Leíró statisztika
(Štatistika)

A matematikai statisztika véletlen kísérletekkel; adatok gyűjtésével – beszerzésével foglalkozik; majd ezekből statisztikai állományok készít. Végül következtetéseket von le ezekből. Ezért a statisztikai tevékenységet néhány résztevékenységre oszthatjuk:

- a statisztikai mintavétel előkészítése
- mintavétel – adatok beszerzése
- a statisztikai minta feldolgozása
- kiértékelés
- az eredmények alkalmazása

statisztikai sokaság (štatistický súbor) – egy véges M halmaz, mely olyan statisztikai egységeket tartalmaz, amik bizonyos tulajdonságaikban megegyeznek, másokban pedig különböznek

pl. a WO Építőipari Szakközépiskola 1.A osztálya a 2019/2020 tanévben 2020.5.1. dátummal
→ felsoroltuk a közös tulajdonságokat, és ezzel egyértelműen meghatároztuk a statisztikai sokaságot

statisztikai egység (štatistická jednotka) – a statisztikai sokaság egy eleme, mely a statisztikai megfigyelés tárgyát képezi

pl. az adott osztály diákjai

statisztikai minta (základný štatistický súbor) – azon statisztikai egységek összessége, melyek a mintavételnél számításba jöhetnek
a statisztikai megfigyelés lehet:

teljes megfigyelés (úplné zisťovanie) – a minta összes egységét vizsgálja (munkaigényes, drága, esetleg megvalósíthatatlan)

kiválasztásos megfigyelés (výberové zisťovanie) – a mintának egy leszűkített egységeit vizsgálja (véletlenszerű választással)

a statisztikai sokaság nagysága (rozsah štatistického súboru) – a statisztikai egységek száma a statisztikai sokaságban: n

pl. n = 24 (az 1.A-nak akkor 24 diákja volt)

statisztikai ismérv/változó (štatistický znak) – azon tulajdonság, mely a vizsgálódás tárgyát képezi: x, y, z

pl. magasság, tömeg, a szem színe

a, ***minőségi ismérv*** (kvalitatívny štatistický znak) – nem lehet számmal kifejezni (mérni): nem, a szem színe
b, ***mennyiségi ismérv*** (kvantitatívny štatistický znak) – számértéket vesz fel: magasság, tömeg

a statisztikai ismérv értéke (hodnota štatistického znaku) – az x ismérv statisztikai egységének az értéke: x_1; x_2; x_3; … ; x_n

pl. x – nem: x_1 = lány; x_2 = fiú; x_3 = fiú; … ; x_{24} = lány

abszolút gyakoriság (absolútna početnosť) – az a szám, ahányszor előfordul a statisztikai sokaságban az x_i érték: n_i

$n_1; n_2; n_3; \dots ; n_r$
érvényes: $r \le n$;
$n_1 + n_2 + n_3 + \dots + n_r = n$

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

relatív gyakoriság (relatívna početnosť) – az abszolút gyakoriság elosztva a statisztikai sokaság nagyságával: $\frac{n_i}{n}$

Helyzetparaméterek/elhelyezkedési mutatók (charakteristiky úrovne/polohy)
A helyzetparaméterek megadják az ismérv „középső” értékét, mellyel az egész statisztikai sokaságot jellemezhetjük. Ide tartoznak a különböző átlagok, a módusz és a medián. Ezen jellemzők értéke mindig az ismérv legkisebb és legnagyobb értéke közé esik.

1, ***a számtani közép*** (aritmetický priemer) – a számok összegét elosztjuk a darabszámmal

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

2, ***a mértani közép*** (geometrický priemer) – a számok szorzatából annyiadik gyököt vonunk, amennyi a tényező

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

3, ***a harmonikus közép*** (harmonický priemer) – a darabszámot elosztjuk a számok reciprokának összegével

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

4, ***a négyzetes/kvadratikus közép*** (kvadratický priemer) – a számok négyzetéből számtani közepet számolunk, majd ebből gyököt vonunk

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

5, ***a módusz*** (modus/modálna hodnota) – az ismérv leggyakoribb értéke, ha létezik (amennyiben két, vagy több értéknek is azonos az abszolút gyakorisága, akkor nem létezik): $\hat{x}$

6, ***a medián/helyzeti középérték*** (medián/stredná hodnota) – az ismérv sorba rendezett értékeiből a középső (páratlan számú statisztikai egység esetén ez egy konkrét érték; ha pedig páros számú, akkor a két középső érték számtani közepét jelenti): $\tilde{x}$

