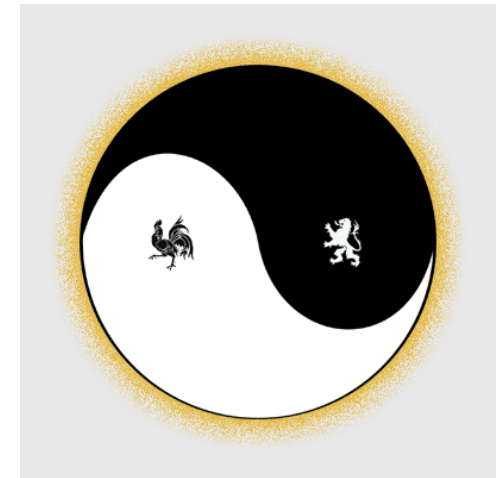


# Igen-nem változás előrejelzése több változó hatásának együttes elemzése alapján

## A többszörös logisztikus regresszió



Dr. Prohászka Zoltán  
*Az MTA doktora*  
Semmelweis Egyetem



Belgyógyászati és Hematológiai Klinika

2023-04-13

[Prohaszka.zoltan@med.semmelweis-univ.hu](mailto:Prohaszka.zoltan@med.semmelweis-univ.hu)

[www.semmelweis.hu/kutlab](http://www.semmelweis.hu/kutlab)

# Ismétlés

- Egyváltozós modellekkel megválaszolható kérdések
  - Egy megfigyelt különbség vagy arány a véletlen hatására alakult-e ki?
- Többváltozós modellekkel megválaszolható kérdések
  - Az adott független változó(k) szignifikáns kapcsolatban vannak-e függő változóval?
  - Mekkora a független változók egymáshoz képest mutatott relatív súlya a megfigyelt különbség (arány) kialakításában?
  - Más szavakkal: segít súlyozni, megítélni, típusba sorolni a független változóinkat.

## A független változó

Típusa	Jellegzetességei	Felismerési lehetőség	Példa
<b>Zavaró faktor (confounder)</b>	Kapcsolatban áll a rizikótényezővel és összefüggésben a kimenetellel	Rétegzett analízis Többváltozós analízis A statisztikai analízis azonban önmagában nem elegendő, hogy a kettő között különbséget lehessen tenni.	Öngyújtó, dohányzás, tüdőrák
<b>Közbeeső változó (intervening variable)</b>	Kapcsolatban áll a rizikótényezővel és okozati összefüggésben a kimenetellel, a rizikótényező általa fejt ki hatását		Doh, cotinine, toxikus metabolitok, tüdőrák
<b>Gátló faktor (supresser)</b>	Kapcsolatban áll a rizikótényezővel és a kimenetellel, a rizikófaktor hatását elfedi	Rétegzett analízis a supresser változó szerint Többváltozós analízis, adjusztálás (illesztés) a supresser változóra	Aspirin, MI, halálozás Orvosi hivatás, válás, öngyilkosság
<b>Hatásmódosító vagy interakciós változó (interaction variable)</b>	Hatásmódosítás. A rizikótényező hatása a kimenetelre egy harmadik változó értékétől függ.	Rétegzett analízis az interakciós változó szerint. Többváltozós analízis „with interaction terms”	Expozíció, inokuláció mértéke, betegség súlyossága

# Többváltozós analízisek

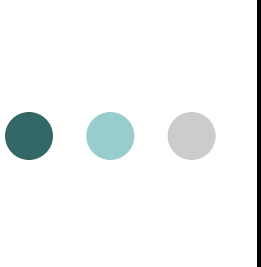
<b>Függő változó</b>	<b>Példa a kimenetelre</b>	<b>A használandó többszörös analitikai módszer</b>
Folyamatos	Vérnyomás, testsúly, hőmérséklet	Többszörös lineáris regresszió
Dichotom (igen-nem)	Halál, betegség kialakulása, felvétel osztályra	Többszörös logisztikus regresszió
Az eseményig eltelt idő	Halálig, dg.-ig eltelt idő	Cox regresszió (proportional hazard analízis)

# Problémabemutató, kérdésfeltevés:

- 195 szívelégtelenségben szenvedő beteg kereste fel az ambulanciát 1 év alatt.
- 92 betegnek súlyos szívelégtelensége volt szubjektív, ön-bevallásos skálán (NYHA)
  - Kik azok a betegek, akik várhatóan a súlyos csoportba tartoznak? *(milyen objektív mutatók alapján jósolható ez meg?)*
  - Ha több ilyen változó is van, akkor vajon ezek egymás hatásán keresztül vagy függetlenül jelzik a **súlyosságot**? *(akik anémiásak és/vagy veseelégtelenségük is van?)*
  - Ha van több független magyarázó változó, akkor ezeknek milyen az egymáshoz viszonyított fontossága, súlya?  
*(kétszer olyan fontos az anémia, mint a veseérintettség?)*

