

1. Диаграмма рассеяния	1
2. Регрессионная модель.....	1
3. Проверка значимости коэффициентов	1
4. Интерпретация.....	1
5. Проверка закона Оукена.....	2
6. Более качественная модель	2
Приложение.....	3

1. Диаграмма рассеяния

Нами была построена диаграмма рассеяния Y от X (Рис 1.), показывающая взаимосвязь двух переменных. График приведен в конце в разделе Таблицы. Диаграмма демонстрирует высокую концентрацию данных вокруг значений 0 - 5 по X и -1.3 - 0.2 по Y , в связи с чем можно предположить, что, скорее всего, линейная зависимость в явном виде будет отсутствовать из-за слабого разброса значений прогнозируемой переменной.

2. Регрессионная модель

С помощью Stata мы построили регрессионную модель вида $Y = \beta_0 + \beta_1 * X$ (Рис 2.). Первое, что можно отметить - довольно низкое значение ESS при сравнительно большом RSS. Это говорит о том, что данная модель плохо объясняет дисперсию целевой переменной. В связи с этим, следовательно, было получено очень низкое значение коэффициента детерминации R^2 - 0.0082. Такое значение является следствием вышеописанных фактов: качество модели довольно низкое. Отрицательный R_{adj} также является индикатором данной проблемы.

3. Проверка значимости коэффициентов

Для проверки значимости коэффициентов регрессии необходимо было проверить гипотезу об отличии коэффициентов от нуля. В общем случае, проверка строится следующим образом: выдвигается гипотеза H_0 о равенстве $\beta_i, i = 0, 1$ нулю. Гипотеза $H_1 \neq \beta_i$ - неравенство. Соответственно, гипотеза является двусторонней. Если при проверке H_0 не будет отвергнута, то мы будем говорить о незначимости коэффициента β_i . В случае, если таковым оказался β_1 - коэффициент при детерминированной переменной - мы можем сделать вывод, что она практически не влияет на таргет.

4. Интерпретация

Нами были протестированы двусторонние гипотезы о значимости коэффициентов регрессии. Для коэффициента β_1 был получен 95% доверительный интервал - (-0.03; 0.08). Как мы видим, 0 попадает в этот интервал. Следовательно, можно говорить о *незначимости* β_1 .

То же самое было сделано для β_0 . Доверительный интервал принимает следующий вид: (-0.54; -0.19). Таким образом, нельзя сделать вывод о незначимости, и мы считаем коэффициент β_0 *значимым*.

На основе полученных выше результатов можно сделать довольно интуитивный вывод: переменная X имеет низкое влияние на таргет - Y. Она почти не объясняет значения целевой переменной.

5. Проверка закона Оукена

Для проверки закона Оукена требуется построить обратную регрессию, поскольку, согласно теоретической модели, переменной “Y” является изменение ВВП, а “X” - изменение безработицы. Регрессия до этого была построена с обратной зависимостью.

Нами было это сделано с помощью пакета Stata. Результаты можно увидеть на Рис. 5-6. Главное, что нам нужно оттуда взять - коэффициент перед детерминированной переменной. Как мы видим, точечная оценка дает положительное значение. Следовательно, из точечной оценки мы можем сделать вывод, что закон Оукена *не соблюдается*. Это можно также увидеть графически на Рис. 6. Однако, исходя из интервальной оценки коэффициента β_1 для данной регрессии, точно утверждать нельзя.

6. Более качественная модель

Исходя из диаграммы рассеяния можно сделать вывод, что в данных есть несколько выбросов. Они портят регрессионную модель, поэтому для лучшей оценки от них стоит избавиться. Мы применили следующие критерии: значение переменной X должно быть в интервале (-5; 7.5), а переменной Y - (-2;1). Таким образом, количество данных сократилось на 12 и стало 73 наблюдения (Рис.3).

После таких “манипуляций” можно заметить, что модель сравнительно сильно улучшилась: во-первых, ESS увеличился почти в 4 раза, а RSS упала в два раза (Рис. 4). Естественно, что улучшился и R^2 - он увеличился примерно в 6 раз. Однако, несмотря на такие цифры, модель все еще остается плохой: $R^2 = 0.05$. Отклонение от константной модели вида $Y = \beta_0$, $\beta_0 - const$ не является существенным.

При этом можно заметить, что коэффициент β_1 больше не является незначимым: 0 не попадает в новый доверительный интервал. β_0 осталась без изменений и также является значимым коэффициентом.

Приложение

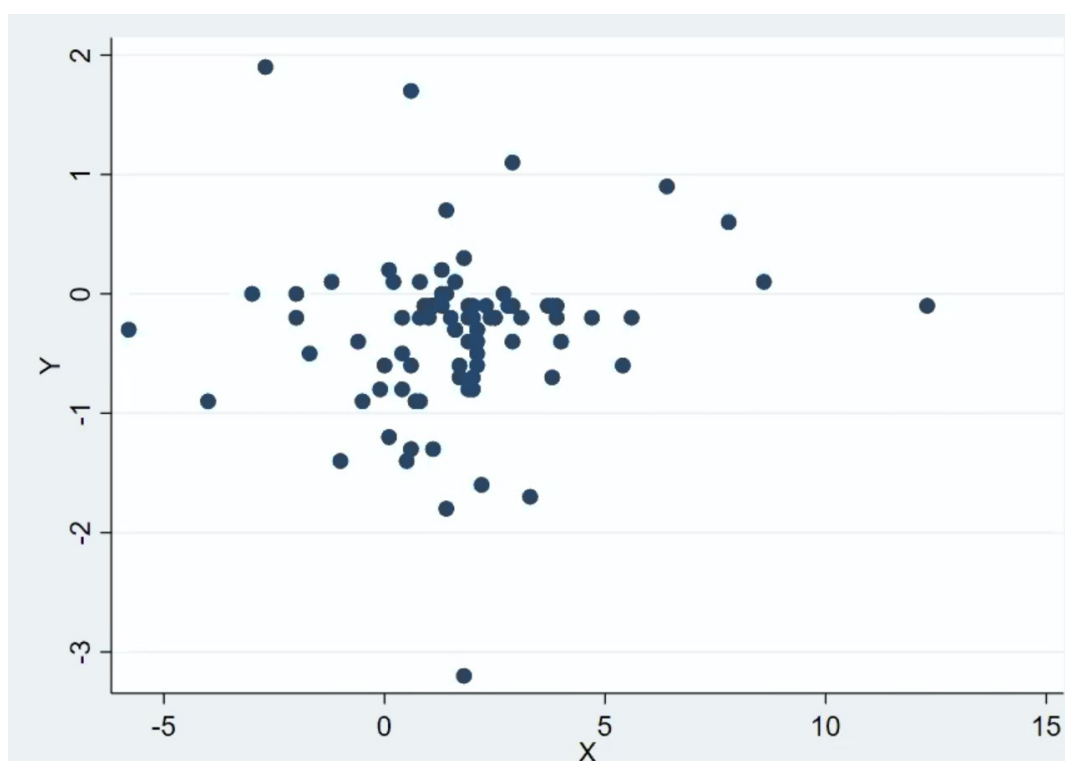


Рис 1. Диаграмма рассеяния.

Source	SS	df	MS	Number of obs	=	85
Model	.326907301	1	.326907301	F(1, 83)	=	0.69
Residual	39.545328	83	.476449735	Prob > F	=	0.4099
				R-squared	=	0.0082
				Adj R-squared	=	-0.0038
Total	39.8722353	84	.474669468	Root MSE	=	.69025

Y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
X	.024948	.0301184	0.83	0.410	-.0349563	.0848524
_cons	-.3700599	.0903039	-4.10	0.000	-.5496709	-.190449

Рис 2. Статистики первой линейной регрессии.

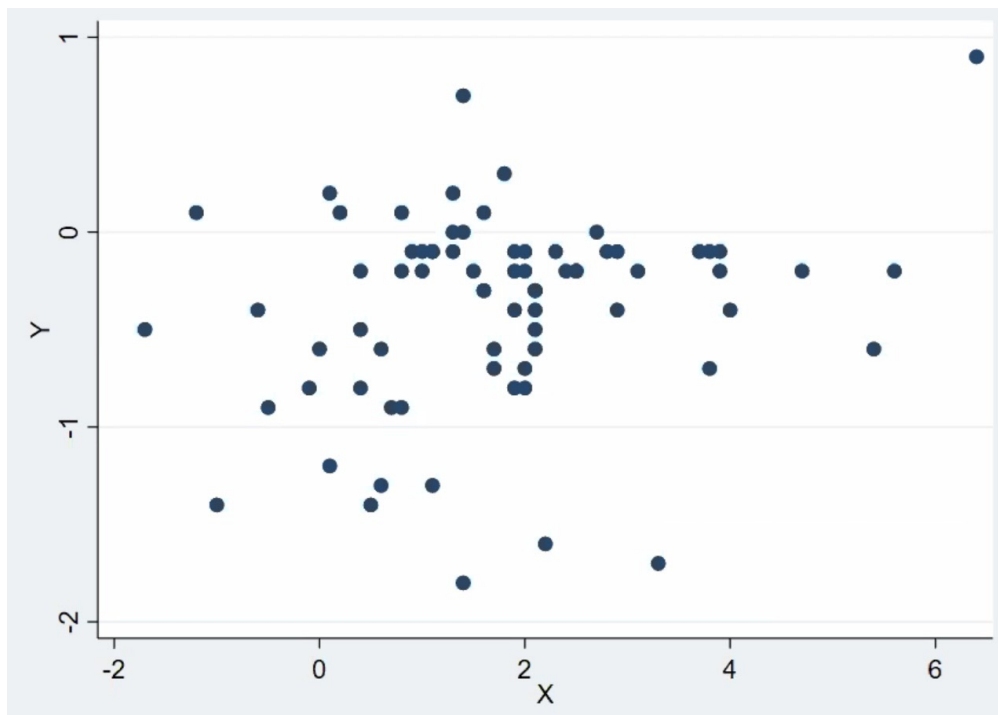


Рис 3. Диаграмма рассеяния без выбросов

Source	SS	df	MS	Number of obs	=	73
Model	1.0247915	1	1.0247915	F(1, 71)	=	4.15
Residual	17.530277	71	.24690531	Prob > F	=	0.0453
				R-squared	=	0.0552
				Adj R-squared	=	0.0419
Total	18.5550685	72	.257709285	Root MSE	=	.4969

Y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
X	.0781809	.038375	2.04	0.045	.0016634	.1546985
_cons	-.5288652	.0889376	-5.95	0.000	-.7062017	-.3515287

Рис 4. Статистики регрессии без выбросов.

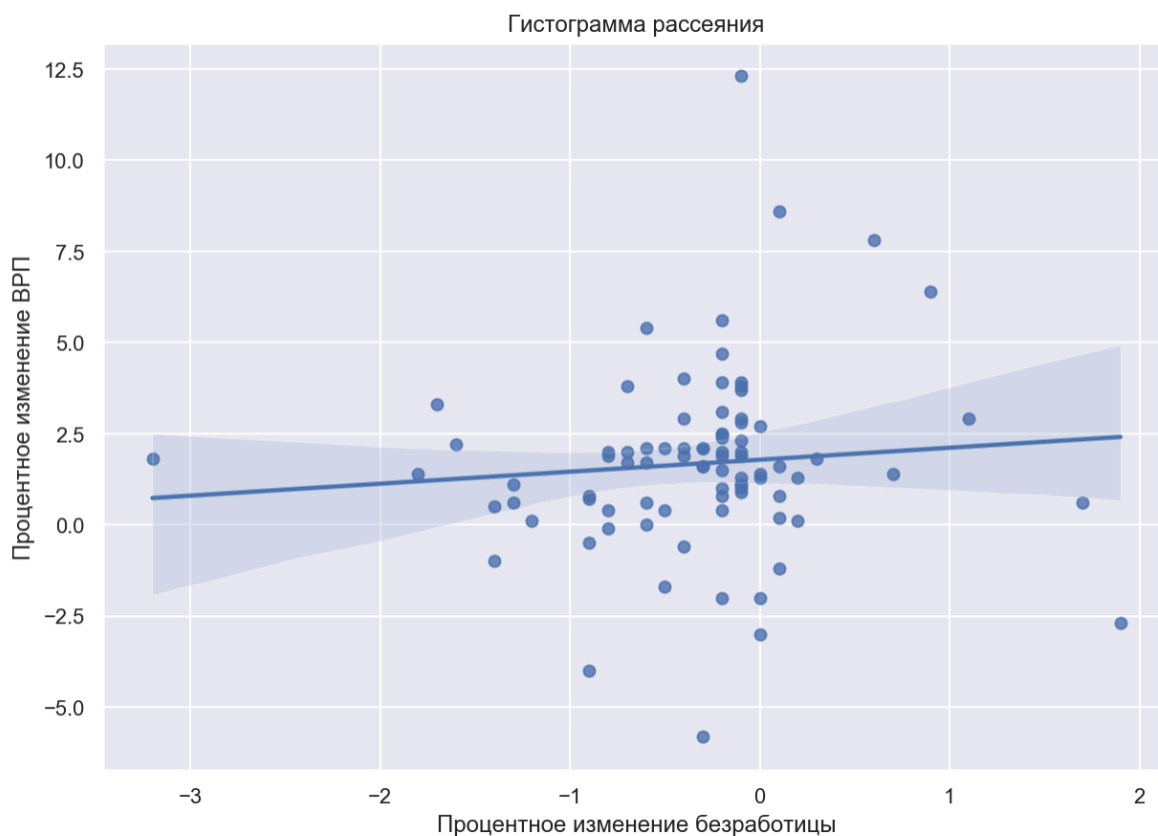


Рис 5. «Правильная» гистограмма рассеяния для проверки закона Оукена с построенными интервалами регрессии.

. reg X Y

Source	SS	df	MS	Number of obs	=	85
Model	4.30631696	1	4.30631696	F(1, 83)	=	0.69
Residual	520.926624	83	6.27622439	Prob > F	=	0.4099
				R-squared	=	0.0082
				Adj R-squared	=	-0.0038
Total	525.232941	84	6.25277311	Root MSE	=	2.5052

X	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Y	.3286379	.3967473	0.83	0.410	-.4604764	1.117752
_cons	1.784341	.3013252	5.92	0.000	1.185017	2.383665

Рис 6. Статистики регрессии для проверки закона Оукена.