

## Содержание:

Исходные данные и цель.....	3
Показатели.....	4
Гистограммы .....	5
Диаграммы рассеивания.....	6
Выводы по диаграммам.....	7
Построение регрессионной модели.....	7
Результаты анализа многофакторной регрессионной модели.....	8
Корреляция и критерий Фишера.....	9
Анализ остатков и исключение выбросов.....	12
Построение улучшенной модели.....	15
Корреляция и критерий Фишера.....	16
ВЫВОДЫ.....	17

## 1. ИССХОДНЫЕ ДАННЫЕ И ЦЕЛЬ

Не смотря на то, что Латвийская Республика уже более четырех лет является членом Европейского Союза, роль Латвии на Европейском рынке до сих пор не стала хоть сколько-нибудь ощутимой. Стоит вспомнить евро-скептиков того времени, утверждавших что Латвия никогда не станет экспортером, а будет вынуждена «захлебываться западным импортом», как становится крайне неприятно из-за того, что их прогнозы вполне могут оправдаться. Разумеется, если латвийские компании не начнут экспортировать продукцию на западные рынки. На данный момент Латвия очень активно экспортирует пиломатериалы (17 %) и лесокругляк (5 %). Почти столько же приходится на экспорт железа и металла в брусках. Практически ничего кроме сырья и полуфабрикатов Латвия не продает за рубеж, а это говорит о невысоком уровне развития промышленности и рынка в целом.

Из написанного выше вполне можно заключить, что для Латвии является крайне важным поднять объемы своего экспорта. Для достижения этой цели сперва следовало выяснить, какие именно факторы в состоянии стимулировать потенциальных экспортеров. После того, как были определены ключевые факторы, была сформулирована также и цель работы.

### **Цель работы**

Выявить взаимосвязь между желанием экспортировать и следующими переменными:

- Уровень заинтересованности в гос.помощи;
- Размер компании (численность служащих);
- Выручка компании;
- Период работы компании на рынке;
- Число продуктов, выпускаемых компанией;
- Обучение сотрудников;
- Опыт руководства в области работы на международном рынке.

**Зависимой** переменной в нашем случае будет желание экспортировать, а остальные переменные будут **независимыми**.

Для построения модели произвольно были выбраны 60 латвийских предприятий разного размера и разного профиля.

## 2. ПОКАЗАТЕЛИ

В таблице представлен фрагмент данных, которые были введены в пакет Statistika:

	1 Total derived employment	2 GNP deflator, 1954 = 100	3 Gross National Product	4 Unemployment	5 Size of armed forces	6 Non-institutional population aged 14 and over	7 year
1	60323	83	234289	2356	1590	107608	1947
2	61122	88,5	259426	2325	1456	108632	1948
3	60171	88,2	258054	3682	1616	109773	1949
4	61187	89,5	284599	3351	1650	110929	1950
5	63221	96,2	328975	2099	3099	112075	1951
6	63639	98,1	346999	1932	3594	113270	1952
7	64989	99	365385	1870	3547	115094	1953
8	63761	100	363112	3578	3350	116219	1954
9	66019	101,2	397469	2904	3048	117388	1955
10	67857	104,6	419180	2822	2857	118734	1956
11	68169	108,4	442769	2936	2798	120445	1957
12	66513	110,8	444546	4681	2637	121950	1958
13	68655	112,6	482704	3813	2552	123366	1959
14	69564	114,2	502601	3931	2514	125368	1960
15	69331	115,7	518173	4806	2572	127852	1961
16	70551	116,9	554894	4007	2827	130081	1962

**Y = Total driven employment** – занятость населения

**GNP deflator** – дефлятор внутреннего валового продукта

**Gross National Product** – внутренний валовой продукт (ВВП)

**Unemployment** - безработные

**Size of armed forces** – размер вооруженных сил

**Non-institutional population aged 14 and over** - население в возрасте от 14 лет и старше

**Year** – год

**EXPORT - Y** – желание экспортировать (в нашем случае, зависимая переменная), измеряется в интервале от 1 до 5, где 1 – желание экспортировать очень низкое и 5 – желание экспортировать очень высокое.

**GOS\_POM** – Уровень заинтересованности в гос.помощи, измеряется в интервале от 1 до 5, где 1 – уровень заинтересованности очень низок и 5 – уровень заинтересованности в гос.помощи очень высок.

**RAZM\_COM** – Размер компании (численность служащих).

**VIRUCHKA** – Выручка компании, в тысячах латов.

**PERIOD** – Период работы компании на рынке, в годах.

**ASSORTIM** – Число продуктов, выпускаемых компанией.

**OBUCHENI** – Обучение сотрудников, 1 – проводится, 0 – не проводится.

**OPIT** – Опыт руководства в области работы на международном рынке, 1 – опыт имеется, 0 – опыт отсутствует.

В таблице представлен фрагмент данных, которые были введены в пакет Statistika:

**Таблица 1**

	1	2	3	4	5	6	7	8
	EXPORT - Y	GOS_POM	RAZM_COM	VIRUCHKA	PERIOD	ASSORTIM	OBUCHENI	OPIT
10	2,000	3,000	28,000	0,900	6,000	2,000	0,000	1,000
11	4,000	1,000	39,000	3,600	7,000	3,000	0,000	1,000
12	3,000	2,000	31,000	4,000	7,000	3,000	1,000	0,000
13	4,000	5,000	65,000	1,000	7,000	9,000	1,000	1,000
14	1,000	4,000	50,000	1,000	7,000	9,000	1,000	1,000
15	4,000	1,000	30,000	2,000	7,500	3,000	1,000	0,000
16	5,000	4,000	58,000	1,000	6,000	5,000	1,000	1,000
17	3,000	4,000	54,000	2,000	6,500	4,000	1,000	1,000
18	4,000	5,000	58,000	1,000	7,000	9,000	1,000	1,000
19	2,000	1,000	37,000	1,000	7,000	9,000	1,000	0,000
20	5,000	1,000	35,000	2,000	9,500	5,000	1,000	0,000
21	4,000	3,000	49,000	2,000	8,500	4,000	1,000	1,000
22	3,000	2,000	37,000	2,000	7,000	2,000	0,000	1,000
23	3,000	4,000	34,000	0,900	6,000	5,000	0,000	1,000
24	2,000	5,000	66,000	1,000	5,500	10,000	1,000	1,000
25	5,000	4,000	50,000	0,300	6,500	6,000	1,000	1,000
26	4,000	3,000	43,000	1,000	7,000	3,000	1,000	0,000
27	4,000	3,000	54,000	1,000	7,000	4,000	1,000	1,000

### 3. ГИСТОГРАММЫ

Ниже представлены гистограммы факторов, которые были выбраны для модели.

