

Използване на иконометрични методи за анализиране на данни от изследването „Образование и обучение на възрастни“, провеждано през 2016 година от Националния статистически институт

Using econometric methods to analyze data from the "Adult Education and Training" survey conducted in 2016 by the National Statistical Institute

Анжелина Ванева Алексиева¹

Абстракт

В отговор на заложените стратегии за Long-life learning, изследването на образованието и обучението на възрастни осигурява информация за постигнатите цели. Чрез използването на статистическата методология и математическия инструментариум се постига анализ на по-високо ниво, което може да доведе до взимането на по-информирани и навременни решения, с които да бъде надградена стратегията за учене през целия живот.

Abstract

As a response to the defined long-life learning strategies, the study of education and training of adults provides information about our set objectives. A more detailed level of analysis is achieved through the application of statistical and mathematical methods. In turn, those enable us to make timely and better informed decisions about improving the set of the aforementioned learning strategies.

Ключови думи: учене през целия живот, логистични модели, висше образование

JEL: C10, I21, I 26

1.

В съвременния динамичен свят е необходимо човек да се усъвършенства през целия си живот. Това е заложено чрез т. нар. „Националната стратегия за учене през целия живот“, както и редица документи на ЕС . В отговор на потребността на качествена и достоверна информация за изпълнението на тази стратегия се организира репрезентативното изследване „Образование и обучение на възрастни“, провеждано от Националния статистически институт. Към днешна дата Националния статистически институт (НСИ) е провел 4 вълни по единна методология с ЕС.

Изследването на образованието и обучението на възрастни обхваща широк кръг от актуални теми: участието в различните форми на УЦЖ (формално, неформално и самостоятелно образование и обучение), пречки за участие в образование и обучение, достъп до информация относно възможностите за образование и обучение, използване на компютри, езикови умения, участие в културни и обществени събития. Методът на анкетиране е чрез индивидуално интервю, т.е. лицето лично отговаря на въпросите, като се изключва възможността друг член на домакинството да отговаря вместо него. Основната цел на изследването е да се оцени участието на населението в активна трудова

¹Докторант към катедра „Статистика и иконометрия“, УНСС, e-mail: anzhelina.aleksieva@abv.bg

възраст (25 - 64 г.) в ученето през целия живот и в частност - в неговите три основни форми - формално образование и обучение, неформално обучение и самостоятелно обучение. Участието в различните форми на УЦЖ се установява за период от 12 месеца преди датата на интервюто. Ученето през целия живот обхваща всички целенасочени учебни дейности, формални и неформални, предприемани с цел подобряване на знанията, уменията и компетентността на отделния човек, както от професионален, така и от общ интерес за лични и социални цели.

Формално образование и/или обучение е това, което се провежда в образователната система или в центрове за професионална квалификация и което води до придобиване на степен на образование и/или степен на професионална квалификация.

Неформалното обучение също е целенасочено и организирано, но успешното му завършване не води до придобиването на степен на образование или степен на професионална квалификация. Успешното завършване на неформалното обучение може да бъде или да не бъде придружено от издаването на някакъв документ - удостоверение, свидетелство или друго.

Самостоятелното обучение е целенасочена (преднамерена) учебна дейност, предприета с цел повишаване на персоналните знания и умения. Най-често срещаните форми (начини) на самостоятелното обучение са обучение с помощта на член от семейството, колега или приятел, ползване на печатни материали/издания - книги, учебници, професионални списания, ръководства, ползване на компютър, чрез гледане на телевизионни образователни програми или аудио-/видеокасети с образователна цел, посещения на музеи, исторически и природни забележителности с участие на екскурзовод, посещения на учебни центрове, библиотеки и читалища отново с образователна цел.

Оценката на вероятността едно лице на възраст 25 и повече години да продължи образованието си в някоя от посочените насоки може да бъде извършена чрез различни статистически методи. Един такъв метод се явява оценката чрез групата на вероятностните модели и преди всичко чрез така наречените логистични вероятностни модели – Logit модели.

С цел да се установи каква е вероятността дадено лице, жител на Република България, на възраст 25 и повече години да реши да запише формално образование, в конкретика, висше образование (бакалавър, магистър или ОНС доктор) ще бъде използван именно такъв модел.

2.

Най-общият преглед на логистичните вероятностни модели (Logit модели) показва, че стъпват на идеята на линейни вероятностните модели (ЛВМ), като същевременно я доразвиват. Те са опит за преодоляване на част от недостатъците на линейните вероятностни модели. Фундаменталната разлика между двата модела е начина на третиране на зависимата променлива. При ЛВМ анализираме дихотомните променливи такива, каквито са. При logit моделите следствието Y^* е ненаблюдаема (скрита) променлива. Това което се наблюдава е фиктивна променлива Y_i , чрез която се прави опит за индикация на Y^* .

