A mérési eredmények jellemzésének matematikai-statisztikai alapjai

A statisztikai módszerek alkalmazásának okai

Ezek a módszerek lehetővé teszik a **mérések értékelését**, bizonytalanság esetén az okokra és a mérés egyéb összefüggéseire vonatkozó **döntéshozatalt**.

A valószínűségszámítás és a matematikai-statisztikai módszerek felhasználásával lehetőségünk nyílik a mérés során fellépő **véletlen hibák** mértékének **becslésére**.

A mérési sorozat statisztika jellemzői

Mérési sorozatunk elemeit a mérendő mennyiség valódi értékének meghatározása érdekében végzett **mintavétel** és **kísérlet** eredményének tekintve, alkalmazhatjuk rájuk a **minták statisztikai jellemzőinek** meghatározására szolgáló **statisztikai módszereket**.

A várható érték becslésére szolgáló jellemzők

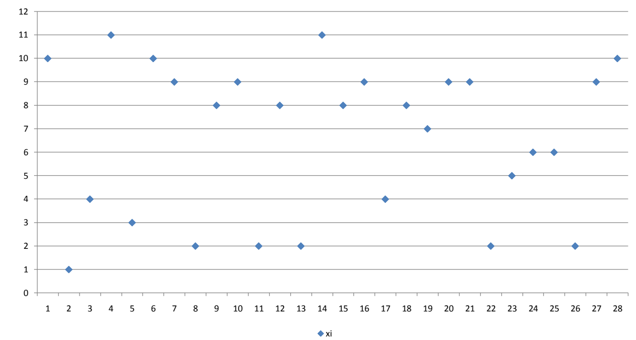
A mérési sorozat **várható értékének** – mint a mérendő mennyiség **esetleges korrigálatlan konvencionális valódi értékének** – becslésére a következő módszerek állnak a rendelkezésünkre:

- számtani átlag,
- geometriai átlag,
- harmonikus átlag,
- tapasztalati medián,
- tapasztalati módusz.

A várható érték becslésére szolgáló jellemzők

n	x_i	n	x_i	n	x_i
1	10	11	2	21	9
2	1	12	8	22	2
3	4	13	2	23	5
4	11	14	11	24	6
5	3	15	8	25	6
6	10	16	9	26	2
7	9	17	4	27	9
8	2	18	8	28	10
9	8	19	7		
10	9	20	9		

A várható érték becslésére szolgáló jellemzők



A számtani átlag

Számítása:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

Tulajdonságai:

- Minden átlagérték közül a számtani átlag a legnagyobb.

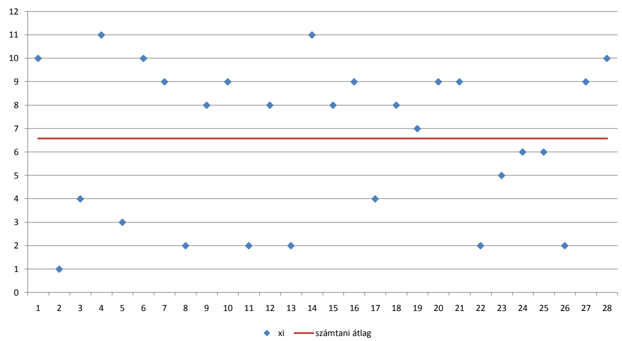
A számtani átlag

- A számtani átlagtól való **eltérések** összege **nulla**, az **eltérés négyzetek összege** pedig **minimális**:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

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

A számtani átlag



A geometriai átlag

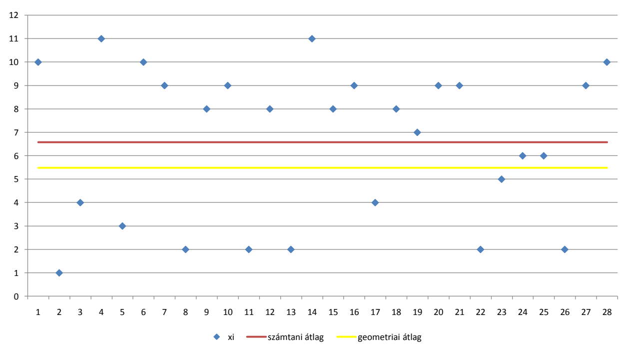
Számítása:

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n}$$

Tulajdonságai:

- A geometriai átlag mindig kisebb a számtani átlagnál.

