

ФАКТОРНИЙ АНАЛІЗ ДЕПОПУЛЯЦІЇ НАСЕЛЕННЯ УКРАЇНИ

Анотація. Проведено факторний аналіз депопуляції населення та побудовано багатofакторну модель чисельності населення України.

Аннотация. Проведен факторный анализ депопуляции населения и построена многофакторная модель численности населения Украины.

Annotation. In this paper the factor analysis of depopulation is performed and a multi-factor model of the population size of Ukraine is constructed.

Ключові слова: факторний аналіз, депопуляція, природний рух населення, механічний рух населення.

Соціально-економічні процеси та явища залежать від великої кількості параметрів, які їх характеризують, що зумовлює труднощі, пов'язані з виявленням структури взаємозв'язків цих параметрів. Це змушує відбирати з них найбільш істотні та вивчати менший набір показників. Таке рішення може бути забезпечене методами зниження розмірності, до яких відноситься факторний аналіз [1]. Оскільки процес депопуляції населення залежить від багатьох соціально-економічних факторів, тому для дослідження впливу цих факторів на депопуляційні процеси доцільно використовувати саме факторний аналіз.

На основі динаміки показників кількості народжених, померлих, прибулих та вибулих за 1994 – 2011 роки побудуємо факторну модель чисельності населення. Варто зауважити, що для дослідження депопуляції населення України розглянуто період депопуляції, тобто період скорочення чисельності населення, який розпочався з 1994 року.

Для визначення кількості факторів, потрібно перевірити три критерії: критерій Кайзера, який полягає у визначенні власних значень факторів – значимими вважаються фактори, власні значення яких більші за одиницю, критерій накопиченого відсотка пояснювальної дисперсії, згідно з яким рекомендовано виділяти стільки факторів, щоб накопичений відсоток пояснювальної ними дисперсії перевищував 80 %, та критерій кам'янистого осипу Кеттела.

Розрахунок власних значень зображений на рис. 1.

Eigenvalues (Факторный анализ (нов).sta)				
Extraction: Principal components				
Value	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	2,340606	58,51515	2,340606	58,51515
2	1,120304	28,00759	3,460910	86,52275

Рис. 1. Власні значення показників руху населення України

У процесі визначення власних значень факторів (рис. 1), які показують, наскільки впливовий той чи інший фактор, а також, наскільки цей фактор впливає на зміну всієї сукупності ознак [2], було виділено 2 фактори із власними значеннями, більшими за одиницю. Накопичений відсоток пояснювальної дисперсії цих двох факторів склав 86,5 %. Тобто, на основі цих двох критеріїв слід виділити 2 фактори. Але після представлення даних у вигляді графіка кам'янистого осипу, було вирішено виділити 2 або 3 фактора відповідно до перегину крутої частини графіка.

Перевіримо варіанти моделі з двома і трьома факторами. Розглянемо факторні навантаження моделі з двох факторів без використання методу обертання. Результат зображений на рис. 2.

Variable	Factor Loadings (Unrotated) (Фактор Extraction: Principal components (Marked loadings are >.700000))	
	Factor 1	Factor 2
Кількість народжених, тис. осіб	0,456974	0,782615
Кількість померлих, тис. осіб	0,521220	-0,706972
Кількість прибулих, тис. осіб	0,981571	0,067614
Кількість вибулих, тис. осіб	0,946905	-0,058628
Expl. Var	2,340606	1,120304
Prp. Totl	0,585152	0,280076

Рис. 2. Факторні навантаження для двох факторів без обертання

На рис. 2. видно чіткий розподіл показників на 2 фактори: в перший увійшли показники кількості прибулих та вибулих, а в другий – народжених та померлих. Спробуємо виконати цю процедуру ще раз але з використанням різних методів обертання. В результаті показник кількості померлих виявився не включеним ні в один фактор. Якщо визначити факторні навантаження показників при розподілі на 3 фактори з використанням методу обертання варімакс, який полягає в максимізації дисперсії нової змінної (фактора) та мінімізації розкиду навколо неї, виявиться, що показники кількості прибулого і вибулого населення об'єднані в один фактор, а показники кількості народжених і померлих розділені на два окремих фактора. Для більш коректної інтерпретації результатів аналізу слід використати розподіл показників на два фактори без обертання. В такому випадку отримаємо 2 фактори: "Механічний рух населення" та "Природний рух населення". Саме на основі такого розподілу формуємо рівняння за факторами.