T. $\bar{x}_H \leq \bar{x}_G \leq \bar{x}_A \leq \bar{x}_Q$
M. Egyenlőség a középértékek között csak akkor áll fenn, ha minden érték (szám) azonos.

példa:

Egy statisztikai ismérv (osztályzatok matematikából) az alábbi értékekkel rendelkezik: 1; 3; 2; 2; 1; 4; 3; 2

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

$$\bar{x}_G = \sqrt[8]{1.3.2.2.1.4.3.2} = \sqrt[8]{288} = 2{,}03$$

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

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

x_i	n_i
1	2
2	3
3	2
4	1

$\hat{x} = 2$

nagyság szerint rendezzük: 1; 1; 2; 2; 2; 3; 3; 4
két középső van → az ő számtani közepét számoljuk

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

Szóródási mutatók (charakteristiky variability/premenlivosti)

A helyzetparaméterek a statisztikai sokaságok legegyszerűbb összehasonlítására szolgálnak. Természetesen, ha csak ezen szempontból vizsgálnánk az egyes ismérveket, akkor az alábbi két ismérvet akár hasonlóknak is mondhatnánk (a számtani közepet számoljuk):

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

Bárki láthatja, hogy ez a két ismérv teljesen különbözik egymástól, nem hasonlítanak egymásra. Ezért valamilyen módon (számokkal) szeretnénk kifejezni, hogy milyen „változatosak” az ismérv értékei. Pontosan ezt fejezik ki a szóródás mérőszámai.

7. ***a szórás terjedelme/mintaterjedelem*** (variačné rozpätie) – az ismérv legnagyobb és legkisebb értéke közötti különbség

8. (***tapasztalati***) ***szórásnégyzet/variancia*** (rozptyl/variancia) – az ismérv átlagos négyzetes eltérése a számtani középtől

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

9. ***korrigált*** (***tapasztalati***) ***szórásnégyzet*** (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. (***tapasztalati***) ***szórás*** (smerodajná/štandardná odchýlka) – a szórásnégyzet négyzetgyöke

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

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

11. ***korrigált*** (***tapasztalati***) ***szórás*** (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. ***relatív szórás/szórási együttható*** (variačný koeficient) – a variabilitás relatív mértéke

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

13. ***átlagos abszolút eltérés*** (priemerná odchýlka) – az ismérv átlagos távolsága a számtani középtől

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

példa:

Határozzuk meg az x és a y ismérvek szóródási mutatóit

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

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

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

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

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

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

$V_x = \frac{\sqrt{2}}{100} = 0{,}014\ 1$

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

y_i	$y_i - \bar{y}$	$\lvert y_i - \bar{y} \rvert$	$(y_i - \bar{y})^2$	
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	összeg
100	0	36	1 600	átlag

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

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

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

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

$s_y = \sqrt{2\ 000} = 44{,}721\ 4$

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

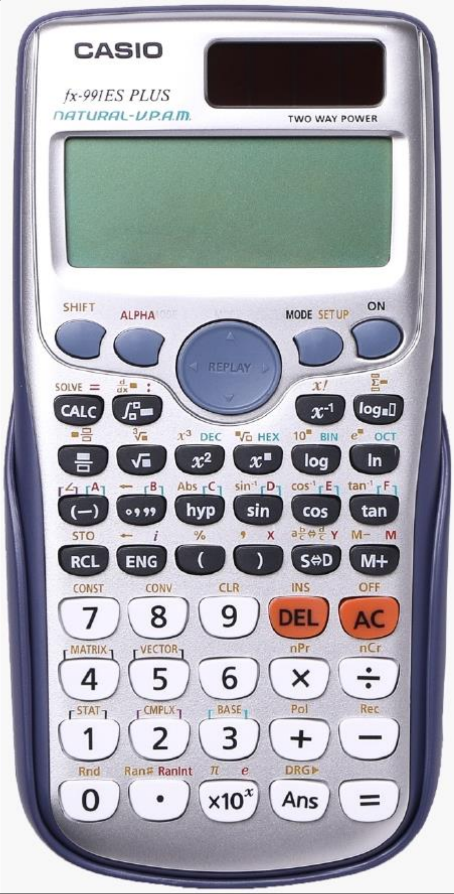
$\bar{d}_{\bar{y}} = \frac{180}{5} = 36$

