

УДК 314.12

Пеленёва Елена Алексеевна, соискатель

Факультет прикладной математики и механики,

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Пермский национальный исследовательский политехнический университет»

e-mail: pelenevaelena7@mail.ru

ОЖИДАЕМАЯ ПРОДОЛЖИТЕЛЬНОСТЬ ЖИЗНИ ПРИ РОЖДЕНИИ В ЮЖНЫХ СТРАНАХ

Аннотация: В ходе работы используются процедуры многомерного статистического анализа. Предложены факторы, влияющие на ожидаемую продолжительность жизни в южных странах. Проведен анализ взаимосвязи ожидаемой продолжительности жизни от различных показателей. Проведен корреляционно-регрессионный анализ данных и построено уравнение регрессии для выявления главных факторов, влияющих на ожидаемую продолжительность жизни. Проведен кластерный анализ данных, с помощью которого разбили совокупность всех объектов на однородные группы (кластеры). Выявлены главные факторы, влияющие на ожидаемую продолжительность жизни.

Ключевые слова: средняя ожидаемая продолжительность жизни, корреляционно-регрессионный анализ, кластерный анализ, уровень значимости, коэффициент корреляции, уравнение регрессии.

Abstract: In the course of the procedure used multivariate statistical analysis. Suggested factors affecting life expectancy in the southern countries. The analysis of the relationship of life expectancy of the various indicators. Correlation and regression analysis of the data and the regression equation is constructed to identify the main factors influencing the life expectancy. Spend a cluster analysis of the data by which the aggregate of all objects broke into homogeneous groups (clusters). It identified the main factors that affect life expectancy.

Keywords: the average life expectancy, correlation and regression analysis, cluster analysis, the level of significance, correlation coefficient, the regression equation.

При изучении статистики различных стран мира невозможно не заметить, что существует колоссальная разница между странами по нескольким показателям. Одним из индикаторов уровня жизни населения той или иной страны является средняя продолжительность жизни населения. Часто ее связывают с уровнем ВВП на душу населения. Во многих случаях это оправдано, но, например, в Индии ВВП на душу населения в 2014 году был на 2190\$ больше, чем в Китае, но средняя продолжительность жизни на семь лет меньше. Становится ясно, что за длительностью жизни стоит гораздо больше факторов.

Целью работы является выявление и анализ факторов, влияющих на ожидаемую продолжительность жизни в южных странах.

Рассмотрим характеристики, которые могут влиять на ожидаемую продолжительность жизни в южных странах:

- Y - Средняя ожидаемая продолжительность жизни при рождении, лет;
- X1 - ВВП в паритетах покупательной способности (на душу населения);
- X2 - Темпы прироста населения по сравнению с предыдущим годом, %;
- X3 - Потребление алкоголя в литрах на душу населения;
- X4 - Коэффициент младенческой смертности на 1000 детей.

Исследуем зависимости средней ожидаемой продолжительности жизни при рождении от параметров с помощью Корреляционно-регрессионного анализа. Построив матрицу парных коэффициентов корреляции (уровень значимости $\alpha = 0,5$) и матрицу частных коэффициентов корреляции, найдем коэффициенты, которые являются значимыми согласно критерию Стьюдента. Таким образом, можно сделать вывод, что на изменение средней ожидаемой продолжительности жизни при рождении (Y) в большей степени влияют ВВП в паритетах покупательной способности (X1), темпы прироста населения по сравнению с предыдущим годом (X2) и Коэффициент младенческой смертности (X4).

Построим уравнения регрессии, выбрав в качестве результативного показателя среднюю ожидаемую продолжительность жизни при рождении (Y) (Таблица 1):

Таблица 1 – Итоги регрессии для зависимой переменной

Итоги регрессии для зависимой переменной: Средняя ожидаемая продолжительность жизни при рождении, лет (Y) R= ,94226065 R2= ,88785514 Скорректир. R2= ,87426183 F(4,33)=65,316 p						
	Бета	Стд.Ош Бета	В	Стд.Ош. В	t(33)	p-уров
Св.член			78,26925	2,134509	36,66849	0,000000
ВВП в паритетах покупательной способности (на душу населения) (X1)	0,178201	0,074154	0,00039	0,000163	2,40312	0,022037
Темпы прироста населения по сравнению с предыдущим годом,% (X2)	-0,291455	0,075900	-2,77114	0,721654	-3,83998	0,000529
Потребление алкоголя в литрах на душу населения (X3)	-0,012796	0,058448	-0,04039	0,184465	-0,21893	0,828050
Коэффициент младенческой смертности на 1000 детей (X4)	-0,612218	0,079031	-0,20704	0,026727	-7,74653	0,000000

