ииминобрнауки россии

федеральное государственное бюджетное образовательное учреждение

высшего образования

«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

|  |  |  |
| --- | --- | --- |
| Институт (факультет) | | Институт информационных технологий |
| Кафедра | Математического и программного обеспечения ЭВМ | |

КУРСОВАЯ РАБОТА

|  |  |
| --- | --- |
| по дисциплине | Прикладная статистика |

|  |  |
| --- | --- |
| на тему | Изучение долговечности и разнообразия выбора пластиковых оконных конструкции от разных производителей |

|  |
| --- |
| Выполнила студент группы |
| 1ПИб-02-3оп-22 |
| направление подготовки (специальности) |
| 09.03.04., Программная инженерия |
| шифр, наименование |
| Бурмистров Дмитрий Сергеевич |
| фамилия, имя, отчество |

|  |
| --- |
| Руководитель |
| Гонторева Ирина Борисовна |
| фамилия, имя, отчество |
| Доцент |
| должность |

|  |
| --- |
| Дата представления работы |
| « » \_\_\_\_\_\_ 2023 г. |
| Заключение о допуске к защите |
|  |
|  |
|  |
| Оценка \_\_\_\_\_, \_\_\_\_\_\_ количество баллов |
|  |
| Подпись преподавателя \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

Череповец, 2023

*год*

Аннотация

Целью данной курсовой работы является изучение долговечности и разнообразия выбора пластиковых оконных конструкции от разных производителей, так же рассматривается возможность монтажа и эксплуатации на территории вологодской области.

В ходе работы использовалась программа Microsoft Office Excel 2016, данные были взяты с официального сайта компании, производящих оконные изделия.

Оглавление

Введение4

Инструменты. Microsoft Office Excel5

Описание предметной области5

Практическая часть8

[1.Группировка](#_Toc12339) данных8

[2.Изучение](#_Toc17741) взаимосвязей12

[3. Ряды динамики 15](#_Toc6491)

[Заключение](#_Toc23950)19

[Список литературы 20](#_Toc8335)

[Приложение 1. Техническое задание](#_Toc30380) 21

Введение

Прикладная статистика – это наука о методах обработки статических данных. Она активно используется в экономике, менеджменте, медицине, социологии и т.д. Термин «статистика» люди понимают по-разному. В обиходе статистика – это набор количественных данных о каком-либо явлении. Другое определение термина «статистика» - функция от результатов наблюдений, используемая для оценивания характеристик и параметров распределений и проверки гипотез. В 1954 г. академик Б.В. Гнеденко дал следующее определение: «Статистика состоит из трёх разделов: сбор статистических сведений, то есть сведений, характеризующих отдельные единицы каких-либо массовых совокупностей; статистическое исследование полученных данных, заключающееся в выяснении тех закономерностей, которые могут быть установлены на основе массового наблюдения; разработка приемов статистического наблюдения и анализа статистических данных».

Задачи прикладной статистики – сбор и обработка, анализ и представление информации о явлении или процессе, например, о выбросах в окружающую среду или о средней заработной плате работников. По типу решаемых задач прикладная статистика делится на разделы: описание данных, оценивание и проверка гипотез.

Эта курсовая работа создана с целью исследования долговечности и разнообразия выбора пластиковых оконных конструкции от разных производителей, так же рассматривается возможность монтажа и эксплуатации на территории вологодской области.

Инструменты. Microsoft Office Excel

Статистический анализ выполнялся с помощью программы Microsoft Office Excel 2016.

Microsoft Excel – это программа для работы с электронными таблицами, созданная корпорацией Microsoft. Программа предоставляет возможности экономико-статистических расчётов, графические инструменты. Входит в состав пакета Microsoft Office. Excel даёт возможность делать расчёты, данные которых могут располагаться в разных областях таблицы и иметь между собой определённую зависимость. Для выполнения расчётов в Excel необходимо вводить данные и формулы в ячейки таблицы. Программа делает вычисления и выдаёт в ячейке с формулой результат. В Excel содержится огромное множество формул – от простых расчётов (деление, умножение, сложение, вычитание) до статистических и финансовых расчётов. В Excel имеется множество функций: финансовые, функции даты и времени, логические, математические, ссылки и массивы, текст, работа с базой данных, мастер диаграмм. Excel совершает перерасчёт при изменении данных в ячейке таблицы. Excel позволяет строить графики на основании данных. Программа обеспечивает лёгкость и сохранность при обращении с данными, позволяет быстро выполнить работу. Также в программе имеется набор инструментов «Анализ данных». В нём присутствуют следующие возможности: корреляция, гистограмма, регрессия, выборка, описательная статистика и др.

Описание предметной области

Профиль – основная деталь любого окна. Чтобы понять, какой профиль лучше для пластиковых окон, стоит углубиться в сравнение доступных вариантов. Разбираемся в разновидностях конструкций, рекомендациях по выбору профиля для пластиковых окон, материалах.

Изначально кажется, что профиль даже от разных заводов-производителей имеет одинаковую структуру, но это ошибочное мнение. Разница качественных характеристик станет ощутимой уже через пару лет использования. Окна из дешёвого сегмента нередко меняют свою геометрию (имеют свойство перекашиваться), в таких профилях образуются щели, что провоцирует возникновение негативных последствий.

Отвечая на вопрос о том, какой профиль для окон лучше, требуется оценить отличительные черты различных конструкций. Профиль – основная часть окна, создаётся из металлопластика, деревянной основы, алюминия и ПВХ.

