

	Review chapter	Модель регрессии
Модель	X неизвестное μ , σ^2	
Оценочная функция	\bar{X}	

Эта последовательность описывает тестирование гипотез, относящихся к коэффициентам регрессии. Он касается только методики проведения, а не теории.

	Review chapter	Модель регрессии
Модель	X неизвестное μ , σ^2	
Оценочная функция	\bar{X}	

Тестирование гипотез составляет основную часть основы эконометрики, и важно иметь четкое понимание теории.

	Review chapter	Модель регрессии
Модель	X неизвестное μ , σ^2	
Оценочная функция	\bar{X}	

Теория, обсуждаемая в разделах R.9-R.11 главы обзора, является нетривиальной и требует тщательного изучения. Эта последовательность является чисто механической и никоим образом не является заменителем.

	Review chapter	Модель регрессии
Модель	X неизвестное μ , σ^2	
Оценочная функция	\bar{X}	

Если вы не понимаете, например, взаимные уступки между размером (уровнем значимости) и мощностью теста, вы должны изучить материал в этих разделах, прежде чем смотреть на эту последовательность.

	Review chapter	Модель регрессии
Модель	X неизвестное μ , σ^2	
Оценочная функция	\bar{X}	

В нашем стандартном примере в главе «Обзор» мы имели случайную переменную X с неизвестным средним населением variance μ и дисперсией σ^2 . Учитывая выборку данных, мы использовали среднее значение выборки как оценку μ .

	Review chapter	Модель регрессии
Модель	X неизвестное μ , σ^2	$Y = \beta_1 + \beta_2 X + u$
Оценочная функция	\bar{X}	$b_2 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}$

В контексте модели регрессии мы имеем параметры β_1 и β_2 и для них были получены оценки b_1 и b_2 . В дальнейшем мы сосредоточимся на β_2 и его оценке b_2 .

	Review chapter	Модель регрессии
Модель	X неизвестное μ , σ^2	$Y = \beta_1 + \beta_2 X + u$
Оценочная функция	\bar{X}	$b_2 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}$
Нулевая гипотеза	$H_0: \mu = \mu_0$	$H_0: \beta_2 = \beta_2^0$
Альтернативная гипотеза	$H_1: \mu \neq \mu_0$	$H_1: \beta_2 \neq \beta_2^0$

В случае случайной величины X , наша стандартная нулевая гипотеза заключалась в том, что μ было равно некоторому определенному значению μ_0 . В случае модели регрессии наша нулевая гипотеза состоит в том, что β_2 равен некоторому конкретному значению β_2^0 .

	Review chapter	Модель регрессии
Модель	X неизвестное μ , σ^2	$Y = \beta_1 + \beta_2 X + u$
Оценочная функция	\bar{X}	$b_2 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}$
Нулевая гипотеза	$H_0: \mu = \mu_0$	$H_0: \beta_2 = \beta_2^0$
Альтернативная гипотеза	$H_1: \mu \neq \mu_0$	$H_1: \beta_2 \neq \beta_2^0$
Статистика испытаний	$t = \frac{\bar{X} - \mu_0}{\text{s.e.}(\bar{X})}$	$t = \frac{b_2 - \beta_2^0}{\text{s.e.}(b_2)}$

Для обеих популяций среднее значение μ случайной величины X и коэффициент регрессии β_2 , тестовая статистика является t статистикой.

	Review chapter	Модель регрессии
Модель	X неизвестное μ, σ^2	$Y = \beta_1 + \beta_2 X + u$
Оценочная функция	\bar{X}	$b_2 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}$
Нулевая гипотеза	$H_0: \mu = \mu_0$	$H_0: \beta_2 = \beta_2^0$
Альтернативная гипотеза	$H_1: \mu \neq \mu_0$	$H_1: \beta_2 \neq \beta_2^0$
Статистика испытаний	$t = \frac{\bar{X} - \mu_0}{\text{s.e.}(\bar{X})}$	$t = \frac{b_2 - \beta_2^0}{\text{s.e.}(b_2)}$

В обоих случаях он определяется как разница между расчетным коэффициентом и его гипотетическим значением, деленная на стандартную ошибку коэффициента.

	Review chapter	Модель регрессии
Модель	X неизвестное μ , σ^2	$Y = \beta_1 + \beta_2 X + u$
Оценочная функция	\bar{X}	$b_2 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}$
Нулевая гипотеза	$H_0: \mu = \mu_0$	$H_0: \beta_2 = \beta_2^0$
Альтернативная гипотеза	$H_1: \mu \neq \mu_0$	$H_1: \beta_2 \neq \beta_2^0$
Статистика испытаний	$t = \frac{\bar{X} - \mu_0}{\text{s.e.}(\bar{X})}$	$t = \frac{b_2 - \beta_2^0}{\text{s.e.}(b_2)}$
Reject H_0 if	$ t > t_{\text{crit}}$	$ t > t_{\text{crit}}$

Мы отклоняем нулевую гипотезу, если абсолютное значение больше критического значения t , учитывая выбранный уровень значимости.

	Review chapter	Модель регрессии
Модель	X неизвестное μ , σ^2	$Y = \beta_1 + \beta_2 X + u$
Оценочная функция	\bar{X}	$b_2 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}$
Нулевая гипотеза	$H_0: \mu = \mu_0$	$H_0: \beta_2 = \beta_2^0$
Альтернативная гипотеза	$H_1: \mu \neq \mu_0$	$H_1: \beta_2 \neq \beta_2^0$
Статистика испытаний	$t = \frac{\bar{X} - \mu_0}{\text{s.e.}(\bar{X})}$	$t = \frac{b_2 - \beta_2^0}{\text{s.e.}(b_2)}$
Reject H_0 if	$ t > t_{\text{crit}}$	$ t > t_{\text{crit}}$
Степени свободы	$n - 1$	$n - k = n - 2$

Есть одно важное различие. При определении критического значения t необходимо учитывать число степеней свободы. В случае случайной величины X это $n - 1$, где n - количество наблюдений в образце.

	Review chapter	Модель регрессии
Модель	X неизвестное μ , σ^2	$Y = \beta_1 + \beta_2 X + u$
Оценочная функция	\bar{X}	$b_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$
Нулевая гипотеза	$H_0: \mu = \mu_0$	$H_0: \beta_2 = \beta_2^0$
Альтернативная гипотеза	$H_1: \mu \neq \mu_0$	$H_1: \beta_2 \neq \beta_2^0$
Статистика испытаний	$t = \frac{\bar{X} - \mu_0}{\text{s.e.}(\bar{X})}$	$t = \frac{b_2 - \beta_2^0}{\text{s.e.}(b_2)}$
Reject H_0 if	$ t > t_{\text{crit}}$	$ t > t_{\text{crit}}$
Степени свободы	$n - 1$	$n - k = n - 2$

В случае модели регрессии число степеней свободы $n - k$, где n - количество наблюдений в выборке, а k - количество параметров (β коэффициентов). Для простой модели регрессии выше $n - 2$.

Пример: $p = \beta_1 + \beta_2 w + u$

Нулевая гипотеза: $H_0: \beta_2 = 1.0$

Альтернативная гипотеза: $H_1: \beta_2 \neq 1.0$

В качестве иллюстрации мы рассмотрим модель, связанную с инфляцией цен для инфляции заработной платы. p - процентный годовой темп роста цен, w - процентный годовой темп роста заработной платы.

Пример: $p = \beta_1 + \beta_2 w + u$

Нулевая гипотеза: $H_0: \beta_2 = 1.0$

Альтернативная гипотеза: $H_1: \beta_2 \neq 1.0$

Мы проверим гипотезу о том, что темп инфляции цен равен ставке инфляции заработной платы. Поэтому нулевая гипотеза $H_0: \beta_2 = 1.0$. (Мы также должны проверить $\beta_1 = 0$.)

Пример: $p = \beta_1 + \beta_2 w + u$

Нулевая гипотеза: $H_0: \beta_2 = 1.0$

Альтернативная гипотеза: $H_1: \beta_2 \neq 1.0$

$$\hat{p} = 1.21 + 0.82w$$

(0.05) (0.10)

Предположим, что результат регрессии показан (стандартные ошибки в скобках). Наша фактическая оценка коэффициента наклона составляет всего 0,82. Мы проверим, следует ли отклонять нулевую гипотезу.

Пример: $p = \beta_1 + \beta_2 w + u$

Нулевая гипотеза: $H_0: \beta_2 = 1.0$

Альтернативная гипотеза: $H_1: \beta_2 \neq 1.0$

$$\hat{p} = 1.21 + 0.82w$$

(0.05) (0.10)

$$t = \frac{b_2 - \beta_2^0}{\text{s.e.}(b_2)} = \frac{0.82 - 1.00}{0.10} = -1.80.$$

Мы вычисляем t статистику путем вычитания гипотетического истинного значения из оценки выборки и деления на стандартную ошибку. Он достигает -1,80.

Пример: $p = \beta_1 + \beta_2 w + u$

Нулевая гипотеза: $H_0: \beta_2 = 1.0$

Альтернативная гипотеза: $H_1: \beta_2 \neq 1.0$

$$\hat{p} = 1.21 + 0.82w$$

(0.05) (0.10)

$$t = \frac{b_2 - \beta_2^0}{\text{s.e.}(b_2)} = \frac{0.82 - 1.00}{0.10} = -1.80.$$

$n = 20$

degrees of freedom = 18

$t_{\text{crit}, 5\%} = 2.101$

В выборке имеется 20 наблюдений. Мы оценили 2 параметра, поэтому существует 18 степеней свободы.

Пример: $p = \beta_1 + \beta_2 w + u$

Нулевая гипотеза: $H_0: \beta_2 = 1.0$

Альтернативная гипотеза: $H_1: \beta_2 \neq 1.0$

$$\hat{p} = 1.21 + 0.82w$$

(0.05) (0.10)

$$t = \frac{b_2 - \beta_2^0}{\text{s.e.}(b_2)} = \frac{0.82 - 1.00}{0.10} = -1.80.$$

$n = 20$

degrees of freedom = 18

$t_{\text{crit}, 5\%} = 2.101$

Критическое значение t с 18 степенями свободы составляет 2,101 на уровне 5%. Абсолютное значение t-статистики меньше этого, поэтому мы не отвергаем нулевую гипотезу.

Модель

$$Y = \beta_1 + \beta_2 X + u$$

На практике обычно не имеют представление о фактическом значении коэффициентов. Очень часто цель анализа состоит в том, чтобы продемонстрировать, что на Y влияет X , без какого-либо конкретного предварительного представления о фактических коэффициентах отношения.

Модель
$$Y = \beta_1 + \beta_2 X + u$$

Нулевая гипотеза: $H_0: \beta_2 = 0$

Альтернативная гипотеза: $H_1: \beta_2 \neq 0$

В этом случае обычно определяется $\beta_2 = 0$ как нулевая гипотеза. На словах нулевая гипотеза состоит в том, что X не влияет на Y . Затем мы попытаемся показать, что нулевая гипотеза ложна.

Модель

$$Y = \beta_1 + \beta_2 X + u$$

Нулевая гипотеза: $H_0: \beta_2 = 0$

Альтернативная гипотеза: $H_1: \beta_2 \neq 0$

$$t = \frac{b_2 - \beta_2^0}{\text{s.e.}(b_2)} = \frac{b_2}{\text{s.e.}(b_2)}$$

For the null hypothesis $\beta_2 = 0$, the t statistic reduces to the estimate of the coefficient divided by its standard error.

Модель

$$Y = \beta_1 + \beta_2 X + u$$

Нулевая гипотеза: $H_0: \beta_2 = 0$

Альтернативная гипотеза: $H_1: \beta_2 \neq 0$

$$t = \frac{b_2 - \beta_2^0}{\text{s.e.}(b_2)} = \frac{b_2}{\text{s.e.}(b_2)}$$

Это отношение обычно называют t статистикой для коэффициента и автоматически распечатывается как часть результатов регрессии. Чтобы выполнить тест для заданного уровня значимости, мы напрямую сравниваем статистику t с критическим значением t для этого уровня значимости.

ИСПЫТАНИЕ ГИПОТЕЗЫ, СВЯЗАННОЙ С КОЭФФИЦИЕНТОМ РЕГРЕССИИ

```
. reg EARNINGS S
```

Source	SS	df	MS			
Model	19321.5589	1	19321.5589	Number of obs =	540	
Residual	92688.6722	538	172.283777	F(1, 538) =	112.15	
Total	112010.231	539	207.811189	Prob > F =	0.0000	
				R-squared =	0.1725	
				Adj R-squared =	0.1710	
				Root MSE =	13.126	

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

Вот результат функции заработка, установленный в предыдущем слайд-шоу, с выделенной статистикой t.

ИСПЫТАНИЕ ГИПОТЕЗЫ, СВЯЗАННОЙ С КОЭФФИЦИЕНТОМ РЕГРЕССИИ

```
. reg EARNINGS S
```

Source	SS	df	MS	Number of obs = 540		
Model	19321.5589	1	19321.5589	F(1, 538)	=	112.15
Residual	92688.6722	538	172.283777	Prob > F	=	0.0000
Total	112010.231	539	207.811189	R-squared	=	0.1725
				Adj R-squared	=	0.1710
				Root MSE	=	13.126

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

Вы можете видеть, что t статистика для коэффициента S огромна. Мы отвергли бы нулевую гипотезу о том, что школьное образование не влияет на заработок на уровне значимости 1% (критическое значение около 2,59).

ИСПЫТАНИЕ ГИПОТЕЗЫ, СВЯЗАННОЙ С КОЭФФИЦИЕНТОМ РЕГРЕССИИ

```
. reg EARNINGS S
```

Source	SS	df	MS			
Model	19321.5589	1	19321.5589	Number of obs =	540	
Residual	92688.6722	538	172.283777	F(1, 538) =	112.15	
Total	112010.231	539	207.811189	Prob > F =	0.0000	
				R-squared =	0.1725	
				Adj R-squared =	0.1710	
				Root MSE =	13.126	

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

В этом случае мы могли бы пойти дальше и отвергнуть нулевую гипотезу о том, что обучение не влияет на доход на уровне значимости 0,1%.

ИСПЫТАНИЕ ГИПОТЕЗЫ, СВЯЗАННОЙ С КОЭФФИЦИЕНТОМ РЕГРЕССИИ

```
. reg EARNINGS S
```

Source	SS	df	MS			
Model	19321.5589	1	19321.5589	Number of obs =	540	
Residual	92688.6722	538	172.283777	F(1, 538) =	112.15	
Total	112010.231	539	207.811189	Prob > F =	0.0000	
				R-squared =	0.1725	
				Adj R-squared =	0.1710	
				Root MSE =	13.126	

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

Преимущество отчетности об отказе на уровне 0,1%, а не на уровне 1%, заключается в том, что риск ошибочного отклонения нулевой гипотезы без эффекта теперь составляет всего 0,1% вместо 1%. Результат, следовательно, еще более убедителен.

ИСПЫТАНИЕ ГИПОТЕЗЫ, СВЯЗАННОЙ С КОЭФФИЦИЕНТОМ РЕГРЕССИИ

```
. reg EARNINGS S
```

Source	SS	df	MS			
Model	19321.5589	1	19321.5589	Number of obs =	540	
Residual	92688.6722	538	172.283777	F(1, 538) =	112.15	
Total	112010.231	539	207.811189	Prob > F =	0.0000	
				R-squared =	0.1725	
				Adj R-squared =	0.1710	
				Root MSE =	13.126	

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

T статистика для перехвата также огромна. Однако, поскольку перехват не имеет никакого значения, для него нет смысла выполнять t-тест.

ИСПЫТАНИЕ ГИПОТЕЗЫ, СВЯЗАННОЙ С КОЭФФИЦИЕНТОМ РЕГРЕССИИ

```
. reg EARNINGS S
```

Source	SS	df	MS			
Model	19321.5589	1	19321.5589	Number of obs =	540	
Residual	92688.6722	538	172.283777	F(1, 538) =	112.15	
Total	112010.231	539	207.811189	Prob > F =	0.0000	
				R-squared =	0.1725	
				Adj R-squared =	0.1710	
				Root MSE =	13.126	

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

Следующий столбец на выходе дает так называемые значения p для каждого коэффициента. Это вероятность получения соответствующей t статистики как случайности, если нулевая гипотеза $H_0: \beta = 0$ истина.

ИСПЫТАНИЕ ГИПОТЕЗЫ, СВЯЗАННОЙ С КОЭФФИЦИЕНТОМ РЕГРЕССИИ

```
. reg EARNINGS S
```

Source	SS	df	MS	Number of obs = 540		
Model	19321.5589	1	19321.5589	F(1, 538)	=	112.15
Residual	92688.6722	538	172.283777	Prob > F	=	0.0000
Total	112010.231	539	207.811189	R-squared	=	0.1725
				Adj R-squared	=	0.1710
				Root MSE	=	13.126

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

Если вы отклоните нулевую гипотезу $H_0: \beta = 0$, это вероятность того, что вы совершили ошибку и сделаете ошибку типа I. Поэтому он дает уровень значимости, при котором нулевая гипотеза будет отвергнута.

ИСПЫТАНИЕ ГИПОТЕЗЫ, СВЯЗАННОЙ С КОЭФФИЦИЕНТОМ РЕГРЕССИИ

```
. reg EARNINGS S
```

Source	SS	df	MS			
Model	19321.5589	1	19321.5589	Number of obs =	540	
Residual	92688.6722	538	172.283777	F(1, 538) =	112.15	
Total	112010.231	539	207.811189	Prob > F =	0.0000	
				R-squared =	0.1725	
				Adj R-squared =	0.1710	
				Root MSE =	13.126	

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

Если $p = 0,05$, нулевая гипотеза может быть просто отвергнута на уровне 5%. Если это было 0,01, его можно было бы просто отклонить на уровне 1%. Если бы он составлял 0,001, его можно было бы просто отклонить на уровне 0,1%. Это предполагает, что вы используете двухсторонние тесты.

ИСПЫТАНИЕ ГИПОТЕЗЫ, СВЯЗАННОЙ С КОЭФФИЦИЕНТОМ РЕГРЕССИИ

TESTING A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

```
. reg EARNINGS S
```

Source	SS	df	MS	Number of obs = 540		
Model	19321.5589	1	19321.5589	F(1, 538)	=	112.15
Residual	92688.6722	538	172.283777	Prob > F	=	0.0000
Total	112010.231	539	207.811189	R-squared	=	0.1725
				Adj R-squared	=	0.1710
				Root MSE	=	13.126

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

В данном случае $p =$ от 0 до трех знаков после запятой для коэффициента S. Это означает, что мы можем отклонить нулевую гипотезу $H_0: b_2 = 0$ на уровне 0,1%, не обращаясь к таблице критических значений t . (Тестирование перехвата не имеет смысла в этой регрессии.)

ИСПЫТАНИЕ ГИПОТЕЗЫ, СВЯЗАННОЙ С КОЭФФИЦИЕНТОМ РЕГРЕССИИ

```
. reg EARNINGS S
```

Source	SS	df	MS	Number of obs = 540		
Model	19321.5589	1	19321.5589	F(1, 538)	=	112.15
Residual	92688.6722	538	172.283777	Prob > F	=	0.0000
Total	112010.231	539	207.811189	R-squared	=	0.1725
				Adj R-squared	=	0.1710
				Root MSE	=	13.126

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

Это более информативный подход к представлению результатов испытаний и широко используемых в медицинской литературе.

ИСПЫТАНИЕ ГИПОТЕЗЫ, СВЯЗАННОЙ С КОЭФФИЦИЕНТОМ РЕГРЕССИИ

```
. reg EARNINGS S
```

Source	SS	df	MS	Number of obs = 540		
Model	19321.5589	1	19321.5589	F(1, 538)	=	112.15
Residual	92688.6722	538	172.283777	Prob > F	=	0.0000
Total	112010.231	539	207.811189	R-squared	=	0.1725
				Adj R-squared	=	0.1710
				Root MSE	=	13.126

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
S	2.455321	.2318512	10.59	0.000	1.999876	2.910765
_cons	-13.93347	3.219851	-4.33	0.000	-20.25849	-7.608444

Однако в экономике стандартная практика заключается в представлении результатов, относящихся к уровням значимости 5% и 1%, а иногда и к уровню 0,1% (когда на этом уровне можно отказаться).