# Тема 6. Прогнозирование на основе использования эконометрических моделей

Что мы знаем:

1. Спецификация эконометрической модели
2. Сбор исходной информации
3. Вычислительный этап:
   Оценка параметров модели (теорема Гаусса-Маркова)
4. Анализ полученных результатов:
   4.1. Тестирование качества спецификации модели (коэффициент $R^2$, F-тест, проверка $H_0$: $a_i=0$)

   4.2 Исследование модели на мультиколлинеарность

# ПОНЯТИЕ МУЛЬТИКОЛЛИНЕАРНОСТИ

Одно из условий возможности применения МНК – это матрица X должна иметь полный ранг.
Это означает, что все столбцы матрицы коэффициентов системы уравнений наблюдений должны быть линейно-независимыми
Данное условие математически можно записать так:

$$\boldsymbol{rank}(\boldsymbol{X}) = \boldsymbol{rank}(\boldsymbol{X}^{T}\boldsymbol{X}) = \boldsymbol{k} \qquad (3.1)$$

где: k – число столбцов матрицы X (Количество регрессоров в модели +1)

Если среди столбцов матрицы X имеются линейно-зависимые, то **rank(X)<k**
Тогда по свойству определителей

$$\mathbf{det}(\boldsymbol{X}^{T}\boldsymbol{X}) = 0 \qquad (3.2)$$

# Понятие мультиколлинеарности

Условие (3.2) приводит к тому, что матрица $(X^TX)^{-1}$ не существует, то есть является вырожденной.

Следовательно, нет возможности воспользоваться процедурами, сформулированными в теореме Гаусса-Маркова для оценки параметров модели и их ковариационной матрицы.

Наличие линейных (иногда функциональных) связей между факторами X1,X2,....Xk, включенными во множественную эконометрическую модель называется **мультиколлинеарностью.**

Существует полная и частичная мультиколлинеарности.

# Полная мультиколлинеарность

Если, регрессоры в модели связаны строгой функциональной зависимостью, то говорят о наличии **полной (совершенной) мультиколинеарности**

Полная мультиколлинеарность не позволяет однозначно оценить параметры исходной модели и разделить вклады регрессоров в эндогенную переменную по результатам наблюдений

Рассмотрим пример

Пусть спецификация модели имеет вид:

$$Y = a_0 + a_1 x_1 + a_2 x_2 + e \qquad (3.3)$$

Предположим, что регрессоры $x_1$ и $x_2$ связаны между собой строгой линейной зависимостью:

$$\boldsymbol{x}_2 = \alpha_0 + \alpha_1 \boldsymbol{x}_1 \qquad (3.4)$$

# Полная мультиколлинеарность

Подставив (3.4) в (3.3), получим уравнение парной регрессии

$$Y = a_0 + a_1 x_1 + a_2(\alpha_0 + \alpha_1 x_1) + e$$

Раскрыв скобки и приведя преобразования, получим модель в виде:

$$Y = (a_0 + a_2\alpha_0) + (a_1 + a_2\alpha_1)\, x_1 + e \qquad (3.5)$$

Уравнение (3.5) можно записать в виде:

$$Y = b_0 + b_1 x_1 + e$$

где:

$$\begin{cases} b_0 = a_0 + a_2\alpha_0 \\ b_1 = a_1 + a_2\alpha_1 \end{cases} \qquad (3.6)$$

По оценкам параметров $b_0$ и $b_1$ невозможно однозначно оценить параметры модели (3.3), так как в системе (3.6) неизвестных больше, чем исходных данных. Такая система, в общем случае, имеет бесчисленное множество решений.

# Частичная мультиколлинеарность

Так как в реальности мы имеем дело с данными, имеющими стохастический характер, то случай полной мультиколлинеарности на практике встречается крайне редко.

На практике мы имеем дело с частичной мультиколлинеарностью.