A geometriai átlag



A harmonikus átlag

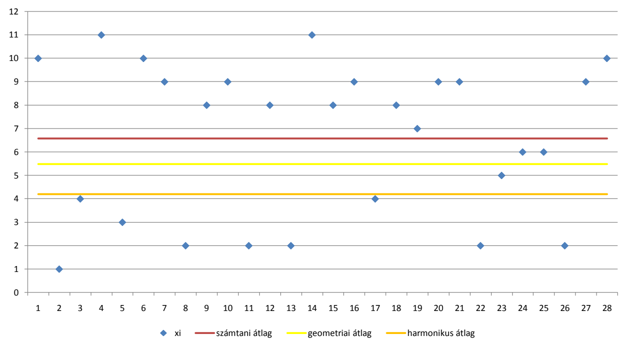
Számítása:

$$\frac{1}{\bar{x}_h} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i} = \frac{1}{n}\cdot\left(\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\ldots+\frac{1}{x_n}\right)$$

Tulajdonságai:

- Az előző három átlag között mindig fennáll a következő összefüggés: $\bar{x}_h \le \bar{x}_g \le \bar{x}$

A harmonikus átlag



A tapasztalati medián

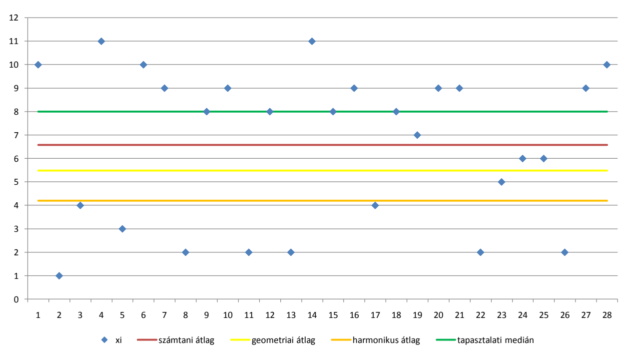
Számítása:

A medián (**Me**), vagy centrális érték a nagyság szerint rendezett mintaelemek **középső eleme**, ha **n páratlan**. Ha **n páros**, akkor a **két középső elem számtani közepe**.

Tulajdonságai:

- A medián csak szimmetrikus eloszlás esetén ad torzítatlan becslést.

A tapasztalati medián



A tapasztalati módusz

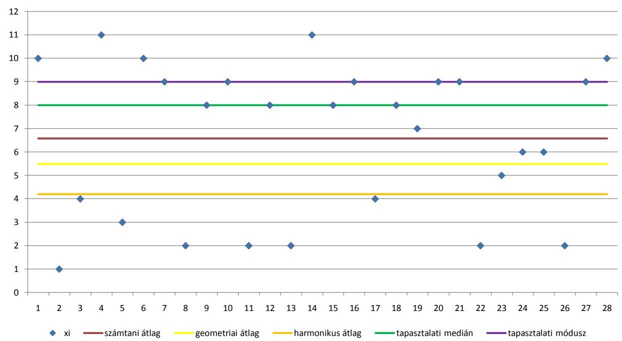
Számítása:

A módusz a mintában leggyakrabban előforduló elem.

Tulajdonságai:

- A módusz értéke nem feltétlenül egyértelmű, mivel ugyanazt a gyakoriságot több különböző érték is elérheti.

A tapasztalati módusz



A várható érték becslésére szolgáló jellemzők használhatósága

- A mérési sorozat várható értékének becslésére szolgáló jellemzők közül a **számtani átlag a legmegbízhatóbb** (torzítatlan, hatékony, konzisztens és elégséges).
- **Minden más átlagérték** (geometriai, harmonikus) – tekintettel arra, hogy kisebb mint a számtani átlag – a várható értékre **torzított becslést ad**.
- A **medián** ugyan szimmetrikus eloszlás esetében torzítatlan becslést ad, de ennek **hatékonysága** a számtani átlaghoz képest csak **63,7 %-os**.

