

УДК 911.3

Ю.С. Брайчевський, к.геогр.н.,
кафедра країнознавства та туризму, географічний факультет,
Київський національний університет імені Тараса Шевченка

ВИКОРИСТАННЯ МЕТОДУ РЕГРЕСІЙНОГО АНАЛІЗУ У СУСПІЛЬНО-ГЕОГРАФІЧНИХ ДОСЛІДЖЕННЯХ НА ПРИКЛАДІ ВИВЧЕННЯ ЕЛЕКТОРАЛЬНОЇ ПОВЕДІНКИ

Ключові слова: *множинна регресія, суспільна географія, причинно-наслідкові зв'язки, статистичні методи.*

Сучасний методичний апарат вітчизняної суспільної географії вражає кількістю складних і різноманітних аналітичних засобів, у т.ч. математико-статистичних прийомів обробки даних. Разом з тим можна констатувати, що вітчизняні дослідження досі лишаються на периферії міжнародного суспільно-географічного дискурсу. Однією з причин цього є певна невідповідність вітчизняного методичного апарату стандартам щодо обґрунтування і доведення наукових висновків, зокрема коли мова іде про обробку емпіричного матеріалу.

Постановка завдання. У даній роботі пропонується звернути увагу на метод регресійного аналізу, що є одним із найбільш поширених прийомів статистичної перевірки причинно-наслідкових зв'язків між досліджуваним об'єктом та гіпотетичними чинниками, що впливають на нього. Можливий спосіб застосування даного методу пропонується розглянути на прикладі стандартного завдання електоральної географії щодо визначення чинників електоральної поведінки населення.

Аналіз публікацій. Проблемам методичного забезпечення суспільно-географічних досліджень присвячено чимало робіт (наприклад Ю.Г.Саушкін, М.Д.Пістун [2], О.Г.Топчієв [4], О.І.Шаблій [6] та ін.). Разом з тим, дослідженню методичних проблем суспільної географії на прикладному рівні приділялося значно менше уваги. Серед вітчизняних робіт, що зверталися до даної проблематики, можна виділити праці О.І. Шаблія [5], О.Г.Топчієва [4], К.В.Мезенцева [1]. Практичні аспекти застосування математико-статистичних методів для потреб перевірки причинно-наслідкових зв'язків у сфері електоральної географії розглядав В.Б.Шишацький [7].

Серед західних авторів проблемам застосування регресійних моделей у суспільних науках приділено чимало уваги, наприклад у роботах А.О.Сайкса [11] Г.Кінга [9] та ін.

Виклад основного матеріалу. Як відомо, приблизно з середини ХХ ст. суспільна географія зіткнулася з проблемою конкуренції з суміжними суспільствознавчими дисциплінами аналітичного спрямування (економікою,

соціологією, політологією тощо). Відправною точкою у цьому стало рішення президента Гарвардського університету Дж. Конанта, котрий у 1948 р. заявив що географія не є дисципліною університетського рівня і виключив її з переліку спеціальностей університету, заклавши відповідний факультет.

Спроби відновити суспільну роль географії у вищій школі та на науковій ниві у країнах Заходу відбувалися у двох паралельних напрямках. Перший передбачав гуманізацію географічного знання і посилення уваги до проблем соціального характеру і критики чинних суспільно-економічних структур (Д.Харві, П.Тейлор, Е.Свінгедоу та ін.). Методологічно даний напрям тяжів до гуманітарних наук і спирався переважно на методи якісного аналізу, тексти часто мали нормативний характер і носили форму суспільного дискурсу. Другий напрям орієнтувався на посилення апарату методів кількісного аналізу, у т.ч. математико-статистичних методів. У змістовому відношенні, найбільш поширеними стали завдання пов'язані з ГІС технологіями, а також політико-географічні, зокрема електорально-географічні дослідження (Дж. О'Лафлін, Р.Джонстон та ін.).

Формування вітчизняних методологічних підходів у суспільній географії у другій половині ХХ ст. відбувалося паралельно і значною мірою незалежно від західної школи. Оскільки провідним напрямом радянської суспільної географії була економічна географія, гуманізація майже не торкнулася радянської географії. Натомість, використання математико-статистичних методів ставало все більш популярним. На момент, коли почала формуватися українська школа суспільної географії, у розпорядженні географів був доволі широкий спектр методичних напрацювань щодо застосування математико-статистичних методів [5].