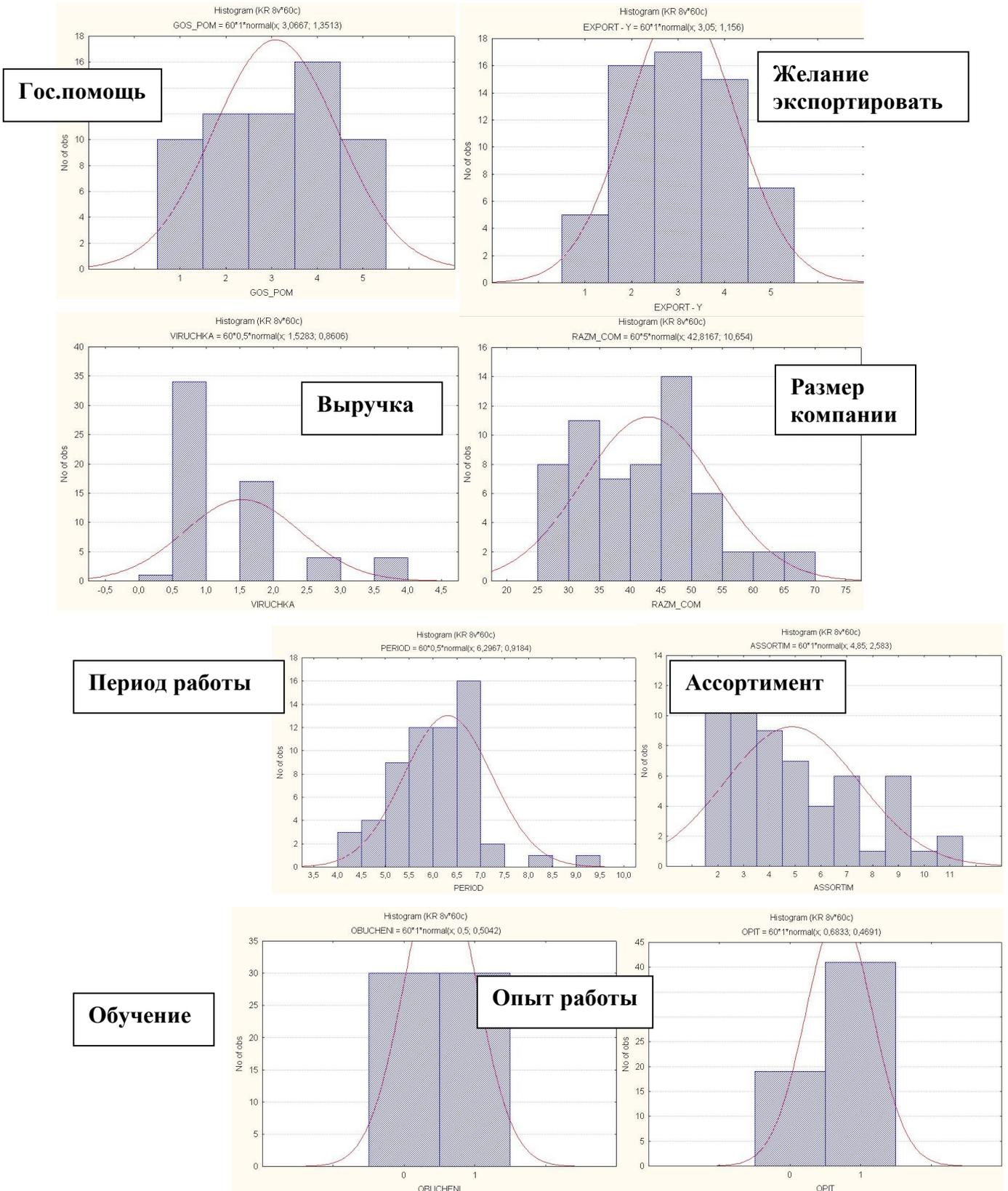


Рис. 3.1.гистограммы

#### 4. ДИАГРАММЫ РАССЕЙВАНИЯ

На рисунке 3.1 представлены диаграммы рассеивания факторов, которые были выбраны для модели.