***Частичная (несовершенная, стохастическая) мультиколлинеарность*** характерна для случаев, когда часть экзогенных факторов (X1, X2, …, Xk) находится в корреляционной связи или образовывает различные линейные комбинации вида

$$x_i = b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$

Для определения степени коррелированности строят матрицу взаимных корреляций регрессоров $R=\{r_{ij}\}$, I,j=1,2,…,k

# Частичная мультиколлинеарность

Если между регрессорами имеется корреляционная связь, соответствующий коэффициент корреляции будет близок к единице $r_{ij} \approx 1$

Матрица $(X^T X)^{-1}$ будет иметь полный ранг, но близка к вырожденной, т.е $\det(X^T X)^{-1} \approx 0$

В этом случае, формально можно получить оценки параметров модели, их точностные показатели, но все они будут **неустойчивыми.**

**Подобная ситуация возникает, если при спецификации модели в качестве факторных признаков одновременно используются такие показатели, как затраты на единицу продукции, себестоимость товара, его цена.**

# Частичная мультиколлинеарность

**Последствия частичной мультиколлинеарности:**

- Увеличение дисперсий оценок параметров. Это расширяет интервальные оценки и снижает их точность;

- Уменьшение значений t-статистик для параметров, что приводит к неправильному выводу о их статистической значимости;

-Неустойчивость оценок МНК-параметров и их дисперсий;

- Возможность получения неверного (с точки зрения теории) знака у оценки параметра.

# Частичная мультиколлинеарность

**Поясним это на примере**

Пусть спецификация модели имеет вид:

$$Y = a_1 x_1 + a_2 x_2 + e$$

Для такой модели значения дисперсий параметров и их ковариация может быть выражена через значение выборочного коэффициента корреляции следующим образом:

$$\sigma^2_{a1} = \frac{\sigma^2}{\sum_{t=1}^{t=n} x^2_{t1}\left(1 - r^2_{12}\right)}, \quad \sigma^2_{a2} = \frac{\sigma^2}{\sum_{t=1}^{t=n} x^2_{t2}\left(1 - r^2_{12}\right)}$$

$$COV(a_1, a_2) = \frac{-\sigma^2 r_{12}}{\left(1 - r^2_{12}\right)\sqrt{\sum_{t=1}^{t=n} x^2_{t1}}\sqrt{\sum_{t=1}^{t=n} x^2_{t2}}}$$

*где*

$$r_{12} = \frac{\sqrt{\sum_{t=1}^{t=n} x_{t1} x_{t2}}}{\sqrt{\sum_{t=1}^{t=n} x^2_{t1}}\sqrt{\sum_{t=1}^{t=n} x^2_{t2}}}$$

# Частичная мультиколлинеарность

Точные количественные критерии для обнаружения частичной мультиколлинеарности отсутствуют.

В качестве признаков ее наличия используют следующие:

- Модуль парного коэффициента корреляции между регрессорами $X_i$ и $X_j$ больше 0,75

- Близость к нулю определителя матрицы $(X^TX)^{-1}$

- Большое количество статистически незначимых параметров в модели

# ПРИЗНАКИ МУЛЬТИКОЛЛЛИНЕАРНОСТИ

К общим признакам наличия мультиколлинеарности в регрессионной модели следует отнести:

1. Небольшое изменение исходных данных (например, добавление новых наблюдений) приводит к существенному изменению оценок параметров модели;

2. Оценки имеют большие стандартные ошибки и малую значимость в то время как модель в целом является значимой (наблюдается высокое значение коэффициента детерминации и соответствующей F-статистики);

3. Оценки параметров имеют неоправданно большие значения или неверные знаки

# СПОСОБЫ ОПРЕДЕЛЕНИЯ НАЛИЧИЯ МУЛЬТИКОЛЛИНЕАРНОСТИ

*1. Величина определителя матрицы* $X^T * X$.. Собственные числа $\lambda_i$ находятся следующим образом:

$$\left|X^T X\right| = \sum_{i=1}^{k} \lambda_i , \qquad \begin{vmatrix} b_{11} - \lambda & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} - \lambda & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{n1} & \dots & \dots & b_{nn} - \lambda \end{vmatrix} = 0 .$$

Чем ближе определитель этой матрицы к нулю, тем большая степень мультиколлинеарности между включенными в модель факторами.

