МУЛЬТКОЛЛИНЕАРНОСТЬ

Лекция 12

Цели лекции

1. Объяснить сущность проблемы мультиколлинеарности
2. Изучить последствия мультиколлинеарности
3. Указать средства обнаружения мультиколлинеарности
4. Обсудить проблему выбора средств борьбы с мультиколлинеарностью

Коллинеарность и мультиколлинеарность

Термин «**коллинеарность**» характеризует линейную связь между двумя объясняющими переменными.

«**Мультиколлинеарность**» означает линейную связь между более чем двумя объясняющими переменными.

На практике всегда используется один термин – мультиколлинеарность.

Термин «мультиколлинеарность» введен Рагнаром Фришем

Виды мультиколлинеарности. Строгая и нестрогая мультиколлинеарность

1. Строгая мультиколлинеарность – наличие линейной функциональной связи между объясняющими переменными (иногда также и зависимой).
2. Нестрогая мультиколлинеарность – наличие сильной линейной корреляционной связи между объясняющими переменными (иногда также и зависимой).

Суть строгой мультиколлинеарности

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_m X_m + \varepsilon$$

$$X_1 = \alpha_0 + \alpha_1 X_2 \quad \Leftrightarrow \quad X_2 = -\frac{\alpha_0}{\alpha_1} + \frac{1}{\alpha_1} X_1$$

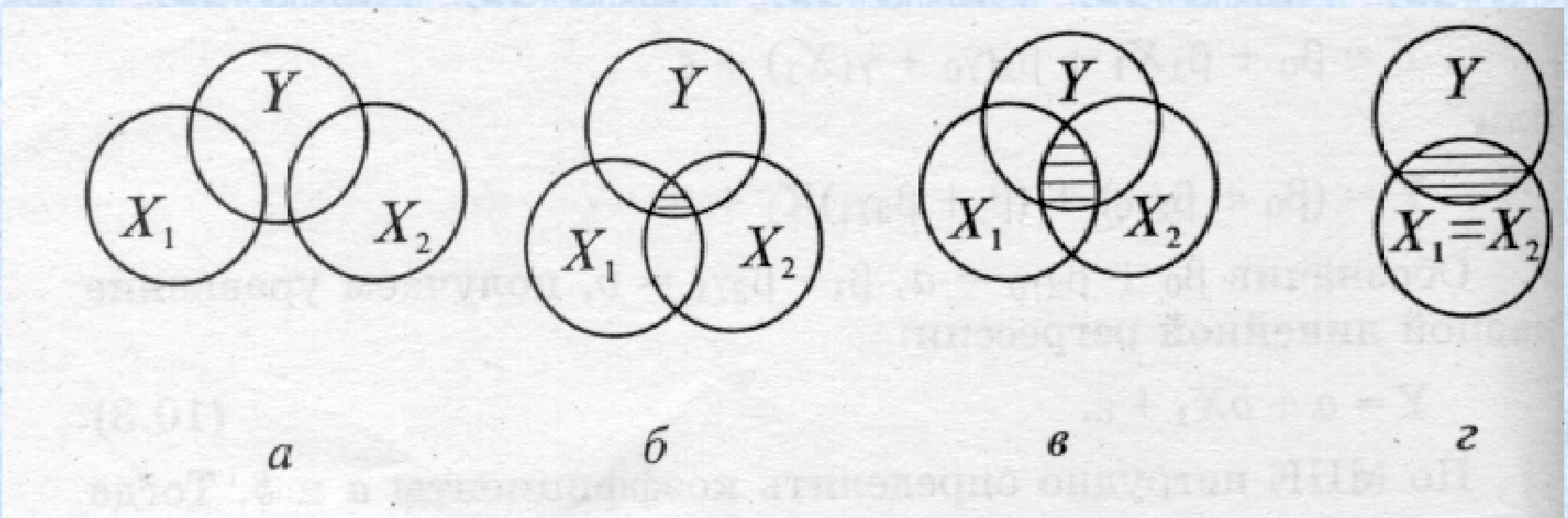
$$Y = (\beta_0 + \beta_1 \alpha_0) + (\beta_2 + \beta_1 \alpha_2) X_2 + \beta_3 X_3 + \ldots + \beta_m X_m + \varepsilon$$

$$Y = \left(\beta_0 - \beta_2 \frac{\alpha_0}{\alpha_1}\right) + \left(\beta_1 + \beta_2 \frac{1}{\alpha_1}\right) X_1 + \beta_3 X_3 + \ldots + \beta_m X_m + \varepsilon$$

Связь между объясняющими переменными – <u>функциональная</u>

Суть строгой мультиколлинеарности. Выводы

Строгая мультиколлинеарность не позволяет однозначно определить коэффициенты регрессии и разделить вклады объясняющих переменных X_1 и X_2 в их влиянии на зависимую переменную Y.

