

Többtenyezős regresszió (az adatelemzésben):

Y közelítése $b_1X_1 + b_2X_2 + \dots + b_JX_J$ alakban,

$$y_n = b_1x_{n,1} + b_2x_{n,2} + \dots + b_Jx_{n,J} + \varepsilon_n \quad , \quad n = 1, 2, \dots, N \quad ,$$

ahol $\varepsilon_1, \dots, \varepsilon_N$ független $\mathcal{N}(0, \sigma^2)$ eloszlású valószínűségi változók, és σ ismeretlen paraméter, és $\{x_{1,1}, x_{2,1}, \dots, x_{N,J}\}$ ismert értékek.

Y kimeneti változó számértékű

X_1, \dots, X_J magyarázó változók

Elterjedt szóhasználat: többváltozós regresszió.

Bináris logisztikus regresszió:

az Y változó bináris, értékei = $\begin{cases} 1 & p_1 \text{ valószínűséggel} \\ 2 & p_2 \text{ valószínűséggel} \end{cases}$

$$\vartheta = \frac{P(Y = 1)}{P(Y = 2)} : \text{esélyhányados}$$

$\log(\vartheta)$ közelítése $a + b_1X_1 + b_2X_2 + \dots + b_JX_J$ alakban,

ahol X_1, \dots, X_J magyarázó változók

Multinomiális logisztikus regresszió:

az Y változó kategoriális, értékei = $\begin{cases} 1 & p_1 \text{ valószínűséggel} \\ 2 & p_2 \text{ valószínűséggel} \\ \dots \\ I & p_I \text{ valószínűséggel} \end{cases}$

$$\vartheta = \frac{P(Y = i)}{P(Y = 1)} : \text{az } i \text{-ik esélyhányados, } i = 2, 3, \dots, I$$

$\log(\vartheta_i)$ közelítése $a_i + b_{1,i}X_1 + b_{2,i}X_2 + \dots + b_{J,i}X_J$ alakban,

ahol X_1, \dots, X_J magyarázó változók

Ordinális logisztikus regresszió (PLUM) :

$$\text{az } Y \text{ változó ordinális, értékei} = \begin{cases} 1 & p_1 \text{ valószínűséggel} \\ 2 & p_2 \text{ valószínűséggel} \\ \dots & \\ I & p_I \text{ valószínűséggel} \end{cases}$$

$$\vartheta = \frac{P(Y \leq i)}{P(Y > i)} : \text{ az } i \text{-ik esélyhányados, } i = 1, 2, 3, \dots, (I - 1)$$

$\log(\vartheta_i)$ közelítése $a_i + b_{1,i}X_1 + b_{2,i}X_2 + \dots + b_{J,i}X_J$ alakban, ahol X_1, \dots, X_J magyarázó változók

Ordinális logisztikus regresszió (CR) :

$$\vartheta = \frac{P(Y = i + 1)}{P(Y \leq i)} : \text{ az } i \text{-ik esélyhányados, } i = 1, 2, 3, \dots, (I - 1)$$

Probit regresszió : Y bináris változó

Az egy magyarázó változós eset: a

$\vartheta = P(Y = 1)$ valószínűséget $P(\xi < X)$ alakban keressük, ahol $\xi \sim \mathcal{N}(\mu, \sigma^2)$ ahol μ, σ ismeretlen paraméterek.

Több magyarázó változó esetében $\Phi^{-1}(\vartheta)$ közelítését keressük $a + b_1X_1 + b_2X_2 + \dots + b_JX_J$ alakban, ahol X_1, \dots, X_J magyarázó változók, Φ a standard normális eloszlásfüggvény.

Poisson regresszió : Y természetes szám értékű változó

$Y \sim \text{Poisson}(\vartheta)$ eloszlású, és $\log(\vartheta)$ közelítése $a + b_1X_1 + b_2X_2 + \dots + b_JX_J$ alakban.

Az **általánosított lineáris modell** segítségével minden eddig bemutatott modell paraméterbecslése elvégezhető. Az utolsó két példa kivételével mind egy speciális link függvényéhez, a $\text{logit}(t) = \log\left(\frac{t}{1-t}\right)$ függvényhez kapcsolódnak.

Bináris logisztikus regresszió

kimeneti változó:

studbin (bináris változó, az ISSP SocIneq v10 változójából létrehozva)

magyarázó változók: rincome és urb

People study to earn a lot of money

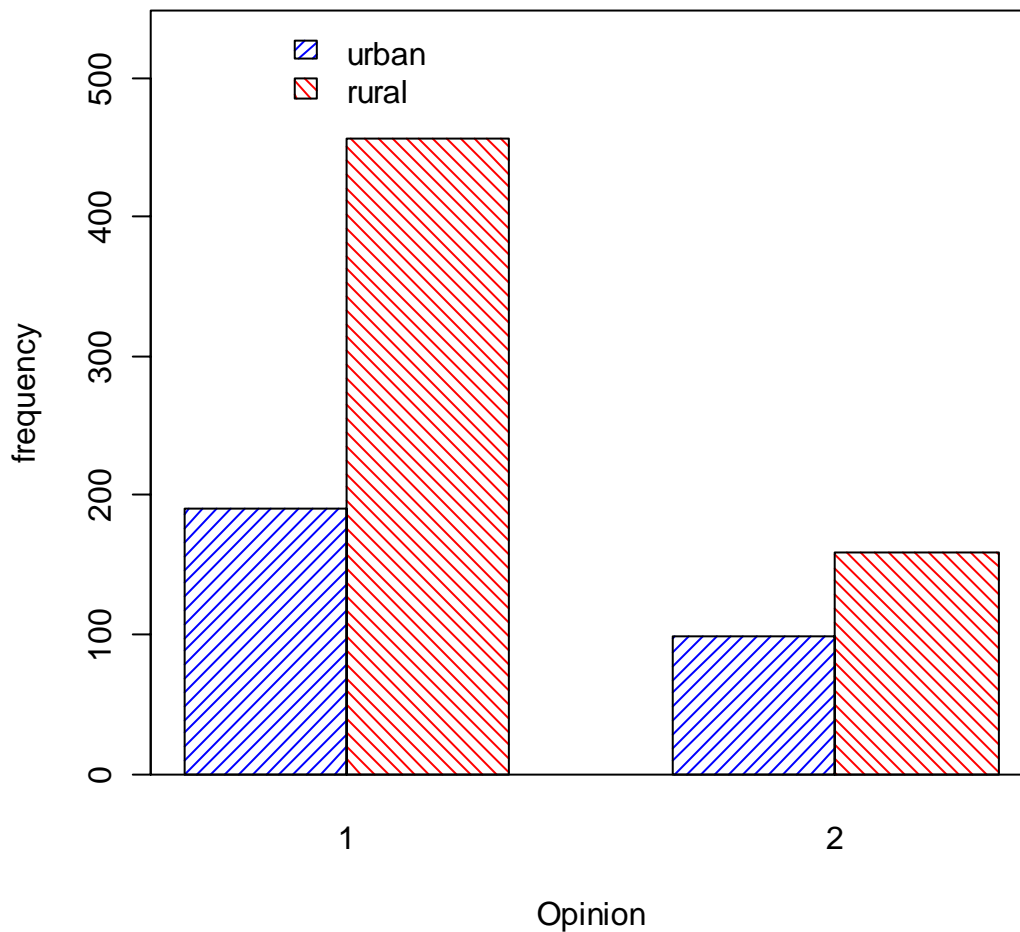
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly agree	220	19.1	19.6	19.6
	2 Agree	576	50.1	51.4	71.1
	3 Neither agree nor disagree	182	15.8	16.2	87.3
	4 Disagree	142	12.3	12.7	100.0
	Total	1120	97.4	100.0	
Missing	8 Cant choose	19	1.7		
	9 No answer	11	1.0		
	Total	30	2.6		
Total		1150	100.0		

studbin

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	796	69.2	71.1	71.1
	2	324	28.2	28.9	100.0
	Total	1120	97.4	100.0	
Missing	System	30	2.6		
Total		1150	100.0		

studbin		B	S.E.	Wald	df	Sig.
Step 1	rincom	.003	.008	.103	1	.749
	urb	-.403	.155	6.766	1	.009
	Constant	-.301	.311	.939	1	.333