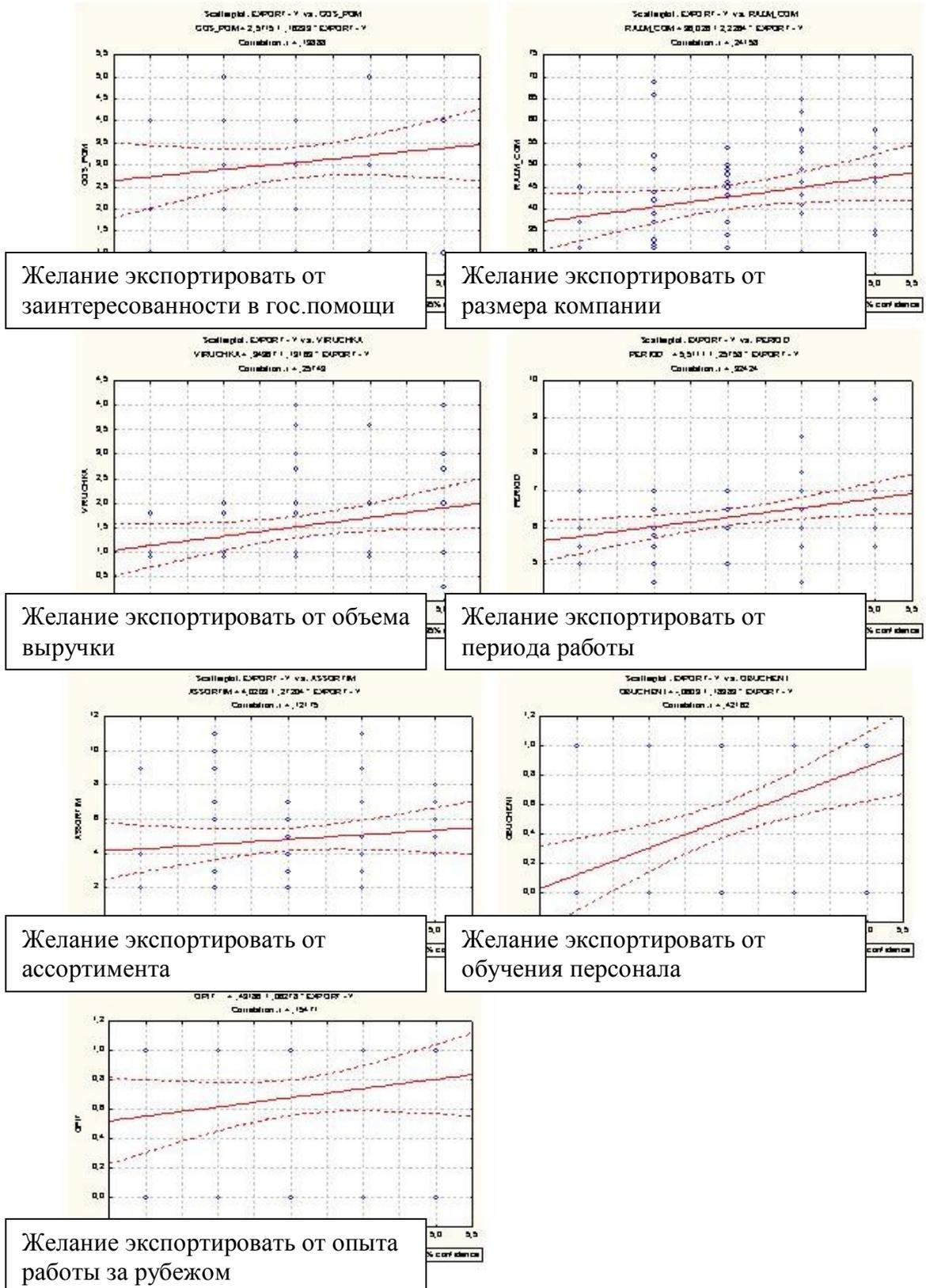


рис. 4.1. диаграммы рассеивания

## 5. ВЫВОДЫ ПО ДИАГРАММАМ

Из приведенных выше гистограмм видно, что многие факторы имеют нормальное распределение.

Из диаграмм рассеивания можно сделать вывод, что наблюдения (синие кружки), у многих факторов выстроились почти вдоль прямой, но не в границах области допустимых значений. Это означает, что между факторами не существует особенно сильной зависимости, однако среди некоторых факторов зависимость просматривается.

## 6. ПОСТРОЕНИЕ РЕГРЕССИОННОЙ МОДЕЛИ

Для модели была выбрана аддитивная линейная форма

$$y_t = \beta_0 + \beta_1 x_t^{(1)} + \dots + \beta_p x_t^{(p)} + \varepsilon_t$$

$y_t$  - значение результирующей (объясняемой) переменной – желание экспортировать.

$x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)}$  - объясняющие переменные.

$\beta_0, \beta_1, \dots, \beta_p$  - параметры, характеризующие значимость влияния  $x$  на  $y$ .

$\varepsilon_t$  - случайная составляющая, характеризующая разницу между модельным и наблюдаемым значениями.

