

УДК 51

ПРОГНОЗИРОВАНИЕ ЧИСЛЕННОСТИ НАСЕЛЕНИЯ КЕМЕРОВСКОЙ ОБЛАСТИ-КУЗБАСС НА ОСНОВЕ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА

Грибанов Е.Н., к.т.н., доцент
Ковальчук Д.А., студент гр. БЭс-201, III курс
Негореев Е.А., студент гр. БЭс-201, III курс
Кузбасский государственный технический университет
имени Т.Ф. Горбачева,
г. Кемерово

Прогнозирование численности населения является важной задачей в области демографии, экономики, социологии, государственного управления и других областей. Эта задача имеет особую важность для государственных органов, так как позволяет планировать социальную и экономическую политику, распределять ресурсы и оценивать потребности населения [1].

Для прогнозирования численности населения используются различные методы, включая корреляционно-регрессионный анализ. В данной статье будет рассмотрен подход, основанный на корреляционно-регрессионном анализе для прогнозирования численности населения.

Корреляционно-регрессионный анализ является методом, который позволяет определить связь между двумя или более переменными. В данном случае мы исследуем связь между численностью населения и другими переменными, такими как рождаемость, смертность, миграция и т.д.

Для проведения корреляционно-регрессионного анализа мы используем статистические показатели, такие как коэффициент корреляции и коэффициент детерминации. Коэффициент корреляции показывает силу и направление связи между двумя переменными. Коэффициент детерминации же показывает, насколько хорошо наша модель подходит для описания зависимости между переменными.

Применение корреляционно-регрессионного анализа позволяет определить связь между численностью населения и другими переменными. Например, мы можем выявить связь между численностью населения и рождаемостью, смертностью, миграцией и другими факторами.

На основе этих связей мы можем построить модель для прогнозирования численности населения. Например, мы можем использовать модель эконометрического прогнозирования, которая основывается на данных о численности населения, рождаемости, смертности и миграции за прошедшие годы.

Для построения модели нам необходимо определить функциональную зависимость между переменными. Например, мы можем использовать линейную мо-

дель, в которой зависимая переменная (численность населения) связана с одной или несколькими независимыми переменными (рождаемость, смертность, миграция) линейным образом. Эта модель может быть представлена в виде уравнения:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

где Y - зависимая переменная (численность населения);

X_1, X_2, \dots, X_n - независимые переменные (рождаемость, смертность, миграция);

$b_0, b_1, b_2, \dots, b_n$ - коэффициенты регрессии.

Для определения оптимальных значений коэффициентов регрессии мы используем метод наименьших квадратов. Этот метод позволяет найти такие значения коэффициентов, при которых сумма квадратов отклонений прогнозируемых значений от фактических значений будет минимальной.

После определения оптимальных значений коэффициентов регрессии мы можем использовать полученную модель для прогнозирования численности населения на будущие периоды. Для этого мы подставляем значения независимых переменных (рождаемость, смертность, миграция) за будущие годы в уравнение модели и получаем прогнозируемые значения зависимой переменной (численности населения).

Рассмотрим применение данной модели на примере населения Кемеровской области-Кузбасса. В качестве выборки будут взяты данные по численности населения, а также основных демографических показателей:

- рождаемости;
- смертности;
- миграции.

В качестве инструмента для проведения анализа можно использовать программу MicrosoftExcel. В ней имеется весь необходимый набор инструментов для построения зависимости между факторами в выборке. Исходные данные для анализа представлены в таблице 1.

Исходя из данных таблицы 1, можно провести регрессионный анализ с целью установления зависимости между численностью населения, смертностью и рождаемостью. Для этого следует воспользоваться пакетом «Анализ данных» во вкладке «Данные» MSExcel. Среди инструментов данного пакета следует выбрать «регрессию» (см. рис.1).

Таблица 1 – Исходные данные для анализа, чел.[2]

Год	Численность населения	Рождаемость	Смертность
-----	-----------------------	-------------	------------

2002	2 917 769	29369	51534
2003	2 893 448	30383	52808
2004	2 862 105	30542	51234
2005	2 832 963	30862	53152
2006	2 805 941	32060	48870
2007	2 785 999	34242	46816
2008	2 780 231	36675	46121
2009	2 776 358	37599	44856
2010	2 772 964	36370	44522
2011	2 761 255	35002	42624
2012	2 750 829	37798	41615
2013	2 742 450	37245	39703
2014	2 734 075	35992	39763
2015	2 724 990	33933	39355
2016	2 717 627	32704	38872
2017	2 708 844	28314	38151
2018	2 694 877	26540	38748
2019	2 674 256	23989	37882
2020	2 657 854	22601	42946
2021	2 633 446	21602	47589
2022	2 604 272	21804	48325

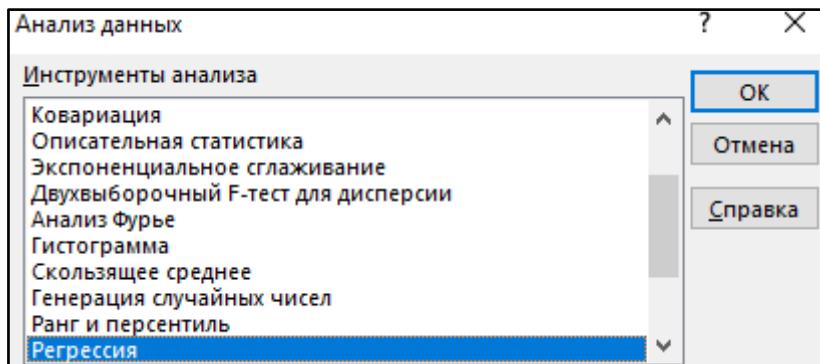


Рисунок 1 – Инструмент «Регрессия» пакета «Анализ данных»

Следующим шагом является установка диапазона для анализа. В нашем примере за X следует взять численность населения, а за диапазон Y–рождаемость и смертность. После этого программа автоматически произведет расчет основных показателей регрессии заданной выборки (см. рис. 2).

