

## Оцінка впливу факторних ознак за допомогою логістичної регресії (аналіз результатів опитування впливу негативних подій на особистісні почуття та успішність українських студентів 2022-2023 років)

Д. Г. Гаврилов<sup>1</sup>, О.В. Мулик<sup>2</sup>

<sup>1,2</sup>Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, <sup>2</sup>ORCID ID: <https://orcid.org/0000-0003-0710-9693>  
(<sup>1,2</sup>National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine)

### Анотація

Дана стаття має на меті продемонструвати можливості використання платформи MS Excel для опрацювання результатів опитування за допомогою побудови багатофакторної моделі логістичної регресії та мінімізації кількості факторних ознак, пов'язаних з вихідною змінною за допомогою методу покрокового виключення змінних. Також відмічені важливі параметри, що характеризують якість та адекватність моделі - інформаційний критерій Акаїке (AIC), байєсівський інформаційний критерій (BIC) та AUC - площа під кривою операційних характеристик моделі (ROC-curve). З отриманих на трьох моделях результату можна зробити висновок, що дійсно - моральний стан студентів під час російського вторгнення має значний вплив на їх подальші наміри продовжувати навчання та фахову роботу в Україні.

**Ключові слова:** логістична регресія, багатофакторна модель, stepwise method, MS Excel, ROC-curve, AIC, BIC, AUC.

**УДК 004.85**

[dengavrylov@gmail.com](mailto:dengavrylov@gmail.com), [mulyk.olena@gmail.com](mailto:mulyk.olena@gmail.com)

©The Author(s)2023. Licensed under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Cite as: D.H.Gavrylov & O.V.Mulyk (2023). Assessment of factor signs through logistic regression (analysis of survey results of the negative events impact on personal feelings and academic performance of Ukrainian students in 2022-2023). *Mathematics in Modern Technical University*, 2025(1), 51-59. 10.20535/mmtu-2025.1-051

## 1 Вступ

Вже майже п'ять років сфера освіти перебуває в стані вимушеної швидкої трансформації, спричиненої як пандемією, так і воєнним станом. Освітні заклади стали своєрідними аренами боротьби за нормальний життєвий процес та навчання молоді. Прискорена адаптація до нових реалій стала необхідністю, що вимагає впровадження інноваційних методик навчання, перегляду програм та підходів до взаємодії зі студентами. Окрім цього, надання психологічної підтримки стало невід'ємною складовою роботи освітніх закладів, оскільки молодь потребує підтримки та розуміння у цей складний період. Важливою умовою успішної організації навчального процесу стає розуміння які саме фактори впливають на бажання студентів продовжувати навчання та фахову роботу в Україні, а які можуть вплинути на рішення поїхати шукати долі до інших країн.

## 2 Основна частина

Дане дослідження спрямовано на спробу прогнозування настроїв студентської молоді про подальшу фахову роботу в Україні в сучасному стані війни і проведено за допомогою побудови багатофакторної моделі логістичної регресії на платформі MS Excel (Real Statistics, Logistic and Probit Regression, <https://real-statistics.com/>). Логістична регресія – це керований алгоритм машинного навчання, який виконує завдання двійкової класифікації шляхом прогнозування ймовірності результату, події або спостереження. Модель забезпечує бінарний або дихотомічний результат, обмежений двома можливими результатами: так/ні (0/1). Тобто прогнозований варіант, що майбутній фахівець планує залишитись в Україні визначено як 1, протилежний варіант 0 («Результат» – дихотомічна змінна  $Y$ ). Формула рівняння математичної моделі прогнозування має вигляд:

$$\ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{p(Y=1)}{p(Y=0)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n,$$

де вираз  $\ln\left(\frac{p}{1-p}\right)$  є «logit», для якого рівняння виражає лінійну залежність логарифму відношення шансів  $\frac{p}{1-p}$  незалежних кількісних та дихотомічних змінних  $X_1, X_2 \dots X_n$ . Для розрахунку значень  $\beta_i, i = \overline{1, n}$  на навчальній множині, як правило, використовується метод найбільшої правдоподібності. В процесі обчислення ймовірність  $p = 0,5$  є параметром прийняття рішення і задається як **Cutoff**. Якщо  $p > 0,5$  тоді прогнозується  $Y = 1$ , (**p-Pred**).