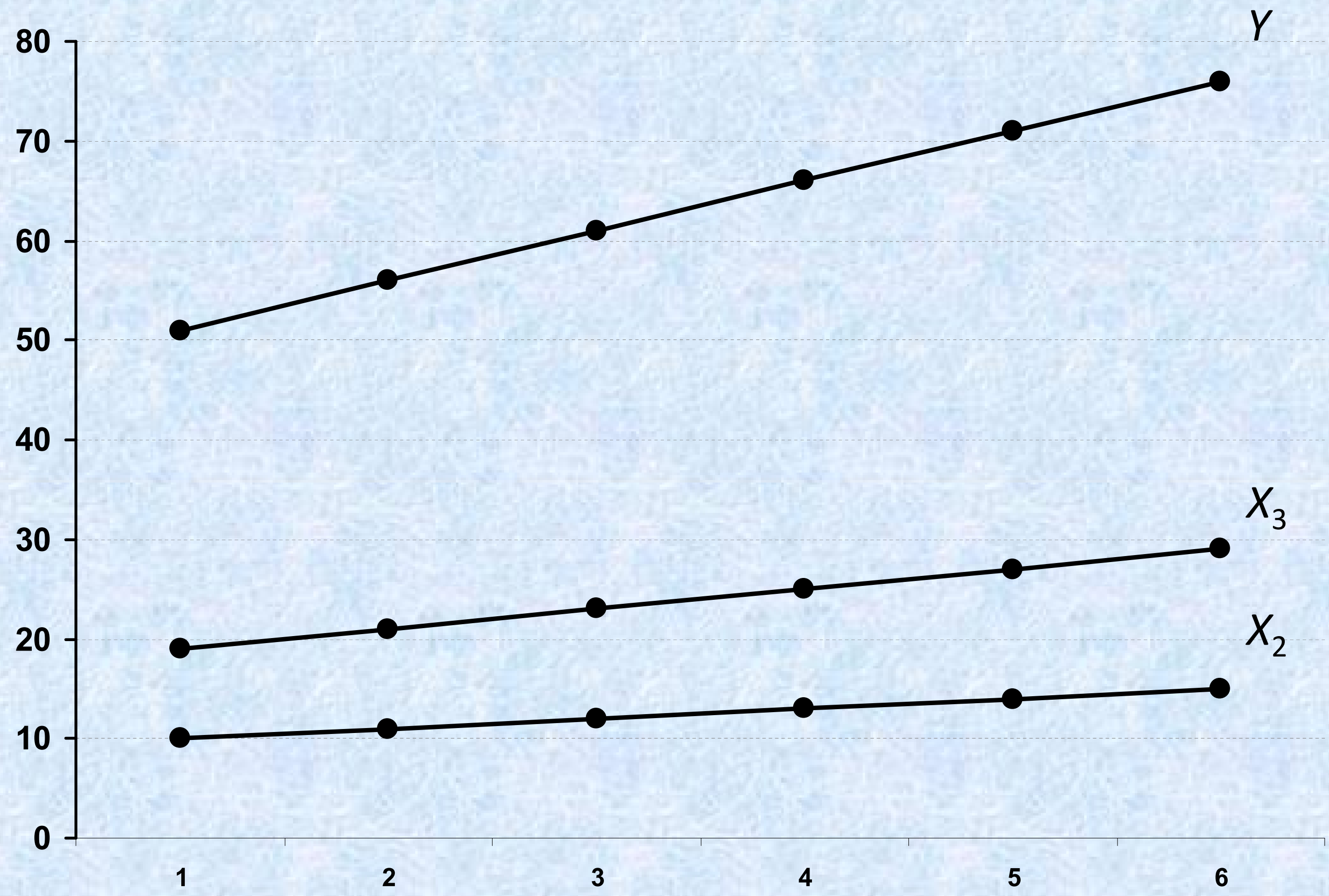


Пример строгой мультиколлинеарности.

$$Y = 2 + 3X_2 + X_3$$
$$X_3 = 2X_2 - 1$$

X_2	X_3	Y
10	19	51
11	21	56
12	23	61
13	25	66
14	27	71
15	29	76

Пример строгой мультиколлинеарности.



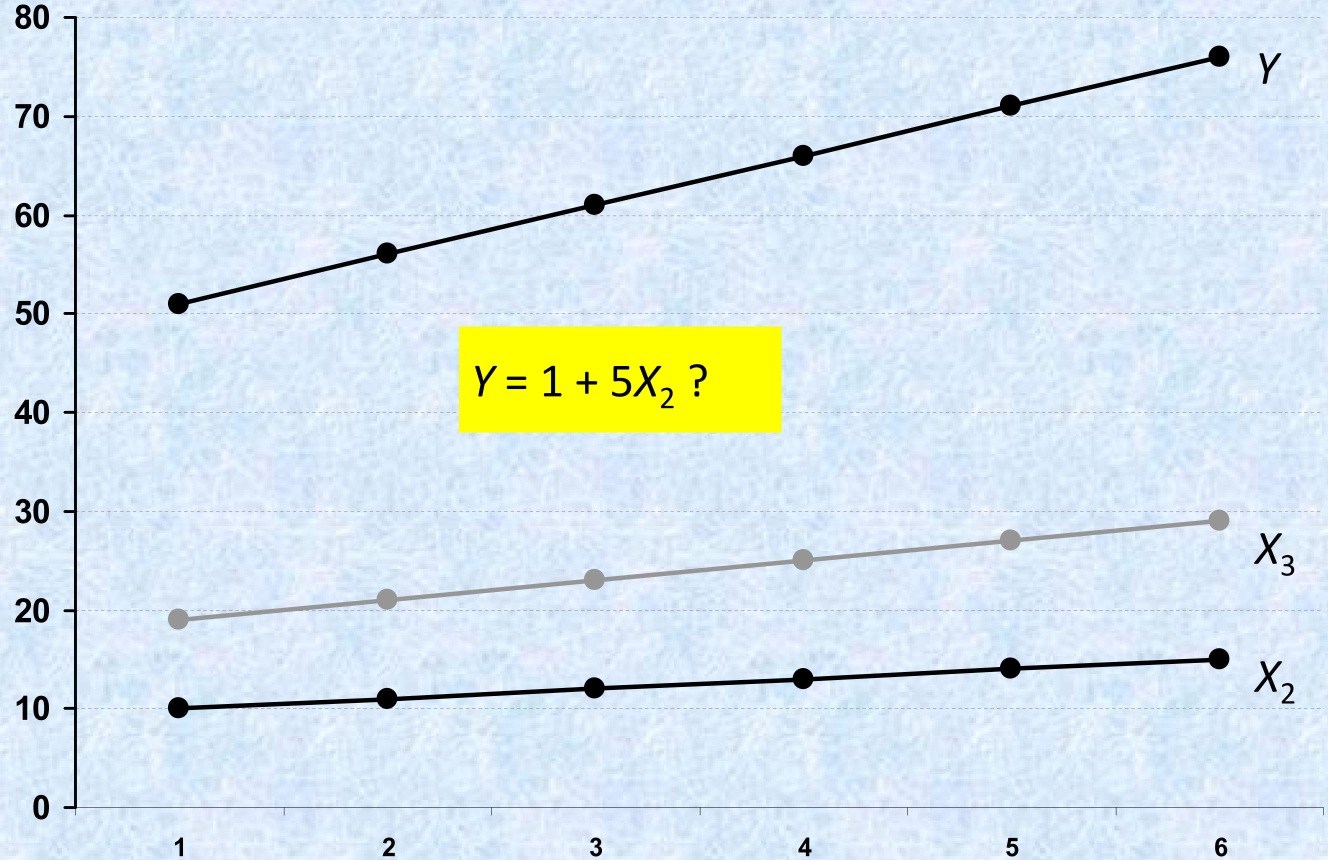
Пример строгой мультиколлинеарности.

$$Y = 2 + 3X_2 + X_3$$

$$X_3 = 2X_2 - 1$$

X_2	X_3	Y	Change in X_2	Change in X_3	Change in Y
10	19	51	1	2	5
11	21	56	1	2	5
12	23	61	1	2	5
13	25	66	1	2	5
14	27	71	1	2	5
15	29	76	1	2	5

Пример строгой мультиколлинеарности.

