

Проверка гипотез, относящихся к коэффициентам регрессии

Пример 1. y – темп общей инфляции в экономике (в %),
 x – темп инфляции, вызванной ростом заработной платы (в %).

$$y_i = ax_i + b + \varepsilon_i, i = \overline{1, n}$$

Гипотеза: общая инфляция увеличивается в той же степени, что и инфляция, вызванная ростом заработной платы

$H_0 : a = 1$ - основная гипотеза

$H_1 : a \neq 1$ - альтернативная гипотеза.

Проверка гипотез, относящихся к коэффициентам регрессии

Пример 2. y – заработная плата,
 x – число лет обучения

$$y_i = ax_i + b + \varepsilon_i, i = \overline{1, n}$$

Гипотеза: заработная плата не зависит от числа лет обучения

$H_0 : a = 0$ - основная гипотеза

$H_1 : a \neq 0$ - альтернативная гипотеза.

Проверка гипотез возможна, только если выполнены 4 условия Гаусса-Маркова, т.е. в классической нормальной линейной модели парной регрессии

$$y_i = ax_i + b + \varepsilon_i, \quad \varepsilon_i \in N(0, \sigma^2), \quad i = \overline{1, n}$$

1. Рассчитываются стандартные ошибки коэффициентов регрессии

$$s_a = \frac{s}{s_x \sqrt{n}} \quad \text{стандартная ошибка коэффициента } a$$

$$s_b = \frac{s \sqrt{x^2}}{s_x \sqrt{n}} \quad \text{стандартная ошибка коэффициента } b$$

$\hat{b}$

Excel делает это автоматически при построении регрессии через Сервис-Анализ данных

Проверка гипотез возможна, только если выполнены 4 условия Гаусса-Маркова, т.е. в классической нормальной линейной модели парной регрессии

$$y_i = ax_i + b + \varepsilon_i, \quad \varepsilon_i \boxtimes N(0, \sigma^2), \quad i = \overline{1, n}$$

2. Выдвигается основная и альтернативная гипотезы о коэффициентах регрессии

$$H_0 : a = a_0 \quad \text{основная}$$

$$H_1 : a \neq a_0 \quad \text{альтернативная}$$

Чаще всего проверяют гипотезы $H_0 : a = 0$ - коэффициент a незначим

$$H_1 : a \neq 0 \quad \text{коэффициент } a \text{ значим}$$

Проверка гипотез возможна, только если выполнены 4 условия Гаусса-Маркова, т.е. в классической нормальной линейной модели парной регрессии

$$y_i = ax_i + b + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = \overline{1, n}$$

3. Задается уровень значимости $\alpha = P(H_1 | H_0)$

Чаще всего $\alpha = 0.05$

$$y_i = ax_i + b + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = \overline{1, n}$$

4. Рассчитывается статистика критерия

$$T_a = \frac{\bar{a} - a_0}{s_a} \quad \text{для проверки гипотезы} \quad H_0 : a = a_0$$

$$T_b = \frac{\bar{b} - b_0}{s_b} \quad \text{для проверки гипотезы} \quad H_0 : b = b_0$$

для гипотез $H_0 : a = 0 \quad H_0 : b = 0$

Excel делает это автоматически при построении регрессии через
Сервис-Анализ данных

$$y_i = ax_i + b + \varepsilon_i, \quad \varepsilon_i \in N(0, \sigma^2), \quad i = \overline{1, n}$$

5. Рассчитывается критическое значение для t статистики с помощью функции Excel

Стьюдраспобр(альфа, n-2)

$$y_i = ax_i + b + \varepsilon_i, \quad \varepsilon_i \in N(0, \sigma^2), \quad i = \overline{1, n}$$

6. Сравнивается t-статистика с критическим значением

Если $|T| \leq t_{krit}$, то принимается основная гипотеза $H_0 : a = a_0$

Иначе альтернативная гипотеза $H_1 : a \neq a_0$

Проверка гипотез, относящихся к коэффициентам регрессии

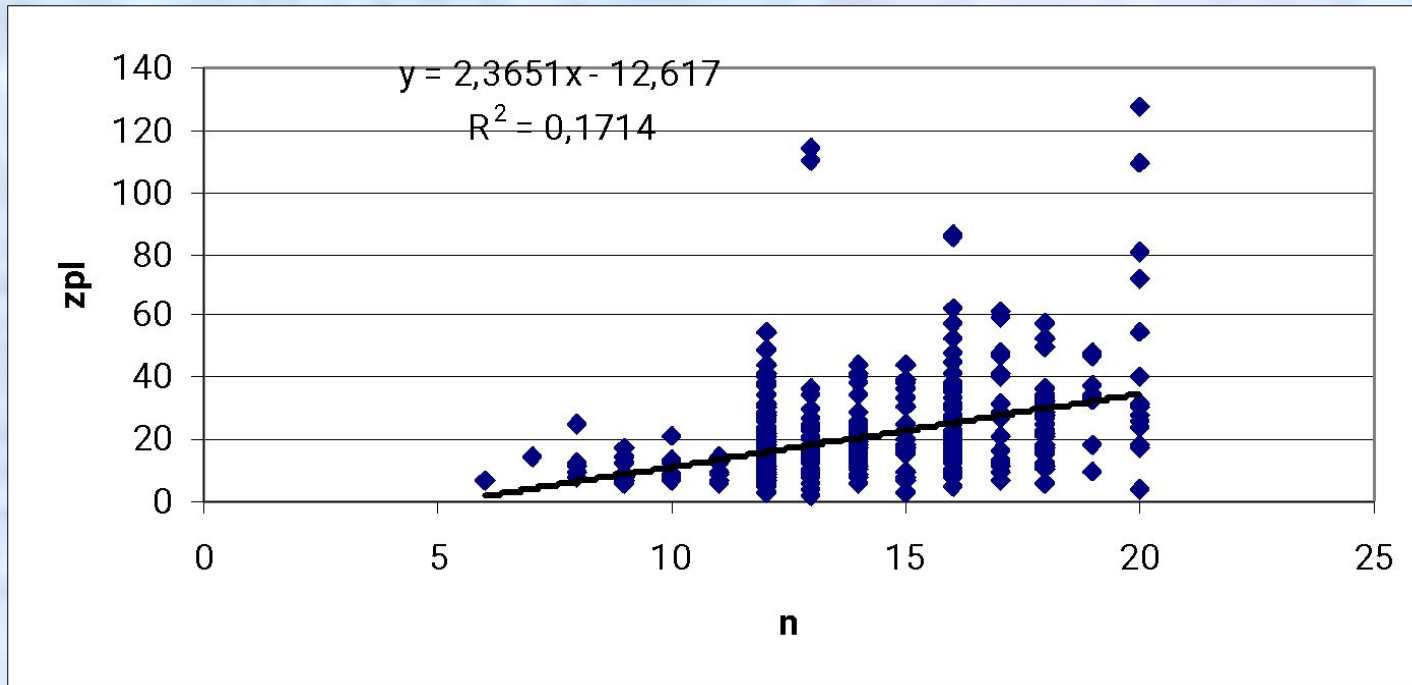
Наиболее часто проверяется гипотеза $H_0 : a = 0$

переменная y не зависит от переменной x . *Уравнение регрессии не значимо.*
В этом случае

$$T_a = \frac{\hat{a}}{s_a}$$

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ

$$Z_{pl} = -12,6188 + 2,3651N$$



$$H_0 : a = 0$$

Уровень образования не влияет на заработную плату.

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ

ВЫВОД ИТОГОВ					
Регрессионная статистика					
Множественный R	0,413988				
R-квадрат	0,171386				
Нормированный R	0,169846				
Стандартная ошибка	13,30384				
Наблюдения	540				
Дисперсионный анализ					
	df	SS	MS	F	значимость
Регрессия	1	19695,21	19695,21	111,2774	8,96E-24
Остаток	538	95221,75	176,9921		
Итого	539	114917			
Коэффициенты					
Y-пересечение	-12,6168	3,118382	-4,04596	5,97E-05	-18,7425
n	2,365094	0,224205	10,54881	8,96E-24	1,92467

стандартные ошибки

t-статистики для гипотез

$$H_0 : a = 0$$

$$H_0 : b = 0$$

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ

ВЫВОД ИТОГОВ					
Регрессионная статистика					
Множественный R	0,413988				
R-квадрат	0,171386				
Нормированный R	0,169846				
Стандартная ошибка	13,30384				
Наблюдения	540				
Дисперсионный анализ					
	df	SS	MS	F	значимость
Регрессия	1	19695,21	19695,21	111,2774	8,96E-24
Остаток	538	95221,75	176,9921		
Итого	539	114917			
Коэффициенты, стандартная ошибка, t-статистика, p-значения, доверительные 95%					
Y-пересечение	-12,6168	3,118382	-4,04596	5,97E-05	-18,7425
x	2,365094	0,224205	10,54881	8,96E-24	1,92467

$$t_{krit} = 1.96$$

=СТЬЮДРАСПОБР(0,05;540-2)

