

ИСПОЛЬЗОВАНИЕ МЕТОДА КАНОНИЧЕСКИХ КОРРЕЛЯЦИЙ В ИЗУЧЕНИИ ВЗАИМОСВЯЗИ ДЕМОГРАФИЧЕСКИХ И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

В связи со сложной демографической ситуацией, важной задачей становится изучение влияния социально-экономических факторов на демографические показатели. Для анализа демографических результатов в странах мира за 2012 год был выбран метод канонических корреляций, сущность которого состоит в анализе связей между двумя множествами: результативным и факторным. Для оценки взаимосвязи демографических и социально-экономических показателей, в статье используется один из показателей ранговой корреляции, коэффициент Спирмена. Он позволяет определить тесноту и направление корреляционной связи между двумя признаками или двумя профилями признаков.

С целью углубленного исследования связей и взаимозависимостей экономических явлений используются статистические методы, которые выражают количественное соотношение между явлениями-результатами и явлениями-причинами.

При анализе взаимосвязей демографических и социально-экономических явлений, как правило, выясняется, что на результативный признак влияет ряд факторных признаков. При этом следует помнить, что влияние всех факторов невозможно учесть по ряду причин: часть факторов может быть неизвестна, по некоторым факторам может отсутствовать достоверная информация или количество включаемых в модель факторов может быть ограничено небольшим объемом выборки.

Формирование результативных демографических показателей связано с воздействием большого числа самых различных факторов, действующих как в одном, так и в противоположном направлении. Для количественной оценки факторов, влияющих на уровень результативных показателей, важную роль играет метод канонических корреляций.

Метод канонических корреляций позволяет находить максимальные корреляционные связи между двумя группами случайных величин. Эта зависимость определяется при помощи новых аргументов канонических величин, вычисленных как линейные комбинации исходных признаков.

Новые канонические величины выбираются таким образом, чтобы новые координаты непосредственно указывали значение корреляции. В каждой группе отыскиваются линейные комбинации исходных величин, имеющие максимальную корреляцию, они и являются первыми координатами новых систем. После этого в каждой группе рассматриваются следующие линейные комбинации, у которых корреляция больше, чем между любыми другими линейными комбинациями, не коррелированными с первыми линейными

комбинациями. Построение продолжается до тех пор, пока не будут полностью получены две новые координатные системы.

Метод канонических корреляций – статистический метод анализа связей между массовыми общественными явлениями и процессами, применяемые в том случае, когда рассматриваются несколько независимых переменных, т.е. канонический корреляционный анализ можно рассматривать как вариант распространения парной корреляции на случай двух многомерных величин [1, с. 139].

Важнейшим достоинством метода канонических корреляция является то, что при его применении не требуется подтверждение отсутствия корреляции как в группе зависимых переменных (Y_k), так и в группе независимых переменных (X_j) [2, с. 63].

Целью применения метода является поиск максимальных корреляционных связей между факторными и результативными переменными.

В настоящее время, в связи со сложной демографической ситуацией, важной задачей становится изучение влияния социально-экономических факторов на демографические показатели. Для анализа демографических результатов в странах мира за 2012 год был выбран метод канонических корреляций, сущность которого состоит в анализе связей между двумя множествами: результативным и факторным. Преимуществами такого анализа являются большой объем выборки, не требуемое подтверждения отсутствия мультиколлинеарности как в группе зависимых, так и в группе независимых переменных.