## 7. РЕЗУЛЬТАТЫ АНАЛИЗА МНОГОФАКТОРНОЙ РЕГРЕССИОННОЙ МОДЕЛИ:

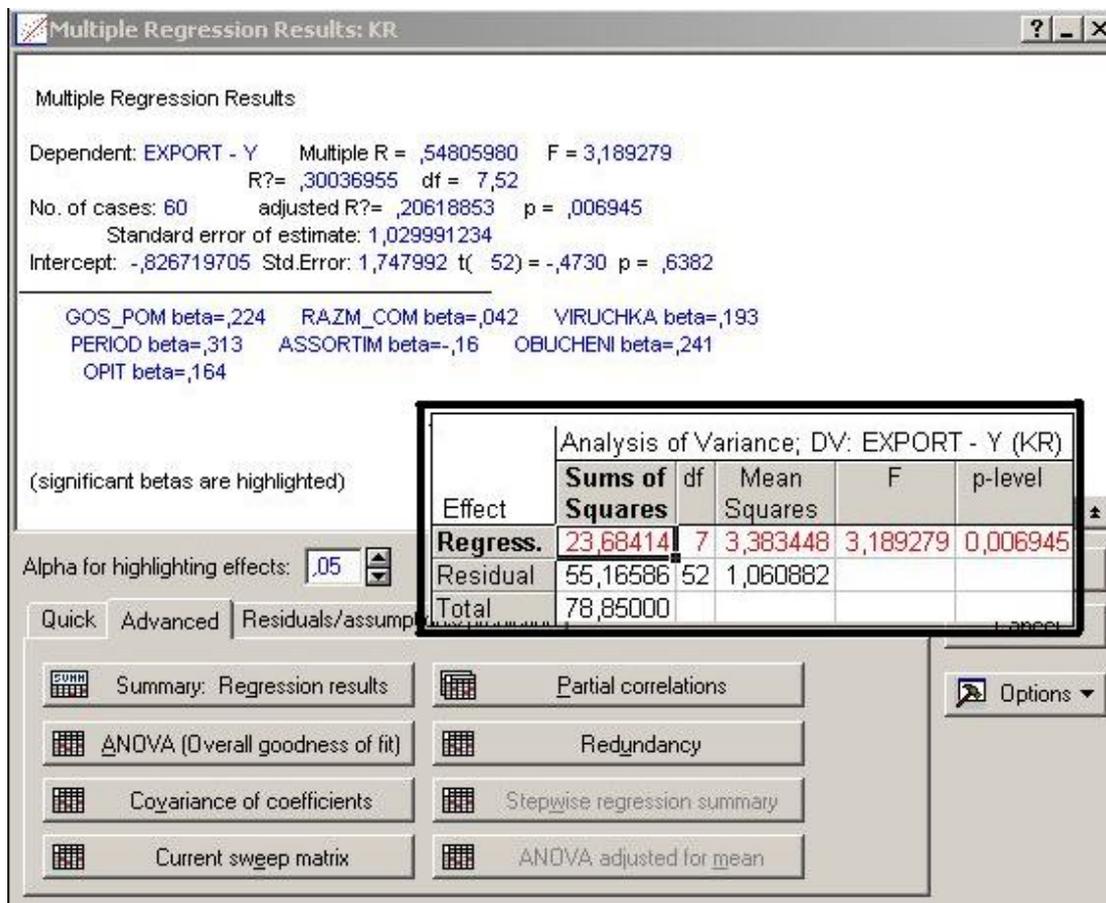


Рис. 7.1

### Показатели качества уравнения в целом:

- Коэффициент множественной детерминации  $R^2 = 0,30$
- Скорректированный коэффициент детерминации  $\check{R}^2 = 0,21$
- Стандартная ошибка регрессии SEE  $= 1,02999$
- Сумма квадратов остатков (СКО): RSS  $= 55,166$
- Критерий Фишера F  $= 3,189$

**Модель:**

Regression Summary for Dependent Variable: EXPORT - Y (KR)						
R= ,54805980 R <sup>2</sup> = ,30036955 Adjusted R <sup>2</sup> = ,20618853						
F(7 ,52)=3,1893 p<,00695 Std.Error of estimate: 1,0300						
N=60	Beta	Std.Err. of Beta	B	Std.Err. of B	t(52)	p-level
<b>Intercept</b>			-0,826720	1,747992	-0,472954	0,638225
GOS_POM	0,224442	0,219772	0,192013	0,188017	1,021251	0,311863
RAZM_COM	0,042205	0,208137	0,004580	0,022585	0,202774	0,840103
VIRUCHKA	0,192695	0,120878	0,258833	0,162367	1,594127	0,116969
PERIOD	0,313491	0,178070	0,394629	0,224158	1,760494	0,084204
ASSORTIM	-0,156986	0,173357	-0,070260	0,077587	-0,905561	0,369345
OBUCHENI	0,240683	0,196591	0,551823	0,450733	1,224279	0,226365
OPIT	0,164024	0,138662	0,404219	0,341715	1,182913	0,242224

Рис.7.2

$$\text{EXPORT} - \hat{Y} = -0,826720 + 0,192 * \text{GOS\_POM} + 0,00458 * \text{RAZM\_COM} + 0,2588 * \text{VIRUCHKA} + 0,178 * \text{PERIOD} - 0,07 * \text{ASSORTIM} + 0,552 * \text{OBUCHENI} + 0,4042 * \text{OPIT}$$

## 8. КОРРЕЛЯЦИЯ И КРИТЕРИЙ ФИШЕРА

Построим корреляционную модель из исследуемых восьми переменных: **EXPORT – Y, GOS\_POM, RAZM\_COM, VIRUCHKA, PERIOD, ASSORTIM, OBUCHENI, OPIT.**

Correlations (KR)								
Marked correlations are significant at p < ,05000								
N=60 (Casewise deletion of missing data)								
Variable	EXPORT - Y	GOS_POM	RAZM_COM	VIRUCHKA	PERIOD	ASSORTIM	OBUCHENI	OPIT
<b>EXPORT - Y</b>	1,00	0,14	0,24	0,26	0,32	0,12	0,42	0,15
GOS_POM	0,14	1,00	0,74	-0,12	-0,45	0,66	0,32	0,43
RAZM_COM	0,24	0,74	1,00	-0,13	-0,19	0,56	0,53	0,49
VIRUCHKA	0,26	-0,12	-0,13	1,00	0,22	-0,08	0,10	-0,05
PERIOD	0,32	-0,45	-0,19	0,22	1,00	-0,16	0,42	-0,31
ASSORTIM	0,12	0,66	0,56	-0,08	-0,16	1,00	0,51	0,30
OBUCHENI	0,42	0,32	0,53	0,10	0,42	0,51	1,00	0,11
OPIT	0,15	0,43	0,49	-0,05	-0,31	0,30	0,11	1,00

Рис.8.1 корреляция

Анализ матрицы парных коэффициентов корреляции показывает, что результативный показатель наиболее тесно связан с показателем OBUCHENI – обучением персонала.