Таким образом, уравнение регрессии по компонентам выглядит следующим образом:

$$Y = 78,26925 + 0,00039 * X1 - 2,77114 * X2 - 0,04039 * X3 - 0,20704 * X4 \quad (*)$$

в безразмерном виде:

$$Y = 0,178201 * X1 - 0,291455 * X2 - 0,012796 * X3 - 0,612218 * X4.$$

Коэффициенты уравнения регрессии (*) показывают на сколько ед. изменится результат при изменении факторов на 1 ед. Коэффициенты при X1, X2, X3, X4 показывают среднее изменение результативного показателя Y. В таблице красным выделены значимые коэффициенты, следовательно, можно сделать вывод что, ВВП в паритетах покупательной способности (X1), темпы прироста населения по сравнению с предыдущим годом (X2) и коэффициент младенческой смертности (X4) влияют на изменение средней ожидаемой продолжительности жизни при рождении (Y) в большей степени.

Поскольку $R^2 (Y, X1, X2, X3, X4, X5, X6, X7) = 0,88785514$, то можно утверждать, что линейная модель хорошо аппроксимирует исходные данные.

Проведем кластеризацию данных. Для этого обезразмерим исходные данные с помощью стандартной нормировки $z = \frac{x - \bar{x}}{\sigma_x}$. Сгруппируем объекты в три кластера

(Рис.1):

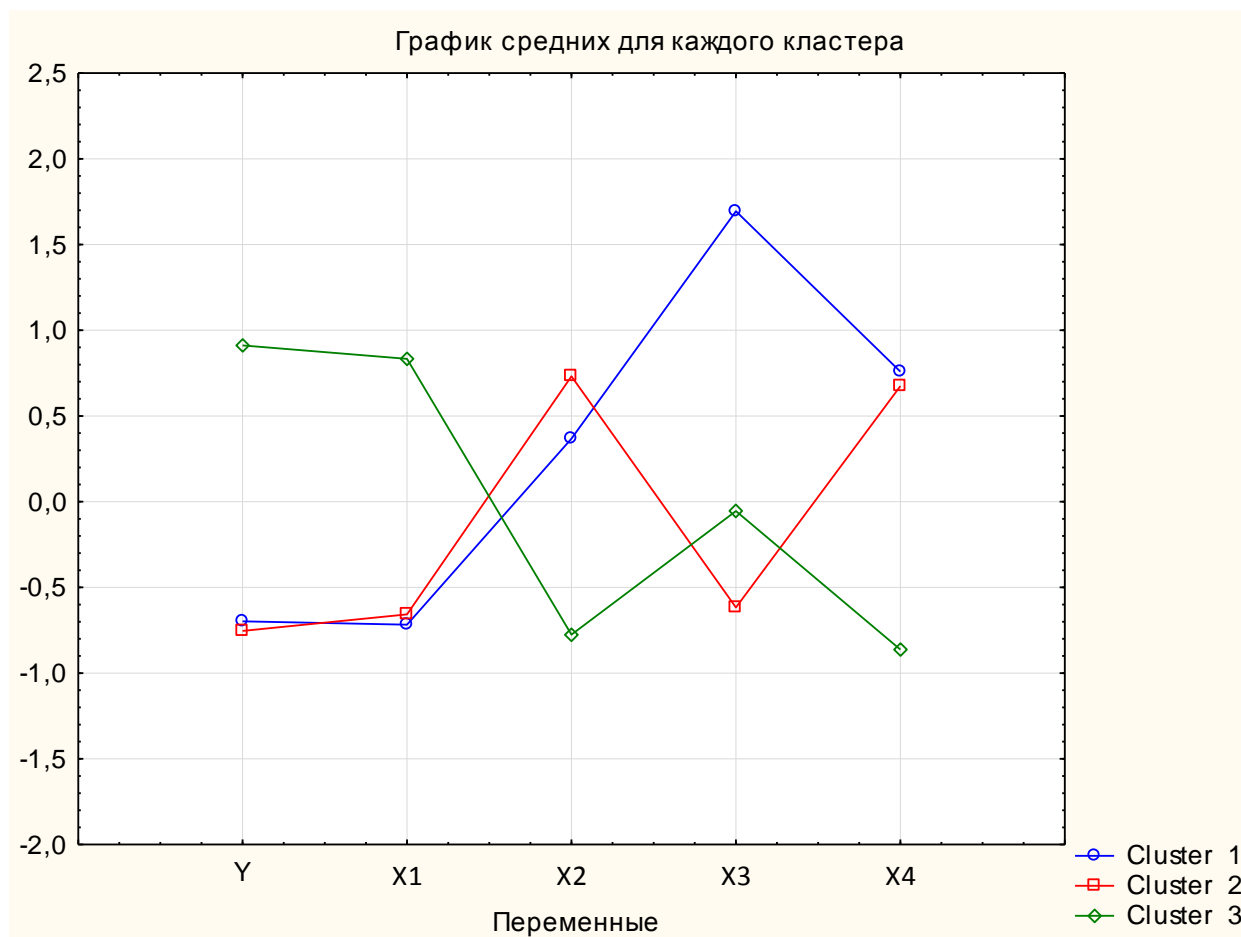


Рисунок 1 – График средних для каждого кластера

По выделенным кластерам выведем зависимости, применяя корреляционно-регрессионный анализ.

Кластер 1 имеет следующие характеристики: ниже среднего ожидаемая продолжительность жизни при рождении, низкий ВВП в паритетах покупательной способности, выше среднего темпы прироста населения по сравнению с предыдущим годом, высокие показатели потребления алкоголя, высокий коэффициент младенческой смертности.

Построив матрицу парных коэффициентов корреляции (уровень значимости $\alpha = 0,5$) и матрицу частных коэффициентов корреляции, найдем коэффициенты, которые являются значимыми согласно критерию Стьюдента. Таким образом, можно сделать вывод, что из матрицы парных коэффициентов корреляции следует, что на изменение средней ожидаемой продолжительности жизни при рождении (Y) влияет Коэффициент младенческой смертности (X4). Но при построение частных коэффициентов корреляции видно, что это ложная корреляция.

Построим уравнения регрессии, выбрав в качестве результативного показателя среднюю ожидаемую продолжительность жизни при рождении (Y) (Таблица 2):

Таблица 2 – Итоги регрессии для зависимой переменной

Итоги регрессии для зависимой переменной: Средняя ожидаемая продолжительность жизни при рождении, лет (Y) R ² =,81040506						
	Бета	Стд.Ош Бета	B	Стд.Ош. B	t(33)	p-уров
Св.член			0,96350	3,203835	0,30073	0,814025
ВВП в паритетах покупательной способности (на душу населения) (X1)	0,221706	0,953077	0,30403	1,306986	0,23262	0,854496
Темпы прироста населения по сравнению с предыдущим годом,% (X2)	0,192439	1,314480	0,14201	0,969984	0,14640	0,907457
Потребление алкоголя в литрах на душу населения (X3)	-0,245308	0,920451	-0,32490	1,219104	-0,26651	0,834190
Коэффициент младенческой смертности на 1000 детей (X4)	-0,962403	0,903587	-1,24549	1,169376	-1,06509	0,479940

Таким образом, уравнение регрессии по компонентам выглядит следующим образом:

$$Y = 0,96350 + 0,30403 * X1 + 0,14201 * X2 - 0,32490 * X3 - 1,24549 * X4 \quad (*)$$

в безразмерном виде:

$$Y = 0,221706 * X1 + 0,192439 * X2 - 0,245308 * X3 - 0,962403 * X4 .$$

Коэффициенты уравнения регрессии (*) показывают на сколько ед. изменится результат при изменении факторов на 1 ед. Коэффициенты при X1, X2, X3, X4 показывают, среднее изменение результативного показателя Y. В таблице нет красных выделений, следовательно можно сделать вывод что, ничего не влияет на Среднюю ожидаемую продолжительность жизни (Y) в большей степени.