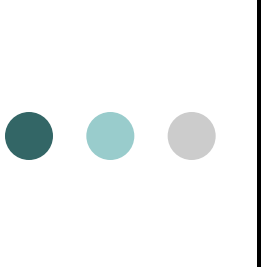
	Tüdőpangás nincs	Tüdőpangás van	<b>Össze n</b>
NYHA 1-2	82	21	<b>103</b>
NYHA 3-4	32	60	<b>92</b>
<b>Összesen</b>	<b>114</b>	<b>81</b>	<b>195</b>

Chi-square, df 38.39, 1  
P value P<0.0001



	Tüdőpan- gás nincs	Tüdőpan- gás van	Össze- se- n
NYHA 1- 2	82	21	<b>103</b>
NYHA 3- 4	32	60	<b>92</b>
Összesen	<b>114</b>	<b>81</b>	<b>195</b>

- A példa kérdésfeltevése szerint („mivel jósolhatom meg, hogy ki a súlyos beteg”):
  - Melyik a független változó?
  - Melyik a függő változó?
  - Hány független változót tudnék elemezni ebben a vizsgálatban?
  - Ki tudná megsaccolni az odds-ot?



	Tüdőpan- gás nincs	Tüdőpan- gás van	Össze- se- n
NYHA 1- 2	82	21	<b>103</b>
NYHA 3- 4	32	60	<b>92</b>
<b>Összesen</b>	<b>114</b>	<b>81</b>	<b>195</b>

○ A példa kérdésfeltevése szerint („mivel jósolhatom meg, hogy ki a súlyos beteg”):

- Melyik a független változó?
- Melyik a függő változó?
- Hány független változót tudnék elemezni ebben a vizsgálatban?
- Ki tudná megsaccolni az odds-ot?

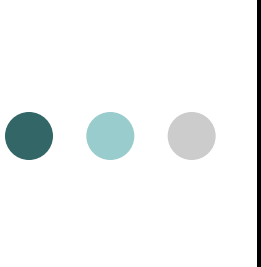
*Függő: NYHA*

*Független (magyarázó): pangás*

*Kisebb csoport 32.... $32/10=kb\ 3$*

*Odds?*





	Tüdőpan- gás nincs	Tüdőpan- gás van	Össze- se n
NYHA 1- 2	82	21	<b>103</b>
NYHA 3- 4	32	60	<b>92</b>
<b>Összesen</b>	<b>114</b>	<b>81</b>	<b>195</b>

- *A példa kérdésfeltevése szerint („mivel jósolhatom meg, hogy ki a súlyos beteg”):*
  - *Melyik a független változó?*
  - *Melyik a függő változó?*
  - *Hány független változót tudnék elemezni ebben a vizsgálatban?*
  - **Ki tudná megsaccolni az esély arányt (odds ratio)?**
  - **„Mennyi az esélye, hogy aki pang, az a súlyosabb csoportba tartozzon, ahhoz képest, mint aki nem pang?”**

	Tüdőpangás nincs	Tüdőpangás van	<b>Össze n</b>
NYHA 1-2	82	21	<b>103</b>
NYHA 3-4	32	60	<b>92</b>
<b>Összesen</b>	<b>114</b>	<b>81</b>	<b>195</b>

Chi-square, df 38.39, 1

P value  $P < 0.0001$

Odds ratio: 7,321, 95% confidence interval 3.847 to 13.93  
*(82/32 osztva 21/60....kb 3 osztva kb. 0.3)*

	Tüdőpangás nincs	Tüdőpangás van	<b>Össze n</b>
NYHA 1-2	82	21	<b>103</b>
NYHA 3-4	32	60	<b>92</b>
<b>Összesen</b>	<b>114</b>	<b>81</b>	<b>195</b>

Chi-square, df 38.39, 1

P value  $P < 0.0001$

Odds ratio: 7,321, 95% confidence interval 3.847 to 13.93

*90%-os CI: 4,26-12,56*

*99%-os CI: 3,14-14,07*

# Logisztikus regresszió

- Számszerűen fejezi ki az összefüggést egy függő és több független változó között
- 2 fő funkció
  - Predikció – egy modell és a független változók birtokában megmondható, milyen eséllyel kerül egy egyén az egyik vagy a mások csoportba
  - Feltérképezés – a változók közötti kapcsolatok és azok erősségének megismerése
- A függő változó (kimeneti változó): dichotom – két lehetséges értéke van – pl.: túlélő/halott, beteg/egészséges stb.

# A független (prediktor, magyarázó) változók

- A független változók lehetnek

- Kategorikusak

- Dichotomok – pl. átesett-e mandulaműtéten vagy sem?
- 0/1-nek kódoljuk – 0 - a referencia csoport, 1- a vizsgált csoport (Dummy változó)

- Ha  $k > 2$  db. kategória akár  $k-1$  dummy változót is csinálhatunk:

Milyen gyakran fogyaszt alkoholt?

- (A) soha
- (B) évente 1-2-szer
- (C) havonta 1-2-szer
- (D) hetente többször

akkor ebből három x-et képezünk:

válasz	$x_1$	$x_2$	$x_3$
A	0	0	0
B	1	0	0
C	0	1	0
D	0	0	1

- Folytonosak (vérnyomás, életkor)

- Egységnyi változáshoz tartozó esélyhányados növekedést adja meg
- Nehezen interpretálható
- → folyamatos változó kategorikussá alakítása (pl.: medián alatt/felet, melyik tercilisbe, decilisbe tartozik)
- folyamatos változó normalizálása (1 SD [Z-score], IQ tartomány)

- Számuk korlátozott – ökölszabály: kb. 6-10x legyen több eset a kisebbik prediktált csoportban, mint a változók száma

# Esély, odds, probability

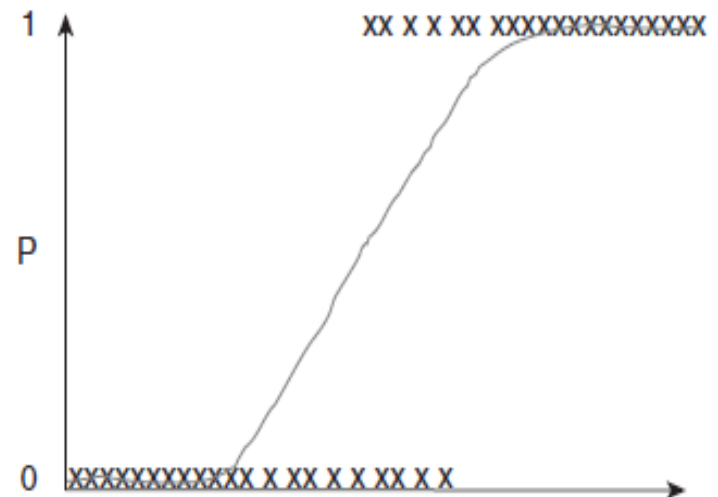
- Probability (valószínűség): milyen arányban látom az eseményt?
  - Tartománya: 0-1 (...%)
- Odds – Esély - dichotom változónál az egyik esemény valószínűsége osztva a másik esemény valószínűségével
- Pl.:  $p=0,8$ ? vagy  $p=0,25$ ? Odds=1?
  - $P=0,8 \rightarrow$  Odds: **4** az 1-hez (0,8/0,2)
  - $P=0,25 \rightarrow 0,33$  (0,25/0,75)
  - Odds =1  $\rightarrow p=0,5$  (0,5/0,5)
  - Odds tartománya: 0 – végtelen, dimenziója nincs
- „hányszor akkora a valószínűsége annak, hogy bekövetkezik, mint annak, hogy nem”
- ...megadja mennyivel *változik* az esélye annak, hogy az eseményes csoportba tartozik egy egyén, ha a *független változó értéke egy egységnyivel változik*

# Esélyhányados, esély arány – Odds ratio

- Két esély hányadosa
- Férfiak: 1000fő – 750 dohányos, 250 nem
- Dohányzás valószínűsége ( $p$ )= $750/1000=75\%$
- Nem dohányzás vsz  $p=250/1000=25\%$
- Dohányos férfi esélye:  $0,75/0,25 = 3$
- Nőknél: 1000 fő – 250 dohányos, 750 nem
- Itt az esély:  $0,33$
- Esélyhányados:  $3/0,33=9$
- Meghatározza, hogy az adott csoportban az események esélye hányszorosa a másik csoportban észlelt esélynek

# Logisztikus regresszió egyenlete

- Mivel nem egy értéket szeretnénk becsülni, mint a lineáris regressziónál, hanem azt, hogy az egyik csoportba tartozik-e az egyén vagy sem (0 vs. 1), inkább valószínűsége vagyunk kíváncsiak
- A valószínűség értéke 0 és 1 között mozog – a predikció során azonban nagyobb értékek is kijöhetnek – transzformációra van szükség
- További érv: lineáris egyenes nem illeszthető a két kategória miatt → logaritmikus transzformáció





# Logisztikus regresszió egyenlete

$$\text{logit } [p(x)] = \log \left[ \frac{p(x)}{1-p(x)} \right] = a + b_1x_1 + b_2x_2 + b_3x_3 \dots$$

- $p$  – a valószínűség, hogy az egyén valamelyik csoportba tartozik
- $a$  – a konstans
- $b$  – az adott prediktor regressziós koefficiense
  - Ez az  $\ln(\text{OR})!$
  - Sok helyen béta ( $\beta$ )
  - $\text{OR} = e^\beta$  – egymásból számíthatóak

# Esélyhányados, odds ratio (OR)

- OR – ezzel számszerűsíti a logisztikus regresszió a függő és független változó közötti összefüggést
- Megadja mennyivel változik az esélye annak, hogy az eseményes csoportba tartozik egy egyén, ha a független változó értéke egy egységnyel nő
- $e$ -t a függő változóhoz tartozó regressziós koefficiens ( $b$ ) értékére emeljük ( $e^b$ )
  - Pl.: a koefficiens ( $b$ ) = 3  $\rightarrow$   $OR=e^3=2,72^3=20,09$
  - Ha a függő változó: él/halott, független: betegség súlyossági stádium  $\rightarrow$
  - „kb. 20x nagyobb eséllyel hal meg az a beteg, akinek egy stádiummal súlyosabb a betegsége, mint akinek kevésbé súlyos.”

# A regressziós együtthatókból kiolvashatjuk...

## Béta (b):

- ... hogy az adott változó hogyan befolyásolja az  $y$  (kimeneteli változó) bekövetkezésének esélyét:
  - $\beta > 0$  – növeli,  $\beta < 0$  – csökkenti,  $\beta = 0$  – nem befolyásolja

## OR:

- Folytonos magyarázó változóknál:
  - ... hogy egy egységnyi növekedés hányszorosára növeli a  $y$  bekövetkeztének esélyét (odds)
- ha az  $x$  magyarázó változó dichotom
  - ... hogy a vizsgált csoportban az  $y$  bekövetkezésének esélye (odds) hányszorososa a referenciacsoportbelinek
- ha a magyarázó változó több, mint 2 kategóriás
  - ... hogy a szóban forgó csoportban az  $y$  bekövetkezésének esélye (odds) hányszorososa a referenciacsoportbelinek

# Wald chi négyzet teszt

- A regressziós egyenletben az egyes prediktorok (x) hatásának szignifikanciáját teszteli
- $H_0$  – a változónak nincs szerepe a függő változó kimenetében
- Ha a OR 95%-os konfidencia intervalluma tartalmazza az 1-es értéket, akkor az adott prediktor nem szignifikáns

**TABLE 7.1**

Frequency of primary PCP prophylaxis among patients whose AIDS-defining diagnosis was PCP.

Characteristic	Did not receive PCP prophylaxis n (%)	Received primary PCP prophylaxis n (%)	Adjusted P value	Adjusted odds ratio	95% confidence limits
Total	212 (65.0)	114 (35.0)			
Age group					
<35 years	61 (69.3)	27 (30.7)		1.0 (ref.)	
≥35 years	151 (63.5)	87 (36.6)	0.55	1.19	0.68, 2.07
Ethnicity					
Nonwhite	74 (77.9)	21 (22.1)		0.49	0.28, 0.87
White	138 (59.7)	93 (40.3)	0.01	1.0 (ref.)	
Sex					
Male	206 (64.6)	113 (35.4)		0.81	0.06, 10.12
Female	6 (85.7)	1 (14.3)	0.87	1.0 (ref.)	
Sexual orientation					
Gay/bisexual man	185 (62.7)	110 (37.3)		3.19	0.78, 13.03
Heterosexual	27 (87.1)	4 (12.9)	0.11	1.0 (ref.)	
Injection drug use					
Yes	35 (72.9)	13 (27.1)		1.11	0.49, 2.54
No	177 (63.7)	101 (36.3)	0.80	1.0 (ref.)	
Insurance					
None	52 (82.5)	11 (17.5)		0.35	0.17, 0.73
Public/private	151 (59.5)	103 (40.6)	0.005	1.0 (ref.)	

Adapted with permission from Schwarcz, S.K., et al. "Prevention of *Pneumocystis carinii* pneumonia: Who are we missing?" *AIDS* 1997; 11:1263-8. Copyright Rapid Science Publishers Ltd.

Wald Chi négyzet

OR a referencia csoportban mindig 1

Ha a OR CI-je nem tartalmazza az 1-et (50-50%), akkor szignifikáns.

Ha  $OR < 1$ , akkor az eseményre a referencia csoportnak van nagyobb esélye

**Table 4.** Risk of tumor multiplicity based on gene promoter methylation in tumor<sup>1</sup>.

	Unadjusted			Adjusted <sup>2</sup>		
	OR	95% CI	P value	OR	95% CI	P value
<b><i>MGMT1-Mp</i></b>						
unmethylated	1 (ref)	-		1 (ref)	-	
methylated	1.58	1.02–2.44	0.03	1.57	1.01–2.43	0.04
<b><i>MGMT2-Enh</i></b>						
unmethylated	1 (ref)	-		1 (ref)	-	
methylated	1.48	1.14–1.93	0.003	1.50	1.14–1.96	0.003
<b><i>CDKN2A</i></b>						
unmethylated	1 (ref)	-		1 (ref)	-	
methylated	1.23	0.92–1.65	0.15	1.23	0.92–1.65	0.16
<b><i>SFRP1</i></b>						
unmethylated	1 (ref)	-		1 (ref)	-	
methylated	1.06	0.87–1.29	0.51	1.06	0.87–1.29	0.55
<b><i>TMEFF2</i></b>						
unmethylated	1 (ref)	-		1 (ref)	-	
methylated	1.13	0.89–1.43	0.29	1.13	0.89–1.44	0.29
<b><i>HS3ST2 (3OST2)</i></b>						
unmethylated	1 (ref)	-		1 (ref)	-	
methylated	1.12	0.90–1.38	0.29	1.12	0.90–1.40	0.29
<b><i>RASSF1A</i></b>						
unmethylated	1 (ref)	-		1 (ref)	-	
methylated	1.95	1.01–3.74	0.04	2.02	1.03–3.93	0.03
<b><i>GATA4</i></b>						
unmethylated	1 (ref)	-		1 (ref)	-	
methylated	1.10	0.89–1.35	0.34	1.10	0.89–1.36	0.34

OR, odds ratio; 95% CI, 95% confidence interval.

<sup>1</sup>Binomial logistic regression analysis.

<sup>2</sup>Adjusted by age and gender.

doi:10.1371/journal.pone.0008777.t004

Referencia csoport

Vizsgálati csoport

A nemre és korra  
illesztés nem  
befolyásolja az OR-t

‘Unadjusted’ és ‘Adjusted’ analízis között különbség, hogy az előzőnél az összes felsorolt változó független változó a modellben (1 független/modell), míg a második oszlopban pedig illesztve van a korra és a nemre (3 független változó/modell).

# Modellépítés - Milyen független változók legyenek?

- A kérdésfeltevés, a függő változó lényeges szempont
- Ha a betegség kimenetelével kapcsolatos:
  - Alap adatok – nem, életkor
  - Betegség súlyosságát jelző mutató
  - Az általunk tesztelni kívánt változó
  - Irodalmi adatokból ismert prediktorok
- Lehetséges irányvonal: egyváltozós modellben szoros összefüggést mutat

# Multikollinearitás

- Ha független változók között szoros korreláció áll fenn. Akkor a legrobosztusabb a modell, ha a független változók a függő változóval és nem egymással korrelálnak
- Az eredményt nem módosítja drasztikusan, de a standard error nagy lesz az érintett változóknak
- Pl.: „Dummy variable trap” – kettőnél több kategóriát ( $k > 2$ ) tartalmazó változót dummy változóvá teszünk és az összes  $(k-1)$  új változót beépítjük a modellbe
- Kerülendő



# „Automatikus” modellépítés

- Forward selection – a legtöbbet hozzáadó beválogatása, egészen addig, amíg javul a modell
- Backward elimination – a legkevésbé szignifikáns változó kivétele a modell javulásáig
- Stepwise regression – nincs semelyik változónak „bérelt helye” – újra ki- vagy bekerülhet
- Biológiaiailag releváns? Benne van a kérdéses változó?
  - Korlátok

# Modell szignifikancia

- Meghatározza, hogy az x-ek (prediktorok) segítségével mennyire pontosan határozható meg az y (függő változó).
- Chi négyzet eloszlást mutat
- Minél nagyobb a statisztika értéke, annál jobb az illeszkedés
- $H_0$  – a modell illeszkedése olyan mint a null modellnek
- ► Ha szignifikáns – arra utal kevés a nem megmagyarázott variabilitás – jó a modell
- Ha nem szignifikáns – a modell nem illeszkedik jól, fel nem tárt magyarázó tényezők is vannak (vagy nincs kapcsolat), „reziduum”

# Statistics//Advanced Linear/Nonlinear Models//Nonlinear Estimation//Quick logit regression

STATISTICA - [Data: adatbzis\_1.sta\* (23v by 195c)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Window Help

Resume... Ctrl+R

to Report Add to MS Word

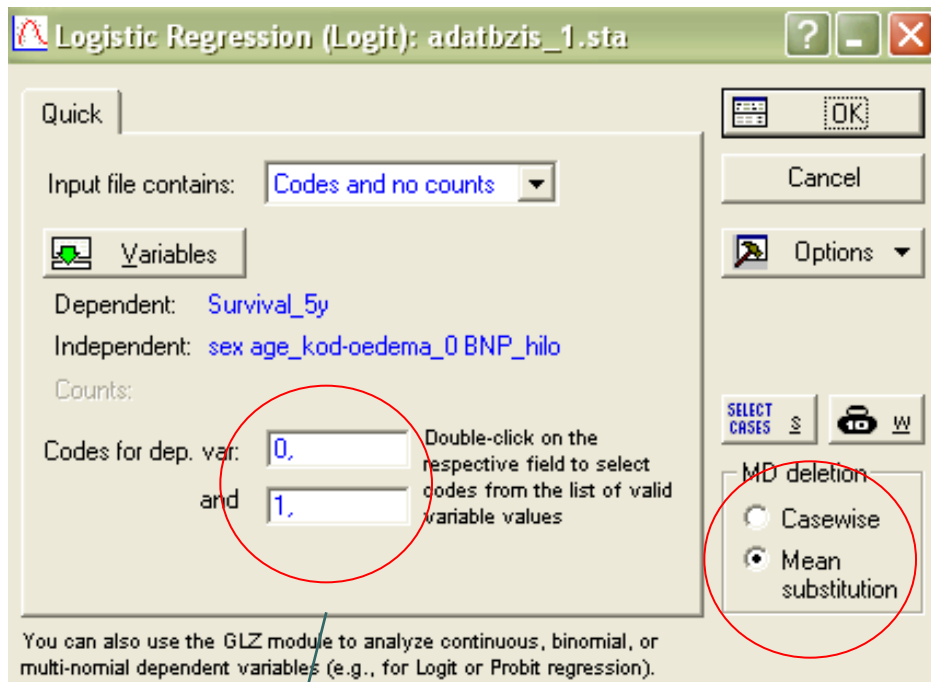
Basic Statistics/Tables  
 Multiple Regression  
 ANOVA  
 Nonparametrics  
 Distribution Fitting

Advanced Linear/Nonlinear Models  
 Multivariate Exploratory Techniques  
 Industrial Statistics & Six Sigma  
 Power Analysis  
 Automated Neural Networks  
 PLS, PCA, Multivariate/Batch SPC  
 Variance Estimation and Precision (VEPAC)

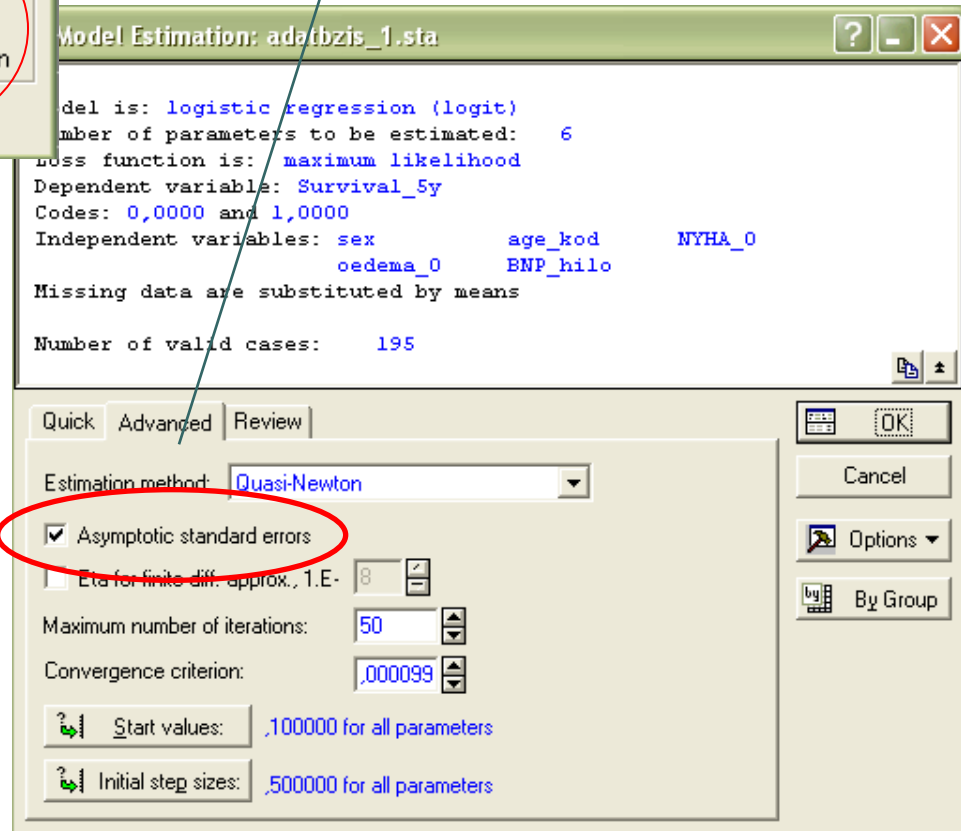
Statistics of Block Data  
 STATISTICA Visual Basic  
 Batch (ByGroup) Analysis  
 Probability Calculator

GLM General Linear Models  
 GLZ Generalized Linear/Nonlinear Models  
 GRM General Regression Models  
 PLS General Partial Least Squares Models  
 NIPALS Algorithm (PCA/PLS)  
 Variance Components  
 Survival Analysis  
 Nonlinear Estimation  
 Fixed Nonlinear Regression  
 Log-Linear Analysis of Frequency Tables  
 Time Series/Forecasting  
 Structural Equation Modeling

	sex	age	ag	Syst_0	Diast_0	BMI_0	loop_diur_0	Na_0				
1	1	83,13						139,1				
2	1	71,05						136,1				
3	1	78,48						139,1				
4	1	66,81						141,1				
5	1	48,79						129,1				
6	1	51,58						135,1				
7	0	75,85						140,1				
8	1	71,23						145,1				
9	0	52,42						138,1				
10	1	74,37						148,1				
11	1	80,83						139,1				
12	1	55,07						119,1				
13	1	68,02						146,1				
14	0	58,16						140,1				
15	1	80,01	3	3	0	1	66,00	120,00	80,00	32,83	1	147,1
16	1	60,60	2	2	0	0	84,00	90,00	60,00	35,06	1	140,1



Wald- chi négyzet értékek és a konfidencia intervallum értéket csak akkor kapunk, ha ezt bejelöljük!!!



Ha itt megcserélem a kódokat az OR reciproka lesz (a kérdésnek, interpretációnak megfelelően választható)

Béta

e a 0,4666-on =  
1,593

Modell illeszkedés:

Itt a H0 – egyetlen független változó sincs kapcsolatban a függő változó log esélyével

N=195

Model: Logistic regression (logit) N of 0's: 85 1's: 110 (adatbzis\_1.sta)  
Dep. var: Survival\_5y Loss: Max likelihood (MS-err. scaled to 1)  
Final loss: 106,27750231 Chi2( 5)=54,558 p=,00000

	Const.B0	sex	age_kod	NYHA_0	oedema_0	BNP_hilo
Estimate	-3,122736	0,4656374	0,1380139	0,8700661	0,6663792	0,9754604
Standard Error	0,7211558	0,3924921	0,1975935	0,2129722	0,3497632	0,3557418
t(189)	-4,330182	1,186438	0,6984738	4,08535	1,905229	2,742046
p-level	0,00002414451	0,2369387	0,4857395	0,00006497258	0,05826845	0,006692511
-95%CL	-4,545284	-0,3085605	-0,2517581	0,4499581	-0,02356211	0,2737258
+95%CL	-1,700188	1,239895	0,5277858	1,290174	1,356321	1,677195
Wald's Chi-square	18,75048	1,407635	0,4878655	16,69008	3,629898	7,518816
p-level	0,00001494387	0,2354583	0,4848861	0,00004411755	0,05675923	0,006109174
Odds ratio (unit ch)	0,04403652	1,593077	1,147991	2,387069	1,947174	2,652388
-95%CL	0,01061715	0,7345035	0,7774328	1,568246	0,9767133	1,314854
+95%CL	0,1826493	3,455252	1,695175	3,633419	3,881883	5,350527
Odds ratio (range)		1,593077	1,51292	13,60175	1,947174	2,652388
-95%CL		0,7345035	0,4698817	3,856941	0,9767133	1,314854
+95%CL		3,455252	4,871284	47,96743	3,881883	5,350527

Kategorikussá alakított  
folytonos változó – 20  
évenkénti kategóriák

Kategorikussá alakított  
folytonos változó (high vs. low)  
– medián értéknél a vágópont.

N=195	Const.B0	BNP_hilo
<b>Estimate</b>	<b>-0,4399513</b>	<b>1,458521</b>
Standard Error	0,2080082	0,3092287
t(193)	-2,115067	4,71664
p-level	0,03570723	0,000004589766
-95%CL	-0,8502124	0,8486192
+95%CL	-0,0296902	2,068423
Wald's Chi-square	4,47351	22,2467
p-level	0,0344319	0,000002407814
Odds ratio (unit ch)	0,6440678	4,299595
-95%CL	0,4273242	2,336418
+95%CL	0,9707462	7,912332
Odds ratio (range)		4,299595
-95%CL		2,336418
+95%CL		7,912332

+  
N  
Y  
H  
A

N=195	Const.B0	Diast_hilo
<b>Estimate</b>	<b>0,6363246</b>	<b>-0,6845298</b>
Standard Error	0,222546	0,2962649
t(193)	2,859295	-2,310533
p-level	0,004712736	0,02191569
-95%CL	0,1973901	-1,268862
+95%CL	1,075259	-0,1001971
Wald's Chi-square	8,175569	5,338561
p-level	0,00424858	0,02086502
Odds ratio (unit ch)	1,889523	0,5043273
-95%CL	1,218219	0,2811513
+95%CL	2,930752	0,9046592
Odds ratio (range)		0,5043273
-95%CL		0,2811513
+95%CL		0,9046592