Главным препятствием на пути эффективного применения регрессионного анализа, является мультиколлинеарность (наличие сильной корреляции между независимыми переменными, входящими в уравнение регрессии  $x_1, x_2, x_3, x_4, x_5, x_6$ ). Наиболее распространенный метод выявления коллинеарности основан на анализе парных коэффициентов корреляции. Он состоит в том, что две или несколько переменных

признаются коллинеарными (мультиколлинеарными), если парные коэффициенты корреляции больше определенной величины. На практике наиболее часто считают, что два аргумента коллинеарны, если парный коэффициент корреляции между ними по абсолютной величине больше 0,8.

В данном примере ни один парный коэффициент корреляции не превышает величины 0,45, что говорит об отсутствии такого явления как мультиколлинеарность.

Кроме этого, сама модель регрессии при проверке с помощью F – критерия оказывается значимой. Так как критерий Фишера, составляющий 3,189279, не входит в ОПГ (0; + 2,191626).

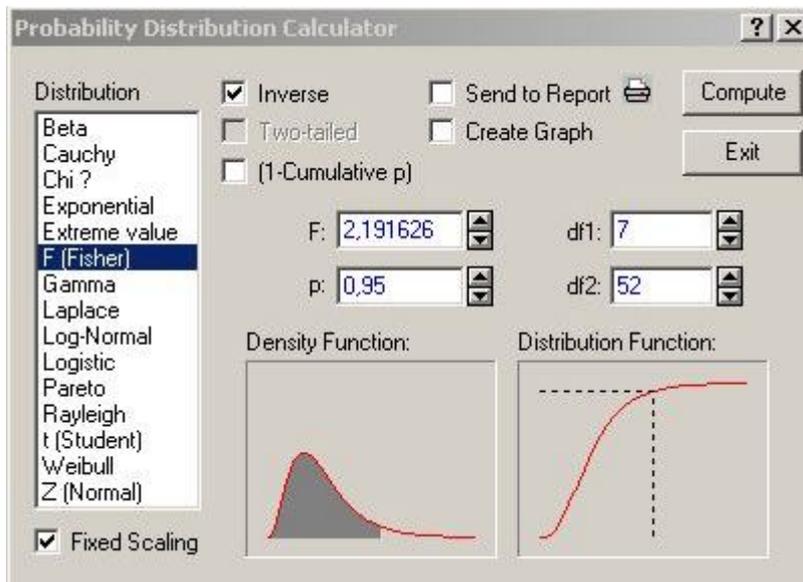


Рис.8.2 критерий Фишера

Чтобы узнать, какая из независимых переменных делает больший вклад в предсказания уровня общих выплат, изучим стандартизованные коэффициенты регрессии (или бэта):

Regression Summary for Dependent Variable: EXPORT - Y (KR)  
 R= ,54805980 R<sup>2</sup>= ,30036955 Adjusted R<sup>2</sup>= ,20618853  
 F(7,52)=3,1893 p<,00695 Std.Error of estimate: 1,0300

N=60	Beta	Std.Err. of Beta	B	Std.Err. of B	t(52)	p-level
<b>Intercept</b>			-0,826720	1,747992	-0,472954	0,638225
GOS_POM	0,224442	0,219772	0,192013	0,188017	1,021251	0,311863
RAZM_COM	0,042205	0,208137	0,004580	0,022585	0,202774	0,840103
VIRUCHKA	0,192695	0,120878	0,258833	0,162367	1,594127	0,116969
PERIOD	0,313491	0,178070	0,394629	0,224158	1,760494	0,084204
ASSORTIM	-0,156986	0,173357	-0,070260	0,077587	-0,905561	0,369345
OBUCHENI	0,240683	0,196591	0,551823	0,450733	1,224279	0,226365
OPIT	0,164024	0,138662	0,404219	0,341715	1,182913	0,242224

В столбце Beta показаны стандартизованные коэффициенты регрессии, а в столбце B – нестандартизованные.

Основные параметры данной модели, на которые стоит обратить внимание это: в первом столбце приведены коэффициенты факторов, где красным цветом выделены статистически значимые; чем ближе R-квадрат (коэффициент детерминации) к единице, это означает, что построенная модель объясняет почти всю изменчивость соответствующих переменных.

Далее убираем три коэффициента, которые дают наименьший вклад в зависимую переменную с большим p-level. В нашем случае это **GOS\_POM** – Уровень заинтересованности в гос.помощи, **RAZM\_COM** – Размер компании (численность служащих), **ASSORTIM** – Число продуктов, выпускаемых компанией. В результате мы видим, что остальные коэффициенты стали незначимыми.

## 9. АНАЛИЗ ОСТАТКОВ И ИСКЛЮЧЕНИЕ ВЫБРОСОВ

В распределении остатков выделяются выбросы (резко выделяющиеся наблюдения). На диаграмме Casewise plot of residuals, фрагмент которой представлен ниже, красным выделены выбросы в модели.

1	.	.	.	.	*	.	.
2	.	.	.	*	.	.	.
3	.	.	*	.	.	.	.
4	.	.	.	.	*	.	.
5	.	.	.	.	*	.	.
6	.	.	*	.	.	.	.
7	.	.	.	.	*	.	.
8	.	.	.	*	.	.	.
9	.	.	.	.	*	.	.
10	.	.	.	*	.	.	.
11	.	.	.	.	*	.	.
12	.	.	.	*	.	.	.
13	.	.	.	.	*	.	.
14	*	.	.	.	.	.	.
15	.	.	.	.	*	.	.
16	.	.	.	.	.	*	.
17	.	.	.	*	.	.	.
18	.	.	.	.	*	.	.
19	.	.	.	*	.	.	.
20	.	.	.	.	*	.	.
21	.	.	.	*	.	.	.
22	.	.	.	*	.	.	.
23	.	.	.	.	*	.	.
24	.	.	*	.	.	.	.
25	.	.	.	.	.	*	.
26	.	.	.	.	*	.	.
27	.	.	.	.	*	.	.
28	.	.	.	*	.	.	.
29	.	.	.	.	*	.	.

Рис.9.1 выбросы, фрагмент

В данной модели довольно много выбросов, для улучшения модели удалим их по одному.

В ходе работы был удален 21 выброс.

Это: V0=3, V0=4, V0=5, V0=12, V0=14, V0=16, V0=19, V0=21, V0=25, V0=26, V0=31, V0=36, V0=37, V0=38, V0=39, V0=40, V0=41, V0=42, V0=52, V0=57, V0=58.

При удалении каждого выброса модель каждый раз улучшалась, незначимые факторы становились постепенно значимыми.

Глядя на рисунок ниже видно, что явных выбросов нет. Хотя остатки и приняты облачный вид, но все показатели значительно улучшились и все факторы стали значимыми. Явных выбросов не наблюдается. Продолжать удалять выбросов не имеет смысла.

24	.	.	*	.	.	.	.
27	.	.	.	.	*	.	.
28	.	.	.	*	.	.	.
29	.	.	.	.	*	.	.
30	.	.	.	.	*	.	.
32	.	.	.	.	*	.	.
33	.	.	.	.	*	.	.
34	.	.	.	*	.	.	.
35	.	.	.	*	.	.	.

Рис.9.2 выбросы

Рис.9.3

Analysis of Variance; DV: EXPORT - Y (KR)					
Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	27,84141	4	6,960354	29,94771	0,000000
Residual	7,90218	34	0,232417		
Total	35,74359				

Рис.9.4

**Показатели качества уравнения в целом:**

- Коэффициент множественной детерминации  $R^2$  – 0,7789
- Скорректированный коэффициент детерминации  $\check{R}^2$  – 0,7529
- Стандартная ошибка регрессии SEE – 0,482
- Сумма квадратов остатков (СКО): RSS – 7,902
- Критерий Фишера F – 29,9477

## 10. ПОСТРОЕНИЕ УЛУЧШЕННОЙ МОДЕЛИ

Модель:

Regression Summary for Dependent Variable: EXPORT - Y (KF)						
R= ,88256472 R <sup>2</sup> = ,77892049 Adjusted R <sup>2</sup> = ,75291113						
F(4,34)=29,948 p<,00000 Std.Error of estimate: ,48210						
N=39	Beta	Std.Err. of Beta	B	Std.Err. of B	t(34)	p-level
<b>Intercept</b>			-2,89567	0,704262	-4,11164	0,000235
VIRUCHKA	0,285660	0,084692	0,32405	0,096073	3,37291	0,001869
PERIOD	0,673202	0,099840	0,72588	0,107653	6,74283	0,000000
OBUCHENI	0,253300	0,089292	0,48903	0,172389	2,83676	0,007626
OPIT	0,269443	0,090472	0,57322	0,192472	2,97821	0,005316

Рис.10.1

$$\text{EXPORT} - \hat{Y} = -2,89567 + 0,324 * \text{VIRUCHKA} + 0,7259 * \text{PERIOD} + 0,489 * \text{OBUCHENI} + 0,5732 * \text{OPIT}$$

Модель полностью значимая, все факторы значимы, это так же подтверждает p-level. Так как он не превышает  $\lambda$  и это говорит о том что гипотеза о незначимости модели отвергается. Так же об улучшение модели говорят показатели, они изменились в лучшую сторону.

