|  |  |  |
| --- | --- | --- |
|  |  |  |
| МИНОБРНАУКИ РОССИИ | | |
| Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА – Российский технологический университет»**  **РТУ МИРЭА** | | |
| Институт информационных технологий | | |
| Кафедра прикладной математики | | |

|  |
| --- |
|  |

**КУРСОВАЯ РАБОТА**

по дисциплине Технологии организации, обработки и хранения статистических данных\_\_\_\_

(наименование дисциплины)

**Тема курсовой работы** Анализ признаков и рисков сахарного диабета для раннего прогнозирования\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Студент группы** ИМБО-02-22 Ким К.С.

(учебная группа, фамилия, имя отчество, студента) (подпись студента)

**Руководитель курсовой работы** доцент к.п.н. Митина О.А.\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(должность, звание, ученая степень) (подпись руководителя)

**Рецензент** (при наличии)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(должность, звание, ученая степень) (подпись рецензента)

Работа представлена к защите «\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2023г.

Допущен к защите «\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2023г.

Москва 2023 г.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| МИНОБРНАУКИ РОССИИ | | |
| Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА – Российский технологический университет»**  **РТУ МИРЭА** | | |
| Институт информационных технологий | | |
| Кафедра прикладной математики | | |

|  |  |
| --- | --- |
| **Утверждаю** | |
| Заведующий кафедрой |  |
|  | *Подпись* |
|  | |
| *ФИО* | |
| «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2023г. | |

**ЗАДАНИЕ**

**на выполнение курсовой работы** по дисциплине

«Технологии организации, обработки и хранения статистических данных»

Студент \_Ким К.С.\_\_\_ Группа \_\_ИМБО-02-22\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |
| --- | --- |
| **Тема:** | Анализ признаков и рисков сахарного диабета для раннего прогнозирования |

**Исходные данные:** Аналитическая платформа Loginom**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Перечень вопросов, подлежащих разработке, и обязательного графического материала:** в чем заключается понятие «рынок труда»; назовите показатели, характеризующие \_\_\_\_\_\_\_ корреляционно-регрессионный анализ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Срок представления к защите курсовой работы:** **до «\_\_»\_\_\_\_\_\_\_\_\_\_\_\_2023г.**

**Задание на курсовую работу выдал** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (\_\_\_\_\_\_\_\_\_\_\_\_\_)

*Подпись руководителя Ф.И.О. руководителя*

**Задание на курсовую работу получил** «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_ 2023г.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (\_\_\_\_\_\_\_\_\_\_\_\_\_)

*Подпись обучающегося* Ф.И.О. обучающегося

Москва 2023г.

**ОТЗЫВ**

**на курсовую работу**

**по дисциплине «Технологии организации, обработки и хранения статистических данных»**

**Студент Ким Кирилл Сергеевич** **ИМБО-02-22**

(ФИО студента) (Группа)

Характеристика курсовой работы

|  |  |  |  |
| --- | --- | --- | --- |
| Критерий | Да | Нет | Не полностью |
| 1. Соответствие содержания курсовой работы указанной теме |  |  |  |
| 2. Соответствие курсовой работы заданию |  |  |  |
| 3. Соответствие рекомендациям по оформлению текста, таблиц, рисунков и пр. |  |  |  |
| 4. Полнота выполнения всех пунктов задания |  |  |  |
| 5. Логичность и системность содержания курсовой работы |  |  |  |
| 6. Отсутствие фактических грубых ошибок |  |  |  |

Замечаний:

Рекомендуемая оценка:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_

Подпись руководителя ФИО руководителя

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 5](#_Toc1693433724)

[1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ 6](#_Toc2052349505)

[1.1 Рынок труда 7](#_Toc1182116011)

[1.2 Корреляционно-регрессионный анализ 13](#_Toc893717917)

[2 ПРАКТИЧЕСКАЯ ЧАСТЬ 18](#_Toc860276641)

[2.1 Корреляционно-регрессионный анализ на примере отрасли добычи полезных ископаемых в США. 19](#_Toc712191868)

[ЗАКЛЮЧЕНИЕ 32](#_Toc1699284128)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 33](#_Toc229578108)

[Теоретическая часть 34](#_Toc1294647169)

[Практическая часть 36](#_Toc1291425966)

[ПРИЛОЖЕНИЯ 36](#_Toc811099868)

[Приложение А 37](#_Toc1225449068)

# ВВЕДЕНИЕ

Современный мир характеризуется динамично развивающимся рынком труда, что требует всестороннего изучения связей между изменениями рынка труда внутри отраслевых структур и экономическими показателями.

Рынок труда функционирует во взаимодействии с другими типами рынков, при этом дисбаланс, возникший в результате каких-либо изменений, передается по цепочке. Имеющиеся прогнозы изменений рынка труда будут преимуществом для предпринимателя. Для прогнозирования изменений на рынке труда в зависимости от каких-либо микроэкономических показателей требуется на основе имеющихся данных найти тесноту связи между изменениями рынка труда и этими показателями, а также построить линию тренда для этой зависимости. В этом может помочь корреляционно-регрессионный анализ.

Таким образом, тема данной курсовой работы является актуальной для сферы рынка труда и экономической сферы.

Цель данной курсовой работы — провести корреляционно-регрессионный анализ влияния основных микроэкономических показателей на структурные изменения рынка труда в отрасли.

Задачи, решаемые в данной курсовой работе:

* + изучение научной и методической литературы о рынке труда и корреляционно-регрессионном анализе;
  + выполнение корреляционно-регрессионного анализа собранных данных;
  + использование знания математической статистики с использованием современных средств обработки данных: аналитической платформы Loginom;
  + обучение оформлению официальных документов.

# 1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

## Сахарный диабет

Сахарный диабет – это хроническое заболевание эндокринной системы человека, характеризующиеся длительным повышением концентрации глюкозы в крови и сопутствующими изменениями процессов обмена веществ. Представляет серьёзную угрозу для здоровья и жизни больного, так как провоцирует развитие тяжелых сопутствующих заболеваний. В основе сахарного диабета лежит инсулиновая недостаточность, что приводит к увеличению сахара в крови и появлению его в моче. Это происходит из –за того, что поджелудочная железа вырабатывает недостаточное количество гормона – инсулина, который регулирует углеводный обмен в организме. При наследственной предрасположенности к заболеванию его возникновение могут провоцировать переедание, злоупотребление сладким, нервное перенапряжение, вирусная инфекция. Диабет может развиваться после краснухи, кори, гриппа и других вирусных заболеваниях. Самостоятельно выявить заболевание невозможно, но некоторые отклонения в организме могут быть первыми тревожными сигналами. К таковым признакам можно отнести: частое и обильное мочеиспускание, ощущение постоянной жажды, кожный зуд, постоянные скачки весовых соотношений (масса тела то резко увеличивается, то моментально снижается), проявление кожных высыпаний. На начальной стадии болезни сахарный диабет практически незаметен, все симптомы усиливаются в тот момент, когда недуг активизировался. Диагностировать болезнь при первичном осмотре и беседе с пациентом невозможно, для того чтобы поставить точный диагноз нужно обязательно провести лабораторные исследования. К сожалению, несвоевременное

лечение сахарного диабета приводит к серьезным осложнениям для больного. Различают 2 типа сахарного диабета:

Сахарный диабет 1 типа (инсулин зависимый диабет) – это деструкция клеток поджелудочной железы, которая приводит к полной инсулиновой недостаточности. При 1 типе диабета практически все клетки поджелудочной железы, которые выделяют инсулин, разрушаются вследствие чего железа не в состоянии продуцировать инсулин. Инфекция сама по себе не разрушает клетки поджелудочной железы, она включает иммунную систему, клетки которой и уничтожают клетки поджелудочной. Данный тип составляет 10% от общего числа больных сахарным диабетом. Болеют дети и взрослые до 30 лет. При диабете 1 типа пациент вынужден постоянно вводить инсулин, который нужен для нормального передвижения глюкозы в организме.

Основной метод лечения – инъекции инсулина. Поэтому и называется 1 тип диабета – инсулин зависимый диабет. Этот тип диабета активно прогрессирует, быстро развиваются осложнения и стадия декомпенсации.

Сахарный диабет 2 типа (инсулин независимый) – это нарушение углеводного обмена с развитием гипергликемии. При данном типе поджелудочная железа не разрушается и продолжает вырабатывать инсулин, но в организме развивается резистентность (снижение чувствительности) клеток к инсулину. В результате этого в клетки не поступает нужного количества глюкозы, даже в присутствии инсулина. При наличии данного типа диабета у кого-то в роду, вероятность развития у потомка на протяжении жизни составляет 40 %. Также, часто его развитию способствуют ожирение, перенапряжение и стрессовые ситуации. Его коррекция может быть достигнута диетой, снижением массы тела и сахароснижающими таблетками [6, с.36].

При диабете 2 типа не нужно вводить инсулин, так как проблема не в выработке инсулина, а именно в усвоении глюкозы тканями [1, с.56]. Но, по мере прогрессирования диабета, выделение инсулина клетками

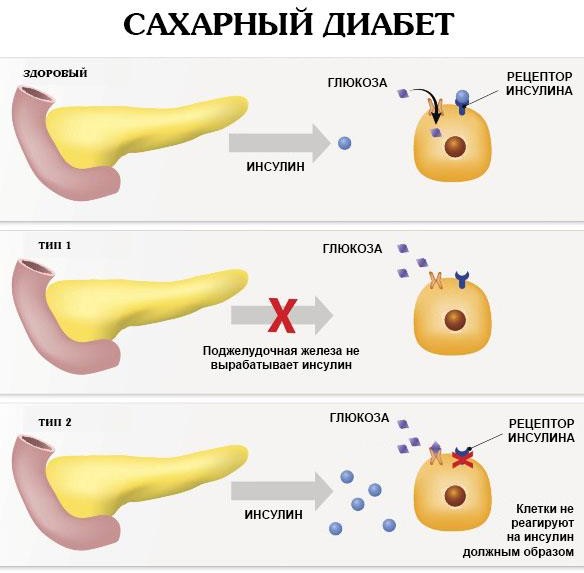
поджелудочной железы снижается и тогда приходиться назначать инсулин. (рис. 1).

Рис.1. Механизм развития сахарного диабета 1 и 2 типа.