Пример за такъв вид фиктивна променлива е ако едно лице е участва във формално образование, то Y_1 = участие, а Y_0 = неучастие

Еднофакторният модел има вида:

$$P_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \epsilon)}} \quad (1)$$

Многофакторният модел има вида:

$$P_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon)}} \quad (2)$$

където

P_i е вероятността за записване на висше образование, e е неперовото число, $\beta_0, \beta_1 \dots \beta_n$ параметри на модела, e – остатъците на модела.

Според спецификата на тези модели тук се търси проявлението на латентна променлива, която може да се представи чрез оценката на т. нар. шансова пропорция, която има вида:

$$\ln \frac{P_i}{1 - P_i} \quad (3)$$

Шансовата пропорция представлява специфична логаритмична трансформация на вероятността P_i за сбъждане на резултативното явление Y , която се нарича логит и от там идва и наименованието логит (logit) модели. Делителят $1 - P_i$ представлява вероятността за $Y=0$ Същността на логита е да измери вероятността за наличие на дадено явление спрямо вероятността за неговото отсъствие.

След математически преобразования логит моделът придобива следния вид:

За еднофакторния модел:

$$L = \ln \frac{P_i}{1 - P_i} = \beta_0 + \beta_1 X_1 + e \quad (4)$$

За многофакторния модел:

$$L = \ln \frac{P_i}{1 - P_i} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \dots + \beta_n X_n + e, \quad (5)$$

Някои отличителни черти на логит моделите:

- 1) Ако вероятността P се изменя в границите $[0; 1]$, то L се изменя от $-\infty$ до $+\infty$.
- 2) Въпреки че L е линеен спрямо X , то вероятностите не са линейни.
- 3) За да се оцени шансът за единица промяна на фактора X е необходимо да се антилогаритмуват получените L .
- 4) Както и при всички вероятностни модели, оценката на параметрите на регресионното уравнение не може да се извърши чрез Метода на най-малките квадрати (МНК) . Предлага се използването на Метода на максималното правдоподобие или претегления МНК.
- 5) Оценяването на параметрите на логит моделите изисква голям брой изследвани случаи. Факторната променлива може да бъде представена с малко на брой разновидности.
- 6) Факторните променливи могат бъдат както представени на интервална скала, така и на категорийна. Когато дадена категорийна променлива участва във регресионното уравнение е необходимо тя да се трансформира в фиктивна (дъми) променлива.

Както и при други иконометрични зависимости, за оценка и диагностика на качествата на използвания модел се преминава през процедурите за оценка на параметрите, проверката за адекватност, изследване степента на взаимодействие. Възможни начини за намиране на оценките на параметрите е посредством претеглен МНМК или метода на максималното правдоподобие (ММП). В SPSS това се осъществява чрез ММП. След получаване на параметрите на логит модела, ако анализираме данни от представителна извадка трябва да проверим дали параметрите β_n са статистически значими. За целта се провежда процедура по статистическа проверка на хипотези (СПХ). Тук съществуват различни методи, като например метода на Валд, методът на правдоподобното отношение и др. В SPSS най-удобно е да се използва методът на Валд, поради наличните резултати.

При проверката за адекватност на модела отново съществуват различни методи например метода на девианса и проверка чрез хи квадрат разпределение.

В SPSS проверката за адекватност на модела се прави въз основа на девианса. При него оценявания модел трябва да се сравни с „пълния модел“. Пълният генерализиран линеен модел е този, за който:

- Зависимата променлива има същото вероятностно разпределение, както в този, който се оценява;
- Функцията на зависимост в пълния модел е същата като тази в оценявания модел;
- Броят на параметрите в пълния модел е равен на броя на наблюденията.

Изследване степента на взаимодействие - при стандартните регресионни модели съществуват редица методи за изчисляване силата на връзката между фактор и резултативна – например R и R². Същите не са приложими при логит моделите, поради характера на зависимата. Въпреки това съществуват някои измерители, които биха могли да дадат ориентир за силата на връзката между фактор и зависима. Това са измерител на Б. Ефрон, Т. Амемийя, псевдо R² на Й. Крег и Р. Улер и др. В SPSS измерител на връзката е Неделкере R². В днешно време някои от съвременните автори не препоръчват използването им, тъй като ги считат за не дотолкова надеждни.

За всички изброени до тук операции по оценка на логистичните вероятностни модели има разработени стандартни статистически програми в SPSS, E-views, STATISTICA и други.

Приложение на логит моделите в изследването „Образование и обучение на възрастни“ чрез специализираният софтуер

Използвани са негрупирани данни от изследването през 2016. Целта е да се установи каква е вероятността едно лице да се запише във формално образование в степен бакалавър, магистър или доктор ако е на 25 г. и повече.

