UDC:336.1:303.724
336.1:352]:303.724
стручен труд

ЛОГИСТИЧКАТА РЕГРЕСИЈА ВО ЈАВНИТЕ И ЛОКАЛНИТЕ ФИНАНСИИ (СТУДИЈА НА СЛУЧАЈ ЗА ПРЕДВИДУВАЊЕ НА ФИСКАЛНИОТ ДИСПАРИТЕТ НА ОПШТИНИТЕ)

Илија Груевски[1], Стеван Габер[2], Еленица Софијанова[3]

[1]Економски факултет, Универзитет „Гоце Делчев“, Штип
ilija.gruevski@ugd.edu.mk
[2]Економски факултет, Универзитет „Гоце Делчев“, Штип
stevan.gaber@ugd.edu.mk
[3]Економски факултет, Универзитет „Гоце Делчев“, Штип
elenica.sofijanova@ugd.edu.mk

Апстракт: Целта на овој труд е да се долови и демонстрира потенцијалната употреба на бинарната логистичка регресија во научните истражувања од областа на јавните и локалните финансии. Како еден вид генерализиран статистички метод, кој содржи исклучиво бинарни (дихотомни) вредности на зависната варијабла, истиот е извонредно погоден за предвидување на одредени економски појави чијашто состојба т.е. status quo се состои од само две вредности (акции со потпросечен принос, акции со натпросечен принос, ликвидни наспроти неликвиднии општини, општини со низок наспроти општини со висок фискален капацитет, држави во банкрот versus држави вон банкрот итн.) Како резултат на тоа, се овозможува да се направи класификација и групирање на опсервациите во соодветни групи со зголемен ризик од конкретната појава (општини со потпросечен приходен капацитет) и групи со низок ризик од појавата (општини со натпросечен капацитет). Друга поволност која ја нуди логистичката регресија е можноста за пресметка на шансите и веројатноста за настапување на одредена состојба (неликвидност или банкрот) и тоа посебно за секоја индивидуална опсервација (општина, држава итн.). Со ваквиот методолошки приод, истражувачот е во состојба објективно да го процени ризикот од настапувањето на конкретната појава, и во таа насока да даде препораки за конкретни дејствија или мерки со цел да се превенира потенцијалното настапување на појавата т.е. редуцира присутниот ризик. Со помош на студија на случај, демонстрирана е примената на логистичката регресија врз статистички примерок составен од македонските општини, од кој се исклучени единствено скопските општини заради недостаток на податоци. Интересно, истражувањето покажа дека како предиктори на фискалниот капацитет на општините во Македонија, освен бројот на жители и наплатените сопствени приходи по глава на жител, се јавуваат и урбано/руралниот и етничкиот статус на општините.

Клучни зборови: *логистичка регресија, фискален капацитет, фискален диспаритет, општина, предвидување, класификација, Република Северна Македонија.*

LOGISTIC REGRESSION IN PUBLIC AND LOCAL FINANCE (A CASE STUDY FOR PREDICTING MUNICIPAL FISCAL DISPARITY)

Ilija Gruevski[1], Stevan Gaber[2], Elenica Sofianova[3]
[1]Faculty of Economics, Goce Delcev University, Stip
ilija.gruevski@ugd.edu.mk
[2]Faculty of Economics, Goce Delcev University, Stip
stevan.gaber@ugd.edu.mk
[3]Faculty of Economics, Goce Delcev University, Stip
elenica.sofijanova@ugd.edu.mk

Abstract: The aim of this paper is to demonstrate the potential use of binary logistic regression in scientific research in the field of public and local finances. As a form of generalized statistical method, which exclusively contains binary (dichotomous) values of the dependent variable, it is exceptionally suitable for predicting certain economic phenomena whose state, i.e., status quo, consists of only two values

(stocks with below-average returns, stocks with above-average returns, liquid versus illiquid municipalities, municipalities with low versus high fiscal capacity municipalities, bankrupt versus non-bankrupt countries, etc.). As a result, it enables the classification and grouping of observations into groups with increased risk of a specific phenomenon (municipalities with below-average revenue capacity) and groups with low risk of the phenomenon (municipalities with above-average capacity). Another advantage of the logistic regression is the ability to calculate the odds coefficients and probabilities of a certain event (illiquidity or bankruptcy), for each individual observation (municipality, country, etc.). With this methodological approach, the researcher is able to assess the risk of the occurrence of a particular phenomenon objectively and, in that regard, provide recommendations for specific actions or measures to prevent the potential realizatation of the phenomenon, in other wortds, to reduce the existing risk. In this case study, the application of logistic regression is demonstrated on a sample composed of Macedonian municipalities, excluding only the municipalities of Skopje City due to lack of data. Interestingly, the research revealed that the population, ethnic statatus, rural/urban status and the level of own revenues per capita predictors could play a significant role in predicting the fiscal disparity and capacity of municipalities in Macedonia.

Keywords: *logistic regression, fiscal capacity, fiscal disparity, municipality, prediction, classification, Republic of North Macedonia.*

1. Вовед

Бинарната логистичка регресија (изворно:binary logistic regression) претставува генерализиран линеарен статистички модел, којшто има потенцијал за широка употреба во економските истражувања. Можностите за негова примена се исклучително големи, почнувајќи од маркетингот и истражувањето на пазарот, преку корпоративните финансии и банкарството, па сè до јавните и локалните финансии. Всушност, секаде каде има потреба од класификација или групирање, објаснување (пронаоѓање на врска) или предвидување на одредена состојба, исход или настан со бинарно детерминирани вредности (како на пример смрт или преживување од некоја болест, купување или продавање на некое добро, победа на демократската или конзервативната партија, акција со натпросечен или потпросечен принос или пак општина со низок или висок приходен капацитет и сл.), логистичката регресија се покажала како релевантен научно-истражувачки метод. Резултатот од примената на моделот се т.н. шанси за остварување на одреден настан (odd ratios – ORs). На пример, колакви се шанските одредена општина да има потпросечен т.е. низок фискален капацитет, или обратно, натпросечен или висок приходен капацитет, ако се имаат во предвид популацијата, сопствените приходи по глава на жител, рурално/урбаниот и етничкиот статус. Исто така, моделот има способност прецизно ја утврди и веројатноста за остварување на обсервираниот настан чии вредности, како што е познато, може да се движат од 0 (сигурно неостварување на конкретниот настан) и 1 (сигурно остварување на настанот). Технички, моделот се состои од најмалку две или повеќе варијабли од кои само една е зависна варијабла, а една или повеќе се независни варијабли (детерминанти или предиктори). Специфична карактеристика на зависната варијабла се нејзините бинарни или дихотомни вредности со кои се врши категоризација на настаните (на пример со 1 се обележува држава во банкрот, општина које е неликвидна, победа на демократската партија и сл. додека со 0 се обележува или категоризира држава вон банкрот, општина која е ликвидна или победа на конзервативната партија). Што се однесува до независните варијабли во моделот, тие можат да се појават како една или повеќе варијабли со дихотомни (бинарни) или континуелни вредности. На пример, доколу целта е да се истражи каква политичка ориентација имаат гласачите според полот или пак расната припадност, тогаш може да се употребат бинарните вредности (1 за маж и 0 за жена, односно 1 за граѓанин од белата раса и 0 за граѓанин од црната раса). Доколку пак, е потребно да се види како би гласале граѓаните според нивната возраст, тогаш континулените варијабли нудат подобро решение (од најмладиот гласач со 18 години до најстариот со на пример 100 години).

