

ПОСТРОЕНИЕ МОДЕЛИ МНОЖЕСТВЕННОЙ РЕГРЕССИИ СОСТОЯНИЯ АТМОСФЕРНОГО ВОЗДУХА РЕСПУБЛИКИ БЕЛАРУСЬ

Проблемы атмосферного загрязнения являются глобальными и требуют совместных усилий со стороны государства и организаций. Изучение факторов, влияющих на состояние атмосферы, помогает разрабатывать эффективные меры по их снижению и предотвращению.

Загрязнение атмосферы вызывает значительные экономические потери, связанные с ухудшением здоровья населения, увеличением расходов на медицинское обслуживание, снижением производительности труда и деградацией окружающей среды. Поэтому такой фактор как загрязнение атмосферного воздуха следует учитывать при разработке экономической политики государства.

Безусловно, существует множество индикаторов, отражающих состояние атмосферного воздуха государства, однако в качестве эндогенной переменной было принято решение использовать такой показатель как выбросы загрязняющих веществ в атмосферный воздух (тонны) – Y . Если говорить об экзогенных переменных (регрессорах), то было выбрано 11 показателей:

- образование отходов производства, тыс. тонн – X_1 ;
- количество организаций промышленности, ед. – X_2 ;
- объем промышленности, тыс. р. – X_3 ;
- расход топливно-энергетических ресурсов (котельно-печное топливо), тыс. т условного топлива – X_4 ;
- расход топливно-энергетических ресурсов (электроэнергия), млн киловатт-часов – X_5 ;
- расход топливно-энергетических ресурсов (теплоэнергия), тысяча гигакалорий – X_6 ;
- производство электроэнергии, млн. киловатт-часов – X_7 ;
- число организаций строительства, ед. – X_8 ;
- количество стационарных источников выбросов, ед. – X_9 ;
- количество транспортных средств, ед. – X_{10} ;
- ВВП на душу населения, долл. США – X_{11} .

Для построения модели был выбран прикладной программный пакет «Gretl». Все данные были взяты с сайта Национального статистического комитета Республики Беларусь, что говорит об их реальности [4]. Выборка осуществлялась по годам (2012-2023).

После выбора переменных, был проведен предварительный анализ данных и сделано предположение: все факторы, за исключением X_4 , должны иметь с переменной Y прямую зависимость. Следует построить графики разброса, чтобы подтвердить или опровергнуть раннее сказанные предположения, они представлены на рис. 1.

На графиках 6 и 8 можно заметить обратную зависимость, что не соответствует экономическому смыслу. Исходя из этого, для построения более точной модели, эти переменные были исключены.

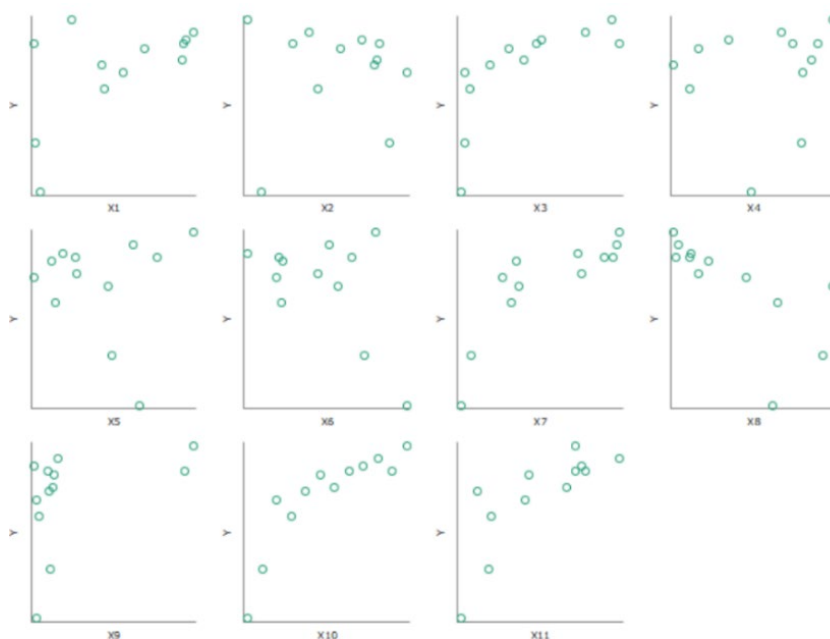


Рис.1. Графики разброса X и Y

Итак, первая модель построена. Ее анализ показал следующее: R-квадрат равен 0,99, в то время как все Р-значения факторов больше уровней значимости (0,01; 0,05; 0,1), на основе чего можно предположить наличие мультиколлинеарности. Для подтверждения ее наличия был проведен анализ корреляционной матрицы (рис. 2).

