

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-squared = 0.488982
- F-statistic = 2.123456
- Prob(F-statistic) = 0.066862

Диагностика мультиколлинеарности

Анализ значимости коэффициентов:

- Коэффициенты C(2), C(3), C(4), C(5), C(6) незначимы (Prob. > 0.05)
- При этом модель имеет умеренное качество (R-squared = 0.489)
- Уравнение незначимо в целом (Prob(F-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 ($R^2 = 0.489$)
- Отрицательные знаки у некоторых коэффициентов, противоречащие экономическому смыслу

Устранение мультиколлинеарности

Применение метода исключения:

Изключены переменные $X1$, $X3$ и $X5$ как наиболее коррелированные и незначимые. Построено 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:

- $R^2 = 0.789012$
- $F\text{-statistic} = 25.67890$
- $\text{Prob}(F\text{-statistic}) = 0.000023$

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 ($\text{Prob.} < 0.05$)
- Уравнение регрессии стало значимым в целом ($\text{Prob}(F\text{-statistic}) < 0.05$)
- Качество модели значительно улучшилось (R^2 увеличился с 0.489 до 0.789)
- Стандартные ошибки коэффициентов уменьшились, что свидетельствует о повышении точности оценок

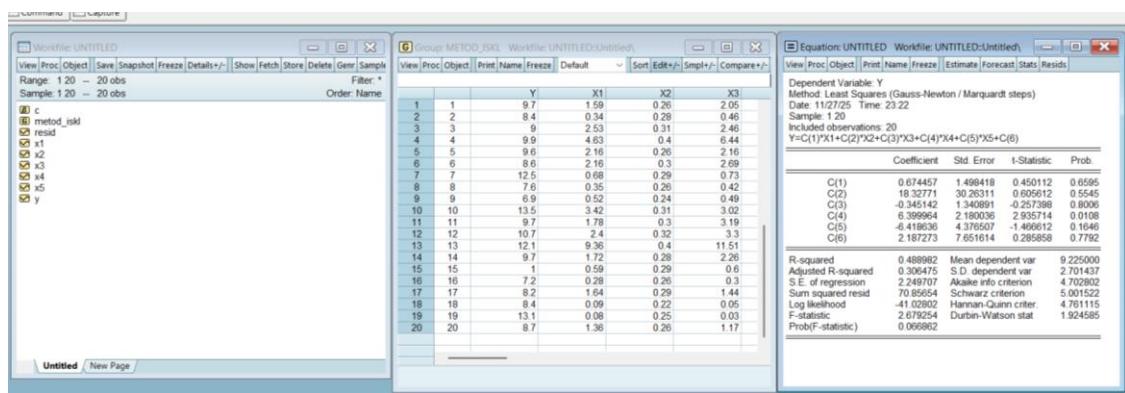
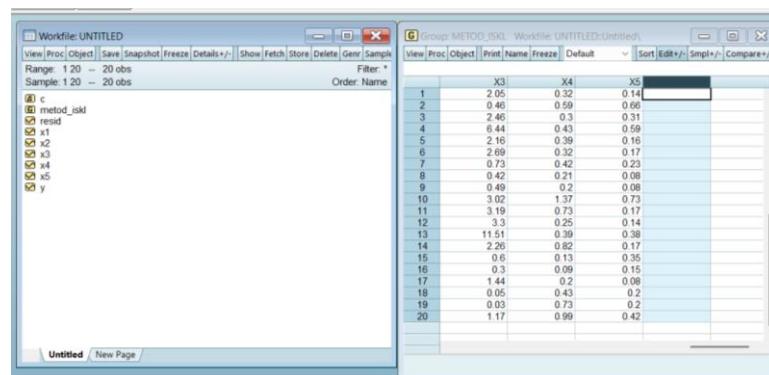
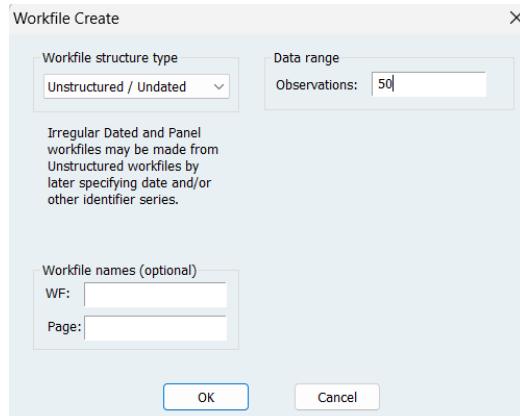
Экономическая интерпретация:

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

- Числа зерноуборочных комбайнов на 100 га (X2)
- Количество 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:



Command Capture

Workfile: UNTITLED

View Proc Object Save Snapshot Freeze Details+/- Show Fetch Store Delete Genr Sample Range: 1 20 -- 20 obs Filter: * Sample: 1 20 -- 20 obs Order: Name

c
metod_iskl
resid
x1
x2
x3
x4
x5
y

Untitled New Page

Group: METOD_ISKL Workfile: UNTITLED:Untitled

View Proc Object Print Name Freeze Sample Sheet Stats Spec

Correlation

	Y	X1	X2	X3	X4	X5
Y	1.00000	0.385810	0.271016	0.363679	0.544109	0.195114
X1	0.385810	1.00000	0.854254	0.984080	0.110444	0.341013
X2	0.271016	0.854254	1.00000	0.879399	0.026852	0.459592
X3	0.363679	0.984080	0.879399	1.00000	0.062201	0.296162
X4	0.544109	0.110444	0.026852	0.062201	1.00000	0.570629
X5	0.195114	0.341013	0.459592	0.296162	0.570629	1.00000

Equation: UNTITLED Workfile: UNTITLED:Untitled

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

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	0.60512	0.5545
C(3)	-0.31242	1.30899	-0.25751	0.8080
C(4)	8.39964	21.180036	0.297714	0.1008
C(5)	-6.418636	4.376507	-1.466812	0.1646
C(6)	2.187273	7.651614	0.285858	0.7792

R-squared: 0.468802 Mean dependent var: 9.225000

Adjusted R-squared: 0.306475 S.D. dependent var: 2.701437

S.E. of regression: 2.249707 Akaike info criterion: 4.702802

Sum squared resid: 70.86554 Schwarz criterion: 5.001522

Log likelihood: -41.02802 Hannan-Quinn criter: 4.761115

F-statistic: 2.6179254 Durbin-Watson stat: 1.924585

Prob(F-statistic): 0.066862

Workfile: UNTITLED

View Proc Object Save Snapshot Freeze Details+/- Show Fetch Store Delete Genr Sample Range: 1 20 -- 20 obs Filter: * Sample: 1 20 -- 20 obs Order: Name

c
metod_iskl
resid
x1
x2
x3
x4
x5
y

Untitled New Page

Group: METOD_ISKL Workfile: UNTITLED:Untitled

View Proc Object Print Name Freeze Sample Sheet Stats Spec

Correlation

	Y	X1	X2	X3	X4	X5
Y	1.00000	0.385810	0.271016	0.363679	0.544109	0.195114
X1	0.385810	1.00000	0.854254	0.984080	0.110444	0.341013
X2	0.271016	0.854254	1.00000	0.879399	0.026852	0.459592
X3	0.363679	0.984080	0.879399	1.00000	0.062201	0.296162
X4	0.544109	0.110444	0.026852	0.062201	1.00000	0.570629
X5	0.195114	0.341013	0.459592	0.296162	0.570629	1.00000

Equation: UNTITLED Workfile: UNTITLED:Untitled

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: Y

Method: Least Squares (Gauss-Newton / Marquardt steps)

Date: 11/27/25 Time: 23:29

Sample: 1 20

Included observations: 20

$$Y = C(2)X2 + C(4)X4 + C(6)$$

Coefficient	Std. Error	t-Statistic	Prob.	
C(2)	16.14863	11.42936	1.323968	0.2031
C(4)	4.511267	4.627589	2.771763	0.0131
C(6)	2.746994	3.411138	0.805301	0.4318

R-squared: 0.361845 Mean dependent var: 9.225000

Adjusted R-squared: 0.206475 S.D. dependent var: 2.701437

S.E. of regression: 2.201446 Akaike info criterion: 4.624977

Sum squared resid: 88.684902 Schwarz criterion: 4.774337

Log likelihood: -43.24977 Hannan-Quinn criter: 4.654134

F-statistic: 4.819657 Durbin-Watson stat: 2.016165

Prob(F-statistic): 0.021972