People study to earn a lot of money

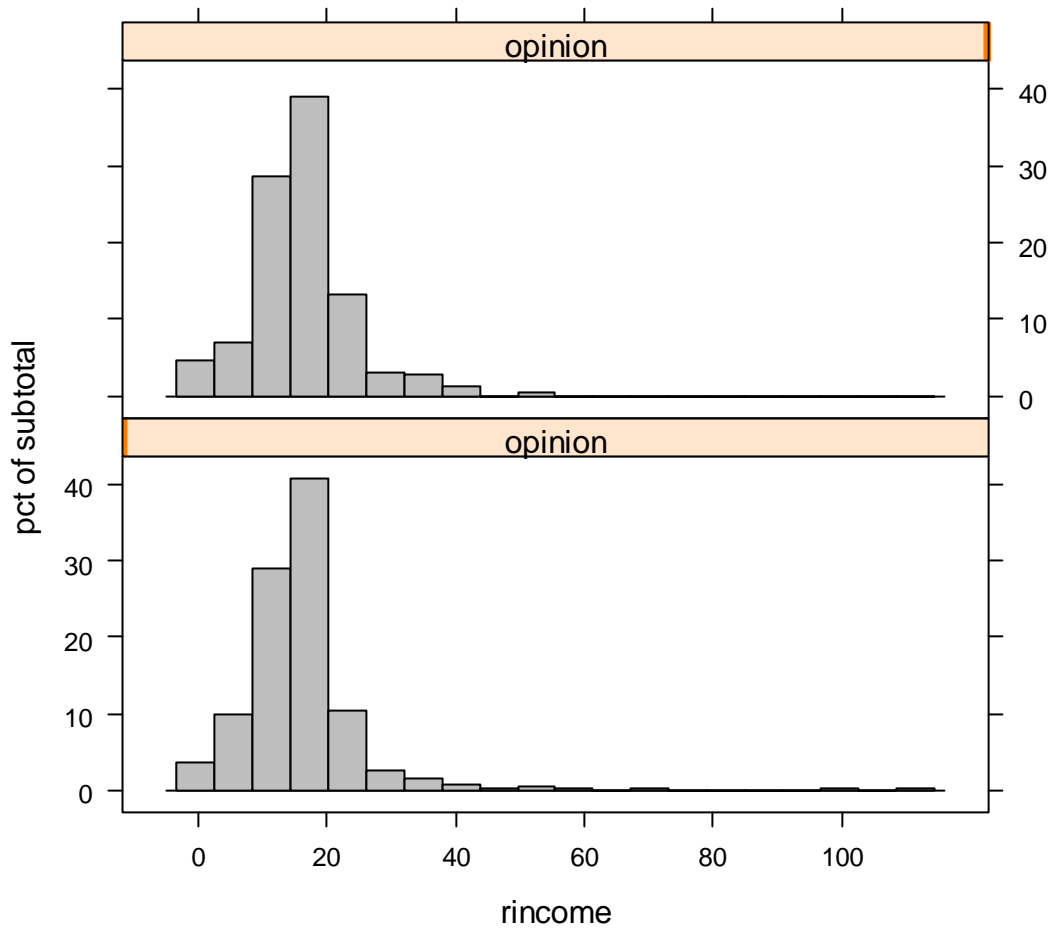


A feltett kérdésre a választ két kimenetelre vontuk össze opt.1 (egyetértő) illetve opt.2 (tagadó). Ez a hisztogram a lakhely típusa szerinti bontásban ábrázolja a válasz gyakoriságokat.

	opt.1	opt.2	össz
URBAN	190	99	289
RURAL	457	159	616

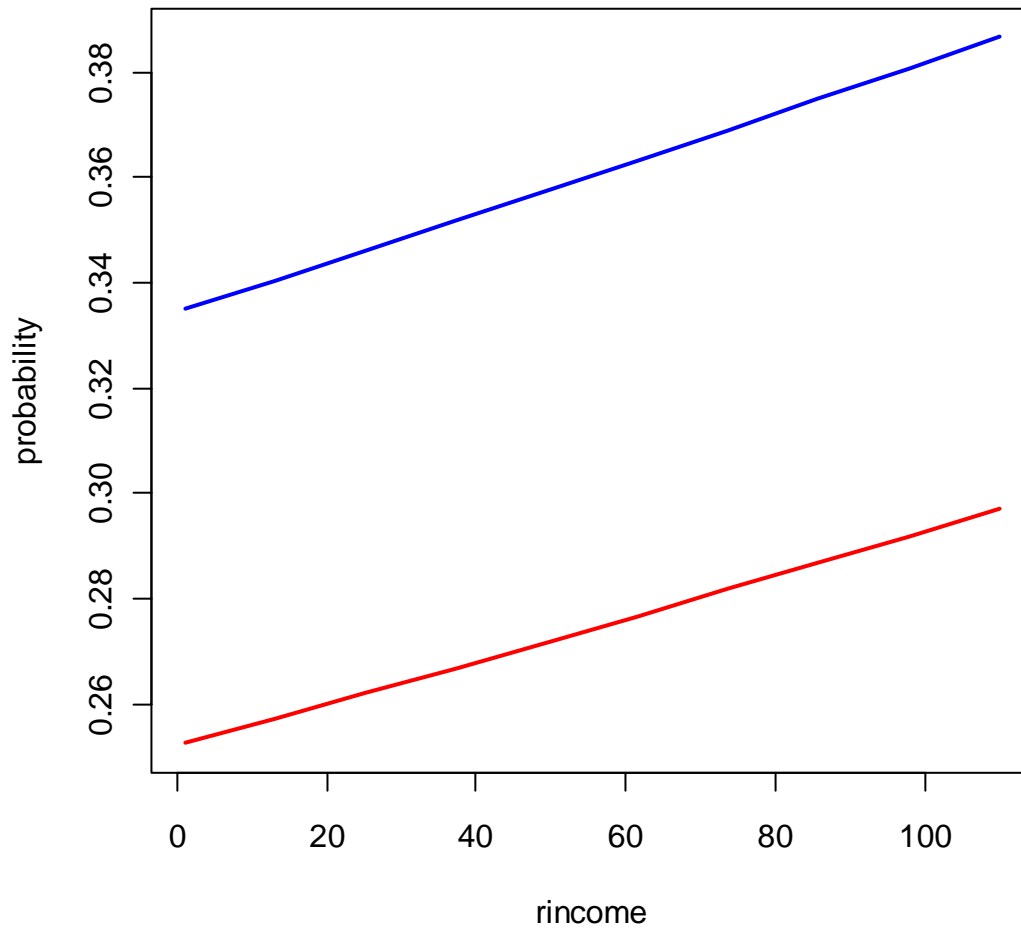
	opt.1	opt.2	össz
URBAN	65.7%	34.3%	100.0%
RURAL	74.2%	25.8%	100.0%

People study to earn a lot of money



A válaszolók jövedelemeloszlása: opt.1 (alsó hisztogram) és opt.2 (felső)

People study to earn a lot of money: opt.2



	opt.1	opt.2	össz
URBAN	65.7%	34.3%	100.0%
RURAL	74.2%	25.8%	100.0%

A bináris logisztikus regressziós modell esély függvénye:
 ábrázoljuk annak a valószínűségét, hogy egy ember a binarizált V10 kérdésre az opt 2. (tagadó) választ adja. A valószínűség a megkérdezett jövedelmének és lakhelyének függvénye.