# СПОСОБЫ ОПРЕДЕЛЕНИЯ НАЛИЧИЯ МУЛЬТИКОЛЛИНЕАРНОСТИ

*2. Минимальное собственное число матрицы $X^T * X$*

$$\lambda_{\underset{i}{\min}}(X^T * X) = \min_i (\lambda_i) = \lambda_1$$

Чем меньше $\lambda_1$', тем сильнее мультиколлинеарность.

*3. Мера обусловленности матрицы $X^T * X$ по Нейману-Голдштейну*:

$$\frac{\lambda_{\max}(X^T * X)}{\lambda_{\min}(X^T * X)}; \quad \lambda_{\max} = \max_i (\lambda_i) = \lambda_k$$

Чем ближе $\frac{\lambda_k}{\lambda_1}$ к бесконечности, тем сильнее мультиколлинеарность. Существует некоторая шкала для данного отношения, позволяющая дать качественную оценку степени мультиколлинеарности по величине меры Неймана-Голдштейна: если это отношение больше 30, то мультиколлинеарность средней степени, если больше 100 – мультиколлинеарность большая.

# СПОСОБЫ ОПРЕДЕЛЕНИЯ НАЛИЧИЯ МУЛЬТИКОЛЛИНЕАРНОСТИ

4 *Максимальная сопряженность.* Для построения этой меры рассчитывается регрессия переменной i на остальные независимые переменные с номерами 1,2, ... i-1, i+1,.... k., определяется коэффициент детерминации $R_i^2$ в регрессии $X_i$ на $X_1, X_2$, ... $X_{i-1}$, $X_{i+1}$, .... $X_k$. В качестве меры сопряженности используется величину $\max\limits_i |R_i|$. Если значение $\max\limits_i |R_i|$ близкое к 1, то совокупность независимых переменных подвержена сильному влиянию эффекта мультиколлинеарности.

# СПОСОБЫ ОПРЕДЕЛЕНИЯ НАЛИЧИЯ МУЛЬТИКОЛЛИНЕАРНОСТИ

5. Метод Феррара-Глобера, который основан на применении трех видов статистических критериев:

1) всего массива независимых переменных ($\chi^2$ критерий);

2) каждой независимой переменной со всеми другими (F-критерий);

3) каждой пары независимых переменных (t-критерий).

Сравнив эти критерии с их критическими значениями, можно сделать вывод о наличии или отсутствии мультиколлинеарности между независимыми переменными.

# ЧАСТНЫЕ КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ

Коэффициент корреляции, очищенный от влияния других факторов, называется **частным коэффициентом корреляции**

Частный коэффициент корреляции определяет степень зависимости между двумя переменными без учета влияния на них других факторов

**Рассмотрим пример.** Пусть спецификация модели имеет вид:

$$Y = a_0 + a_1 x_1 + a_2 x_2 + e \qquad (3.6)$$

**Задача.** Определить корреляцию между Y и $X_1$, исключив влияние переменной $X_2$

# ЧАСТНЫЕ КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ

Алгоритм решения заключается в следующем:

1. Строится регрессия Y на $X_2$

$$\tilde{Y} = \tilde{\alpha}_0 + \tilde{\alpha}_2 x_2$$

2. Строится регрессия $X_1$ на $X_2$

$$x_1 = \tilde{\beta}_0 + \tilde{\beta}_2 x_2$$

3. Для удаления влияния $X_2$ вычисляются остатки:

$$\varepsilon_Y = Y - \tilde{Y}, \qquad \varepsilon_{x1} = x_1 - \tilde{x}_1$$

4. Значение частного коэффициента корреляции между переменными Y и $X_1$ вычисляется по формуле:

$$r(Y, x_1 | x_2) = r(\varepsilon_Y, \varepsilon_{x1})$$

# ЧАСТНЫЕ КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ

Частные коэффициенты корреляции могут быть вычислены по значениям парных коэффициентов

$$r(Y,x_1|x_2)=\frac{r(Y,x_1)-r(Y,x_2)r(x_1,x_2)}{\sqrt{1-r^2(x_1,x_2)}\sqrt{1-r^2(Y,x_2)}} \qquad (3.7)$$