## 11. КОРРЕЛЯЦИЯ И КРИТЕРИЙ ФИШЕРА

Корреляция, согласно представленной ниже таблице, практически не имеет высоких значений:

Correlations (KR)					
Marked correlations are significant at $p < ,05000$					
N=39 (Casewise deletion of missing data)					
Variable	EXPORT - Y	VIRUCHKA	PERIOD	OBUCHENI	OPIT
EXPORT - Y	1,00	0,48	0,75	0,54	0,00
VIRUCHKA	0,48	1,00	0,31	0,10	-0,12
PERIOD	0,75	0,31	1,00	0,35	-0,38
OBUCHENI	0,54	0,10	0,35	1,00	0,09
OPIT	0,00	-0,12	-0,38	0,09	1,00

Рис.11.1 корреляция

Мультиколлинеарность – нормальное явление. Однако в нашей модели данного явления не наблюдается.

Модель регрессии при проверке с помощью F – критерия оказывается значимой, поскольку критерий Фишера, составляющий 29,9477, не входит в ОПГ (0; + 2,64989).

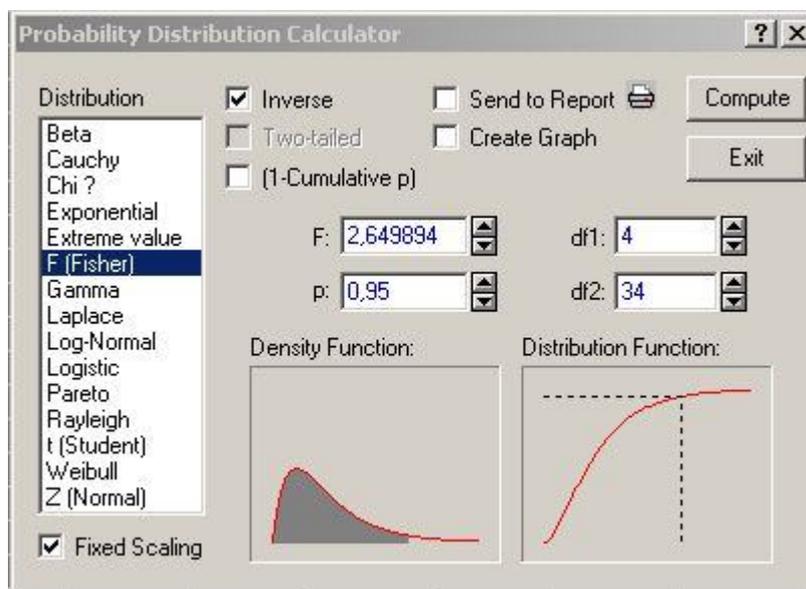


Рис.6.1.2 критерий Фишера

## 12. ВЫВОДЫ

Если сравнить модель в первом в виде и в последующем, видно, что она ощутимо улучшилась за счет удаления выбросов и отказа от части коэффициентов. Все факторы стали значимые и показатели качества модели значительно улучшились:

- Коэффициент множественной детерминации изменился с  $R^2 = 0,30$  до  $R^2 = 0,7789$
- Скорректированный коэффициент детерминации изменился с  $\check{R}^2 = 0,21$  на  $\check{R}^2 = 0,7529$
- Стандартная ошибка регрессии с  $SEE = 1,02999$  на  $SEE = 0,482$
- Сумма квадратов остатков (СКО): с  $RSS = 55,166$  на  $RSS = 7,902$
- Критерий Фишера с  $F = 3,189$  на  $29,9477$

Исходя из результатов проведенного анализа, можно сделать вывод, что для того, чтобы у предприятия появилось или развивалось желание экспортировать, необходимо, чтобы компания обладала достаточной выручкой, проработала достаточное время на рынке, обладала опытом работы на зарубежных рынках и следила за обучением своего персонала. Следовательно, именно на эти факторы и надо обратить пристальное внимание нашему государству.