Во рамки на статистичкото моделирање е познато дека, обичните мултифакторски модели на линеарна регресија, коишто сорджат континуелни (непрекинати) вредности на зависната варијабла (на пример БДП, обем на инвестиции, обем на потрошувачка, големина на јавниот долг, вредност на буџеткиот дефицити сл., сите изразени во апсолутни бројки или проценти), нормална карактеристика е претпоставената линеарна врска помеѓу зависната варијабла и нејзините детерминанти. Тоа е така, бидејќи дистрибуцијата

на таквите вредности на голем примерок е нормална, односно, наликува на симетрична парабола. Меѓутоа, кога зависната варијабла се состои исклучиво од бинарни, дихотомни вредности, што впрочем е одлика на логистичката регресија, тогаш добиените коефициенти на регресија не ја изразуваат соодветно линеарната врска на зависната варијабла со факторските детерминанти. Причината за тоа се биноминалните карактеристики на вредностите 0 и 1 од зависната варијабла, чијашто дистрибуција во рамки на големите статистички примероци исто така е биноминално распоредена.

Се поставува прашањето „На кој начин може да се одржи линеараната врска во рамки на бинараната логистичка регресија, ако се имаат во предвид горните констатации?". Одговорот лежи во т.н. функција на шанси (odds function) која има способност да ја трансформира функцијата на веројатност (probability function) со вредности во интервалот од 0 до 1 во еквивалентна функција со вредности во интервал од 0 до ∞ [1]. Интерпретирано со други зборови, на овој начин се извршува трансформација на бинарните вредности на зависната варијабла во континуелни, распоредот, односно дитрибуцијата на вредностите од биноминални добива нормални карактеристики, а врската помеѓу зависната варијабла и независните варијабли станува линеарна. Исто така, може да се насети дека како логична последница на претходново се наметнува и врската помеѓеу овие две функции, а оттаму и поврзаноста помеѓу шансите и веројатностите за настапување на одреден бинарно определен настан (состојба на банкрот, состојба без банкрот). Бидејќи конкретната трансформација на вредностите се изведува со логаритмирање на шансите за тој настан или log(odds), воспостевената релација е наречена logit function, а регресијатата е именувана како логистичка.

1.1 Определување на општиот модел на логистичка регресија

Од погоре изнесеното за одликите и карактеристиките на теоретските постулати на т.н. logit функција , општиот модел на бинарната логистичка регресија [2], може да се опише со следниот израз (1):

$$ln\ odds(e) = ln\left(\frac{p(e)}{1-p(e)}\right) = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n$$

Во оваа формула, елементот од левата страна ln(p/(1-p)) е всушност логаритамот на шансите за настанување на одреден настан ln(odds(e)) и истиот има улога на зависна варијабла, b_0 е пресекот или интерцептот, b_1, b_2 и b_n се регресионите коефициенти на независните варијабли и x_1, x_2 и x_n се вредностите на независните варијабли.

Меѓутоа, за да се конвертира оваа вредност во чиста вредност на шансите за некој настан odds(e) [3], потребно е да се пресмета експоненцијалната вредност на изразот од десната страна (2):

$$Odds(\text{e}) = \exp\left(b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n\right)$$

Тоа не е случајно, ако се има предвид поврзаноста на логаритамската и експоненцијалната функција, бидејќи едната е всушност, инверзна вредност на другата и обратно. Всушност [4], природниот логаритам на било кој број X или ln(X) е еднаква на базата на природниот логаритам со експонент од истиот број X или exp(X), или со други зборови, ln(X) = exp(X). Така на пример, ако го го земеме бројот 100, тогаш ln(100) = 4,605, и обратно, exp(4,605) = 100.

Слична е постапката и со поединечните коефициенти b_0...b_n, од горната формула. Имено, и нивната трансформација се добива со примена на операцијата exp, односно, odds(b_0) = exp(b_0), ..., odds(b_n) = exp(b_n).

Останува уште да се изведе веројатноста од настанување на одреден настан p(e). И тука на помош доаѓа поврзаноста на овие два концепти. Теоријата на веројатност укажува дека веројатноста определен настан да се реализира може да се престави како однос помеѓу бројот на извесни настани и вкупниот број на можни настани. На пример, ако во статистичкиот примерок од вкупно 100 општини има 10 банкротирани општини, веројатноста определена општина да прогласи банкрот изнесува 0,10, односно 10/100. Од друга страна, пак, шансите за настанување на истиот набљудуван настан odds(e) може да се

дефинираат како однос помеѓу веројатноста за настанување на настанот p(e) и веројатноста за ненастанување на настанот 1-p(e) [5], односно (3):

$$Odds(e) = \frac{p(e)}{1 - p(e)}$$

Со цел повратно да ја изведеме p(e), ако е позната вредноста на odds(e), потребно е двете страни од горниов израз (3), да се помножат со фактор 1-p(e), а потоа добиеното равенство да се реаранжира (изолира) според p(e) [6]. Крајниот резултат би бил со следниот облик (4):

$$p(e) = \frac{odds(e)}{1 + odds(e)}$$

И за крај, со замена на горнниот израз (2) во изразот (4), ќе се добие следната формула за пресметка на веројатноста на набљудуваниот настан p(e) (5):

$$p(e) = \frac{\exp\left(b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n\right)}{1 + \exp\left(b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n\right)}$$

2. Класификација и групирање на опсервациите

Формулата (3), која служи за пресмета на веројатноста на некој настан, овозможува да се проценат веројатностите т.е. ризиците за настапување на одреден исход (неликвиднсот, банкрот, намален фискален капцитет и сл.) кај било која опсервација од статистичкиот примерок, којшто е од предмет и интерес на емпириските истражувања. Понатаму, а врз основа на одредена стапка на веројатност, а тоа е обично стапката на половична веројатност од 0,5, сите опсервации се групираат т.е. систематизираат, или во групата опсервации со низок ризик, или пак, во групата на опсервации со зголемен ризик за настапување на настанот. Веројатноста со која се врши бинарна класификација на предвидените настани кај опсервациите од статистичкиот прикерок на точни (true) или неточни (false), а во случајов тоа е веројатноста од 0,5, се нарекува граница или маргина на веројатност (cut-off point, probability threshold) [7]. Инаку, маргината на веројатност може да се оптимизира, на пример преку симулација, со цел да се пронајде посоодветна стапка на веројатност со која би се унапредила парцијалната точност во прогнозата (сензитвност и специфичност) или пак, вкупната (генерална) точност во прогнозата.

