



**ФИНАНСОВЫЙ УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ**

Департамент экономической теории

Центр макроэкономических исследований

Балацкий Евгений, Екимова Наталья, Юревич Максим

**Математическое моделирование в
экономических публикациях**

Москва

28 февраля 2018 года

ОБЩАЯ МОДЕЛЬ

$$y = a_0 + a_1 * x_1 + \dots + a_n * x_n$$

y – выходная переменная; x – входные переменные;

a – отыскиваемые параметры

ОПРЕДЕЛИТЬ КОЛИЧЕСТВО СВИНЕЙ, УНИЧТОЖЕННЫХ ВСЛЕДСТВИЕ ИХ ЗАБОЛЕВАНИЯ АЧС В РЕГИОНАХ РОССИИ С 2012-2016 ГОДЫ

№	Регион	Год наблюдения	Число уничтоженных свиней (Y), ед.	Число заболевших свиней (X), ед.
1	Краснодарский край	2016	85170	3153
2	Липецкая область	2016	16442	1405
3	Кабардино-Балкарская Республика	2016	38053	690
4	Рязанская область	2016	27098	690
5	Саратовская область	2016	2008	500
6	Республика Крым	2016	1193	314
7	Воронежская область	2016	44317	260
8	Волгоградская область	2016	7128	241
9	Пензенская область	2016	895	79
10	Московская область	2016	12430	68
11	Смоленская область	2016	65	64
12	Владимирская область	2016	240	62
13	Тамбовская область	2016	135	60
14	Курская область	2016	18336	30
15	Орловская область	2015	90	60
....				
28	Тверская область	2013	27692	54
29	Смоленская область	2013	83	49
30	Краснодарский край	2012	169447	3091
31	Ростовская область	2012	925	230
32	Волгоградская область	2012	2467	121
33	Тверская область	2012	33527	60

МЕНЮ: ФАЙЛ | ГЛАВНАЯ | ВСТАВКА | РАЗМЕТКА СТРАНИЦЫ | ФОРМУЛЫ | **ДАнные** | РЕЦЕНЗИРОВАНИЕ | ВИД

Получение внешних данных: Из Access, Из Интернета, Из текста, Из других источников, Существующие подключения, Обновить все, Подключения: Подключения, Свойства, Изменить связи

Сортировка и фильтр: Сортировка, Фильтр, Очистить, Повторить, Дополнительно

Работа с данными: Мгновенное заполнение, Удалить дубликаты, Проверка данных, Консолидация, Анализ "что если", Отношения

Структура: Группировать, Разгруппировать, Промежуточный итог

Анализ: Анализ данных, Поиск решения

F8

№	Регион	Год наблюдения	Число уничтоженных свиней (Y), ед.	Число заболевших свиней (X), ед.
1	Краснодарский край	2016	85170	3153
2	Липецкая область	2016	16442	1405
3	Кабардино-Балкарская Республ	2016	38053	690
4	Рязанская область	2016	27098	690
5	Саратовская область	2016	2008	500
6	Республика Крым	2016	1193	314
7	Воронежская область	2016	44317	260
8	Волгоградская область	2016	7128	241
9	Пензенская область	2016	895	79
10	Московская область	2016	12430	68
11	Смоленская область	2016	65	64
12	Владимирская область	2016	240	62
13	Тамбовская область	2016	135	60
14	Курская область	2016	18336	30
15	Орловская область	2015	90	60
.....				
18	Тверская область	2013	27692	54
19	Смоленская область	2013	83	49
20	Краснодарский край	2012	169447	3091
21	Ростовская область	2012	925	230
22	Волгоградская область	2012	2467	121
23	Тверская область	2012	33527	60

Анализ данных

Инструменты анализа

- Ковариация
- Описательная статистика
- Экспоненциальное сглаживание
- Двухвыборочный F-тест для дисперсии
- Анализ Фурье
- Гистограмма
- Скользящее среднее
- Генерация случайных чисел
- Ранг и перцентиль
- Регрессия**

OK, Отмена, Справка

Экранная панель Excel с вкладками: ФАЙЛ, ГЛАВНАЯ, ВСТАВКА, РАЗМЕТКА СТРАНИЦЫ, ФОРМУЛЫ, ДАННЫЕ, РЕЦЕНЗИРОВАНИЕ, ВИД. Вкладка ДАННЫЕ активна.