A szóródás becslésére szolgáló jellemzők

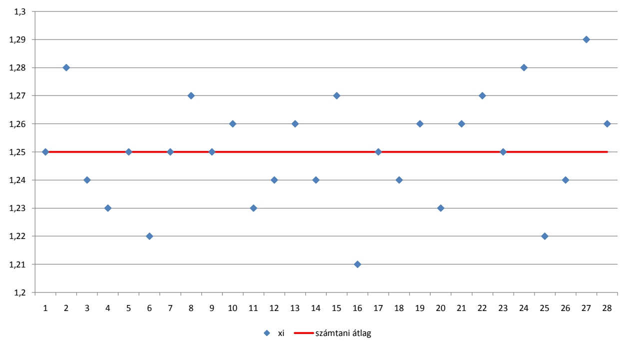
A mérési sorozat **szóródásának**– mint a mérés **hibájának** – becslésére a következő módszerek állnak a rendelkezésünkre:

- tapasztalati szórás,
- korrigált tapasztalati szórás,
- relatív szórás,
- a mérési sorozat átlagának korrigált tapasztalati szórása,
- mintaterjedelem.

A szóródás becslésére szolgáló jellemzők

n	x_i	n	x_i	n	x_i
1	1,25	11	1,23	21	1,26
2	1,28	12	1,24	22	1,27
3	1,24	13	1,26	23	1,25
4	1,23	14	1,24	24	1,28
5	1,25	15	1,27	25	1,22
6	1,22	16	1,21	26	1,24
7	1,25	17	1,25	27	1,29
8	1,27	18	1,24	28	1,26
9	1,25	19	1,26		
10	1,26	20	1,23		

A szóródás becslésére szolgáló jellemzők



A tapasztalati szórás

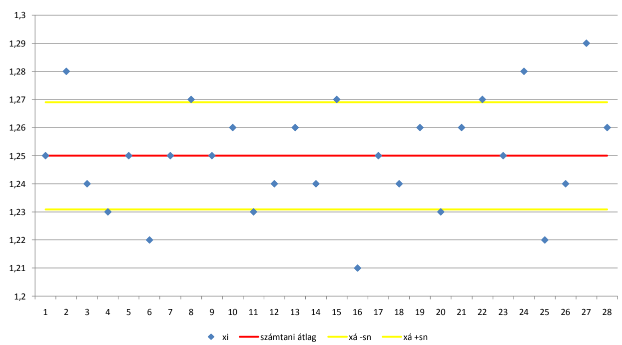
Számítása:

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

Tulajdonságai:

- A tapasztalati szórás nem más, mint az egyes eredmények matematikai átlagtól való eltéréseinek négyzetéből képzett matematikai átlag négyzetgyöke.
- A **tapasztalati szórás nem ad torzítatlan becslést** az alapsokaság szóródására vonatkozóan.

A tapasztalati szórás



A korrigált tapasztalati szórás

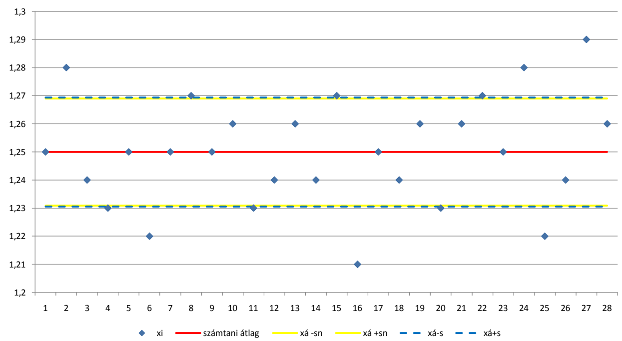
Számítása:

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

Tulajdonságai:

- A korrigált tapasztalati szórás mindig **nagyobb** a **tapasztalati szórásnál**, a különbség azonban n növekedésével csökken.
- A korrigált tapasztalati szórás **torzítatlan és konzisztens becslését** adja az alapsokaság szóródásának.