стандартные ошибки

t-статистики для гипотез

$$H_0 : a = 0$$

$$H_0 : b = 0$$

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ

ВЫВОД ИТОГОВ					
Регрессионная статистика					
Множественный R	0,413988				
R-квадрат	0,171386				
Нормированный R	0,169846				
Стандартная ошибка	13,30384				
Наблюдения	540				
Дисперсионный анализ					
	df	SS	MS	F	значимость
Регрессия	1	19695,21	19695,21	111,2774	8,96E-24
Остаток	538	95221,75	176,9921		
Итого	539	114917			
Коэффициенты регрессии, стандартная ошибка, t-статистика, p-значения, доверительные интервалы 95%					
Y-пересечение	-12,6168	3,118382	-4,04596	5,97E-05	-18,7425
n	2,365094	0,224205	10,54881	8,96E-24	1,92467

$$t_{krit} = 1.96$$

=СТЬЮДРАСПОБР(0,05;540-2)

Так как $|-4.05| > 1.96$, принимаем $H_1 : b \neq 0$

Так как $|2.37| > 1.96$, принимаем $H_1 : a \neq 0$

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ

	Коэффициент ы	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	-12,6168	3,118382	-4,04596	5,97E-05
n	2,365094	0,224205	10,54881	8,96E-24

Если гипотетическое значение 0, то проверку гипотезы $H_0 : a = 0$
и $H_0 : b = 0$ можно выполнить проще по P-значению