Результативное множество включает основные показатели характеризующие демографическую ситуацию ста семидесяти стран мира за 2012 год:

- Y_1 – суммарный коэффициент рождаемости, чел.
- Y_2 – общий коэффициент смертности, ‰;
- Y_3 – суммарный коэффициент демографической зависимости, на 100 чел. в возрасте 15–64 лет;

Факторное множество состоит из следующих социально-экономических показателей:

- X_1 – ожидаемая продолжительность жизни при рождении, лет;
- X_2 – валовой национальный доход на душу населения, в долл. США 2005 г. по ППС;
- X_3 – средняя продолжительность обучения, лет;
- X_4 – доля занятых в общей численности населения;

Начинать анализ, целесообразнее с расчета первого канонического коэффициента корреляции и проверки его значимости с помощью χ^2 -критерия Бартлетта. Критерий сферичности Бартлетта – показатель, с помощью которого проверяют, отличаются ли корреляции от 0. Если g близко к нулю, то выбранная переменная не взаимосвязана с другими. Значимость меньше 0,05 указываает, на то что проведение метода канонических корреляций приемлемо.

На следующем этапе оценим коэффициенты канонических переменных для стандартизованных значений исходных переменных X_1, X_2, X_3, X_4 и Y_1, Y_2, Y_3 .

На основании результатов расчетов, можно определить какая из линейных комбинаций индексов позволяет наилучшим образом предсказать сводный индекс результативных показателей. Все необходимые расчеты выполнены с использованием пакета программ STATISTICA.

В программе Statistica метод канонических корреляций реализуется с помощью модуля Canonical Analis. В первую очередь необходимо выбрать, имена переменных принадлежащих первому и второму множеству анализируемых признаков.

В данном случае, к первому множеству были отнесены независимые переменные, а именно: X_1, X_2, X_3, X_4 , а ко второму – зависимые переменные (Y_1, Y_2, Y_3).

Одна из основных задач анализа канонических корреляций, заключается в том, чтобы найти такую пару значений канонических переменных, которой будет соответствовать максимальный канонический коэффициент корреляции.

Для вычисления канонических коэффициентов корреляции необходимо, прежде всего, определить матрицы ковариаций исходных переменных.

Таким образом, перед началом анализа необходимо просмотреть матрицу парных корреляций для исходных переменных. Из нее видны первые результаты канонического анализа: максимальный канонический коэффициент корреляции $R = 0,899$, а его оценка по χ^2 -критерию ($\chi^2 = 366,79$). Таким образом, показатели результативного множества на 89,9 % зависят от показателей факторного множества.

Для дальнейшего анализа, необходимо последовательно выбрать процедуры Eigenvalues (собственные значения), Canonical weights, left & right set (канонические веса для левого и правого множества) [2, с. 63].

С помощью данной операции были вычислены три собственных числа матрицы C : $\lambda_1^2 = 0,809$ $\lambda_2^2 = 0,313$ и $\lambda_3^2 = 0,109$. Кроме того, показаны два варианта (Root 1 и Root 2) канонических весовых коэффициентов для переменных обоих множеств.

Результатом анализа является оцененная теснота связи между новыми каноническими переменными U и V , представляющими собой линейные комбинации исходных факторных и зависимых переменных [1, с.140].

$$\begin{cases} U = a_1X_1 + a_2X_2 + \dots + a_pX_p; \\ V = b_1Y_1 + b_2Y_2 + \dots + b_pY_p. \end{cases} \quad (1)$$

Также, на основании рассчитанных данных можно записать, как выглядят канонические переменные и соответствующие им канонические коэффициенты корреляции.

Итак, максимальный коэффициент канонической корреляции равен 0,899. Ему соответствует пара канонических переменных:

$$\begin{cases} U_1 = -0,876\hat{X}_1 + 0,513\hat{X}_2 - 0,489\hat{X}_3 - 0,645\hat{X}_4 \\ V_1 = 1,296\hat{Y}_1 + 0,461\hat{Y}_2 - 0,622\hat{Y}_3; \end{cases} \quad \chi^2=366,79;$$

Второму коэффициенту канонической корреляции $r_2 = 0,56$ соответствует пара канонических переменных:

$$\begin{cases} U_2 = -1,211 \hat{X}_1 + 0,244 \hat{X}_2 + 1,321 \hat{X}_3 - 0,23 \hat{X}_4 \\ V_2 = 1,16 \hat{Y}_1 + 1,037 \hat{Y}_2 - 1,062 \hat{Y}_3; \end{cases} \quad \chi^2=83,97;$$