A korrigált tapasztalati szórás



A mérési sorozat átlagának korrigált tapasztalati szórása

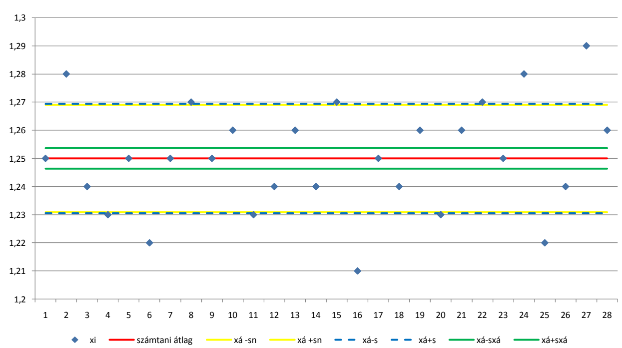
Számítása:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Tulajdonságai:

- A mérési sorozatok számtani átlagaiból képzett minta korrigált tapasztalati szórása **mindig kisebb** mint az **eredeti mérési sorozat korrigált tapasztalati szórása**.
- Ezt az összefüggést felhasználva elvileg bármilyen mértékben **csökkenthetjük** a mérésünk **szóródását** (**véletlen hibáját**, **bizonytalanságát**)

A mérési sorozat átlagának korrigált tapasztalati szórása



A relatív szórás

Számítása:

$$CV = \frac{s}{\bar{x}}$$

Tulajdonságai:

- A relatív szórás, vagy variációs együttható a szórás értékét az összehasonlíthatóság érdekében a számtani átlag arányában adja meg, általában 100-al megszorozva százalékos formában.

A mintaterjedelem

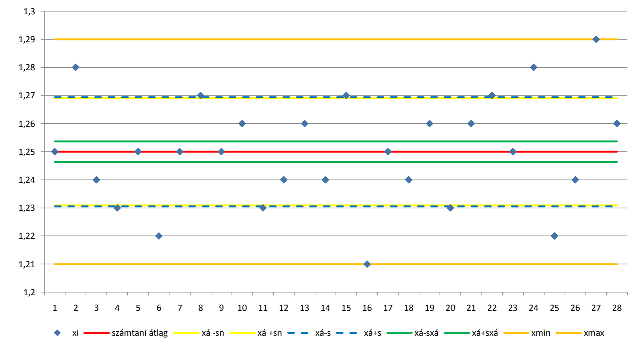
Számítása:

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

Tulajdonságai:

- A minta legnagyobb és legkisebb elemének különbsége.
- A mintaterjedelem lehetőséget ad a **szórás közelítő becslésére**.

A mintaterjedelem



A szórás közelítő becslése

Számítása:

$$s_{becs} = f \cdot R$$

n	f	E
2	0,886	1,00
4	0,486	0,95
6	0,395	0,93
8	0,351	0,89
10	0,325	0,85
20	0,268	0,70

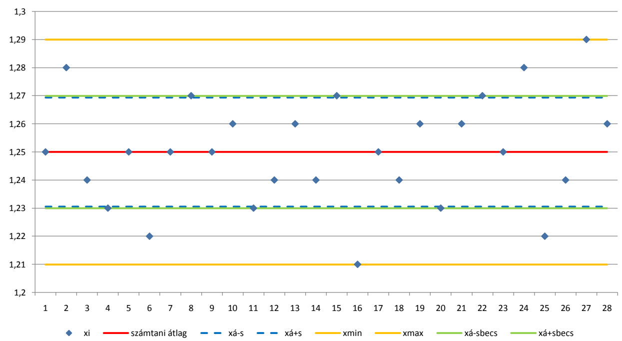
E: a becslés hatékonysága.

A szórás közelítő becslése

Tulajdonságai:

- A becslés **torzított** és **hatékonysága** (E) a mintaszám növelésével **romlik**.
- A szórás a **mintaterjedelemből** egy **szorzófaktor** (f) segítségével becsülhető, aminek értéke függ a **mintaelemek (n) számától**.
- A mintaterjedelem alapján történő becslés **nagy információveszteséggel** jár, de **gyors tájékozódást** tesz lehetővé.

