počte štatistických jednotiek je to konkrétna hodnota; ak je počet páry, potom sa rovná aritmetickému priemeru dvoch prostredných hodnôt): \tilde{x}

V. $\bar{x}_H \leq \bar{x}_G \leq \bar{x}_A \leq \bar{x}_Q$

P. Rovnosť medzi priermi iba vtedy nastane, ak všetky hodnoty (čísla) sú rovnaké.

príklad:

Jeden štatistický znak (známky z MAT) má nasledujúce hodnoty: 1; 3; 2; 2; 1; 4; 3; 2

$$\bar{x}_A = \frac{1+3+2+2+1+4+3+2}{8} = \frac{18}{8} = 2,25$$

$$\bar{x}_G = \sqrt[8]{1 \cdot 3 \cdot 2 \cdot 2 \cdot 1 \cdot 4 \cdot 3 \cdot 2} = \sqrt[8]{288} = 2,03$$

$$\bar{x}_H = \frac{8}{\frac{1}{1} + \frac{1}{3} + \frac{1}{2} + \frac{1}{2} + \frac{1}{1} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2}} = \frac{8}{\frac{53}{12}} = \frac{96}{53} = 1,81$$

$$\bar{x}_Q = \sqrt{\frac{1^2+3^2+2^2+2^2+1^2+4^2+3^2+2^2}{8}} = \sqrt{\frac{1+9+4+4+1+16+9+4}{8}} = \sqrt{\frac{49}{8}} = 2,47$$

x_i	n_i
1	2
2	3
3	2
4	1

$$\hat{x} = 2$$

usporiadame podľa veľkosti: 1; 1; 2; 2; 2; 3; 3; 4

sú dve prostredné → počítame ich aritmetický priemer

$$\tilde{x} = \frac{2+2}{2} = 2$$

Charakteristiky variability (premenlivosti)

Charakteristiky úrovne slúžia na najjednoduchšie porovnávanie štatistických súborov. Samozrejme keby sme skúmali jednotlivé štatistické znaky iba z tohto hľadiska, možno by sme posúdili nasledujúce dva znaky za podobné (počítame aritmetický priemer):

x_i	y_i
98	40
99	70
100	110
101	130
102	150
$\bar{x} = 100$	$\bar{y} = 100$

Každý vidí, že sú úplne iné tie dva štatistické znaky, ktoré sa nepodobajú. Preto potrebujeme nejakým spôsobom (číslami) vyjadriť aj to, že aké „pestré“ sú tieto hodnoty znaku. Práve to vyjadrujú charakteristiky variability.

7. **variálné rozpätie** – rozdiel medzi najväčšou a najmenšou hodnotou štatistického znaku

$$R_x = x_{\max} - x_{\min}$$

8. **rozptyl (variancia)** – priemerná kvadratická odchýlka od aritmetického priemeru štatistického znaku

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

9. **výberový rozptyl**

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

10. **smerodajná (štandardná) odchýlka** – odmocnina rozptylu

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

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

11. **výberová smerodajná odchýlka**

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

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

12. **variálny koeficient** – relatívna miera variability

$$V_x = \frac{\sigma_x}{\bar{x}}$$

13. **priemerná odchýlka** – priemerná vzdialenosť od aritmetického priemeru štatistického znaku

$$\bar{d}_{\bar{x}} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

príklad:

Vypočítajme pre štatistické znaky x a y charakteristiky variability

x_i	$x_i - \bar{x}$	$ x_i - \bar{x} $	$(x_i - \bar{x})^2$	
98	-2	2	4	
99	-1	1	1	
100	0	0	0	
101	1	1	1	
102	2	2	4	
500	0	6	10	súčet
100	0	1,2	2	priemer

$$R_x = x_{\max} - x_{\min} = 102 - 98 = 4$$

$$\sigma_x^2 = \frac{10}{5} = 2$$

$$s_x^2 = \frac{10}{4} = 2,5$$

$$\sigma_x = \sqrt{2} = 1,414 2$$

$$s_x = \sqrt{2,5} = 1,581 1$$

$$V_x = \frac{\sqrt{2}}{100} = 0,0141$$

$$\bar{d}_x = \frac{6}{5} = 1,2$$

y_i	$y_i - \bar{y}$	$ y_i - \bar{y} $	$(y_i - \bar{y})^2$	
40	-60	60	3 600	
70	-30	30	900	
110	10	10	100	
130	30	30	900	
150	50	50	2 500	
500	0	180	8 000	súčet
100	0	36	1 600	priemer

$$R_y = y_{\max} - y_{\min} = 150 - 40 = 110$$

$$\sigma_y^2 = \frac{8\,000}{5} = 1\,600$$

$$s_y^2 = \frac{8\,000}{4} = 2\,000$$

$$\sigma_y = \sqrt{1\,600} = 40$$

$$s_y = \sqrt{2\,000} = 44,7214$$

$$V_y = \frac{40}{100} = 0,4$$

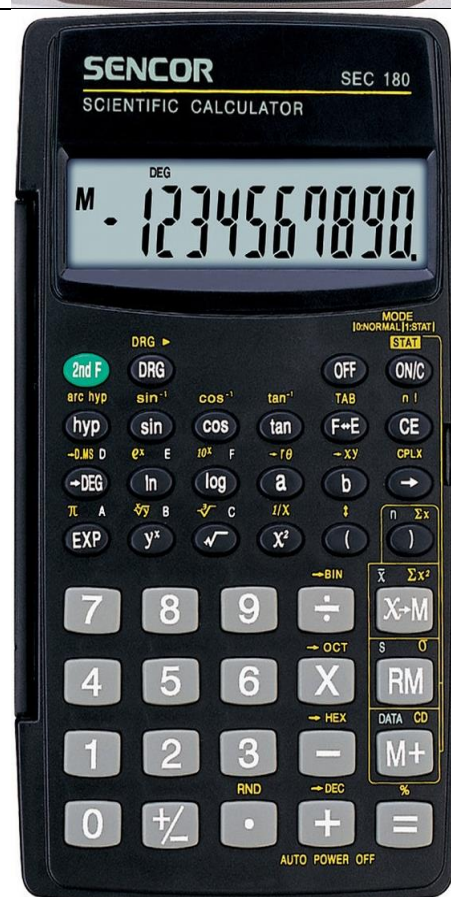
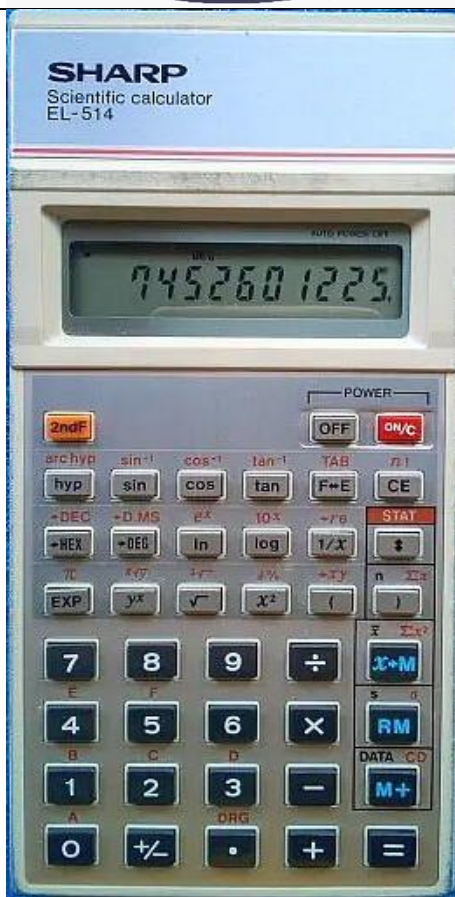
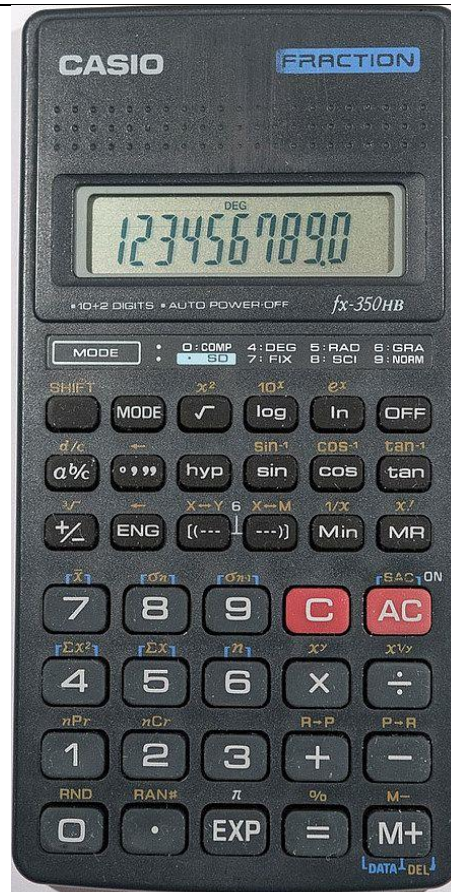
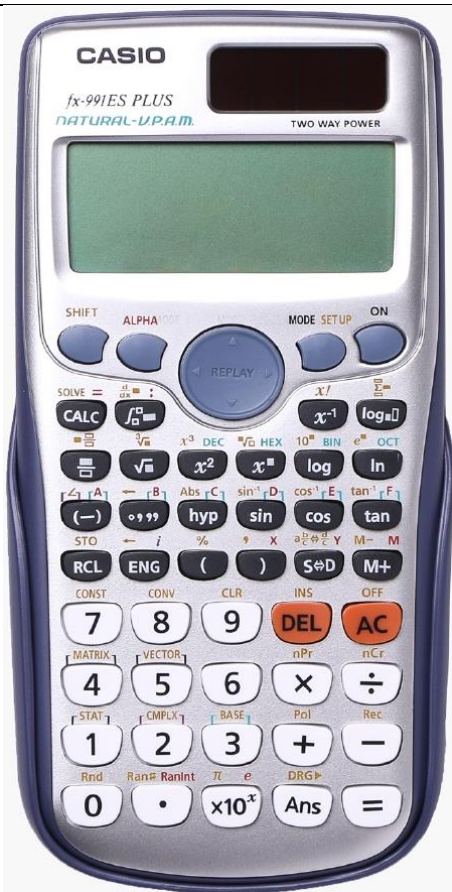
$$\bar{d}_y = \frac{180}{5} = 36$$