Если P-значение меньше α , то принимается $H_1 : a \neq 0$

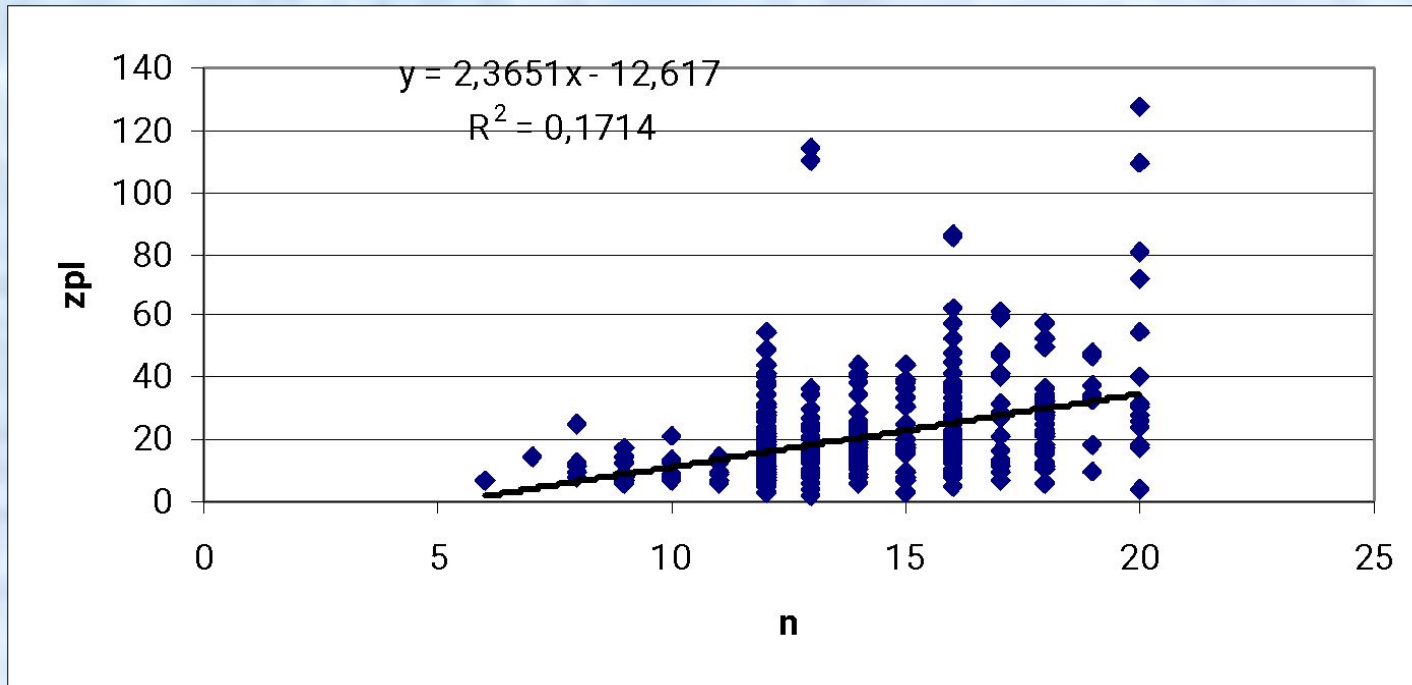
иначе $H_0 : a = 0$

P-значение - это вероятность того, что соответствующая переменная не влияет на зависимую переменную y. При P-значении больше 0,05 обычно считают, что соответствующая переменная незначима и ее можно исключить из уравнения регрессии.

Замечание. Константу из уравнения регрессии удалять нельзя, даже если она незначима.

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ

$$Z_{pl} = -12,6188 + 2,3651N$$



$$H_0 : a = 2$$

Каждый дополнительный год обучения добавляет 2 доллара к заработной плате.

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ

ВЫВОД ИТОГОВ					
Регрессионная статистика					
Множественный R	0,413988				
R-квадрат	0,171386				
Нормированный R	0,169846				
Стандартная ошибка	13,30384				
Наблюдения	540				
Дисперсионный анализ					
	df	SS	MS	F	Значимость
Регрессия	1	19695,21	19695,21	111,2774	8,96E-24
Остаток	538	95221,75	176,9921		
Итого	539	114917			
Коэффициенты					
Y-пересечение	-12,6168	3,118382	-4,04596	5,97E-05	-18,7425
n	2,365094	0,224205	10,54881	8,96E-24	1,92467

стандартные ошибки

$$T_a = \frac{\hat{a} - a_0}{s_a} = \frac{2.365 - 2}{0.224} = 1.63$$

ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ

ВЫВОД ИТОГОВ					
Регрессионная статистика					
Множественный R	0,413988				
R-квадрат	0,171386				
Нормированный R	0,169846				
Стандартная ошибка	13,30384				
Наблюдения	540				
Дисперсионный анализ					
	df	SS	MS	F	значимость
Регрессия	1	19695,21	19695,21	111,2774	8,96E-24
Остаток	538	95221,75	176,9921		
Итого	539	114917			
Коэффициенты регрессии, стандартная ошибка, t-статистика, p-значения, доверительные интервалы 95%					
Y-пересечение	-12,6168	3,118382	-4,04596	5,97E-05	-18,7425
n	2,365094	0,224205	10,54881	8,96E-24	1,92467

$$t_{krit} = 1.96$$

Так как $|1.63| < 1.96$
принимается

$$H_0 : a = 2$$

$$T_a = \frac{\hat{a} - a_0}{s_a} = \frac{2.365 - 2}{0.224} = 1.63$$

Пример проверки гипотезы

Пример 2. y – темп общей инфляции в экономике (в %),
 x – темп инфляции, вызванной ростом заработной платы (в %).

$$\hat{y} = 0,82x - 1,21 \quad \text{В скобках стандартные ошибки}$$

(0,10) (0,05)

$n=20$ наблюдений

Гипотеза: общая инфляция увеличивается в той же степени, что и инфляция, вызванная ростом заработной платы

$H_0 : a = 1$ - основная гипотеза

$H_1 : a \neq 1$ - альтернативная гипотеза.

$$\hat{y} = 0,82x - 1,21 \quad n=20 \text{ наблюдений}$$

(0,10) (0,05)

$$H_0 : a = 1 \quad - \text{основная гипотеза}$$

$$T_a = \frac{\hat{a} - a_0}{s_a} = \frac{0,82 - 1}{0,10} = -1,80$$

$$\alpha = 0,05 \quad = \text{СТЮДРАСПОБР}(0,05; 20-2)$$

$$t_{krit} = 2.1$$

$$|-1.8| \leq 2.1 \quad \text{принимаем} \quad H_0 : a = 1$$

общая инфляция незначимо отличается
от инфляции, вызванной
ростом заработной платы