Веќе е јасно дека тежиштето на предвидувањето на некој настан е ставено врз сензитивноста и специфичноста во прогнозата, кои ќе ги дефинираме во продолжение. Пред тоа ја претставуваме стандардната табела за класификација која се користи во истражувањата, и која ја имаат имплементирано повеќето софтверски пакети за статистичка анализа.

Слика 1. Табела за класификација

Picture 1. Classification table

Табела за класификација			
	*Го **има** настанот*	*Го **нема** настанот*	
Позитивен	**Точно позитивен**	**Погрешно позитивен**	*Ред за **позитивно** предвидени настани*
Негативен	**Погрешно негативен**	**Точно негативен**	*Ред за **негативно** предвидени настани*
	*Колумна на **сензитивност***	*Колумна на **специфичност***	

Извор: Адаптирано според: Trevethan R., (2017) "Sensitivity, specificity and predictive values: Foundations, pliabilities and pitfalls in research and practice"

Со **сензитивност** на прогнозата се означува односот помеѓу бројот на опсервации со точно остварена прогноза на одреден настан (true positive) и збирот од опсервациите со точно остварена прогноза (true positive) и опсервациите со погрешно негативна прогноза на настанот (false negative) [8], односно (6):

$$\text{Сензитивност} = \frac{\text{Точно Позитивни}}{\text{Точно Позитивни} + \text{Погрешно Негативни}}$$

Специфичноста пак се изразува со пропорцијата од бројот на опсервации со точна негативна прогноза на набљудуваниот настан и збирот од опсервациите со точна негативна и погрешна позитивна прогноза на соодветниот настан [9], или (6):

$$\text{Специфичност} = \frac{\text{Точно Негативни}}{\text{Точно Негативни} + \text{Погрешно Позитивни}}$$

Со пресметката на сензитивноста и специфичноста се постигнува т.н. парцијална евалуација на точноста на прогнозата. Доколку пак тие се интегрираат во единствен показател се добива т.н. **генерална** (вкупна) точност на прогнозата, како што е покажано во долниот израз (7):

$$\text{Вкупна Точност} = \frac{\text{Точно Позитивни} + \text{Точно Негативни}}{\text{Точно Позитивни} + \text{Точно Негативни} + \text{Погрешно Позитивни} + \text{Погрешно Негативни}}$$

Дополнителна и продлабочена оценка на предиктабилната способност на моделот на логистичка регресија може да се постигне со анализа на т.н. ***ROC крива*** *(Receiver Operating Characteristic Curve – ROC curve)*. Всушност, тоа е крива со која се графира стапката на сензитивност во однос на разликата помеѓу 1 и стапката на специфичност, за сите можни степени на маргината на веројатност [10]. Таа има исклучителна дијагностичка моќ за евалуација на точноста во прогнозата кај статистичките модели со кои се врши определана класификација.на опсервациите. Основната алатка која се генерира со кривата е т.н. област или површина под кривата *(Area Under the Curve – AUC),* која може да има различни вредности. На пример, доколку AUC = 1, тоа значи дека моделот врши перфектна прогноза и класификација на опсервациите од статистичкиот примерок. Од друга страна на ектстремот е AUC со вредност од 0,5, а тоа значи случајна класификација, односно, не постојат разлики во групирањето на опсервациите во моделот. Меѓутоа, AUC може да има било каква вредност помеѓу овие две гранични вредности, а рангирањето на способноста за предвидување на статистичкиот модел може да се види од долната табела.

Табела 1. Рангирање на способноста за предвидување и класификација на статистичкиот модел
Table 1. Scoring of the predictive ability and classification of the statistical model

Вредност на AUC	**Квалитет (точност за класификација)**
1	Перфектна класификација (Perfect classifier)
0,9 - 1,0	Одлична (Excellent)
0,8 - 0,9	Многу добра (Very Good)
0,7 - 0,8	Добра (Good)
0,6 - 0,7	Задоволителна (Satisfactory)
0,5 - 0,6	Незадоволителна (Unsatisfactory)
0,5	Случајна класификација (Random Classifier)

Извор: mljourney.com;

Инаку, сите овие калкулации и алгоритми, иако навидум сложени и комлицирани, се изведуваат прилично автоматизирано во рамки на стандардните статистички софтверски пакети. Нашите калкулации за потребите на истражувањето, односно, демонстративната примена на моделот на логистичка регресија се реализираа преку програмскиот пакет *xrealstat*.

3. Студија на случај: Примена на моделот на логистичка регресија за класифицирање на општините со низок и висок фискален капацитет

Мораме најпрвин да спомнеме дека, податоците врз кои се базира оваа студија на случај датираат од поодамна, поточно од 2010 година, кога постоеше различна териториjална поделба на државата по општини. Меѓутоа, застареноста на податоците не го намалува научно-истражувачкото значење на демонстраниот статистички модел, единствено, можеби е доведена во прашање актуелноста на податоците. Имено, статистичкиот примерок се сотои од вкупно 74 општини, во кои не се вклучени скопските општини и градот Скопје. Отсуството на овие опсервации не е тенденциозно напревено, туку, нив не можевме да ги пронајдеме ниту во рамки на оригиналниот извор на податоци, а тоа е студијата на UNDP од 2012, која беше наменета како водиич и прирачник за процесот на фискалната децентрализација во РМ, насловена. *Fiscal Decentralization for Local Development: An Integral Study.*

Инаку, фискалниот капацитет (revenue capacity) го претставува потенцијалот на локалната заедница да генерира сопствени приходи од сопствената даночна основа како и да обезбеди административна поддршка на процесот на наплата [11]. Претходново сугерира дека можноста за наплата на сопствени приходи (фискалниот капацитет) може да се разграничи на даночен капацитет (ширина на даночната основа) и административен капацитет (обезбедување технички, логистички и кадровски предуслови за наплата на даноци). Бидејќи разликите во административниот капацитет можат да се решат со територијално окрупнување или со ангажирање агенции за напата (во чија улога можат да се најдат централната влада или екстерните субјекти), фискалниот капацитет примарно се изедначува со ширината на даночната основа. Разликата во наплатените сопствени приходи по глава на жител, пак, се нарекува фискален диспаритет (fiscal or revenue disparity), без разлика на природата на изворот на таквите разлики [12].

Иако се познати повеќе методи за проценка или мерење на фискалниот капацитет, како што се на пример, историскиот метод, методот на репрезантитивен даночен систем, методот со макроекономски индикатори и сл., авторите на студијата го застапуваат, а воедно го употребуваат во случајот на РМ, релативно комплицираниот, иновативен метод на т.н. proxy варијабли. И сето тоа, заради едноставна причина - недостаток на податоци. Со користење варијабли кои ја апроксимизираат локалната даночна основа и со употреба на просечната ефективна даночна стапка, целта на методот е да се измерат потенцијалните сопствени приходи на локалните заедници и на тој начин да го реплицираат репрезентативниот даночен систем [13].