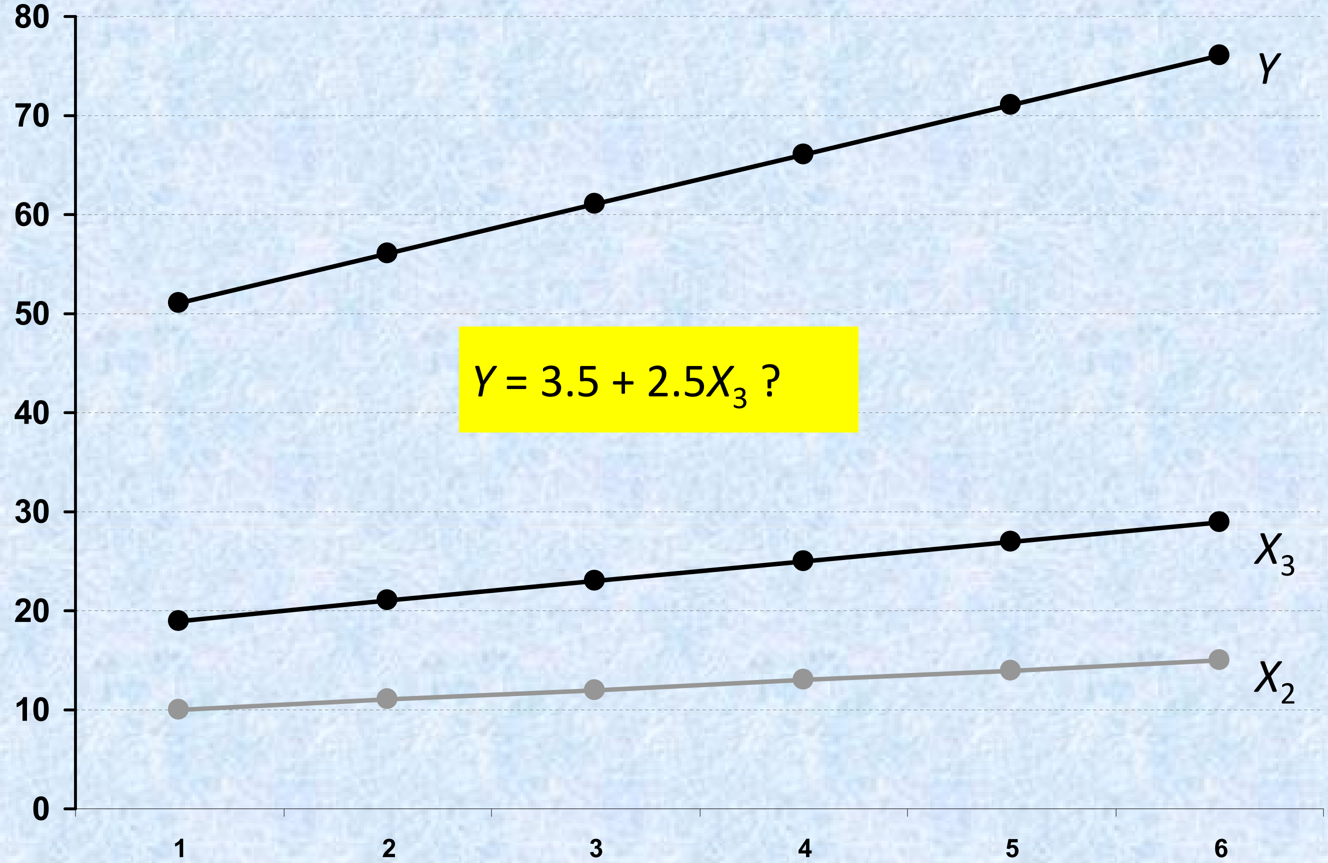


Пример строгой мультиколлинеарности.

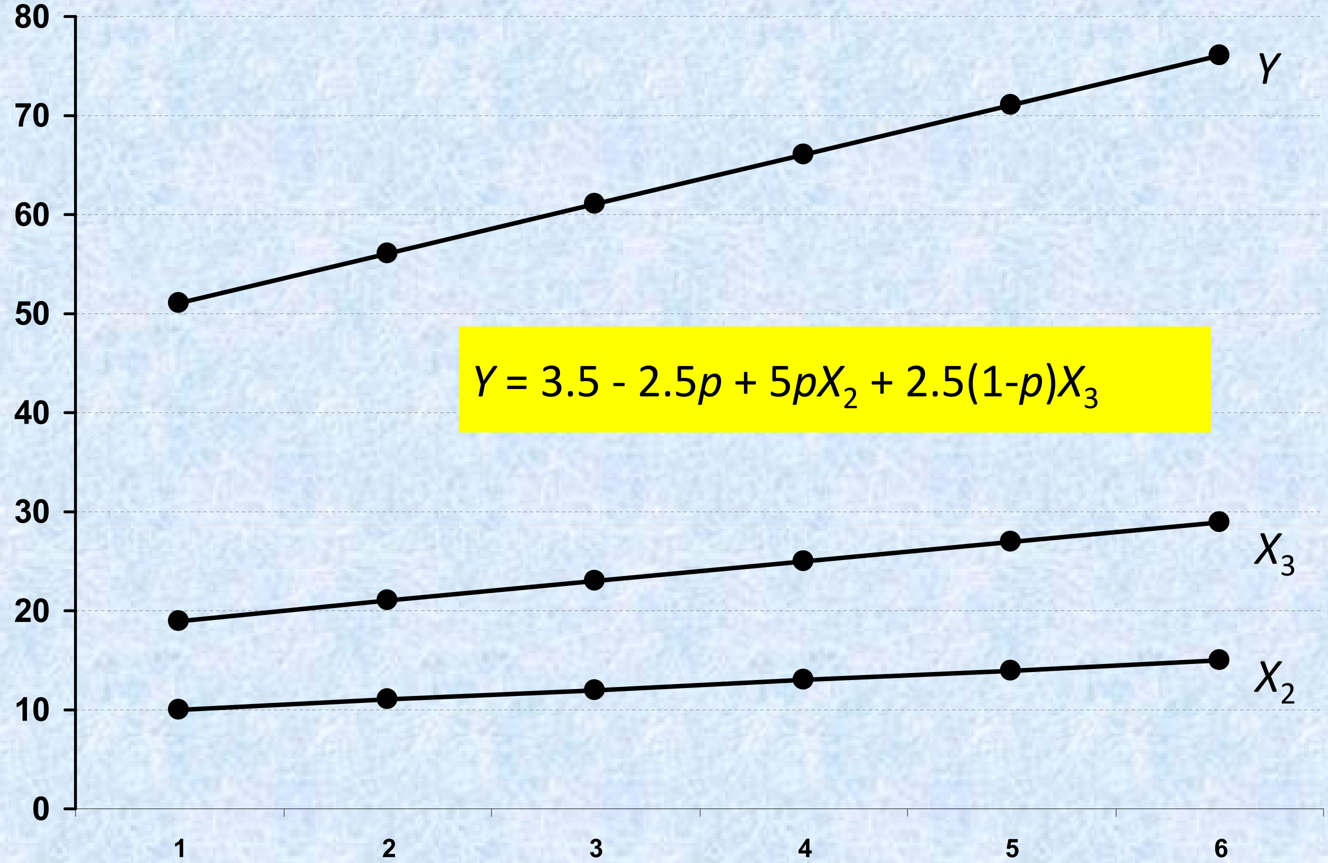
$$Y = 2 + 3X_2 + X_3$$
$$X_3 = 2X_2 - 1$$

X_2	X_3	Y	Change in X_2	Change in X_3	Change in Y
10	19	51	1	2	5
11	21	56	1	2	5
12	23	61	1	2	5
13	25	66	1	2	5
14	27	71	1	2	5
15	29	76	1	2	5

Пример строгой мультиколлинеарности.