Multinomiális logisztikus regresszió

a kimeneti változó: ISSP SocIneq v4 változója (öt kategóriás),

magyarázó változók: rincome és urb

Importance: coming from a wealthy famil ^a		B	Std. Error	Wald	df	Sig.
2 Very important	Intercept	.788	.499	2.496	1	.114
	rincom	.073	.033	4.766	1	.029
	[urb=1]	-.433	.493	.771	1	.380
	[urb=2]	0 ^b	.	.	0	.
3 Fairly important	Intercept	1.666	.480	12.058	1	.001
	rincom	.068	.032	4.427	1	.035
	[urb=1]	.211	.461	.210	1	.647
	[urb=2]	0 ^b	.	.	0	.
4 Not very important	Intercept	1.444	.485	8.870	1	.003
	rincom	.059	.033	3.222	1	.073
	[urb=1]	.299	.465	.413	1	.520
	[urb=2]	0 ^b	.	.	0	.
5 Not important at all	Intercept	.437	.530	.680	1	.410
	rincom	.056	.035	2.548	1	.110
	[urb=1]	-.279	.520	.287	1	.592
	[urb=2]	0 ^b	.	.	0	.

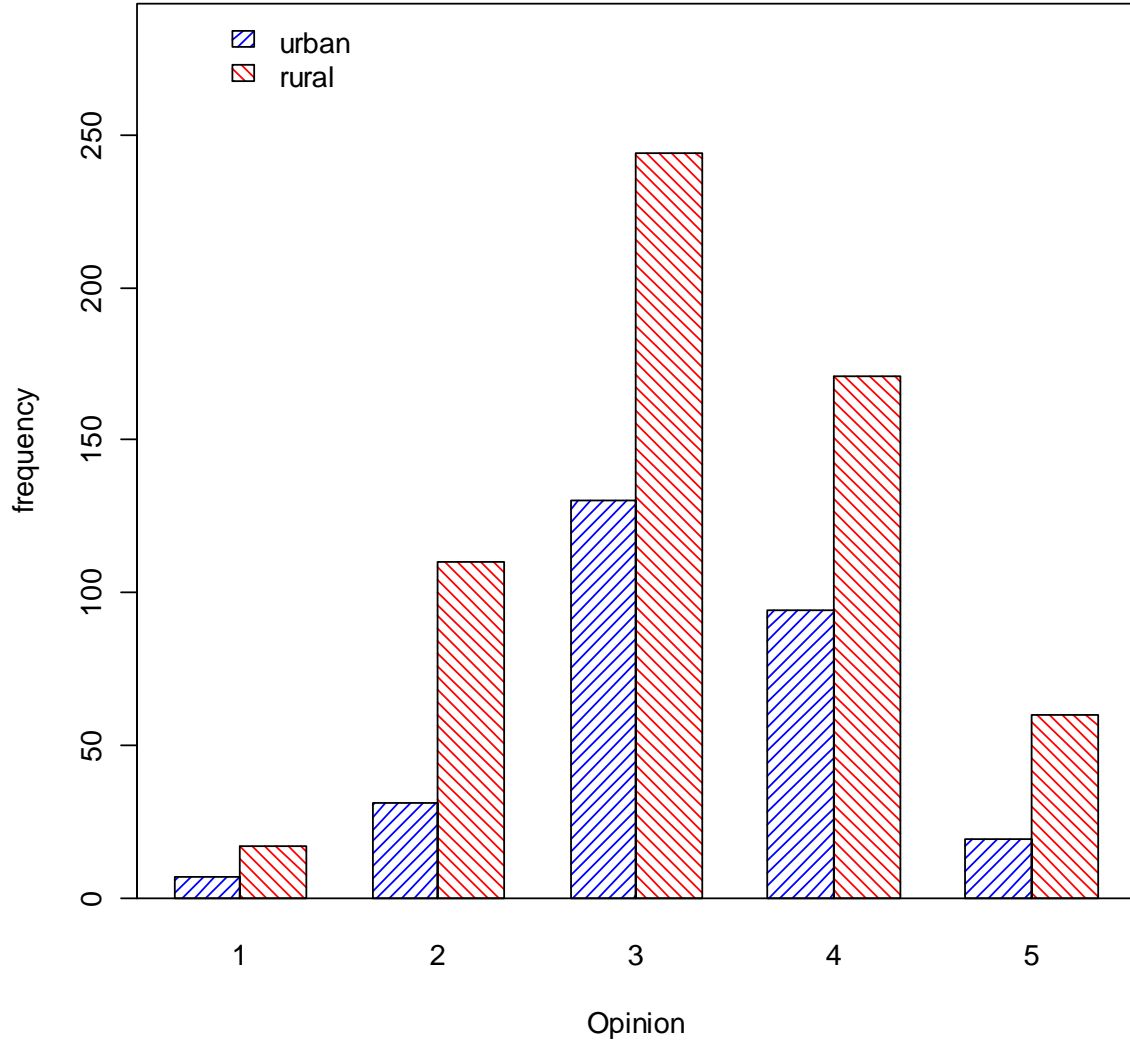
a. The reference category is: 1 Essential.

b. This parameter is set to zero because it is redundant.