Со оглед дека во Македонија не постојат релевантни податоци за големината на даночните основи на доминантните извори на локални даноци, а тоа се данокот на имот и фирмарината, авторите предлагаат алтернативни proxy варојабли кои имаат висок степен на корелација со сопствените приходи. Така, како proxy варијабла, за основата на данокот на имот, тие ја предлагаат површината (квадратурата) на недвижниот имот на домаќинствата и становите (измерен степен на корелација со сопствените приходи од 0,28). Како апроксимативна варијабла за потенцијалните приходи од данокот на комерцијален имот (фирмарина), тие ги предлагаат приходите од заедничките даноци по основ на ПДД (корелација со сопствените приходи од 0,72).

Понатаму, тие развиваат формула за пресметка на фискалниот капацитет која ја применуваат на македонските општини, која го има следниот облик (8):

$$Capacity_i = \left(1 + 0{,}72 * \frac{PIT_{per\ capita\ (i)} - PIT_{per\ capita(N)}}{PIT_{per\ capita(N)}} + 0{,}28 * \frac{Kvadratura_{per\ capita\ (i)} - Kvadratura_{per\ capita(N)}}{Kvadratura_{per\ capita(N)}}\right) * POP_i * OR_{per\ capita(N)}$$

При што симболите го имаат следново значење [14]:
$Capacity_i$ – фискален капацитет на општина i,
$OR_{per\ capita(N)}$ – сопствени приходи по глава на жител (на национално ниво),
$PIT_{per\ capita(i)}$ – приходи од ПДД по глава на жител во општина i,
$PIT_{per\ capita(N)}$ – приходи од ПДД по глава на жител (на национално ниво),
$Kvadratura_{per\ capita(i)}$ – квадратура на имот по глава на жител во општина,
$Kvadratura_{per\ capita(N)}$- квадратура на имот по глава на жител (на национално ниво),
POP_i – број на жители на општина i.

Применувајќи ги чекорите од горната формула, авторите го пресметуваат фискалниот капацитет на општините во Македонија за 2010 година, но како што се гледа целиот тој процес е доста макотрпен, бара собирање на многу податоци од катастарот, УЈП, Министарството за финансии, општините и Министерството за локална самоуправа итн.

Во продолжение, ќе демонстрираме како со употреба на логистичката регресија, би можеле да го заобиколиме целокупниот процес на калкулации со тешко достапни информации и податоци, и да предвидеме со релативно голема точност, кои и колку од македонските општини би имале понозок, а кои и колку од нив повисок приходен капацитет во однос на просекот.

3.1. Применетиот модел на логистичка регресија

Применетиот модел на логистичка регресија во нашето демонстративно истражување за предвидување и класифицирање на општини со потпросечен и натпросечен фискален капацитет, на статистичкиот примерок на македонските општини од 2010 година, може да се опише како:

$$ln\left(\frac{p(fiskalen\ kapacitet)}{1 - p(fiskalen\ kapacitet)}\right) = b_0 + b_1 Populacija + b_2 Etnikum + b_3 Rural/Urban + b_4 Sopstveni\ Prihodi\ per\ Capita$$

Така што:
p(fiskalen kapacitet) – веројатност за потпросечен фискален капацитет на општината
ln(p(fiskalen kapcitet)/(1-p(fiskalen kapacitet) – Логаритам од шансите за потпросечен фискален капацитет на општината;
Populacija – Популација или број на жители на општината;
Etnikum – Етнички статус на општината;
Rural/Urban – Рурално/Урбан статус на општината;
Sopstveni Prihodi per Capita – Сопствени приходи по глава на жител;
b_0 – Интерцепт (коефициент на контролната или референтна група);
b_1 – Коефициент на варијаблата Populacija;
b_2 – Коефициент на варијаблата Etnikum;
b_3 – Коефициент на варијаблата Rural/Urban и
b_4 – Коефициент на варијаблата Sopstveni Prihodi per Capita.

Од елементите на моделот може да се види дека истиот се состои од вкупно 5 варијабли, од кои една е зависна, а останатите 4 се независни варијабли (детерминанти, предиктори или фактори на влијание). Зависната варијабла, а тоа е ***фискалниот капацитет***, содржи бинарани (дихотомни) вредности, кои се кодирани на начин што доколку конкретната општина има негативен фискален диспаритет се доделува 0, во спротивно се доделува 1. Всушност, фискалниот капацитет е детерминиран од фискалниот диспаритет, така што општините со негативен диспаритет имаат потпроселен фискален капацитет, и обратно општините со позитивен диспаритет имаат натпросечен фискален капацитет. Кај независните варијабли, две од нив се бинарни, а две со континуелни вредности. Конкретно, со варијблата ***етникум*** се кодира етничкиот статус на општината, така што со 0 се бележи општина со доминантно

македонско население, во спротивно се бележи 1. Бинарни карактристики содржи уште варијаблата ***рурал/урбан*** со која се кодира рурално-урбаниот статус општината, така што со 0 се бележи рурална општина, а со 1 урбана. Варијаблите ***популација*** и ***сопствени приходи per capita*** за разлика од претходните две, имаат континуелни вредности, кои теоретски би можеле да содржат вредности од 0 до ∞. Имајќи го предвид претходново, можеме да заклучиме дека ***референтната (споредбената) или контролна општина*** *е од рурален карактер, со доминантно македонско население, со број на население од 0 и остварува 0 денари сопствени приходи по глава на жител*.

3.2. Резултати од моделот

Подолу се презентирани резултатите од применетиот модел на логистичка регресија, добиени со софтверскиот пакет xrealstat.