Пример строгой мультиколлинеарности.



$Y = 3.5 - 2.5p + 5pX_2 + 2.5(1-p)X_3$

Неопределенность коэффициентов при строгой мультиколлинеарности.

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \qquad X_3 = \lambda + \mu X_2$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, Y)\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - \left[\mathrm{Cov}(X_2, X_3)\right]^2}$$

$$= \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(\lambda + \mu X_2) - \mathrm{Cov}([\lambda + \mu X_2], Y)\mathrm{Cov}(X_2, [\lambda + \mu X_2])}{\mathrm{Var}(X_2)\mathrm{Var}(\lambda + \mu X_2) - \left[\mathrm{Cov}(X_2, [\lambda + \mu X_2])\right]^2}$$

$$= \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(\mu X_2) - \mathrm{Cov}(\mu X_2, Y)\mathrm{Cov}(X_2, \mu X_2)}{\mathrm{Var}(X_2)\mathrm{Var}(\mu X_2) - \left[\mathrm{Cov}(X_2, \mu X_2)\right]^2}$$

Неопределенность коэффициентов при строгой мультиколлинеарности.

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \qquad X_3 = \lambda + \mu X_2$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, Y)\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - \left[\mathrm{Cov}(X_2, X_3)\right]^2}$$

$$= \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(\lambda + \mu X_2) - \mathrm{Cov}([\lambda + \mu X_2], Y)\mathrm{Cov}(X_2, [\lambda + \mu X_2])}{\mathrm{Var}(X_2)\mathrm{Var}(\lambda + \mu X_2) - \left[\mathrm{Cov}(X_2, [\lambda + \mu X_2])\right]^2}$$

$$= \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(\mu X_2) - \mathrm{Cov}(\mu X_2, Y)\mathrm{Cov}(X_2, \mu X_2)}{\mathrm{Var}(X_2)\mathrm{Var}(\mu X_2) - \left[\mathrm{Cov}(X_2, \mu X_2)\right]^2}$$

Неопределенность коэффициентов при строгой мультиколлинеарности.

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u \qquad X_3 = \lambda + \mu X_2$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(\mu X_2) - \mathrm{Cov}(\mu X_2, Y)\mathrm{Cov}(X_2, \mu X_2)}{\mathrm{Var}(X_2)\mathrm{Var}(\mu X_2) - \left[\mathrm{Cov}(X_2, \mu X_2)\right]^2}$$

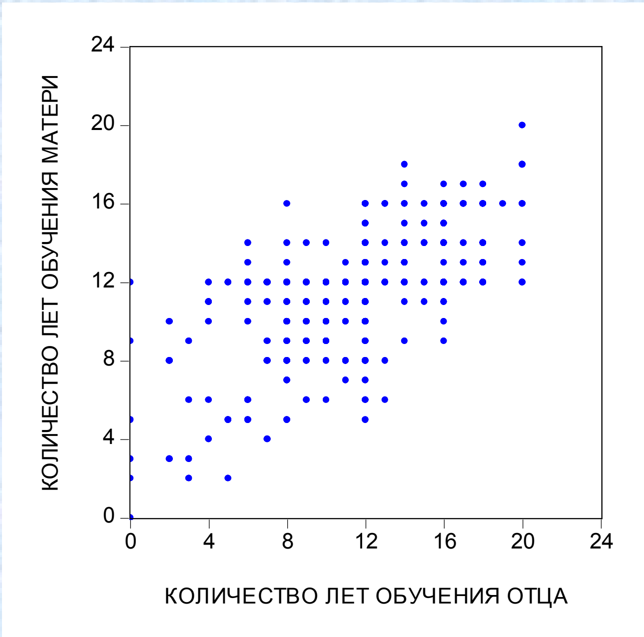
$$= \frac{\mathrm{Cov}(X_2, Y)\mu^2\mathrm{Var}(X_2) - \mu\mathrm{Cov}(X_2, Y)\mu\mathrm{Cov}(X_2, X_2)}{\mathrm{Var}(X_2)\mu^2\mathrm{Var}(X_2) - \left[\mu\mathrm{Cov}(X_2, X_2)\right]^2}$$

$$= \frac{\mu^2\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_2) - \mu^2\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_2)}{\mu^2\mathrm{Var}(X_2)\mathrm{Var}(X_2) - \left[\mu\mathrm{Var}(X_2)\right]^2} = \frac{0}{0}$$

Нестрогая мультиколлинеарность

$$X_{i1} = \alpha_0 + \alpha_1 X_{i2} + \eta_i$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{im} + \varepsilon_i$$



Связь между объясняющими переменными – <u>корреляционная</u>

Пример нестрогой мультиколлинеарности

```
. reg EARNINGS S ASVABC ASVAB5

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =   27.66
   Model |  4909.11468     3  1636.37156           Prob > F      =  0.0000
Residual |  33487.9224   566  59.1659406           R-squared     =  0.1279
---------+------------------------------           Adj R-squared =  0.1232
   Total |  38397.0371   569  67.4816117           Root MSE      =  7.6919

------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .7115506   .1612235      4.413   0.000       .3948811     1.02822
  ASVABC |   .1104595   .0504223      2.191   0.029       .0114219    .2094972
  ASVAB5 |   .0770794   .0463868      1.662   0.097      -.0140319    .1681908
   _cons |  -5.944977   2.161409     -2.751   0.006      -10.19034   -1.699616
------------------------------------------------------------------------------
```

Пример нестрогой мультиколлинеарности

```
. reg EARNINGS S ASVABC

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  2,   567) =   39.98
   Model |  4745.74965     2  2372.87483           Prob > F      =  0.0000
Residual |  33651.2874   567  59.3497133           R-squared     =  0.1236
---------+------------------------------           Adj R-squared =  0.1205
   Total |  38397.0371   569  67.4816117           Root MSE      =  7.7039

------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .7390366   .1606216      4.601   0.000       .4235506    1.054523
  ASVABC |   .1545341   .0429486      3.598   0.000       .0701764    .2388918
   _cons |  -4.624749     2.0132     -2.297   0.022      -8.578989   -.6705095
------------------------------------------------------------------------------
```