В общем случае связь между частными и обычными коэффициентами корреляции осуществляется следующим образом:

$$r^*_{ij}=\frac{-c_{ij}}{\sqrt{c_{ii}}\sqrt{c_{jj}}} \qquad i=1,2,\dots,k, \qquad j=1,2,\dots,k$$

где :

$$C=R^{-1}, \qquad R=\begin{pmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & \dots & 1 \end{pmatrix}, \qquad r_{ij}=COR(x_i,x_j) \qquad (3.8)$$

# ЧАСТНЫЕ КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ

**Пример 1**. Вычислить частный коэффициент корреляции $r(Y,X_1 | X_2)$ между переменными модели (3.6)
Пусть матрица R имеет вид:

$$\boldsymbol{R} = \begin{pmatrix} 1 & \boldsymbol{r}_{01} & \boldsymbol{r}_{02} \\ \boldsymbol{r}_{10} & 1 & \boldsymbol{r}_{12} \\ \boldsymbol{r}_{20} & \boldsymbol{r}_{21} & 1 \end{pmatrix},$$

**где :**

$$\boldsymbol{r}_{01} = \boldsymbol{COR}(\boldsymbol{Y}, \boldsymbol{x}_1), \quad \boldsymbol{r}_{02} = \boldsymbol{COR}(\boldsymbol{Y}, \boldsymbol{x}_2), \quad \boldsymbol{r}_{10} = \boldsymbol{COR}(\boldsymbol{x}_1, \boldsymbol{Y}),$$

$$\boldsymbol{r}_{20} = \boldsymbol{COR}(\boldsymbol{x}_2, \boldsymbol{Y}), \quad \boldsymbol{r}_{12} = \boldsymbol{COR}(\boldsymbol{x}_1, \boldsymbol{x}_2), \quad \boldsymbol{r}_{21} = \boldsymbol{COR}(\boldsymbol{x}_2, \boldsymbol{x}_1),$$

$$\boldsymbol{r}_{00} = \boldsymbol{r}_{11} = \boldsymbol{r}_{22} = 1$$

Тогда частный коэффициент корреляции $r(Y,X_1 | X_2)$ вычисляется с помощью (3.7)

$$\boldsymbol{r}_{01|2} = \frac{\boldsymbol{r}_{10} - \boldsymbol{r}_{20}\boldsymbol{r}_{12}}{\sqrt{1 - \boldsymbol{r}_{21}\boldsymbol{r}_{12}}\sqrt{1 - \boldsymbol{r}_{02}^2}}$$

# ЧАСТНЫЕ КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ

**Пример 2**. В таблице приведены данные об объеме импорта Y (млрд. дол), ВНП $X_1$ (млрд.дол) и индексе цен $X_2$ в США за период 1964-1979 гг

Вычислить элементы матрицы взаимных корреляций модели:

$$Y = a_0 + a_1 x_1 + a_2 x_2 + e$$

| Годы | Y | $X_1$ | $X_2$ |
|---|---|---|---|
| 1964 | 28,4 | 635,7 | 92,9 |
| 1965 | 32,0 | 688,1 | 94,5 |
| 1966 | 37,7 | 753,0 | 97,2 |
| 1967 | 40,6 | 796,3 | 100,0 |
| 1968 | 47,7 | 868,5 | 104,2 |
| 1969 | 52,9 | 935,5 | 109,8 |
| 1970 | 58,5 | 982,4 | 116,3 |
| 1971 | 64,0 | 1063,4 | 121,3 |
| 1972 | 75,9 | 1171,1 | 125,4 |
| 1973 | 94,4 | 1306,6 | 133,1 |
| 1974 | 131,9 | 1412,9 | 137,7 |
| 1975 | 126,9 | 1528,8 | 161,2 |
| 1976 | 155,4 | 1702,2 | 170,5 |
| 1977 | 185,8 | 1899,5 | 181,5 |
| 1978 | 217,5 | 2127,6 | 195,4 |
| 1979 | 260,9 | 2368,5 | 217,4 |

Решение.

