

Kvantitatív kutatási módszertanok 4

Ruzsa Zoltán
ruzsa.zoltan@emk.semmelweis.hu

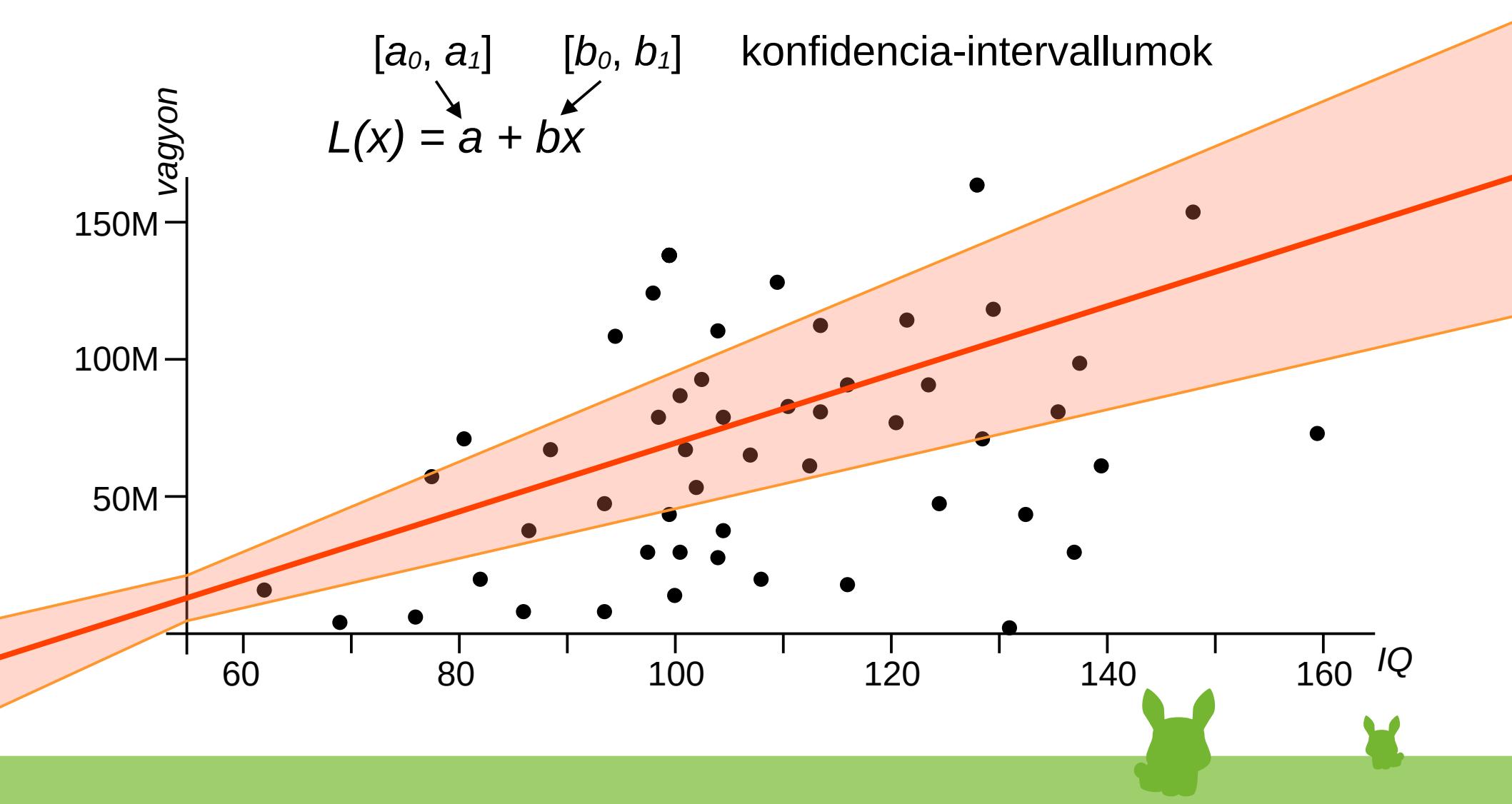
<https://ruzsaz.github.io/stat4.pdf>



Tematika

- Leíró statisztika (átlag, medián, módusz, szórás...)
- Paraméterbecslés (mennyi?)
- Hipotézisvizsgálat (igaz vagy nem?)
- Regressziószámítás, modellezés (hogyan befolyásol?)





HF volt: várható élettartam elemzése

<https://ruzsaz.github.io/elettartam.csv>

- 1) Készítsünk lineáris modellt, rajzot: élettartam ~ gdp

Make a linear regression to explain life_expectancy with gdp from the loaded data frame.

Show the confidence intervals.

Draw the data points.

Draw the 95% confidence band.



HF volt: várható élettartam elemzése



```
Call:
lm(formula = life_expectancy ~ gdp, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-19.562 -4.541  1.462  5.222 15.649 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 6.746e+01 3.027e-01 222.88 <2e-16 ***
gdp         2.594e-04 1.213e-05   21.39 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.909 on 764 degrees of freedom
Multiple R-squared:  0.3746,    Adjusted R-squared:  0.3737 
F-statistic: 457.5 on 1 and 764 DF,  p-value: < 2.2e-16
                2.5 %      97.5 % 
(Intercept) 6.686499e+01 6.805334e+01 
gdp         2.356418e-04 2.832639e-04
```

HF volt: várható élettartam elemzése

```
Call:  
lm(formula = life_expectancy ~ gdp, data = df)  
  
Residuals:  
    Min      10     Median      30      Max  
-19.562   -4.541    1.462    5.222   15.649  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 6.746e+01 3.027e-01 222.88 <2e-16 ***  
gdp         2.594e-04 1.213e-05   21.39 <2e-16 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 6.909 on 764 degrees of freedom  
Multiple R-squared:  0.3746,  Adjusted R-squared:  0.3737  
F-statistic: 457.5 on 1 and 764 DF,  p-value: < 2.2e-16  
              2.5 %    97.5 %  
(Intercept) 6.686499e+01 6.805334e+01  
gdp         2.356418e-04 2.832639e-04
```

95%-os konfidencia intervallum

Élettartam = $67.4 + 0.00025 * \text{gdp}$

T-próba, H_0 : együttható = 0

szignifikancia

illeszkedés várható hibája

modell fittsége: az élettartam szórásának ennyi részét magyarázza a regresszió (0-1)

F-próba, H_0 : minden együttható* = 0
(*: a konstans tag nincs beleértve)

Lineáris? regresszió

$$L(x) = a + bx$$

$$S(x) = a + bx + cx^2$$

$$Q(x) = a + bx + cx^2 + dx^3$$

$$Ln(x) = a + b\ln(x)$$

...



HF volt: várható élettartam elemzése

<https://ruzsaz.github.io/elettartam.csv>

- 1) Készítsünk lineáris modellt, rajzot: élettartam ~ gdp
- 2) Négyzetes modell? Más függvény?
- 3) Nézzük meg a többi lehetséges magyarázó változót is.
- 4) Mely magyarázó változókkal lesz a modell a legjobb?
(R: step eljárás automatikusan megkeresi)



1. feladat: egészséges életmód elemzése

<https://ruzsaz.github.io/egeszseg.csv>

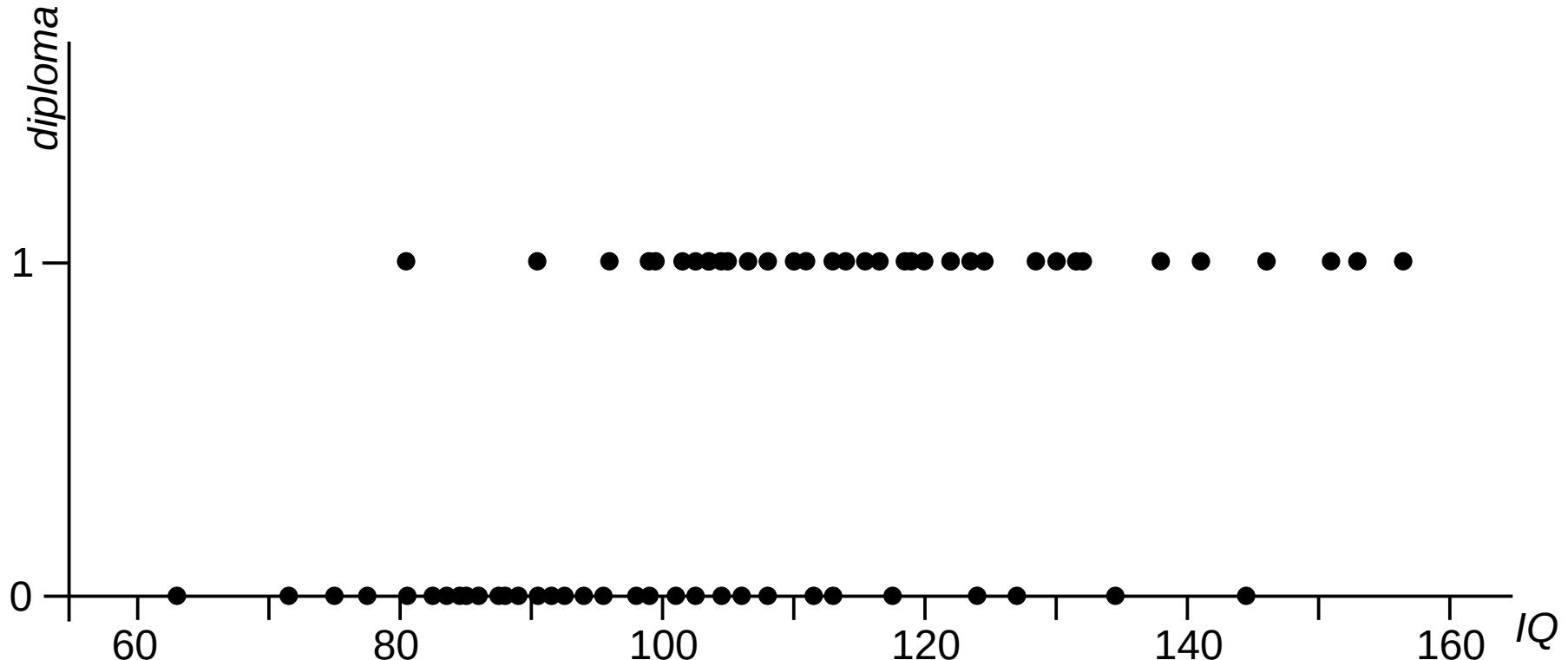
Célváltozó: health (0-100)

Magyarázó változók:

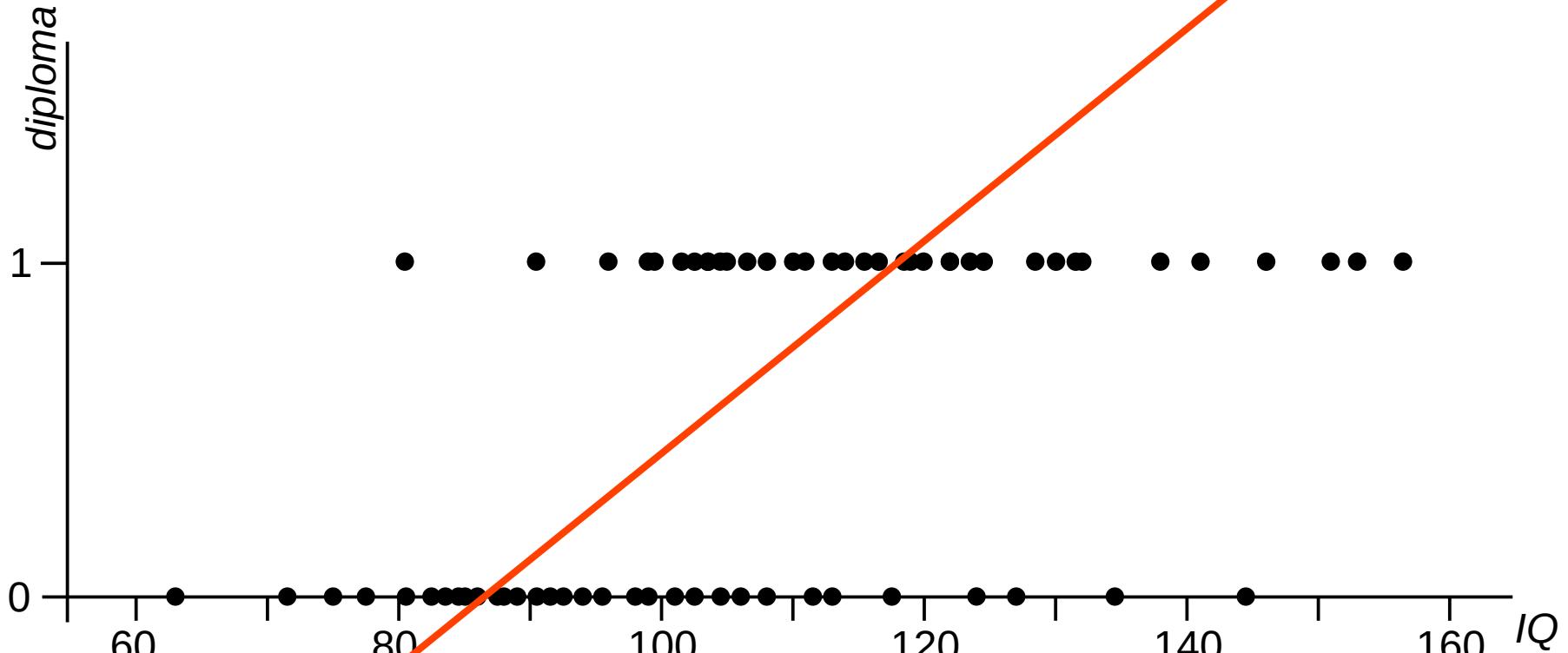
- age, bmi, exercise (0-7), diet (0-100), sleep (óra)
- smoke, sex: (kategória változók: 0, 1)