Пример нестрогой мультиколлинеарности

```
. reg EARNINGS S ASVABC ASVAB5

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =   27.66
   Model |  4909.11468     3  1636.37156           Prob > F      =  0.0000
Residual |  33487.9224   566  59.1659406           R-squared     =  0.1279
---------+------------------------------           Adj R-squared =  0.1232
   Total |  38397.0371   569  67.4816117           Root MSE      =  7.6919

---------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t
---------+-----------------------------------------
       S |   .7115506   .1612235      4.413
  ASVABC |   .1104595   .0504223      2.191
  ASVAB5 |   .0770794   .0463868      1.662
   _cons |  -5.944977   2.161409     -2.751
---------------------------------------------------
```

```
. cor ASVABC ASVAB5
(obs=570)

        |   ASVABC   ASVAB5
--------+------------------
  ASVABC|   1.0000
  ASVAB5|   0.6371   1.0000
```

Пример нестрогой мультиколлинеарности

```
. reg EARNINGS S ASVABC ASVAB5
```

EARNINGS	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.7115506	.1612235	4.413	0.000	.3948811	1.02822
ASVABC	.1104595	.0504223	2.191	0.029	.0114219	.2094972
ASVAB5	.0770794	.0463868	1.662	0.097	-.0140319	.1681908
_cons	-5.944977	2.161409	-2.751	0.006	-10.19034	-1.699616

```
. reg EARNINGS S ASVABC
```

EARNINGS	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.7390366	.1606216	4.601	0.000	.4235506	1.054523
ASVABC	.1545341	.0429486	3.598	0.000	.0701764	.2388918
_cons	-4.624749	2.0132	-2.297	0.022	-8.578989	-.6705095

Причины возникновения мультиколлинеарности

1. Ошибочное включение в уравнение регрессии двух или более линейно зависимых переменных.

2. Две (или более) объясняющих переменных, в нормальной ситуации слабо коррелированные, становятся в конкретной выборке сильно коррелированными.

3. В модель включается объясняющая переменная, сильно коррелирующая с зависимой переменной (такая переменная называется доминантной).

Мультиколлинеарность как результат логической ошибки

Ошибочное признание независимыми содержательно зависимых переменных:

$$Y = P \cdot S \quad \Leftrightarrow \quad \ln Y = \ln P + \ln S$$

Y – *сбор урожая*, P – *урожайность*,
S – *площадь*

$$\ln Y_i = \beta_0 + \beta_1 \ln P_i + \beta_2 \ln S_i + \varepsilon_i$$

Оценка коэффициентов уравнения невозможна!

Мультиколлинеарность из-за доминантной переменной

Доминантная переменная «забивает» влияние остальных:

$$Y_i = \alpha_0 + \alpha_1 X_{i1}$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{im} + \varepsilon_i$$

Влияние остальных переменных становится незначимым

Неизбежность мультиколлинеарности

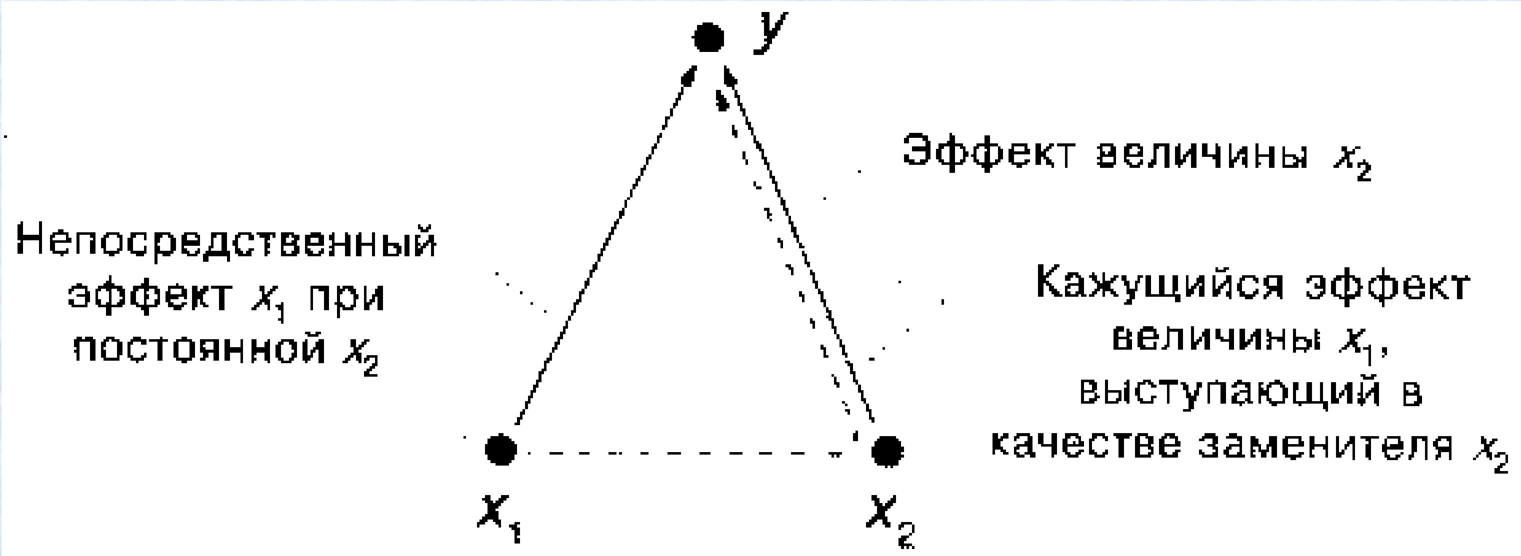
1. Мультиколлинеарность – нормальное явление.
2. Практически любая модель содержит мультиколлинеарность.
3. Мы не обращаем внимания на мультиколлинеарность до появления явных симптомов.
4. Только чрезмерно сильные связи становятся помехой.

Механизм действия мультиколлинеарности

Мультиколлинеарность проявляется в совместном действии факторов:

1. Построить модель – значит определить вклад каждого фактора.
2. Если два или более фактора изменяются только совместно, их вклад по отдельности становится невозможно различить.
3. Чем более сильно коррелированы переменные, тем труднее различить их вклад.

Механизм действия мультиколлинеарности



Из-за наличия вторичных связей качество оценок страдает - оценки оказываются менее эффективными.

В случае исключения значимой переменной X2 часть изменений Y за счет X2 будет приписана X1 , если переменная X1 может замещать X2. В результате оценка значения β1 будет смещена.