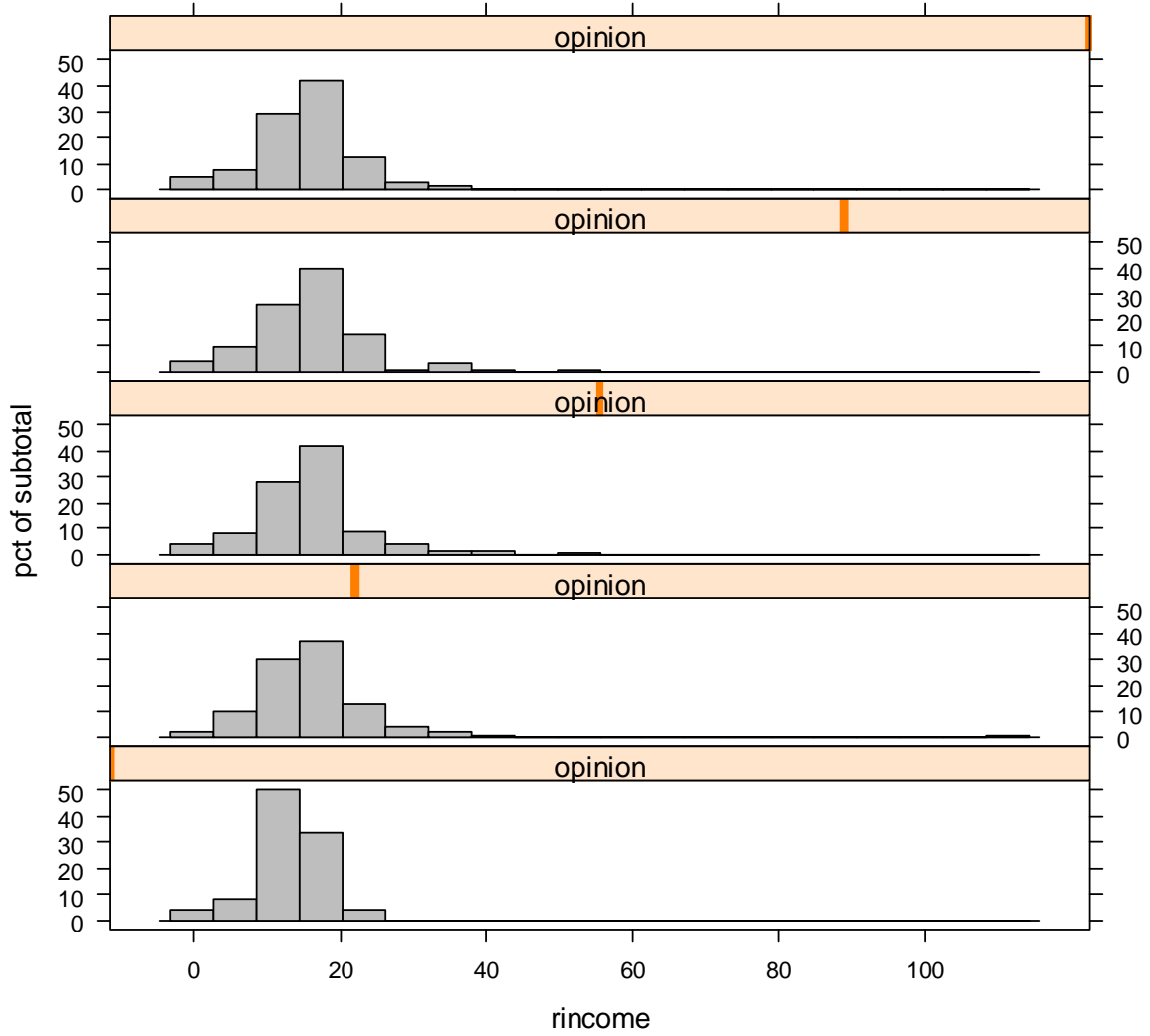
Importance: coming from a wealthy famil

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Essential	36	3.1	3.3	3,3
	2 Very important	171	14.9	15.7	19,0
	3 Fairly important	453	39.4	41.5	60,4
	4 Not very important	333	29.0	30.5	90,9
	5 Not important at all	99	8.6	9.1	100,0
	Total	1092	95.0	100.0	
Missing	8 Cant choose	22	1.9		
	9 No answer	36	3.1		
	Total	58	5.0		
Total		1150	100.0		

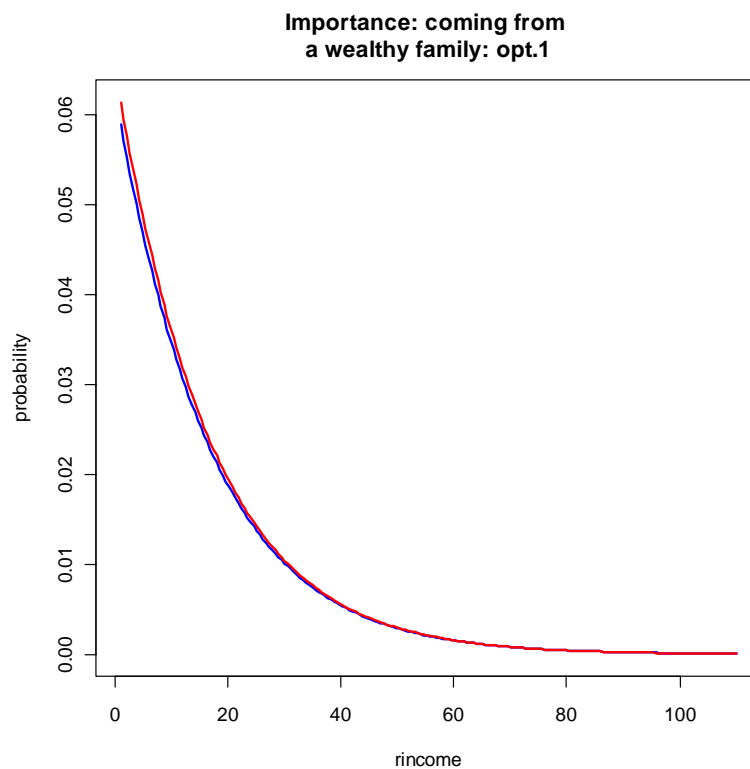
Importance: coming from a wealthy family



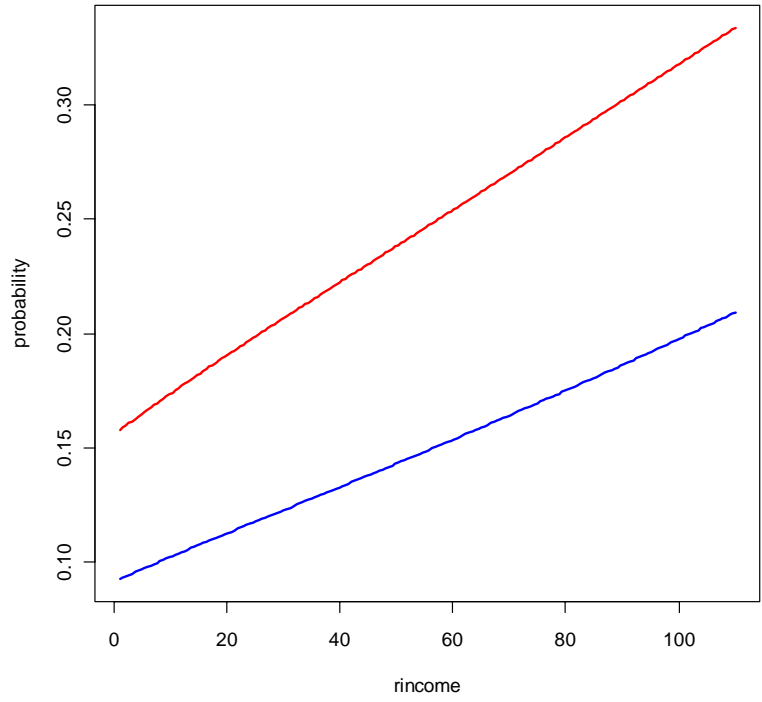
Importance: coming from a wealthy family