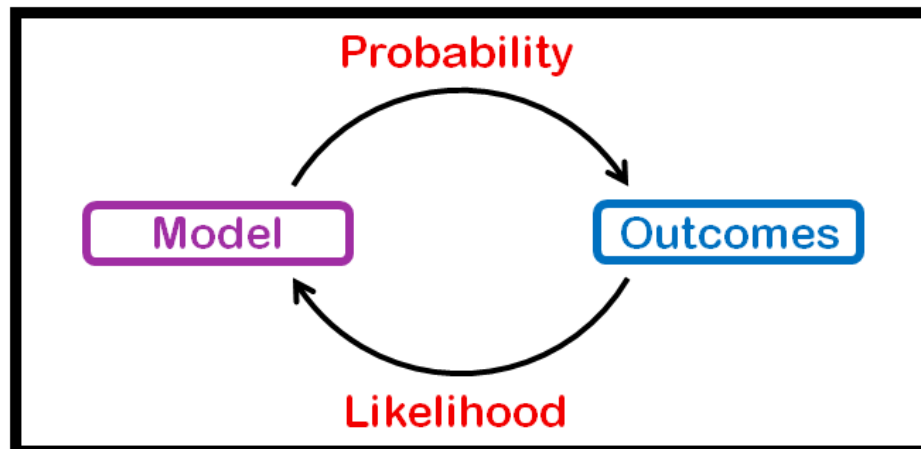
N=195	Const.B0	NYHA_0	BNP_hilo
<b>Estimate</b>	<b>-2,439296</b>	<b>0,9607115</b>	<b>0,9446079</b>
Standard Error	0,4911867	0,2061979	0,3399505
t(192)	-4,966127	4,659172	2,778663
p-level	0,000001504267	0,000005922739	0,006000536
-95%CL	-3,408111	0,5540075	0,2740906
+95%CL	-1,470481	1,367416	1,615125
Wald's Chi-square	24,66242	21,70788	7,720967
p-level	0,0000006865427	0,000003187608	0,005461497
Odds ratio (unit ch)	0,08722226	2,613555	2,571805
-95%CL	0,03310368	1,740213	1,315334
+95%CL	0,2298149	3,925193	5,028517
Odds ratio (range)		17,85234	2,571805
-95%CL		5,269958	1,315334
+95%CL		60,47599	5,028517

N=195	Const.B0	Diast_hilo	NYHA_0
<b>Estimate</b>	<b>-2,127318</b>	<b>-0,3136609</b>	<b>1,093131</b>
Standard Error	0,5495127	0,3312247	0,2010591
t(192)	-3,871281	-0,9469731	5,436863
p-level	0,0001483232	0,3448428	0,0000001634818
-95%CL	-3,211175	-0,9669672	0,6965626
+95%CL	-1,043461	0,3396455	1,489699
Wald's Chi-square	14,98682	0,8967581	29,55948
p-level	0,0001084797	0,3436596	0,00000005462294
Odds ratio (unit ch)	0,1191564	0,7307668	2,983601
-95%CL	0,04030922	0,3802345	2,006843
+95%CL	0,3522334	1,40445	4,435762
Odds ratio (range)		0,7307668	26,55965
-95%CL		0,3002345	0,002393
+95%CL		1,40445	87,27798

Ez alapján a BNP hilo tényleges prediktor, míg a diasztolés vérnyomás csak confounder vagy közbeeső változó a NYHA súlyossági stádiumok mellett

# Modellilleszkedés meghatározása

- Nincs olyan jól interpretálható mutató, mint az  $R^2$  a lineáris regressziónál
- A legnagyobb valószínűség (maximum likelihood, LL), hasonlóan a lineáris regresszió legkisebb négyzetek által meghatározott regressziós egyenes becsléséhez. Itt: azon paraméterek megtalálása, amelyek mellett legvalószínűbb, hogy éppen a megfigyelt értékeket kapjuk.
- $-2LL$  –  $-2\log$ -likelihood
  - Olyan regressziós B értékeket keres, ahol a legnagyobb a LL – iterációs eljárás
  - Chi négyzet eloszlású
  - Minél nagyobb a  $-2LL$ , annál rosszabb hatékonyságú az előrejelzés
- Nagelkerke  $R^2$ – tökéletes illeszkedés esetén értéke 1: a kiindulási (null modell) és az aktuális modell között LL értéket hasonlítja össze és osztja a lehetséges maximális értékkel >> a modell illeszkedés javulását/ romlását mutatja



# Modell diagnosztika

- A modell illeszkedése >> maximum likelihood becslés alapján ( $R^2$  értelmezése problémás lenne)
- Reziduumok analízise >> A megfigyelt és a prediktált érték közötti eltérés (folytonos vált.)
- Találatmátrix: logisztikus regresszió esetén a helyesen klasszifikált esetek aránya (binomiális vált.)

The screenshot shows the SPSS Results window for a logistic regression model. The main text displays the following statistics:

- Model is: **logistic regression (logit)**
- No. of 0's: 110,0000
- No. of 1's: 85,00000
- Dependent variable: **Survival\_5y**
- Independent variables: 3
- Loss function is: **maximum likelihood** Final value: 109,3231
- 2\*log(Likelihood): for this model=218,6462 intercept only
- Chi-square = 48,46719 df = 3 p = ,0000000

Below the text is a classification matrix titled "Classification of Cases (adatbzis\_1) Odds ratio: 4,3974 Perc. correct: 68,21%".

	Pred. 1,000000	Pred. 0,000000	Percent Correct
Observed 1,000000	84	26	76,36364
Observed 0,000000	36	49	57,64706

The bottom part of the screenshot shows the "Review" tab of the SPSS dialog box. The "Classification of cases & odds ratio" option is selected and circled in red. Other options include "Observed, predicted, residual vals", "Normal probability plot of residuals", "Save predicted and residual values", "Histogram of residuals", "Half-normal probability plot", "Predicted vs. observed values", and "Predicted vs. residual values".



# Köszönöm a figyelmet!