Сахарный диабет является сложным заболеванием, которое трудно поддается лечению. При его развитии в организме происходит нарушение углеводного обмена и снижение синтеза инсулина поджелудочной железой, в результате чего глюкоза перестает усваиваться клетками и оседает в крови в виде микрокристаллических элементов. Точные причины, по которым начинает развиваться данный недуг, ученым установить до сих пор не удалось. Но благодаря им были выявлены факторы риска, которые могут спровоцировать возникновение этого заболевания у людей разностной возрастной категории.

Прежде чем рассматривать факторы риска развития сахарного диабета, необходимо сказать, что это заболевание, как уже было сказано выше, имеет два типа, и каждый из них имеет свои особенности. Диабет 1 типа характеризуется системными изменениями в организме, при которых нарушается не только углеводный обмен, но и функциональность поджелудочной железы. По каким – то причинам ее клетки перестают вырабатывать инсулин в нужном количестве, в результате чего сахар, проникающий в организм вместе с пищей, не подвергается процессам расщепления и, соответственно, не может усваиваться клетками[3, с.66].

Самыми распространенными осложнениями этой болезни являются следующие состояния:

* гипергликемия – повышение уровня сахара в крови за пределы нормы (свыше 7 ммоль/л);
* гипогликемия – снижение уровня глюкозы в крови за пределы нормы (ниже 3,3 ммоль/л);
* гипергликемическая кома – повышение уровня сахара в крови свыше 30 ммоль/л;
* гипогликемическая кома – снижение уровня глюкозы в крови ниже 2,1 ммоль/л;
* диабетическая стопа – снижение чувствительности нижних конечностей и их деформация;
* диабетическая ретинопатия – снижение остроты зрения;
* тромбофлебит – образование в стенках сосудов бляшек;
* гипертония – повышение артериального давления;
* гангрена – некроз тканей нижних конечностей с последующим развитием абсцесса;
* инсульт и инфаркт миокарда (рис.2).

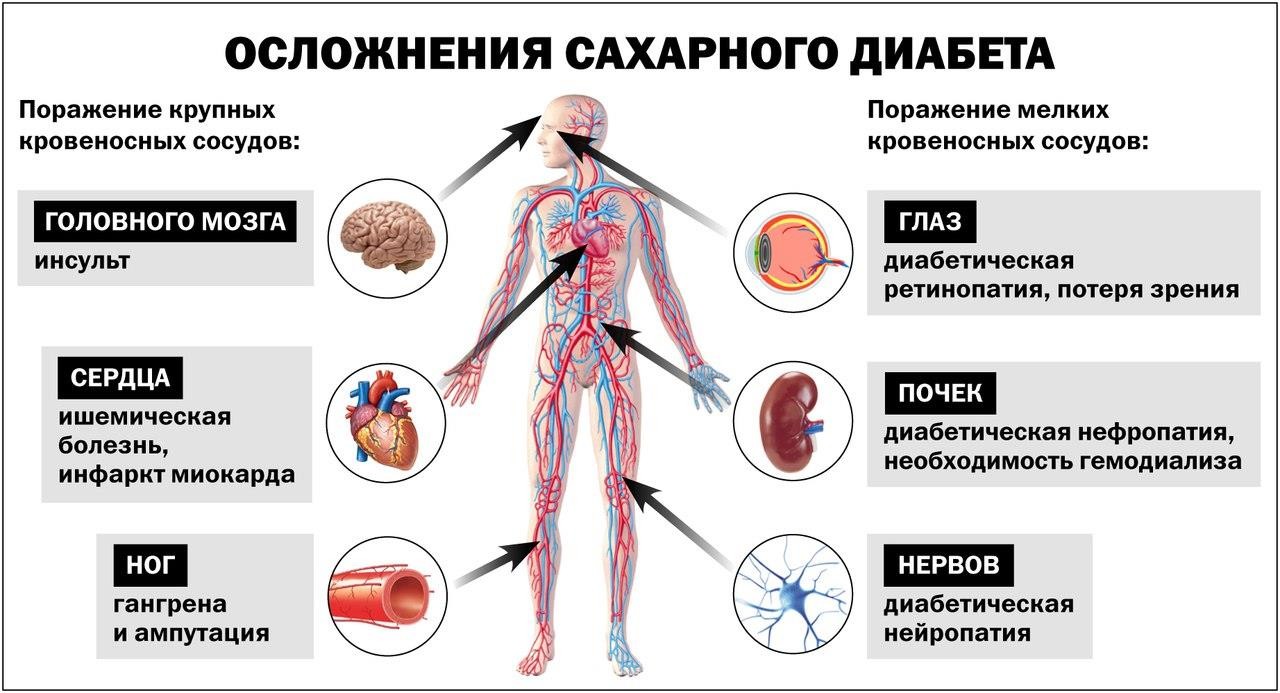


Рис.2. Осложнения сахарного диабета.

Основными факторами его развития являются:

* наследственная предрасположенность;
* вирусные заболевания;
* интоксикация организма;
* неправильное питание;
* частые стрессы.

В возникновении СД1 главную роль играет наследственная предрасположенность. Если кто-то из членов семьи страдает от этого недуга, то риски его развития у последующего поколения составляют примерно 10- 20%.

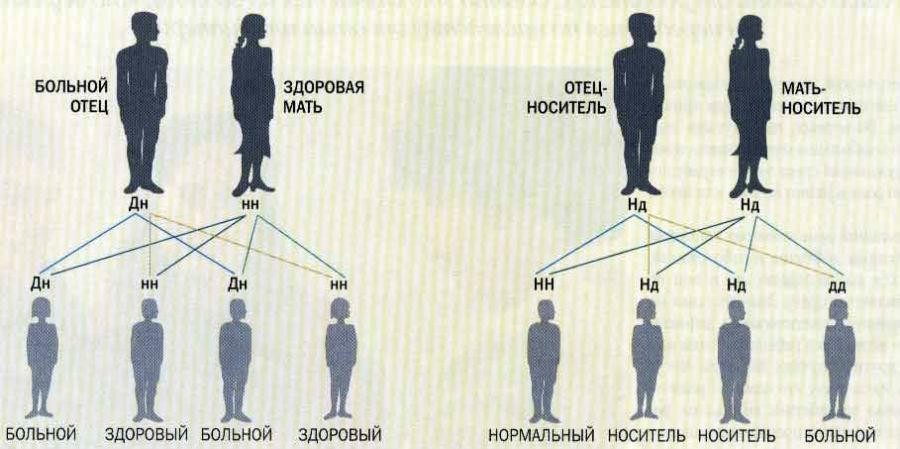
При этом следует отметить, что в данном случае речь идет не об установленном факте, а о предрасположенности. То есть если мать или отец болеют СД1, это вовсе не значит, что у их детей также будет диагностировано это заболевание. Предрасположенность говорит о том, что если человек не будет проводить профилактические мероприятия и будет вести неправильный образом жизни, то у него есть большая вероятность стать диабетиком в течение нескольких лет (рис.3).

Рис.3. Риск возникновения наследственной предрасположенности сахарного диабета у детей.

Однако и в этом случае необходимо учитывать, что если от диабета страдают сразу оба родителя, то вероятность возникновения его у их ребенка значительно повышается. И нередко именно в таких ситуациях это заболевание диагностируется у детей еще в школьном возрасте, хотя они еще

не имеют вредных привычек и ведут активный образ жизни.

Вирусные заболевания – еще одна причина, по которой может развиваться СД1. Особенно опасными в этом случае являются такие болезни, как паротит и краснуха. Учеными уже давно было доказано, что эти заболевания негативно сказываются на работе поджелудочной железы и приводят к повреждению ее клеток, снижая, таким образом, уровень инсулина в крови.

Следует отметить, что в зоне риска не только уже рожденные дети, но и те, которые еще находятся в утробе матери. Любые вирусные заболевания, которые переносит беременная женщина, могут спровоцировать развитие у ее ребенка сахарного диабета 1 типа.

Многие люди, работающие на заводах и предприятиях, где используются химические вещества, также подвержены большому риску. Действие химикатов негативно сказывается на работе всего организма, в том числе и на функциональности поджелудочной железы.

Химиотерапии, которые проводятся для лечения различных онкологических заболеваний, также оказывают токсичное действие на клетки организма, поэтому их проведение тоже в несколько раз увеличивает вероятность развития СД1 у человека [9, с.78].

Неправильное питание является одним из самых распространенных причин развития сахарного диабета 1 типа. Ежедневный рацион современного человека содержит в себе огромное количество жиров и углеводов, что оказывает сильную нагрузку на пищеварительную систему, в том числе и на поджелудочную. Со временем ее клетки повреждаются, и синтез инсулина нарушается. Необходимо также отметить, что из-за неправильно питания СД1 может развиваться и у детей в возрасте 1-2 лет.

Симптомы обоих типов сахарного диабета проявляются тогда, когда уровень глюкозы в крови превышает 6,0 ммоль/л и тогда, когда глюкоза попадает в мочу, где ее в норме быть не должно.

Для сахарного диабета 1 типа характерно быстрое развитие симптомов

(в течение нескольких недель, месяцев).

Сахарный диабет 2 типа может длительное время себя не проявлять [12, с.5]. Его можно выявить случайно, при очередном медицинском осмотре или сдаче анализов по поводу другого заболевания или при медицинском осмотре. Симптомы могут развиваться годами, быть мало выраженными.

Основные симптомы являются:

1. Полиурия – повышенное мочеотделение. По мере увеличения уровня глюкозы в крови, повышается ее содержание и в моче. Почки реагируют первыми и начинают выделять больше жидкости из организма, для того чтобы разбавить концентрацию глюкозы в моче. Характерно усиление мочеотделения в ночное время. В результате этого происходит усиленное мочевыделение (до 2 литров мочи в сутки).
2. Полидипсия – это неутолимая жажда, сухость во рту, что является следствием выделения большого количества жидкости из организма в виде мочи. Больные начинают много пить, чтобы утолить жажду и восполнить потерю жидкости с мочой [4, с.77].
3. Полифагия – это постоянное чувство голода. Это связано с нарушением обмена веществ, точнее сказать из – за того, что клетки не способны поглощать и перерабатывать глюкозу без помощи инсулина. Резкое снижение веса, похудение особенно характерны для больных с диабетом 1 типа. Это связано с повышенным разрушением белков и жиров из-за отсутствия глюкозы в энергетическом обмене клеток. Парадоксально то, что похудение развивается, несмотря на повышенный аппетит больного.

Симптомы, которые могут сопровождать диабет, но не являются обязательными:

* сухость кожных покровов;
* сухость во рту;
* мышечная слабость;
* гнойничковые образования и заболевания кожи;
* длительное и плохое заживление ран и рубцов;
* быстрая утомляемость;
* частые головные боли;
* зуд половых органов (чаще после мочеиспускания).

Выделяются три степени тяжести заболевания:

1. степень (легкая) – при этом повышение глюкозы в крови не превышает 8 ммоль/л натощак. При этом нет значительных колебаний глюкозы в течение суток, допустимы следы глюкозы в моче до 20 г/л. Возможны начальные проявления осложнений (ангионейропатии).
2. степень (средняя) – уровень глюкозы в крови натощак достигает 14ммоль/л. Глюкоза в моче увеличивается до 40 г/л. Компенсация состояния достигается диетой и приемом сахароснижающих лекарственных препаратов. Возможны проявления осложнений [2, с.35].
3. степень (тяжелое течение) – уровень глюкозы натощак более 14 ммоль/л, в моче уровень глюкозы превышает 50 г/г. В этой стадии больные нуждаются в постоянной инсулинотерапии и ярко выражены сопутствующие осложнения.

Рассмотрим разновидности фаз компенсации сахарного диабета:

* Фаза компенсации, при которой больной чувствует себя хорошо, а с помощью терапии легко можно добиться нормальных цифр глюкозы в крови. В моче глюкозы не содержится.
* Фаза субкомпенсации. При данной фазе не удается снизить уровень глюкозы в крови ниже 13,9 ммоль/л. Появляется сахар в моче. Ацетона в моче не содержится.
* Фаза декомпенсации (самая тяжелая) – проводимая терапия не дает эффекта, и уровень сахара поднимается выше 14,0 ммоль/л. Количество глюкозы в моче увеличивается и появляется ацетон. Возможно развитие гипергликемической комы.

## 1.2 Корреляционно-регрессионный анализ

Корреляционно-регрессионный анализ является наиболее широко распространенным и гибким приемом обработки статистической информации. [11]

Корреляционно-регрессионный анализ — это один из самых распространенных методов изучения отношений между численными величинами. Его основная цель состоит в нахождении зависимости между двумя параметрами и ее степени с последующим выведением   
уравнения [1.12]. То есть, корреляционно-регрессионный анализ представляет из себя объединение методов корреляционного и регрессионного анализов.

Задачами корреляционно-регрессионного анализа являются:

* + установление типа уравнения регрессии;
  + определение параметров уравнения регрессии и оценка значимости параметров;
  + оценка тесноты и направления связи между переменными; - оценка значимости уравнения регрессии;
  + определение прогнозных значений зависимой переменной и оценка полученного прогноза. [1.13]

Так как в корреляционно-регрессионном анализе используются методы корреляционного и регрессионного анализа, рассмотрим эти методы подробней.

Корреляционный анализ — раздел математической статистики, в котором изучаются задачи выявления статистических зависимостей между случайными величинами путем оценок различных коэффициентов корреляции. Методы корреляционного анализа дают хорошие результаты тогда, когда данные эксперимента можно считать выбранными из генеральной совокупности, распределенной по многомерному нормальному закону. [1.14]

Невозможно управлять явлениями, предсказывать их развитие без изучения характера, силы и других особенностей связей. Поэтому методы исследования, изменения связей составляют чрезвычайно важную часть методологии научного исследования, в том числе и статистическую.

Связи между изучаемыми переменными подразделяются на функциональные и статистические. При функциональной связи определенному значению одной переменной величины соответствует строго определенное значение другой переменной.

При изменении одной из них на определенную величину, другая переменная изменяется на величину, в соответствии с видом функции, связывающей переменные.

Статистической называется связь между переменными или признаками, когда определенному значению факторного признака соответствует несколько различных значений результативного признака. Частным случаем статистической связи является корреляционная, которая проявляется в среднем, в массе наблюдений, как статистическая закономерность.

При корреляционной связи с изменением факторного признака на определенную величину изменяется среднее значение результативного признака. Обычно корреляционная зависимость представляется как функциональная зависимость между переменными в виде уравнения регрессии. [1.13]

Корреляционной связью называют важнейший частный случай статистической связи, состоящий в том, что разным значениям одной переменной соответствуют различные средние значения другой. С изменением значения признака х закономерным образом изменяется среднее значение признака у; в то время как в каждом отдельном случае значение признака y (с различными вероятностями) может принимать множество различных значений. [1.15]

Тесноту связи изучаемых явлений оценивает Коэффициент   
Пирсона (Kn).

Коэффициент Пирсона используется для изучения связи между двумя качественными признаками, каждый из которых состоит более чем из двух групп. Вычисляют по формуле:

,

где — показатель взаимной сопряженности:

,

где — объемы признака X по группам;

— объемы признака Y по группам;

— объемы выборок, относящихся к X и Y одновременно. [1.16]

Корреляционный коэффициент Пирсона может принимать значения в диапазоне .

По значению эмпирического корреляционного отношения судят о тесноте связи между признаками. Обычно придерживаются следующей шкалы:

— связь слабая;

— связь заметная;

— связь умеренно тесная;

— связь тесная;

— связь очень тесная. [1.17]

После того как с помощью корреляционного анализа выявлено наличие статистических связей между переменными и оценена степень их тесноты, обычно переходят к математическому описанию зависимостей, то есть к регрессионному анализу.

Регрессионный анализ применяется в тех случаях, когда необходимо отыскать непосредственно вид зависимости х и у. При этом предполагается, что независимые факторы являются не случайными величинами, а результативный показатель у имеет постоянную, независимую от факторов дисперсию и стандартное отклонение. [1.18]

Рассмотрим метод линейной регрессии.

Под линейностью имеется в виду, что переменная у предположительно находится под влиянием переменной х в зависимости

,

где b — постоянная величина (или свободный член уравнения);

a — коэффициент регрессии, определяющий наклон линии, вдоль которой рассеяны данные наблюдения. Это показатель, характеризующий изменение переменной yi при изменении значения xi на единицу. Если a > 0, переменные xi и yi положительно коррелированные, если a < 0 — отрицательно коррелированные;

εi — независимая нормально распределенная величина — остаток с нулевым математическим ожиданием и постоянной дисперсией:

,

где σ — среднее квадратическое отклонение. Отражает тот факт, что изменение yi будет неточно описываться изменением xi: присутствуют другие факты, не учтенные в данной модели. [1.19]

Построение линейной регрессии сводится к оценке параметров a и b. Классический подход к оцениванию параметров линейной регрессии основан на методе наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров a и b, при которых сумма квадратов отклонений фактических значений результативного признака y от расчетных (теоретических) , минимальна:

,

где yi — фактические значения результативного признака y;

— расчетные значения результативного признака y.

То есть из всего множества линий линия регрессии на графике выбирается так, чтобы сумма квадратов расстояний по вертикали между точками и этой линией была бы минимальной. [1.20]

Таким образом, рассмотрено понятие корреляционно-регрессионного анализа, а также методы корреляционно-регрессионного анализа: нахождение корреляционной связи с помощью коэффициента Пирсона; построение парной линейной регрессии с помощью метода наименьших квадратов (МНК).

Рассмотрим применение корреляционно-регрессионного анализа на примере отрасли добычи полезных ископаемых в США.

# 2 ПРАКТИЧЕСКАЯ ЧАСТЬ

## 2.1 Корреляционно-регрессионный анализ на примере отрасли добычи полезных ископаемых в США.

В практической работе проведём корреляционно-регрессионный анализ зависимости количества работников в отрасли от выпуска валовой продукции в этой отрасли на примере отрасли добычи полезных ископаемых и отрасли гостиничного хозяйства и общественного питания в США.

В качестве исследуемых данных подаётся на вход подаётся наборы данных «Микроэкономические показатели по отраслям в США» и «Количество занятых по отраслям в США».

Набор данных «Микроэкономические показатели по отраслям   
в США» включает в себя следующие столбцы данных:

* 1. Industry — Отрасль.
  2. Year — Год.
  3. Gross Output — Валовой выпуск продукции в миллиардах долларов.
  4. Intermediate Inputs — Промежуточные затраты в миллиардах долларов.
  5. Value Added — Добавленная стоимость в миллиардах долларов.

Набор данных «Количество занятых по отраслям в США» включает в себя следующие столбцы данных:

* 1. Industry — Отрасль.
  2. Year — Год.
  3. Employees — Количество работников в тысячах.

Для обработки данных и проведения корреляционно-регрессионного анализа используем аналитическую low-code платформу Loginom. Импортируем наши наборы данных в сценарий пакета. Для этого в рабочую область сценария добавляем два узла «Excel файл» и для каждого в настройках указываем путь к нужному набору данных.

На Рисунке 2.1 указана визуализация в виде OLAP-куба набор данных «Микроэкономические показатели по отраслям в США».

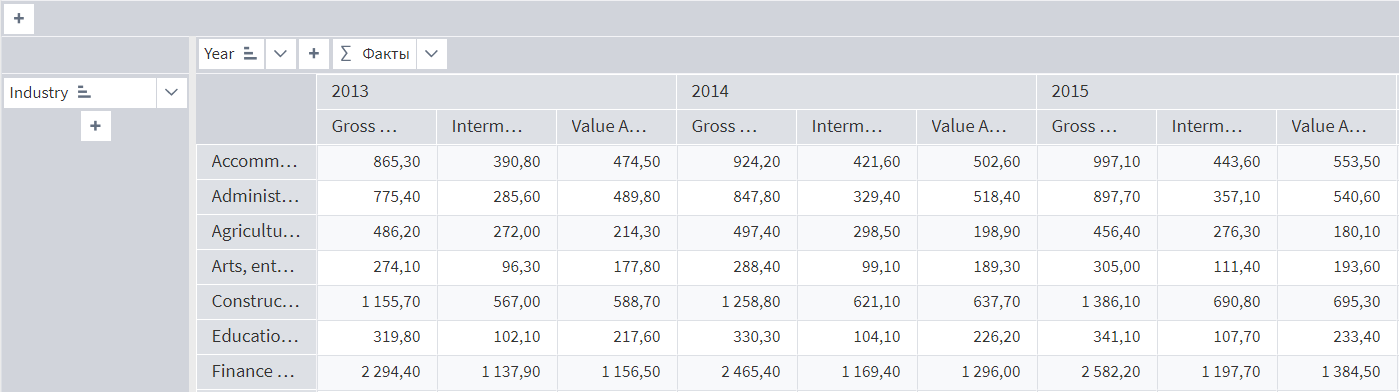


Рисунок 2.1 — Фрагмент визуализации в виде OLAP-куба набора данных «Микроэкономические показатели по отраслям в США»

На Рисунке 2.2 указана визуализация в виде OLAP-куба наборов данных «Количество занятых по отраслям в США».

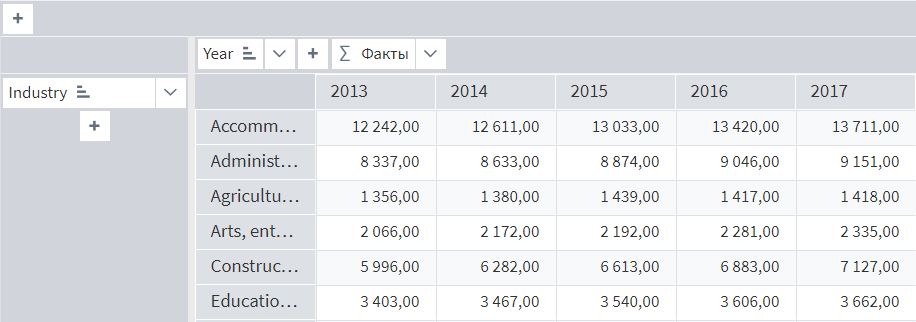


Рисунок 2.2 — Фрагмент визуализации в виде OLAP-куба набора данных «Количество занятых по отраслям в США»

Работать с двумя наборами данных, у каждого из которых имеются столбцы «Отрасль» и «Год», неудобно. Соединим их в один набор данных с помощью узла «Слияние». Визуализация результата в виде OLAP-куба представлена на Рисунке 2.3.

Мы объединили два исходных набора данных и получили новый набор данных, визуализация которого представлена на Рисунке 2.3

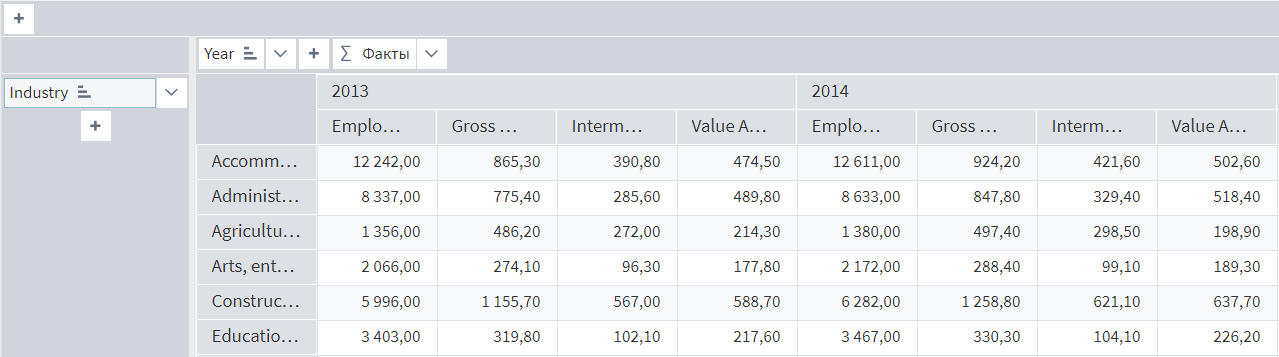


Рисунок 2.3 — Фрагмент визуализации в виде OLAP-куба нового набора данных

Для проведения корреляционно-регрессионного анализа какой-либо отрасли требуется отфильтровать набор данных, выделив из него нужную нам отрасль. Нам требуется выделить отрасль «Mining» — добыча полезных ископаемых. Воспользуемся узлом «Фильтр строк». Визуализация результата в виде таблицы представлена на Рисунке 2.4.

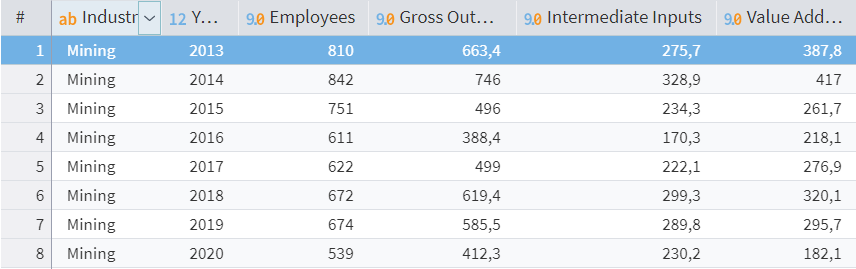


Рисунок 2.4 — Визуализация в виде таблицы набора данных с отфильтрованной отраслью «Добыча полезных ископаемых»

На Рисунке 2.4 показан отфильтрованный набор данных, над которым предстоит провести корреляционно-регрессионный анализ.

Покажем зависимость показателя «Сотрудники» от показателей «Валовой выпуск продукции», «Промежуточные затраты» и «Добавленная стоимость».

На Рисунке 2.5 показана зависимости каждого значения показателя «Сотрудники» от показателя «Валовой выпуск продукции».

По оси y — значения «Сотрудники», по оси x — значения «Валовой выпуск продукции».

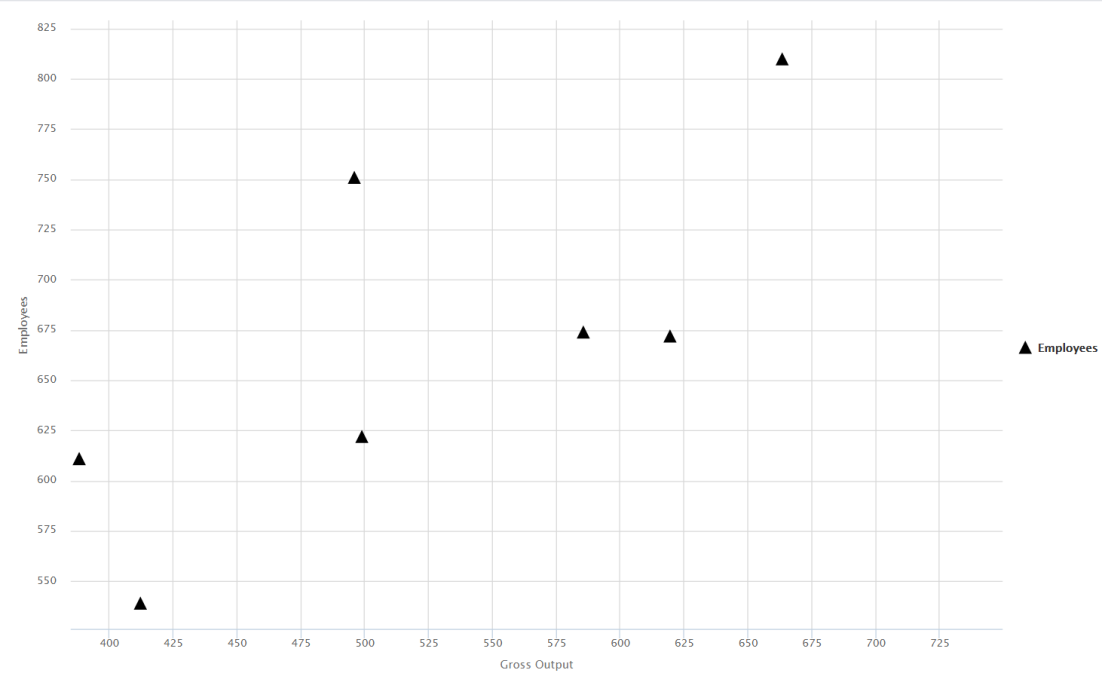


Рисунок 2.5 — Диаграмма зависимости показателя «Сотрудники» от показателя «Валовой выпуск продукции» данных с отфильтрованной строкой «Добыча полезных ископаемых»

На Рисунке 2.6 показаны зависимости каждого значения показателя «Сотрудники» от показателя «Промежуточные затраты».

По оси y — значения «Сотрудники», по оси x — значения «Промежуточные затраты».

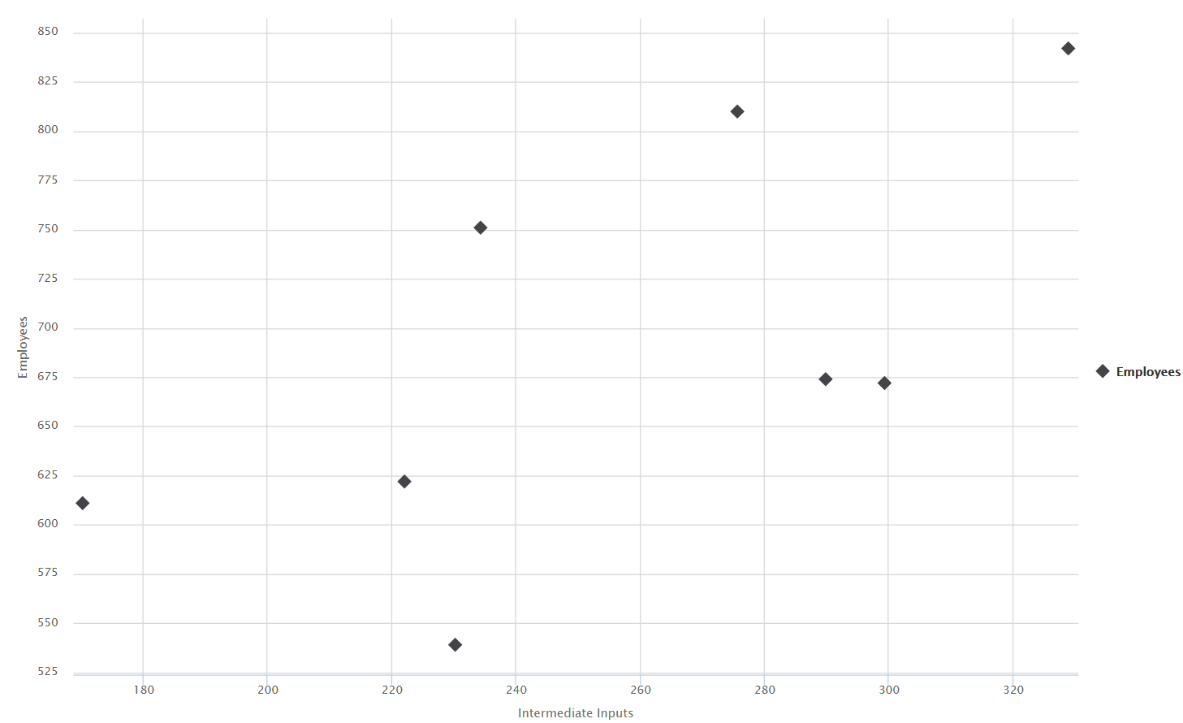


Рисунок 2.6 — Диаграмма зависимости показателя «Сотрудники» от показателя «Промежуточные затраты» данных с отфильтрованной строкой «Добыча полезных ископаемых»

На Рисунке 2.7 показаны зависимости каждого значения показателя «Сотрудники» от показателя «Добавленная стоимость».