Вивчалися результати анонімного опитування у травні 2023 року для студентів 1-3 курсів. Мета - виявлення найбільш впливових факторних ознак (з 19 поставлених запитань, на які відповіді давались в різних шкалах, для дослідження відібрані 11 запитань за шкалою від 1 до 5), що можуть стати визначальними на вибір студентів продовжувати навчання і фахову роботу в Україні.

На явно поставлене запитання «Оцініть свої подальші перспективи, щодо фахової роботи в Україні після закінчення війни з московитами перемогою України» були відібрані відповіді  $Y_i > 3$  (IF(A3>3;1;0)). Таких студентів було 43 з 153. Щоб отримати мінімальний набір факторних ознак, пов'язаних з вихідною змінною було використано метод покрокового виключення змінних (stepwise method) на платформі MS Excel порівнянням значення  $p$  – value, покроково відбирались фактори, для яких  $p < 0,05$ .

	<i>coeff b</i>	<i>s.e.</i>	<i>Wald</i>	<i>p-value</i>	<i>exp(b)</i>	<i>lower</i>	<i>upper</i>
<b>Intercept</b>	-4,574	1,709	7,161	0,007	0,010		
<b>H</b>	0,083	0,293	0,081	0,776	1,087	0,612	1,931
<b>I</b>	-0,266	0,183	2,114	0,146	0,766	0,535	1,097
<b>J</b>	-0,127	0,221	0,332	0,565	0,880	0,571	1,358
<b>K</b>	-0,094	0,185	0,262	0,609	0,910	0,634	1,307
<b>L</b>	-0,168	0,268	0,392	0,531	0,846	0,500	1,429
<b>M</b>	0,066	0,265	0,062	0,803	1,068	0,636	1,795
<b>N</b>	0,321	0,229	1,970	0,160	1,379	0,880	2,159
<b>O</b>	0,141	0,212	0,445	0,505	1,152	0,760	1,746
<b>P</b>	-0,047	0,204	0,053	0,818	0,954	0,639	1,424
<b>Q</b>	-0,035	0,315	0,012	0,912	0,966	0,521	1,792
<b>R</b>	1,371	0,336	16,663	4,5E-05	3,941	2,040	7,614

Рис. 1. Багатофакторна логістична модель регресії MS Excel 1 крок.  
(Stepwise selection based on  $p$  – value)

	coeff b	s.e.	Wald	p-value	exp(b)	lower	upper
Intercept	-4,76188	1,45396	10,7263	0,00106	0,00855		
I	-0,28968	0,17776	2,65577	0,10317	0,7485	0,52831	1,06047
J	-0,15552	0,21317	0,53227	0,46565	0,85597	0,56365	1,2999
L	-0,1133	0,21992	0,26544	0,60641	0,89288	0,58022	1,37401
M	0,07484	0,25707	0,08476	0,77095	1,07771	0,65115	1,78371
N	0,2875	0,21189	1,84095	0,17484	1,33308	0,88003	2,01938
O	0,15088	0,20033	0,56728	0,45134	1,16286	0,78525	1,72205
R	1,38458	0,30209	21,0064	4,6E-06	3,99314	2,20888	7,21867

Рис. 2. Багатофакторна логістична модель регресії MS Excel, кроки 4 та 6.

	coeff b	s.e.	Wald	p-value	exp(b)	lower	upper
Intercept	-4,87868	1,17868	17,1321	3,5E-05	0,00761		
I	-0,32679	0,16555	3,89646	0,04839	0,72124	0,52139	0,99769
N	0,3618	0,18299	3,90917	0,04802	1,43591	1,00315	2,05535
R	1,3961	0,28976	23,2141	1,4E-06	4,03944	2,28915	7,12799

Рис. 3. Багатофакторна логістична модель регресії MS Excel, 8 крок.

