REGRESSZIÓANALÍZIS I.

Y függőváltozó

$X_1, X_2, \ldots X_p$ független változók

$Y \approx f(X_1, X_2, \ldots X_p)$ becslés $f \in F$

$$E(Y - f^*(X_1, X_2, \ldots X_p))^2 = \min_{f \in F} E(Y - f(X_1, X_2, \ldots X_p))^2$$

A legkisebb négyzetek módszere

$$h(a,b,c,\ldots) = \sum_{i=1}^{n} (Y_i - f(X_{1i}, X_{2i}, \ldots X_{pi}, a,b,c,\ldots))^2 \rightarrow \min_{a,b,c,\ldots}$$

REGRESSZIÓANALÍZIS II.

- Lineáris regresszió $f(X) = B_0 + B_1 X$
- Többváltozós lineáris regresszió

$$f(X_1, X_2, \ldots, X_p) = B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_p X_p$$

- Polinomiális regresszió

$$f(X_1, X_2, \ldots, X_p) = B_0 + B_1 X + B_2 X^2 + \ldots + B_p X^p$$

$$X_1 = X,\ X_2 = X^2,\ \ldots,\ X_p = X^p$$

- Kétparaméteres (lineárisra visszavezethető) regresszió

pl. $Y = f(X) = B_o \cdot e^{B_1 X} \Rightarrow \ln Y = B_1 X + \ln B_o$

REGRESSZIÓANALÍZIS III.

- Nemlineáris regresszió

$f(X) = B_1 + B_2 \exp(B_3 X)$	aszimptotikus I.
$f(X) = B_1 - B_2 \cdot (B_3)^X$	aszimptotikus II.
$f(X) = (B_1 + B_2 X)^{-1/B_3}$	sűrűség
$f(X) = B_1 \cdot (1 - B_3 \cdot \exp(B_2 X^2))$	Gauss
$f(X) = B_1 \cdot \exp(-B_2 \exp(-B_3 X^2))$	Gompertz
$f(X) = B_1 \cdot \exp(-B_2/(X + B_3))$	Johnson-Schumacher

REGRESSZIÓANALÍZIS IV.

• Nemlineáris regresszió

$f(X) = (B_1 + B_3 X)^{B_2}$	log-módosított
$f(X) = B_1 - \ln(1 + B_2 \exp(- B_3 X))$	log-logisztikus
$f(X) = B_1 + B_2 \exp(- B_3 X)$	Metcherlich
$f(X) = B_1 \cdot X / (X + B_2)$	Michaelis Menten
$f(X) = (B_1 B_2 + B_3 X^{B_4})/(B_2 + X^{B_4})$	Morgan-Merczer-Florin
$f(X) = B_1 /(1+B_2 \exp(- B_3 X + B_4 X^2 + B_5 X^3))$	Peal-Reed

REGRESSZIÓANALÍZIS V.

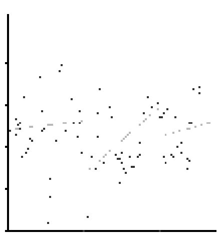
• Nemlineáris regresszió

$f(X) = (B_1 + B_2 X + B_3 X^2 + B_4 X^3)/ B_5 X^3$	köbök aránya
$f(X) = (B_1 + B_2 X + B_3 X^2)/ B_4 X^2$	négyzetek aránya
$f(X) = B_1/((1+B_3 \cdot \exp(B_2 X))^{(1/B_4)}$	Richards
$f(X) = B_1/((1+B_3 \cdot \exp(B_2 X))$	Verhulst
$f(X) = (B_1^{(1-B_4)} \cdot B_2 \exp(- B_3 X))^{1/(1-B_4)}$	Von Bertalanffy
$f(X) = B_1 - B_2 \exp(-B_3 X^{B_4})$	Weibull
$f(X) = 1/(B_1 + B_2 X + B_3 X^2)$	Yield sűrűség

REGRESSZIÓANALÍZIS VI.

• Szakaszonkénti lineáris regresszió

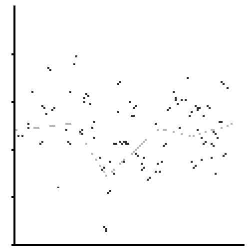
$$f(X) = \begin{cases} A_1 X + B_1 & X \in [a_0, a_1] \\ A_2 X + B_2 & X \in [a_1, a_2] \\ \vdots & \vdots \\ A_k X + B_k & X \in [a_{k-1}, a_k] \end{cases}$$



REGRESSZIÓANALÍZIS VII.