1. Вычисляем матрицу взаимных корреляций

| | *Y* | *X1* | *X2* |
|---|---|---|---|
| Y | 1,0000 | | |
| X1 | 0,9932 | 1,0000 | |
| X2 | 0,9885 | 0,9957 | 1,0000 |

2. Вычисляется обратная матрица

| | | |
|---|---|---|
| 73,764 | -76,936 | 3,689 |
| -76,625 | 196,433 | -119,845 |
| 3,379 | -119,537 | 116,683 |

# ЧАСТНЫЕ КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ

**Пример 2.** (Продолжение)

3. Вычисляются оценки частных коэффициентов корреляции с помощью (3.8)

Обратная матрица $R^{-1}$

| | | |
|---|---|---|
| 73,764 | -76,936 | 3,689 |
| -76,625 | 196,433 | -119,845 |
| 3,379 | -119,537 | 116,683 |

Выражение (3.8)

$$r^*_{ij} = \frac{-c_{ij}}{\sqrt{c_{ii}}\sqrt{c_{jj}}} \qquad i = 1,2,\dots,k, \quad j = 1,2,\dots,k$$

Тогда:

$$r(Y, x_1 | x_2) = \frac{76.936}{\sqrt{73.764}\sqrt{196.433}} = 0{,}639$$

$$r(Y, x_2 | x_1) = \frac{-3.689}{\sqrt{73.764}\sqrt{116.683}} = -0{,}0398$$

$$r(x_1, x_2 | Y) = \frac{119.845}{\sqrt{196.433}\sqrt{116.683}} = 0{,}792$$

Проверка гипотезы $H_0$: $r(x_1,x_2 | Y)=0$

$$t = \frac{\left|r(x_1, x_2 | Y)\right|\sqrt{n-k-1}}{\sqrt{1-r(x_1,x_2|Y)^2}} = \frac{0.792\sqrt{16-2-1}}{\sqrt{1-0.792^2}} = 7.661 > t_{кр}$$

# МЕТОДЫ УСТРАНЕНИЯ МУЛЬТИКОЛЛИНЕАРНОСТИ

**Существуют следующие группы методов устранения мультиколлинеарности в уравнениях регрессии:**

- Методы исключения переменных модели;
- Методы, которые используют внешнюю информацию;
- Методы, предполагающие переход к смещенным оценкам параметров модели;
- Методы преобразования данных;
- Метод главных компонент

# МЕТОДЫ УСТРАНЕНИЯ МУЛЬТИКОЛЛИНЕАРНОСТИ

Основным методом устранения мультиколлинеарности заключается в исключении переменных

Существует несколько способов решения этой задачи

1. **Метод дополнительных регрессий**

Алгоритм метода заключается в следующем:

1. Строятся уравнения регрессии, которые связывают каждый из регрессоров со всеми оставшимися
2. Вычисляются коэффициенты детерминации $R^2$ для каждого уравнения регрессии
3. Проверяется статистическая гипотеза $H_0$: $R^2=0$ с помощью F теста

**Вывод**: если гипотеза $H_0$: $R^2=0$ не отклоняется, значит данный регрессор не приводит к мультиколлинеарности

# Методы устранения мультиколлинеарности

**Пример.** Рассмотрим предыдущую задачу и определим, приводит ли регрессор $X_1$ к мультиколлинеарности

Исходные данные

| Годы | Y | $X_1$ | $X_2$ |
|---|---|---|---|
| 1964 | 28,4 | 635,7 | 92,9 |
| 1965 | 32,0 | 688,1 | 94,5 |
| 1966 | 37,7 | 753,0 | 97,2 |
| 1967 | 40,6 | 796,3 | 100,0 |
| 1968 | 47,7 | 868,5 | 104,2 |
| 1969 | 52,9 | 935,5 | 109,8 |
| 1970 | 58,5 | 982,4 | 116,3 |
| 1971 | 64,0 | 1063,4 | 121,3 |
| 1972 | 75,9 | 1171,1 | 125,4 |
| 1973 | 94,4 | 1306,6 | 133,1 |
| 1974 | 131,9 | 1412,9 | 137,7 |
| 1975 | 126,9 | 1528,8 | 161,2 |
| 1976 | 155,4 | 1702,2 | 170,5 |
| 1977 | 185,8 | 1899,5 | 181,5 |
| 1978 | 217,5 | 2127,6 | 195,4 |
| 1979 | 260,9 | 2368,5 | 217,4 |