Третьему коэффициенту канонической корреляции $r_3 = 0,33$ соответствует пара канонических переменных:

$$\begin{cases} U_3 = -0,061 \hat{X}_1 + 1,167 \hat{X}_2 - 0,496 \hat{X}_3 + 0,363 \hat{X}_4 \\ V_3 = 2,594 \hat{Y}_1 - 0,147 \hat{Y}_2 - 2,748 \hat{Y}_3; \end{cases} \quad \chi^2=19,7;$$

Для проверки гипотезы о значимости канонических коэффициентов необходимо сравнить расчетные значения, которые представлены на рисунке 1, с табличными для уровня значимости $\alpha = 0,01$ $\chi^2_{кр.} = 26,2$ для двенадцати степеней свободы, $\chi^2_{кр.} = 16,80$ для шести степеней свободы и $\chi^2_{кр.} = 9,2$ для двух степеней свободы, т.е. проверяемая гипотеза о равенстве канонического коэффициента корреляции нулю отвергается. Следовательно, все канонические коэффициенты корреляции значимы. Связь между множествами переменных (Y_1, Y_2, Y_3) и (X_1, X_2, X_3, X_4) сильная, так как $r_1 = 0,899$.

| Root Removed | Chi-Square Tests with Successive Roots Removed (Spreadsheet) | | | | | |
|--------------|--|------------------|----------|----|----------|--------------|
| | Canonical R | Canonical R-sqr. | Chi-sqr. | df | p | Lambda Prime |
| 0 | 0.899284 | 0.808711 | 366.7943 | 12 | 0,000000 | 0,117069 |
| 1 | 0.559712 | 0.313277 | 83.9655 | 6 | 0,000000 | 0,611998 |
| 2 | 0.329869 | 0.108814 | 19.6995 | 2 | 0,000053 | 0,891186 |

Рисунок 1 – Канонические коэффициенты корреляции и их оценки

В результате анализа, был выбран только один корень (вариант системы канонических переменных), так как он значим по критерию χ^2 и связь между множествами переменных сильная.

Полученная система канонических переменных имеет вид:

$$\begin{cases} U_1 = -0,876 \hat{X}_1 + 0,513 \hat{X}_2 - 0,489 \hat{X}_3 - 0,645 \hat{X}_4 \\ V_1 = 1,296 \hat{Y}_1 + 0,461 \hat{Y}_2 - 0,622 \hat{Y}_3; \end{cases}$$

Итак, исходя из проведенного канонического корреляционного анализа можно сделать вывод, что на демографические показатели в большей мере оказывают влияние ожидаемая продолжительность жизни при рождении и доля занятых в общей численности населения, а в наименьшей – валовой национальный доход на душу населения и средняя продолжительность обучения.

Наиболее информативными переменными из результативного множества оказались суммарный коэффициент рождаемости и суммарный коэффициент демографической зависимости, а общий коэффициент смертности имеет меньшую информационную нагрузку.

Также, для оценки взаимосвязи демографических и социально-экономических показателей используются показатели ранговой корреляции.

Один из показателей ранговой корреляции, коэффициент Спирмена позволяет определить тесноту и направление корреляционной связи между двумя признаками или двумя профилями признаков.

Для подсчета ранговой корреляции Спирмена необходимо располагать двумя рядами значений, которые могут быть проранжированы. Такими рядами значений могут быть:

1. Два признака, измеренные в одной и той же группе испытуемых;
2. Две индивидуальные иерархии признаков, выявленные у двух испытуемых по одному и тому же набору признаков;
3. Две групповые иерархии признаков;
4. Индивидуальная и групповая иерархии признаков [3, с. 35].

Вначале показатели ранжируются отдельно по каждому из признаков. Как правило, меньшему значению признака начисляется меньший ранг.