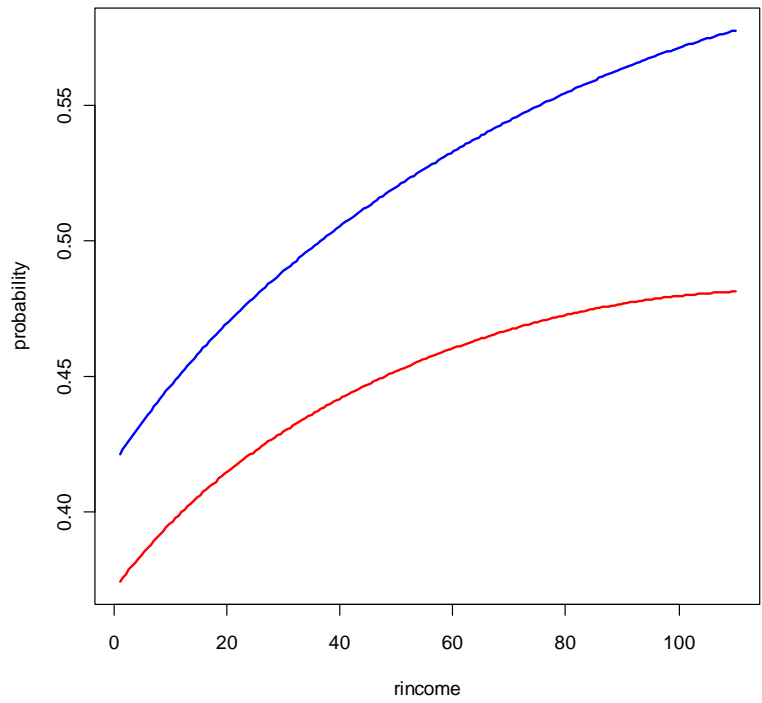
Az eddigi ábrákból azt látjuk, hogy a jövedelemeloszlásban jól látható különbségek vannak a válasz osztályok között, viszont a lakhely típusa szerinti különbségek nem szembeötlők. A Multinomiális logisztikus modell paramétereinek szignifikancia szintjei alapján hasonló következtetésre jutunk. Most megnézzük a modell esély függvényeit (az egyes válasz opciók valószínűségét a válaszadó jövedelme és lakhely típusa függvényében).



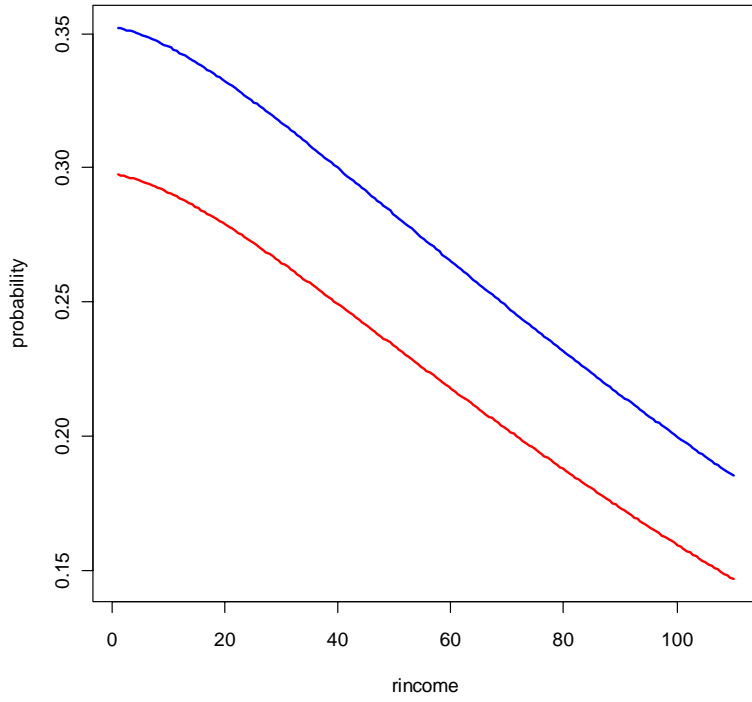
Importance: coming from a wealthy family: opt.2



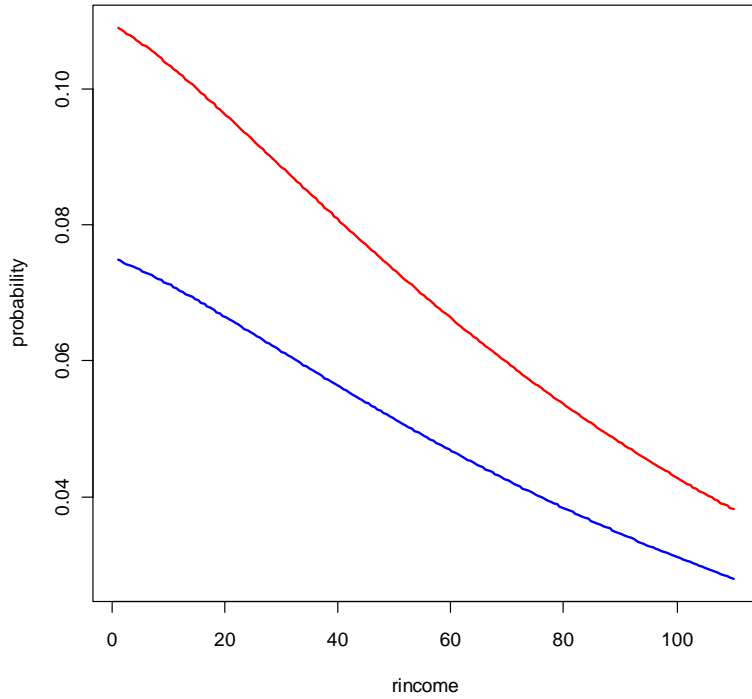
Importance: coming from a wealthy family: opt.3



Importance: coming from a wealthy family: opt.4



Importance: coming from a wealthy family: opt.5



Multinomiális logisztikus regresszió

a kimeneti változó: ISSP SocIneq v10 változója (négy kategóriára összevonva),

magyarázó változók: rincome és urb

People study to earn a lot of money ^a		B	Std. Error	Wald	df	Sig.
2 Agree	Intercept	.716	.197	13.274	1	.000
	rincom	.010	.011	.873	1	.350
	[urb=1]	.238	.198	1.452	1	.228
	[urb=2]	0 ^b	.	.	0	.
3 Neither agree nor disagree	Intercept	-.622	.243	6.523	1	.011
	rincom	.012	.013	.857	1	.355
	[urb=1]	.703	.238	8.721	1	.003
	[urb=2]	0 ^b	.	.	0	.
4 Disagree	Intercept	-.771	.269	8.186	1	.004
	rincom	.008	.015	.287	1	.592
	[urb=1]	.395	.266	2.209	1	.137
	[urb=2]	0 ^b	.	.	0	.

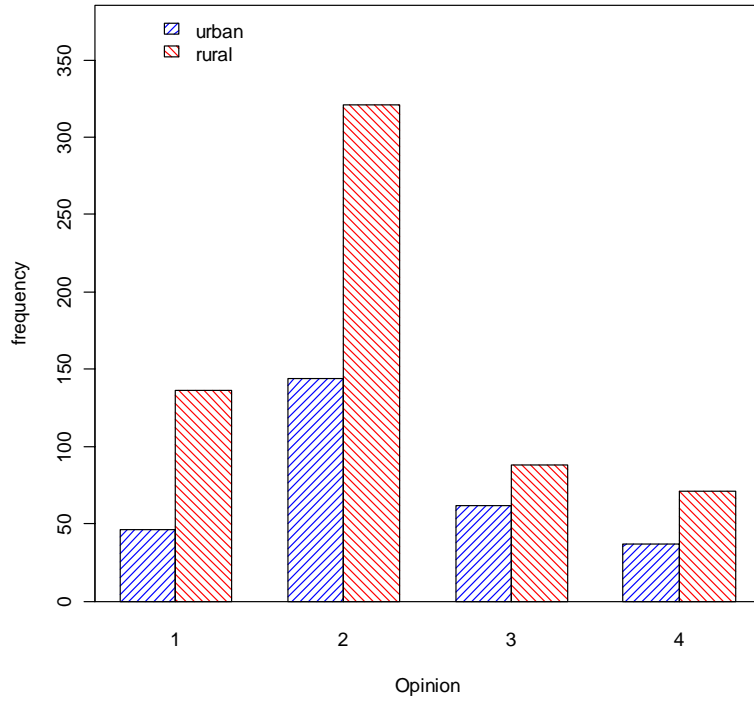
a. The reference category is: 1 Strongly agree.

b. This parameter is set to zero because it is redundant.

