

*Блинов А.В.,
заместитель заведующего кафедрой «Техника и технологии»
Южно-Уральский технологический университет*

Россия, г. Челябинск

Шахвалеева С.Т.

студент

1 курс, кафедра «Техника и технологии»

Информационные системы и технологии

Россия, г. Челябинск

РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ В РОССИЙСКОЙ ФЕДЕРАЦИИ

***Аннотация:** необходимость контроля за прогрессирующими отраслями и новаторской деятельностью, развитие системы показателей, оценивающих инновационную активность организаций Российской Федерации, ставит в приоритет решение актуальной задачи в виде разработки математической модели развития инновационной деятельности в стране.*

***Ключевые слова:** инновация, математическая модель, математическое моделирование, корреляционный анализ, регрессионный анализ.*

***Annotation:** the need to monitor progressive industries and innovative activities, the development of a system of indicators evaluating the innovative activity of organizations in the Russian Federation, prioritizes the solution of an urgent task in the form of developing a mathematical model for the development of innovative activities in the country.*

Key words: *innovation, mathematical model, mathematical modeling, correlation analysis, regression analysis.*

В представленной работе для построения математической модели инновационной деятельности был использован метод корреляционно-регрессионного анализа для установления связей между переменными и отражения соответствующих причинно-следственных отношений.

На рисунке 1 приведены входные статистические данные для проведения корреляционно-регрессионного анализа, взятые из официальных источников государственной статистики «Росстат» и «Госкомстат России».

	X1	X2	X3 (млрд)	X4 (млн)	Y (млрд)
1994	3968	1106250	4	5146,1	39,8
1995	4059	1061044	9	12149,5	39,8
1996	4122	990743	11,1	19393,9	35,3
1997	4137	934637	18	24449,7	55
1998	4019	855190	10,5	250082,1	45,8
1999	4089	872400	24	480050,5	84,4
2000	4099	887729	17,4	76697,1	154,7
2001	4037	885568	23,7	105260,7	181,9
2002	3906	870878	31,1	135004,5	206,3
2003	3797	858470	41,6	169862,4	312,7
2004	3656	839338	47,5	196039,9	435,1
2005	3566	813207	76,9	230785,2	545,6
2006	3622	807066	97,4	288805,2	777,4
2007	3957	801135	132,7	371080,3	959
2008	3666	761252	162,1	431073,2	1103,3
2009	3536	742433	219,1	485834,3	934,6
2010	3492	736540	237,6	523377,2	1243,7
2011	3682	735273	313,9	610426,7	2106,7
2012	3566	726318	355,9	699869,8	2872,9
2013	3605	727029	425,3	749797,6	3507,9
2014	3604	732274	437,3	847527	3580
2015	4175	738857	439,4	914669,1	3843,4
2016	4032	722291	402,7	943815,2	4364,3
2017	3944	707900	377,9	1019152,5	4167
2018	3950	682580	420,5	1028247,7	4516,2
2019	4051	682456	489,1	1134786,7	4863,4
2020	4175	679333	549,6	1174534,3	5189,1
2021	4175	662702	626,6	1301490,9	6003,3

Рисунок.1. Исходная таблица данных

Зависимой переменной (Y) был взят объем инновационных товаров, работ и услуг, в качестве независимых переменных взяты: организации, выполнявшие исследования и разработки (X1); персонал, занятый исследованиями и разработками (X2); финансирование науки из средств федерального бюджета (X3); внутренние затраты на исследования и разработки (X4).

С помощью внутренних функций программы для работы с электронными таблицами Microsoft Excel, были построены точечные диаграммы для каждой факторной переменной X.

По построении точечных диаграмм факторных переменных, можно заметить некоторые точки данных, которые значительно отличаются от среднего диапазона в статистической выборке – выбросы. После фильтрации, в таблице остаются соответствующие данные указанные на рисунке 2.

	X1	X2	X3 (млрд)	X4 (млн)	Y (млрд)
1994	3968	1106250	4	5146,1	39,8
1995	4059	1061044	9	12149,5	39,8
1996	4122	990743	11,1	19393,9	35,3
1997	4137	934637	18	24449,7	55
2000	4099	887729	17,4	76697,1	154,7
2001	4037	885568	23,7	105260,7	181,9
2002	3906	870878	31,1	135004,5	206,3
2003	3797	858470	41,6	169862,4	312,7
2004	3656	839338	47,5	196039,9	435,1
2005	3566	813207	76,9	230785,2	545,6
2006	3622	807066	97,4	288805,2	777,4
2008	3666	761252	162,1	431073,2	1103,3
2009	3536	742433	219,1	485834,3	934,6
2010	3492	736540	237,6	523377,2	1243,7
2011	3682	735273	313,9	610426,7	2106,7
2012	3566	726318	355,9	699869,8	2872,9
2013	3605	727029	425,3	749797,6	3507,9
2014	3604	732274	437,3	847527	3580
2017	3944	707900	377,9	1019152,5	4167
2018	3950	682580	420,5	1028247,7	4516,2
2019	4051	682456	489,1	1134786,7	4863,4
2020	4175	679333	549,6	1174534,3	5189,1
2021	4175	662702	626,6	1301490,9	6003,3