Шанси и веројатност. Во долната табела од десната страна на сликата истакнати се логаритмите на шансите на фискалниот капацитет, за секоја одделна независна варијаба, обележани со coeff b. Во истата табела во колоната exp(b), презентирани се шансите на фискалниот капацитет по варијабли, автоматски изведени од претходните коефициенти според веќе познатите математички образци. Бројките покажуваат дека сите детерминанти имаат одредено влијание врз зависната варијабла, некои поголемо, некои помало, од коишто единствено варијаблата етникум не е статистички сигнификантана (p>0,05). Така на пример, шансите урбаните општини да имаат потпросечен фискален капацитет изнесуваат 0,08, односно тие се за 92% помали ((1-0,08)x100)) во споредба со шансите на руралните општини, со интервал на доверба од 0,017 до 0,404. Од аспект на популацијата, за секој поединечен жител шансите на општината да има потпросечен приходен капацитет се зголемуваат за 1,000039 пати (интервал на доверба од 1,000002 до 1,000076). Иако побројните општини, по некоја логика, би требало да имаат поголема способност да генерираат приходи, претходните факти говорат во обратна насока за македонските општини. Што се однесува до сопствените приходи per capita, со секој собран денар сопствени приходи по глава на жител, шансите општината да има низок фискален капацитет се зголемуваат за 0,998 пати, т.е. се намалуваат за 0,2%((1-0,998)x100)), со интервал на доверба од 0,996 до 0,999. Иако варијаблата етникум е статистички незначајна ако се имаат во предвид нејзите p-вредности и интервал на доверба, тоа не мора да значи не постои воопшто ефект т.е. врска, туку едноставно значи дека утврдената врска од примерокот не може да се генерализира за целата статистичка популација. Бидејќи ефектот е очигледен, тука интерпретираме дека шансите една општината да има потпросечен приходен капацитет се за цели 765637491,3 пати поголеми, доколку нејзиниот етнички состав е со мнозинско немакедонско население, во однос на општините со доминантно македонско население! Ова не треба да изненадува ако се знае дека сите општини од статистичкиот примерок со ваков етнички статус, всушност, имаат навистина голем дисперитет помеѓу собраните приходи на локално ниво и националниот просек.

Слика 2. Резултати од статистичкиот модел
Picture 2. Results from the statistical model

Coeff		LL0	-50,8596	Covariance Matrix						Converge	
		LL1	-26,6603	0,99128	7,46E-06	-0,08117	-0,25891	-0,0006		1,29E-08	
2,787905				7,46E-06	3,59E-10	-3,1E-06	-7,3E-06	-7,3E-09		0,000482	
3,87E-05		Chi-Sq	48,39856	-0,08117	-3,1E-06	77264803	-0,0824	4,36E-05		1,29E-08	
20,45622		df	4	-0,25891	-7,3E-06	-0,0824	0,631881	6,21E-05		1,05E-08	
-2,46265		p-value	7,79E-10	-0,0006	-7,3E-09	4,36E-05	6,21E-05	4,76E-07		2,11E-05	
-0,00166		alpha	0,05								
		sig	yes								
					coeff b	*s.e.*	*Wald*	*p-value*	*exp(b)*	*lower*	*upper*
		R-Sq (L)	0,475805	Intercept	2,787905	0,995631	7,840783	0,005108	16,24694		
		R-Sq (CS)	0,480056	Populacija	3,87E-05	1,89E-05	4,17546	0,041013	1,000039	1,000002	1,000076
		R-Sq (N)	0,642598	Etnikum	20,45622	8790,04	5,42E-06	0,998143	7,66E+08	0	#NUM!
		AIC	63,32066	Rural/Urba	-2,46265	0,79491	9,597772	0,001948	0,085209	0,017941	0,404681
		BIC	74,84099	Sopstveni	-0,00166	0,00069	5,80154	0,016012	0,998339	0,996989	0,99969

Извор: Сопствени калкулации во xrealstat.

Да видиме сега, како се интерпретираат шансите во однос на контролната група на општини. Ако се потсетиме, референтната општина е од рурален карактер, со доминантно македонско население, со број на население од 0 и остварува 0 денари сопствени приходи по глава на жител. Шансите на општина со ваков профил да има потпросечен капацитет се гледаат од интерцептот, а од табелата се гледа дека тие изнесуваат 16,246. Да претпоставиме сега дека некоја општина е од градски (урбан) карактер, со доминантно немакедонско население, има 10.000 жители и остварува 2.000 денари сопствени приходи по глава на жител. Шансите општина со ваков профил да има потпросечен капацитет изнесуваат: 16,246 x 0,08 x 765637491,3 x $1,000039^{10.000}$ x $0,998^{2.000}$ или 26437384,6, што би значело со 100%-на сигурност, дека хипотетичката општина би имала низок приходен капацитет (или негативен фискален диспаритет). Ако претпоставиме сега дека општината ги има истите карактеристики, единствено се разликував по етничкиот состав тогаш би имале 16,246 x 0,08 x 1 x $1,000039^{10.000}$ x $0,998^{2.000}$, односно, шансите за реализација на настанот кај општината би биле само 0,034.

Сега вниманието ќе го насочиме од шансите кон веројатноста. Користејќи ги вредностите за шансите, лесно може да се изведе и веројатноста според утврдениот образец odds/(1+odds). На пример, очекуваната веројатност одредена градска општина да има низок фискален капацитет изнесува 0,07 или 0,08./(1+0,08), што значи од вкупно 100 градски општини, 7 би биле со негативен диспаритет. Очекуваната веројатност општина со немакедонски етнички статус да биде со низок капацитет е 0,99, односно скоро 1, што значи од 100 општини со ваков профил, речиси сите би биле со потпросечен капацитет. Очекуваната веројатност општина со карактериски на хипотетичката контролна група, да биде класифицирана како фискално ризична изнесува 0,94, т.е. 16,246 / (1+16,246), односно од вкупно 100 општини со профил како на контролната група, дури 94 би биле фискално ризични. Ако претпоставиме општина од градски (урбан) карактер, со доминантно македонско население, популација од 10.000 жители, која остварува 2.000 денари сопствени приходи per capita, очекуваната веројатност таа да има потпросечен капацитет би била 0,032 или 0,034/(1+0,034). Ова значи од 100 општини со вакви карактеристики, само 3 би остариле потпросечен фискаленн капацитет.

Што се однесува до сигнификантноста на моделот, таа претставува корисна алатка за евалуација на можноста за вклопување на моделот во рамки на статистичката популација (model fit), но и добар приод за разбирање колку добро избраниот модел го рефлектира она што е опсервирано во рамки на податоците од статистичкиот примерок. Сигнификантноста се одредува со помош на хи-квадрат тест (chi-square statistics), кој автоматски се пресметува кај повеќето софтверски пакети. Во нашиот пример, резултатите се истакнати на Слика 1, од каде се гледа дека хи-квадрат тестот е статистички сигнификантен, со хи-квадрат вредност од 48,398, степени на слобода 4 и p-вредност помала од 0,001 (7,79E-10).

Класификација и анализа на ROC крива. Софтверот автоматски ги пресметува стапаките на сензитивност, специфичност и генаралната точност на прогнозата, генерирајќи ја табелата за класификација на опсервацииите според шематскиот приказ од Слика 1. Во продолжение се прикажани резултатите од моделот заедно со нивната интерпретација.