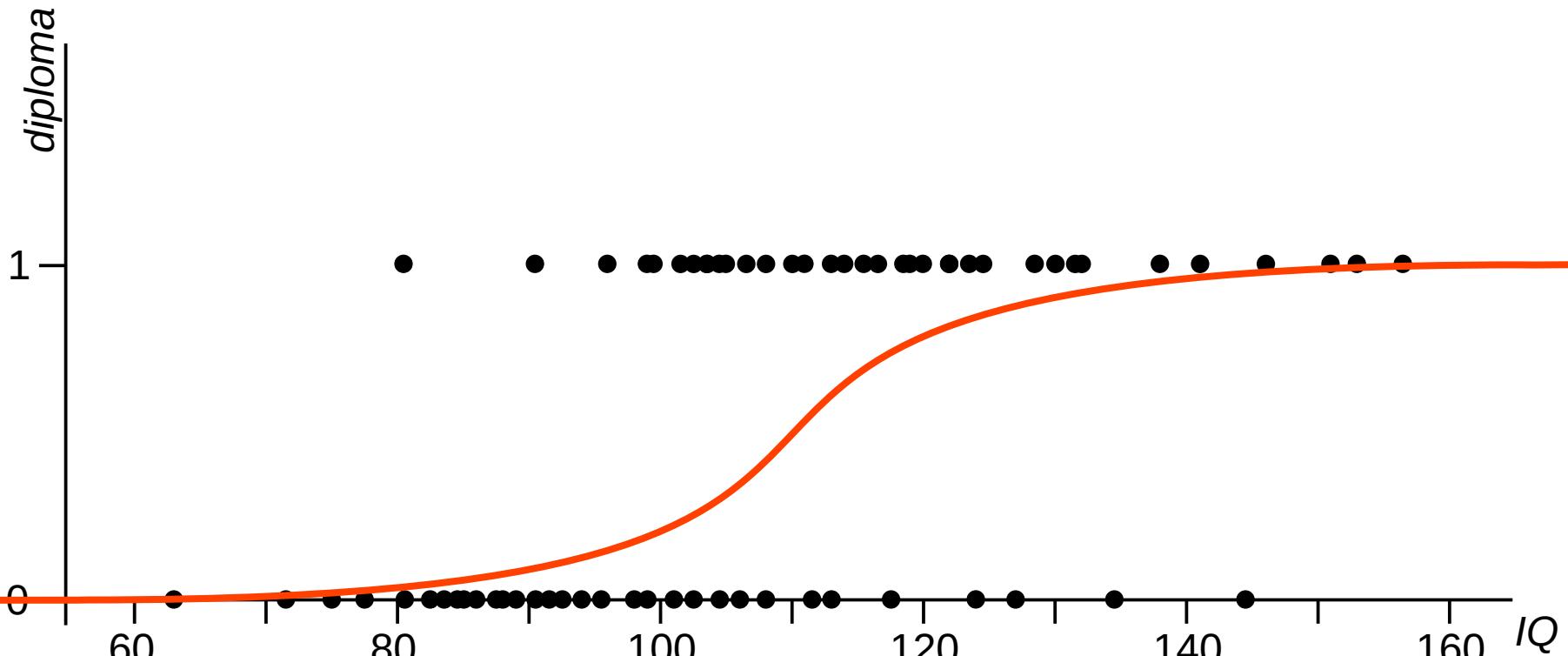
Logisztikus regresszió

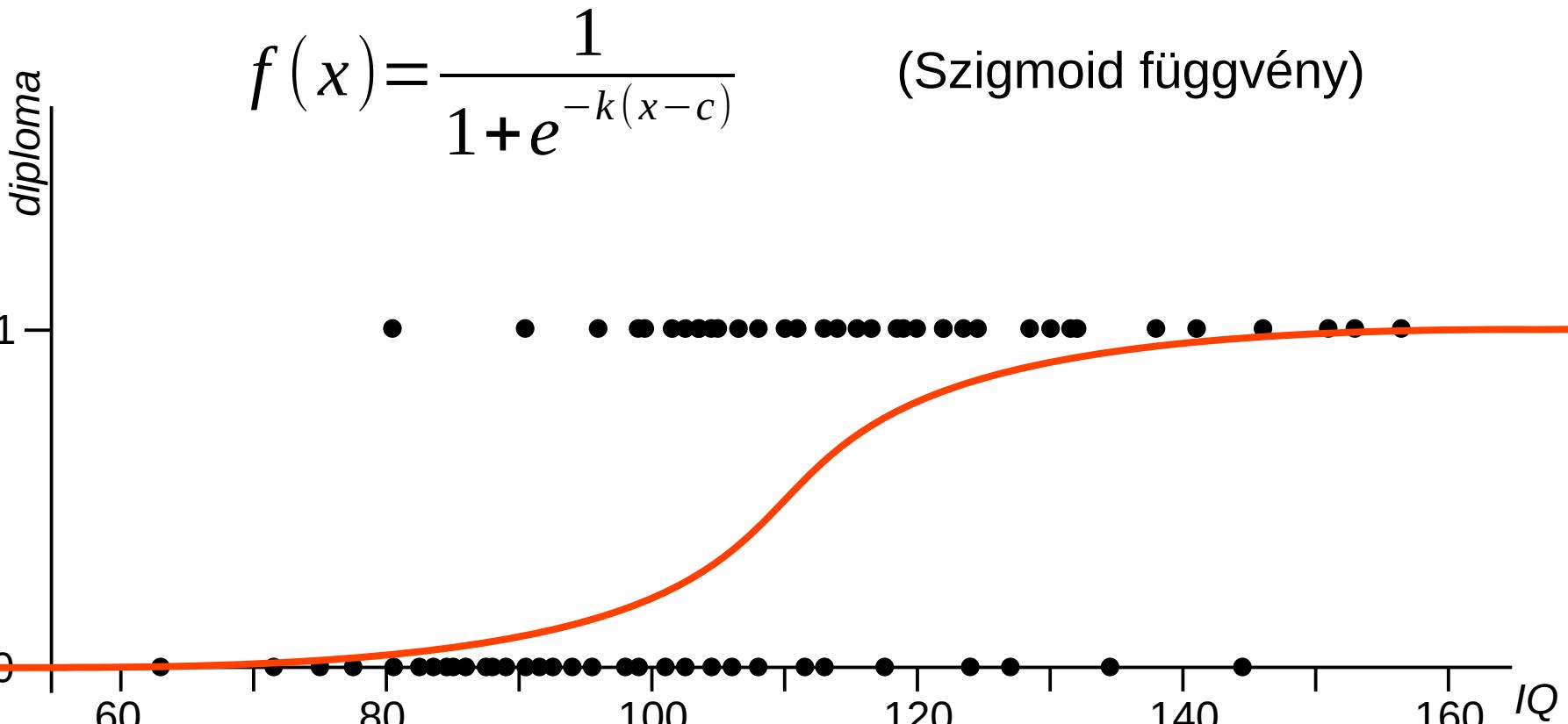


Logisztikus regresszió



Logisztikus regresszió





diploma

$$f(x) = \frac{1}{1 + e^{-k(x - c)}}$$

1

0

60

80

100

120

140

160

IQ



diploma

$$f(x) = \frac{1}{1 + e^{-k(x - c)}}$$

1

60

80

100

120

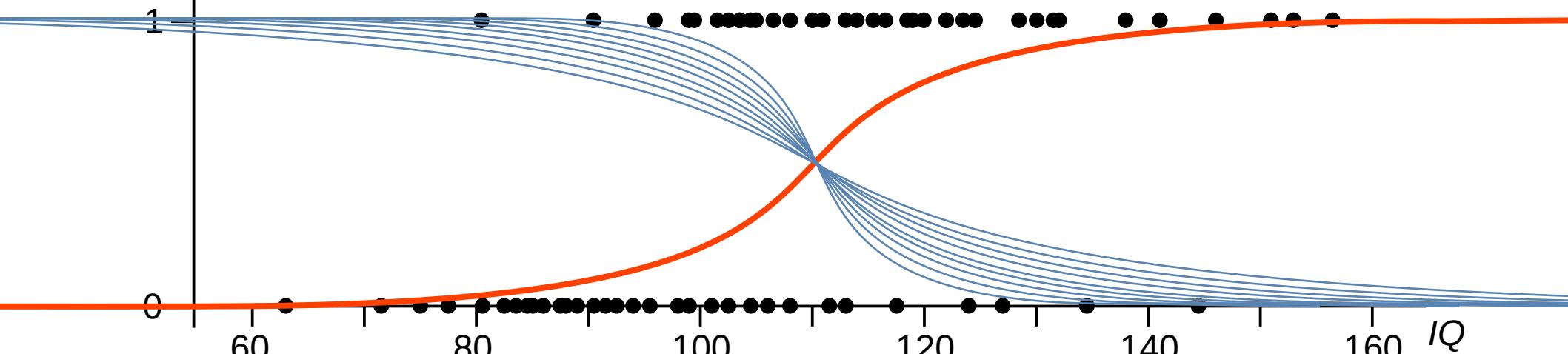
140

160

IQ



$$f(x) = \frac{1}{1 + e^{-k(x - c)}}$$

diploma

$$f(x) = \frac{1}{1 + e^{-k(x - c)}}$$

[k_0, k_1] [c_0, c_1]