По оси y — значения «Сотрудники», по оси x — значения «Добавленная стоимость».

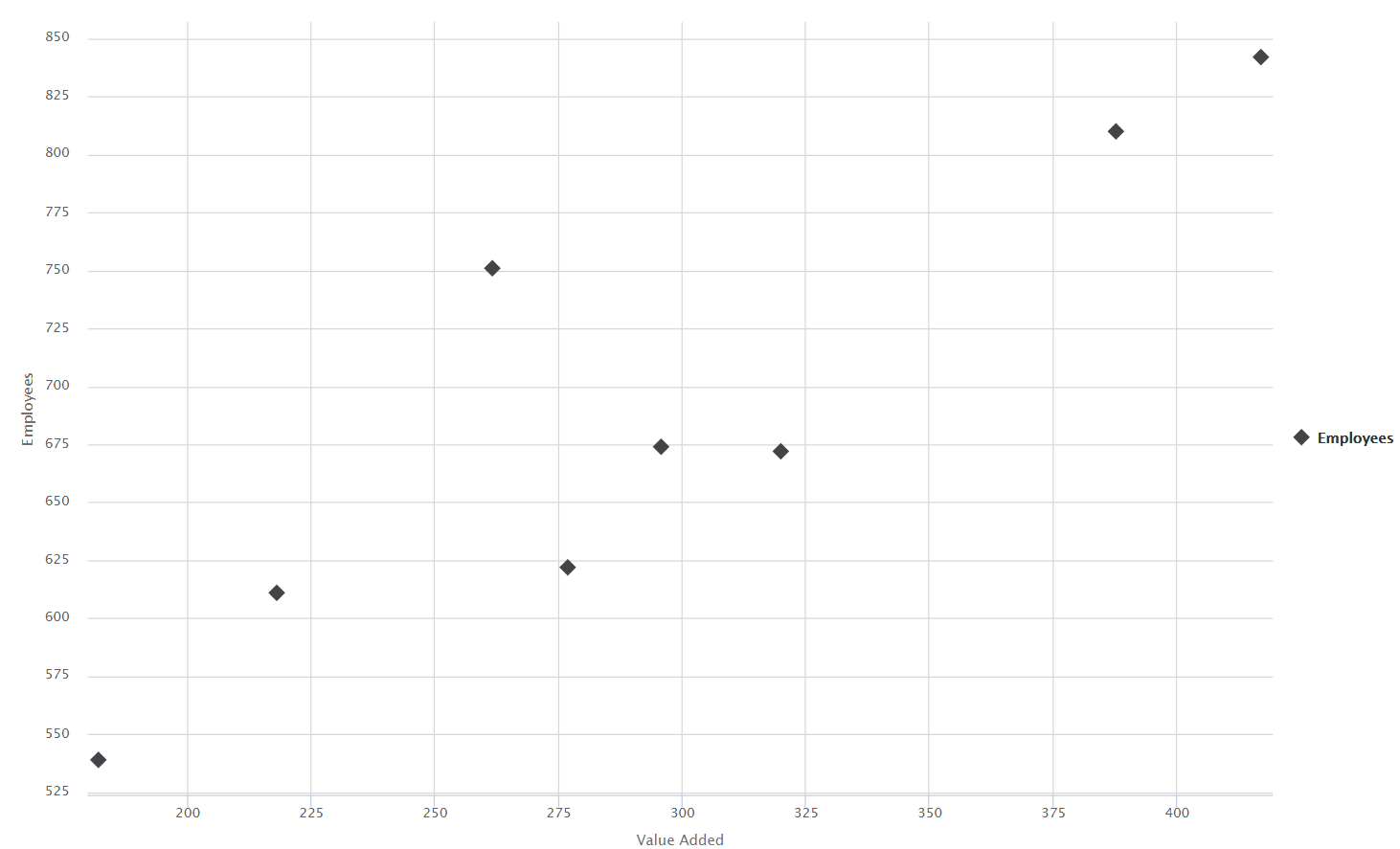


Рисунок 2.7 — Диаграмма зависимости показателя «Сотрудники» от показателя «Добавленная стоимость» данных с отфильтрованной строкой «Добыча полезных ископаемых»

Для проведения корреляционно-регрессионного анализа необходимо разделить показатели на независимые (x) и зависимые (y) переменные. В данном случае зависимой переменной будет показатель «Сотрудники», а независимыми — показатели «Валовой выпуск продукции», «Промежуточные затраты» и «Добавленная стоимость». Столбцы «Отрасль» и «Год» в корреляционно-регрессионном анализе не участвуют.

Сначала нужно найти значение коэффициента Пирсона и по нему определить тесноту связи между зависимой и независимыми переменными. Для вычисления коэффициента Пирсона используем узел «Корреляционный анализ». Результат представлен на Рисунке 2.8.

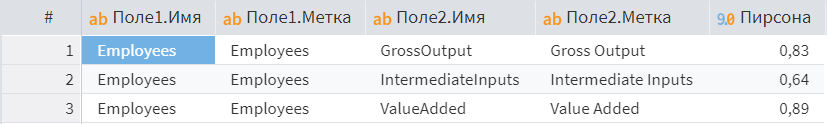


Рисунок 2.8 — Результат вычисления коэффициента Пирсона

Из Рисунка 2.8 мы видим, что коэффициент корреляции Пирсона между показателем «Сотрудники» и показателем «Валовой выпуск продукции» равен 0,83; между показателем «Сотрудники» и показателем «Добавленная стоимость» — 0,89. Это означает, что связь между показателем «Сотрудники» и этими двумя показателями является тесной.

Коэффициент корреляции Пирсона же между показателем «Сотрудники» и показателем «Промежуточные затраты» равен 0,64. Это означает, что связь между показателем «Сотрудники» и показателем «Промежуточные затраты» является умеренно тесной.

Для проверки данных утверждений построим графики зависимости показателя «Сотрудники» от показателей «Валовой выпуск продукции», «Промежуточные затраты» и «Добавленная стоимость».

Графики изображены на Рисунке 2.9, Рисунке 2.10 и Рисунке 2.11.

По оси y — значения «Сотрудники», по оси x — значения Валовой выпуск продукции», «Промежуточные затраты» и «Добавленная стоимость».

На Рисунке 2.9 показано, что график зависимости показателя «Сотрудники» от показателя «Валовый выпуск продукции» имеет небольшой разброс значений показателя «Сотрудники», но с увеличением валового выпуска продукции увеличивается и количество сотрудников в отрасли.

Это означает, что показатель «Валовый выпуск продукции» имеет влияние на показатель «Сотрудники», и это соответствует значению коэффициента Пирсона, вычисленного между показателем «Сотрудники» и показателем «Валовой выпуск продукции». Коэффициент равен 0,83.

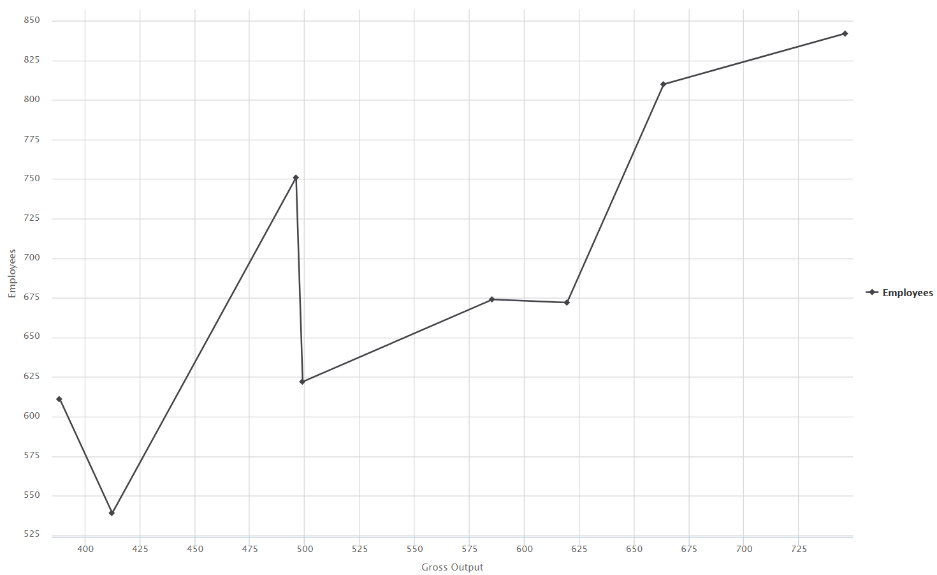


Рисунок 2.9 — График зависимости показателя «Сотрудники» от показателя «Валовой выпуск продукции»

На Рисунке 2.10 показан график зависимости показателя «Сотрудники» от показателя «Добавленная стоимость».