Vidíte, v týchto hodnotách už sú veľké rozdiely. Hoci aritmetické priemery vychádzajú na rovnaké čísla, tieto hodnoty ukazujú, že medzi hodnotami štatistického znaku y sú veľké, rádovo väčšie rozdiely.

V poslednom čase ale na papieri tieto charakteristiky nikto nepočíta. Aj preto, lebo rozsah štatistického súboru, ktoré spracúvajú, je veľký. Firmy, ktoré sa zaoberajú štatistickou činnosťou si nemôžu dovoliť aby celý proces trval niekoľko týždňov – hlavne ak konkurencia používa iné pomôcky (špeciálny softvér na počítači).

Pomôcka môže byť kalkulačka, ale už ani to nevyužívajú. Jednak iba málo charakteristických údajov vypočítajú, a na druhej strane nemôžeme kontrolovať správnosť vložených údajov – ak sa pomýlime, a nezbadáme chybu, potom ju nevieme korigovať. Ale na menšie výpočty stačí aj kalkulačka.

Vedecké kalkulačky už dávno majú zabudovaný štatistický režim.



Prvý krok: prepnutie do režimu štatistických výpočtov – tento krok väčšinou automaticky maže štatistické pamäte, ak nie, hľadajte funkciu (pravdepodobne sekundárnu funkciu) **SAC** / **SC** (pravdepodobne sekundárna funkcia). Na niektorých kalkulačkách si musíme zvoliť aj typ štatistického súboru pri režime štatistických výpočtov.