konfidencia-intervallumok

1

0

60

80

100

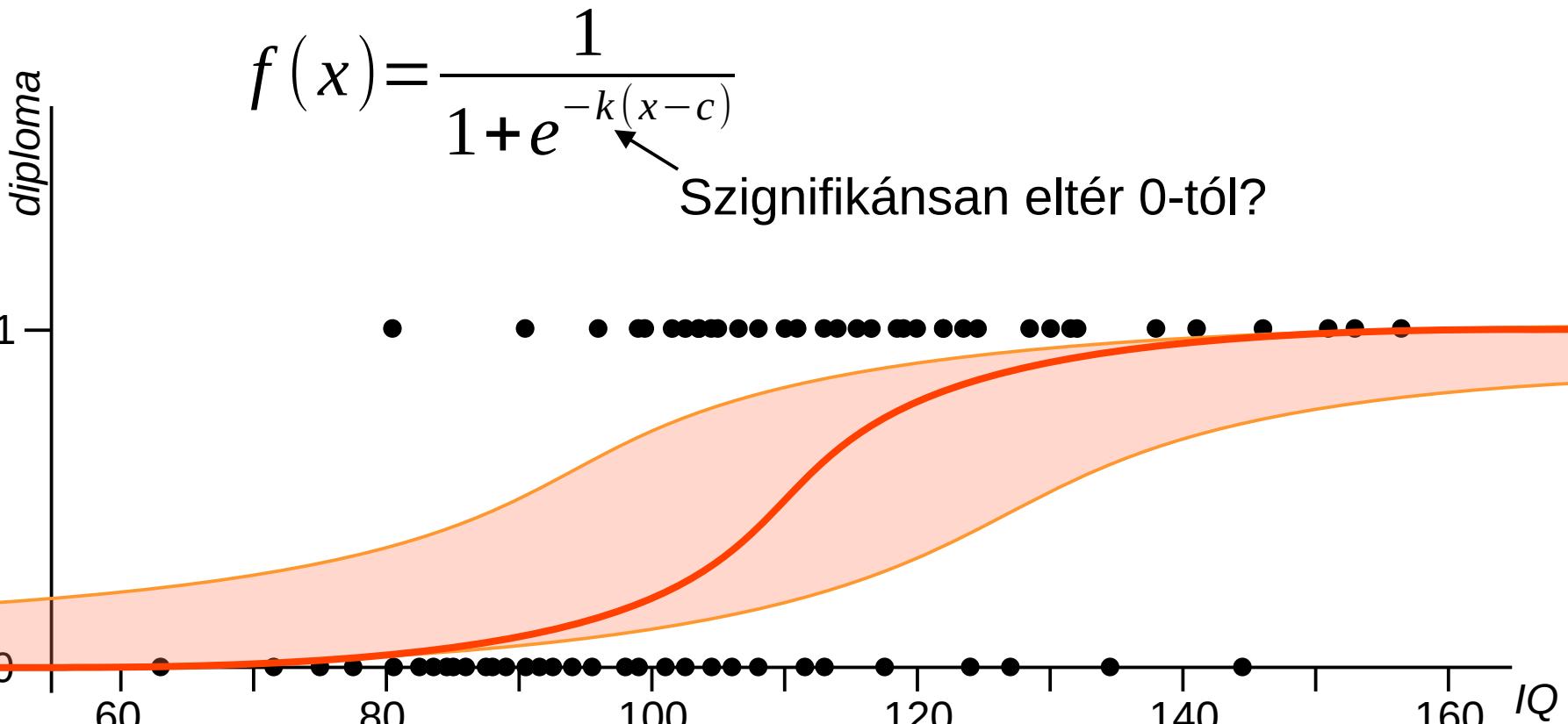
120

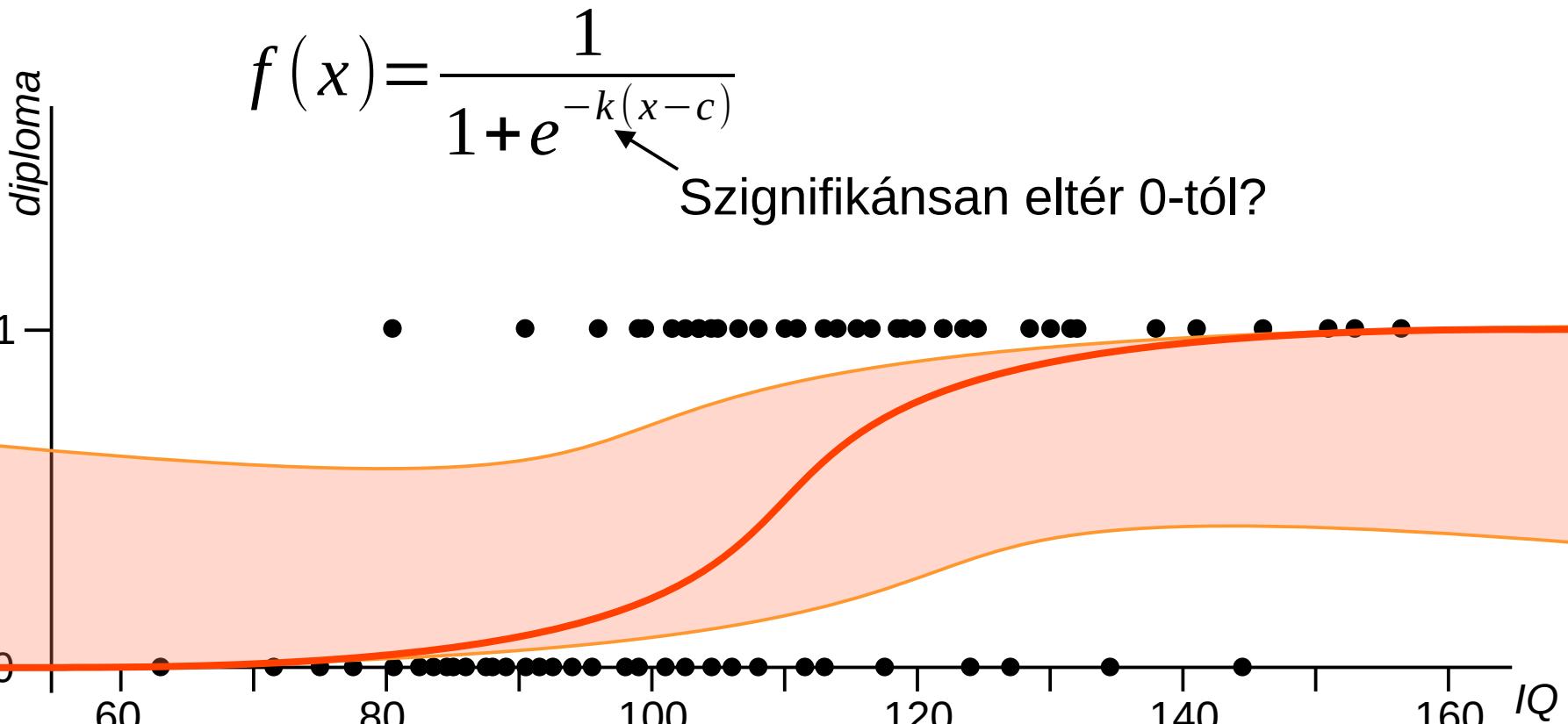
140

160

IQ







Mitől logisztikus?

$$p = \frac{1}{1 + e^{-k(x-c)}}$$



$$\frac{1}{p} = 1 + e^{-k(x-c)}$$

Mitől logisztikus?

$$p = \frac{1}{1 + e^{-k(x-c)}}$$

2

1

0



Mitől logisztikus?

$$\frac{1}{p} = 1 + e^{-k(x-c)}$$

$$\frac{1}{p} - 1 = e^{-k(x-c)}$$

$$p = \frac{1}{1 + e^{-k(x-c)}}$$

2

1

0



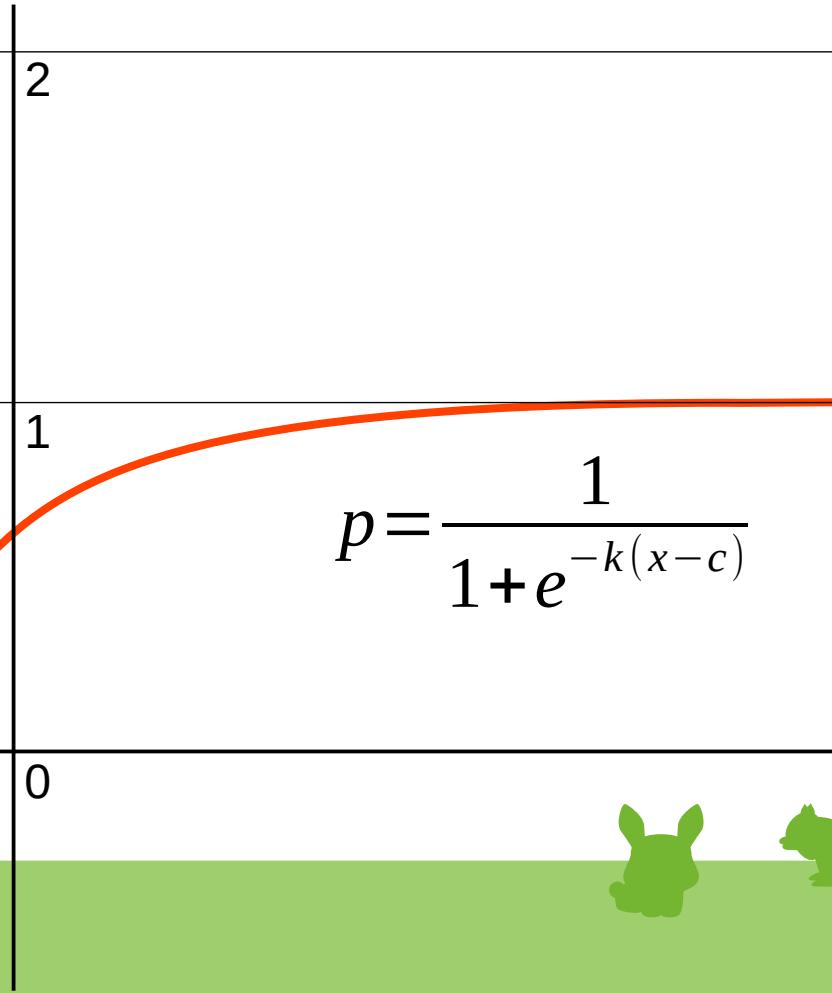
Mitől logisztikus?

$$\frac{1}{p} = 1 + e^{-k(x-c)}$$

$$\frac{1}{p} - 1 = e^{-k(x-c)}$$

$$\frac{1-p}{p} = e^{-k(x-c)}$$

$$p = \frac{1}{1 + e^{-k(x-c)}}$$



Mitől logisztikus?

