<u>Korreláció-, és regresszióanalízis</u>

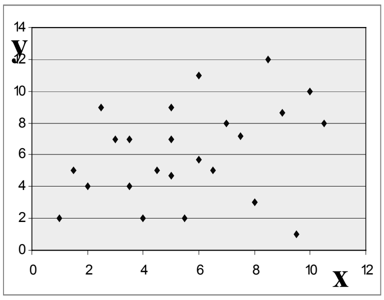
Két mérhető (kvantitatív) mennyiség közötti kapcsolat
(ha egyáltalán létezik) lehet → **determinisztikus** (pl. T=a·b
→ **sztochasztikus** ,azaz a véletlen is befolyásolja, X minden értékéhez Y különböző kimenetelei tartoznak (pl. geodéziai mérések ismétlései)

Az ilyen sztochasztikus kapcsolatok elemzésére szolgál a korrelációanalízis és a regresszióanalízis.

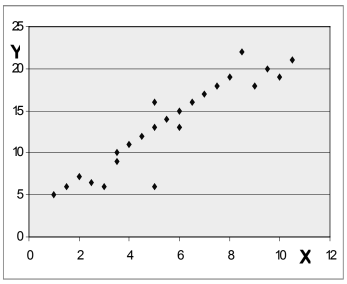
<u>Korreláció</u> - a lineáris, vagy arra visszavezethető kapcsolatok **<u>szorosságát</u>** méri.

Regresszió - az X és Y valószínűségi változók közötti kapcsolatot „legjobban” leíró függvényt keresi meg.

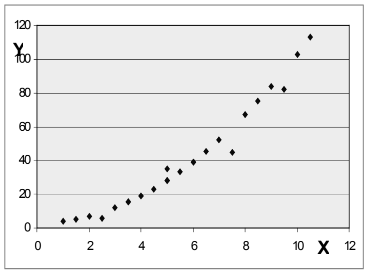
Ha kellő számú összetartozó mérést végeztünk X és Y-ra, akkor először célszerű egy megfelelő léptékű koordinátarendszerben ábrázolni őket és alaposan megnézni.



Elkent pontfelhő a kép⇒ valószínűleg X és Y **<u>függetlenek</u>**



Határozott tendencia mutatkozik X és Y között (itt jól sejthetően **<u>lineáris</u>**)



Határozott tendencia mutatkozik X és Y között (itt jól sejthetően **<u>nem lineáris</u>**)

I. Lineáris korreláció számítás

$$\rho = \frac{E(XY) - E(X)E(Y)}{D(X)D(Y)}$$ **az elméleti korrelációs együttható**

ez méri az X és Y változók közötti lin. kapcsolat szorosságát

$$\mathbf{r} = \frac{\sum_{i=1}^{n} x_i \cdot y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right] \cdot \left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right]}} = \frac{SP_{xy}}{\sqrt{SQ_x \cdot SQ_y}}$$

r a **tapasztalati korrelációs együttható**, ezzel becsüljük a ρ értékét a konkrét x_i és y_i mérési párok segítségével.

Tulajdonságai:

1. **r** értéke mindig -1 és +1 közötti,
2. teljes függetlenség esetén 0, ha pedig minden pont azonos egyenesen van +1 vagy - 1,
3. minél szorosabb a kapcsolat r absz. értéke annál közelebb van 1-hez,
4. előjele + akkor növekvő, ha - akkor csökkenő a kapcsolat tendenciája.

A lineáris korrelációs együttható próbája

Ha r nem 0, akkor is lehet nagy valószínűséggel inkább független a kapcsolat és az a határ, hogy miként döntünk n-től és ε-tól is nagyban függ.

H_0: ρ=0 (azaz X és Y lineárisan függetlenek)

H_1: ρ≠0 (azaz X és Y lineárisan összefüggnek és van értelme a legjobban illeszkedo lin. fv-t megkeresni)

Próba statisztika:

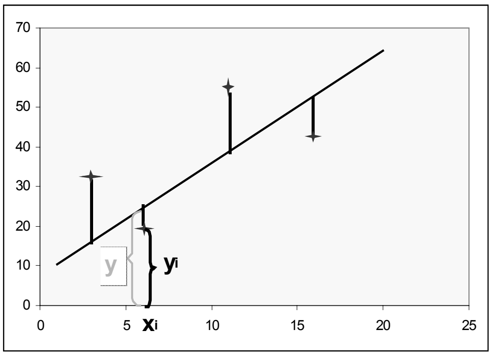
A X. táblázatból n és ε ismeretében kikeressük a kritikus értéket és **$K: \{|r| > r_\varepsilon\}$ Szab. fok: n-2 !** alapján döntünk.

Ha r ∈K, akkor H_0 -t elvetjük és megkeressük a legjobban illeszkedo függvényt, ha r∉K, akkor nincs igazolható kapcsolat.