Группы инструментов:

- Получение внешних данных: Из Access, Из Интернета, Из текста, Из других источников, Существующие подключения, Обновить все, Изменить связи.
- Подключения: Подключения, Свойства.
- Сортировка и фильтр: Сортировка, Фильтр, Очистить, Повторить, Дополнительно.
- Работа с данными: Мгновенное заполнение, Консолидация, Анализ "что если", Удалить дубликаты, Проверка данных, Отношения.
- Структура: Группировать, Разгруппировать, Промежуточный итог.
- Анализ: Анализ данных, Поиск решения.

Q18 : X ✓ fx

№	Регион	Год наблюдения	Число уничтоженных свиней (Y), ед.	Число заболевших свиней (X), ед.
1	Краснодарский край	2016	85170	3153
2	Липецкая область	2016	16442	1405
3	Кабардино-Балкарская Республ	2016	38053	690
4	Рязанская область	2016	27098	690
5	Саратовская область	2016	2008	500
6	Республика Крым	2016	1193	314
7	Воронежская область	2016	44317	260
8	Волгоградская область	2016	7128	241
9	Пензенская область	2016	895	79
10	Московская область	2016	12430	68
11	Смоленская область	2016	65	64
12	Владимирская область	2016	240	62
13	Тамбовская область	2016	135	60
14	Курская область	2016	18336	30
15	Орловская область	2015	90	60
....				
18	Тверская область	2013	27692	54
19	Смоленская область	2013	83	49
20	Краснодарский край	2012	169447	3091
21	Ростовская область	2012	925	230
22	Волгоградская область	2012	2467	121
23	Тверская область	2012	33527	60

Регрессия

Входные данные

Входной интервал Y: \$Q\$18

Входной интервал X: \$D\$1:\$D\$34

Метки Константа - ноль

Уровень надежности: 95 %

Параметры вывода

Выходной интервал: []

Новый рабочий лист: []

Новая рабочая книга

Остатки

Остатки График остатков

Стандартизованные остатки График подбора

Нормальная вероятность

График нормальной вероятности

Кнопки: OK, Отмена, Справка

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Что вы хотите сделать?

Вырезать Копировать Вставить Формат по образцу Буфер обмена

Calibri 11 A A Шрифт

Перенести текст Объединить и поместить в центре Выравнивание

Общий Число

Условное форматирование Форматировать как таблицу

Обычный Ввод Нейтральный Вывод Плохой Вычисление Хороший Контрольна...

Ячейки Вставить Удалить Формат

Автосумма Заполнить Очистить Редактирование

Сортировка и фильтр Найти и выделить

Вывод ИТОВОГ																	
<i>Регрессионная статистика</i>																	
Множественный R			0,768704819														
R-квадрат			0,590907099														
Нормированный R-квадрат			0,577710554														
Стандартная ошибка			23769,93589														
Наблюдения			33														
Дисперсионный анализ																	
			<i>df</i>		<i>SS</i>		<i>MS</i>		<i>F</i>		<i>Значимость F</i>						
Регрессия			1		25299677170		25299677170		44,77740887		1,74396E-07						
Остаток			31		17515305420		565009852,3										
Итого			32		42814982590												
			<i>Коэффициенты</i>		<i>Стандартная ошибка</i>		<i>t-статистика</i>		<i>P-Значение</i>		<i>Нижние 95%</i>		<i>Верхние 95%</i>		<i>Нижние 95,0%</i>		<i>Верхние 95,0%</i>
Y-пересечение			7079,643141		4813,485456		1,470793504		0,15142547		-2737,525172		16896,81145		-2737,525172		16896,81145
Число заболевших свиней (*), ед.			36,91827168		5,517113037		6,691592402		1,74396E-07		25,66604545		48,1704979		25,66604545		48,1704979
Вывод ОСТАТКА																	
			<i>Наблюдение</i>		<i>Предсказанное Число уничтоженных свиней (K), ед.</i>		<i>Остатки</i>										
			1		123482,9537		-38312,95374										
			2		58949,81485		-42507,81485										
			3		32553,2506		5499,749403										
			4		32553,2506		-5455,250597										
			5		25538,77898		-23530,77898										
			6		18671,98045		-17478,98045										
			7		16678,39378		27638,60622										
			8		15976,94661		-8848,946615										
			9		9996,186603		-9101,186603										
			10		9590,085615		2839,914385										
			11		9442,412528		-9377,412528										
			12		9368,575985		-9128,575985										
			13		9294,739441		-9159,739441										
			14		8187,191291		10148,80871										
			15		9294,739441		-9204,739441										
			16		8740,965366		-8126,965366										
			17		8261,027834		-6683,027834										
			18		43739,48692		14152,51308										
			19		33993,06319		32282,93681										
			20		31593,37553		-4325,375534										