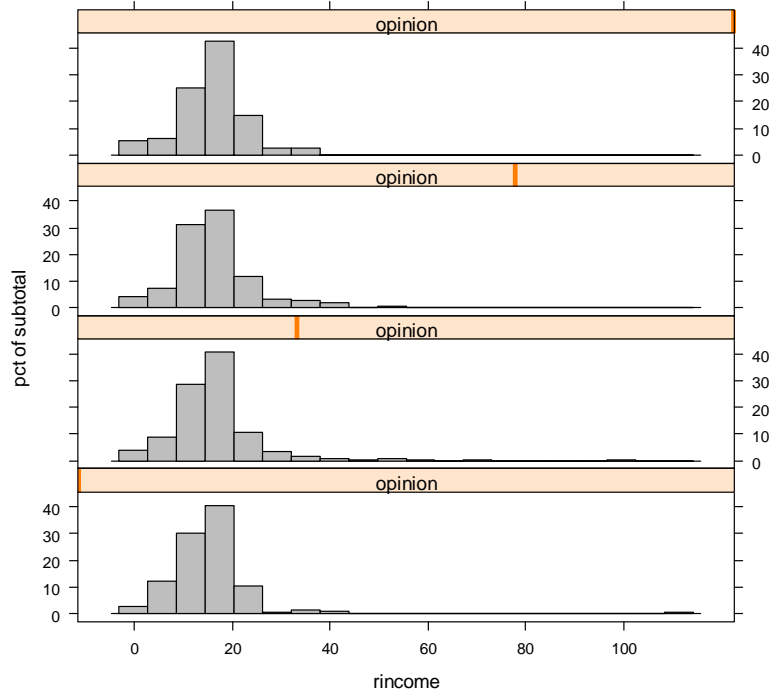
People study to earn a lot of money

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly agree	220	19.1	19.6	19,6
	2 Agree	576	50.1	51.4	71,1
	3 Neither agree nor disagree	182	15.8	16.2	87,3
	4 Disagree	142	12.3	12.7	100,0
	Total	1120	97.4	100.0	
Missing	8 Cant choose	19	1.7		
	9 No answer	11	1.0		
	Total	30	2.6		
Total		1150	100.0		

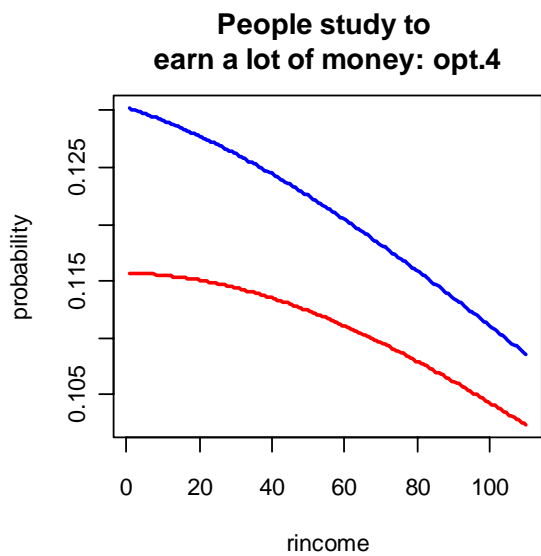
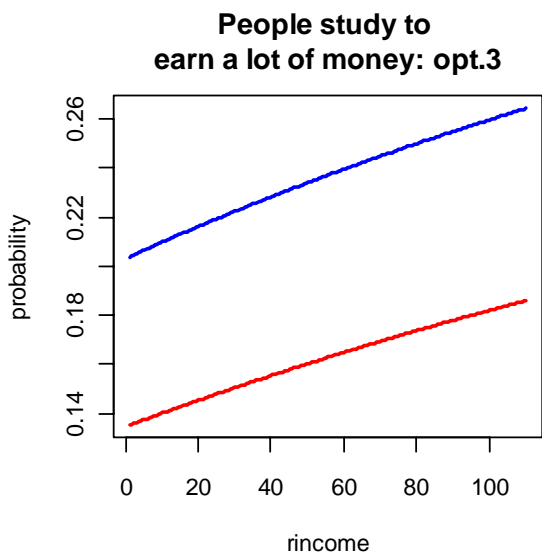
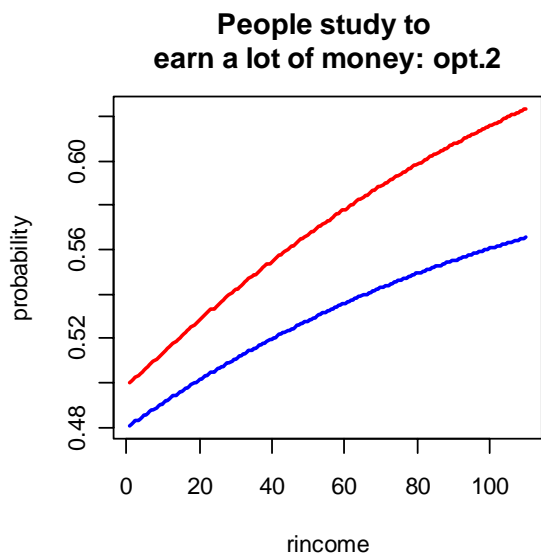
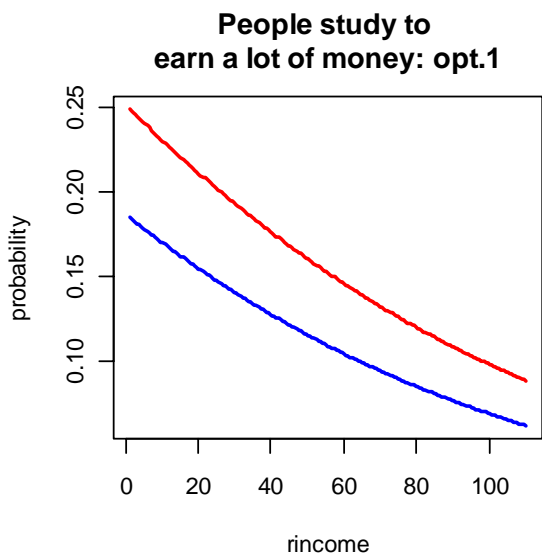
People study to earn a lot of money



People study to earn a lot of money



A fenti ábrákból nem sok különbséget látunk a jövedelemeloszlásban, de a lakhely típusa szerinti különbségek szembeötlők. A Multinomiális logisztikus modell paramétereinek szignifikancia szintjei alapján hasonló következtetésre jutunk. Most megnézzük a modell esély függvényeit (az egyes válasz opciók valószínűségét a válaszadó jövedelme és lakhely típusa függvényében).



