

*Малахова О.С.*

*студент магистратуры, Самарский национальный исследовательский университет имени академика С.П. Королева (Самарский университет)*

*Самара, Россия*

**КОРРЕЛЯЦИОННО-РЕГРЕССИОННОЙ  
АНАЛИЗ ЗАВИСИМОСТИ ВЫРУЧКИ ОТ ФАКТОРОВ  
ВНЕШНЕЭКОНОМИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ  
(НА ПРИМЕРЕ ООО «ЦМТ «АНАЛИТИКА»)**

***Аннотация:** Определяющим фактором работы любого коммерческого предприятия являются его денежные результаты. Выручка предприятий, развивающихся за пределами страны, тесно связана со значениями факторов внешнеэкономической деятельности, которые подвержены постоянным изменениям. Применение методов стохастического моделирования позволяет определить не только степень влияния каждого из факторов, но и построить прогнозные модели.*

***Ключевые слова:** статистическое моделирование, эконометрическая модель, корреляционно-регрессионный анализ, мультиколлинеарность, медицинская техника и оборудование, построение прогнозной модели, фактор, корреляция, регрессия, выручка.*

***Annotation:** The determining factor in the operation of any commercial enterprise is its monetary results. The revenue of enterprises developing outside the country is closely related to the values of the factors of foreign economic activity, which are subject to constant changes. The use of methods of stochastic modeling allows us to determine not only the degree of influence of each of the factors, but also to build predictive models.*

**Keywords:** *statistical modeling, econometric model, correlation and regression analysis, multicollinearity, medical technology and equipment, building a predictive model, factor, correlation, regression, revenue.*

Объектом исследования являются результаты экономической деятельности ООО «ЦМТ «Аналитика» за 95 периодов (с 2011 по 2018 годы). Компания ведет деятельность по импорту, продаже и обслуживанию медицинского оборудования в России и за рубежом.

Деятельность предприятия ООО «ЦМТ «Аналитика» подразумевает наличие широких внешнеэкономических связей, а также зависимость выручки от множества факторов, объектов и субъектов внутренней и внешней экономической деятельности. Исходя из этого, на первом этапе для дальнейшего анализа результативного фактора – выручки, был отобран ряд показателей, отражающих как внешние, так и внутренние изменения рынка (таблица 1).

Таблица 1. Введенные обозначения для показателей деятельности компании

№	Наименование фактора, ед.	Введенное обозначение
1	списочная численность сотрудников, (чел.)	$x_1$
2	курс \$, (руб.)	$x_2$
3	курс €, (руб.)	$x_3$
4	число новых открытых представительств и филиалов за период, (ед.)	$x_4$
5	наименований услуг в номенклатуре компании на конец периода, (ед.)	$x_5$
6	число поставщиков на конец периода, (ед.)	$x_6$
7	выручка от продажи медицинского оборудования за период, (тыс. руб)	$x_7$
8	выручка от деятельности по содержанию и ремонту оборудования за период, (тыс. руб)	$x_8$
9	выручка от прочих видов деятельности за период, (тыс. руб)	$x_9$
10	таможенные пошлины на ввоз оборудования за период,	$x_{10}$

	(тыс. руб)	
11	импорт за период, (т)	X <sub>11</sub>
12	число мер протекционизма, применяемых к РФ за период, (ед.)	X <sub>12</sub>
13	число мер либерализации ВЭД применяемых к РФ за период, (ед.)	X <sub>13</sub>

В качестве результирующего значения модели выбран показатель Y – выручка предприятия (тыс.руб).

На первом этапе, с целью выявления факторов, оказывающих наиболее высокое влияние на выручку – Y, произведен корреляционный анализ исходных данных. С помощью встроенного анализа в табличном процессоре Microsoft Excel построена матрица корреляционной зависимости факторов и целевой функции (таблица 2).