Зависимость мультиколлинеарности от выборки

Мультиколлинеарность – явление, проявляющееся на уровне выборки:

1. В одной выборке мультиколлинеарность может быть сильной, в другой слабой.
2. Выборочные данные следует предварительно всесторонне исследовать.
3. Полезен расчет выборочных коэффициентов корреляции, ковариационной матрицы и ее определителя, собственных чисел

Зависимость мультиколлинеарности от смысла задачи

Мультиколлинеарность может быть выявлена при содержательном анализе задачи и данных

Пример. Номинальная (*in*) и реальная величина (*ir*) процента (*inf* – темп инфляции)

$$in_t = ir_t + \inf_t$$

$$Y_t = \beta_0 + \beta_1 in_t + \beta_2 ir_t + \ldots + \beta_m X_{tm} + \varepsilon_i$$

При строгом контроле за ценами в некоторый период возникает строгая мультиколлинеарность

$$in_t = ir_t + const$$

Истинная мультиколлинеарность при отсутствии парных зависимостей

Пример. Рассмотрим три ряда данных:

$$X_1 = (1, 1, 1, 1, 1, 0, 0, 0, 0, 0)$$

$$X_2 = (0, 0, 0, 0, 0, 1, 1, 1, 1, 1)$$

$$X_3 = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$$

В данной выборке существует строгая мультиколлинеарность

$$X_3 = X_1 + X_2$$

но

$$r_{12} = -\frac{1}{3}; \quad r_{13} = r_{23} = \frac{1}{\sqrt{3}} \approx 0{,}58$$

Последствия мультиколлинеарности: диагностика и прогноз

1. Оценки коэффициентов остаются несмещенными.
2. Стандартные ошибки коэффициентов увеличиваются.
3. Вычисленные t-статистики занижены.
4. Неустойчивость оценок. Добавление или исключение малого количества информации (например, только одного наблюдения) может привести к очень сильному изменению оценок коэффициентов. При этом резко уменьшается и точность предсказания по модели.

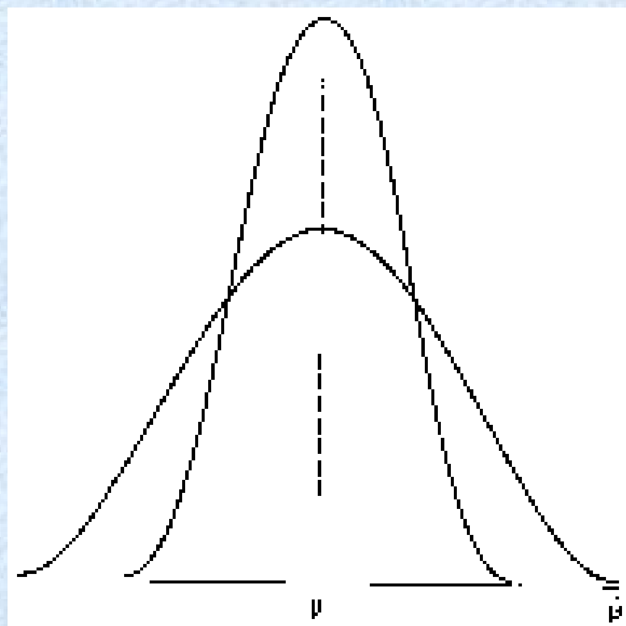
Последствия мультиколлинеарности: диагноз и прогноз

5. Численная неустойчивость процедуры оценивания, вызванная ошибками машинного округления и накоплением этих ошибок.
6. Высокая коррелированность коэффициентов лишает смысла их интерпретацию.
7. Общее качество уравнения, а также оценки переменных, не связанных мультиколлинеарностью, Остаются незатронутыми.
8. Чем ближе мультиколлинеарность к строгой (совершенной), тем серьезнее ее последствия.

Последствия мультиколлинеарности: увеличение стандартных ошибок коэффициентов

Для уравнения с объясняющими переменными X_1 и X_2

$$S_{b_1} = \sqrt{\frac{\sum e_i^2 /(n-3)}{\sum (x_{1i} - \overline{x}_1)^2 (1 - r_{12}^2)}}$$



Последствия мультиколлинеарности: увеличение стандартных ошибок коэффициентов

Практически отсюда следует возможность получить незначимый коэффициент или «неправильный» знак

Типичная ситуация.

1. Оба коэффициента в теоретической модели положительны.

2. Оба парных коэффициента корреляции между объясняющими и зависимой переменной положительны.

3. Парный коэффициент корреляции между объясняющими переменными положителен, причем корреляция между ними сильнее, чем у каждой с зависимой переменной.

В этой ситуации у одного из коэффициентов практически всегда «неправильный» знак

Обнаружение мультиколлинеарности. Основной признак

Внешним признаком наличия мультиколлинеарности служат слишком большие значения элементов матрицы $(\mathbf{X}^T\mathbf{X})^{-1}$

Основной признак мультиколлинеарности: определитель корреляционной матрицы **R** близок к нулю:

$$|\mathbf{R}| = \begin{vmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{vmatrix} \to 0$$

Если все объясняющие переменные не коррелированы между собой,

то $|\mathbf{R}| = 1$. В противном случае $0 \le |\mathbf{R}| < 1$

Обнаружение мультиколлинеарности. Дополнительные признаки

1. Высокие R^2 и F-статистика, но некоторые (или даже все) коэффициенты незначимы, т.е. имеют низкие t-статистики.

2. Высокие парные коэффициенты корреляции.

3. Высокие частные коэффициенты корреляции.

4. Высокие значения коэффициента *VIF* («фактор инфляции вариации»).

5. Знаки коэффициентов регрессии противоположны ожидаемым.

6. Добавление или удаление наблюдений из выборки сильно изменяют значения оценок.

Фактор инфляции вариации как оценка эффекта мультиколлинеарности

$$Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ij} + \varepsilon_i \quad \Rightarrow \quad X_{i1} = \alpha_0 + \sum_{j=2}^{m} \alpha_j X_{ij} + \eta_i \;\rightarrow\; R_1^2$$

$$VIF(X_1) = \frac{1}{1 - \sqrt{R_1^2}}$$

Если *VIF* > 10, то объясняющие переменные $X_i, i = \overline{1,m}$ считаются мультиколлинеарными.

Методы устранения мультиколлинеарности

1. Включить более подходящие регрессоры.
2. Изменить или увеличить выборку
3. Увеличить вариацию регрессоров
4. Преобразовать мультиколлинеарные переменные:
- использовать агрегаты (линейные комбинации переменных);
- очищать переменные от влияния друг на друга
- использовать нелинейные формы;
- использовать первые разности вместо самих переменных.
5. Исключить из модели одну из переменных.
6. Использовать при оценке коэффициентов метод главных компонент или другие специальные процедуры расчета коэффициентов при плохой обусловленности $\mathbf{X}^T\mathbf{X}$
7. Ничего не делать!