$$Y = 7079,64 + 36,91X$$

(4813,49) (5,52)

$$N = 33; R^2 = 0,591; F = 44,78$$

ОЦЕНКА ЗНАЧИМОСТИ ПАРАМЕТРОВ

Суть проверки общего качества уравнения регрессии – оценить насколько хорошо эмпирическое уравнение регрессии согласуется со статистическими данными

Основные показатели качества:

1. Коэффициент детерминации R^2 (Коэффициент корреляции r_{xy})
 2. Значение F -статистики (критерий Фишера)
 3. Стандартная ошибка регрессии (T -статистика)
 4. Средняя ошибка аппроксимации
5. Критерий Дарбина-Уотсона (в случае временных рядов)

КОЭФФИЦИЕНТ ДЕТЕРМИНАЦИИ R^2

Коэффициент R^2 показывает долю объясненной вариации зависимой переменной

Используется для предварительной оценки качества модели и как основа для расчета других показателей

$$0 \leq R^2 \leq 1$$

Чем ближе R^2 к 1, тем лучше регрессия аппроксимирует статистические данные, тем теснее линейная связь между зависимой и объясняющими переменными

Критические значения коэффициентов корреляции (r) для уровней значимости 0,05 и 0,01

df	$\alpha = 0,05$	$\alpha = 0,01$	df	$\alpha = 0,05$	$\alpha = 0,01$
1	0,9969	0,9998	17	0,4555	0,5751
2	0,9950	0,9900	18	0,4438	0,5614
3	0,8783	0,9587	19	0,4329	0,5487
4	0,8114	0,9172	20	0,4227	0,5368
5	0,7545	0,8745	25	0,3809	0,4869
6	0,7067	0,8343	30	0,3494	0,4487
7	0,6664	0,7977	35	0,3246	0,4182
8	0,6319	0,7640	40	0,3044	0,3932
9	0,6021	0,7348	45	0,2875	0,3721
10	0,5760	0,7079	50	0,2732	0,3541
11	0,5529	0,6835	60	0,2500	0,3248
12	0,5324	0,6614	70	0,2319	0,3017
13	0,5139	0,6411	80	0,2172	0,2830
14	0,4973	0,6226	90	0,2050	0,2673
15	0,4821	0,6055	100	0,1946	0,2540
16	0,4683	0,5897			

$$r = \sqrt{R^2}$$

$$df = n - m - 1$$

n – количество наблюдений,
 m – число объясняющих переменных (число параметров модели регрессии без учета свободного члена)

$$r = 0,7687$$

$$df = 33 - 1 - 1 = 31$$

$$r_{\text{табл}} = 0,3494$$

F-СТАТИСТИКА

F-тест состоит в проверке гипотезы H_0 о статистической незначимости уравнения регрессии и показателя тесноты связи

Для этого выполняется сравнение фактического $F_{\text{факт}}$ и критического (табличного) $F_{\text{табл}}$ значений **F-критерия Фишера**

$F_{\text{табл}} < F_{\text{факт}}$ – H_0 -гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность

$F_{\text{табл}} > F_{\text{факт}}$ – H_0 -гипотеза не отклоняется и признается статистическая незначимость и ненадежность уравнения регрессии

Таблица значений F-критерия Фишера на уровне значимости 0,05

Уровень значимости $\alpha = 0,05$												
k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,50	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

k_1 – число независимых факторов

$$k_2 = n - m - 1$$

$$F = 44,78$$

$$k_1 = 1$$

$$k_2 = 31$$

$$F_{\text{табл}} = 4,17$$

СТАНДАРТНЫЕ ОШИБКИ РЕГРЕССИИ

Стандартные ошибки коэффициентов регрессии – это средние квадратические отклонения коэффициентов регрессии от их истинных значений.