Рівняння першого фактора матиме вигляд:

$$\Phi_{(1)} = 0,9815x_3 + 0,9469x_4 \quad (1)$$

Рівняння другого фактора матиме вигляд:

$$\Phi_{(2)} = 0,7826x_1 - 0,7069x_2 \quad (2)$$

Використовуючи рівняння (1) та (2), розрахуємо значення факторів та на їх основі побудуємо багатфакторну регресійну модель. Результати наведені на рис. 3.

Regression Summary for Dependent Variable: Чисельність насе						
R= ,99998439 R²= ,99996878 Adjusted R²= ,99996462						
F(2,15)=2402E2 p<0,0000 Std.Error of estimate: 11,863						
N=18	Beta	Std.Err. of Beta	B	Std.Err. of B	t(15)	p-level
Intercept			2228,863	141,3209	15,7716	0,000000
Механічний рух населення (фактор 1)	0,041959	0,003053	0,257	0,0187	13,7433	0,000000
Природний рух населення (фактор 2)	0,962809	0,003053	1,227	0,0039	315,3612	0,000000

Рис. 3. Побудова багатофакторної моделі депопуляції населення України

Згідно з результатами, зображеними на рис. 3, модель має вигляд:

$$y = 2228,8630,257\Phi_{(1)} + 1,227\Phi_{(2)} \quad (3)$$

Як видно з рис. 3, результати моделі мають таку інтерпретацію: $R = 0,9999$ – коефіцієнт множинної кореляції наближений до одиниці показує, що між двома факторними і результативною ознаками існує щільний зв'язок; $R^2 = 0,9998$ – коефіцієнт детермінації моделі свідчить про те, що варіація чисельності населення на 99,98 % зумовлюється двома факторами – природним та механічним рухами. Це означає, що ці фактори суттєво впливають на досліджуваний показник; параметр Adjusted $R^2 = 0,9999$ – скоригований на кількість спостережень і кількість параметрів коефіцієнт детермінації показує, що на 99,99 % дані пояснюють модель регресії; $F(2,15) = 2402$ – критерій адекватності Фішера в даному випадку перевищує теоретичне значення ($F_{\text{теор.}} = 3,68$) при рівні значущості $\alpha = 0,05$, можна зробити висновок про статистичну значущість рівняння регресії; Std. Error of estimate = 11,863 – середньоквадратичне відхилення помилок моделі. Середньоквадратичне відхилення параметрів моделі показане у стовпці Std. Error of B, а у стовпці t (15) – значимість параметрів моделі за критерієм Ст'юдента. Оскільки всі розраховані значення критерія Ст'юдента перевищують допустиме критичне значення 4,22 при рівні значущості $\alpha = 0,05$, то параметри моделі слід вважати значущими. Отже, модель можна вважати якісною та адекватною.

Факторний аналіз дозволив просто і точно описати спостережувані вихідні дані, структуру і характер взаємозв'язків між ними. Стискання інформації вийшло за рахунок того, що використовуване число факторів або головних компонент для побудови багатофакторної регресійної моделі скоротилося без впливу на її адекватність. Отже, для регулювання депопуляційних процесів необхідно звертати увагу саме на природний та механічний рух населення та чинники, що їх зумовлюють.

Наук. керівн. Аксьонова І. В.

Література: 1. Буреева М. М. Многомерный статистический анализ с использованием ППП STATISTICA / Буреева М. М. – М. : Н. Новгород, 2007. – 112 с. 2. Боровиков В. П. STATISTICA. Искусство анализа данных на компьютере / Боровиков В. П. – М. : Питер, 2003. – 688 с.