На виділених трьох факторних ознаках була побудована логістична модель прогнозування для якої при порівнянні її прогностичних характеристик з моделями побудованими на всіх одинадцяти змінних не виявлено кардинального погіршення. Також, відслідковувались інформаційний критерій Акаїке (AIC) та байєсівський інформаційний критерій (BIC), кінцева модель має бути з найменшими значеннями AIC та BIC.

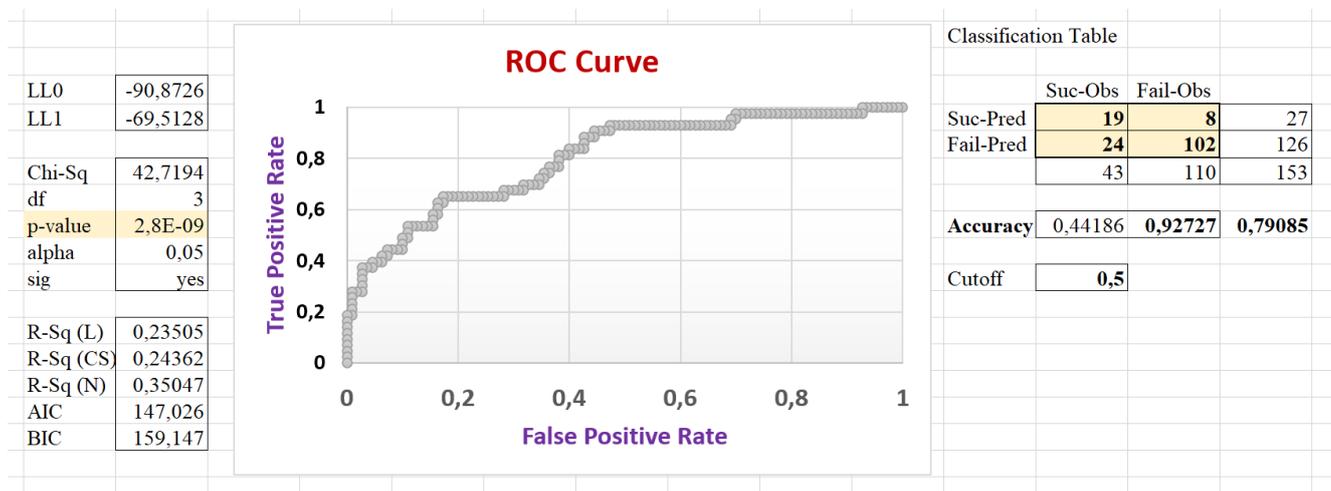


Рис. 4. Трифакторна модель (AUC=0,81).

Для даної моделі AIC=147, BIC=159, та площею під кривою операційних характеристик моделі (ROC-curve) AUC=0,81, це статистично значимо ( $p < 0.05$ ) перевищує 0.5, що є свідченням адекватності побудованої моделі та, за класифікацією, є дуже добрим показником. Розраховані коефіцієнти моделі для трьох факторів (Рис.3) дають наступну формулу прогнозування:

$$\ln\left(\frac{p}{1-p}\right) = -4,88 - 0,327X_1 + 0,361X_2 + 1,396X_3.$$

Потрібно відмітити, що MS Excel виводить чотирипільну таблицю спряженості (Classification Table), яка побудована на основі результатів класифікації в математичній моделі, вона є стандартною і в процесі розрахунків значення в ній не змінюються (для даної моделі, див. Рис.4):

Таблиця 1.