Рисунок.2. Таблица данных по удалении выбросов

После редактирования исходных данных из получившейся таблицы построены новые точечные диаграммы, на которых теперь отсутствуют явно видимые выбросы.

Посредством построенных диаграмм определяется положительная линейная корреляция на графиках X3 и X4, X2 имеет отрицательную линейную корреляцию, на графике X1 прослеживается явная нелинейная корреляция.

С помощью функции анализ данных в Microsoft Excel проведён корреляционный анализ, результаты которого показаны на рисунке 3.

	X1	X2	X3 (млрд)	X4 (млн)	Y (млрд)
X1	1	0,300250111	0,003510514	0,030477109	0,1484142
X2	0,300250111	1	-0,835826323	-0,861996387	-0,785542484
X3 (млрд)	0,003510514	-0,835826323	1	0,983305921	0,976701689
X4 (млн)	0,030477109	-0,861996387	0,983305921	1	0,983557798
Y (млрд)	0,1484142	-0,785542484	0,976701689	0,983557798	1

Рисунок.3. Результаты корреляционного анализа

Линейное уравнение регрессии в данном случае имеет следующий вид:

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 \quad (1)$$

где Y – объем инновационных товаров, работ и услуг, эндогенная переменная;

X1 – организации, выполнявшие исследования и разработки, экзогенная переменная;

X2 – персонал, занятый исследованиями и разработками, экзогенная переменная;

X3 – финансирование науки из средств федерального бюджета, экзогенная переменная;

X4 – внутренние затраты на исследования и разработки, экзогенная переменная;

a_0 – свободная переменная;

a_1, a_2, a_3, a_4 – коэффициенты модели при факторных переменных.

Корреляционная связь определяется коэффициентом корреляции – это показатель, величина которого варьируется в пределах от -1 до $+1$. Если показатель близок к -1 , то переменные имеют сильную отрицательную связь, когда показатель близок к 0 , можно сказать, что обе переменные независимы друг от друга, если же показатель корреляции стремится к $+1$, то связь переменных является сильной положительной.

Исходя из результатов корреляционного анализа можно сделать следующие выводы:

1) факторная переменная X_1 слабо влияет на результирующую переменную Y , так как коэффициент корреляции равен 0,1484142, потому её можно исключить из расчётов;

2) факторная переменная X_2 имеет сильную отрицательную корреляцию в отношении результирующей переменной, она равна - 0,785542484, это значит, что у обоих показателей есть обратная зависимость. Связь переменной X_2 с другими факторными переменными является мультиколлинеарной и также имеет сильную отрицательную корреляцию.

3) факторная переменная X_3 и зависимая переменная Y имеют очень высокую положительную корреляцию равную 0,976701689, что свидетельствует о значительном влиянии переменной X_3 на результирующую. Так же X_3 имеет мультиколлинеарность с переменными X_2 и X_4 , как и в предыдущем случае, переменную можно исключить из модели, оставив в расчётах лишь одну факторную переменную – X_4 , так как она имеет наибольший коэффициент корреляции, который равен 0,983557798.

	X_4 (млн)	Y (млрд)
X_4 (млн)	1	0,983557798
Y (млрд)	0,983557798	1

Рисунок.4. Корреляционная матрица, после исключения из модели мультиколлинеарности

Исключив из корреляционной матрицы соответствующие факторные переменные, линейное уравнение регрессии выглядит следующим образом:

$$Y = a_0 + a_4 X_4 \quad (2)$$

По значимым переменным, определённым с помощью корреляционного анализа, происходит расчёт регрессионной статистики и дисперсионный анализ (табл. 1,2,3).

Таблица 1

Результат регрессионного анализа

Множественный R	0,983557798
R-квадрат	0,967385941
Нормированный R-квадрат	0,965832891
Стандартная ошибка	368,0319849
Наблюдения	23

Таблица 2

Результат дисперсионного анализа

	df	SS	MS	F	Значимость F
Регрессия	1	84369474,15	84369474,1	622,894096	4,29515E-17
Остаток	21	2844398,38	135447,542		
Итого	22	87213872,53			

Таблица 3

Дисперсионный анализ данных

	Коэффициенты	Стандартная ошибка	t- статистика	P-Значение
Y-пересечение	-382,392321	118,2804805	-3,2329284	0,003987
X4 (млн)	0,004584565	0,000183692	24,9578464	4,3E-17

По результатам регрессионной статистики и дисперсионного анализа можно сделать следующие выводы:

1. показатель R-квадрат = 0,967385941. R-квадрат является коэффициентом линейной детерминации. Данный показатель является одной из наиболее эффективных оценок адекватности регрессионной модели. В данном случае показатель равен 96,7, следовательно это может утверждать о высокой точности;

2. значимость F – достоверность модели по уровню значимости критерия Фишера. Для положительной оценки значимости, показатель не

должен превышать значение 0,05. В проведенном дисперсионном анализе результат равен $4,29515E-17$, из чего следует, что показатель удовлетворяет критерию Фишера и является значимым;

3. Р-значение является значением вероятности, эта статистическая величина выражается числом от 0 до 1, используемое для проверки достоверности гипотезы. Уровень значимости, для положительного результата не должен превышать 5% или 0,05. Результатом дисперсионного анализа стал результат Р равный 0,003987, что соответственно меньше 0,05 и является показателем достоверности гипотезы.

Исходя из данных таблицы 3 и рассчитанных в ней коэффициентов представляется возможным составить уравнение многомерной линейной регрессии для определения факторных переменных:

$$Y = 0,046X - 382,39 \quad (3)$$

На основании проведенного исследования, можно сделать следующие выводы:

1. три из четырёх факторных переменных, а именно: X2 – персонал, занятый исследованиями и разработками, X3 – финансирование науки из средств федерального бюджета, X4 – внутренние затраты на исследования и разработки, в значительной мере влияют на результирующую переменную Y – объем инновационных товаров, корреляционная связь каждой переменной с Y составила более 0,78 (учитывая сильную отрицательную корреляцию X2, составляющую -0,78);

2. между тремя факторными переменными: X2 - персонал, занятый исследованиями и разработками, X3 - финансирование науки из средств федерального бюджета, X4 - внутренние затраты на исследования и разработки, отобранными для корреляционно-регрессионного анализа присутствовала мультиколлинеарность, которая свидетельствует о тесной корреляционной зависимости переменных друг от друга, что значительно ухудшает качество разрабатываемой модели. Наличие мультиколлинеарность представляло бы

матрицу слабо обусловленной, потому из трёх имеющихся факторов, тесно связанных друг с другом, были удалены два, наименее влияющих на результирующую переменную. В связи с этим, повторный корреляционный и регрессионный анализ проводился на основе одной факторной переменной;

3. по результатам регрессионного анализа показатель R-квадрат равен 0,967385941, что свидетельствует о высокой точности построенной модели.

Таким образом, построенная математическая модель инновационной деятельности в виде линейного уравнения регрессии позволяет производить расчёты факторной переменной, не прибегнув к повторному анализу статистических данных.

Перспективы дальнейшего исследования проблемы видятся в более детальном построении математической модели, посредством расширения числа факторных переменных и исследования развития инновационной деятельности иными видами регрессионного анализа.

Использованные источники:

1. Денис Соловьев, Станислав Кузора. Применение математического моделирования в инновационной деятельности. [Электронный ресурс]. URL: https://www.researchgate.net/publication/333625431_Primerenie_matematicheskogo_modelirovania_v_innovacionnoj_deatelnosti (дата обращения: 10.12.2022).

2. Корреляционный анализ. [Электронный ресурс]. URL: https://e.vyatsu.ru/pluginfile.php/462616/mod_resource/content/3/Теоретический%20материал_корреляционный%20анализ.pdf (дата обращения: 19.11.2022).

3. Назимова И.Р. Экономико-математическое моделирование инновационного развития Российской Федерации методом главных компонент. [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/ekonomiko-matematicheskoe-modelirovanie->

innovatsionnogo-razvitiya-rf-metodom-glavnyh-komponent/viewer (дата обращения: 20.12.2022).

4. Основы анализа данных. Регрессионный анализ. [Электронный ресурс]. URL: <https://intuit.ru/studies/courses/6/6/lecture/172?page=4> (дата обращения: 1.12.2022).

5. С.В. Звонарев. Основы математического моделирования. [Электронный ресурс]. URL: https://elar.urfu.ru/bitstream/10995/68494/1/978-5-7996-2576-4_2019.pdf (дата обращения: 1.12.2022).

6. Статистическое исследование инновационной деятельности в экономике. [Электронный ресурс]. URL: https://conference.osu.ru/assets/files/conf_info/conf5/37.pdf (дата обращения: 31.12.2022).