Ярко-красные и ярко-синие значения (за исключением диагонали) соответствуют большим по модулю значениям коэффициентов парной корреляции, а это как раз и свидетельствует о вероятном наличии мультиколлинеарности. Помимо анализа корреляционной матрицы также использовался метод инфляционных факторов. Значения всех факторов оказались выше 10, что в очередной раз предполагает наличие мультиколлинеарности. Для того чтобы получить надежные результаты, необходимо было исключить мультиколлинеарность, а точнее ликвидировать избыточные переменные.

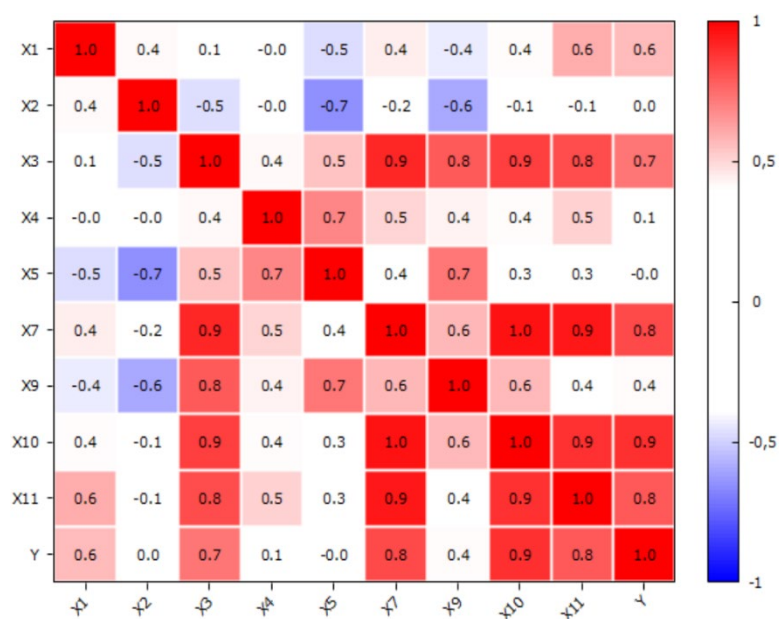


Рис. 2. Корреляционная матрица

Было протестировано множество моделей и, путем исключения факторов X11, X3, X5, X7, X10, была выбрана лучшая. В новой модели можно говорить об отсутствии мультиколлинеарности, что подтверждается с помощью метода инфляционных факторов VIF, представленного на рис. 3 (все значения < 10).

Метод инфляционных факторов	
Минимальное возможное значение = 1.0	
Значения > 10.0 могут указывать на наличие мультиколлинеарности	
X1	1,325
X2	1,761
X4	1,371
X9	2,310

Рис. 3. Метод инфляционных факторов

Новая модель (рис. 4) имеет следующие показатели: коэффициент детерминации близок к 1, что свидетельствует об адекватности модели, Р-значение Фишера меньше всех уровней значимости, на основе чего можно сделать вывод о значимости модели в целом. Р-значения факторов также меньше всех уровней значимости (0,01; 0,05; 0,1), что позволяет сделать заключение о значимости всех коэффициентов.

Модель 5: МНК, использованы наблюдения 2012-2023 (T = 12)					
Зависимая переменная: Y					
	коэффициент	ст. ошибка	t-статистика	p-значение	
const	276,759	41,9581	6,596	0,0003	***
X1	0,00129804	0,000117421	11,05	1,10e-05	***
X2	0,0764840	0,0204359	3,743	0,0072	***
X4	-0,00569463	0,00149674	-3,805	0,0067	***
X9	0,000871871	8,19734e-05	10,64	1,42e-05	***
Среднее завис. перемен 462,5000 Ст. откл. завис. перемен 12,14683					
Сумма кв. остатков 59,26485		Ст. ошибка модели 2,909709			
R-квадрат 0,963484		Исправ. R-квадрат 0,942618			
F(4, 7) 46,17470		P-значение (F) 0,000041			
Лог. правдоподобие -26,60992		Крит. Акаике 63,21984			
Крит. Шварца 65,64437		Крит. Хеннана-Куинна 62,32219			
параметр rho -0,015130		Стат. Дарбина-Уотсона 1,904937			
обратите внимание на сокращенные обозначения статистики					