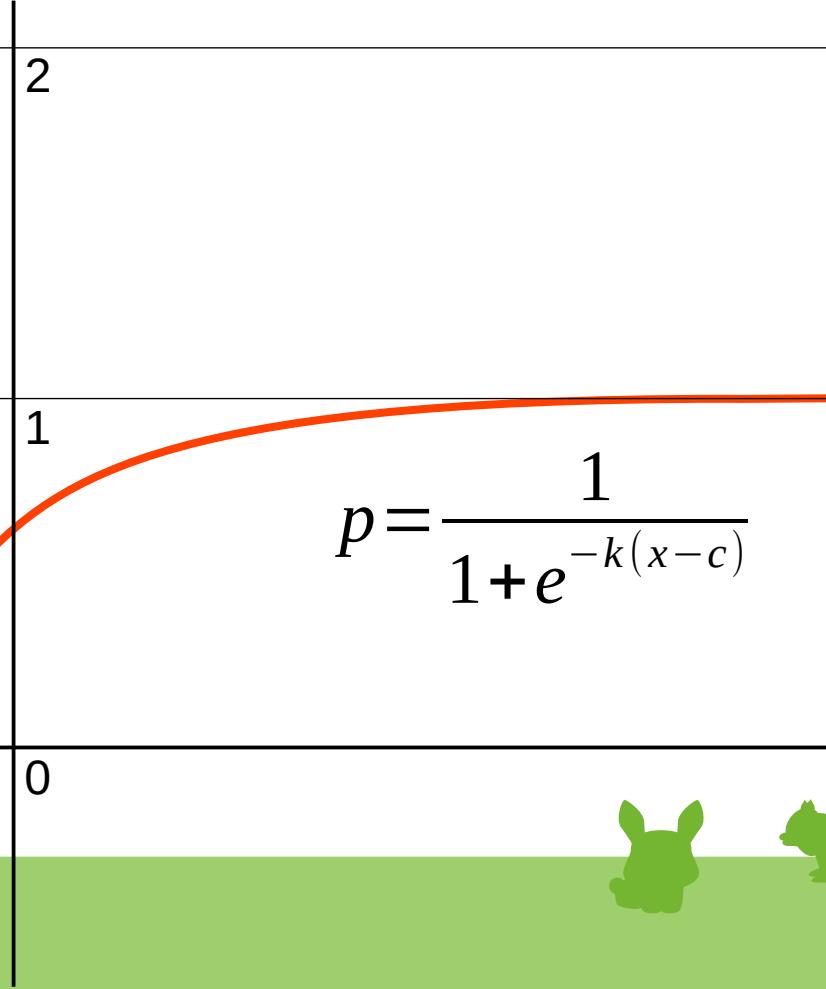
$$\frac{1}{p} = 1 + e^{-k(x-c)}$$

$$\frac{1}{p} - 1 = e^{-k(x-c)}$$

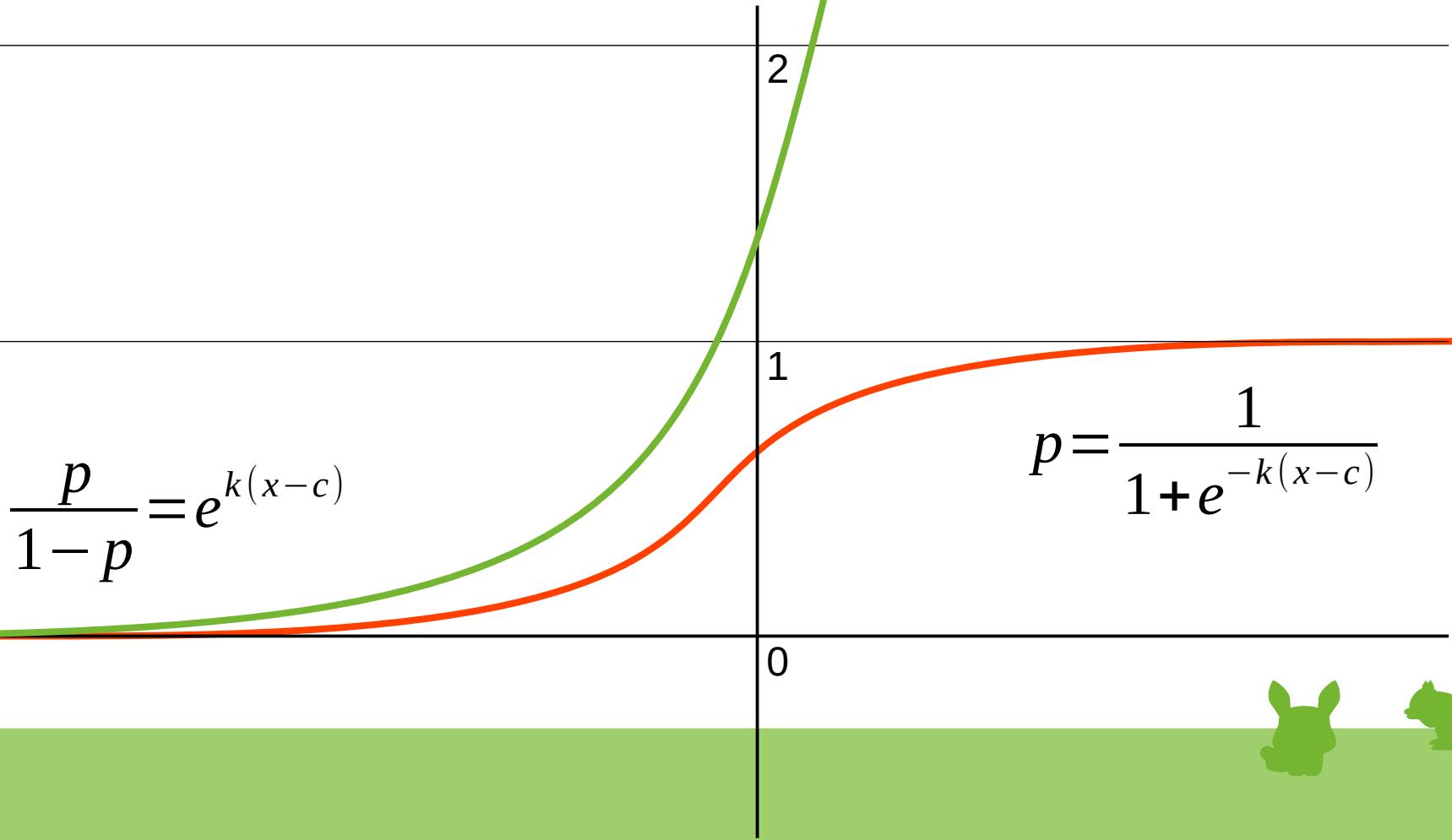
$$\frac{1-p}{p} = e^{-k(x-c)}$$

$$\frac{p}{1-p} = e^{k(x-c)}$$

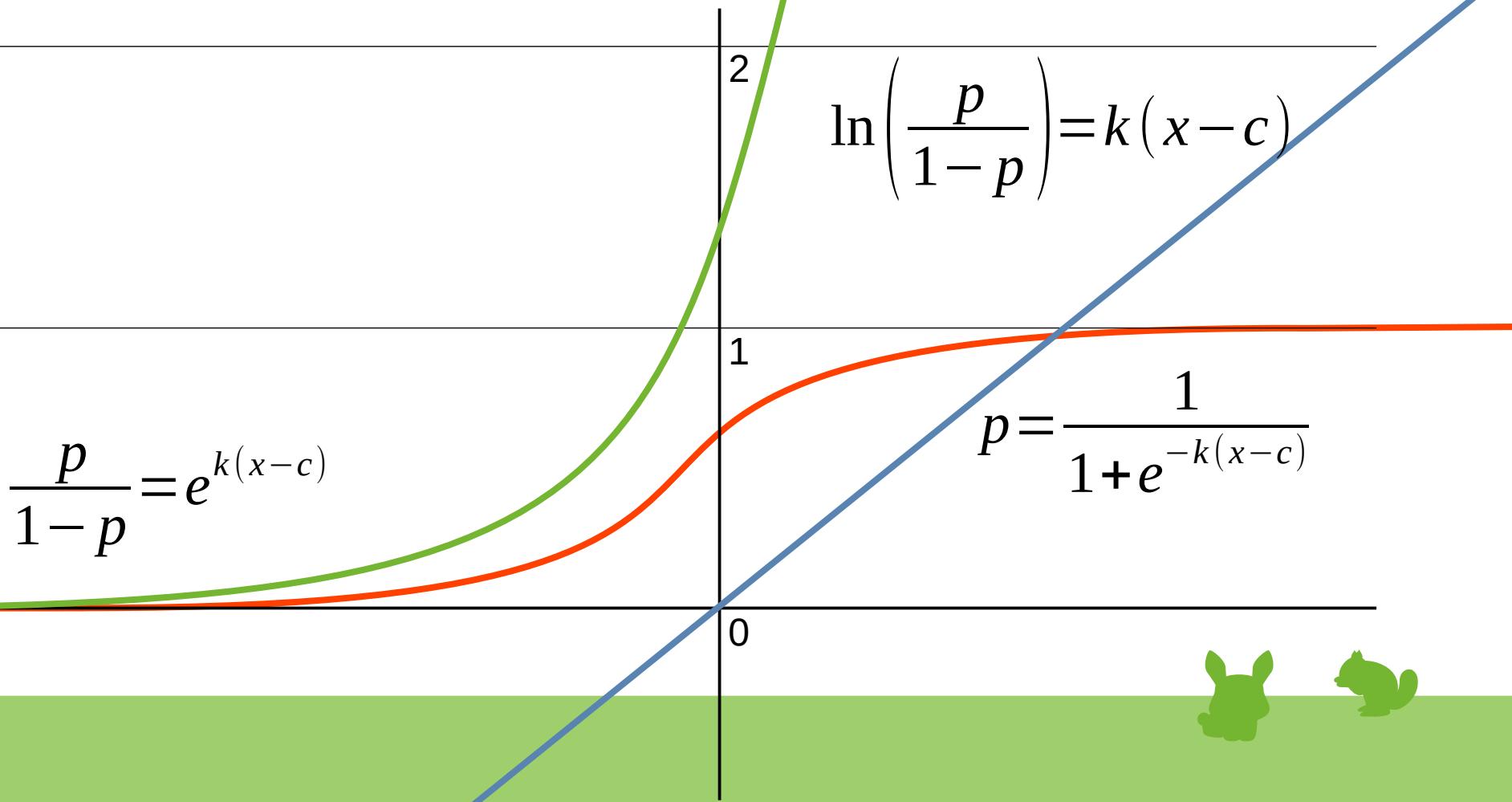
$$p = \frac{1}{1 + e^{-k(x-c)}}$$



Mitől logisztikus?



Mitől logisztikus?



Mitől logisztikus?

Logisztikus regresszió p -re

II

Lineáris regresszió $\ln\left(\frac{p}{1-p}\right)$ -re

$$\frac{p}{1-p} = e^{k(x-c)}$$

$$\ln\left(\frac{p}{1-p}\right) = k(x-c)$$

$$p = \frac{1}{1+e^{-k(x-c)}}$$

0

1

2



Logit-függvény

Logisztikus regresszió p -re

II

Lineáris regresszió $\ln\left(\frac{p}{1-p}\right)$ -re

$$\frac{p}{1-p} = e^{k(x-c)}$$

$$\ln\left(\frac{p}{1-p}\right) = k(x-c)$$

$$p = \frac{1}{1+e^{-k(x-c)}}$$

0

1

2



Odds

Logisztikus regresszió p -re

II

Lineáris regresszió $\ln\left(\frac{p}{1-p}\right)$ -re

$$\frac{p}{1-p} = e^{k(x-c)}$$

$$\ln\left(\frac{p}{1-p}\right) = k(x-c)$$

$$p = \frac{1}{1+e^{-k(x-c)}}$$

2

1

0



Valószínűség?

- p
- Kedvező / összes
- $p = 1/5 = 0.2$
 $p = 3/5 = 0.6$
- $[0, 1]$

Odds?

- $p/(1-p)$
- Kedvező : kedvezőtlen
- „1 a 4-hez” = $1 : 4 = 0.25$
„3 a 2-höz” = $3 : 2 = 1.5$
(3x nyer valaki, 2x veszít)
- $[0, \infty]$





Háromszorosára növeli az elhízás kockázatát a házasság a férfiaknál, de a feleségükknél nem

Dóka Boglárka | 2025.03.13. 13:40



Egy friss lengyel kutatás szerint a házasság jelentősen növeli az elhízás esélyét a férfiaknál, míg a nőknél nem figyelhető meg hasonló hatás. A kutatás rámutat arra is, hogy az életkor előrehaladtával mindenkor nem esetében nő az elhízás kockázata.

A kutatásban 2405, átlagosan 50 éves résztvevő egészségügyi adatait elemezték.

Az eredmények szerint a házas férfiak 3,2-szer nagyobb eséllyel válnak elhízottá, mint a nőtlen férfiak.

Ezzel szemben a házasság nem növelte a nőknél az elhízás kockázatát.

A lengyel kutatás azt is kimutatta, hogy az életkor előrehaladtával mindenkor nemnél nő a túlsúly és az elhízás kockázata. A férfiaknál évente 3 százalékkal nő a túlsúly, míg az elhízás kockázata 4 százalékkal emelkedik.

A nőknél ezek az arányok még magasabbak: a túlsúly esélye évente 4 százalékkal, az elhízásé pedig 6 százalékkal növekszik.

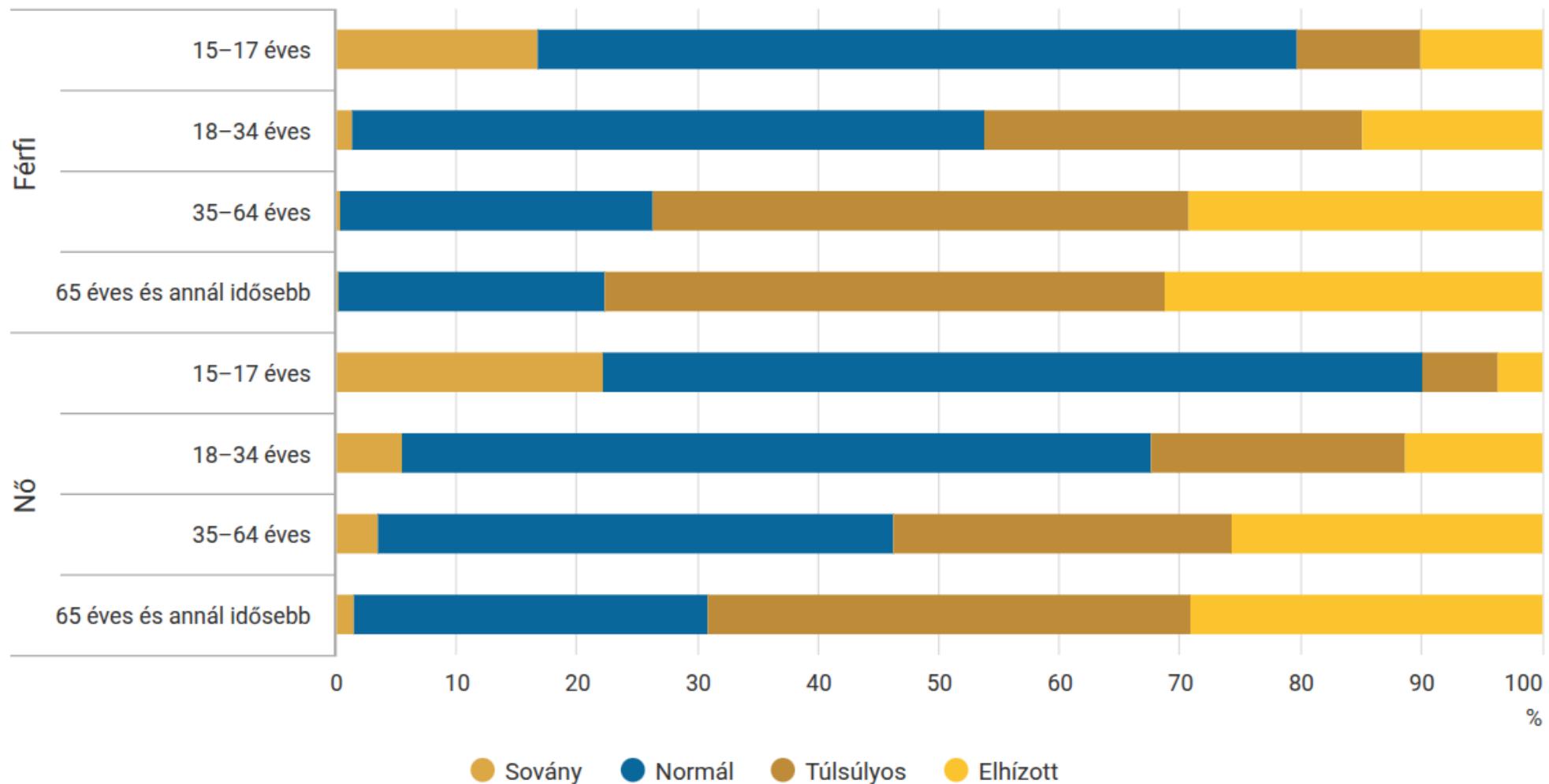
Érdekesség, hogy bizonyos tényezők kizárolag a nőknél növelték az elhízás kockázatát. A depresszió megduplázta, míg az egészségügyi ismeretek hiánya 43 százalékkal növelte az elhízás esélyét. Emellett a kisebb településeken élő nők körében is gyakoribb volt az elhízás, míg ezek a tényezők a férfiakat nem érintették hasonló mértékben.

<https://hirado.hu/extra/eltmod/cikk/2025/03/13/háromszorosára-növeli-az-elhízás-kockázatát-a-házasság-kockázatát-a-férfiaknál-de-a-feleségüknel-nem>

Hivatkozás:

[https://www.thelancet.com/journals/lancet/article/PII-S0140-6736\(25\)00397-6/fulltext](https://www.thelancet.com/journals/lancet/article/PII-S0140-6736(25)00397-6/fulltext)

Tápláltság nem és korcsoport szerint, 2019





Háromszorosára növeli az elhízás kockázatát a házasság a férfiaknál, de a feleségüknél nem

Dóka Boglárka | 2025.03.13. 13:40



Egy friss lengyel kutatás szerint a házasság jelentősen növeli az elhízás esélyét a férfiaknál, míg a nőknél nem figyelhető meg hasonló hatás. A kutatás rámutat arra is, hogy az életkor előrehaladtával mindenkorban nő az elhízás kockázata.

A kutatásban 2405, átlagosan 50 éves résztvevő egészségügyi adatait elemezték.

Az eredmények szerint a házas férfiak 3,2-szer nagyobb eséllyel válnak elhízottá, mint a nőtlen férfiak.

Ezzel szemben a házasság nem növelte a nőknél az elhízás kockázatát.

A lengyel kutatás azt is kimutatta, hogy az életkor előrehaladtával mindenkorban nő a túlsúly és az elhízás kockázata. A férfiaknál évente 3 százalékkal nő a túlsúly, míg az elhízás kockázata 4 százalékkal emelkedik.

A nőknél ezek az arányok még magasabbak: a túlsúly esélye évente 4 százalékkal, az elhízásé pedig 6 százalékkal növekszik.

Érdekkesség, hogy bizonyos tényezők kizárolag a nőknél növelték az elhízás kockázatát. A depresszió megduplázta, míg az egészségügyi ismeretek hiánya 43 százalékkal növelte az elhízás esélyét. Emellett a kisebb településeken élő nők körében is gyakoribb volt az elhízás, míg ezek a tényezők a férfiakat nem érintették hasonló mértékben.