Результаты расчета

| | | |
|---|---|---|
| $a_i$ | 13,59 | -568,32 |
| $s_i$ | 0,34 | 47,35 |
| $R^2$ | 0,99 | 51,07 |
| Fтест | 1616,97 | 14,00 |
| | 4217961 | 36519,9 |

Значение Fтест =1616.97 > Fкрит
Следовательно, гипотеза о равенстве нулю коэффициента детерминации отклоняется
**Вывод:** регрессор $X_1$ вызовет в модели мультиколлинеарность

# Методы устранения мультиколлинеарности

**2. Метод последовательного присоединения (пошаговая регрессия)**

В отличие от рассмотренного, метод последовательного присоединения регрессоров позволяет выявить набор регрессоров, который ни только не приводит к мультиколлинеарности, но и обеспечивает наилучшее качество спецификации модели

Алгоритм метода следующий:

1. Строится регрессионная модель с учетом всех предполагаемых регрессоров. По признакам делается вывод о возможном присутствии мультиколлинеарности
2. Рассчитывается матрица корреляций  и выбирается регрессор, имеющий наибольшую корреляцию с эндогенной переменной
3. К выбранному регрессору последовательно в модель добавляется каждый из оставшихся регрессоров и вычисляются скорректированные коэффициенты детерминации для каждой из моделей.

   К модели присоединяется тот регрессор, который обеспечивает наибольшее значение скорректированного $R^2$

# Методы устранения мультиколлинеарности

4. К паре выбранных регрессоров последовательно присоединяется третий из числа оставшихся  Строятся модели, вычисляется скорректированный $R^2$, добавляется тот регрессор, который обеспечивает наибольшее значение скорректированного $R^2$

Процесс присоединения регрессоров прекращается, когда значение скорректированного $R^2$ становится меньше достигнутого на предыдущем шаге

**Замечание.** Каким бы образом не осуществлялся отбор факторов, уменьшение их числа приводит к улучшению обусловленности матрицы $(X^TX)^{-1}$, а, следовательно, к повышению качества оценок параметров модели

# Методы устранения мультиколлинеарности

**Пример 2.**

Исследуется зависимость урожайности зерновых культур Y от следующих факторов производства:

$X_1$ – число тракторов на 100га

$X_2$ – число зерноуборочных комбайнов на 100га

$X_3$ – Число орудий поверхностной обработки почвы на 100 га

$X_4$ - количество удобрений, расходуемых на гектар (т/га)

$X_5$ – количество химических средств защиты растений (т/га)

# Методы устранения мультиколлинеарности

## Исходные данные

| Номер района | Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|---|
| 1 | 9,70 | 1,59 | 0,26 | 2,05 | 0,32 | 0,14 |
| 2 | 8,40 | 0,34 | 0,28 | 0,46 | 0,59 | 0,66 |
| 3 | 9,00 | 2,53 | 0,31 | 2,46 | 0,30 | 0,31 |
| 4 | 9,90 | 4,63 | 0,40 | 6,44 | 0,43 | 0,59 |
| 5 | 9,60 | 2,16 | 0,26 | 2,16 | 0,39 | 0,16 |
| 6 | 8,60 | 2,16 | 0,30 | 2,69 | 0,32 | 0,17 |
| 7 | 12,50 | 0,68 | 0,29 | 0,73 | 0,42 | 0,23 |
| 8 | 7,60 | 0,35 | 0,26 | 0,42 | 0,21 | 0,08 |
| 9 | 8,90 | 0,52 | 0,24 | 0,49 | 0,20 | 0,08 |
| 10 | 13,50 | 3,42 | 0,31 | 3,02 | 1,37 | 0,73 |
| 11 | 9,70 | 1,78 | 0,30 | 3,19 | 0,73 | 0,17 |
| 12 | 10,70 | 2,40 | 0,32 | 3,30 | 0,25 | 0,14 |
| 13 | 12,20 | 9,36 | 0,40 | 11,51 | 0,39 | 0,38 |
| 14 | 9,70 | 1,72 | 0,28 | 2,26 | 0,82 | 0,17 |
| 15 | 7,00 | 0,59 | 0,29 | 0,60 | 0,13 | 0,35 |
| 16 | 7,20 | 0,28 | 0,26 | 0,30 | 0,09 | 0,15 |
| 17 | 8,20 | 1,64 | 0,29 | 1,44 | 0,20 | 0,08 |
| 18 | 8,40 | 0,09 | 0,22 | 0,05 | 0,43 | 0,2 |
| 19 | 13,10 | 0,08 | 0,25 | 0,03 | 0,73 | 0,2 |
| 20 | 8,70 | 1,36 | 0,26 | 0,17 | 0,99 | 0,42 |