• Poligoniális regresszió

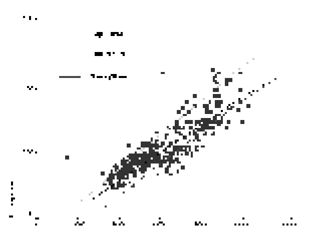
$$f(X)=\begin{cases} A_1X+B_1 & X\in[x_0,x_1] \\ A_2X+B_2 & X\in[x_1,x_2] \\ \vdots & \vdots \\ A_kX+B_k & X\in[x_{k-1},x_k] \end{cases} \qquad A_ix_i+B_i=A_{i+1}x_i+B_{i+1}$$



REGRESSZIÓANALÍZIS VIII.

• Többváltozós lineárisis regresszió kategória-változóval

$$f(X)=\begin{cases} B_0+B_1X & \text{ha } K=1 \\ (B_0+B_2)+(B_1+B_{c+1})X & \text{ha } K=2 \\ (B_0+B_3)+(B_1+B_{c+2})X & \text{ha } K=3 \\ \vdots & \vdots \\ (B_0+B_c)+(B_1+B_{2c-1})X & \text{ha } K=c \end{cases}$$



REGRESSZIÓANALÍZIS IX.

• Logisztikus regresszió

Y dichotóm Y= { 1, ha az A esemény bekövetkezik; 0, ha az A esemény nem következik be

A

• A választó fog szavazni
• A páciensnek szívinfarktusa lesz
• Az üzletet meg fogják kötni

X_1 , X_2 ,...,X_p **ordinális szintű független változók**

• eddig hányszor ment el, kor, iskola, jövedelem
• napi cigi, napi pohár, kor, stressz
• ár, mennyiség, piaci forgalom, raktárkészlet

REGRESSZIÓANALÍZIS X.

• Logisztikus regresszió

$$P(Y=1) = P(A) » \frac{1}{1 - e^{-Z}}$$

$$Z = B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_p X_p$$

$$ODDS = \frac{P(A)}{1 - P(A)} » e^{Z} \quad \Rightarrow$$

$$\log(ODDS) = Z = B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_p X_p$$

REGRESSZIÓANALÍZIS XI.

• Logisztikus regresszió

A legnagyobb valószínűség elve

$$L(e_1, e_2, \ldots, e_n) = P(Y_1 = e_1, Y_2 = e_2, \ldots, Y_n = e_n) =$$

$$= P(Y_1 = e_1)\, P(Y_2 = e_2)\ L\ P(Y_n = e_n) »$$

$$» \frac{1}{1 - e^{-Z_1}} \cdot \frac{1}{1 - e^{-Z_2}} \cdot L \cdot \frac{1}{1 - e^{-Z_n}}$$

$$\ln L(e_1, e_2, \ldots, e_n) = S \ln\left(\frac{1}{1 - \exp(B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_p X_p)}\right)$$

REGRESSZIÓANALÍZIS XII.

• Lineáris regresszió

A lineáris kapcsolat kitüntetett:

(1) a legegyszerűbb és leggyakoribb.

(2) két dimenziós normális eloszlás esetén a kapcsolat nem is lehet más (vagy lineáris vagy egyáltalán nincs).

$$B_0 = \bar{Y} - B_1 \bar{X}, \quad B_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

REGRESSZIÓANALÍZIS

XIII.

- Lineáris regresszió

A teljes négyzetösszeg

[illegible]

A maradékösszeg

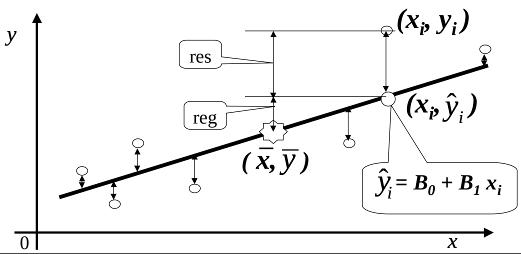
[illegible]

A regressziós összeg

[illegible]

A lineáris regresszió

$\boldsymbol{Q = Q_{res} + Q_{reg}}$



A lineáris regresszió

A teljes négyzetösszeg felbontása:

$Q = Q_{res} + Q_{reg}$

f_{reg} szabadsági foka *n-2*, mert *n* tagú az összeg, de ezek között két összefüggés van.

f_{res} szabadsági foka mindössze 1, mert az átlag konstans

Ha nincs lineáris regresszió, a varianciák hányadosa (*1*, *n-2*) szabadsági fokú *F* eloszlást követ.

$$F = \frac{s_{reg}^2}{s_{res}^2} = \frac{\frac{Q_{reg}}{f_{reg}}}{\frac{Q_{res}}{f_{res}}} = \frac{Q_{reg}(n-2)}{Q_{res}}$$

A lineáris regresszió

A ***legkisebb négyzetek módszere*** alapelve: