

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

Что вы можете сказать о мультиколлинearности, если вы столкнулись с этим? Мы обсудим некоторые возможные меры, рассматривая модель с двумя объясняющими переменными.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

Прежде чем сделать это, необходимо подчеркнуть два важных момента. Во-первых, мультиколлинеарность не приводит к смещению коэффициентов регрессии.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

Проблема в том, что они имеют неудовлетворительно большие отклонения.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

Во-вторых, стандартные ошибки и t-тесты остаются в силе. Стандартные ошибки больше, чем они были бы в отсутствие мультиколлинеарности, предупреждая нас, что оценки регрессии неустойчивы.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n\text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

Проблема мультиколлинеарности обусловлена тем, что дисперсии коэффициентов неудовлетворительно велики. В этой последовательности мы рассмотрим возможные прямые методы их сокращения. В следующей последовательности мы рассмотрим возможные косвенные методы.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

Мы сосредоточимся на коэффициенте наклона и рассмотрим различные компоненты его дисперсии.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

Мы сосредоточимся на коэффициенте наклона и рассмотрим различные компоненты его дисперсии.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

Мы сосредоточимся на коэффициенте наклона и рассмотрим различные компоненты его дисперсии.



$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

Мы сосредоточимся на коэффициенте наклона и рассмотрим различные компоненты его дисперсии.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

Начнем с  $n$ , количества наблюдений.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Source	SS	df	MS			
Model	1235.0519	3	411.683966	Number of obs =	500	
Residual	2518.9701	496	5.07856875	F( 3, 496) =	81.06	
Total	3754.022	499	7.52309018	Prob > F =	0.0000	
				R-squared =	0.3290	
				Adj R-squared =	0.3249	
				Root MSE =	2.2536	

  

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.242527	.123587	10.05	0.000	.999708	1.485345
SM	.091353	.0459299	1.99	0.047	.0011119	.1815941
SF	.2028911	.0425117	4.77	0.000	.1193658	.2864163
_cons	10.59674	.6142778	17.25	0.000	9.389834	11.80365

Предположим, что исследуются детерминанты образовательного уровня. На выходе показан результат установки базовой спецификации с использованием набора данных EAWWE 21. Зависимая переменная S - это годы обучения респондента.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕЛИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Source	SS	df	MS			
Model	1235.0519	3	411.683966	Number of obs =	500	
Residual	2518.9701	496	5.07856875	F( 3, 496) =	81.06	
				Prob > F =	0.0000	
				R-squared =	0.3290	
				Adj R-squared =	0.3249	
Total	3754.022	499	7.52309018	Root MSE =	2.2536	

  

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.242527	.123587	10.05	0.000	.999708	1.485345
SM	.091353	.0459299	1.99	0.047	.0011119	.1815941
SF	.2028911	.0425117	4.77	0.000	.1193658	.2864163
_cons	10.59674	.6142778	17.25	0.000	9.389834	11.80365

Пояснительными переменными являются ASVABC, SM и SF. ASVABC - это оценка по когнитивной способности, масштабируемая таким образом, что она имеет среднее значение 0 и стандартное отклонение 1. SM и SF являются годами обучения матери и отца респондента, соответственно.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Source	SS	df	MS	Number of obs = 500		
Model	1235.0519	3	411.683966	F( 3, 496)	=	81.06
Residual	2518.9701	496	5.07856875	Prob > F	=	0.0000
				R-squared	=	0.3290
				Adj R-squared	=	0.3249
Total	3754.022	499	7.52309018	Root MSE	=	2.2536

  

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.242527	.123587	10.05	0.000	.999708	1.485345
SM	.091353	.0459299	1.99	0.047	.0011119	.1815941
SF	.2028911	.0425117	4.77	0.000	.1193658	.2864163
_cons	10.59674	.6142778	17.25	0.000	9.389834	11.80365

Все коэффициенты регрессии являются положительными, как и ожидалось, но для SM достаточно просто на уровне 5 процентов.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Source	SS	df	MS			
Model	1235.0519	3	411.683966	Number of obs = 500		
Residual	2518.9701	496	5.07856875	F( 3, 496) = 81.06		
				Prob > F = 0.0000		
				R-squared = 0.3290		
				Adj R-squared = 0.3249		
				Root MSE = 2.2536		
S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.242527	.123587	10.05	0.000	.999708	1.485345
SM	.091353	.0459299	1.99	0.047	.0011119	.1815941
SF	.2028911	.0425117	4.77	0.000	.1193658	.2864163
_cons	10.59674	.6142778	17.25	0.000	9.389834	11.80365

Материнское образование можно разумно считать мощным детерминантом образовательного уровня. Соответственно, предельное значение коэффициента SM является неожиданным.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Source	SS	df	MS	. cor SM SF (obs=500)		
Model	1235.0519	3	411.683966			
Residual	2518.9701	496	5.07856875			
Total	3754.022	499	7.52309018			

  

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.242527	.123587	10.05	0.000	.999708	1.485345
SM	.091353	.0459299	1.99	0.047	.0011119	.1815941
SF	.2028911	.0425117	4.77	0.000	.1193658	.2864163
_cons	10.59674	.6142778	17.25	0.000	9.389834	11.80365

  

	SM	SF
SM	1.0000	
SF	0.5312	1.0000

Наиболее вероятным объяснением является многоколлинеарность, учитывая, что из-за assortативного спаривания корреляция между SM и SF составляет 0,53.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Source	SS	df	MS	Number of obs = 2274		
Model	5939.99328	3	1979.99776	F( 3, 2270) =	397.47	
Residual	11308.1092	2270	4.98154589	Prob > F =	0.0000	
Total	17248.1025	2273	7.58825449	R-squared =	0.3444	
				Adj R-squared =	0.3435	
				Root MSE =	2.2319	

  

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.23488	.0556274	22.20	0.000	1.125794	1.343966
SM	.1477984	.0222765	6.63	0.000	.1041139	.1914829
SF	.1527509	.0197091	7.75	0.000	.1141011	.1914007
_cons	10.28846	.2834715	36.29	0.000	9.732573	10.84435

Вот результат регрессии с той же спецификацией, используя все 2 274 наблюдения в базе данных EAWE.



# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Number of obs = 2274

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.23488	.0556274	22.20	0.000	1.125794	1.343966
SM	.1477984	.0222765	6.63	0.000	.1041139	.1914829
SF	.1527509	.0197091	7.75	0.000	.1141011	.1914007
_cons	10.28846	.2834715	36.29	0.000	9.732573	10.84435

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

Number of obs = 500

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.242527	.123587	10.05	0.000	.999708	1.485345
SM	.091353	.0459299	1.99	0.047	.0011119	.1815941
SF	.2028911	.0425117	4.77	0.000	.1193658	.2864163
_cons	10.59674	.6142778	17.25	0.000	9.389834	11.80365

Сравнивая этот результат с использованием набора данных 21, мы видим, что стандартные ошибки намного меньше, как и ожидалось.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Number of obs = 2274

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.23488	.0556274	22.20	0.000	1.125794	1.343966
SM	.1477984	.0222765	6.63	0.000	.1041139	.1914829
SF	.1527509	.0197091	7.75	0.000	.1141011	.1914007
_cons	10.28846	.2834715	36.29	0.000	9.732573	10.84435

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

Number of obs = 500

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.242527	.123587	10.05	0.000	.999708	1.485345
SM	.091353	.0459299	1.99	0.047	.0011119	.1815941
SF	.2028911	.0425117	4.77	0.000	.1193658	.2864163
_cons	10.59674	.6142778	17.25	0.000	9.389834	11.80365

Как следствие, статистика  $t$  всех переменных велика и коэффициенты существенно отличаются от нуля на уровне 0,1 процента. Проблема мультиколлинеарности исчезла.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Number of obs = 2274

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.23488	.0556274	22.20	0.000	1.125794	1.343966
SM	.1477984	.0222765	6.63	0.000	.1041139	.1914829
SF	.1527509	.0197091	7.75	0.000	.1141011	.1914007
_cons	10.28846	.2834715	36.29	0.000	9.732573	10.84435

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

Number of obs = 500

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.242527	.123587	10.05	0.000	.999708	1.485345
SM	.091353	.0459299	1.99	0.047	.0011119	.1815941
SF	.2028911	.0425117	4.77	0.000	.1193658	.2864163
_cons	10.59674	.6142778	17.25	0.000	9.389834	11.80365

Этот пример является искусственным, потому что набор данных 21 является подмножеством базы данных EAWЕ. На практике вы всегда будете использовать весь набор данных, если только он не будет настолько огромным, как в случае переписи населения, что затраты на его обработку полностью не могут быть оправданы.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(1) Увеличьте количество наблюдений.

Опросы: увеличение бюджета, использование кластеризации.

Если вы работаете с данными поперечного сечения (люди, домашние хозяйства, предприятия и т. Д.), И вы проводите опрос, вы можете увеличить размер выборки, обсудив больший бюджет на стадии планирования.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(1) Увеличьте количество наблюдений.

Опросы: увеличение бюджета, использование кластеризации.

Кроме того, вы можете сделать фиксированный бюджет пойти дальше, используя метод, известный как кластеризация. Вы географически разбиваете страну по почтовому индексу или почтовой области.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(1) Увеличьте количество наблюдений.

Опросы: увеличение бюджета, использование кластеризации.

You select a number of these randomly, perhaps using stratified random sampling to make sure that metropolitan, other urban, and rural areas are properly represented.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(1) Увеличьте количество наблюдений.

Опросы: увеличение бюджета, использование кластеризации.

You then confine the survey to the areas selected. This reduces the travel time and cost of the fieldworkers, allowing them to interview a greater number of respondents.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

**(1) Увеличьте количество наблюдений.**

**Опросы: увеличение бюджета, использование кластеризации.**

**Временные ряды: используйте ежеквартально вместо годовых данных.**

**Если вы работаете с данными временных рядов, вы можете увеличить выборку, работая с более короткими интервалами времени для данных, например ежеквартальные или даже ежемесячные данные, а не годовые данные.**



$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

- (2) Уменьшите,  $\sigma_u^2$  включив в модель дополнительные релевантные переменные

Мы могли бы уменьшить дисперсию, добавив в модель больше переменных и уменьшив  $\sigma_u^2$  дисперсию условия нарушения.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Source	SS	df	MS			
Model	1291.03572	4	322.758931	Number of obs =	500	
Residual	2462.98628	495	4.97572985	F( 4, 495) =	64.87	
Total	3754.022	499	7.52309018	Prob > F =	0.0000	
				R-squared =	0.3439	
				Adj R-squared =	0.3386	
				Root MSE =	2.2306	

  

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.224768	.1224438	10.00	0.000	.9841945	1.465342
SM	.1028448	.0455914	2.26	0.025	.0132684	.1924213
SF	.1942085	.0421586	4.61	0.000	.1113766	.2770405
<b>MALE</b>	<b>-.6720436</b>	<b>.2003524</b>	<b>-3.35</b>	0.001	-1.06569	-.2783976
_cons	10.89728	.6145927	17.73	0.000	9.689752	12.10482

Здесь мы добавили в спецификацию фиктивную переменную MALE. Использование фиктивных переменных будет подробно описано в главе 5.d спецификации.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Source	SS	df	MS			
Model	1291.03572	4	322.758931	Number of obs =	500	
Residual	2462.98628	495	4.97572985	F( 4, 495) =	64.87	
Total	3754.022	499	7.52309018	Prob > F =	0.0000	
				R-squared =	0.3439	
				Adj R-squared =	0.3386	
				Root MSE =	2.2306	

  

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.224768	.1224438	10.00	0.000	.9841945	1.465342
SM	.1028448	.0455914	2.26	0.025	.0132684	.1924213
SF	.1942085	.0421586	4.61	0.000	.1113766	.2770405
<b>MALE</b>	<b>-.6720436</b>	<b>.2003524</b>	<b>-3.35</b>	0.001	-1.06569	-.2783976
_cons	10.89728	.6145927	17.73	0.000	9.689752	12.10482

Коэффициент MALE указывает на то, что, контролируя другие характеристики, мужчины, как правило, имеют менее 0,67 года меньше школьного образования, чем женщины.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕЛИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Source	SS	df	MS			
Model	1291.03572	4	322.758931	Number of obs = 500		
Residual	2462.98628	495	4.97572985	F( 4, 495) = 64.87		
				Prob > F = 0.0000		
				R-squared = 0.3439		
				Adj R-squared = 0.3386		
				Root MSE = 2.2306		
S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.224768	.1224438	10.00	0.000	.9841945	1.465342
SM	.1028448	.0455914	2.26	0.025	.0132684	.1924213
SF	.1942085	.0421586	4.61	0.000	.1113766	.2770405
<b>MALE</b>	<b>-.6720436</b>	<b>.2003524</b>	<b>-3.35</b>	0.001	-1.06569	-.2783976
_cons	10.89728	.6145927	17.73	0.000	9.689752	12.10482

Значение t и значение p показывают, что эффект очень значителен, поэтому включение MALE, похоже, улучшило спецификацию.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

Source	SS	df	MS	
Model	1291.03572	4	322.758931	Number of obs = 500
Residual	2462.98628	495	4.97572985	F( 4, 495) = 64.87
Total	3754.022	499	7.52309018	Prob > F = 0.0000
				R-squared = 0.3439
				Adj R-squared = 0.3386
				Root MSE = 2.2306

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

Source	SS	df	MS	
Model	1235.0519	3	411.683966	Number of obs = 500
Residual	2518.9701	496	5.07856875	F( 3, 496) = 81.06
Total	3754.022	499	7.52309018	Prob > F = 0.0000
				R-squared = 0.3290
				Adj R-squared = 0.3249
				Root MSE = 2.2536