ВЫВОД ИТОГОВ					
<i>Регрессионная статистика</i>					
Множественный R	0,808887952				
R-квадрат	0,654299718				
Нормированный R-квадрат	0,615888576				
Стандартная ошибка	50025,68351				
Наблюдения	21				
<i>Дисперсионный анализ</i>					
	df	SS	MS	F	Значимость F
Регрессия	2	85258083702	42629041851	17,03411	7,05176E-05
Остаток	18	45046242192	2502569011		
Итого	20	1,30304E+11			

Рисунок 2 – Показатели регрессии выборки

Как следует из рисунка 2, значение критерия R-квадрат составило 0,65, что указывает на значительное влияние рассмотренных факторов. При этом можно рассмотреть коэффициент корреляции численности населения с каждым из факторов. Для этого, помимо стандартной формулы расчета, можно воспользоваться функцией КОРПРЕЛ. Результаты расчета будут представлены на рисунке 3.

D25					=КОРПРЕЛ(С3:С23;D3:D23)
	A	B	C	D	E
1					
2		Год	Численность населения	Рождаемость	Смертность
3		2002	2 917 769	29369	51534
4		2003	2 893 448	30383	52808
5		2004	2 862 105	30542	51234
6		2005	2 832 963	30862	53152
7		2006	2 805 941	32060	48870
8		2007	2 785 999	34242	46816
9		2008	2 780 231	36675	46121
10		2009	2 776 358	37599	44856
11		2010	2 772 964	36370	44522
12		2011	2 761 255	35002	42624
13		2012	2 750 829	37798	41615
14		2013	2 742 450	37245	39703
15		2014	2 734 075	35992	39763
16		2015	2 724 990	33933	39355
17		2016	2 717 627	32704	38872
18		2017	2 708 844	28314	38151
19		2018	2 694 877	26540	38748
20		2019	2 674 256	23989	37882
21		2020	2 657 854	22601	42946
22		2021	2 633 446	21602	47589
23		2022	2 604 272	21804	48325
24				r	r
25				0,47	0,60

Рисунок 3 – Коэффициенты корреляции между каждым фактором выборки и численностью населения

Из данных рисунка 3 следует, что большее влияние на численность населения оказала смертность, на что указывает значение коэффициента корреляции 0,6, в то время как у рождаемости он составил 0,47. Поэтому, исходя из этого, можно спрогнозировать численность населения Кемеровской области-Кузбасса на 2030 год. Для этого понадобится составить уравнение линейной регрессии. Удобнее всего это сделать с использованием пакета «Анализ данных» [3].

ВЫВОД ИТОГОВ	
<i>Регрессионная статистика</i>	
Множественный R	0,472642089
R-квадрат	0,223390544
Нормированный R-квадрат	0,182516362
Стандартная ошибка	72980,01588
Наблюдения	21
<i>Дисперсионный анализ</i>	
	<i>df</i>
Регрессия	1
Остаток	19
Итого	20
<i>Коэффициенты</i>	
Y-пересечение	2531955,74
Переменная X 1	7,11

Рисунок 4 – Коэффициенты уравнения линейной регрессии

Таким образом, уравнение регрессии будет иметь вид $2531955,7 + 7,11 \cdot x_1$, где в качестве переменной X выступает искомый год. Следовательно, прогнозная численность населения в 2030 году составит:

$$2531955,74 + 7,11 \cdot 2030 = 2546389 \text{ чел.}$$

Подводя итоги, следует отметить, что прогнозирование численности населения на основе корреляционно-регрессионного анализа является важным инструментом для планирования социально-экономической политики и распределения ресурсов. Корреляционно-регрессионный анализ позволяет выявить связь между численностью населения и другими переменными, а метод наименьших квадратов позволяет определить оптимальные значения коэффициентов регрессии для прогнозирования численности населения на будущие периоды. Этот подход может быть полезен для государственных органов, научных исследователей и других заинтересованных в принятии решений, связанных с населением, процессами социально-экономического развития и планирования бюджета.

Однако необходимо понимать, что модель, построенная на основе корреляционно-регрессионного анализа, может быть неполной или недостаточно точной. В этом случае необходимо использовать дополнительные методы прогнозирования и учитывать другие факторы, влияющие на численность населения. Кроме то-

го, необходимо учитывать возможные изменения в экономической, политической и социальной сферах, которые могут повлиять на рождаемость, смертность и миграцию.

Таким образом, прогнозирование численности населения на основе корреляционно-регрессионного анализа является важным инструментом для планирования социально-экономической политики и распределения ресурсов. Однако, для достижения более точных результатов необходимо учитывать другие факторы, влияющие на численность населения, и проводить регулярное обновление и анализ модели для корректировки прогнозов.

Список литературы:

1. Эконометрика : учебник для вузов / И. И. Елисеева [и др.] ; под редакцией И. И. Елисеевой. — Москва : Издательство Юрайт, 2023. — 449 с. — (Высшее образование). — ISBN 978-5-534-00313-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/510472> (дата обращения: 27.03.2023).
2. Официальные статистические показатели. — Текст : электронный // ЕМИСС: государственная статистика : [сайт]. — URL: <https://fedstat.ru/> (дата обращения: 27.03.2023).
3. КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ В EXCEL: ИНСТРУКЦИЯ ВЫПОЛНЕНИЯ. — Текст : электронный // EXCEL table : [сайт]. — URL: <https://exceltable.com/otchety/korrelacionno-regressionnyy-analiz> (дата обращения: 27.03.2023).