Како што се гледа, добиените резултати од предвидувањето се добиени со користење на стандардната маргина на веројатност од 0,5. Притоа, стапката на сензитивнот изнесува 82,9%, односно (34/41)x100, што значи од вкупно 41 општина во рамки на статистичкиот примерок кои имаат негативен фискален капацитет, моделот точно ја предвиди набљудуваната состојба кај 34 општини (34/41)x100 = 82,9%. Стапката на специфичност пак, изнесува 81,8% или (27/33)x100, од која може да се констатира дека од вкупно 33 општини со позитивен фискален диспаритет, моделот извршил успешна прогноза кај 27 од нив. Ако ги интегрираме сензитивноста и специфичноста, ја добиваме генерланата точност на предвидувањето. Нејзината стапка изнесува 82,4% или ((34+27)/74)x100, со интерпретација дека од вкупно 74 општини, колку што изнесува целиот статистички примерок, моделот извршил точно предвидување кај вкупно 66 општини. И покрај тоа што резултатот од предвидувањето е за доволителен, постојат можности за оптимизација на маргината на веројатност, со цел унапредвивање на точноста во прогнозата. Маргината може да се модифицира и во зависност од целите на истражувањето, дали фокусот е да се предвидат позитивноте или пак негативните нестани, или пак е потребно да се унапреди

општата точност на прогнозата во целина. Во литературата можат да се најдат повеќе методи за оптимизација на маргината на веројатност, како што се на пример, правилото за максимален збир и правилото за максимален производ на стапките на сензитивност и специфичност, како и правилото за максимална вредност на индексот youden [15]. Инаку оптимизацијата на т.н. cut-off margin се изведува со помош на симулација.

Табела 2. Резултати од класификацијата во моделот
Table 2. Classification results in the model

Табела за класификација						
				Маргина	0,5	
	*Го **има** настанот*		*Го **нема** настанот*			
					Позитивно	
					предвидени настани	
Позитивен	34		6			*40*
					Негативно	
					предвидени настани	
Негативен	7		27			34
	41		33			74
	0,829268		0,818182			0,824324
	Сензитивност		***Специфичност***		***Вкупна Точност***	

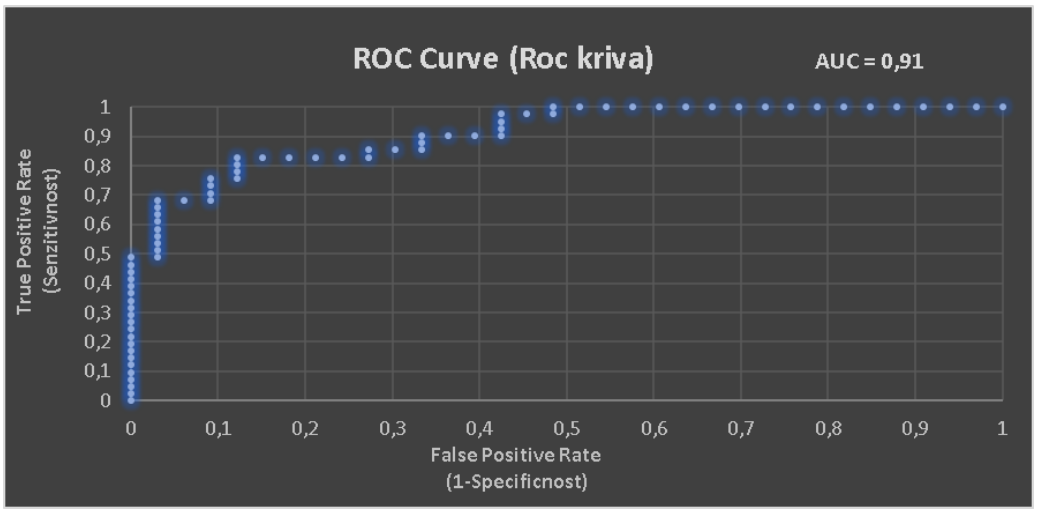
Извор: Сопствени калкулации во xrealstat

Друга дијагностичка алатка која ја нуди логистичката регресија од аспект на проценката на у

спешноста во прогнозата на одредени бинарно определени настани е таканаречената ROC крива. Како што е спомнато, со неа се мапира стапката на сензитивност (true positive rate) во однос на разликата помеѓу 1 и стапката на специфичност (false positive rate), за сите можни комбинации на маргината на веројатност [10]. Во нашиот модел таа е автоматски добиена и го има изгледот како што е прикажано на графиконот подолу.

Главното аналитичко средство во рамкии на кривата е површината под кривата (Area Under the Curve – AUC), која има вредност од 0,91. Ако направиме евалуација според скалилата за квалитет на моделот од Табела 1, може да му се додели одлична оценка (excellent), која се доделува на модели со AUC во интервал од 0,9 до 1,0. Од аспект на предвидувачката моќ на моделот, ова значи дека тој поседува висок капацитет и точност за предвидување и класификација на општини со потпросечен фискален капацитет, односно негативен фискален диспаритет

Графикон 1. ROC крива од статистичкиот модел
Graph 1. ROC curve from the statistical model



Извор: Сопствена изработка на авторите

4. Дискусија

Проценката и мерењето на фискалниот капацитет на општините претставува сложен и комплициран процес, кој во себе инкорпорира мноштво податоци, кои можат да бидат недостапни или несредени, а во истовреме иницира употреба на современи и иновативни методи за пресметка. Со мерење на фискалниот капацитет, а особено со утврдување на фискалниот диспаритет помеѓу потенцијалната даночна основа и остварените приходи на општината, централните власти добиваат претстава колку треба да изнесуваат еквилизирачките трансфери за премостување на фискалниот јаз. Моделот од студијата на UNDP токму тоа и го постигнува, но со ограничувањата кои погоре ги образложивме. Мотивот, односно идејата за истражувањето е да се види дали без прецизно мерење на диспаритетот и фискалниот капацитет, може да се идентификуваат и разграничат општините кои во тој поглед би остварile сопствени приходи помали во однос на нивниот потенцијален фискален капацитет. Односно, да се предвидат општините со негативен фискален диспаритет и тоа со употреба на лесно достапни и едноставни податоци. На овој начин не би се постигнала квантификација на потребните средства за еквилизација, единствено би се овозможило детекција на општините на кои им е неопходна таа.

За оваа цел ја избравме бинарната логостичка регресија, која ја има таа способност да врши класифицирање на зависните варијабли со бинарно определени вредности. Како зависна варијабла се определи диспаритетот на општините (со назив фискалан капацитет), а податоците беа позајмени од студијата на UNDP. Притоа, општините со негативен фискален диспаритет (помали наплатени сопствени приходи во однос на потенцијалниот фискален капацитет на општината) беа кодирани со 1, додека оние со позитивен добија код 0. Во поглед на независните варијабли, изборот беше малку проблематичен. Поаѓајќи од воспостевниот принцип во локалните финансии, дека средствата треба да ги следат надлежностите или со други зборови, дека фискалниот капацитет треба да одговара на фискалните потреби, изборот на независни варијабли беше направен врз популацијата, територијата на општината и бројот на населени места во општината (со помош на овие фактори се врши расптеделбата на дотациите по основ на ДДВ согласно фискалните потреби). Меѓутоа, прелиминарните резултати не беа задоволителни, што впрочем и не е случајно, со оглед дека потребите се едно, додека капацитетот сосема друго, а нивната изедначеност доаѓа само во идеален случај (во светот речиси не постои општина која своите надлежности ги финансира исклучиво од сопствени приходи). Заради тоа, од моделот се исфрлија детерминантите територија и број на населени места, а во истовреме беа додадени детерминантите за рурално/урбаниот статус, бидејќи се забележуваше драстична разлика во диспаритетот помеѓу градските и селските општини, како и факторот сопствени и приходи по глава на жител. Исто така, и бројот на жители се задржа како предиктор, водејќи се од девизата дека населението претставува генератор на приход. Тестирањето на моделот со претходниве варијабли покажа значително подобрување, сепак се чувствуваше потребата од дополнитено калибрирање заради постигнување одличен степен на предикција на моделот. Тоа се направи со интергрирање на уште една варијабла, а тоа е етничкиот стаус на општината, откако се констатира негативен приходен диспаритет кај сите општини со доминантно немакедонско население.