Самое главное – выбрать правильное средство

Методы устранения мультиколлинеарности

Чтобы уменьшить стандартные ошибки коэффициентов следует уменьшить вариацию случайного фактора

$$\operatorname{Var} b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n\operatorname{Var}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(1) Уменьшение $\boldsymbol{\sigma_u^2}$ за счет включения в модель более подходящих регрессоров

Методы устранения мультиколлинеарности

```
. reg EARNINGS S ASVABC ASVAB5

  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  3,   566) =   27.66
   Model |  4909.11468     3  1636.37156               Prob > F      =  0.0000
Residual |  33487.9224   566  59.1659406               R-squared     =  0.1279
---------+------------------------------               Adj R-squared =  0.1232
   Total |  38397.0371   569  67.4816117               Root MSE      =  7.6919

------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .7115506   .1612235      4.413   0.000       .3948811     1.02822
  ASVABC |   .1104595   .0504223      2.191   0.029       .0114219    .2094972
  ASVAB5 |   .0770794   .0463868      1.662   0.097      -.0140319    .1681908
   _cons |  -5.944977   2.161409     -2.751   0.006      -10.19034   -1.699616
------------------------------------------------------------------------------
```

Методы устранения мультиколлинеарности

```
. reg EARNINGS S ASVABC ASVAB5 TENURE MALE URBAN

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  6,   563) =   23.60
   Model |  7715.87322     6  1285.97887           Prob > F      =  0.0000
Residual |  30681.1638   563  54.4958505           R-squared     =  0.2009
---------+------------------------------           Adj R-squared =  0.1924
   Total |  38397.0371   569  67.4816117           Root MSE      =  7.3821

------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .8137184   .1563975      5.203   0.000       .5065245    1.120912
  ASVABC |   .0442801    .049716      0.891   0.373      -.0533714    .1419317
  ASVAB5 |   .1113769   .0458757      2.428   0.016       .0212685    .2014853
  TENURE |    .287038   .0676471      4.243   0.000       .1541665    .4199095
    MALE |   3.123929     .64685      4.829   0.000       1.853395    4.394463
   URBAN |   2.061867   .7274286      2.834   0.005       .6330618    3.490672
   _cons |  -10.60023   2.195757     -4.828   0.000      -14.91311   -6.287358
------------------------------------------------------------------------------
```

Методы устранения мультиколлинеарности

```
. reg EARNINGS S ASVABC ASVAB5

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =   27.66
   Model |  4909.11468     3  1636.37156           Prob > F      =  0.0000
Residual |  33487.9224   566  59.1659406           R-squared     =  0.1279
---------+------------------------------           Adj R-squared =  0.1232
   Total |  38397.0371   569  67.4816117           Root MSE      =  7.6919


. reg EARNINGS S ASVABC ASVAB5 TENURE MALE URBAN

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  6,   563) =   23.60
   Model |  7715.87322     6  1285.97887           Prob > F      =  0.0000
Residual |  30681.1638   563  54.4958505           R-squared     =  0.2009
---------+------------------------------           Adj R-squared =  0.1924
   Total |  38397.0371   569  67.4816117           Root MSE      =  7.3821
```

Методы устранения мультиколлинеарности

```
. reg EARNINGS S ASVABC ASVAB5
```

EARNINGS	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.7115506	.1612235	4.413	0.000	.3948811	1.02822
ASVABC	.1104595	.0504223	2.191	0.029	.0114219	.2094972
ASVAB5	.0770794	.0463868	1.662	0.097	-.0140319	.1681908
_cons	-5.944977	2.161409	-2.751	0.006	-10.19034	-1.699616

```
. reg EARNINGS S ASVABC ASVAB5 TENURE MALE URBAN
```

EARNINGS	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.8137184	.1563975	5.203	0.000	.5065245	1.120912
ASVABC	.0442801	.049716	0.891	0.373	-.0533714	.1419317
ASVAB5	.1113769	.0458757	2.428	0.016	.0212685	.2014853
TENURE	.287038	.0676471	4.243	0.000	.1541665	.4199095
MALE	3.123929	.64685	4.829	0.000	1.853395	4.394463
URBAN	2.061867	.7274286	2.834	0.005	.6330618	3.490672
_cons	-10.60023	2.195757	-4.828	0.000	-14.91311	-6.287358

Стандартные ошибки не изменились, а коэффициенты повели себя нестабильно

Методы устранения мультиколлинеарности

Чтобы уменьшить стандартные ошибки коэффициентов следует уменьшить вариацию случайного фактора

$$\operatorname{Var} b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n\operatorname{Var}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(2) Уменьшение $\boldsymbol{\sigma_u^2}$ за счет включения в модель бол ьшего количества наблюдений, стратифицированной выборки, использования квартальных данных вместо годовых и т.п.

Методы устранения мультиколлинеарности

```
. reg EARNINGS S ASVABC ASVAB5
                                                  Number of obs =      570
------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .7115506   .1612235      4.413   0.000       .3948811    1.02822
  ASVABC |   .1104595   .0504223      2.191   0.029       .0114219    .2094972
  ASVAB5 |   .0770794   .0463868      1.662   0.097      -.0140319    .1681908
   _cons |  -5.944977   2.161409     -2.751   0.006      -10.19034   -1.699616
------------------------------------------------------------------------------
```

```
. reg EARNINGS S ASVABC ASVAB5
                                                  Number of obs =     2868
------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   1.002693   .0787447     12.733   0.000       .8482905    1.157095
  ASVABC |   .1448345   .0241135      6.006   0.000        .097553    .1921161
  ASVAB5 |   .0483846   .0218352      2.216   0.027       .0055703     .091199
   _cons |  -9.654593   1.033311     -9.343   0.000       -11.6807   -7.628485
------------------------------------------------------------------------------
```

Методы устранения мультиколлинеарности

Чтобы уменьшить стандартные ошибки коэффициентов следует уменьшить вариацию случайного фактора

$$\mathrm{Var}\, b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n\mathrm{Var}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(3) Уменьшение $\boldsymbol{\sigma_u^2}$ за счет увеличения вариации регрессора

Методы устранения мультиколлинеарности

Чтобы уменьшить стандартные ошибки коэффициентов следует уменьшить вариацию случайного фактора

$$\mathrm{Var}\, b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n\mathrm{Var}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(4) Уменьшение $\boldsymbol{\sigma_u^2}$ за счет уменьшения корреляции между регрессорами

Методы устранения мультиколлинеарности

Чтобы уменьшить стандартные ошибки коэффициентов следует уменьшить вариацию случайного фактора

$$\text{Var } b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n\text{Var}(X_2)} \times \frac{1}{1 - r_{X_2,X_3}^2}$$

(5) Уменьшение $\boldsymbol{\sigma_u^2}$ за счет комбинирования связанных регресоров

Методы устранения мультиколлинеарности

```
. reg EARNINGS S ASVAB2 ASVAB3 ASVAB4

  Source |       SS       df       MS
---------+------------------------------
   Model |  5906.47726     4  1476.61931
Residual |  32490.5598   565  57.5054156
---------+------------------------------
   Total |  38397.0371   569  67.4816117
```

```
. cor ASVAB2 ASVAB3 ASVAB4
(obs=570)

        |   ASVAB2   ASVAB3   ASVAB4
--------+---------------------------
  ASVAB2|   1.0000
  ASVAB3|   0.6916   1.0000
  ASVAB4|   0.6536   0.7628   1.0000
```

```
------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .7362439   .1586812      4.640   0.000       .4245668    1.047921
  ASVAB2 |   .2472668   .0472249      5.236   0.000        .154509    .3400246
  ASVAB3 |   .0137422    .058716      0.234   0.815      -.1015861    .1290705
  ASVAB4 |  -.1051868   .0544682     -1.931   0.054      -.2121716     .001798
   _cons |  -4.734303    2.06706     -2.290   0.022      -8.794363   -.6742428
------------------------------------------------------------------------------
```