Multinomiális logisztikus regresszió

a kimeneti változó: ISSP SocIneq v34 változója (négy kategóriára összevonva),

magyarázó változók: rincome és urb

Differences in income are too large ^a		B	Std. Error	Wald	df	Sig.
2 Agree	Intercept	-.276	.205	1.811	1	.178
	rincom	.033	.013	6.467	1	.011
	[urb=1]	.431	.185	5.414	1	.020
	[urb=2]	0 ^b	.	.	0	.
3 Neither agree nor disagree	Intercept	-1.642	.257	40.791	1	.000
	rincom	.062	.015	17.118	1	.000
	[urb=1]	.868	.213	16.583	1	.000
	[urb=2]	0 ^b	.	.	0	.
4 Disagree	Intercept	-3.425	.328	108.716	1	.000
	rincom	.125	.017	56.458	1	.000
	[urb=1]	.855	.263	10.524	1	.001
	[urb=2]	0 ^b	.	.	0	.

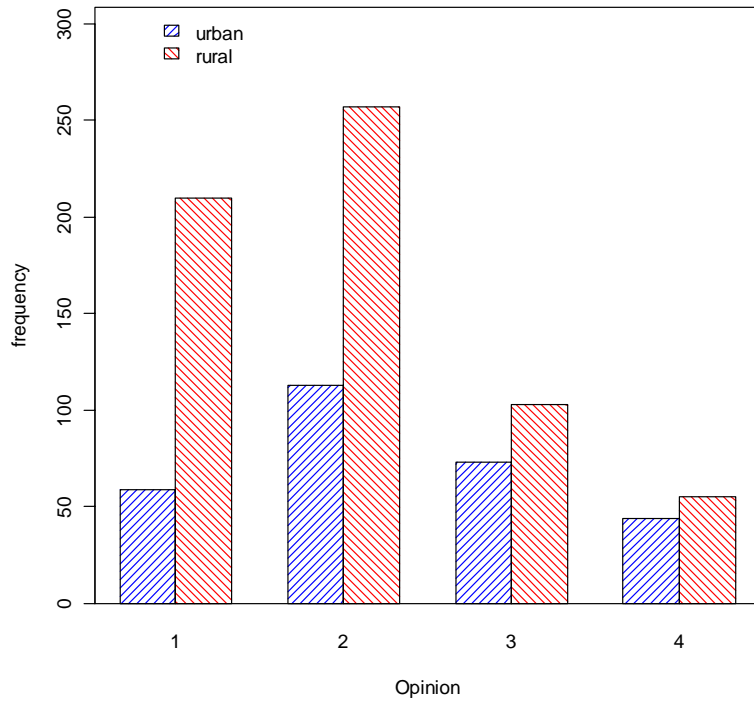
a. The reference category is: 1 Strongly agree.

b. This parameter is set to zero because it is redundant.

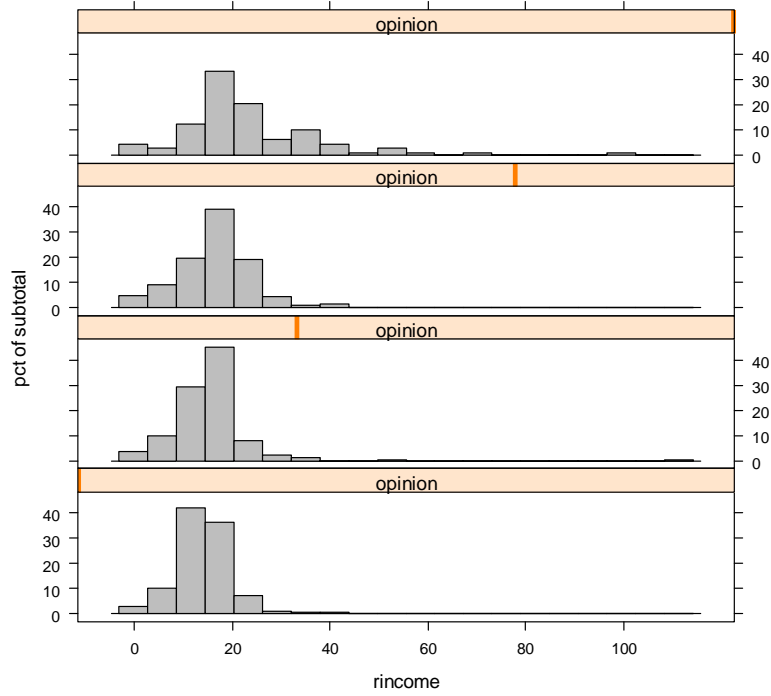
Differences in income are too large

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Strongly agree	331	28.8	29.2	29,2
	2 Agree	474	41.2	41.9	71,1
	3 Neither agree nor disagree	205	17.8	18.1	89,2
	4 Disagree	122	10.6	10.8	100,0
	Total	1132	98.4	100.0	
Missing	8 Cant choose	14	1.2		
	9 No answer	4	.3		
	Total	18	1.6		
Total		1150	100.0		

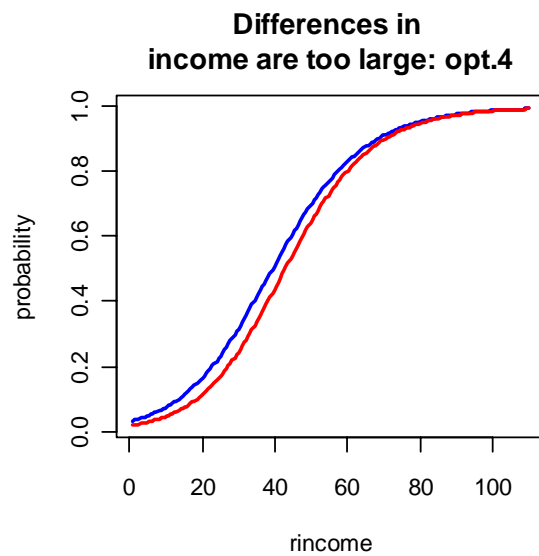
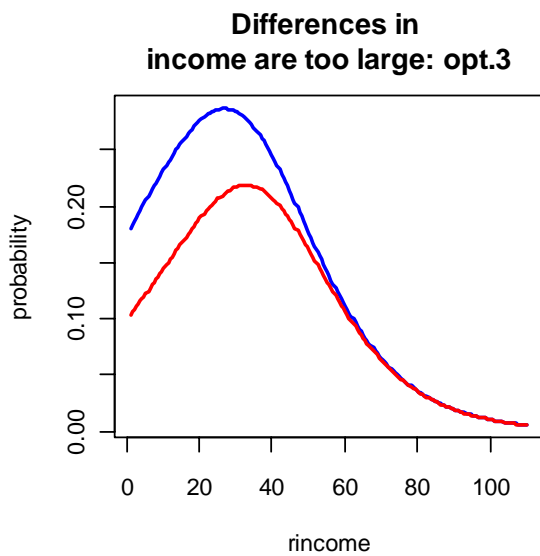
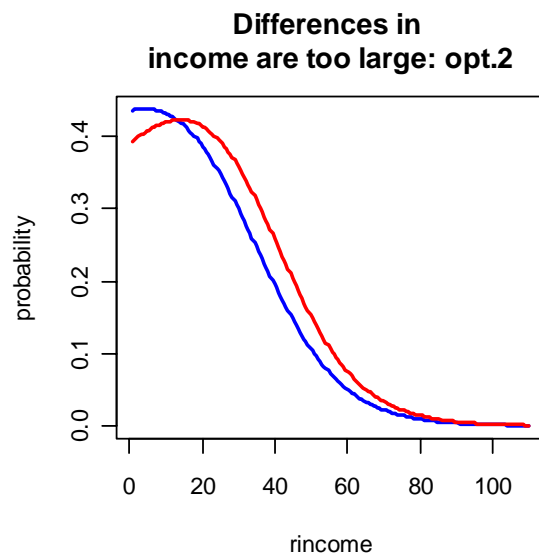
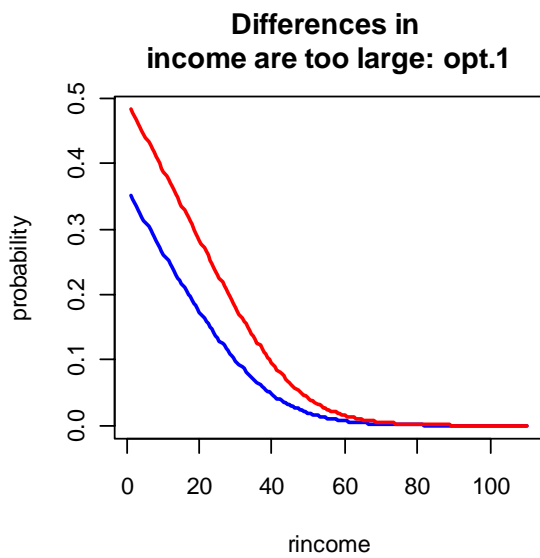
Differences in income are too large