Порівняння методичного апарату західної та вітчизняної суспільної географії звертає увагу на важливі відмінності між ними. Застосування математико-статистичних методів вітчизняними географами зводиться, переважно, до репрезентації просторових форм організації досліджуваних об'єктів і явищ, або їх типології та класифікації. Іншими словами, розвиваються різні методи картографічного моделювання, вивчення просторової концентрації/диференціації об'єктів та явищ, групування об'єктів або їх характеристик. В останні роки все більш популярними стають факторний та кластерний аналіз, що використовуються з метою суспільно-географічного районування.

Натомість західна суспільна географія, якщо не враховувати спеціалізовані напрями пов'язані з ГІС, переважно використовує математичні методи для статистичної перевірки змістових гіпотез аналогічно до того, як це робиться у суміжних суспільствознавчих дисциплінах. Здебільшого, мова іде про побудову багатфакторних моделей множинної регресії.

Цікаво, що попри значно ширший спектр методів обробки статистики, використовуваний українськими географами, тестування гіпотез за допомогою моделей множинної регресії практично не застосовується. До певної міри, це пояснюється домінуванням системно-структурного підходу, та орієнтацією на типізацію та класифікацію як основне завдання досліджень. З іншого боку, стандартне застосування регресійної моделі не має безпосередньої прив'язки до змінних простору, тому може інтерпретуватися як негеографічне.

Поширений серед українських географів метод просторово-регресійного аналізу ґрунтується на тій же математичній основі, що і метод багатофакторної регресії. Відмінність полягає лише у тому, що у просторово-регресійному аналізі незалежними змінними є географічні координати, а отже метод перетворюється зі способу перевірки причинно-наслідкових зв'язків на ще одну форму генералізованого картографічного представлення статистичних даних.

До проблеми статистичної перевірки гіпотез зверталися вітчизняні електоральні географи, зокрема В.Б. Шишацький [7, 8]. Здебільшого, у таких випадках застосовувався простий кореляційний аналіз, тобто обчислення парних кореляцій між гіпотетичним чинником та явищем, що вивчається. Наприклад, обрахунок парних кореляцій між голосуванням за політичну партію та окремими соціально-демографічними характеристиками населення.

Перевага моделі множинної регресії перед парними кореляціями полягає у можливості будувати багатофакторні моделі і визначати відносну силу впливу кожного чинника. Основи використання даного методу для виявлення причинно-наслідкових зв'язків добре розроблені серед західних дослідників (наприклад, А.О.Сайкс [11]). Напрями сучасного пошуку щодо застосування методу множинної регресії пов'язані з вирішенням проблем недосконалих вихідних даних, нелінійності тощо (наприклад, праці Г. Кінга та ін. [9]).

Нижче пропонується короткий приклад застосування моделі множинної регресії у електорально-географічному аналізі результатів виборів до Верховної ради України у 2012 р. Даний приклад є ілюстративним і не ставить за мету надати вичерпну відповідь щодо головних чинників, що визначали хід голосування.

Загальний вигляд моделі множинної регресії подається наступним чином:

$$Y = \alpha + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n + \xi,$$

Де Y – результат голосування за певну партію, X_n – гіпотетичний чинник, вплив якого перевіряється, β - коефіцієнт, що визначає інтенсивність впливу конкретного чинника X на підтримку даної партії, α - константа (мінімальна кількість голосів, що віддаються за дану партію не залежно від зовнішніх чинників), ξ - статистичний «шум», вплив випадкових чинників, непередбачуваних обставин, або чинників, які з певних причин не можуть бути включені до моделі, або пропущені дослідником. Множина територіальних одиниць, по яких наявні дані голосування і соціально-демографічні

характеристики, вплив яких аналізується, складає масив спостережень, на підставі яких будується регресійна модель. Якщо припустити, що ми врахували найбільш істотні чинники, то величина ξ є випадковою, а отже при достатньо великій кількості спостережень її математичне очікування прямуватиме до нуля. Відповідно, завданням регресійного аналізу є визначення параметрів β , що визначають рівень впливу гіпотетичних чинників (соціально-демографічних характеристик) на голосування за ту чи іншу партію, а також перевірити, наскільки статистично значимими є одержані результати. Для перевірки статистичної значимості використовується показник p , що відображає, яка частка спостережень (у даному випадку територіальних одиниць, по яких вивчається голосування) не вписується в побудовану модель. Зазвичай, вважається, що при значенні p , що не перевищує 0.05, результати є статистично значимими, тобто отримані результати у достатньому наближенні описують понад 95% спостережень.