- a, prvý typ: stlačiť kláves **MODE** (niekedy aj viackrát – potom *na displeji* sa objaví SD a s tým súvisiace číslo) / *na kalkulačke uvedené číslo* alebo kláves, čo potom treba stlačiť **STAT** / **SD** [**MODE** → 3: **STAT** / **MODE** →]

b, druhý typ: kláves **2nd F** a **STAT / SD** [sekundárna funkcia klávesov: **2ndF** → \updownarrow / **2ndF** → **On/C**]

P. Pozor! Ak opustíte režim štatistických výpočtov, strácate zadané údaje (väčšinou).

Druhý krok: zadávanie údajov:

a, prvý typ: údaje píšeme do stĺpca a za údajmi stlačíme kláves **=**; zadávanie údajov ukončíme klávesom **AC** (vidíme všetky zadané dáta v stĺpci – hýbeme sa šípkami hore-dole)

b, druhý typ: za údajmi stlačíme kláves **DATA / DT** [primárna funkcia **M+**]

P. Ak urobíme chybu, zlé číslo zadáme môžeme opraviť hneď, alebo aj neskôršie:

a, prvý typ: šípkami prejdeme na zlý údaj, zadáme správne číslo a stlačíme **=** – tým prepíše údaj

b, naľukáme zlú hodnotu a stlačíme funkciu **DEL / CD** [**SHIFT** → **M+** / **2ndF** → **M+**] – tým ju vymažeme; potom horeuvedeným spôsobom zadáme dobré číslo (2. krok)

Tretí krok: vyvolávanie štatistických charakteristík:

aritmetický priemer

a, **SHIFT** → **1**: **STAT** → **4**: **Var** → **2**: \bar{x} → **=**

b, **SHIFT** → **7**: \bar{x}

c, **x** → **M**: \bar{x}

smerodajná odchýlka

a, **SHIFT** → **1**: **STAT** → **4**: **Var** → **3**: σ_X → **=**

b, **SHIFT** → **8**: σ_n

c, **2ndF** → **RM**: σ

výberová smerodajná odchýlka

a, **SHIFT** → **1**: **STAT** → **4**: **Var** → **4**: s_X → **=**

b, **SHIFT** → **9**: σ_{n-1}

c, **RM**: s

Sú tam ešte pomocné hodnoty z ktorých sa počítajú tieto charakteristiky: Σx ; Σx^2 ; n .

Lepší spôsob na jednoduchú štatistickú analýzu ponúka nám počítač. Konkrétne **Excel**.

Hodnoty daného štatistického znaku môžeme umiestniť napríklad do jedného stĺpca tabuľky. Takže aj ďalšie stĺpce na ďalšie znaky môžeme pritom používať.

aritmetický priemer: **AVERAGE**

geometrický priemer: **GEOMEAN**

harmonický priemer: **HARMEAN**

modus: **MODE**

medián: **MEDIAN**

rozptyl: **VAR.P** (**VARP** – staršia verzia)

výberový rozptyl: **VAR.S** (**VAR** – staršia verzia)

smerodajná odchýlka: **STDEV.P** (**STDEVP** – staršia verzia)

výberová smerodajná odchýlka: **STDEV.S** (**STDEV** – staršia verzia)

priemerná odchýlka: **AVEDEV**

Môžete použiť aj inú bezplatnú aplikáciu, napríklad **Libre Office** – presne tie isté príkazy.

aritmetický priemer: **AVERAGE**

geometrický priemer: **GEOMEAN**

harmonický priemer: **HARMEAN**

modus: **MODE**

medián: **MEDIAN**

rozptyl: **VAR.P**

výberový rozptyl: **VAR.S**

smerodajná odchýlka: **STDEV.P**

výberová smerodajná odchýlka: **STDEV.S**

priemerná odchýlka: **AVEDEV**