Differences in income are too large



A fenti ábrákból jelentős különbségeket látunk a jövedelemeloszlásban, és a lakhely típusa szerinti különbségek is szembeötlők. A Multinomiális logisztikus modell paramétereinek szignifikancia szintjei alapján hasonló következtetésre jutunk. Most megnézzük a modell esély függvényeit (az egyes válasz opciók valószínűségét a válaszadó jövedelme és lakhely típusa függvényében).



Ordinális logisztikus regresszió

a kimeneti változó: ISSP SocIneq v34 változója (öt kategóriás),

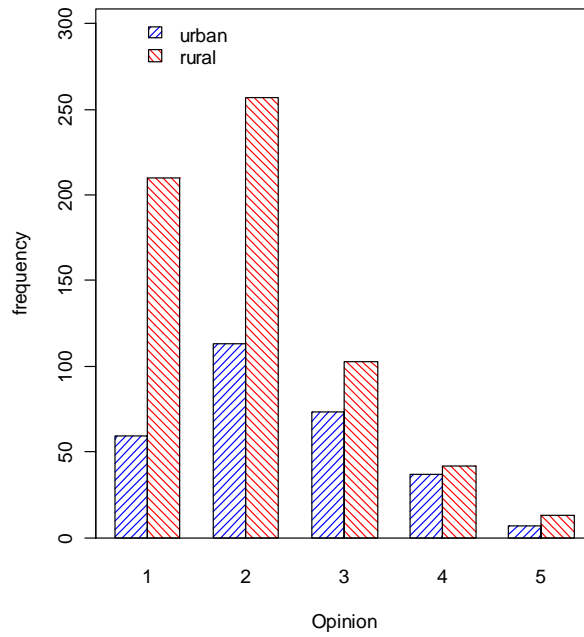
magyarázó változók: rincome és urb.

Differences in income are too large		Estimate	Std. Error	Wald	df	Sig.
Threshold	[v34 = 1]	.227	.140	2.618	1	.106
	[v34 = 2]	2.053	.154	177.061	1	.000
	[v34 = 3]	3.435	.185	345.111	1	.000
	[v34 = 4]	5.247	.290	326.773	1	.000
Location	rincom	.063	.008	63.855	1	.000
	[urb=1]	.546	.131	17.254	1	.000
	[urb=2]	0 ^a	.	.	0	.

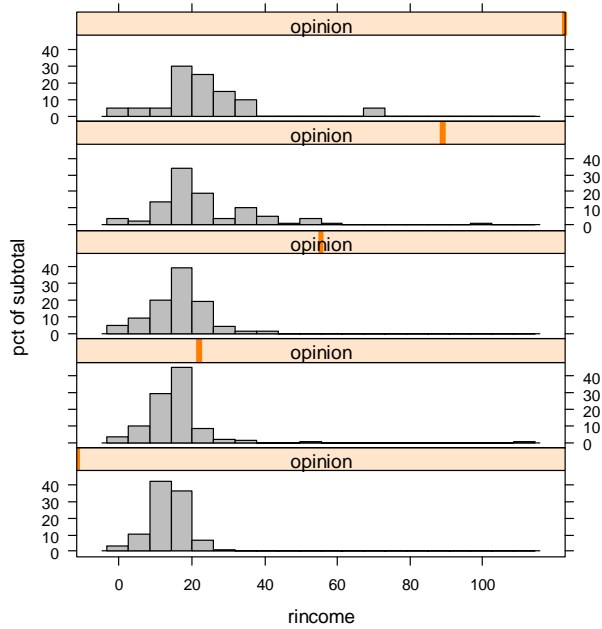
Link function: Logit.

a. This parameter is set to zero because it is redundant.

Differences in income are too large



Differences in income are too large

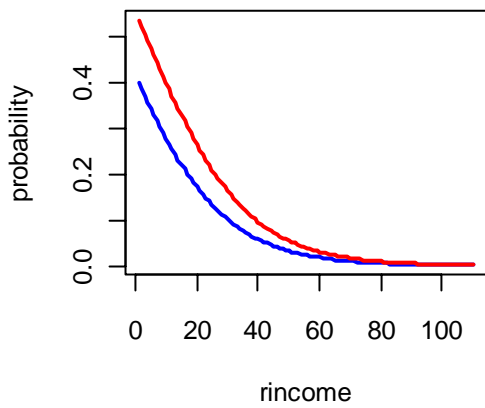


Ordinális logisztikus regresszió

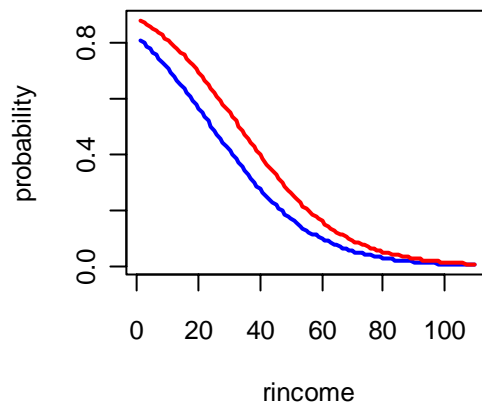
a kimeneti változó: ISSP SocIneq v34 változója (öt kategóriás),

magyarázó változók: rincome és urb

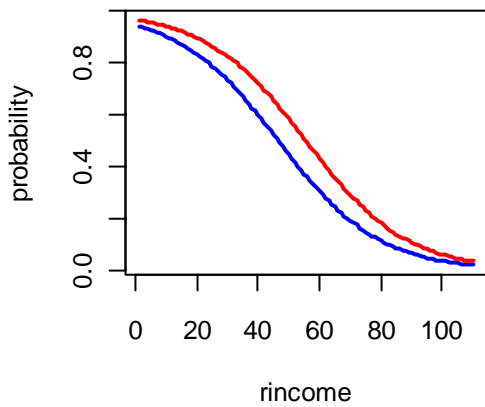
**Differences in
income are too large: thresh.1**



**Differences in
income are too large: thresh.2**



**Differences in
income are too large: thresh.3**



**Differences in
income are too large: thresh.4**