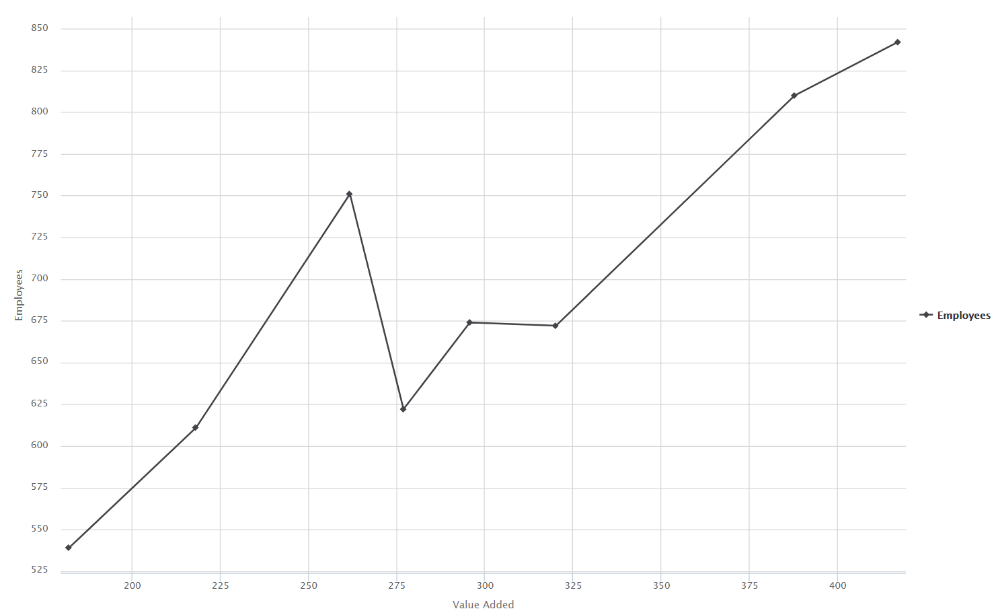


Рисунок 2.10 — График зависимости показателя «Сотрудники» от показателя «Добавленная стоимость»

На Рисунке 2.10 показано, что график зависимости показателя «Сотрудники» от показателя «Валовый выпуск продукции» на большей части значений принимает близкий к линейному вид. Лишь при значениях показателя «Добавленная стоимость» в промежутке от 255 до 300 график имеет небольшой разброс значений показателя «Сотрудники», что значительно уменьшило значение коэффициента Пирсона, вычисленного между показателем «Сотрудники» и показателем «Добавленная стоимость».

Это означает, что показатель «Добавленная стоимость» имеет сильное влияние на показатель «Сотрудники», и это соответствует значению коэффициента Пирсона, вычисленного между показателем «Сотрудники» и показателем «Добавленная стоимость». Коэффициент равен 0,89

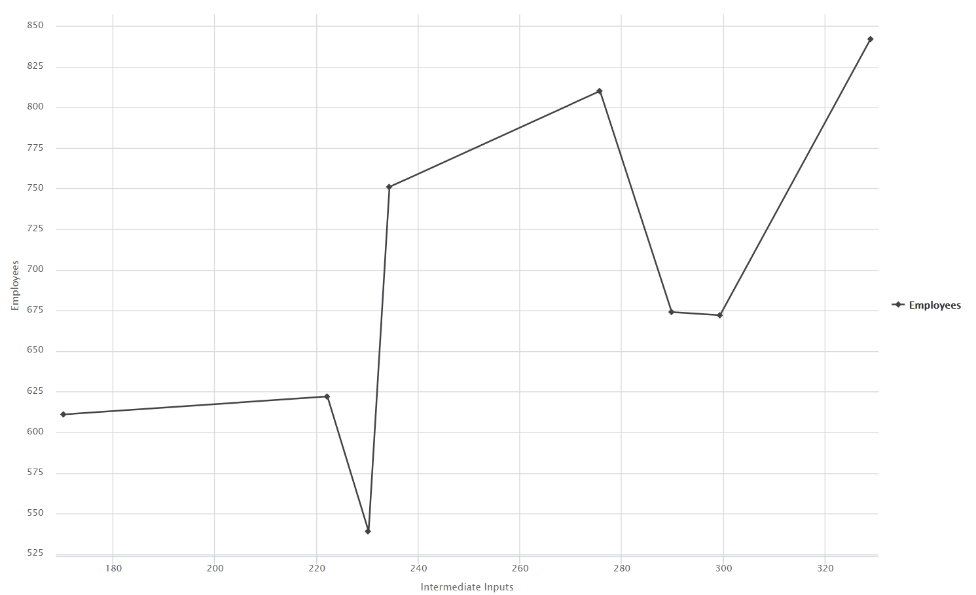


Рисунок 2.11 — График зависимости показателя «Сотрудники» от показателя «Промежуточные затраты»

На Рисунке 2.11 показано, что график зависимости показателя «Сотрудники» от показателя «Валовый выпуск продукции» имеет довольно большой разброс значений показателя «Сотрудники», но с увеличением валового выпуска продукции увеличивается и количество сотрудников в отрасли.

Это означает, что показатель «Промежуточная стоимость» имеет влияние на показатель «Сотрудники», и это соответствует значению коэффициента Пирсона, вычисленного между показателем «Сотрудники» и показателем «Промежуточная стоимость». Коэффициент равен 0,64.

Значение данного коэффициента является наименьшим среди вычисленных, что заметно при сравнении трёх представленных на   
Рисунке 2.9, Рисунке 2.10 и Рисунке 2.11 графиков, где график зависимости показателя «Сотрудники» от показателя «Промежуточные затраты» имеет наибольший разброс значений показателя «Сотрудники».

Посмотрим на сравнение значений коэффициента Пирсона всех независимых переменных. Визуализация представлена на Рисунке 2.12.

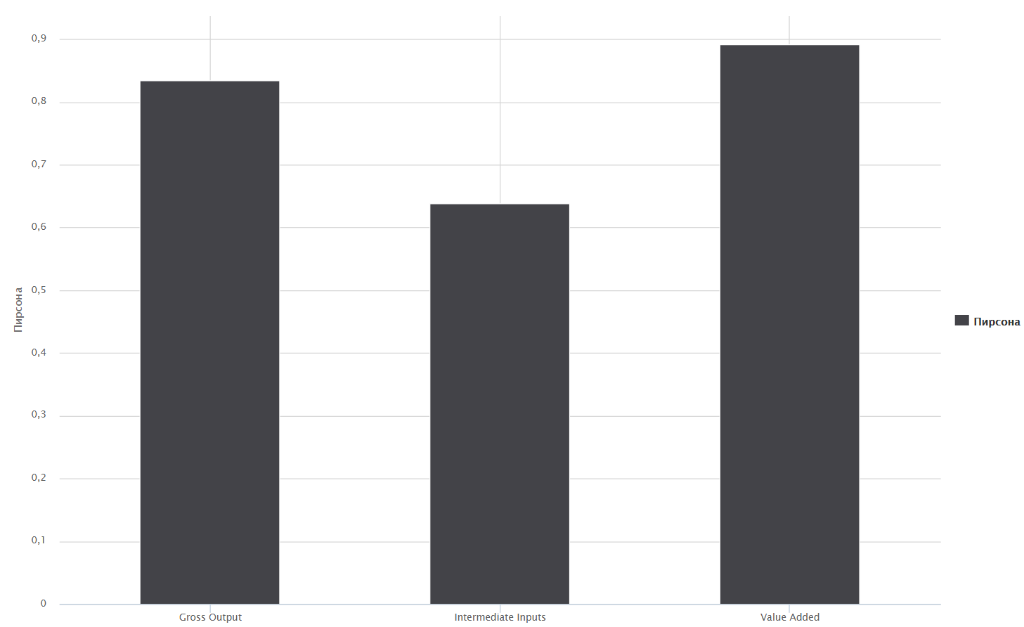


Рисунок 2.12 — Сравнение коэффициентов Пирсона

На Рисунке 2.12 показано, что хоть коэффициенты Пирсона разнятся между собой, они всё еще показывают довольно высокие значения. Из этого можно сделать вывод о том, что микроэкономические показатели по отраслям оказывают влияние на изменение количества работников по отраслям. Это ещё раз доказывает корректность сделанных выводов.

Теперь, когда известна степень тесноты связи между показателем «Сотрудники» и остальными показателями, можно проводить поиск значений коэффициентов уравнений регрессии. В нашем случае уравнения имеют линейный вид: .

То есть нам требуется найти значения неизвестных a и b. y— зависимая переменная, то есть показатели «Сотрудники», а x — независимая переменная, то есть показатели «Валовой выпуск продукции», «Промежуточные затраты» и «Добавленная стоимость».

Для поиска значений уравнения воспользуемся узлом «Линейная регрессия». Результаты выполнения узла показаны на Рисунке 2.13, Рисунке 2.14 и Рисунке 2.15.

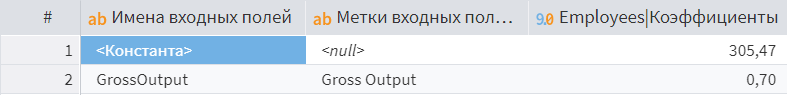


Рисунок 2.13 — Результат вычисления коэффициентов уравнения линейной регрессии, где x — показатель «Валовой выпуск продукции»

На рисунке 2.13 показано, что неизвестные уравнения   
 (показатель «Сотрудники|Коэффициенты» в первой строке), а (показатель «Сотрудники|Коэффициенты» во второй строке). То есть, уравнение принимает вид .

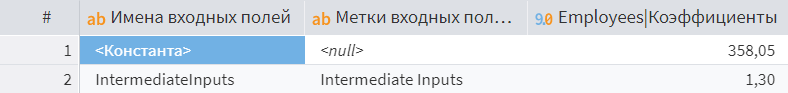


Рисунок 2.14 — Результат вычисления коэффициентов уравнения линейной регрессии, где x — показатель «Промежуточные затраты»

На рисунке 2.14 видно, что неизвестны уравнения   
 (показатель «Сотрудники|Коэффициенты» в первой строке), а (показатель «Сотрудники|Коэффициенты» во второй строке). То есть, уравнение принимает вид.

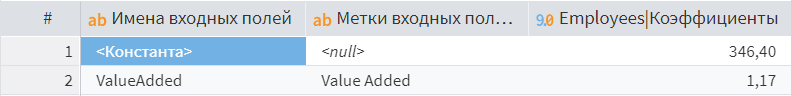


Рисунок 2.15 — Результат вычисления коэффициентов уравнения линейной регрессии, где x — показатель «Добавленная стоимость»

На рисунке 2.15 показано, что неизвестны уравнения (показатель «Сотрудники|Коэффициенты» в первой строке), а (показатель «Сотрудники|Коэффициенты» во второй строке). То есть, уравнение принимает вид .

С помощью найденных коэффициентов можно построить линии регрессии для каждого независимого показателя. Узел «Линейная регрессия» автоматически вычисляет значения y в зависимости от x для найденного уравнения. Результаты вычисления значений y представлены на Рисунке 2.16, Рисунке 2.17 и Рисунке 2.18.

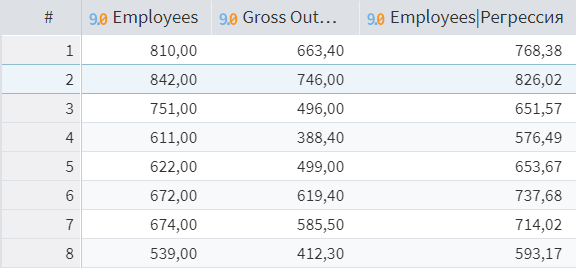


Рисунок 2.16 — Результат вычисления значений y в зависимости от значений x уравнения линейной регрессии, где x— показатель «Валовой выпуск продукции»

Из Рисунка 2.16 мы можем увидеть, какие значения принимает уравнение линейной регрессии (показатель «Сотрудники|Регрессия») в зависимости от x (показатель «Валовой выпуск продукции»).

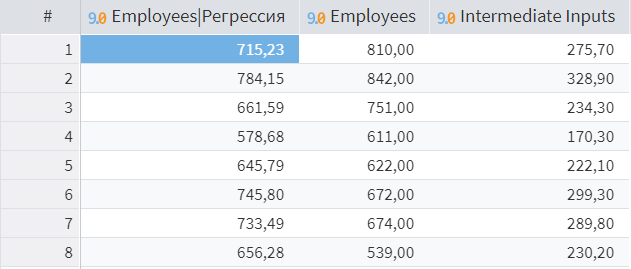


Рисунок 2.17 — Результат вычисления значений y в зависимости от значений x уравнения линейной регрессии, где x — показатель «Промежуточные затраты»

Из Рисунка 2.17 мы можем увидеть, какие значения принимает уравнение линейной регрессии (показатель «Сотрудники|Регрессия») в зависимости от x (показатель «Промежуточные затраты»).

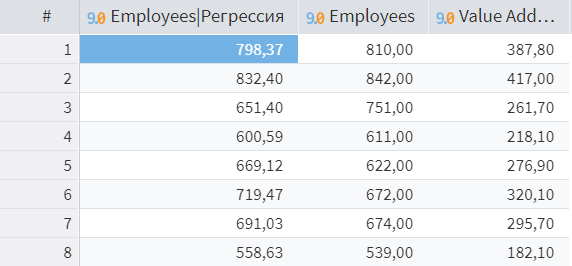


Рисунок 2.18 — Результат вычисления значений y в зависимости от значений x уравнения линейной регрессии, где x — показатель «Добавленная стоимость»

Из Рисунка 2.18 мы можем увидеть, какие значения принимает уравнение линейной регрессии (показатель «Сотрудники|Регрессия») в зависимости от x (показатель «Добавленная стоимость»).

Так как мы знаем значения уравнений в зависимости от x, можно построить линии регрессии с помощью визуализатора «Диаграмма» узла «Линейная регрессия», где осью X будут показатели «Валовой выпуск продукции», «Промежуточные затраты» и «Добавленная стоимость» а осью Y — показатель «Сотрудники». Результат построения линий представлен на Рисунке 2.19, Рисунке 2.20 и Рисунке 2.21.

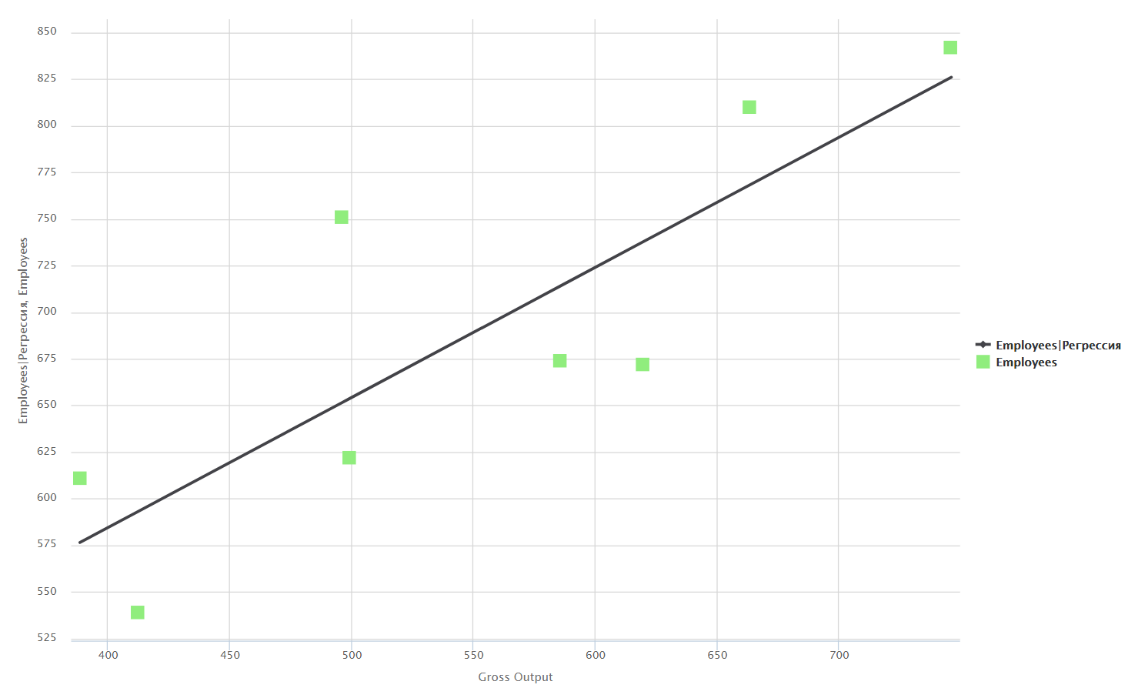


Рисунок 2.19 — Диаграмма построения линии регрессии с помощью уравнения линейной регрессии, где x— показатель «Валовой выпуск продукции»

На Рисунке 2.20 показана диаграмма построения линии регрессии с помощью уравнения линейной регрессии, где x — показатель «Промежуточные затраты» а осью Y — показатель «Сотрудники».

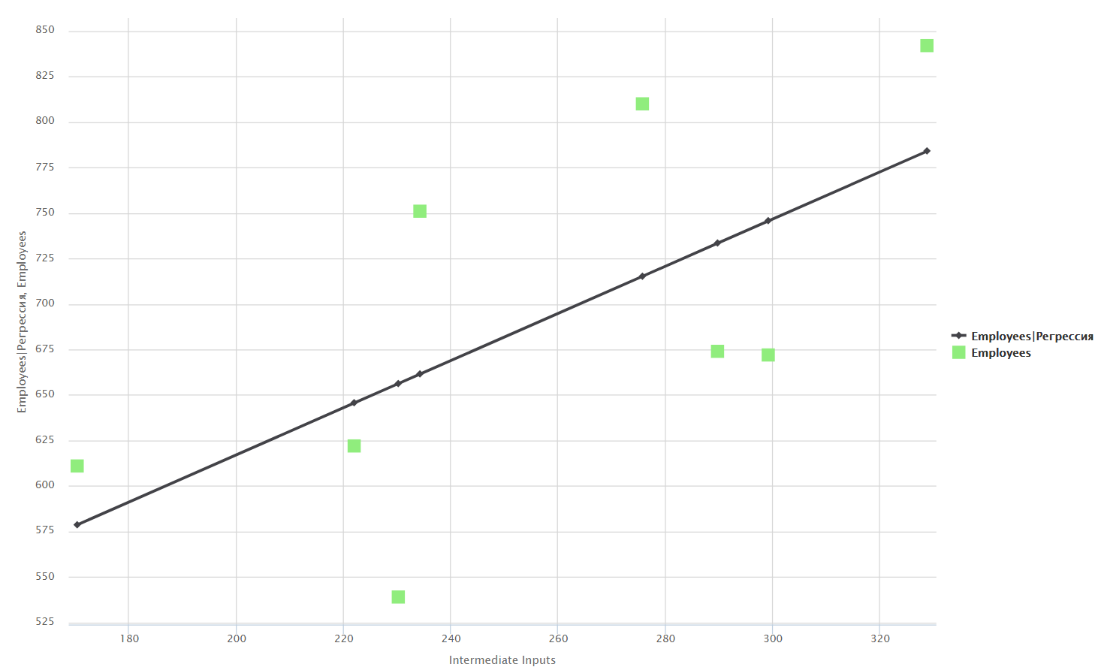


Рисунок 2.20 — Диаграмма построения линии регрессии с помощью уравнения линейной регрессии, где x — показатель «Промежуточные затраты»

На Рисунке 2.21 показана диаграмма построения линии регрессии с помощью уравнения линейной регрессии, где x — показатель «Добавленная стоимость»

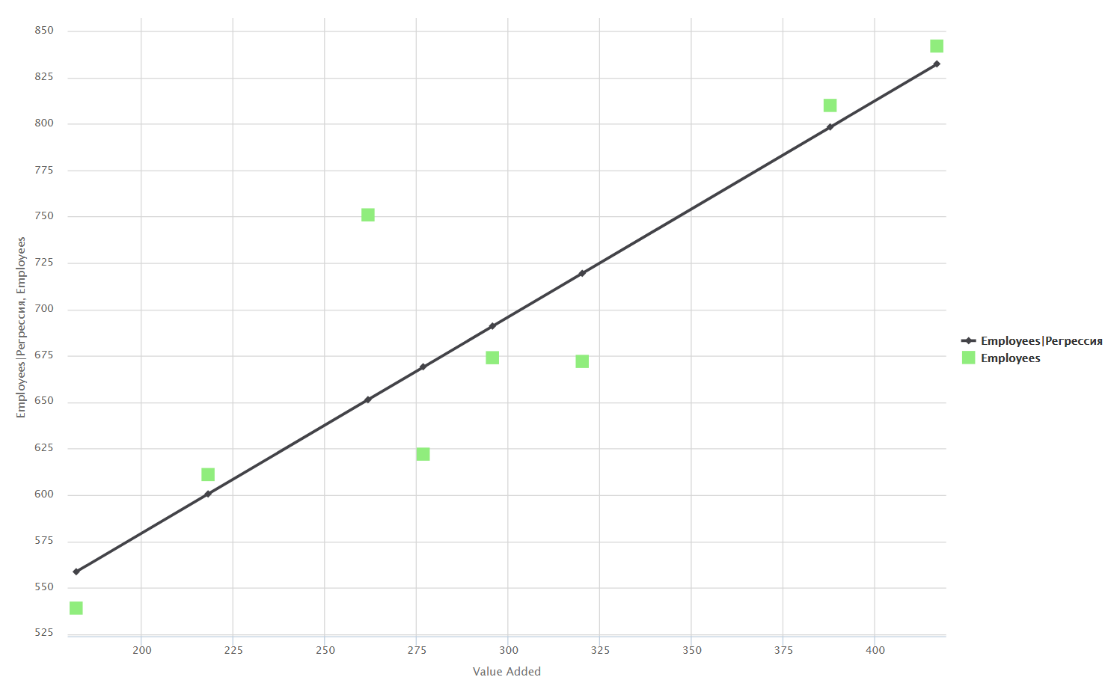


Рисунок 2.21 — Диаграмма построения линии регрессии с помощью уравнения линейной регрессии, где x — показатель «Добавленная стоимость»

На Рисунке 2.19, Рисунке 2.20 и Рисунке 2.21 видно, что линии регрессии проходят сквозь поля скопления точек, обозначающих зависимость количества работников в отрасли от соответствующего микроэкономического показателя валовой выпуск продукции, промежуточные затраты и добавленная стоимость в отрасли по годам.

Таким образом, проведя корреляционно-регрессионный анализ влияния микроэкономических показателей валовой выпуск продукции, промежуточные затраты и добавленная стоимость в отрасли на количество работников в этой отрасли на примере отрасли добычи полезных ископаемых, мы выяснили, что микроэкономические показатели валовой выпуск продукции, промежуточные затраты и добавленная стоимость и количество работников в отрасли тесно друг с другом связаны.

Количество работников растет вместе с ростом любого из микроэкономических показателей валовой выпуск продукции, промежуточные затраты и добавленная стоимость, что означает повышение востребованности работников в данной отрасли. Мы вывели уравнения регрессии и построили линии регрессии, с помощью которых есть возможность предсказания следующих значений количества работников в зависимости от микроэкономических показателей. Что даёт нам возможность предпринять различные действия, для предотвращения или большего улучшения ситуации в будущем, которое мы предсказали с помощью уравнений регрессии и видимыми диаграммами линий регрессии на Рисунке 2.19, Рисунке 2.20 и Рисунке 2.21 что будет сказываться в лучшую сторону для страны, которая будет преждевременно готова к различным ситуациям.

Можно сделать вывод о том, что благодаря всем показателям из данной курсовой работы можно подтвердить что основные микроэкономические показатели линейно влияют на структурные изменения рынка труда.

# ЗАКЛЮЧЕНИЕ

Исследование состояния рынка труда, как одного из важнейших элементов экономической системы любого государства и анализ основных данных о структурных изменениях показателей рынка труда является важнейшей задачей для любого предприятия.

Проверенные и проанализированные данные о связи экономических показателей с показателями рынка труда позволят любому предприятию предсказать возможные изменения, связанные с рынком труда и экономическими показателями в будущем.

Отсюда возникает потребность в анализе влияния экономических показателей на структурные изменения рынка труда, для чего проводится корреляционно-регрессионный анализ данного влияния.

Цель данной курсовой работы — провести корреляционно-регрессионный анализ влияния основных микроэкономических показателей на структурные изменения рынка труда в отрасли — достигнута.

В ходе выполнения данной курсовой работы проведен корреляционно-регрессионный анализ влияния выпуска валовой продукции в отрасли на количество работников в отрасли с использованием low-code платформы Loginom.

Задачи, выполненные в данной курсовой работе:

* + изучены научная и методическая литературы о рынке труда и корреляционно-регрессионном анализе;
  + выполнен корреляционно-регрессионный анализ собранных данных;
  + использованы знания математической статистики с использованием современных средств обработки данных: аналитической платформы Loginom;
  + пройдено обучение оформлению официальных документов.

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

## Теоретическая часть

* 1. Поварич, И.П. Рынок труда: учебное пособие / И.П. Поварич, М.Д. Поварич, Т.А. Кузнецова. — Кемерово: КемГУ, 2019. — 291 с.
  2. Рофе, А.И. Рынок труда: учебник / Рофе А.И. — Москва: КноРус, 2018. — 259 с.
  3. Ермолаева, С.Г. Рынок труда: учебное пособие / С.Г. Ермолаева. — Екатеринбург: Изд-во Урал. ун-та, 2016. — 108 с.
  4. Корнейчук, Б.В. Рынок труда: учебник для академического бакалавриата / Б.В. Корнейчук. — Москва: Издательство Юрайт, 2018. — 263 с.
  5. Дубовик, А.К. Рынок труда: учебно-методическое пособие для студентов специальности 1-26 02 02 «Менеджмент», направления специальности 1-26 02 02-02 «Менеджмент (социально-административный)» / А.К. Дубовик. — Минск: БНТУ, 2017. — 109 с.
  6. Особенности и характеристики внешнего рынка труда / Проект HR-Portal [Электронный ресурс]. https://hr-portal.ru/article/osobennosti-i-harakteristiki-vneshnego-rynka-truda
  7. Сегментация рынков труда: внутренние и профессиональные рынки / Официальный сайт кадрового агентства «КАУС» [Электронный ресурс]. https://www.kaus-group.ru/knowledge/300-articles/category/market/material/45/
  8. Васяйчева, В.А. Рынок труда и трудовая миграция: учеб. пособие / В.А. Васяйчева, Д.Г. Слатов. — Самара: Изд-во Самарского университета, 2018. — 96 с.
  9. Шигапова, Д.К. Конспект лекций по курсу «Рынок труда» / Автор-разработчик: доцент Шигапова Д.К. — 2021. — 30 с.
  10. Райгородская, В.С. Организация, нормирование и оплата труда на предприятиях отрасли: учеб. пособие / В.С. Райгородская, Д.В. Зайцев, В.А. Гусейналиев. — Москва: МАДИ, 2018. — 120 с
  11. Корреляционно-регрессионный анализ / Научный словарь-справочник «Справочник24» [Электронный ресурс] <https://spravochnick.ru/ekonomicheskiy_analiz/korrelyacionno-regressionnyy_analiz/>
  12. Корреляционно-регрессионный анализ: пример, задачи, применение. Метод корреляционно-регрессионного анализа / Интернет-портал BusinessMan.ru [Электронный ресурс]. https://businessman.ru/new-korrelyacionno-regressionnyj-analiz-primer-zadachi-primenenie.html
  13. П.С. Бондаренко, Г.В. Горелова, И.А. Кацко, А.Е. Жминько, Т.В. Соловьёва, С.А. Кацко, С.К. Куижева, А.А. Митус, Н.Б. Паклин, А.Е. Сенникова. Эконометрика. Практикум : учебно-практическое пособие / коллектив авторов ; под ред. И.А. Кацко. — Москва: КНОРУС, 2019. — 218 с.
  14. Александрова, О.В. Статистические методы решения технологических задач : учебное пособие / О.В. Александрова, Т.А. Мацеевич, Л.В. Кирьянова, В.Г. Соловьев. – Москва: Издательство МИСИ – МГСУ, 2017. — 154 с.
  15. Круценюк, К.Ю. Корреляционно-регрессионный анализ в эконометрических моделях : учебное пособие / К. Ю. Круценюк. – Норильск: НГИИ, 2018. — 108 с.
  16. Ганичева, А. В. Прикладная статистика: учебное пособие для спо / А. В. Ганичева. — Санкт-Петербург: Лань, 2021. — 164 с.
  17. Полякова, В.В. Основы теории статистики : [учеб. пособие] / В. В. Полякова, Н. В. Шаброва. – Екатеринбург: Изд-во Урал. ун-та, 2016. — 148 с.
  18. Метод корреляционно-регрессионного анализа / Студенческие реферативные статьи и материалы "Studref" [Электронный ресурс]. <https://studref.com/591347/ekonomika/metod_korrelyatsionno_regressionnogo_analiza>
  19. Кийко, П. В. Эконометрика. Регрессионные модели : учебное пособие / П. В. Кийко, Н. В. Щукина. — Омск: Омский ГАУ, 2021. — 83 с.
  20. Герасимова, Е. А. Эконометрика: регрессионный анализ : учебно-методическое пособие / Е. А. Герасимова, М. Ю. Карышев. — Самара: СамГУПС, 2016. — 98 с.

## Практическая часть

* 1. Interactive Access to Industry Economic Accounts Data / U.S. Bureau of Economic Analysis (BEA) [Электронный ресурс]. <https://apps.bea.gov/iTable/iTable.cfm?reqid=150&step=2&isuri=1&categories=gdpxind>
  2. National Income and Product Accounts / U.S. Bureau of Economic Analysis (BEA) [Электронный ресурс]. <https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=2&isuri=1&1921=survey>

# ПРИЛОЖЕНИЯ

Приложение А — Графический материал.

## Приложение А

На рисунке 3.1 представлен сценарий проекта в аналитической платформе Loginom, полученный в результате выполнения данной курсовой работы.

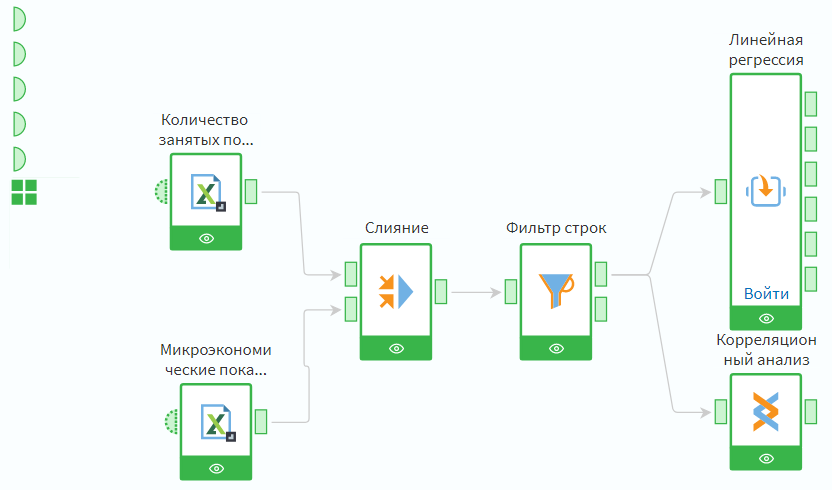


Рисунок 3.1 — Сценарий проекта в аналитической платформе Loginom

Этапы выполнения сценария:

1. Импорт исходных данных (Excel файлы).
2. Предобработка исходных данных.
3. Корреляционно-регрессионный анализ.