Рішення, прийняте в моделі (тесті)	Фактичний результат		
	Випадок	Не випадок	
Позитивне	TP=19	FP=8	27
Позитивне	FN=24	TN=102	126
	43	110	153

Також, MS Excel розраховує показники: чутливість (Sensitivity) моделі - відношення кількості студентів, що хочуть залишитись в Україні для яких модель дала правильний результат до всіх студентів, відмічених  $Y_i = 1$ :

$$\text{Sensitivity} = \frac{TP}{NP + FN} = \frac{19}{43} = 0,44.$$

Специфічність (Specificity) моделі - кількість студентів, які не бачать перспективи фахової роботи в Україні, що були правильно розпізнані моделлю до всіх позначених як  $Y_i = 0$ :

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{102}{110} = 0,93.$$

Прогностичність позитивного результату (PPV):  $PPV = \frac{TP}{TP+FP} = 0,704$ .

Прогностичність негативного результату (NPV):  $NPV = \frac{TN}{TN+FN} = 0,81$ .

З даної моделі робимо висновки, що найбільш впливові фактори на можливий вибір студентів є: 1) вибір професії, на який вторгнення суттєво не вплинуло;  
2) проблеми у переживанні позитивних емоцій (неможливість відчувати радість або любов);  
3) моральний настрій щодо навчання під час військового стану.

Встановлено, що з підвищенням проблем у переживанні позитивних емоцій та моральному настрої знижується бажання продовжувати навчання та фахову роботу в Україні. Але, якщо фактори вибору професії та емоцій приблизно можуть компенсувати один одного у впливу на відношення шансів (ВШ), то моральний настрої впливає набагато значніше (див. Табл.2)

Таблиця 2.

Показник	Значення коефіцієнту моделі $\beta_i \pm Se$	Рівень значимості відмінності коефіцієнту від 0, p-value	Показник ВШ (95% ВІ)
Intercept	-4,88 ± 1,17	3,49 · E – 05	
I	0,327 ± 0,17	0,048	0,72(0,52 – 0,99)
N	0,361 ± 0,18	0,048	1,44(1,0 – 2,05)
R	1,396 ± 0,29	1,45 · E – 06	4,04(2,9 – 7,13)

Якщо ще раз розглянути результати даного анонімного опитування де було задано три подібних запитання з підтекстом продовження можливості залишатись на навчання та фахову роботу в Україні та розглянути середнє значення за цими відповідями  $Y_i > 3$  ( $IF(A3>3;1;0)$ ) та  $Y_i \geq 3$  ( $IF(A3\geq 3;1;0)$ ). Таких студентів стало 73 з 153 та 98 з 153. Знову було двічі використано метод покрокового виключення змінних (stepwise method) порівнянням значення p – value, покроково відбиралось  $p < 0,05$ . Отримані результати частково повторюють попередні результати, але, що природньо, модель має інші характеристики, див. Рис. 5 та 6.

	<i>coeff b</i>	<i>s.e.</i>	<i>Wald</i>	<i>p-value</i>	<i>exp(b)</i>	<i>lower</i>	<i>upper</i>
Intercept	-0,36392	0,7395	0,24218	0,62263	0,69494		
P	-0,29345	0,13781	4,53437	0,03322	0,74569	0,56919	0,97692
R	0,47694	0,20098	5,63145	0,01764	1,61113	1,08657	2,38893