Поскольку R² (Y,X1,X2,X3,X4,X5,X6,X7)= 0,81040506, то можно утверждать, что линейная модель хорошо аппроксимирует исходные данные.

Кластер 2 имеет следующие характеристики: низкие показатели средней ожидаемой продолжительности жизни при рождении, низкие показатели ВВП в паритетах покупательной способности, высокие темпы прироста населения по сравнению с предыдущим годом, низкие показатели потребления алкоголя, высокий коэффициент младенческой смертности.

Построив матрицу парных коэффициентов корреляции (уровень значимости $\alpha = 0,5$) и матрицу частных коэффициентов корреляции, найдем коэффициенты, которые являются значимыми согласно критерию Стьюдента. Таким образом, можно сделать вывод, что на изменение средней ожидаемой продолжительности жизни при рождении (Y) в большей степени влияет Коэффициент младенческой смертности (X4).

Построим уравнения регрессии, выбрав в качестве результативного показателя среднюю ожидаемую продолжительность жизни при рождении (Y) (Таблица 3):

Таблица 3 – Итоги регрессии для зависимой переменной

Итоги регрессии для зависимой переменной: Средняя ожидаемая продолжительность жизни при рождении, лет (Y) R2=,70004584						
	Бета	Стд.Ош Бета	В	Стд.Ош. В	t(33)	p-уров
Св. член			-0,158585	0,327424	-0,48434	0,638577
ВВП в паритетах покупательной способности (на душу населения) (X1)	0,160667	0,199573	0,238483	0,296233	0,80505	0,439511
Темпы прироста населения по сравнению с предыдущим годом,% (X2)	-0,358550	0,200001	-0,345717	0,192842	-1,79274	0,103265
Потребление алкоголя в литрах на душу населения (X3)	-0,211154	0,212377	-0,333423	0,335355	-0,99424	0,343554
Коэффициент младенческой смертности на 1000 детей (X4)	-0,721924	0,177145	-0,580308	0,142396	-4,07532	0,002231

Таким образом, уравнение регрессии по компонентам выглядит следующим образом:

$$Y = -0,158585 + 0,238483 * X1 - 0,345717 * X2 - 0,333423 * X3 - 0,580308 * X4 \quad (*)$$

в безразмерном виде:

$$Y = 0,160667 * X1 - 0,358550 * X2 - 0,211154 * X3 - 0,721924 * X4 .$$

Коэффициенты уравнения регрессии (*) показывают на сколько ед. изменится результат при изменении факторов на 1 ед. Коэффициенты при X1, X2, X3, X4 показывают среднее изменение результативного показателя Y. В таблице красным выделены значимые коэффициенты, следовательно, можно сделать вывод что Коэффициент младенческой смертности (X4) влияют на изменение Средней ожидаемой продолжительности жизни при рождении (Y) в большей степени.

Поскольку $R^2 (Y, X_1, X_2, X_3, X_4, X_5, X_6, X_7) = 0,70004584$, то можно утверждать, что линейная модель хорошо аппроксимирует исходные данные.

Кластер 3 имеет следующие характеристики: высокая продолжительность жизни при рождении, высокий ВВП в паритетах покупательной способности, низкий темп прироста населения по сравнению с предыдущим годом, среднее потребление алкоголя, низкий коэффициент младенческой смертности.

Построив матрицу парных коэффициентов корреляции (уровень значимости $\alpha = 0,5$) и матрицу частных коэффициентов корреляции, найдем коэффициенты, которые являются значимыми согласно критерию Стьюдента. Таким образом, можно сделать вывод, из матрицы парных коэффициентов корреляции видно, что на изменение средней ожидаемой продолжительности жизни при рождении (Y) в большей степени влияют ВВП в паритетах покупательной способности (X_1) и Коэффициент младенческой смертности (X_4). Но построив матрицу частных, видно что ложная корреляция присутствует между Средней ожидаемой продолжительности жизни при рождении (Y_1) и Коэффициента младенческой смертности (X_4).