Окремої уваги заслуговує характеристика параметру β . Він може обчислюватися безпосередньо на основі вихідних даних, і у такому разі коефіцієнт β може використовуватися безпосередньо для прогнозних розрахунків. Водночас, якщо множина незалежних змінних X_i представлена різнорозмірними величинами, тобто такими, що мають різні одиниці виміру, зведення коефіцієнтів β у одну формулу сукупного впливу буде некоректним. Тому часто у суспільних науках застосовують стандартизоване значення цього параметру. У програмних пакетах, що виконують множинну регресію, часто подаються два значення, просте і стандартизоване, тобто таке, що враховує лише відносну роль чинника стосовно інших факторів впливу. Зокрема у пакеті STATISTICA розраховує одночасно параметри B (просте значення) і β – стандартизоване.

Розглянемо спосіб застосування цього методу на прикладі результатів голосування за політичні партії на виборах до Верховної ради України у 2012 р. на рівні областей. Зазначимо, що чим більший розмір масиву спостережень, тим більш статистично значимими є результати. Тому аналіз на рівні виборчих округів був би суттєво точнішим, ніж на рівні областей. Проте, для потреб ілюстрації можливостей методу, з урахуванням легшого доступу до статистичних даних, аналіз подається на обласному рівні.

Гіпотетичними соціально-демографічними характеристиками населення, які могли б мати вплив на хід голосування за політичні партії, пропонується взяти: кількість осіб з вищою освітою (на 1000 мешканців), відсоток населення, що вважає рідну мову українською, середній дохід на душу населення, відсоток населення у віці старше 50 років (становлення світогляду яких відбувалося у радянський час), населення виборчого віку до 35 р. (покоління, світогляд яких формувалася після проголошення незалежності). Перевіряється статистичний

зв'язок цих характеристик із голосуванням за Партію регіонів, «Батьківщину», «Удар» та КПУ (табл.1).

Таблиця 1

Регресійний аналіз результатів голосування за окремі політичні партії, 2012р.

Гіпотетичні чинники	ПР		"Батьківщина"		УДАР	
	β	р-рівень	β	р-рівень	β	р-рівень
<i>Вища освіта</i>	-0.70	0.00	0.03	0.89	-0.35	0.11
<i>Рідна українська</i>	-1.10	0.00	0.91	0.00	0.73	0.00
<i>Дохід на д.н.</i>	0.13	0.33	-0.04	0.81	0.02	0.90
<i>Літні люди</i>	0.48	0.04	-0.03	0.91	0.68	0.03
<i>Молодь</i>	0.57	0.06	0.00	0.99	1.28	0.00

(продовження таблиці 1)

Гіпотетичні чинники	КПУ		"Свобода"	
	β	р-рівень	β	р-рівень
<i>Вища освіта</i>	0.19	0.50	0.96	0.01
<i>Рідна українська</i>	-0.72	0.00	0.83	0.00
<i>Дохід на д.н.</i>	-0.20	0.38	0.05	0.84
<i>Літні люди</i>	0.37	0.34	-1.36	0.00
<i>Молодь</i>	0.04	0.93	-1.41	0.02

Обчислено на основі даних ЦВК, Всеукраїнського перепису населення та Державного комітету статистики України

Отже виявлено, що для Партії регіонів найбільш статистично значущими чинниками є відсоток літніх людей, а також зворотна залежність від української мови як рідної та вищої освіти. «Батьківщина» здобула найбільшу підтримку в україномовних регіонах. «УДАР» здобув підтримку у областях з відносно вищим відсотком як молоді так і літніх людей, тоді як «Свобода», вочевидь, отримала підтримку від виборців середнього віку. Обидві партії підтримуються україномовним населенням. Крім того, свободу більше підтримують регіони з більшою часткою громадян з вищою освітою.