1. Построяваме модела на връзката

$$L = \ln \frac{P_1}{1-P_1} = \beta_0 + \beta_1 X_1 + e,$$

Уравнението придобива вида:

$$L = b_0 + b_1 X_1,$$

където b_0 и b_1 са оценки на параметрите β_0 и β_1

С помощта на SPSS оценяваме параметрите на модела.

Така за $b_0=0,806$; $b_1=-0,112$ Стандартната грешка на $b_0 = 0,010$, а на $b_1=0,371$.

Необходимо е да се проведат статистически тестове за значимост на параметрите. Провежда се отделна процедура по статистическа проверка на хипотези (СПХ) за всеки параметър, при която:

H_0 гласи, че оценките $b_n=0$, т.е. оценените параметри на регресионното уравнение са статистически незначими. Те не отразяват обективно съществуващата връзка.

H_1 ще гласи, че оценките $b_n \neq 0$, т.е. оценените параметри на регресионното уравнение са статистически значими и отразяват обективно съществуващата връзка.

$\alpha = 0,05$

сравнява се α със sig; $\alpha > \text{sig} \Rightarrow$ можем да отхвърлим нулевата хипотеза. За вярна се приема алтернативната, следователно и двата параметъра са статистически значими.

Получените резултати от проведеното спецификацията на модела са представени в табл. 1.

Таблица 1 Коефициенти на параметрите

Variables in the Equation							95% C.I. for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step AGE	-,112	,010	122,949	1	,000	,894	,877	,912
1 ^a Constant	,806	,371	4,715	1	,030	2,238		

a. Variable(s) entered on step 1: AGE.

Оценката на параметрите на логистични вероятности модели е извършена по Метода на максималното правдоподобие. Въз основа на така получените данни регресионното уравнение има вида:

$$L = \ln \frac{P_1}{1-P_1} = 0,806 - 0,112X,$$

Въз основа на така записаното уравнение тълкуването ще бъде следното:

При увеличение на възрастта с единица шансовата пропорция за записване на висше образование на човек на 25 години и повече, жител на Република България средно ще намалее 0.112 пъти или с 11.2 %. Ако решим да оценим вероятността за различните стойности на фактора X (единични възрасти от 25 до 65 г.) ще използваме уравнение №1. В таблица №2 са представени част от изчислените вероятности. Така например, за жител на Република България, на 25 г., оценената вероятност за записване на висше образование е 12%, за лице на 26 г. – 10,9%, а за лице на 65 г. е 0,2%

Таблица 2 Изчислени вероятности за записване на висше образование по навършени години спрямо логит модела

навършен и години	вероятност
25	0,120
26	0,109
27	0,098
28	0,089
29	0,080
30	0,072
31	0,065
32	0,059
33	0,053
**	**
65	0,002

Таблица 3 Проверка за адекватност на логит модела

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	175,556	1	,000
Block	175,556	1	,000
Model	175,556	1	,000

Таблица №3 показва резултатите от проведен тест за адекватност на модела. На този етап се прави проверка между „базовият модел“, включващ само константа и оценениият. От получените резултати може да се направи изводът, че така конструираният модел е адекватен.

В настоящия доклад бе посочена една възможност за използване на иконометрични методи в областта на образованието. Разбира се, обхватът на тези методи позволява още по-задълбочено търсене и оценяване на статистически модели, не само в сферата на образованието.

References

1. Статистическо изследване на зависимости - проф. д. ик. н Ив. Съйкова, доц. д-р А. Стойкова-Къналиева и ст.н.с.д-р Св. Съйкова УИ „Стопанство“, 2002
2. Статистически анализ на категорийни данни с SPSS, доц д-р Соня Чипева, УИ
3. Стопанство“, 2011
4. Въведение в иконометрията - доц д-р Соня Чипева, доц д-р Венелин Бошнаков, Издателски комплекс УНСС, 2015
5. Статистическо изследване на зависимости - проф. д. ик. н Ив. Съйкова, доц. д-р А. Стойкова-Къналиева и д-р Св. Съйкова УИ „Стопанство“, 2002
6. Статистически данни – Национален статистически институт, отдел „Статистика на образованието и културата“
7. Baguley, Thom. (2022). Re: What is the minimum acceptable limit of Nagelkerke R Square value in binary logit regression model?. Retrieved from: https://www.researchgate.net/post/What_is_the_minimum_acceptable_limit_of_Nagelkerke_R_Square_value_in_binary_logit_regression_model/62cdf153cf8f941b3a0b6d65/citation/download.
8. An Introduction to Categorical Data Analysis, Second Edition ALAN AGRESTI Department of Statistics University of Florida Gainesville, Florida, Wiley 2007
9. Basic Econometrics (5th Edition) by Damodar N. Gujarati, Dawn C. Porter, Porter Gujarathi, The McGraw-Hill Companies, Inc.