


BINARY LOGISTIC REGRESSION,  
RANDOM FOREST



Article

# Consumer Evaluation of the Role of Functional Food Products in Disease Prevention and the Characteristics of Target Groups

Brigitta Plasek <sup>1</sup>, Zoltán Lakner <sup>1</sup>, Gyula Kasza <sup>2</sup> and Ágoston Temesi <sup>1,\*</sup> 

<sup>1</sup> Department of Food Economics, Faculty of Food Science, Szent István University, Villányi str. 29-43, 1118 Budapest, Hungary; plasek.brigitta@etk.szie.hu (B.P.); lakner.zoltan@etk.szie.hu (Z.L.)

<sup>2</sup> Directorate for Food Safety Risk Assessment, National Food Chain Safety Office, Keleti Károly str. 24, 1024 Budapest, Hungary; kaszagy@nebih.gov.hu

\* Correspondence: temesi.agoston@etk.szie.hu; Tel.: +36-1-305-7100 (ext. 6178)

Received: 2 December 2019; Accepted: 23 December 2019; Published: 26 December 2019



**Abstract:** Our research explores the methods consumers would consider using in order to prevent non-communicable diseases, with consuming functional food products being one of these methods. Previous research has pointed out the importance of missing information such as which diseases worry consumers and what they would use to avoid them. We conducted a personal survey with 1027 people in Hungary about 13 diseases and four prevention methods. We analyzed the results with descriptive statistical methods, binary logistic regression, and random forest. According to our results, the highest proportion of worried respondents think it is justified to consume functional foods for the prevention of digestive problems, a weakened immune system, and high cholesterol level. Our results help to characterize the target group for these three diseases. Completed education plays a key role in choosing functional foods to prevent a weakened immune system. Those with tertiary education are the most likely to choose this prevention method. With the other two diseases, age played a crucial role. All age groups over 36 would be more likely to choose functional foods to prevent digestive problems, whereas in case of high cholesterol level, the 36–50 age group would be more likely to apply this method.

**Keywords:** consumer research; willingness to purchase; non-communicable diseases; binary logistic regression; random forest

# Binary logistic regression

**regresszió egy speciális formája**, amely nem egy metrikus függő mérőszám, hanem egy **bináris (kétcsoportos) kategorikus változó előrejelzésére** és magyarázatára szolgál

A regressziós mérőszám egyetlen többváltozós kapcsolatot reprezentál, az egyes prediktív változók relatív hatását jelző regressziószerű együtthatókkal.

Két okból részesíthető előnyben:

- Robosztusabb, mint a diszkriminancia-analízis, melynek szigorúbb feltételei vannak az alkalmazhatósága érdekében
- Hasonló a multiple-regression módszerhez - **egyszerű statisztikai tesztekkel, hasonló megközelítésekkel rendelkezik**

# A bináris logisztikus regresszió célja

A logisztikus regresszió a diszkriminancia analízissel azonos az általa kezelhető alapvető célok tekintetében. A logisztikus regresszió két kutatási cél megvalósítására alkalmas a legjobban:

- Azon **független változók azonosítása**, amelyek hatással vannak a függő változó csoporttagságára.
- A logisztikus modellen alapuló osztályozási rendszer létrehozása a csoporttagság meghatározására.

# A bináris logisztikus regresszió alkalmazása

Kódolás: 1-0

Teljes minta mérete: 400-nál nagyobb

Kategórián belüli minta mérete: legalább 10 adat az egyes megfigyelésekre

# Saját példánk

Teljes minta: 1027 fő

**Table 1.** Grouping of respondents based on variables ( $n = 1027$ ).

Characteristics of Survey Respondents		%
Gender	Female	53.5%
	Male	46.5%
Age brackets	<25	47.7%
	25–35	18.8%
	36–50	14.2%
	50<	19.3%
Place of living	Budapest	39%
	Another city	48.6%
	Village	12.4%
Characteristics of Survey Respondents		%
Highest accomplished qualification	Primary school/vocational school	15.1%
	Secondary technical school/secondary school diploma	40.0%
	Tertiary education	44.9%
Subjective estimation of income position	Below average	35.5%
	Average	53.6%
	Above average	10.9%
Presence of a child under 18 in the household	Yes	79.5%
	No	20.5%
Overweight or underweight status of the respondent (based on BMI) *	Undernourished	6.8%
	Normal	61.3%
	Obese/overweight	31.9%
It is important for me to do something for my health	Important	88.9%
	Less important/not important	11.1%
I pay attention to healthy eating	Important	72.8%
	Less important/not important	27.2%
Grocery shopper	Respondent	25.7%
	Another person	22.7%
	Shared	51.6%

\* Based on World Health Organization (WHO) body mass index (BMI) [24].

# Bináris logisztikus regresszió eredményeinek értelmezése

Pozitív  $\beta$ : Ha a  $\beta$  pozitív, akkor  $e^\beta > 1$ , ami azt jelenti, hogy az adott független változó növekedése **növeli az esemény bekövetkezésének esélyét.**

Negatív  $\beta$ : Ha a  $\beta$  negatív, akkor  $e^\beta < 1$ , ami azt jelenti, hogy az adott független változó növekedése **csökkenti az esemény bekövetkezésének esélyét.**

$\beta=0$ : Ha  $\beta=0$ , akkor  $e^\beta=1$ , ami azt jelenti, hogy az adott független változó nincs hatással az esemény bekövetkezésének esélyére.

# Bináris logisztikus regressziós eredményeink

**Table 4.** Results of the binary logistic regression.

Number of Cases	Characteristics of the Respondent (Independent Variable)	Level of Attribute	Would Buy Functional Food to Avoid the Specified Disease (Dependent Variable)						
			B	SE	Wald	Sig.	Exp(B)	95% CI for EXP (B)	
								Lower	Upper
High cholesterol level									
289	Constant		-0.752	0.189	15.773	0.000	0.471		
	Age	36–50	0.919	0.346	7.056	0.008	2.508	1.273	4.942
Cancer-related diseases									
498	Constant		-0.745	0.371	4.024	0.045	0.475		
	Gender	Men	-0.469	0.208	5.108	0.024	0.626	0.416	0.940
	Level of education	College/university	0.992	0.356	7.749	0.005	2.697	1.341	5.423
	Grocery shopper	Another person	-0.862	0.296	8.498	0.004	0.422	0.236	0.754
Skin diseases									
293	Constant		-0.455	0.251	3.301	0.069	0.634		
	Grocery shopper	Another person	-0.743	0.362	4.205	0.040	0.476	0.234	0.968
		Shared		-0.797	0.321	6.170	0.013	0.451	0.240
Memory lapse/concentration problems									
284	Constant		-1.871	0.483	15.006	0.000	0.154		
	Age	25–35	0.875	0.359	5.933	0.015	2.399	1.186	4.852
	Level of education	College/university	1.004	0.494	4.132	0.042	2.728	1.037	7.181
Digestion problems									
291	Constant		-1.103	0.396	7.758	0.005	0.332		
	Age	36–50	1.055	0.362	8.478	0.004	2.872	1.412	5.844
		50+		0.940	0.332	8.013	0.005	2.559	1.335



# Random forest

- egy **gépi tanulási módszer**, felülmúlja a logisztikus regressziót, ha a prediktív változók között nem lineáris jellemzők és összetett kölcsönhatások állnak fenn.
- **osztályozási és regressziós feladatokra egyaránt alkalmaznak**
- **"erdő"** sok döntési fából
- Előnye a döntési fákhhoz képest: **döntési fa sokkal érzékenyebb** a „training data”-ra
- Predikcióra alkalmas
- “OOB Estimate of Error Rate” alkalmazása – megmutatja, hogy az algoritmus mennyire jól teljesített az OOB mintákon, amelyek nem kerültek kiválasztásra a bootstrap folyamat során, és így nem lettek felhasználva a fa tanításához – pl. 10% - algoritmus átlag 10% hibát követ el az előrejelzéseknél
- 70% of the total sample was used for training and 30% was used to validate the models
- MeanDecreaseGini értékek - segít megérteni, hogy mely változók mennyire fontosak a modell szempontjából

# Eredményeink a random forest módszerrel - részlet

**Table 5.** Results obtained with the help of random forest.

Disease	OOB Estimate of Error Rate	Relative Importance	
			MeanDecreaseGini
Cardiovascular disease	33.66%	age	23.52
		grocery shopper	19.92
		education	19.40
High cholesterol level	40.89%	age	27.12
		education	22.02
		subjective estimation of income position	21.54
Cancer	38.11%	age	21.88
		education	20.85
		place of living	19.60
Mood disorders/Apnoea	26.15%	age	20.15
		place of living	17.57
		education	17.19
Respiratory diseases	31.15%	age	20.74
		education	18.90
		place of living	18.22
Skin diseases, eczema	33.38%	age	24.67
		education	19.89
		place of living	19.89

# Trianguláció

	Concerned, Ready to Make a Financial Sacrifice to Prevent It		Form of Financial Sacrifice			
			Medicine	Dietary Supplement	Functional Food	Lifestyle Changes
	<i>n</i>	%	%	%	%	%
Digestive problems	327	100%	13%	29%	43%	55%
Weakened immune system	473	100%	22%	36%	39%	56%
High cholesterol level	322	100%	12%	19%	37%	64%
Diabetes	418	100%	36%	17%	35%	59%
Cancer	552	100%	47%	19%	31%	48%

Cancer-related diseases									
498	Constant		-0.745	0.371	4.024	0.045	0.475		
	Gender	Men	-0.469	0.208	5.108	0.024	0.626	0.416	0.940
	Level of education	College/university	0.992	0.356	7.749	0.005	2.697	1.341	5.423
	Grocery shopper	Another person	-0.862	0.296	8.498	0.004	0.422	0.236	0.754

Cancer	38.11%	age	21.88
		education	20.85
		place of living	19.60

KÖSZÖNÖM A FIGYELMET!