Férfi:

17 éves: 10%

+50 év: $*(1.04)^{50} = 71\%$

+házas: $*3.2 = 227.2\%$

Nő:

17 éves: 4%

+50 év: $*(1.06)^{50} = 73.6\%$

+depi: $*2 = 147.2\%$

+buta: $*1.43 = 210.5\%$





ODDS & RETURNS

Fractional	Decimal	Payout*	Fractional	Decimal	Payout*
1/5	1.20	£2.40	7/2	4.50	£9.00
2/5	1.40	£2.80	4/1	5.00	£10.00
1/2	1.50	£3.00	9/2	5.50	£11.00
3/5	1.60	£3.20	5/1	6.00	£12.00
4/5	1.80	£3.60	6/1	7.00	£14.00
1/1	2.00	£4.00	7/1	8.00	£16.00
6/5	2.20	£4.40	8/1	9.00	£18.00
7/5	2.40	£4.80	9/1	10.00	£20.00
6/4	2.50	£5.00	10/1	11.00	£22.00
8/5	2.60	£5.20	15/1	16.00	£32.00
9/5	2.80	£5.60	20/1	22.00	£44.00
2/1	3.00	£6.00	25/1	26.00	£52.00
5/2	3.50	£7.00	30/1	31.00	£62.00
3/1	4.00	£8.00	50/1	51.00	£102.00

* Payouts based on a £2 stake

MOBILE // WINS SPORTS

Valószínűség = p odds = $p/(1-p)$



$$p = 0.67 \quad \text{odds} = 0.67/0.33 = 2:1$$

$$p = 0.5 \quad \text{odds} = 0.5/0.5 = 1:1$$

$$p = 0.33 \quad \text{odds} = 0.33/0.67 = 1:2$$

$$p = 0.25 \quad \text{odds} = 0.25/0.75 = 1:3$$



ODDS & RETURNS

Fractional	Decimal	Payout*	Fractional	Decimal	Payout*
1/5	1.20	£2.40	7/2	4.50	£9.00
2/5	1.40	£2.80	4/1	5.00	£10.00
1/2	1.50	£3.00	9/2	5.50	£11.00
3/5	1.60	£3.20	5/1	6.00	£12.00
4/5	1.80	£3.60	6/1	7.00	£14.00
1/1	2.00	£4.00	7/1	8.00	£16.00
6/5	2.20	£4.40	8/1	9.00	£18.00
7/5	2.40	£4.80	9/1	10.00	£20.00
6/4	2.50	£5.00	10/1	11.00	£22.00
8/5	2.60	£5.20	15/1	16.00	£32.00
9/5	2.80	£5.60	20/1	22.00	£44.00
2/1	3.00	£6.00	25/1	26.00	£52.00
5/2	3.50	£7.00	30/1	31.00	£62.00
3/1	4.00	£8.00	50/1	51.00	£102.00

* Payouts based on a £2 stake

MOBILE // WINS SPORTS

2. feladat: szívbetegség előrejelzése

<https://ruzsaz.github.io/sziv.csv>

- Célváltozó: chd – 10 koronaér-betegség 10 éven belül
(0: nincs, 1: van)
- Magyarázó változó: cholesterol – koleszterin szint

Végezzük el a logisztikus regressziót, rajzoljuk le az eredményt!



```
# prompt: use logistic regression to explain "chd" from "cholesterol"

# Fit the logistic regression model
model <- glm(chd ~ cholesterol, data = df, family = "binomial")

# Summarize the model
summary(model)

# Calculate odds ratios and confidence intervals
exp(cbind(OR = coef(model), confint(model)))
```

```
Call:
glm(formula = chd ~ cholesterol, family = "binomial", data = df)

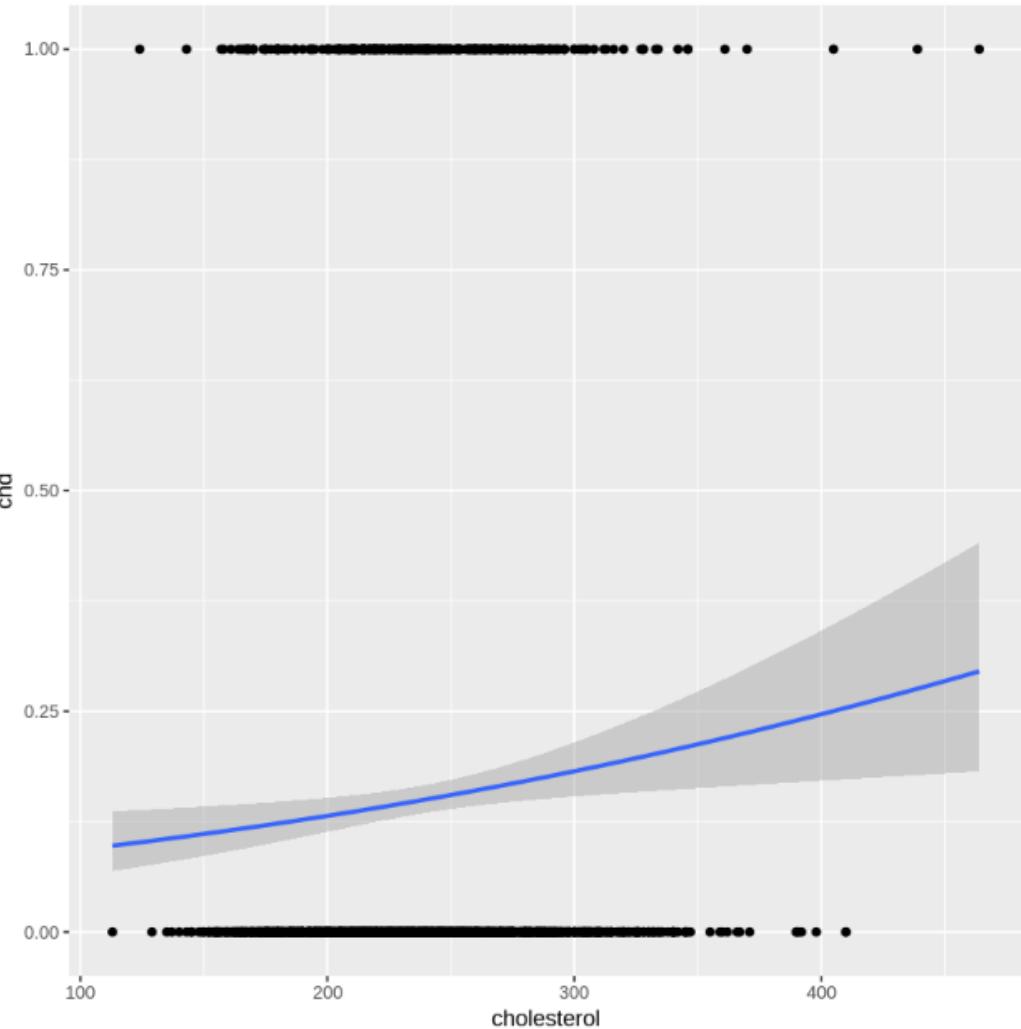
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.658310   0.348271 -7.633  2.3e-14 ***
cholesterol  0.003854   0.001420   2.713  0.00666 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1687.4 on 1999 degrees of freedom
Residual deviance: 1680.1 on 1998 degrees of freedom
AIC: 1684.1

Number of Fisher Scoring iterations: 4
Waiting for profiling to be done...
```

	A matrix: 2 × 3 of type dbl		
	OR	2.5 %	97.5 %
(Intercept)	0.07006656	0.03531783	0.1384529
cholesterol	1.00386139	1.00105543	1.0066495



```
# prompt: use logistic regression to explain "chd" from "cholesterol"

# Fit the logistic regression model
model <- glm(chd ~ cholesterol, data = df, family = "binomial")

# Summarize the model
summary(model)

# Calculate odds ratios and confidence intervals
exp(cbind(OR = coef(model), confint(model)))
```

```
Call:
glm(formula = chd ~ cholesterol, family = "binomial", data = df)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.658310  0.348271 -7.633 2.3e-14 ***
cholesterol   0.003854  0.001420   2.713 0.00666 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1687.4 on 1999 degrees of freedom
Residual deviance: 1680.1 on 1998 degrees of freedom
AIC: 1684.1

Number of Fisher Scoring iterations: 4
Waiting for profiling to be done...
```

```
A matrix: 2 × 3 of type dbl
      OR    2.5 %   97.5 %
(Intercept) 0.07006656  0.03531783  0.1384529
cholesterol 1.00386139 1.00105543  1.0066495
```

1-el nő a cholesterol, 0.003854-el nő a log odds

Z-teszt a szignifikanciára

Model illeszkedése a null-modelhez képest

Akaike Information Criterion
(minél kisebb, annál több információt tartalmaz a model)

1-el nő a cholesterol, 1.0038 szorosára változik az odds

2. feladat: szívbetegség előrejelzése

<https://ruzsaz.github.io/sziv.csv>

- Célváltozó: chd – 10 koronaér-betegség (0: nincs, 1: van)
- Magyarázó változók
 - Kategória:
male, education, smoker, cigarettes, hypertension, diabetes
 - Számszerű:
age, cigarettes, cholesterol, bloodpressure, bmi, heartrate
- Végezzük el az elemzést!



Odds-arány számolása

460613 gyermek egészségügyi adatainak elemzésével azt szeretnénk megállapítani, hogy mennyivel növeli a gyermekkor (0-3 év) alsólégúti betegség a későbbi (7 éves kori) asztma kockázatát.

	Későbbi asztma	Nincs asztma
Alsólégúti betegség	6550	18803
Betegség nélkül	43803	391457

A kontingencia-táblázat elemzésével számoljuk ki az odds-arány és a kockázat növekedést!



	Eredmény Pozitív (Betegség)	Eredmény Negatív (Egészséges)	Összesen
Expozíció Van (A)	$\$a = 6550\$$	$\$b = 18803\$$	$\$a+b = 25353\$$
Expozíció Nincs (B)	$\$c = 43803\$$	$\$d = 391457\$$	$\$c+d = 435260\$$ $\$N =$

Gemini 2.5

1. Odds Arány (Odds Ratio, OR) és 95% CI 🎲

Odds Arány (OR)

$$OR = \frac{a \cdot d}{b \cdot c} = \frac{6550 \cdot 391457}{18803 \cdot 43803} \approx 0.3112$$

???

$$\frac{a \cdot d}{b \cdot c} = \frac{6550 \cdot 391457}{18803 \cdot 43803} = 0.3112$$

2. Kockázati Arány (Risk Ratio, RR) és 95% CI 📈

Kockázati Arány (RR)

- Kockázat A (Exponált): $P_A = \frac{6550}{25353} \approx 0.25835\$$
 - Kockázat B (Nem exponált): $P_B = \frac{43803}{435260} \approx 0.10064\$$
- $$RR = \frac{P_A}{P_B} = \frac{0.25835}{0.10064} \approx 2.5670\$$$

Összefoglalás és Értelmezési Megjegyzés ➤

Mutató	Érték	95% Konfidencia Intervallum	Kockázat/Odds Növekedés
Odds Arány (OR)	$0.3112\$$	$[0.3022, 0.3209]\$$	Az odds 68.9% -kal alacsonyabb
Kockázati Arány (RR)	$2.5670\$$	$[2.5241, 2.6098]\$$	A kockázat 156.7% -kal magasabb



	Eredmény Pozitív (Betegség)	Eredmény Negatív (Egészséges)	Összesen
Expozíció Van (A)	\$a = 6550\$	\$b = 18803\$	\$a+b = 25353\$
Expozíció Nincs (B)	\$c = 43803\$	\$d = 391457\$	\$c+d = 435260\$ \$\bar{N} =

Gemini 2.5

1. Odds Arány (Odds Ratio, OR) és 95% CI

Odds Arány (OR)

$$OR = \frac{a \cdot d}{b \cdot c} = \frac{6550 \cdot 391457}{18803 \cdot 43803} \approx 0.3112$$

2. Kockázati Arány (Risk Ratio, RR) és 95% CI

Kockázati Arány (RR)

- Kockázat A (Exponált): $P_A = \frac{6550}{25353} \approx 0.25835$
 - Kockázat B (Nem exponált): $P_B = \frac{43803}{435260} \approx 0.10064$
- $$RR = \frac{P_A}{P_B} = \frac{0.25835}{0.10064} \approx 2.5670$$

Összefoglalás és Értelmezési Megjegyzés

Mutató	Érték	95% Konfidencia Intervallum	Kockázat/Odds Növekedés
Odds Arány (OR)	\$0.3112\$	[\$0.3022, 0.3209]\$	Az odds \$68.9%\$-kal alacsonyabb
Kockázati Arány (RR)	\$2.5670\$	[\$2.5241, 2.6098]\$	A kockázat \$156.7%\$-kal magasabb