A szórás közelítő becslése



A mérési sorozat eloszlása

Ahhoz, hogy az előzőekben meghatározott **statisztikai jellemzőket** fel tudjuk használni a mérési sorozat **statisztikai kiértékelésére**, - legalább közelítőleg - ismernünk kell a mérési sorozat **eloszlását**.

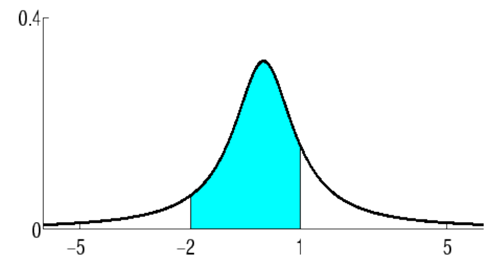
A **mérési sorozat eloszlásán** azt a **tulajdonságot** értjük, hogy hogy a mérési sorozat bármely **lehetséges érték-intervallumához** hozzá tudjuk rendelni annak a **valószínűségét**, hogy egy mérési eredmény ebbe az intervallumba esik.

A mérési sorozat eloszlása

A mérési sorozat eloszlását a **sűrűségfüggvénnyel** és az **eloszlásfüggvénnyel** tudjuk jellemezni.

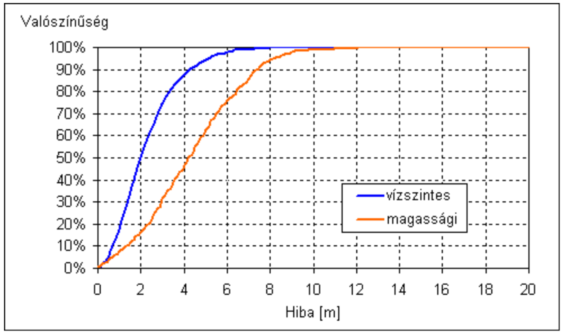
- A **sűrűségfüggvény** azt mutatja meg, hogy a mérési sorozat **egyes lehetséges értékei** mekkora **valószínűséggel** esnek egy kijelölt **intervallumba**.
- Az **eloszlásfüggvény** a sűrűségfüggvény **integrálfüggvénye**, és azt mutatja meg, hogy mennyi annak a **valószínűsége**, hogy a mérési eredmény egy **adott értéknél kisebb**.

A sűrűségfüggvény



Cauchy-eloszlás sűrűségfüggvénye

Az eloszlásfüggvény



A mérési sorozat eloszlásának tulajdonságai

- Az eloszlások egy részénél a **sűrűségfüggvény** és az **eloszlásfüggvény matematikai függvény** formájában leírható.
- A matematikai függvénnyel leírható eloszlások sűrűségfüggvényében szereplő **konstansok** és az azokból leszármaztatható mennyiségek az **eloszlás paraméterei**.
- A mérési sorozat statisztikai kiértékelésével az **eloszlásfüggvény alakját** és a függvényben szereplő **paramétereket** akarjuk meghatározni.

Az eloszlás paraméterei

- Az eloszlás **várható értéke**.
- Az eloszlás **mediánja** (az az érték, amelyre igaz, hogy a kisebb és nagyobb érték valószínűsége egyaránt 50%).
- Az eloszlás **módusza** (a sűrűségfüggvény maximumának helye).
- Az eloszlás **varianciája** (a mérési eredmények várható értéktől való eltérésének a mértéke).

Az eloszlásfüggvény alakja

Az eloszlásokat az **eloszlásfüggvény alakja** szerint szokás **csoportosítani**. Azok az eloszlások kerültek azonos csoportba, amelyek sűrűségfüggvénye csak a **paraméterek (konstansok) értékében tér el egymástól**.

Ismert eloszlások például: Normál (Gauss) eloszlás, binomiális eloszlás, Poisson eloszlás, Cauchy eloszlás, stb.

A normál eloszlás

A normál eloszlás **sűrűségfüggvénye** egy szimmetrikus, harang alakú görbe:

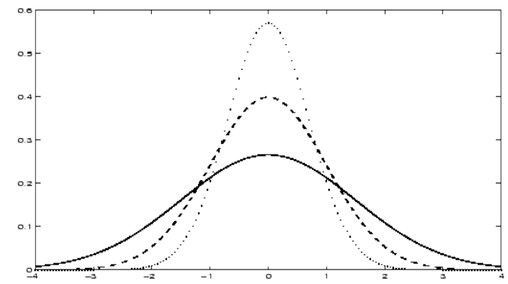
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

ahol az eloszlás paraméterei:

- **μ** az eloszlás várható értéke (egyúttal mediánja és módusza),

- **σ** a variancia négyzetgyöke.

A normál eloszlás sűrűségfüggvénye



μ=0, három különböző σ esetén

A centrális határeloszlás tétele

Bármely eloszlású sokaság esetén az *n* elemű minta **számtani középértékének eloszlása** *n* növelésével egy olyan **normál eloszláshoz tart**, melynek várható értéke megegyezik az eredeti eloszlás várható értékével.

Mivel a **mérések eredménye** az esetek döntő többségében ilyen **átlagolás eredménye** (a mutató tehetetlensége, a skálára való kerekítés, stb. miatt), így a mérési sorozatok a **legtöbb esetben normál eloszlást mutatnak**.

A normál eloszlás paramétereinek becslése

- **A várható érték (μ) becslése:**

  A **számtani átlag** torzítatlan, hatékony, konzisztens és elégséges becslése az alapsokaság várható értékének.

  A **medián** szintén konzisztens, és növekvő *n* esetén torzítatlan becslése a várható értéknek, de nagyobb a szórása mint a számtani átlagnak.

A normál eloszlás paramétereinek becslése

- **A variancia négyzetgyökének (σ) becslése:**

  A **korrigált tapasztalati** szórás torzítatlan és konzisztens becslést ad.

  A **tapasztalati szórás** és a **terjedelem** jelentősen torzított becslést eredményez.

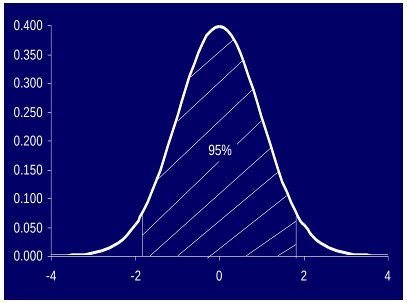
A normál eloszlás tulajdonságai

- A **μ** adja meg a haranggörbe x tengelyre vonatkoztatott **helyzetét**, a **σ** pedig az **alakját**. (Minél nagyobb a σ, annál laposabb a görbe.)
- Pontosan meghatározott, hogy hogyan alakul *k* függvényében annak **valószínűsége (α)**, hogy a mérési sorozat egy **értéke** a

$$\mu \pm k \cdot \sigma$$

tartományon **kívülre esik**.

A normál eloszlás alkalmazása



μ=0, σ=1 (standardizált normál eloszlás sűrűségfüggvénye)

$$\mu \Leftarrow \bar{x}$$

$$\sigma \Leftarrow s, \left(s_{\bar{x}} = \frac{s}{\sqrt{n}} \right)$$

$$\mu \pm k \cdot \sigma$$

Megbízhatósági szint	α %	k
95 %	5	1,96
99 %	1	2,58
99,9 %	0,1	3,29
99,99 %	0,01	4

A lineáris regresszió

A regresszió számítás során matematikai összefüggést keresünk egy függvény változói között.

A metrológiában leggyakrabban alkalmazott regresszió az egyetlen független változós **lineáris regresszió**:

$$y = m \cdot x + a$$

A lineáris regresszió

- *y*: **függő változó**
- *x*: **független változó**
- *m*, *a*: a függvény keresett **konstansai**

  A regresszió számítás során annak az **egyenesnek** a **meredekségét** (*m*) és **eltolását** (*a*) keressük, amely a legközelebb halad el az *x*-*y* **értékpárokkal** meghatározott **koordináta-pontokhoz**.