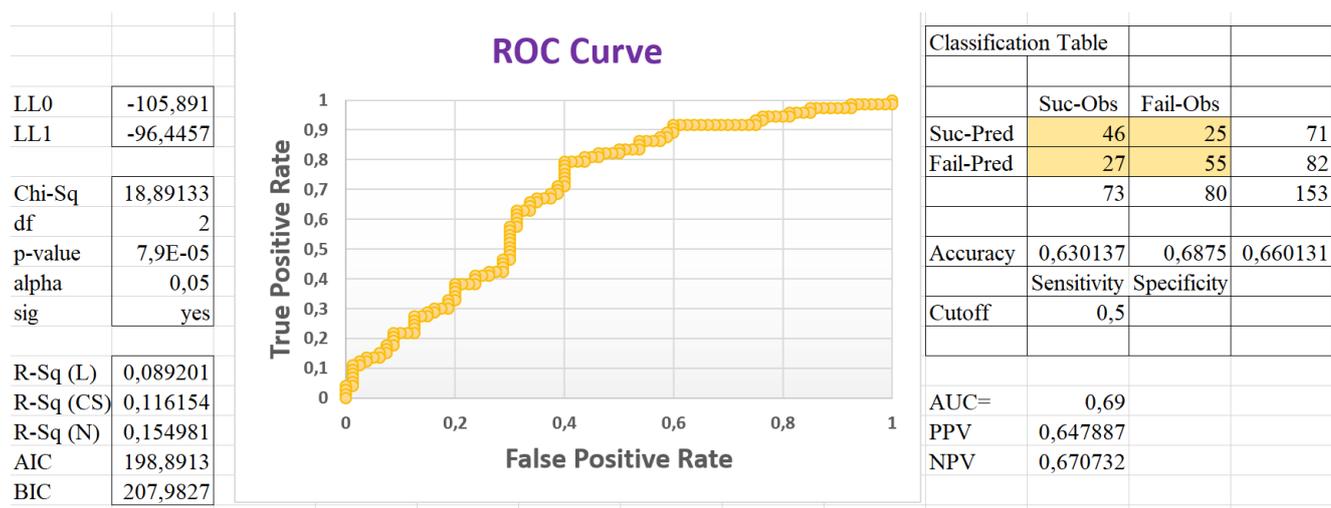


Рис. 5. Двофакторна модель (AUC=0,69).

Рівняння моделі  $Y_i = 1$  для 73 студентів з 153 має вигляд:

$$\ln\left(\frac{p}{1-p}\right) = -0,36 - 0,29X_1 + 0,477X_2.$$

З даної двофакторної моделі робимо висновки, що найбільш впливові фактори на можливий вибір студентів є: 1) втрата інтересу до цієї активності (діяльності), яка раніше приносила задоволення; 2) моральний настрій щодо навчання під час військового стану.

Для моделі  $Y_i = 1$  для 98 студентів з 153 рівняння має вигляд:

$$\ln\left(\frac{p}{1-p}\right) = -1,23 + 0,398X_1 - 0,29X_2 + 0,46X_3.$$

	<i>coeff</i>	<i>s.e.</i>	<i>Wald</i>	<i>p-value</i>	<i>exp(b)</i>	<i>lower</i>	<i>upper</i>
intercept	-1,22716	0,903868	1,843277	<b>0,174567</b>	0,293125		
H	0,397688	0,182431	4,752134	<b>0,029262</b>	1,488379	1,040946	2,128134
I	-0,29084	0,136509	4,539393	<b>0,033123</b>	0,747633	0,572126	0,976978
R	0,464355	0,202692	5,248389	<b>0,021967</b>	1,590987	1,069387	2,367001

