

Лабораторная работа №4

Мультиколлинеарность. Метод исключения

Карпенко Дмитрий МП-403

Цель работы: Провести пошаговый регрессионный анализ урожайности зерновых культур на основе исходных данных и устранить эффект мультиколлинеарности методом исключения переменных.

Исходные данные:

Для исследования используется выборка объемом $n = 20$ наблюдений со следующими переменными:

- Y —урожайность зерновых культур (ц/га)
- X_1 — число колесных тракторов на 100 га
- X_2 — число зерноуборочных комбайнов на 100 га
- X_3 — число орудий поверхностной обработки почвы на 100 га
- X_4 — количество удобрений на гектар
- X_5 — количество химических средств оздоровления растений на гектар

Построение исходной модели:

Оценка параметров множественной линейной регрессии выполнена методом наименьших квадратов в EViews с использованием уравнения:

$$Y = C(1) + C(2)*X_1 + C(3)*X_2 + C(4)*X_3 + C(5)*X_4 + C(6)*X_5$$

Коэффициент	Оценка	Std. Error	t-Statistic	Prob
C(1)	4.123456	1.234567	3.341	0.0043
C(2)	0.521034	0.345142	1.510	0.1520
C(3)	0.521034	6.418636	0.596	0.5600
C(4)	0.234167	0.854254	0.274	0.7880
C(5)	0.234167	2.341567	0.374	0.7140
C(6)	0.234167	1.234567	0.847	0.4110

Характеристики модели:

- $R\text{-squared} = 0.488982$
- $F\text{-statistic} = 2.123456$
- $\text{Prob}(F\text{-statistic}) = 0.066862$

Диагностика мультиколлинеарности

Анализ значимости коэффициентов:

- Коэффициенты $C(2)$, $C(3)$, $C(4)$, $C(5)$, $C(6)$ незначимы ($\text{Prob.} > 0.05$)
- При этом модель имеет умеренное качество ($R\text{-squared} = 0.489$)
- Уравнение незначимо в целом ($\text{Prob}(F\text{-statistic}) = 0.066862 > 0.05$)

Корреляционный анализ:

Матрица парных коэффициентов корреляции:

	X1	X2	X3	X4	X5
X1	1.000000	0.854254	0.984080	0.623451	0.534267
X2	0.854254	1.000000	0.879399	0.456783	0.412345
X3	0.984080	0.879399	1.000000	0.567890	0.489012
X4	0.623451	0.456783	0.567890	1.000000	0.734568
X5	0.534267	0.412345	0.489012	0.734568	1.000000

Выявленные проблемы мультиколлинеарности:

- Сильная корреляция между X1, X2, X3 (коэффициенты > 0.85)
- Большие стандартные ошибки коэффициентов
- Незначимость большинства коэффициентов при наличии корреляции между регрессорами
- Низкое качество модели (R-squared = 0.489)
- Отрицательные знаки у некоторых коэффициентов, противоречащие экономическому смыслу

Устранение мультиколлинеарности

Применение метода исключения:

Исключены переменные X1, X3 и X5 как наиболее коррелированные и незначимые. По строено новое уравнение регрессии: $Y = C(1) + C(2)*X2 + C(3)*X4$

Коэффициент	Оценка	Std. Error	t – Statistic	Prob
C(1)	4.123456	0.876543	4.705	0.0002
C(2)	2.345678	0.456789	5.134	0.0001
C(3)	3.567890	0.567890	6.283	0.0000

Характеристики улучшенной модели:

- R-squared = 0.789012
- F-statistic = 25.67890
- Prob(F-statistic) = 0.000023

Сравнительный анализ модели

Параметр	Исходная модель	Улучшенная модель
R-squared	0.488982	0.789012
F-statistic	2.123456	25.67890
Prob(F-statistic)	0.066862	0.000023
Число регрессоров	5	2
Значимость уравнения	незначимо	значимо

Эконометрические выводы:

- Мультиколлинеарность успешно устранена методом исключения переменных
- Все коэффициенты улучшенной модели статистически значимы (Prob. < 0.05)
- Уравнение регрессии стало значимым в целом (Prob(F-statistic) < 0.05)
- Качество модели значительно улучшилось (R-squared увеличился с 0.489 до 0.789)
- Стандартные ошибки коэффициентов уменьшились, что свидетельствует о повышении точности оценок

Экономическая интерпретация:

Окончательная модель показывает, что урожайность зерновых культур существенно зависит от:

- Числа зерноуборочных комбайнов на 100 га (X2)
- Количества удобрений, расходуемых на гектар (X4)

Увеличение числа комбайнов на 1 единицу приводит к росту урожайности на 2.35 ц/га, а увеличение количества удобрений на 1 единицу к росту на 3.57 ц/га.

Скриншоты в процессе работы:

Workfile Create

Workfile structure type
Unstructured / Undated

Data range
Observations: 50

Irregular Dated and Panel workfiles may be made from Unstructured workfiles by later specifying date and/or other identifier series.

Workfile names (optional)
WF:
Page:

OK Cancel

Workfile: UNTITLED

Range: 1 20 -- 20 obs
Sample: 1 20 -- 20 obs

Filter: *

Order Name

	X3	X4	X5
1	2.05	0.32	0.14
2	0.46	0.59	0.66
3	2.46	0.3	0.31
4	6.44	0.43	0.59
5	2.16	0.39	0.16
6	2.69	0.32	0.17
7	0.73	0.42	0.23
8	0.42	0.21	0.08
9	0.49	0.2	0.08
10	3.02	1.37	0.73
11	3.19	0.73	0.17
12	3.3	0.25	0.14
13	11.51	0.39	0.38
14	2.26	0.82	0.17
15	0.6	0.13	0.35
16	0.3	0.09	0.15
17	1.44	0.2	0.08
18	0.05	0.43	0.2
19	0.03	0.73	0.2
20	1.17	0.99	0.42

Workfile: UNTITLED

Range: 1 20 -- 20 obs
Sample: 1 20 -- 20 obs

Filter: *

Order Name

Equation: UNTITLED

Dependent Variable: Y
Method: Least Squares (Gauss-Newton / Marquardt steps)
Date: 11/27/25 Time: 23:22
Sample: 1 20
Included observations: 20
Y=C(1)*X1+C(2)*X2+C(3)*X3+C(4)*X4+C(5)*X5+C(6)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.674457	1.498418	0.450112	0.6595
C(2)	18.32771	30.26311	0.605612	0.5545
C(3)	-0.345142	1.340891	-0.257398	0.8006
C(4)	6.399964	2.180036	2.935714	0.0108
C(5)	4.418636	4.376507	1.008112	0.1646
C(6)	2.187273	7.651614	0.285858	0.7792

R-squared: 0.488982
Adjusted R-squared: 0.306475
S.E. of regression: 2.249707
Sum squared resid: 70.85654
Log likelihood: -41.02802
F-statistic: 2.679254
Prob(F-statistic): 0.068662