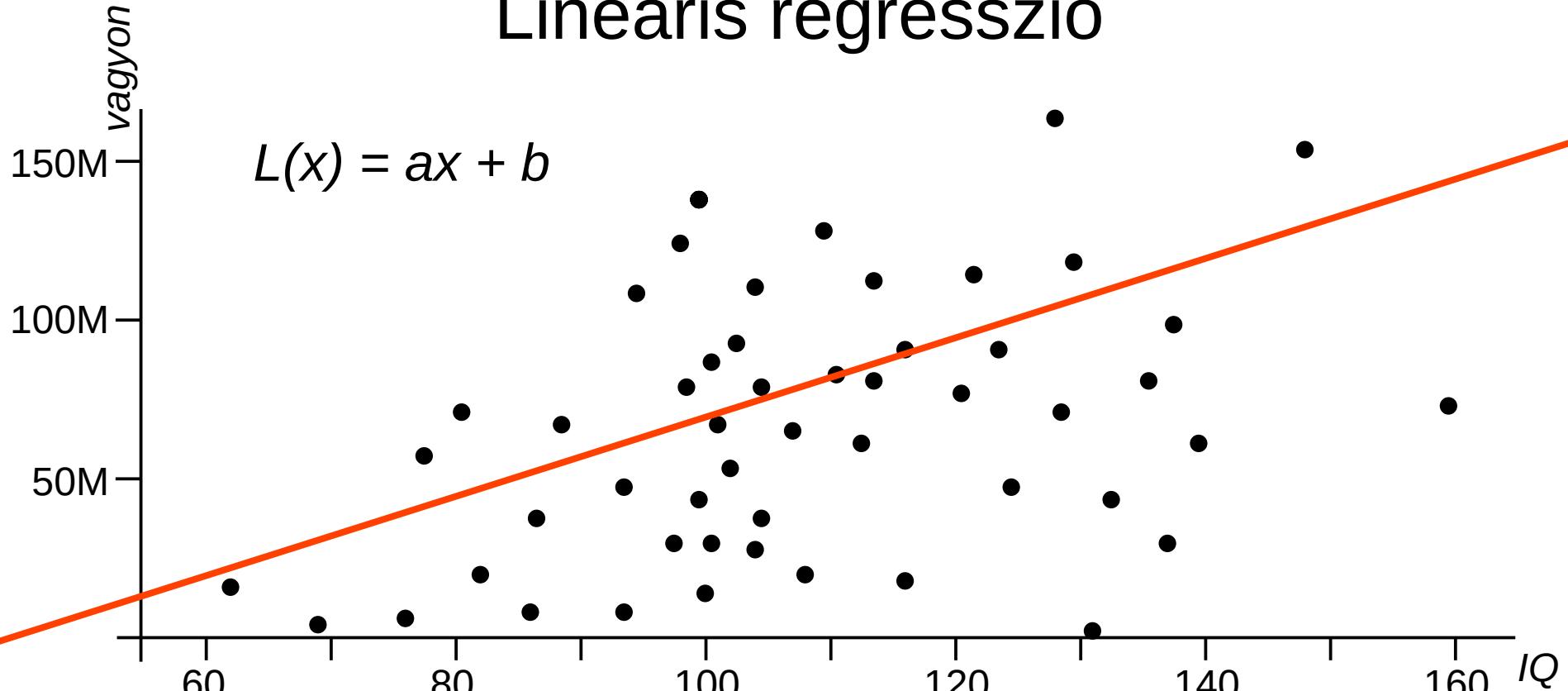


Jön az AI!

Mi szükség van regresszióra, amikor az AI eszközökkel sokkal pontosabb előrejelzést készíthetünk?



Lineáris regresszió



$$L(x) = ax + b$$

IQ

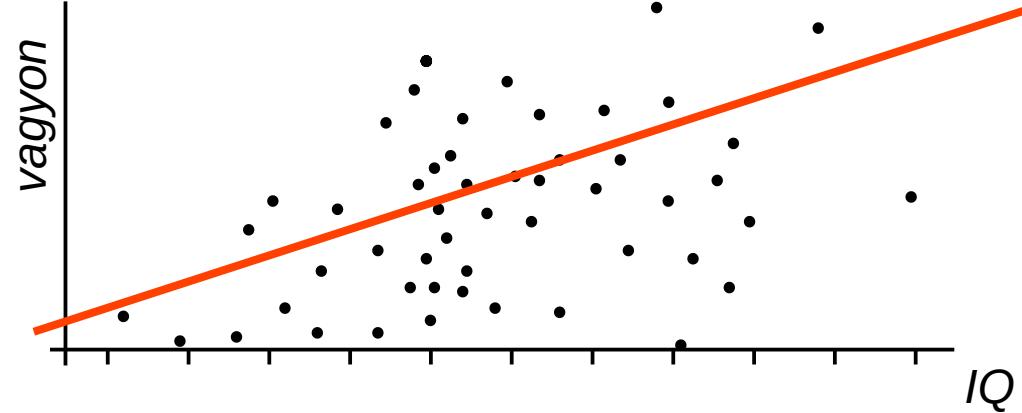


$$\begin{bmatrix} 107 \\ 98 \\ 157 \\ 80 \\ 121 \\ \dots \end{bmatrix}$$

[*a*]

$$+ \begin{bmatrix} b \\ b \\ b \\ b \\ b \\ \dots \end{bmatrix} \approx$$

$$\begin{bmatrix} 123M \\ 47M \\ 239M \\ 447M \\ 3M \\ \dots \end{bmatrix}$$



vagyon



$$L(x) = a + bx$$

IQ

$$\begin{bmatrix} a \end{bmatrix}$$

107

98

157

80

121

...

$$\begin{bmatrix} 107a \end{bmatrix}$$

$$\begin{bmatrix} 98a \end{bmatrix}$$

$$\begin{bmatrix} 157a \end{bmatrix}$$

$$\begin{bmatrix} 80a \end{bmatrix}$$

$$\begin{bmatrix} 121a \end{bmatrix}$$

...

+

$$\begin{bmatrix} b \\ b \\ b \\ b \\ b \end{bmatrix}$$

...

\approx

$$\begin{bmatrix} 123M \end{bmatrix}$$

$$\begin{bmatrix} 47M \end{bmatrix}$$

$$\begin{bmatrix} 239M \end{bmatrix}$$

$$\begin{bmatrix} 447M \end{bmatrix}$$

$$\begin{bmatrix} 3M \end{bmatrix}$$

...

vagyon

IQ



vagyon

Multilineáris regresszió

$$\begin{array}{c} \text{kor} \quad \text{IQ} \\ \downarrow \quad \downarrow \\ \left[\begin{array}{cc} 51 & 107 \\ 77 & 98 \\ 47 & 157 \\ 81 & 80 \\ 17 & 121 \\ \dots & \dots \end{array} \right] + \left[\begin{array}{c} a_0 \\ a_1 \\ b \\ b \\ b \\ b \\ \dots \end{array} \right] \approx \left[\begin{array}{c} 123M \\ 47M \\ 239M \\ 447M \\ 3M \\ \dots \end{array} \right] \end{array}$$

vagyon



Multilineáris regresszió

$$\begin{array}{l} \text{nem} \quad \text{kor} \quad \text{IQ} \quad \left[\begin{array}{c} a_0 \\ a_1 \\ a_2 \end{array} \right] \\ \downarrow \quad \downarrow \quad \downarrow \\ \left[\begin{array}{ccc} 1 & 51 & 107 \\ 1 & 77 & 98 \\ 2 & 47 & 157 \\ 1 & 81 & 80 \\ 2 & 17 & 121 \\ \dots & \dots & \dots \end{array} \right] + \left[\begin{array}{c} b \\ b \\ b \\ b \\ b \\ \dots \end{array} \right] \approx \left[\begin{array}{c} 123M \\ 47M \\ 239M \\ 447M \\ 3M \\ \dots \end{array} \right] \end{array}$$

vagyon



nem kor IQ

1	51	107
1	77	98
2	47	157
...		
1	81	80
2	17	121
...

$$\begin{bmatrix} \dots \\ a_{n-2} \\ a_{n-1} \\ a_n \end{bmatrix}$$

$$+ \begin{bmatrix} b \\ b \\ b \\ b \\ b \\ \dots \end{bmatrix}$$

≈

$$\begin{bmatrix} 123M \\ 47M \\ 239M \\ 447M \\ 3M \\ \dots \end{bmatrix}$$

vagyon



input mátrix

1	51	107
1	77	98
2	47	157
...		
1	81	80
2	17	121
...

$$\begin{bmatrix} \dots \\ a_{n-2} \\ a_{n-1} \\ a_n \end{bmatrix}$$

súlyok

$$+ \begin{bmatrix} b \\ b \\ b \\ b \\ b \\ \dots \end{bmatrix}$$

eltolás (bias)

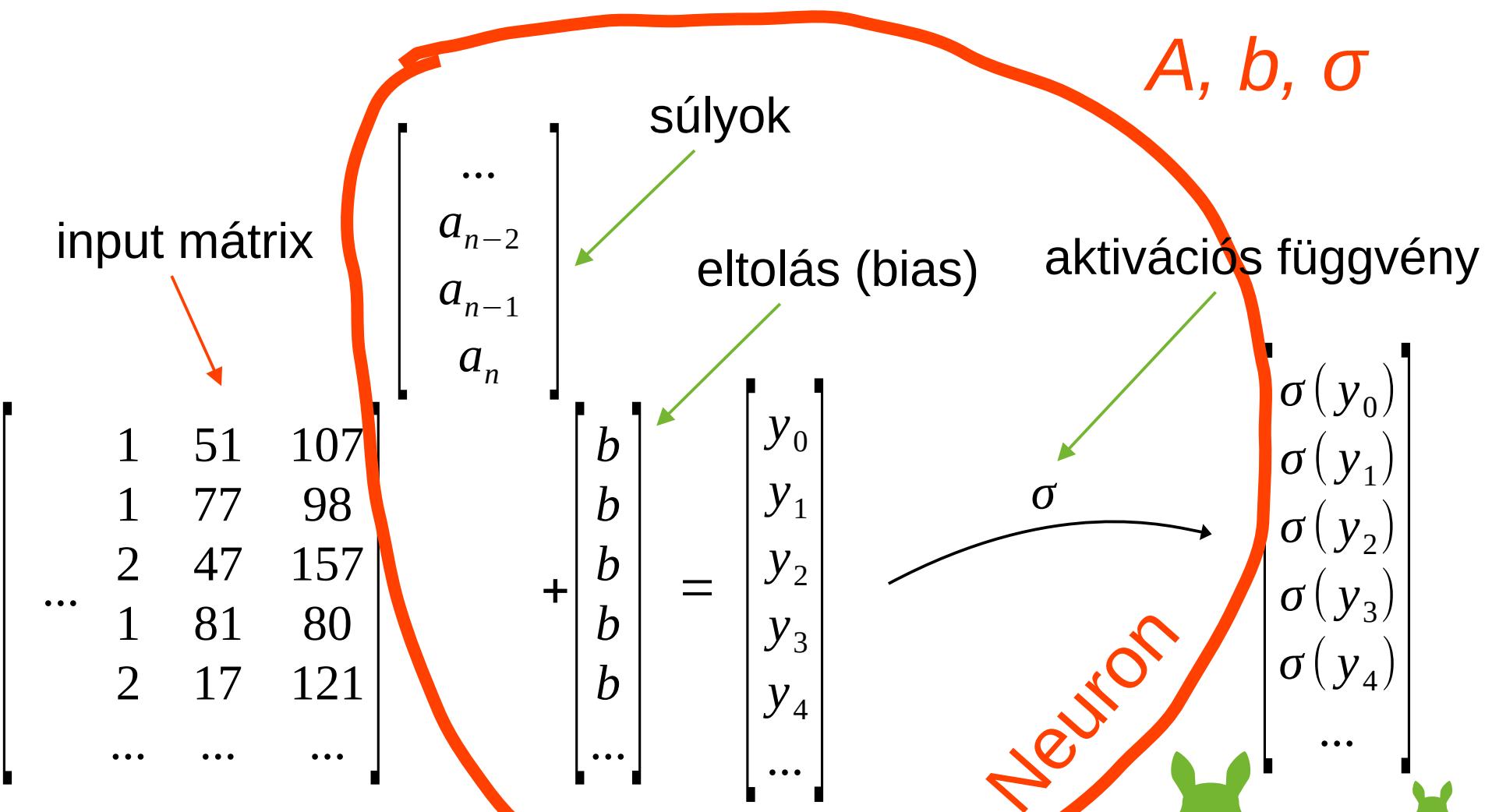
$$= \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ \dots \end{bmatrix}$$

aktivációs függvény

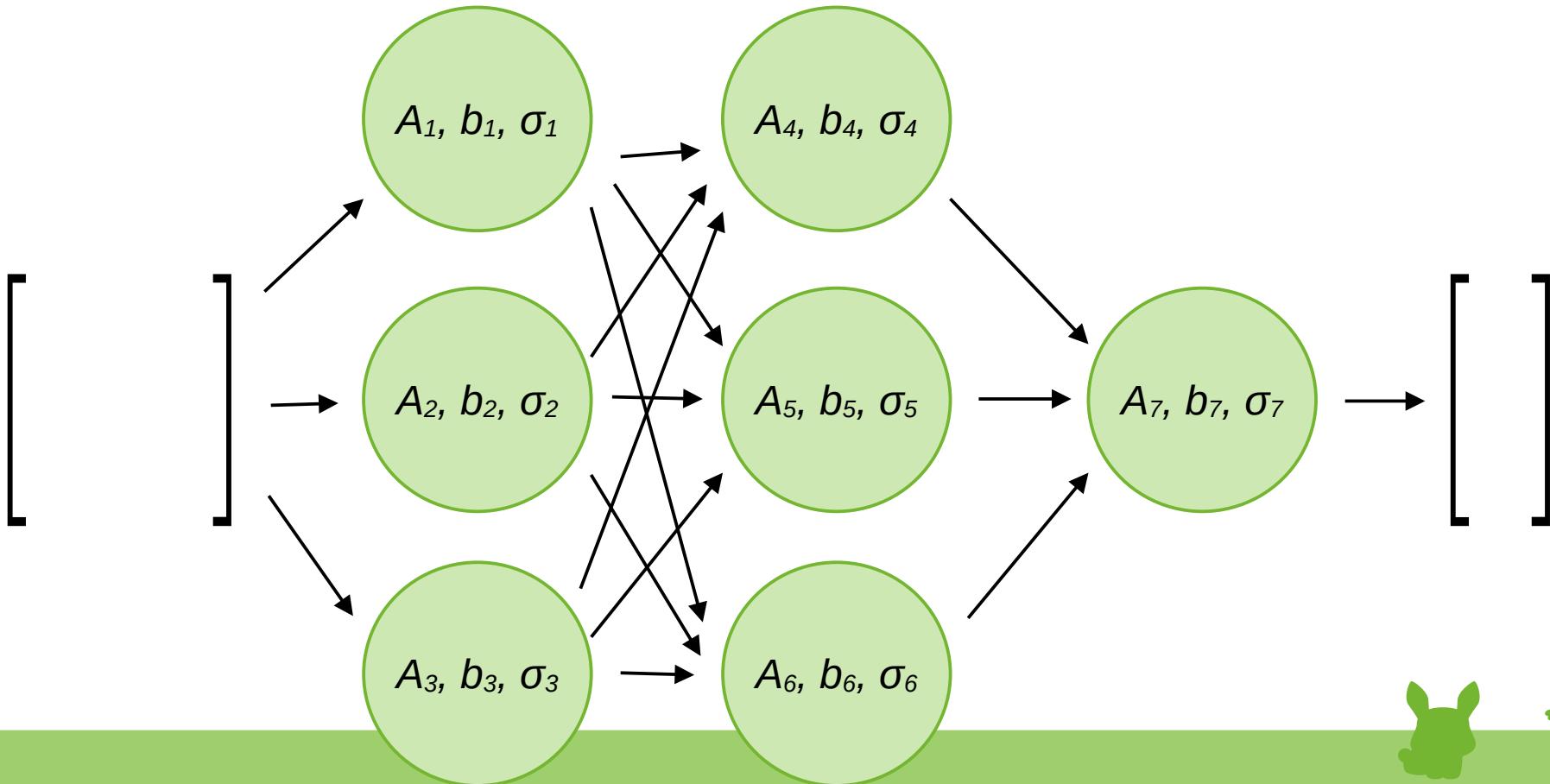
σ

$$\begin{bmatrix} \sigma(y_0) \\ \sigma(y_1) \\ \sigma(y_2) \\ \sigma(y_3) \\ \sigma(y_4) \\ \dots \end{bmatrix}$$

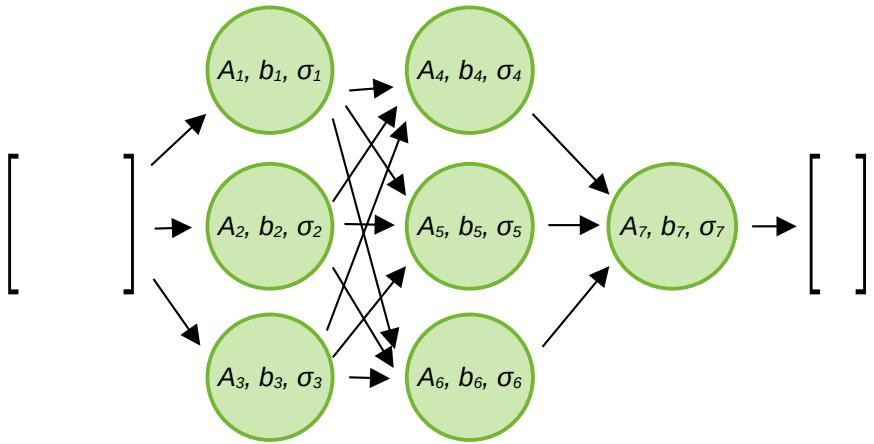




Neurális háló



Neurális háló

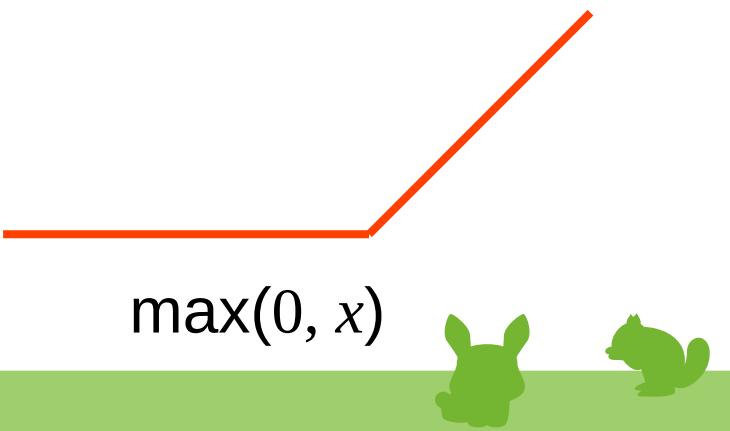


Tanítás: A_i, b_i beállítása, hogy a tanító halmazon a kimenet minél jobban megközelítse az elvártat.

Az összes σ_i gyakran azonos, pl. szigmoid, vagy ReLU.

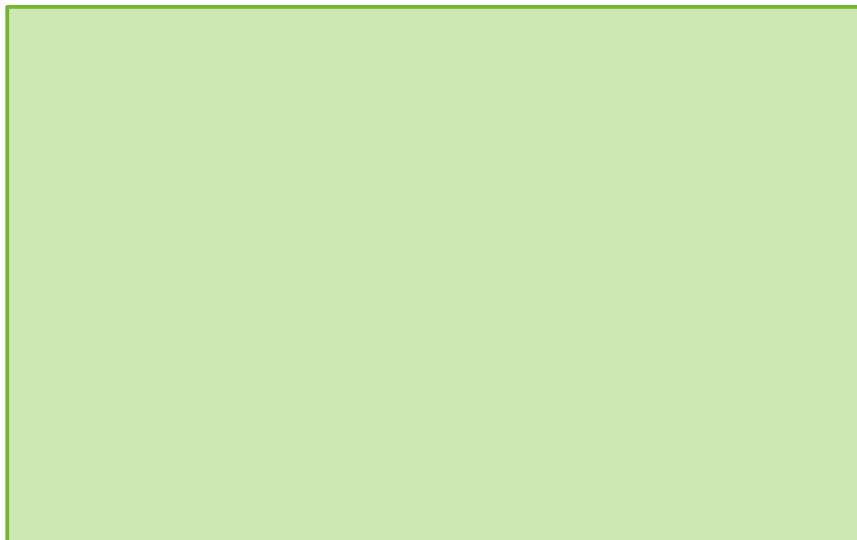
A graph of the sigmoid function, also known as the logistic function. The curve is red and passes through the point (0, 0.5). The equation for the sigmoid function is:

$$\frac{1}{1+e^{-k(x-c)}}$$



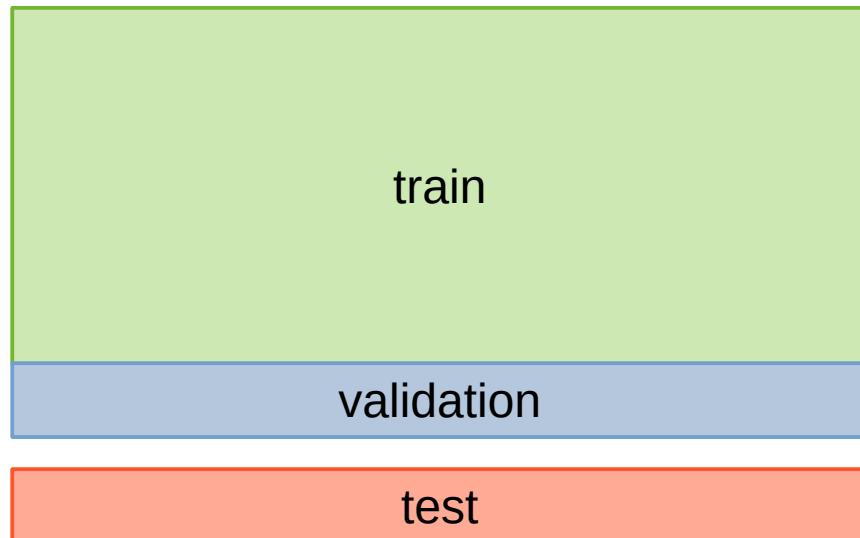
Tanító adathalmaz

Olyan adathalmaz, amelyen ismert a célváltozó pontos értéke.



Tanító adathalmaz

Olyan adathalmaz, amelyen ismert a célváltozó pontos értéke.



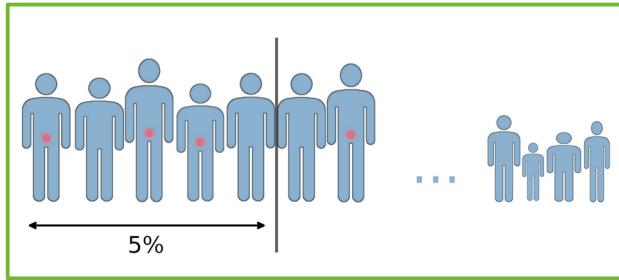
„A háború a politika folytatása más eszközökkel”
(Carl von Clausewitz, 1892)



„Az AI a statisztika folytatása más eszközökkel”



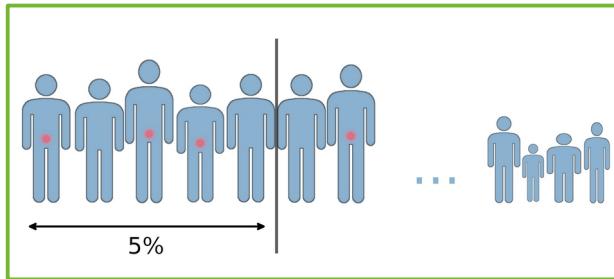
Példa neurális hálóra: prosztata szűrési stratégia



Szűrési stratégia: az összes férfi sorrendbe állítása



Példa neurális hálóra: prosztata szűrési stratégia



Szűrési stratégia: az összes férfi sorrendbe állítása

Hagyományos: 50-70 közötti férfiak elől, a többiek utánuk.

AI alapú: döntsön az AI az összes férfi összes egészségügyi rekordját feldolgozva.



Példa neurális hálóra: prosztata szűrési stratégia

AI alapú: döntsön az AI az összes férfi összes egészségügyi rekordját feldolgozva.



BNO, OENO, TTT kódok
+ életkor, lakóhely

prosztata rosszindulatú daganata



Példa neurális hálóra: prosztata szűrési stratégia

AI alapú: döntsön az AI az összes férfi összes egészségügyi rekordját feldolgozva.

♣ ♦ ♠ ♥ ♣ ♦ ♠ -> 1

♦ ♠ ♠ ♠ -> 0

♣ ♦ -> 0

.

.

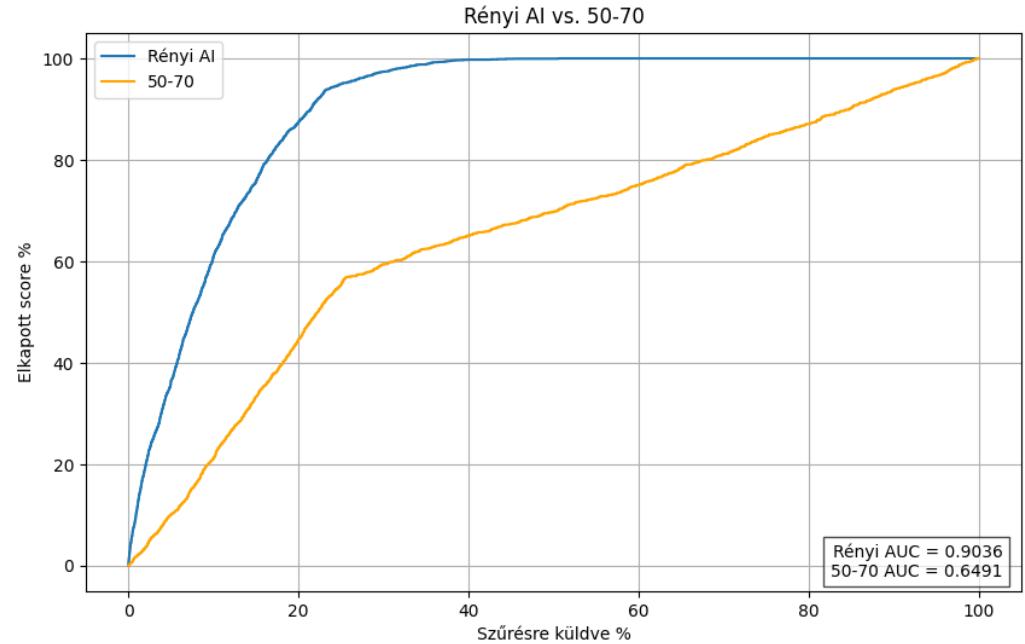
.