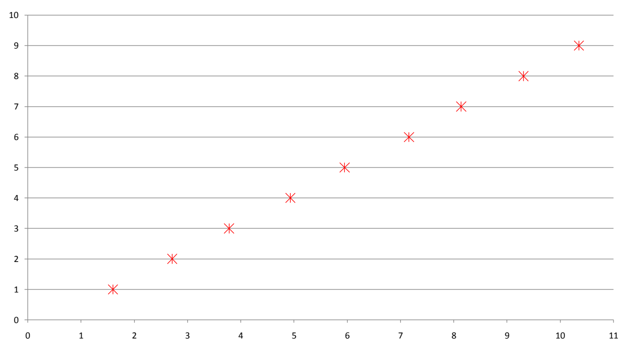
A lineáris regresszió

A mintapélda x-y értékpárjai:

y	1	2	3	4	5	6	7	8	9
x	1,6	2,71	3,78	4,93	5,95	7,16	8,14	9,31	10,35

m=? a=?

A lineáris regresszió



A lineáris regresszió

A lineáris regresszió (Gauss nevéhez fűződő) matematikai apparátusa a **legkisebb négyzetek elvén** alapszik.

Az összefüggés ***m*** és ***a*** paramétereit úgy határozzuk meg, hogy a mért és az összefüggésből számított ***y*** értékek **eltérésének négyzete minimális legyen**:

$$\sum_{i=1}^{n} \left[y_i - (a + m \cdot x_i) \right]^2 \rightarrow \min$$

A lineáris regresszió

A minimalizálási feladat eredménye:

$$m = \frac{\sum_{i=1}^{n} x_i \cdot y_i - \frac{\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left[\sum_{i=1}^{n} x_i\right]^2}{n}}$$

$$a = \frac{\sum_{i=1}^{n} y_i}{n} - m \cdot \frac{\sum_{i=1}^{n} x_i}{n}$$

ahol *n* az *x*-*y* adatpárok száma.

A lineáris regresszió

A megoldás gyakorlati menete:

1. Négyzetösszegek meghatározása:

$$SQ_x = \sum_{i=1}^{n} x_i^2 - \frac{\left[\sum_{i=1}^{n} x_i\right]^2}{n}$$

$$SP = \sum_{i=1}^{n} x_i \cdot y_i - \frac{\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n}$$

A lineáris regresszió

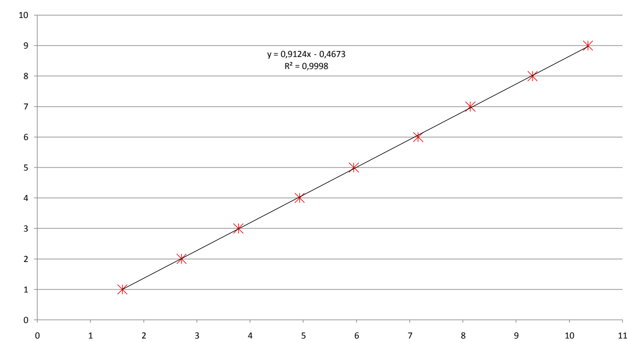
2. Átlagértékek meghatározása:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

3. Az egyenes egyenletének együtthatói:

$$m = \frac{SP}{SQ_x} \qquad a = \bar{y} - m \cdot \bar{x}$$

A lineáris regresszió



m=0,9124 *a*=-0,4673 R^2=0,9998

A korrelációs együttható

A mérési pontok és a számított egyenes közötti **illeszkedés szorosságát** a **korellációs együttható (R)** adja meg.

A korrelációs együttható értéke 0 és 1 között változik. A **0-hoz közeli** érték **gyenge kapcsolatra** utal, míg **teljes függvénykapcsolat** esetén **R=1**.

A korrelációs együttható

Számítása a gyakorlatban:

$$SQ_y = \sum_{i=1}^{n} y_i^2 - \frac{\left[\sum_{i=1}^{n} y_i\right]^2}{n}$$

$$R = \frac{SP}{\sqrt{SQ_x SQ_y}}$$