Методы устранения мультиколлинеарности

```
. reg EARNINGS S ASVABC

------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .7390366   .1606216      4.601   0.000       .4235506    1.054523
  ASVABC |   .1545341   .0429486      3.598   0.000       .0701764    .2388918
   _cons |  -4.624749     2.0132     -2.297   0.022      -8.578989   -.6705095
------------------------------------------------------------------------------
```

```
. reg EARNINGS S ASVAB2 ASVAB3 ASVAB4

------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .7362439   .1586812      4.640   0.000       .4245668    1.047921
  ASVAB2 |   .2472668   .0472249      5.236   0.000        .154509    .3400246
  ASVAB3 |   .0137422    .058716      0.234   0.815      -.1015861    .1290705
  ASVAB4 |  -.1051868   .0544682     -1.931   0.054      -.2121716     .001798
   _cons |  -4.734303    2.06706     -2.290   0.022      -8.794363   -.6742428
------------------------------------------------------------------------------
```

Методы устранения мультиколлинеарности

Чтобы уменьшить стандартные ошибки коэффициентов следует уменьшить вариацию случайного фактора

$$\operatorname{Var} b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n\operatorname{Var}(X_2)} \times \frac{1}{1 - r_{X_2,X_3}^2}$$

(6) Уменьшение $\boldsymbol{\sigma_u^2}$ за счет удаления незначимого регрессора

Методы устранения мультиколлинеарности

```
. reg EARNINGS S ASVAB2

------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .6449415   .1519755      4.244   0.000       .3464378    .9434452
  ASVAB2 |   .2019724   .0376567      5.364   0.000       .1280086    .2759361
   _cons |  -5.796398   1.957987     -2.960   0.003      -9.642191   -1.950605
------------------------------------------------------------------------------

. reg EARNINGS S ASVAB2 ASVAB3 ASVAB4

------------------------------------------------------------------------------
EARNINGS |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .7362439   .1586812      4.640   0.000       .4245668    1.047921
  ASVAB2 |   .2472668   .0472249      5.236   0.000        .154509    .3400246
  ASVAB3 |   .0137422    .058716      0.234   0.815      -.1015861    .1290705
  ASVAB4 |  -.1051868   .0544682     -1.931   0.054      -.2121716     .001798
   _cons |  -4.734303    2.06706     -2.290   0.022      -8.794363   -.6742428
------------------------------------------------------------------------------
```

Методы устранения мультиколлинеарности

Чтобы уменьшить стандартные ошибки коэффициентов следует уменьшить вариацию случайного фактора

$$\mathrm{Var}\, b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n\mathrm{Var}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(7) Уменьшение $\boldsymbol{\sigma_u^2}$ за счет очищения влияния одних регрессоров на другие

Методы устранения мультиколлинеарности

Чтобы уменьшить стандартные ошибки коэффициентов следует уменьшить вариацию случайного фактора

$$\operatorname{Var} b_2 = \sigma_{b_2}^2 = \frac{\sigma_u^2}{n \operatorname{Var}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

8) Уменьшение $\boldsymbol{\sigma_u^2}$ за счет теоретического ограничения на связь регрессоров

Методы устранения мультиколлинеарности

```
. reg S ASVABC SM SF

  Source |       SS       df       MS              Number of obs =     570
---------+------------------------------           F(  3,   566) =  110.83
   Model |  1278.24153     3  426.080508           Prob > F      =  0.0000
Residual |  2176.00584   566  3.84453329           R-squared     =  0.3700
---------+------------------------------           Adj R-squared =  0.3667
   Total |  3454.24737   569  6.07073351           Root MSE      =  1.9607

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1295006   .0099544     13.009   0.000       .1099486    .1490527
      SM |    .069403   .0422974      1.641   0.101       -.013676     .152482
      SF |   .1102684   .0311948      3.535   0.000       .0489967    .1715401
   _cons |   4.914654   .5063527      9.706   0.000       3.920094    5.909214
------------------------------------------------------------------------------
```

```
. cor SM SF
(obs=570)
        |       SM       SF
--------+------------------
      SM|   1.0000
      SF|   0.6391   1.0000
```

Методы устранения мультиколлинеарности

8) Уменьшение $\boldsymbol{\sigma}_u^2$ за счет теоретического ограничения на связь регрессоров

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + \beta_4 SF + u$$

$$\beta_2 = \beta_3$$

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 (SM + SF) + u$$

$$= \beta_1 + \beta_2 ASVABC + \beta_3 SP + u$$

Методы устранения мультиколлинеарности

```
. g SP=SM+SF

. reg S ASVABC SP

  Source |       SS       df       MS                  Number of obs =     570
---------+------------------------------               F(  2,   567) =  166.22
   Model |  1276.73764     2  638.368819               Prob > F      =  0.0000
Residual |  2177.50973   567  3.84040517               R-squared     =  0.3696
---------+------------------------------               Adj R-squared =  0.3674
   Total |  3454.24737   569  6.07073351               Root MSE      =  1.9597

------------------------------------------------------------------------------
       S |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ASVABC |   .1295653   .0099485     13.024   0.000       .1100249    .1491057
      SP |    .093741   .0165688      5.658   0.000       .0611973    .1262847
   _cons |   4.823123   .4844829      9.955   0.000       3.871523    5.774724
------------------------------------------------------------------------------
```

Методы устранения мультиколлинеарности

```
. g SP=SM+SF
```

```
. reg S ASVABC SP
```

S	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ASVABC	.1295653	.0099485	13.024	0.000	.1100249	.1491057
SP	.093741	.0165688	5.658	0.000	.0611973	.1262847
_cons	4.823123	.4844829	9.955	0.000	3.871523	5.774724

```
. reg S ASVABC SM SF
```

S	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ASVABC	.1295006	.0099544	13.009	0.000	.1099486	.1490527
SM	.069403	.0422974	1.641	0.101	-.013676	.152482
SF	.1102684	.0311948	3.535	0.000	.0489967	.1715401
_cons	4.914654	.5063527	9.706	0.000	3.920094	5.909214

Устранение мультиколлинеарности. Преобразование переменных

Пусть эмпирическое уравнение регрессии имеет вид:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

где X_1, X_2 – коррелированные переменные

В моделях:

$$\frac{\hat{Y}}{X_1} = b'_0 + b'_1 \frac{X_2}{X_1} \qquad \frac{\hat{Y}}{X_2} = b''_0 + b''_1 \frac{X_1}{X_2}$$

мультиколлинеарность отсутствует.

Конец лекции