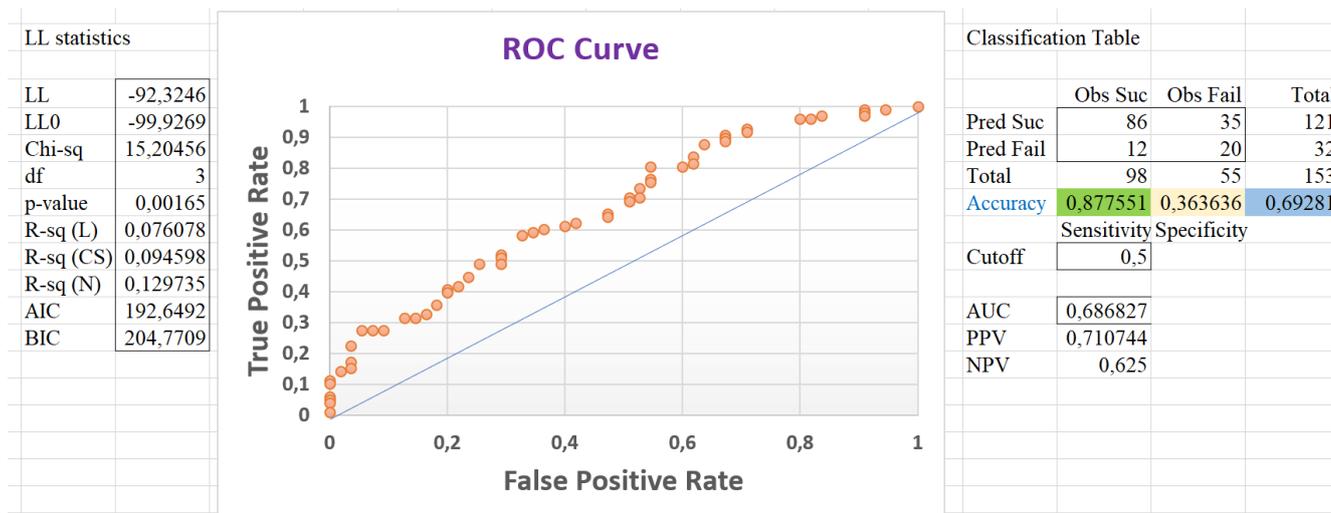


Рис. 6 Трифакторна модель (AUC=0,68).

Трифакторна модель показала, що найбільш впливові фактори на можливий вибір студентів є: 1) значний стресовий досвід, пов'язаний із російським вторгненням в Україну; 2) вибір професії, на який вторгнення суттєво не вплинуло; 2) моральний настрій щодо навчання під час військового стану.

### 3 Висновки

Всі три моделі позначили, як найбільш впливовий фактор, моральний настрій щодо навчання під час військового стану. Зміна цього фактора найбільше впливає на відношення шансів прийняття рішення, про вибір студентів продовжувати свою діяльність в Україні, відношення шансів в залежності від моделі має діапазон від 1,07 до 4,04 (p-value  $p < 0,05$ ). Слід зауважити, що основна практична цінність логістичної моделі регресії полягає у визначенні факторних ознак, що пов'язані з результируючою змінною  $Y_i, i = \overline{1, n}$ , для оцінки сили незалежного впливу цих ознак (при урахуванні інших факторів) на результат, а не для прогнозування результату на практиці.

## References:

1. Посібник з біостатистики. Аналіз результатів медичних досліджень у пакеті EZR (R–statistics) / В. Г. Гур'янов, Ю. Є. Лях, В. Д. Парій, О. В. Короткий, О. В. Чалий, К. О. Чалий, Я. В. Цехмістер : Навчальний посібник. – К. : Вістка, 2018. – 208 с. ISBN 978-617-7157-67-9
2. McDonald, J.H. Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland, 2014. 305 p.
3. The data analysis for this paper was generated using the Real Statistics Resource Pack software (Release 8.9.1). Copyright (2013 – 2023) Charles Zaiontz. [www.real-statistics.com](http://www.real-statistics.com)

---

D.H.Gavrylov & O.V.Mulyk (2025). Assessment of factor signs through logistic regression (analysis of survey results of the negative events impact on personal feelings and academic performance of Ukrainian students in 2022-2023). *Mathematics in Modern Technical University*, 2025(1), 51-59.

*Submitted: 2023-06-21*

*Accepted: 2023-07-17*

**Abstract.** This article aims to demonstrate the possibility of using MS Excel platform to process survey results by constructing a logistic regression Multifactor Model and minimizing factor signs quantity that are associated with output change through stepwise method. Important parameters that characterize the quality and adequacy of the model are also marked: The Akaike information criterion (AIC); The Bayesian Information Criterion (BIC) and AUC stands for "Area under the ROC Curve." From the results obtained on three models, we can conclude that the morale of students during the Russian invasion has a significant impact on their further intentions to continue their studies and professional work in Ukraine.

**Key words:** logistic regression, multivariate model, stepwise method, MS Excel, ROC-curve, AIC, BIC, AUC, p-value.