II. Lineáris regresszió elve és kiszámítása

Keressük az Y=a+b·X lineáris függvény a és b együtthatóját úgy, hogy a **__mérési pontok és az illesztett egyenes függőleges távolságainal négyzetösszege a lehető legkisebb legyen__**.
Ezt hívjuk: **legkisebb négyzetek módszerének** (Least Squered Method)

Azaz:



$$\sum_{i=1}^{n}(y_i - y)^2 \rightarrow \textbf{minimumát keressük!}$$

$$\sum_{i=1}^{n}(y_i - a - bx_i)^2 \textbf{ -nek milyen a és b}$$

esetén van szélsőértéke?

$$f:(a;b) \mapsto \sum_{i=1}^{n}(y_i - a - bx_i)^2 \textbf{ kétváltozós függvénynek}$$

akkor lehet szélsőértéke, **ha** $f'_a = 0$ **és** $f'_b = 0.$

Ez akkor következik be, ha

$$b = \frac{\sum_{i=1}^{n} x_i \cdot y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]} = \frac{SP_{xy}}{SQ_x}$$

$$a = \frac{\sum_{i=1}^{n} y_i}{n} - b\frac{\sum_{i=1}^{n} x_i}{n} = \overline{y} - b \cdot \overline{x}$$

y=a+b·x az egyenes egyenlete,

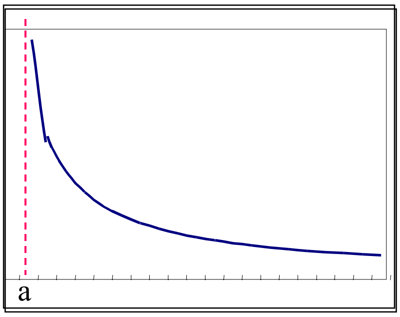
$$r = \frac{SP_{xy}}{\sqrt{SQ_x \cdot SQy}}$$

pedig méri az illeszkedés szorosságát.

Lineárisra visszavezethető nemlineáris regressziós függvények

1.Hiperbolikus függvények

a/



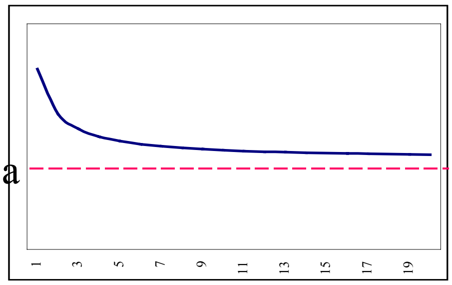
$$y = \frac{1}{a + bx}$$

Linearizálás:

$$\frac{1}{y} = a + bx$$

$$y^* = a + bx$$

b/



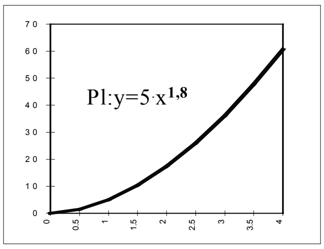
$$y = a + b\frac{1}{x}$$

$$x^*$$

Linearizálás:

$$y = a + b \cdot x^*$$

2. Hatvány függvények



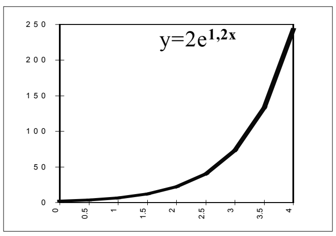
$$\mathbf{y = c \cdot x^b}$$

Linearizálása: $\ln y = \ln c + b \cdot \ln x$

$$y^* = a + b \cdot x^*$$

tehát az eredeti adatok helyett az ln -jeik és $c = e^a$ -ból adódik az együttható

3. Exponenciális függvény



$$\mathbf{y = c \cdot e^{bx}}$$

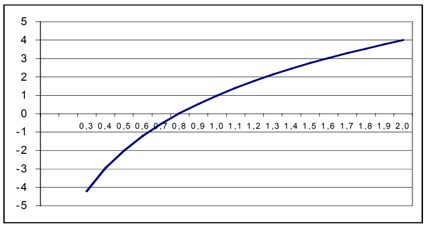
Linearizálva: $\ln y = \ln c + bx \cdot \ln e$

$$1$$

$$y^* = a + b \cdot x$$

4. Logaritmus függvény

$y = a + b \cdot \log x$, ahol log tetszőleges alapú lehet.
Linearizálva: $y = a + b \cdot x^*$
$x^* = \log x$ a transzformáció.



5. Telítődési függvény

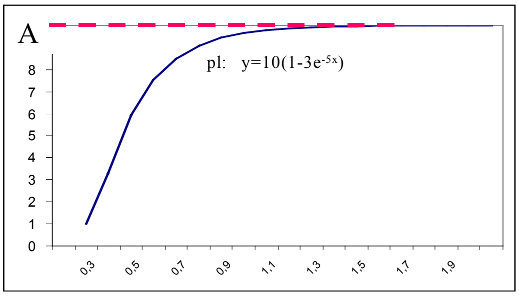
$y = A(1 - ce^{bx})$, $b<0$
és A-t az adatokból becsülni kell.

$A ce^{bx} = A - y$

$ce^{bx} = \frac{A-y}{A}$

$\ln c + \ln e^{bx} = \ln \frac{A-y}{A}$

$a + bx = y^*$



6. Logisztikus fv. (szigmoid görbe)

$$y = \frac{A}{1 + ce^{bx}}$$

ahol $b<0$ és A becsülendő az adatokból.

$$\frac{A-y}{y} = ce^{bx}$$

$$\ln \frac{A-y}{y} = \ln c + bx$$

$$y^* = a + bx$$