Таблица 2. Корреляционная матрица факторов для ООО «ЦМТ «Аналитика»

	Y	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>	x <sub>8</sub>	x <sub>9</sub>	x <sub>10</sub>	x <sub>11</sub>	x <sub>12</sub>	x <sub>13</sub>
Y	1,00													
x <sub>1</sub>	<b>0,70</b>	1,00												
x <sub>2</sub>	<b>0,82</b>	0,63	1,00											
x <sub>3</sub>	<b>0,82</b>	0,63	0,99	1,00										
x <sub>4</sub>	0,01	0,05	- 0,02	- 0,01	1,00									
x <sub>5</sub>	<b>0,90</b>	0,75	0,87	0,89	- 0,02	1,00								
x <sub>6</sub>	<b>0,92</b>	0,77	0,86	0,89	- 0,04	0,97	1,00							
x <sub>7</sub>	<b>0,90</b>	0,76	0,66	0,66	- 0,01	0,83	0,82	1,00						
x <sub>8</sub>	<b>0,86</b>	0,55	0,79	0,76	0,00	0,75	0,72	0,69	1,00					
x <sub>9</sub>	- 0,07	0,03	- 0,13	- 0,11	- 0,07	- 0,02	- 0,03	- 0,01	- 0,03	1,00				
x <sub>10</sub>	0,32	0,43	0,29	0,27	0,09	0,26	0,21	0,38	0,41	0,32	1,00			
x <sub>11</sub>	0,17	0,33	0,12	0,11	0,03	0,15	0,12	0,25	0,25	0,59	0,88	1,00		
x <sub>12</sub>	- 0,15	0,02	- 0,05	- 0,07	- 0,06	- 0,15	- 0,18	- 0,15	- 0,08	- 0,08	0,18	0,10	1,00	
x <sub>13</sub>	0,21	0,16	0,15	0,14	0,01	0,20	0,18	0,16	0,19	- 0,07	0,04	- 0,01	0,08	1,00

Наблюдается высокая степень зависимости сразу между несколькими факторами. Ячейки, содержащие информацию о мультиколлинеарности факторов, выделены цветом. Означает это, что объясняющие переменные модели тесно связаны между собой.

Первый столбец матрицы отображает зависимость результирующего параметра  $Y$  от  $x_i$ . Наиболее весомые полученные значения обозначены в таблице жирным и подчеркнутым шрифтом. Применяя метод главных компонент для борьбы с мультиколлинеарностью факторов, определим, что в наибольшей мере выручка -  $Y$  связана с переменными факторами:  $x_1$ ,  $x_2$ ,  $x_3$ ,  $x_5$ ,  $x_6$ ,  $x_7$ ,  $x_8$ . Как видно из таблицы 2, модели регрессионных зависимостей – парные.

На следующем этапе был произведён регрессионный анализ данных. Пример вычислений для  $x_1$  отображен на рисунках 1 и 2.

Вывод итогов								
<b>Регрессионная статистика</b>								
Множественный	0,700092							
R-квадрат	0,490129							
Нормированный	0,484646							
Стандартная ошибка	29991,32							
Наблюдения	95							
<b>Дисперсионный анализ</b>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>значимость F</i>			
Регрессия	1	8,04E+10	8,04E+10	89,39899	2,91E-15			
Остаток	93	8,37E+10	8,99E+08					
Итого	94	1,64E+11						
	<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	-79189,5	15934	-4,96984	3,04E-06	-110831	-47547,7	-110831	-47547,7
$x_1$	1328,578	140,5144	9,455104	2,91E-15	1049,544	1607,612	1049,544	1607,612

Рисунок 1. Регрессионный анализ  $Y$  и  $x_1$

Как это видно из рисунка 3, коэффициент  $b_0$  равен -79189,5, что говорит нам о пересечении с осью координат. С экономической точки зрения данный параметр  $x_1$  не может быть интерпретирован, поскольку при нулевом числе сотрудников выручка не может быть отрицательной. Поскольку  $Y$  – пересечение меняет свой знак, построим данную итерацию без свободного члена (рисунок 4).

Вывод итогов							
<b>Регрессионная статистика</b>							
Множественный R	0,909333						
R-квадрат	0,826886						
Нормированный R-квадрат	0,816248						
Стандартная ошибка	33559,75						
Наблюдения	95						
<b>Дисперсионный анализ</b>							
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>значимость F</i>		
Регрессия	1	5,05683E+11	5,06E+11	448,9947	2,29E-37		
Остаток	94	1,05868E+11	1,13E+09				
Итого	95	6,11551E+11					
<b>Коэффициенты</b>							
	<i>Коэффициент</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	<i>Верхние 5%</i>	<i>Нижние 5%</i>	<i>Верхние 95,0%</i>
Y-пересечение	0	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
x1	643,3886	30,36356372	21,18949	1,43E-37	583,101	703,6761	583,101 703,6761

*Рисунок 4 – Регрессионная зависимость Y от x<sub>1</sub>  
при условии отсутствия свободного члена*

Таким образом, регрессионная модель принимает следующий вид:

$$Y = 643,39 x_1.$$

Экономический смысл данной зависимости отображает, что 1 сотрудник приносит 643,39 руб. выручки компании.