С включением MALE RSS упал с 2519 до 2463. Как следствие, упал с 5,08 до 4,98.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.224768	.1224438	10.00	0.000	.9841945	1.465342
<b>SM</b>	<b>.1028448</b>	<b>.0455914</b>	<b>2.26</b>	0.025	.0132684	.1924213
SF	.1942085	.0421586	4.61	0.000	.1113766	.2770405
MALE	-.6720436	.2003524	-3.35	0.001	-1.06569	-.2783976
_cons	10.89728	.6145927	17.73	0.000	9.689752	12.10482

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

ASVABC	1.242527	.123587	10.05	0.000	.999708	1.485345
<b>SM</b>	<b>.091353</b>	<b>.0459299</b>	<b>1.99</b>	0.047	.0011119	.1815941
SF	.2028911	.0425117	4.77	0.000	.1193658	.2864163
_cons	10.59674	.6142778	17.25	0.000	9.389834	11.80365

Хотя это шаг в правильном направлении, он очень мал. Как следствие, уменьшение стандартной погрешности коэффициента SM очень мало. Увеличение ее статистики объясняется главным образом увеличением оценочного коэффициента.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.224768	.1224438	10.00	0.000	.9841945	1.465342
<b>SM</b>	<b>.1028448</b>	<b>.0455914</b>	<b>2.26</b>	0.025	.0132684	.1924213
SF	.1942085	.0421586	4.61	0.000	.1113766	.2770405
MALE	-.6720436	.2003524	-3.35	0.001	-1.06569	-.2783976
_cons	10.89728	.6145927	17.73	0.000	9.689752	12.10482

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

ASVABC	1.242527	.123587	10.05	0.000	.999708	1.485345
<b>SM</b>	<b>.091353</b>	<b>.0459299</b>	<b>1.99</b>	0.047	.0011119	.1815941
SF	.2028911	.0425117	4.77	0.000	.1193658	.2864163
_cons	10.59674	.6142778	17.25	0.000	9.389834	11.80365

Этот результат на самом деле довольно типичен. Вероятно, вы уже включили все основные переменные в исходную спецификацию, поэтому те, которые останутся в вашем наборе данных, скорее всего, будут незначительными.

# ВОЗМОЖНЫЕ ПРЯМЫЕ МЕРЫ, КАСАЮЩИЕСЯ АЛЛЕВИРОВАНИЯ МУЛЬТИКОЛИНЕЙНОСТИ

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

S	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ASVABC	1.224768	.1224438	10.00	0.000	.9841945	1.465342
<b>SM</b>	<b>.1028448</b>	<b>.0455914</b>	<b>2.26</b>	0.025	.0132684	.1924213
SF	.1942085	.0421586	4.61	0.000	.1113766	.2770405
MALE	-.6720436	.2003524	-3.35	0.001	-1.06569	-.2783976
_cons	10.89728	.6145927	17.73	0.000	9.689752	12.10482

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

ASVABC	1.242527	.123587	10.05	0.000	.999708	1.485345
<b>SM</b>	<b>.091353</b>	<b>.0459299</b>	<b>1.99</b>	0.047	.0011119	.1815941
SF	.2028911	.0425117	4.77	0.000	.1193658	.2864163
_cons	10.59674	.6142778	17.25	0.000	9.389834	11.80365

Этот подход может даже иметь эффект, противоположный намеченному. Стандартные ошибки существующих переменных в спецификации могут фактически увеличиваться, если новая переменная коррелирует с ними.



$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(3) Увеличить  $\text{MSD}(X_2)$ .

Третьим возможным способом уменьшения проблемы многоколлинеарности может быть увеличение вариации объясняющих переменных. Это возможно только на этапе проектирования обследования.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(3) Увеличить  $\text{MSD}(X_2)$ .

Например, если вы планируете опрос домашних хозяйств с целью изучения того, как структура расходов варьируется в зависимости от дохода, вы должны убедиться, что в выборку включены относительно богатые и относительно бедные домашние хозяйства, а также домохозяйства со средним доходом.

$$\sigma_{\hat{\beta}_2}^2 = \frac{\sigma_u^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2, X_3}^2} = \frac{\sigma_u^2}{n \text{MSD}(X_2)} \times \frac{1}{1 - r_{X_2, X_3}^2}$$

(4) Уменьшить  $r_{X_2, X_3}$

Другая возможность может заключаться в уменьшении корреляции между объясняющими переменными. Опять же, это возможно только на этапе проектирования опроса, и даже тогда это редко возможно с экономическими данными.