# Методы устранения мультиколлинеарности

## Шаг 2. Построение матрицы корреляций

| | Y | X1 | X2 | X3 | X4 | X5 |
|---|---|---|---|---|---|---|
| Y | 1 | | | | | |
| X1 | 0,42 | 1 | | | | |
| X2 | 0,34 | 0,85 | 1 | | | |
| X3 | 0,4 | 0,98 | 0,88 | 1 | | |
| X4 | 0,56 | 0,11 | 0,03 | 0,03 | 1 | |
| X5 | 0,29 | 0,34 | 0,46 | 0,28 | 0,57 | 1 |

Наибольшую корреляцию эндогенная переменна Y имеет с $X_4$

**Вывод:** в модель необходимо включить регрессор $X_4$ и к нему присоединять остальные

Шаг 3. Рассматриваем следующие спецификации моделей:

1. $Y = a_0 + a_4 x_4 + a_1 x_1 + e_1$
2. $Y = a_0 + a_4 x_4 + a_2 x_2 + e_2$
3. $Y = a_0 + a_4 x_4 + a_3 x_3 + e_3$
4. $Y = a_0 + a_4 x_4 + a_5 x_5 + e_4$

| | $X_4, X_1$ | $X_4, X_2$ | $X_4, X_3$ | $X_4, X_5$ |
|---|---|---|---|---|
| $R^2$ | 0,4113 | 0,3814 | 0,4232 | 0,272 |

Наибольший R2 в модели 3

**Вывод:** Продолжаем присоединение к модели 3

# Методы устранения мультиколлинеарности

Шаг 4. Рассматриваем следующие спецификации моделей:

1. $Y = a_0 + a_4 x_4 + a_3 x_3 + a_1 x_1 + e_1$

2. $Y = a_0 + a_4 x_4 + a_3 x_3 + a_2 x_2 + e_2$

3. $Y = a_0 + a_4 x_4 + a_3 x_3 + a_5 x_5 + e_3$

| | $X_4, X_1, X_3$ | $X_4, X_3, X_2$ | $X_4, X_3$, X5 |
|---|---|---|---|
| $R^2$ | 0,3911 | 0,392 | 0,4169 |

Наибольший коэффициент детерминации соответствует модели 3.

Однако его значение меньше, чем было достигнуто ранее: $R^2$=0,4232

**Выводы:**

1. Не имеет смысл рассматривать спецификацию 3.
2. Для построения следует принять спецификацию модели в виде:

$$Y = a_0 + a_4 x_4 + a_3 x_3 + e$$

# ПРОБЛЕМА МУЛЬТИКОЛЛИНЕАРНОСТИ

Выводы:

1. Последствием мультиколлинеарности является потеря устойчивости вычисления оценок параметров модели
2. Наличие мультиколлинеарности приводит к завышенным значениям СКО оценок
3. Отсутствуют строгие критерии тестирования наличия мультиколлинеарности
4. Подозрением наличия мультиколлинеарности служит большое количество незначимых факторов в модели
5. Для устранения мультиколлинеарности необходимо удалить из спецификации модели факторы, ее вызывающие
6. Для получения спецификации модели, не имеющей мультиколлинеарности можно воспользоваться методом присоединения регрессоров или методом исключения регрессоров