Построим уравнения регрессии, выбрав в качестве результативного показателя среднюю ожидаемую продолжительность жизни при рождении (Y) (Таблица 4):

Таблица 4 – Итоги регрессии для зависимой переменной

Итоги регрессии для зависимой переменной: Средняя ожидаемая продолжительность жизни при рождении, лет (Y) $R^2 = 0,70004584$						
	Бета	Стд.Ош Бета	В	Стд.Ош. В	t(33)	p-уров
Св.член			0,255949	0,173651	1,47393	0,166244
ВВП в паритетах покупательной способности (на душу населения) (X_1)	0,559733	0,180899	0,213451	0,068985	3,09417	0,009290
Темпы прироста населения по сравнению с предыдущим годом, % (X_2)	-0,284589	0,188358	-0,156608	0,103652	-1,51089	0,156692
Потребление алкоголя в литрах на душу населения (X_3)	0,178625	0,193440	0,072731	0,078763	0,92341	0,373988
Коэффициент младенческой смертности на 1000 детей (X_4)	-0,400187	0,199926	-0,417492	0,208571	-2,00168	0,068453

Таким образом, уравнение регрессии по компонентам выглядит следующим образом:

$$Y = 0,255949 + 0,213451 * X_1 - 0,156608 * X_2 + 0,072731 * X_3 - 0,417492 * X_4 (*)$$

в безразмерном виде:

$$Y=0,559733*X1-0,284589*X2+0,178625*X3-0,400187*X4.$$

Коэффициенты уравнения регрессии (*) показывают на сколько ед. изменится результат при изменении факторов на 1 ед. Коэффициенты при X1, X2, X3, X4 показывают среднее изменение результативного показателя Y. В таблице красным выделены значимые коэффициенты, следовательно, можно сделать вывод, что ВВП в паритетах покупательной способности (X1) влияют на изменение Средней ожидаемой продолжительности жизни при рождении (Y) в большей степени.

Поскольку $R^2 (Y, X1, X2, X3, X4, X5, X6, X7) = 0,67207873$, то можно утверждать, что линейная модель хорошо аппроксимирует исходные данные.

Одним из индикаторов уровня жизни населения той или иной страны является средняя продолжительность жизни населения. Часто ее связывают с ВВП на душу населения. Несомненно, высокоразвитая экономика влияет на продолжительность жизни, но в южных странах, как видно из корреляционно-регрессионного анализа, ситуация складывается несколько иначе. Самыми важными являются высокие показатели ВВП в паритетах покупательной способности, Темп прироста населения по сравнению с предыдущим годом и Коэффициента младенческой смертности. Поделив страны на группы (с помощью кластерного анализа) видно, что ВВП в паритетах покупательной способности (X1) и коэффициента младенческой смертности (X4) влияют на изменение Средней ожидаемой продолжительности жизни при рождении (Y) в большей степени.

Библиографический список

1. Википедия. Список стран по ВВП на душу населения. URL: [https://ru.wikipedia.org/wiki/%D0%A1%D0%BF%D0%B8%D1%81%D0%BE%D0%BA_%D1%81%D1%82%D1%80%D0%B0%D0%BD_%D0%BF%D0%BE_%D0%92%D0%92%D0%9F_\(%D0%9F%D0%9F%D0%A1\)_%D0%BD%D0%B0_%D0%B4%D1%83%D1%88%D1%83_%D0%BD%D0%B0%D1%81%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D1%8F](https://ru.wikipedia.org/wiki/%D0%A1%D0%BF%D0%B8%D1%81%D0%BE%D0%BA_%D1%81%D1%82%D1%80%D0%B0%D0%BD_%D0%BF%D0%BE_%D0%92%D0%92%D0%9F_(%D0%9F%D0%9F%D0%A1)_%D0%BD%D0%B0_%D0%B4%D1%83%D1%88%D1%83_%D0%BD%D0%B0%D1%81%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D1%8F)
2. Все рейтинги. URL: <http://total-rating.ru/24-ozhidaemaya-prodolzhitelnost-zhizni-v-stranah-mira-na-2014-god.html>
3. Все рейтинги. URL: <http://total-rating.ru/23-mladencheskaya-smernost-v-stnanah-mira-2014.html>