Предиктори на фискалниот диспаритет. Моделот на логистичка регресија се состои од 4 независни варијабли, предиктори на фискалниот диспаритет кој ја претставува зависната варијабла.

Првата независна варијабла се однесува на ***популацијата*** или бројот на жители, која традиционално се смета за генератор на приходи на општината, но истовремено и детерминанта на фискалните потреби. Имено, колку е поголем бројот на жители, толку поголема е можноста повеќе приходи од ПДД да се распределат во општината, согласно принципот на потекло на приходите. Поголемаа популација значи и повеќе семејства, домаќинства и недиижен имот кој е предмет и основа на данокот на имот. Во истражувањето моделот детектира скромна позитивна врска помеѓу бројот на жители и негативниот фискален диспаритет, односно спротивно на нашите очекувања, кај понаселените општини шансите да имаат потпросечен фискален капацитет се поголеми. За популацијата може

да се констатира дека на примерокот на македонските општини, таа претствува посилна детерминанта на фискалните потреби, отколку на фискалниот капацитет.

Со втората независна варијабла ***етникум*** се означува етничкиот статус на општината. Интересно е што моделот открива мнгу силна позитивна корелација меѓу општините со немакедонско доминантно население и негативниот фискален диспаритет, Имено, шансите некоја општина со ваков профил, без ралика дали е урбана или рурална, да има потпросечен капацитет за генерирање приходи се скоро стопроцентни, па логично се поставува прашањето: која е причината за хипервисокиот фискален диспаритет на немакедонските општини? Дали е тоа заради скромниот административен капацитет, нискиот даночен напор или пак тое е едноставно навиката да се чека централната власт да ги покрие „фискалните дупки" на локалната власт (т.н. ефект на леплива хартија).

Со третата детерминанта ***рурал/урбан*** се обележува рурално/урбаниот статус на општината, која во рамки на моделот на логистичката регресија манифестира силна негативна врска помеѓу урбаните, градски општини и негативниот фискален диспаритет. Конкретно, урбаните општини имаат низок ризик да остварат приходи пониски од својот приходен потенцијал, за разлика од руралните општини. Искусвото покажува дека помалите селски општини имаат скромен административен капацитет за наплата и собирање на даноци, а постои дефицит на знаење и стручни кадри да организираат даночни одделенија и финансиски служби. Голем дел од овие општини имаат карактеристика на т.н. монотипни општини, т.е. општини со голем обем на делегирани надлежности, но мала даночна база. Оваа појава беше резултат на претходната територијална поделна на Републиката, која продуцира преголем број на мали рурални општини, па така решенијата за овој проблем се огледаат во територијалното окрупнување на општините, овластувањето на големите општини за администрирање на приходите на малите оптшини, па дури и во ангажирање на надворешни консултанти или агенти за потребите на даночната администрација. Други потенцијални причини за скромниот фискален потенцијал на малите и рурални општини се помалата квадратура и вредност на недвижниот имот, како и помалиот број на вработени лица, додека вработените лица се претежно со ниска квалификација и ниска продуктивност на трудот која не креира висока додадена вредност. Сето тоа носи скромни приходи за овој тип на општини, како по основ на ПДД, така и по основ на данокот на имот.

Последната четврта независна варијабла е предикторот ***сопствени приходи per capita***, која се однесува, како што самото име покажува, на износот на наплатени сопствени приходи по глава на жител. Врската која ја разоткрива моделот помеѓу сопствените приходи и негативниот фискален диспаритет е скромна, но сепак негативна и значајна. Така, општините со поголем износ на сопствени приходи per capita имаат помали шанси да остварат негативен фискален диспаритет и така да манифестираат потпросечен капацитет за наплата на приходи.

Од сите четири варијабли, најдинамични промени по години покажуваат сопствените приходи и популацијата, додека етничкиот и рурално/урбаниот статус се прилично статични, односно, можно е да помине подолго време додека се променат статистичко-правните дефиниции кои го детерминираат таквиот статус. Покрај тоа што бројот на население се менува релативно брзо како резултат на иселувањето и миграциските трендови, сепак тие промени не можат да се доловат на годишно ниво од проста причина што попис на населелние се спроведува доста ретко. Последниов факт го анулира динамичниот карактер на факторот популација и со тоа му придава статичен белег. Од ова следува дека сопствените приходи per capita го претставуваат примарниот фактор од кој би зависеле промените во предвидувањето на фискалниот диспаритет на општините од година во година.

На крајот би дале препораки за понатамошни и продлабочени истражувања со примена на моделот на логистичката регресија. Добро би било на пример, да се интергрира во моделот ефектот од квадратурата на недвижниот имот како и наплатениот ПДД, кои претставуваат proxy варијабли на фискалниот капацитет во оригиналната студија на UNDP. Пожелно е да направи истражување со кое ќе се креира модел за предвидување на ризикот од презадолжување или банкрот на општините, и со помош на логистичката регресија да се иднтификуваат потенцијалните предикатори кои имаат најголемо влијание врз таквата состојба.

5. Заклучок

Логистичката регресија може да биде навистина корисна алатка во рамки на истражувањата од областа на локалните и јавните финансии, но и во поширокото поле на финансиите. Методолошката вредност на логистичката регресија доаѓа од нејзината способност да изврши идентификување, групирање и класификација на опсервациите во статистичкиот примерок според некој бинарно определен критериум. Дополнително имплементираните алатки, каква што е ROC кривата, дозволуваат да се направи евалуација на точноста во прогнозата, преку пресметка на стапките на сензитивност и специфичност итн. AUC. Аналитичката моќ на логистичката регерсија добива на уште поголема димензија, со генерирањето на шанските и веројатноста од настапување на бинарно определниот настан. Имено, преку шансите се определува интензитетот на врската помеѓу независната и зависната варијабла, додека веројатноста го детерминира инхерентниот ризик од настапување на набљудуваниот настан.