Попри очевидну зручність використання даного методу, існує цілий ряд практичних проблем з його застосуванням. Найбільш поширеними проблемами є не обов'язкова наявність причинних зв'язків за умови кореляційних залежностей, складність із пошуком підходящих вихідних даних (так, у запропонованому прикладі соціально-демографічні характеристики взяті із Всеукраїнського перепису населення 2001 р., тобто фактично є застарілими,

тоді як свіжіші джерела є недоступними). Ще однією проблемою, згаданою Дж. О'Лафліном [10], є невідповідність територіальних одиниць, по яких збирається статистика, тим, що становлять інтерес для дослідника-географа. Часто проблемою є нелінійний характер залежності або недостатня кількість спостережень для статистичного обґрунтованих результатів. Вирішенню цих проблем присвячуються роботи ряду провідних методологів світу [9].

Висновки. Наведений приклад відображає лише часткові можливості моделі множинної регресії, зокрема оминаються увагою можливості перевірки якісних характеристик за допомогою так званих «фіктивних змінних»; також, поза увагою даної роботи лишаються більш складні прийоми застосування цього методу, у випадках, коли стандартна процедура не може бути коректно використана. Разом з тим, вважаємо, що дана модель має серйозний потенціал для виведення вітчизняних географічних досліджень на якісно новий рівень і перехід суспільно-географічних студій від описово-репрезентативної до пояснювально-прогностичної функції.

Джерела:

1. Мезенцев К.В. Про використання факторного аналізу в регіональних дослідженнях. // Економічна та соціальна географія. – Випуск 53. – 2002. – с. 21-27.
2. Пістун М. Д. Основи теорії суспільної географії: Навч. посіб.: — К.: Вища шк., 1996. – 231 с.
3. Саушкин Ю.Г. Экономическая география: история, теория, методы, практика. - М.: Мысль, 1973 – 559 с.
4. Топчієв О.Г. Суспільно-географічні дослідження: методологія, методи, методика. Навчальний посібник. — Одеса: Астропринт, 2005. — 632с.
5. Шаблій О.І. Математичні методи в соціально-економічній географії. – Львів: Світ, 1994. – 304 с.
6. Шаблій О.І. Суспільна географія: теорія, історія, українознавчі студії. – Львів : Львівський національний університет імені Івана Франка, 2001. – 744 с.
7. Шишацький В.Б. Вплив етнолінгвістичного фактору на електорально-географічний простір України. // Регіональні проблеми України: Збірник наукових праць. — Херсон, 2007. – с.340-349.
8. Шишацький В.Б. Електоральні регіони України: підходи до виділення та делімітації // «Наукові записки» інституту етнонаціональних досліджень НАН України. — К., 2004. Випуск 27. – с. 198-209.
9. Blackwell, M., Honaker J., and King G. In Press. A Unified Approach to Measurement Error and Missing Data: Overview // *Sociological Methods and Research*. In press. [Електр. ресурс]. – Режим доступу: <http://gking.harvard.edu/publications/multiple-overimputation-unified-approach-measurement-error-and-missing-data>

10. O'Loughlin J. The regional factor in contemporary Ukrainian politics: scale, place, space, or bogus effect? // *Post Soviet Geography And Economics* . – 2001. – Vol. 42., no.1: 1-33.
11. Sykes, A. O. "An introduction to regression analysis." (1993). [Електр. ресурс]. – Режим доступу: http://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1050&context=law_and_economics

Брайчевський Ю.С. Використання методу регресійного аналізу у суспільно-географічних дослідженнях на прикладі вивчення електоральної поведінки

На прикладі електорально-географічного аналізу поведінки виборців, розглядається спосіб застосування методу множинної регресії для потреб суспільної географії як засобу статистичної перевірки причинно-наслідкових зв'язків між об'єктом дослідження та гіпотетичними чинниками, що мають на нього вплив.

Ключові слова: множинна регресія, суспільна географія, причинно-наслідкові зв'язки, статистичні методи.

Брайчевский Ю.С. Использование метода регрессионного анализа в общественно-географических исследованиях на примере изучения электорального поведения

На примере электорально-географического анализа поведения избирателей рассматривается способ применения метода множественной регрессии для нужд общественной географии как средства статистической проверки причинно-следственных связей между объектом исследования и гипотетическими факторами, имеющими на него влияние.

Ключевые слова: множественная регрессия, общественная география, причинно-следственные связи, статистические методы.

Braychevskyy Y. The Use of Regression Analysis in Human-Geographic Studies: The Case of Electoral Behavior Analysis

Using electoral behavior analysis as a case, the article demonstrates the potential of regression analysis in human-geographic studies as a means to statistically test causal relations between the object of study and the hypothetical factors that influence it.

Key words: multiple regression, human geography, causal relations, statistical methods.