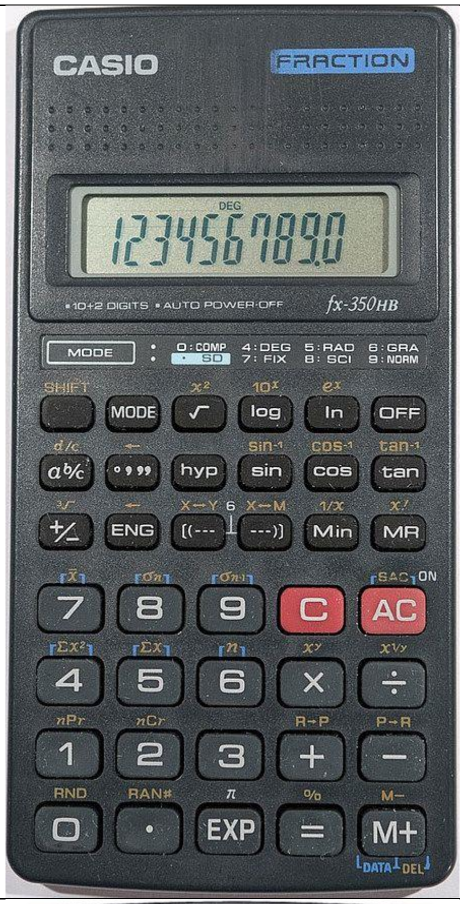
Láthatjuk, hogy ezekben a számokban már jelentős eltérések tapasztalhatók. Bár a számtani közepeik megegyeznek, ezek a mérőszámok már mutatják, hogy az y ismérv értékei között nagy, nagyságrendekkel nagyobb az eltérés.

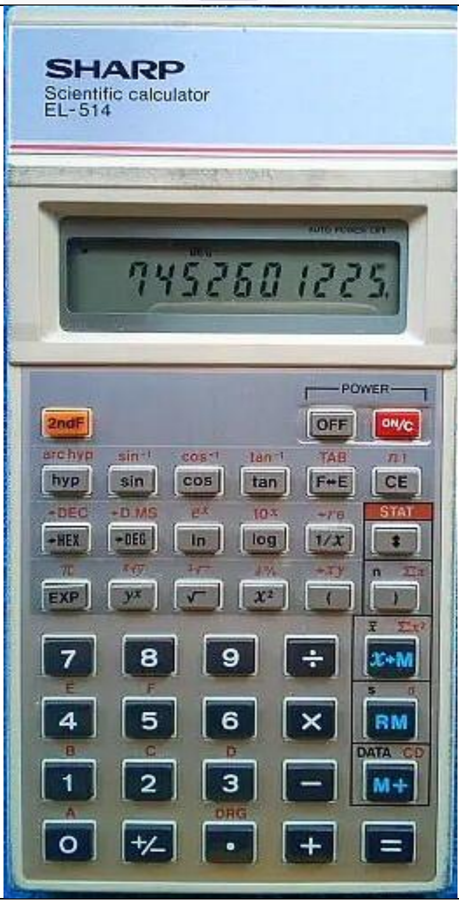
Az utóbbi időben ezeket a jellemzőket papíron már senki nem számolja. Azért is, mert a statisztikai sokaság, ami feldolgozásra kerül, hatalmas. Azok a cégek, akik statisztikai tevékenységgel foglalkoznak, nem engedhetik meg maguknak, hogy az egész folyamat hetekig tartson – főleg, ha a konkurencia egyéb segédeszközöket használ (speciális szoftvert a számítógépen).

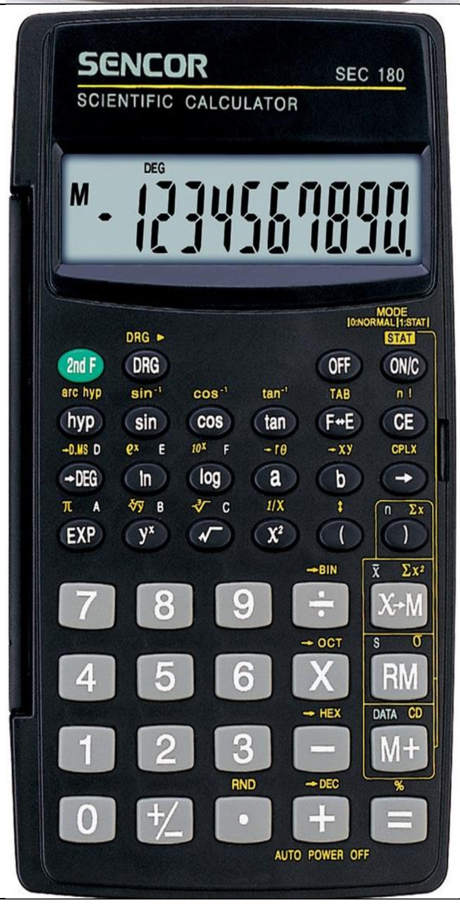
Segítségünkre lehet a számológép, de már azt sem igazán használják. Egyrészt kevés statisztikai mutatót lehet velük számolni, másrészt nem ellenőrizhető a bevitt adatok helyessége – ha megtévedünk, és nem vesszük ezt észre, később már nem tudjuk kijavítani. De a kisebb számításokhoz elegendő a számológép is.

A tudományos számológépekben régóta megtalálható a statisztika üzemmód.









Az első lépés: átkapcsolni statisztikai számítások üzemmódjába – ez általában automatikusan törli a statisztikához használt memóriákat. Ha nem, akkor keressétek a SAC / SC függvényt (valószínűleg másodlagos funkció). Némelyik számológépen a statisztikai számítások üzemmódban még a sokaság típusát is meg kell választani.

a, első típus: a MODE megnyomása (akár többször is – ekkor *a kijelzőn* megjelenik az SD és a hozzá tartozó szám) / a *számológépen feltüntetettszám szám* vagy *billentyű*, amit utána kell megnyomi STAT / SD [MODE → 3: STAT / MODE → ·]
b, második típus: a 2nd F billentyű és a STAT / SD [a billentyűzet másodlagos funkciója: 2ndF → ↕ / 2ndF → On/C]

M. Vigyázzunk! Ha kilépünk a statisztikai számítások üzemmódjából, elvesznek a bevitt adataink (általában).

A második lépés: az adatok bevitele:
a, első típus: az adatokat oszlopba visszük be, és minden adat után az = billentyűvel lépünk a következő sorba; végül az AC billentyűvel fejezzük be az adatbevitelt (láthatjuk az összes bevitt adatot az oszlopban – nyilakkal fel-le haladva)
b, második típus: az adat bevitele után megnyomjuk a DATA / DT billentyűt [az M+ az elsődleges funkció]

M. Ha hibázunk, és rossz adatot viszünk be, azonnal kijavíthatjuk, esetleg később is, csak tudnunk kell, hogy mit javítsunk mire:
a, első típus: a nyilak segítségével rámegyünk a rosszul bevitt számra, beírjuk a jó adatot, majd megnyomjuk az =-t – ezzel átírja a számot
b, második típus: beírjuk a rosszul megadott számot, majd megnyomjuk a DEL / CD [SHIFT → M+ / 2ndF → M+] függvényt – ezzel töröljük őt; ezek után a fenti módon bevisszük a jó számot (2. lépés)

A harmadik lépés: a statisztikai mutatók előhívása:
számtani közép
a, SHIFT → 1: STAT → 4: Var → 2: $\bar{x}$ → =
b, SHIFT → 7: $\bar{x}$
c, x→M: $\bar{x}$
szórás
a, SHIFT → 1: STAT → 4: Var → 3: σX → =
b, SHIFT → 8: σ_n
c, 2ndF → RM: σ
korrigált szórás
a, SHIFT → 1: STAT → 4: Var → 4: sX → =
b, SHIFT → 9: σ_{n-1}
c, RM: s

Még megtalálható néhány segédérték, amikből számolja a számológép ezen mutatókat: Σx; Σx^2; n.

Ennél jobb eszközt kínál a számítógép az egyszerű statisztikai elemzéshez. Konkrétan az Excel.
Egy statisztikai ismérv értékeit például elrendezhetjük a táblázat egy oszlopába. Vagyis további ismérveket további oszlopokban adhatunk meg.
számtani közép: AVERAGE
mértani közép: GEOMEAN
harmonikus közép: HARMEAN
módusz: MODE
medián: MEDIAN
szórásnégyzet: VAR.P (VARP – régebbi verzió)
korrigált szórásnégyzet: VAR.S (VAR – régebbi verzió)
szórás: STDEV.P (STDEVP – régebbi verzió)
korrigált szórás: STDEV.S (STDEV – régebbi verzió)
átlagos abszolút eltérés: AVEDEV

Használhatunk más, ingyenes applikációt is, például Libre Office – pontosan ugyanazok a parancsok.
számtani közép: AVERAGE
mértani közép: GEOMEAN
harmonikus közép: HARMEAN
módusz: MODE
medián: MEDIAN
szórásnégyzet: VAR.P
korrigált szórásnégyzet: VAR.S

szórás: STDEV.P
korrigált szórás: STDEV.S
átlagos abszolút eltérés: AVEDEV