

Секция «Модели и технологии цифровой трансформации экономики»

КАНОНИЧЕСКИЙ АНАЛИЗ МЕДИЦИНСКИХ И ЭКОЛОГИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Г.О. Абросимова

Научный руководитель А.Ю.Трусова
Самарский национальный исследовательский университет
имени академика С.П. Королева.

Актуальность работы: Влияние показателей загрязнений окружающей среды на количество заболеваний.

Научная новизна: Применения метода канонического анализа для исследования взаимосвязей между медицинскими и экологическими показателями.

Практическая значимость: Изучить взаимосвязь медицинских показателей по заболеваемости и экологических показателей по загрязнению окружающей среды.

Цель исследования: Изучить взаимосвязь между количеством заболеваний и экологическими показателями.

Задачи исследования:

1. Исследовать масштабы загрязнении окружающей среды и уровень заболеваемости людей.
2. Сформировать исходные массивы.
3. Изучить алгоритм канонического анализа.
4. Проанализировать полученные результаты.

Канонический анализ – способ получения информации из матриц взаимной корреляции.

$$(1) C = R_{22}^{-1} R_{21} R_{11}^{-1} R_{12},$$

$$(2) A = \frac{S_{-1}^{11} S_{12} B}{\lambda},$$

$$(3) \chi^2 = -n - 1 - k - \frac{1}{2} (p + q + 1) \ln W_k,$$

$$(4) W_k = \prod_{i=k+1}^p (1 - \lambda_i^2),$$

$$(5) a_n = \frac{b_n}{\sigma X_n},$$

$$(6) b_n = \frac{b_n}{\sigma Y_n},$$

X_1 - Объем сброса загрязненных сточных вод по Российской Федерации, млн. м²

X_2 - Отходы от производства по Российской Федерации, тыс. тонн

X_3 - Выбросы, загрязняющие атмосферу по Российской Федерации, тыс. тонн

X_4 - Выбросы парниковых газов по Российской Федерации, CO₂e

Y_1 - Болезни крови, кроветворных органов по Российской Федерации, тыс. чел

Y_2 - Болезни органов дыхания по Российской Федерации, тыс. чел

Y_3 - Врожденные аномалии и хромосомные нарушения по Российской Федерации,

тыс. чел

Таблица 1. Исходный массив данных

Года	X1	X2	X3	X4	Y1	Y2	Y3
2000	20,3	2613,5	32301	1970,6	551	46170	214
2001	19,8	2644,3	33291	2019,4	563	43012	211
2002	19,8	3035,5	33930	2100,9	731	43005	241
...
2019	12,6	9017,3	22735	2225,3	613	52278	292
2020	11,8	8409,0	22228	2236,9	480	54273	243
2021	11,6	8721,2	22300	2248,6	514	59382	253

Таблица 2. Матрица парных коэффициентов корреляции.

	X1	X2	X3	X4	X5	X6	X7
X1	1,00	-0,96	0,75	-0,92	0,29	-0,88	-0,55
X2	-0,96	1,00	-0,77	0,91	-0,35	0,90	0,50
X3	0,75	-0,77	1,00	-0,70	0,66	-0,83	-0,08
X4	-0,92	0,91	-0,70	1,00	-0,18	0,84	0,60
X5	0,29	-0,35	0,66	-0,18	1,00	-0,54	0,50
X6	-0,88	0,90	-0,83	0,84	-0,54	1,00	0,34
X7	-0,55	0,50	-0,08	0,60	0,50	0,34	1,00

Таблица 3. Матрица C, собственное значение и собственные векторы.

$C = R_{22}^{-1} R_{21} R_{11}^{-1} R_{12}$		
0,39	-0,127	0,282
-0,264	0,733	0,401
0,141	0,188	0,361
		r
λ_1	0,573	0,757
λ_2	0,877	0,936
λ_3	0,029	0,17
b1	b2	b3
1,53	-0,16	-1,1
-0,008	3,003	-0,97
1	1	1

Таблица 4. Векторы коэффициентов A.

$A = \frac{S_{-1}^{11} S_{12} B}{\lambda}$		
A1	A2	A3
-1,873	-0,999	-7,242
-1,517	1,507	0,133
3,96	-0,447	1,336
2,981	1,003	-6,342

$$U_1 = -1,87 X_1 - 1,51 X_2 + 3,96 X_3 + 2,98 X_4$$

$$V_1 = 1,53 Y_1 - 0,008 Y_2 + Y_3$$

$$U_2 = -0,99 X_1 + 1,50 X_2 - 0,44 X_3 + 1,57 X_4$$

$$V_2 = -0,16 Y_1 + 3,003 Y_2 + Y_3$$

$$U_3 = -7,24 X_1 + 0,13 X_2 + 1,33 X_3 - 6,34 X_4$$

$$V_3 = -1,1 Y_1 - 0,97 Y_2 + Y_3$$

Таблица 5. Оценка значимости канонических корреляций.

Но:r1=0		Но:r2=0		Но:r3=0	
W0	0,051	W1	0,123	W2	0,97
X ² расп	50,59	X ² расп	33,53	X ² расп	1,756
X ² табл	5,23	X ² табл	1,64	X ² табл	1,64
$\chi_{расп}^2 > \chi_{табл}^2$		$\chi_{расп}^2 > \chi_{табл}^2$		$\chi_{расп}^2 > \chi_{табл}^2$	

Таблица 6. Расчет коэффициентов канонических переменных.

σ	$a_n = \frac{b_n}{\sigma X_n} \quad b_n = \frac{b_n}{\sigma Y_n}$			
2,52	a1'	-0,743	-0,396	-2,873
2039,78	a2'	-0,001	0,001	0,000
3970,27	a3'	0,001	0,000	0,000
78,64	a4'	0,038	0,013	-0,081
76,28	b1'	0,020	-0,002	-0,014
4579,83	b2'	0,000	0,001	0,000
30,67	b3'	0,033	0,033	0,033

Выводы по работе:

1. Исследованы сведения о загрязнении окружающей среды и количестве заболевших людей.
2. Сформированы исходные массивы дынных.
3. Изучен алгоритм канонического анализа.
4. Проанализированы полученные результаты.

Анализ канонических переменных позволил выявить тесную взаимосвязь между:

1. $X_4 = 1,57$ (выбросы парниковых газов) и $Y_2 = 3,003$ (болезнями органов дыхания), $r = 0,757$
2. $X_3 = 3,69$ (выбросы загрязняющих атмосферу) и $Y_1 = 1,53$ (болезни крови, кроветворных органов), $r = 0,936$
3. $X_3 = 1,33$ (выбросы загрязняющих атмосферу) и $Y_3 = 1$ (врожденные аномалии и хромосомные нарушения), $r = 0,170$

Список использованных источников

1. Сошникова Л.А., Тамашевич В. Н., Уебе Г., Шефер М. Многомерный статистический анализ в экономике: Учеб. пособие для вузов // Под ред. проф. В. Н. Тамашевича. – М.: ЮНИТИ-ДАНА, 1999. – 598 с.
2. Федеральная служба государственной статистики. – 2019. [Электронный ресурс] URL: <https://rosstat.gov.ru/statistics/incomparisons>