

КОРРЕЛЯЦИЯ

Что это такое и с чем ее едят

ВЫБЕРИ Y (ЗАВИСИМАЯ ПЕРЕМЕННАЯ)



ВЫБЕРИ X -Ы (НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ)



ОПРЕДЕЛИ КОРРЕЛЯЦИЮ ФАКТОРОВ (X -ОВ) НА Y



ОПРЕДЕЛИ МУЛЬТИКОЛЛЕНИАРНОСТЬ МЕЖДУ
ФАКТОРАМИ (X -МИ)



ПОЗДРАВЛЯЮ, ТЫ ВЫБРАЛ X -Ы ДЛЯ РЕГРЕССИИ

Пример У и Х-ов для построения ЭММ

У - объем кредитования
нефтегазового бизнеса

- Добыча сырой нефти X1, тыс.тонн;
- Ключевая ставка Банка России X2,%;
- Средневзвешенная процентная ставка по кредитам, свыше 1 года X3,%;
- Цена на нефть марки БРЕНТ X4, долларов США за баррель;
- Цена на нефть марки ЮРАЛС X5, долларов США за баррель;
- Нефтегазовые доходы X6, млрд.руб.;
- Задолженность по кредитам X7, млн.руб.;
- Просроченная задолженность X8, млн.руб.;
- Курс доллара США X9, за единицу;
- Добыча природного газа X10, млн.м³.

Исходные данные по У и по Х-ам

Исходные данные для проведения корреляционно-регрессионного анализа

Дата	Объем кредитования добычи топливно-энергетических полезных ископаемых, млн.руб.	Добыча сырой нефти (с учетом газового конденсата), тыс.тонн	Ключевая ставка Банка России, %	Средневзвешенные процентные ставки по кредитам свыше 1 года, %	Цена на нефть марки Brent, долларов за баррель	Цена на нефть марки URALS, долларов за баррель	Нефтегазовые доходы, млрд.руб.	Задолженность по кредитам, млн.руб.	Просроченная задолженность, млн.руб.	Курс доллар США, за единицу	Добыча природного газа, млн.м ³
01.01.2014	291618	44993	5,5	10,64	107,75	106,43	615,0411	349654	3955	32,6587	66100
01.02.2014	22231	40364	5,5	11,11	108,85	107,42	1233,784	356009	4138	35,18	58000
01.03.2014	42085	44675,4	5,5	10,6	107,66	106,68	1826,6733	361383	3740	36,1847	56900
01.04.2014	133227	43148,3	7	10,97	108,08	106,56	2516,6648	429951	2948	35,6053	52200
01.05.2014	147310	44573	7,5	11,23	110,02	107,7	3148,6977	416376	2807	35,7227	53100
01.06.2014	179223	43198,9	7,5	11,67	112,5	108,93	3703,4386	421062	2837	34,7352	46600
01.07.2014	210159	43968	7,5	11,92	105,6	105,4	4313,8677	409728	2711	33,8434	43000
01.08.2014	248243	44585	8	11,83	102,95	101,09	4887,3417	428467	2694	35,4438	43100
01.09.2014	331337	43529,3	8	12,05	94,77	107,91	5494,7339	476934	2844	37,2945	45900
01.10.2014	383554	44999	8	12,24	86,02	86,38	6113,1296	512980	4657	39,3836	57200
01.11.2014	470263	43532,1	8	12,56	70,15	78,33	6718,763	528665	4566	41,9627	57200
01.12.2014	524542	45088	9,5	12,94	57,35	61,07	7433,8058	538204	8308	51,8068	62600
01.01.2015	588370	45071	17	15,09	49,15	46,58	520,77732	517579	23174	56,2376	60900
01.02.2015	65249	40731,4	17	16,36	60,75	57,3	1046,3351	523530	23229	68,9291	53500
01.03.2015	123318	45210,6	15	16,45	55,18	54,53	1545,5566	523052	22974	61,2718	55300
01.04.2015	155806	43668,3	14	15,8	66,78	59,16	2018,4796	493253	23507	57,65	51300

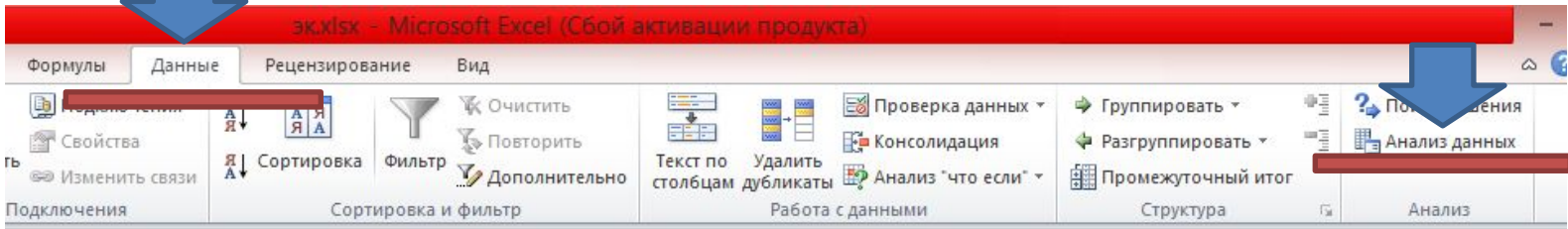
КОРРЕЛЯЦИЯ

ЧТО ЭТО ТАКОЕ?!



Показатель корреляции демонстрирует, как изменение значения одного фактора влияет на изменение значения другого, влияет ли вообще и какова теснота их связи. Сам коэффициент варьируется от -1 до $+1$. В случае, когда наблюдается положительная корреляция, то изменение одного фактора влечет за собой изменение другого. Когда происходит отрицательная корреляция, то рост одного показателя приводит к снижению другого.

КАК ЕЕ СДЕЛАТЬ?!



	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
цена на нефть марки	цена на нефть марки	нефтегазо	Задолжен	Просроченная	курс	Добыча									

Выбираешь «Корреляция»

	B	C	D	E	F	G	H	I	J	K	L
2014	291618										
2014	22231										
2014	42085										
2014	133227										
2014	147310										
2014	179223										
2014	210159										
2014	248243										
2014	331337										
2014	383554										
2014	470263	43532,1	8	12,56	70,15	78,33	6718,763	528665	4566	41,9627	57200
2014	524542	45088	9,5	12,94	57,35	61,07	7433,8058	538204	8308	51,8068	62600
2015	588370	45071	17	15,09	49,15	46,58	520,77732	517579	23174	56,2376	60900

Корреляция

Входные данные:

Группирование: по столбцам по строкам

Метки в первой строке

Параметры вывода: Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Входной интервал – все цифры с названием столбцов (ДАТЫ НЕ БРАТЬ)

Обязательно группирование по столбцам и МЕТКИ

Выход данных можно как на новый лист, так и на этом же листе на пустой ячейке

На выходе получаем матрицу коэффициентов корр-ии

A	B	C	D	E	F	G	H	I	J	K	L
	объем кредитования добычи топливно-энергет	Добыча сырой нефти (с учетом газового конденс	ключевая ставка Банка России, %	средневзвешенные процентные ставки по	цена на нефть марки Brent, долларо в за баррель	цена на нефть марки URALS, долларо в за баррель	нефтегазовые доходы, млрд.руб	Задолженность по кредитам, млн.руб.	Просроченная задолженность, млн.руб.	курс доллар США, за единицу	Добыча природного газа, млн.м3
объем кредитования	1,000										
Добыча сырой нефти	0,661	1,000									
ключевая ставка Банка	0,106	0,083	1,000								
средневзвешенные	-0,119	-0,034	0,914	1,000							
цена на нефть марки	-0,532	-0,448	-0,667	-0,592	1,000						
цена на нефть марки	-0,504	-0,441	-0,676	-0,591	0,990	1,000					
нефтегазовые доходы	0,393	0,201	-0,270	-0,237	0,049	0,119	1,000				
Задолженность по	0,621	0,487	0,251	0,075	-0,699	-0,713	-0,071	1,000			
Просроченная задол	-0,111	0,050	0,845	0,901	-0,588	-0,596	-0,225	-0,004	1,000		
курс доллар США, за	0,400	0,356	0,654	0,560	-0,949	-0,965	-0,198	0,740	0,555	1,000	
Добыча природного	0,396	0,189	-0,034	-0,137	-0,231	-0,233	-0,073	0,170	-0,122	0,206	1,000

Чем ближе коэф.корр к 1, тем лучше, тем сильнее степень влияния X на Y. Однако, нужно брать во внимание и показатели, значение которых идет от 0,5 и выше, т.к. идет средняя степень взаимосвязи.

НО НА ЭТОМ КОРРЕЛЯЦИЯ НЕ ЗАКОНЧИЛАСЬ

Необходимо проверить мультиколлениарность между X-ми (факторами, которые оказали влияние на Y)
В примере они выделены желтым цветом.

Мультиколлениарность есть тоже, что и корреляция, только уже не влияние X на Y, а влияние X на X. Когда между факторами возникает мультиколлениарность, это значит, что значения этих факторов зависят друг от друга, и брать оба фактора в дальнейшее моделирование экономически НЕ ЦЕЛЕСООБРАЗНО. Значение коэффициента более и равное 0,7 между X-ми говорит о наличии мультиколлениарности.

В модель пойдет тот X, у которого с Y будет большая корреляция.

A	B	C	D	E	F	G	H	I	J	K	L
	объем кредитования добычи топливно-энергет	Добыча сырой нефти (с учетом газового конденс	ключевая ставка Банка России, %	средневзвешенные процентные ставки по	цена на нефть марки Brent, долларо в за баррель	цена на нефть марки URALS, долларо в за баррель	нефтегазовые доходы, млрд.руб	Задолженность по кредитам, млн.руб.	Просроченная задолженность, млн.руб.	курс доллар США, за единицу	Добыча природного газа, млн.м3
объем кредитован	1,000										
Добыча сырой неф	0,661										
ключевая ставка Ба	0,106	0,083	1,000								
средневзвешенные	-0,119	-0,034	0,914	1,000							
цена на нефть марн	-0,532	-0,448	-0,667	-0,592	1,000						
цена на нефть марн	-0,504	-0,441	-0,676	-0,591	0,990	1,000					
нефтегазовые дохо	0,393	0,201	-0,270	-0,237	0,049	0,119	1,000				
Задолженность по	0,621	0,487	0,251	0,075	-0,699	-0,713	-0,071	1,000			
Просроченная зад	0,111	0,050	0,845	0,901	-0,588	-0,596	-0,225	-0,004	1,000		
курс доллар США, з	0,000	0,356	0,654	0,560	-0,949	-0,965	-0,198	0,740	0,555	1,000	
Добыча природног	0,096	0,189	-0,034	-0,137	-0,231	-0,233	-0,073	0,170	-0,122	0,206	1,000

ВЫБИРАЕТСЯ ОБЪЕМ НЕФТЕДОБЫЧИ (X1) и ЗАДОЛЖЕННОСТЬ (X7)

Проверим значимость данных коэффициентов согласно тесту Стьюдента по формуле, где r – коэффициент корреляции, n – число наблюдений:

$$t = r * \sqrt{n - 2} / \sqrt{1 - r^2}$$

r – коэф.корреляции (выделено желтым)
(находим фактическое значение)

Табличное значение находится путем ввода в ячейку
=СТЮДРАСПОБР

Степень свободы – число периодов (36 месяцев) – 2.

Уровень значимости 0,05 или 0,01 (это вероятность 95%
ми 99% соответственно)

Расчетные значения ДОЛЖНЫ БЫТЬ больше табличного