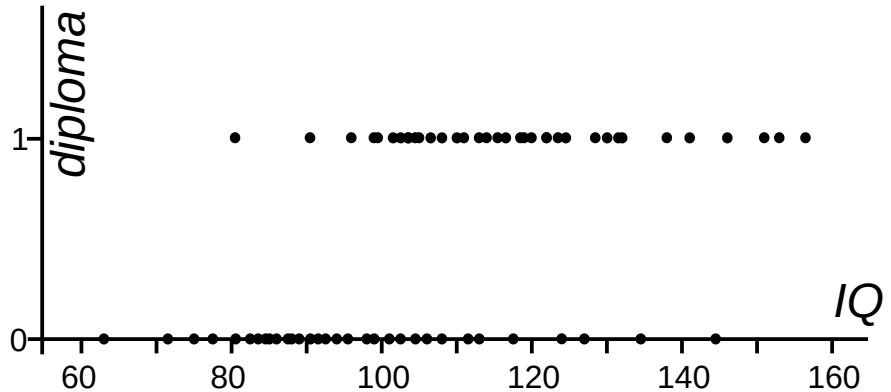
Példa neurális hálóra: prosztata szűrési stratégia

AI alapú: döntsön az AI az összes férfi összes egészségügyi rekordját feldolgozva.

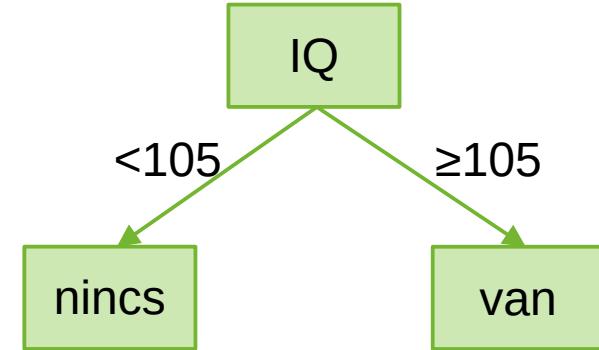
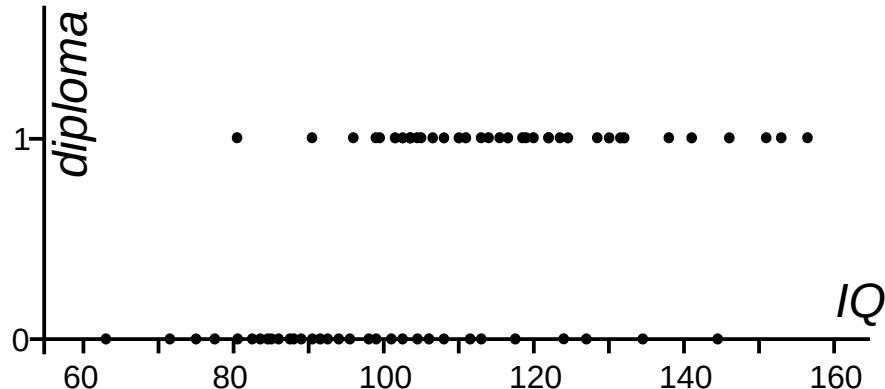
♣ ♦ ♠ ♥ ♣ ♦ ♠	-> 1
♦ ♠ ♠ ♠	-> 0
♣ ♦	-> 0
.	.
.	.
.	.



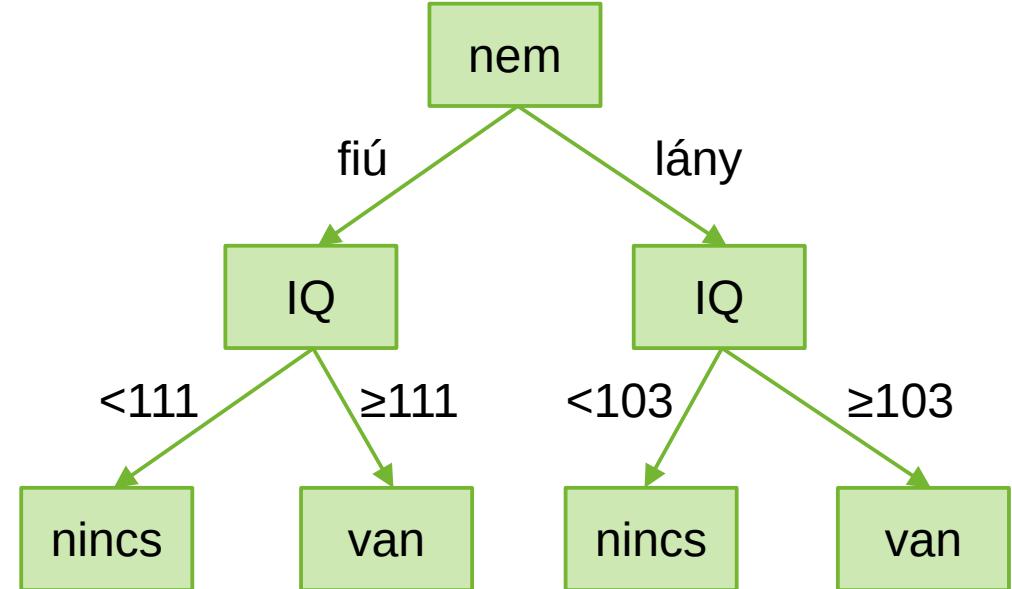
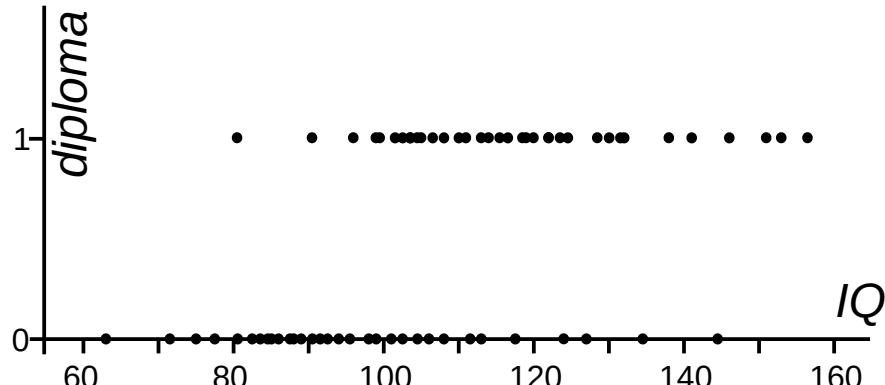
Döntési fa



Döntési fa

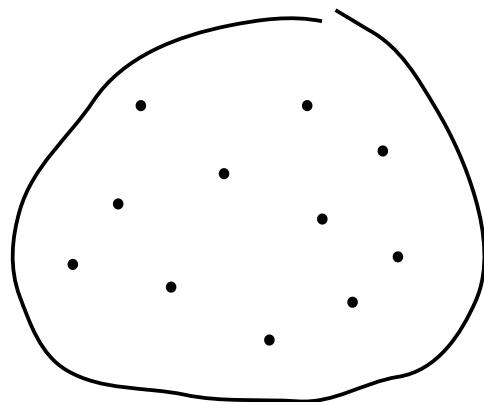


Döntési fa



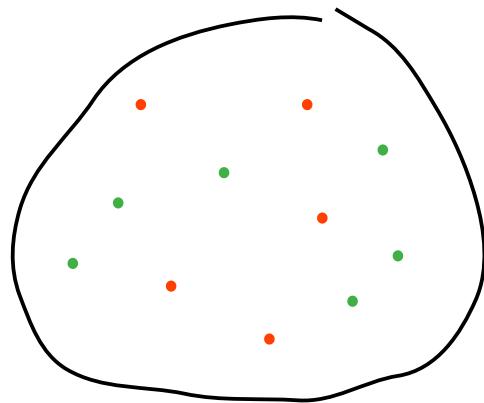
Véletlen döntési fa

véletlen tanító halmaz



Véletlen döntési fa

véletlen tanító halmaz



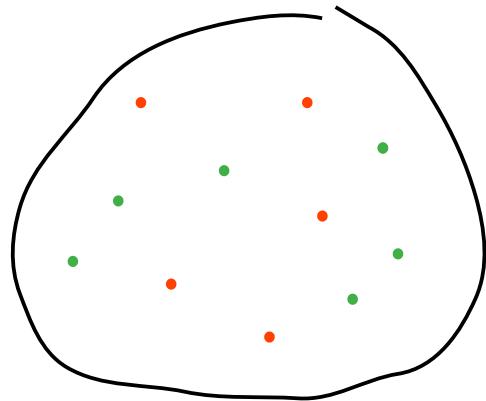
van

nincs



Véletlen döntési fa

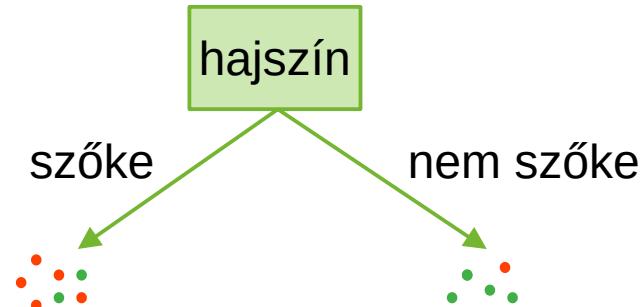
véletlen tanító halmaz



van

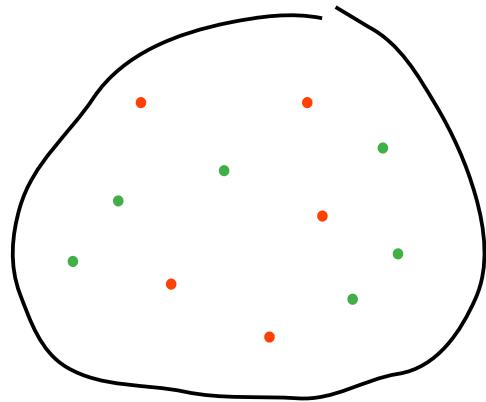
nincs

véletlen tulajdonság(ok)



Véletlen döntési fa

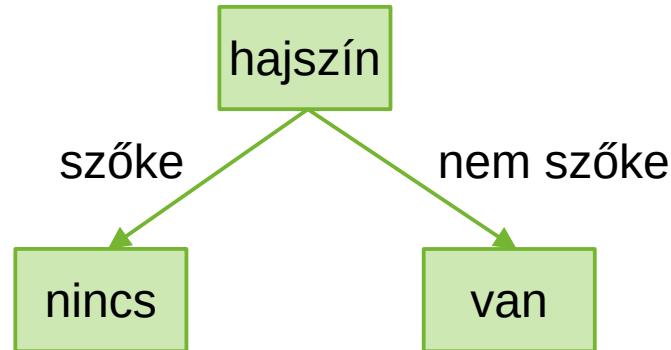
véletlen tanító halmaz



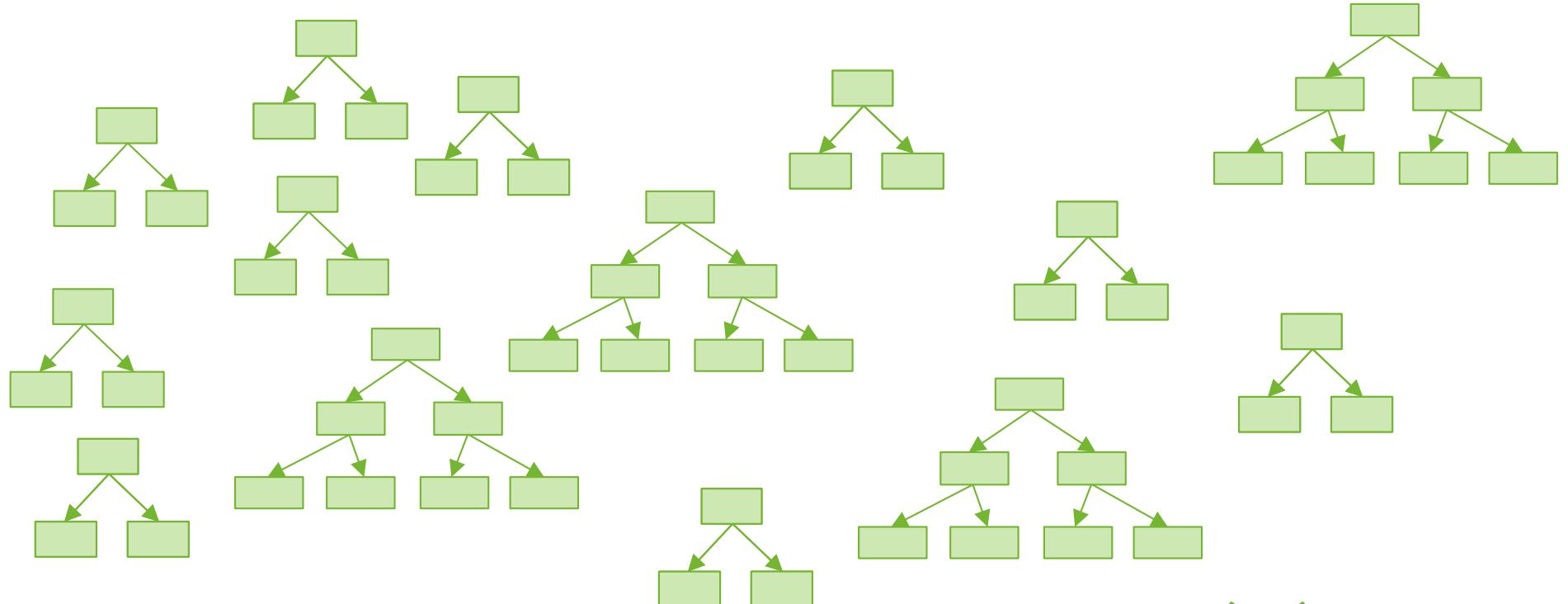
van

nincs

véletlen tulajdonság(ok)



Random forest



Sok véletlen döntési fa, többségi szavazással