Ограничения коэффициента ранговой корреляции:

1. По каждой переменной должно быть представлено не менее 5 наблюдений;
2. Коэффициент ранговой корреляции Спирмена при большом количестве одинаковых рангов по одной или обоим сопоставляемым переменным дает огрубленные значения. В идеале оба коррелируемых ряда должны представлять собой две последовательности несовпадающих значений.

Практический расчет коэффициента ранговой корреляции Спирмена включает следующие этапы:

1. Сопоставить каждому из признаков их порядковый номер (ранг) по возрастанию (или убыванию).
2. Определить разности рангов каждой пары сопоставляемых значений.
3. Возвести в квадрат каждую разность и суммировать полученные результаты.
4. Вычислить коэффициент корреляции рангов по формуле [3, с. 37]:

$$r = \frac{6 \sum d^2}{n(n^2 - 1)} \quad (2)$$

где $\sum d^2$ – сумма квадратов разностей рангов,
 n – число парных наблюдений.

При использовании коэффициента ранговой корреляции условно оценивают тесноту связи между признаками, считая значения коэффициента равные 0,3 и менее, показателями слабой тесноты связи; значения более 0,4, но менее 0,7 – показателями умеренной тесноты связи, а значения 0,7 и более – показателями высокой тесноты связи.

Мощность коэффициента ранговой корреляции Спирмена несколько уступает мощности параметрического коэффициента корреляции.

В качестве исходных данных, выступают индекс человеческого развития и суммарный коэффициент рождаемости по ста семидесяти девяти странам мира.

В результате анализа, оказалось, что коэффициент Спирмена равен $-0,814$. Таким образом, можно сделать вывод, что зависимость между индексом человеческого развития и суммарным коэффициентом рождаемости высокая и обратная. То есть, в слаборазвитых странах коэффициент рождаемости значительно выше, нежели в высокоразвитых.

Это связано в первую очередь с тем, что сегодня женщины высокоразвитых стран рожают в более позднем возрасте. Жительницы городов и мегаполисов, сначала хотят выйти замуж, построить дом, путешествовать по миру, и получать хорошую заработную плату.

Женщины также хотят владеть собственностью, и в идеале жить недалеко от своих родителей. Им нужен хороший автомобиль и хорошая должность на работе до создания семьи, говорит новое исследование. Однако главный приоритет традиционный – вступить в брак до рождения ребенка.

Можно привести целый ряд факторов, которые приводят к снижению рождаемости при повышении уровня развития страны:

1. Рост уровня образования женщин: процесс образования и построения карьеры отнимает у женщины значительную часть репродуктивного периода.
2. Повышение уровня жизни: родители не решаются заводить столько детей, как в прошлом, потому что опасаются, что не смогут надлежащим образом их обеспечить.
3. Пенсионное обеспечение: людям больше не нужно, чтобы их старость обеспечивали многочисленные дети.
4. Низкий уровень детской смертности: люди перестают рожать «про запас».
5. Новые смыслы и интересы: разнообразие общественной жизни позволяет части людей найти иные пути самореализации помимо рождения детей.

В целом этот процесс снижения рождаемости совершенно нормален. Он соответствует переходу земной цивилизации и стадии экспоненциального роста к стадии стабилизации, поскольку размеры земного шара ограничены. В развивающихся странах, возможно, в ближайшие десятилетия тоже произойдет демографический переход и рождаемость сократится.

Список использованных источников

1. Многомерный статистический анализ в экономике: учеб. пособие для вузов / Л.А. Сошникова [и др.]; под ред. проф. В.Н. Тамашевича. – М.:ЮНИТИ-ДАНА, 1999. – 598 с.
2. Многомерная обработка экономических данных с использованием интегрированной системы STATISTICA: учеб.-метод. пособие / В.В. Лабоцкий. – Мн.:БГЭУ, 2002. – 95 с.
3. Хоменко Л.Н. Социальная статистика: учеб.-метод. пособие / Л.Н. Хоменко. – Минск: БГЭУ, 2009. – 179 с.