Со цел да се промовира поширока употреба на логистичката регресија, во овој труд извршивме демонстрација како овој статистички метод може да се примени за предвидување на општините со негативен фискален диспаритет, на примерокот на македонските општини од 2010 година. Притоа, како главни детерминанти во моделот беа земени бројот на жители, етничкиот статус на општината, рурално/урбаниот статус и сопствените приходи по глава на жител. Кај сите е утврдено статистички значајно влијание врз фискалниот диспаритет, освен кај варијаблата етникум.

Имајќи ја предвид демонстративната и едукативна намена на трудот, во негови рамки се потенцирани само најважните елементи кои треба да ги содржи еден стандарден труд во кој се применува логистичка регресија. Заради тоа, во продолжение на овој заклучок, ја презентираме чек-листата на постапките и чекорите за кои е пожелно да бидат имплементирани заради успешна реализација на еден ваков истражувачки труд:

- **Определување на целта и предметот на истажувањето**. Логистичката регресија не може да се употреби за секаков вид истражување. Потребно е да има компатибилност на целите на истражувањето со природата на логистичката регресија, а тоа е вредностите на зависната варијабла да бидат бинарно определани. Во оваа фаза се определува и статистичкиот примерок, како дел и репрезент на целата статистичка популација.
- **Плотирање и визуална анализа на податоците.** Бидејќи е линеарен метод, мора да се утврди прелиминарно барем една потенцијална линеарна врска помеѓу логаритамот на шансите и вредностите на независните варијабли. Особено треба да се проверат варијаблите со континуелни вредности, со оглед дека варијаблите со дихотомни вредности, поради нивната биноминална дистрибуција не можат да обезбедат линерана врска.
- **Калибрирање на моделот.** Моделот на логистичка регресија треба пробно да се провери. Во овој чекор е можно дел од варијаблите со најлоши перформанси да се исфрлат, а на нивно место да се инкорпорираат други независни варијабли. Потребно е да се испита можноста од егзистенција на т.н. скриена, латентна варијабла (lurking variable). Во овој процес се отстрануваат и потенцијалните оутлајери, т.е. опсервациите со премногу дисперзирана вредност во однос на медијаната.
- **Дескриптивна статистичка анализа.** Пред употребата на главниот модел, потребно е да се направи дескриптивна статистичка анализа, со која се опишуваат статистичките карактеристики на податоците од примерокот, како на пример просечната вредност, медијаната, максимална и минимална вредност, интервал на вредност, интер-квартилен ранг на вредност итн.
- **Презентирање и интерпретација на резултатите од моделот.** Постојат повеќе софтвери кои содржат алгоритми од логистичката регресија како на пример SPSS, Phyton, Stata, xrealstat итн. Тука како резултати се мисли на шансите и веројатностите по варијабли, нивната интерпретација и статистичка значајност.

- **Оценка на статистичката значајност и подобност на моделот (significance and model fit).** Се презентираат и интерпретираат накратко параметрите од пресметките кои укажуваат дека моделот е статистички значаен и подобен.
- **Тестирање на статистичките претпоставки на моделот**. Различните модели се базираат на одредни претпоставки кои мора да бидат задоволени со цел резултатите да сметаат за валидни. Најбитните претпоставки на логистичката регресија се однесуваат на линеарноста (мора да постои линерана врска помеѓу логаритамот на шансите и вредноста на континуелните експланаторни варијабли), мултиколинеарноста (не смее да постои перфектна мултиколинеарност помеѓу независните варијабли), независност (опсервациите во рамки на статистичкиот примерок мора да бидат независни) и големина на статистичкиот примерок (бројот на опсервации мора да биде доволно голем за да се извлечат релевантни, но и генерализирани заклучоци)[15], [16]. Притоа, сите овие претпоставки мора да се тестираат, за што постојат соодветни статистички тестови. Така на пример, мултиколинеарноста помеѓу експланаторните варијабли на моделот се тестира со критариумот фактор на инфлација на варијансата (Variance Inflation Factor – VIF).
- **Класификација на опсервациите и оценка на предвидувачката моќ на моделот.** Се презентираат и интерпретираат резултатите од предвидувањето и класификацијата на опсервациите. Се анализираат сензитивноста, специфичноста и генералната (вкупна) точност на прогнозата. Доколку е потребно, се врши оптимизација на маргината на веројатност со примена на симулација. Понатаму се прави анализа на ROC кривата и се дава оценка на способноста на моделот за предвидување и класификација врз основа на вредноста на AUC.
- **Дискусија и заклучок.** Истражувачкиот труд мора да содржи дискусија во која е многу важно да се истражат и аргументираат причините за констатираните врски и релации помеѓу експланаторните варијабли и зависната варијабла. Во рамки на заклучокот се даваат заклучните согледувања, но обопштено без да се навлегува во детали. Тука е пожелно и да се апострофираат некои препораки и сугестии поврзани со истражувачкиот труд.

Користена литература

[1] Zaiontz C. (2024) Basic concepts of logistic regression, Real-statistics.com;
[2] Agresti A.(2008). Categorical Data Analysis, Third Edition, University of Florida, John Wiley & Sons;
[3] Zaiontz C. (2024) Basic concepts of logistic regression, Real-statistics.com;
[4] Zaiontz C. (2024) Exponentials and Logs, Real-statistics.com;
[5] Agresti A.(2008). Categorical Data Analysis, Third Edition, University of Florida, John Wiley & Sons;
[6] Kleinbaum D.G. (1998) Logistic Regression: A Self-Learning Text, Springer;
[7] DataCadamia, (2024) Threshold/cut-off of binary classification, datacadamia.com;
[8] Baratloo A., Hosseini M., Negida A., El Ashal G., (2015) Part 1: Simple definition and calculation of accuracy, sensitivity and specificity", PMC;
[9] Baratloo A., Hosseini M., Negida A., El Ashal G., (2015) Part 1: Simple definition and calculation of accuracy, sensitivity and specificity", PMC;
[10] Ilker U., (2017) Defining an optimal cut-point value in ROC analysis: An alternative approach, PMC;
[11] Груевски И., Габер С., (2023) Основи на локални финансии, Економски факултет, Универзитет Гоце Делчев, Штип;
[12] Груевски И., Габер С., (2023) Основи на локални финансии, Економски факултет, Универзитет Гоце Делчев, Штип;

[13] Груевски И., Габер С., (2023) Основи на локални финансии, Економски факултет, Универзитет Гоце Делчев, Штип;
[14] Musharraf, R.C., Martinez-Vazquez J., Timofeev A., (2012). Fiscal Decentralization for Loca Development: An Integral Study, International Center for Public Policy, Atlanta, USA;
[15] Ruopp M.D., Perkins N.J., Whitcomb B.W., Schisterman E.F., (2008) Youden index and optimal cut-off point estimated from observations affected by a lower limit of detection" PMC;
[16] Harris J.K., (2021) "Primer on binary logistic regression", PMC PubMed Central;
[17] Bobbitt Z. (2020) "The 6 assumptions of logistic regression", Statology.com.