Рис. 4. Усовершенствованная модель

Полученное уравнение регрессии имеет следующий вид:

$$Y = 276,759 + 0,0013 \cdot X1 + 0,0765 \cdot X2 - 0,0057 \cdot X4 + 0,0009 \cdot X9$$

$a_0 = 276,759$, это означает, что загрязнение атмосферного воздуха увеличивается на 276,759 тыс. тонн под влиянием других факторов, не включенных в модель множественной регрессии.

$a_1 = 0,0013$, показывает, что при увеличении образования отходов производства на 1 ед. загрязнение атмосферного воздуха увеличивается на 0,0013 тыс. тонн.

$a_2 = 0,0765$, это означает, что при увеличении количества организация промышленности на 1 ед., загрязнение атмосферного воздуха увеличивается на 0,0765 тыс. тонн.

$a_4 = -0,0057$, показывает, что при увеличении расхода топливно-энергетических ресурсов (котельно-печного топлива) на 1 ед. загрязнение атмосферного воздуха уменьшается на 0,0057 тыс. тонн.

$a_9 = 0,0009$, показывает, что при увеличении количества стационарных источников выбросов на 1 ед. загрязнение атмосферного воздуха увеличивается на 0,0009 тыс. тонн.

Модель была проверена на правильность выбранной спецификации, гомоскедастичность и автокорреляцию. По результатам проведенных тестов все Р-значения оказались больше уровней значимости (0,01; 0,05; 0,1), поэтому все нулевые гипотезы были приняты: спецификация была выбрана верно (является линейной), гетероскедастичность и автокорреляция отсутствуют.

Для анализа расчетных и наблюдаемых значений был построен график, а также использован анализ средней абсолютной процентной ошибки, ее значение составляет 3,67 %, на основе чего можно сказать, что модель хорошо аппроксимирует фактические данные.

На последнем этапе проведен анализ остатков, для этого был построен график (рис. 5), который является стохастическим. Также был проведен тест на нормальное распределение, и нулевая гипотеза подтвердилась. Эти два фактора свидетельствуют о нормальности распределения остатков.



Рис. 5. График остатков

Построив модель множественной регрессии состояния атмосферного воздуха Республики Беларусь и проанализировав ее, нами было выяснено, что наибольшее влияние на загрязнение воздуха оказывают такие факторы, как образование отходов производства, количество организаций промышленности, расход топливно-энергетических ресурсов (котельно-печное топливо) и количество стационарных источников выбросов. Однако также значительное влияние на состояние воздуха оказывают объем промышленности, количество транспортных средств и ВВП на душу населения, но ввиду мультиколлинеарности эти факторы пришлось исключить, чтобы повысить качество модели.

Список использованных источников

1. Беларусь в цифрах: сайт [Электронный ресурс]. – Режим доступа: https://www.belstat.gov.by/ofitsialnaya/statistika/publications/izdania/public_compilation/index_69861. – Дата доступа: 24.04.2024.

2. Министерство природных ресурсов и охраны окружающей среды Республики Беларусь [Электронный ресурс]. – Режим доступа: <https://minpriroda.gov.by/ru>. – Дата доступа: 24.04.2024.
3. Национальный статистический комитет Республики Беларусь [Электронный ресурс]. – Минск, 1998. – Режим доступа: <https://www.belstat.gov.by/ofitsialnaya-statistika>. – Дата доступа: 24.04.2024.
4. Национальный статистический комитет Республики Беларусь [Электронный ресурс]. – Режим доступа: <https://www.belstat.gov.by/ofitsialnaya-statistika>. – Дата доступа: 24.04.2024.
5. Хацкевич, Г. А. Эконометрика / Г. А. Хацкевич, Т. В. Русилко. – Минск : РИВШ, 2021. – 452 с. – ISBN 978-985-586-480-7.
6. Эколог. бай [Электронный ресурс]. – Режим доступа: <https://www.ecolog.by>. – Дата доступа: 24.04.2024.
7. Экологический портал Республики Беларусь [Электронный ресурс]. – Режим доступа: <https://ecoportal.gov.by>. – Дата доступа: 24.04.2024.