Коэффициент детерминации данной модели равен 0,827. Что соответствует примерно 83%. Это означает, что численность сотрудников оказывает значительное влияние на показатели выручки и подтверждает правильность выбора фактора для построения данной модели. Чем выше коэффициент детерминации, тем качественнее модель. Рассчитанный уровень значимости (значимость F из рисунка 4) мал. Данный показатель означает допустимую вероятность ошибки первого рода (ложноположительного решения), поскольку она близка к 0, то данный критерий подтверждает полученную значимость  $R^2$ .  $S = 30,36$  – стандартная ошибка оценки. Стандартная ошибка регрессии рассматривается в качестве меры разброса данных наблюдений от смоделированных значений. Чем меньше значение стандартной ошибки регрессии, тем качество модели выше.

Для проверки сравним табличное значение с фактическим, представленным на рисунке 4.  $F_{\text{табл}} - 3,98$ , это меньше фактического значения. Следовательно, переменная является значимой.

Найденное по данным наблюдения значение t-критерия – 21,18 (его еще

называют наблюдаемым или фактическим) сравниваем с табличным (критическим) значением, определяемым по таблицам распределения Стьюдента - 1,984. Если значение статистики по абсолютной величине выше критического значения, то отличие коэффициента является статистически значимым (неслучайным), как в нашем случае. Если же наоборот, то незначимым (случайным), то есть истинный коэффициент вероятно равен или очень близок к предполагаемому значению.

Поля Нижние 95% – Верхние 95% на рисунке 4 означают доверительный интервал для параметра, т.е. с надежностью 0,95 этот коэффициент лежит в интервале от 583,10 до 703,68.

Аналогично произведем расчеты для отобранных факторов  $x_i$ . Результаты вычислений объединим в таблице 3.

*Таблица 3. Результаты регрессионного анализа факторов*

$x_i$	Уравнение регрессии	$R^2$	Критерий Фишера	Критерий Стьюдента	Доверительный интервал
$x_1$	$Y = 643,39 x_1$	0,83	448,99	21,18	583,10 – 703,68
$x_2$	$Y = 1531,212 x_2$	0,89	794,73	28,19	1423,36 – 1639,06
$x_3$	$Y = 1284,212 x_3$	0,87	646,69	25,43	1183,81 – 1384,32
$x_5$	$Y = 7192,84 + 419,88 x_5$	0,82	416,02	20,39	379,00 – 460,76
$x_6$	$Y = 98,55 + 4352,95 x_6$	0,84	479,41	21,89	3958,16 – 4747,74
$x_7$	$Y = 2,0784 x_7$	0,94	1388,93	37,27	1,97 – 2,19
$x_8$	$Y = 28177,88 + 2,1 x_8$	0,73	259,10	16,10	1,84 – 2,36

Как видно из таблицы, значения коэффициентов детерминации для полученных уравнений регрессии достаточно высоки, что говорит о значительном весе данных факторов в полученных моделях. Значения расчетных критериев Фишера в разы больше, чем критическое, полученное из таблицы – 3,98. Аналогично, критерий Стьюдента превосходит значение

критическое – 1,984. Это позволяет говорить нам о том, что переменные являются статистически значимыми и их включение необходимо.

Используя данные факторы и регрессионные модели, мы сможем прогнозировать результаты деятельности компании в будущих периодах, минимизировать риски за счёт детерминации источников дохода компании.

### **Использованные источники:**

1 Экономические термины [Электронный ресурс]. – URL: <http://abc.informbureau.com> (дата обращения: 04.04.2018).

2 Елисеева, И.И. Общая теория статистики [Текст] / И.И. Елисеева, М.М. Юзбашев. – М.: Финансы и статистика, 2014. — 656 с.

3 Котенко, А.П. Эконометрика: курс лекций [Текст]: учебное пособие / А.П. Котенко – Самара: Изд-во Самар. гос. аэрокосм. ун-та им. С.П. Королева, 2012. – 83 с.