Сравнивая значение коэффициента с его стандартной ошибкой, можно судить о значимости коэффициента: он является значимым, если есть достаточно высокая вероятность того, что его истинное значение отлично от нуля

Для стандартных ошибок оценок нет таблиц критических уровней – для точного суждения используются t -статистики:

$$t = \text{коэффициент} / \text{стандартная ошибка}$$

T-СТАТИСТИКА

T-статистика (t-критерий Стьюдента) используется для проверки значимости каждого фактора регрессионной модели

Значение t-статистики сравнивается с критическим, имеющим распределение Стьюдента с **(n-m-1)** степенями свободы

ПОРЯДОК РАБОТЫ ПРИ ПРОВЕРКЕ ЗНАЧИМОСТИ КОЭФФИЦИЕНТА ПО t-СТАТИСТИКЕ:

1. Выбираем уровень значимости α (1% или 5%)
2. Вычисляем число степеней свободы
3. По таблицам распределения Стьюдента определяем критическое значение t
4. Если модуль t -статистики больше критического значения, то коэффициент является значимым на уровне значимости α
5. В противном случае коэффициент не значим (на данном уровне α)

Критические значения t-критерия на уровнях значимости 0,10; 0,05; 0,01

Число средней свободы <i>df</i>	α			Число средней свободы <i>df</i>	α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

T-СТАТИСТИКА

При оценке значимости коэффициентов регрессии можно использовать следующие «грубые» правила (почти всегда работает при $n > 10$), позволяющее не прибегать к таблицам:

Если $|t| \leq 1$, то коэффициент *не может быть признан значимым*

Если $1 < |t| \leq 2$, то найденная оценка может рассматриваться как относительно *(слабо) значимая*

Если $2 < |t| \leq 3$, то *доверительная вероятность составляет 0,95-0,99*

Если $|t| > 3$, то это почти гарантия значимости коэффициента

Если коэффициент $>$ стандартной ошибки в 2 и более раза, то это почти гарантия значимости коэффициента

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Вырезать Копировать Вставить Формат по образцу Буфер обмена

Calibri 11 Шрифт

Перенести текст Объединить и поместить в центре Выравнивание

Общий Число

Условное форматирование Форматировать как таблицу

Обычный Нейтральный Плохой Хороший Ввод Вывод Вычисление Контрольная...

Вставить Удалить Формат Ячейки

Автосумма Заполнить Очистить Сортировка и фильтр Найти и выделить Редактирование

Вывод итогов								
Регрессионная статистика								
Множественный R			0,768704819					
R-квадрат			0,590907099					
Нормированный R-квадрат			0,577710554					
Стандартная ошибка			23769,93589					
Наблюдения			33					
Дисперсионный анализ								
	df	SS	MS	F	Значимость F			
Регрессия	1	25299677170	25299677170	44,77740887	1,74396E-07			
Остаток	31	17515305420	565009852,3					
Итого	32	42814982590						
Коэффициенты		Стандартная ошибка	t-статистика	P-значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	7079,643141	4813,485456	1,470793504	0,15142547	-2737,525172	16896,81145	-2737,525172	16896,81145
Число заболевших свиней (I*), ед.	36,91827168	5,517113037	6,691592402	1,74396E-07	25,66604545	48,1704979	25,66604545	48,1704979
Вывод остатка								
Наблюдение	Предсказанное	Число унцтоженных свиней (K), ед.	Остатки					
1	123482,9537	123482,9537	-38312,95374					
2	58949,81485	58949,81485	-42507,81485					
3	32553,2506	32553,2506	5499,749403					
4	32553,2506	32553,2506	-5455,250597					
5	25538,77898	25538,77898	-23530,77898					
6	18671,98045	18671,98045	-17478,98045					
7	16678,39378	16678,39378	27638,60622					
8	15976,94661	15976,94661	-8848,946615					
9	9996,186603	9996,186603	-9101,186603					
10	9590,085615	9590,085615	2839,914385					
11	9442,412528	9442,412528	-9377,412528					
12	9368,575985	9368,575985	-9128,575985					
13	9294,739441	9294,739441	-9159,739441					
14	8187,191291	8187,191291	10148,80871					
15	9294,739441	9294,739441	-9204,739441					
16	8740,965366	8740,965366	-8126,965366					
17	8261,027834	8261,027834	-6683,027834					
18	43739,48692	43739,48692	14152,51308					
19	33993,06319	33993,06319	32282,93681					
20	31593,37553	31593,37553	-4325,375534					

$t_{табл} = 2,04$

«ОШИБКИ» РЕГРЕССИИ

1. Невключение в модель всех объясняющих переменных
2. Неправильный выбор функциональной зависимости модели
3. Агрегирование данных
4. Ограниченность статистических данных
5. Прочее (человеческий фактор, ошибки измерений и т.п.)

ОШИБКИ ПЕРВОГО И ВТОРОГО РОДА; УРОВЕНЬ ЗНАЧИМОСТИ

		Верная гипотеза	
		H_0	H_1
Результат применения критерия	H_0	верно принята	неверно принята (<i>Ошибка второго рода</i>)
	H_1	неверно отвергнута (<i>Ошибка первого рода</i>)	верно отвергнута

Поиск оптимальной регрессионной модели: первичная оценка

Вывод итогов								
<i>Регрессионная статистика</i>								
Множественный R	0,810194369							
R-квадрат	0,656414916							
Нормированный R-квадрат	0,647490628							
Стандартная ошибка	15453,48811							
Наблюдения	80							
<i>Дисперсионный анализ</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
Регрессия	2	35130789394	17565394697	73,55375829	1,4E-18			
Остаток	77	18388392696	238810294,8					
Итого	79	53519182090						
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	2245,058788	2013,527545	1,11498787	0,268323551	-1764,386786	6254,504363	-1764,386786	6254,504363
Заболело (гол.)	39,69734355	3,273110382	12,12832411	1,52767E-19	33,17974811	46,21493898	33,17974811	46,21493898
Плотность свиней, голов на км2	1,451982104	86,53266101	0,016779585	0,986655859	-170,8565595	173,7605237	-170,8565595	173,7605237

Поиск оптимальной регрессионной модели: исключение незначимых переменных

Вывод итогов								
<i>Регрессионная статистика</i>								
Множественный R	0,810193594							
R-квадрат	0,656413659							
Нормированный R-квадрат	0,652008706							
Стандартная ошибка	15354,13581							
Наблюдения	80							
<i>Дисперсионный анализ</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
Регрессия	1	35130722156	35130722156	149,0171737	8,88207E-20			
Остаток	78	18388459935	235749486,3					
Итого	79	53519182090						
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	2258,992378	1822,53146	1,239480595	0,218883984	-1369,389037	5887,373793	-1369,389037	5887,373793
Заболело (гол.)	39,69678684	3,251900091	12,20725906	8,88207E-20	33,22275107	46,1708226	33,22275107	46,1708226

Поиск оптимальной регрессионной модели: исключение из выборки экстремальных значений

Регрессия

Входные данные

Входной интервал Y: SAS3:SAS82

Входной интервал X: SB\$3:SB\$82

Метки Константа - ноль

Уровень надежности: 95 %

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Остатки

Остатки График остатков

Стандартизованные остатки График подбора

Нормальная вероятность

График нормальной вероятности

OK

Отмена

Справка

Вывод итогов		
Регрессионная статистика		
Множественный	0,810193594	
R-квадрат	0,656413659	
Нормированный	0,652008706	
Стандартная ош	15354,13581	
Наблюдения	80	

Вывод остатка		
Наблюдение	Предсказанное Y	Остатки
1	29649,7753	8403,224705
2	2616,26346	-2611,26346
3	5395,038538	-4500,038538
4	2735,35382	-2731,35382
5	2457,476312	-2421,476312
6	2616,26346	-2606,26346
7	58032,97788	-41590,97788
8	22107,3858	-20099,3858
9	29649,7753	-2551,775295
10	4958,373883	7471,626117
11	2576,566673	-2566,566673

Поиск оптимальной регрессионной модели: исключение из выборки экстремальных значений

ВЫВОД ИТОГОВ								
<i>Регрессионная статистика</i>								
Множественный R	0,768704819							
R-квадрат	0,590907099							
Нормированный R-квадрат	0,577710554							
Стандартная ошибка	23769,93589							
Наблюдения	33							
<i>Дисперсионный анализ</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
Регрессия	1	25299677170	25299677170	44,77740887	1,74396E-07			
Остаток	31	17515305420	565009852,3					
Итого	32	42814982590						
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	7079,643141	4813,485456	1,470793504	0,15142547	-2737,525172	16896,81145	-2737,525172	16896,81145
Заболело (гол.)	36,91827168	5,517113037	6,691592402	1,74396E-07	25,66604545	48,1704979	25,66604545	48,1704979

Поиск оптимальной регрессионной модели: введение дополнительных переменных

В выборку вводится **фиктивная переменная**, отражающая низкий, средний и высокий уровни опасности эпидемии, принимающая значения 0,1 и 2 соответственно.

Вывод итогов								
<i>Регрессионная статистика</i>								
Множественный R	0,887657406							
R-квадрат	0,78793567							
Нормированный R-квадрат	0,773798048							
Стандартная ошибка	17396,86812							
Наблюдения	33							
<i>Дисперсионный анализ</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
Регрессия	2	33735451982	16867725991	55,73325335	7,88886E-11			
Остаток	30	9079530608	302651020,3					
Итого	32	42814982590						
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	-10779,64923	4884,067351	-2,207104951	0,035098859	-20754,24546	-805,0530062	-20754,24546	-805,0530062
Число заболевших свиней, ед.	31,81118261	4,152149943	7,661376166	1,51677E-08	23,33136115	40,29100408	23,33136115	40,29100408
Уровень неблагополучия региона	20135,81601	3813,974775	5,279483268	1,058E-05	12346,64038	27924,99164	12346,64038	27924,99164

Поиск оптимальной регрессионной модели: изменение функциональной зависимости

Гипотеза: зависимость между переменными имеет логарифмический характер.

Вывод итогов								
<i>Регрессионная статистика</i>								
Множественный R	0,887657406							
R-квадрат	0,78793567							
Нормированный R-квадрат	0,773798048							
Стандартная ошибка	17396,86812							
Наблюдения	33							
<i>Дисперсионный анализ</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
Регрессия	2	33735451982	16867725991	55,73325335	7,88886E-11			
Остаток	30	9079530608	302651020,3					
Итого	32	42814982590						
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	-10779,64923	4884,067351	-2,207104951	0,035098859	-20754,24546	-805,0530062	-20754,24546	-805,0530062
Число заболевших свиней, ед.	31,81118261	4,152149943	7,661376166	1,51677E-08	23,33136115	40,29100408	23,33136115	40,29100408
Уровень неблагополучия региона	20135,81601	3813,974775	5,279483268	1,058E-05	12346,64038	27924,99164	12346,64038	27924,99164

Поиск оптимальной регрессионной модели: оценка качества модели

Средняя ошибка аппроксимации - среднее отклонение расчетных значений от фактических:

$$\bar{A} = \frac{\sum |y_i - y_x| : y_i}{n} 100\%$$

ВЫВОД ОСТАТКА				
Наблюдение	Предсказанное Y	Остатки	Фактическое значение Y	Yф-Yп /Yф
1	11,22	0,13	11,35	0,01
2	13,49	-1,45	12,04	0,12
3	9,30	0,41	9,71	0,04
4	11,18	-0,21	10,97	0,02
5	10,89	0,21	11,10	0,02
6	10,85	-0,30	10,55	0,03
7	8,52	1,69	10,21	0,17
8	8,49	1,73	10,21	0,17
9	5,98	1,63	7,60	0,21
10	5,77	1,31	7,08	0,19
.....				
33	10,12	-0,31	9,82	0,03
				9,9%

Заключение

Области применения регрессионного анализа:

1. Межпространственные и межвременные выборки
2. Модели на основе панельных данных
3. Предсказание наступления определенных событий (logit- и probit-модели)

Специальные программные приложения:

1. Eviews, Staitistica
2. RStudio

БЛАГОДАРИМ
ЗА
ВНИМАНИЕ!