В среднем покупатели заказывают ПВХ профиль, так как он дополнительно оснащается стальными вставками, которые обеспечивают повышенные параметры долговечности, прочности и жёсткости. Сама конструкция обязательно имеет полости с воздухом, являющиеся основой формирования высокого уровня звуко- и теплоизоляции. Часть заводов-производителей полости заполняют специальным азотом, что позволяет исключить запотевание стекла при температурных скачках.

Существует классификация оконных профилей:

* **Класс A**. Размер внешней стеночки такого профиля равен 2.5 мм, толщина внутреннего слоя не будет более 2.6 мм. Благодаря таким особенностям конструкции формируется отличная защита от негативного влияния погодных условий, создаются идеальные параметры сохранности тепла.
* **Класс B**. В этом варианте стенки будут менее толстыми. Внутренняя – не свыше 2,1 мм, а внешняя – 2.6 мм. Такой вариант идеален для применения в зонах с тёплым климатом. Из-за более низких параметров толщины стенок увеличивается риск деформационных изменений конструкции (на 16%).
* **Класс C**. В этом варианте производители самостоятельно определяют толщину стенок, так как сертификация изделий отсутствует, что исключает строгие требования к готовой продукции. Эти варианты имеют популярность из-за низкой цены, но качество их является невероятно низким.

При выборе профиля для покупателя очень большую роль играет материал изделия, поэтому в большинстве организации было принято решение создавать разные ламинированные покрытия на пластиковых окнах или менять цвет пластика полностью, для расширения ассортимента продаваемой продукции.

Основные термины, связанные с работой:

**Долговечность** — срок службы объекта (в данном случае- срок службы профиля без потери качественных признаков с момента монтажа).

**Монтаж** — установка готового оконного изделия в проем с соблюдением норм.

**Соответствие климатическим нормам вологодской области** — по данным теплоизоляции, установочной глубине и максимальной морозостойкости проверяется возможность практичной установки при учете минимальных температур и влажности климата определенной области.

**Наличие в вологодской области** — проверка наличия компании, готовых произвести сборку готового оконного изделия по заказу.

**Оконный профиль** — каркас пластикового окна, его несущий элемент.

**Количество профилей** — тип, размер, наличие ламинации, цвет профиля и другие его минимальные отличия создают собой совокупность количества профилей.

Практическая часть

1.Группировка данных

В прикладной статистике существует несколько методов группировки данных, которые помогают упорядочить информацию и сделать ее более понятной и удобной для анализа. Ниже представлены некоторые из самых популярных методов группировки данных.

* Дискретизация и интервалы. Один из наиболее распространенных способов группировки. Состоит в разделении непрерывной переменной на интервалы или категории. Это может быть полезным, когда нужно визуализировать данные или проводить сравнительный анализ.
* Агрегация данных. Этот метод заключается в суммировании или усреднении значений переменных, чтобы получить общую информацию для группы.
* Факторизация. Этот метод заключается в группировке данных по категориальной переменной.
* Кластеризация. Данный метод позволяет автоматически группировать данные на основе их схожести или различия. Кластеризация может использоваться, когда неизвестно заранее, какие группы данных существуют.
* Группировка по времени. При анализе данных, связанных с временем, может быть полезно группировать данные по периодам, например, по дням, неделям или месяцам. Это позволяет проводить временные анализы и обнаруживать тренды и циклы.

Для данной курсовой работы был выбран метод группировки факторизация и сформирована таблица с исходными данными самых популярных компании, производящих профиля с сортировкой по разнообразию продукции (рис. 1).

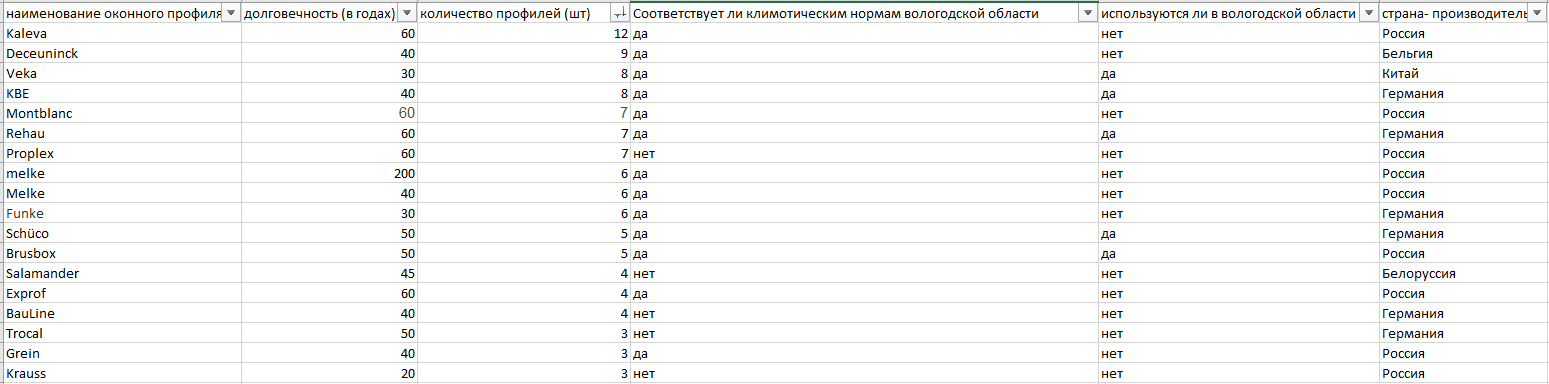


Рис. 1. Таблица с исходными данными

Таблица со сводными данными представлена на рис. 2.

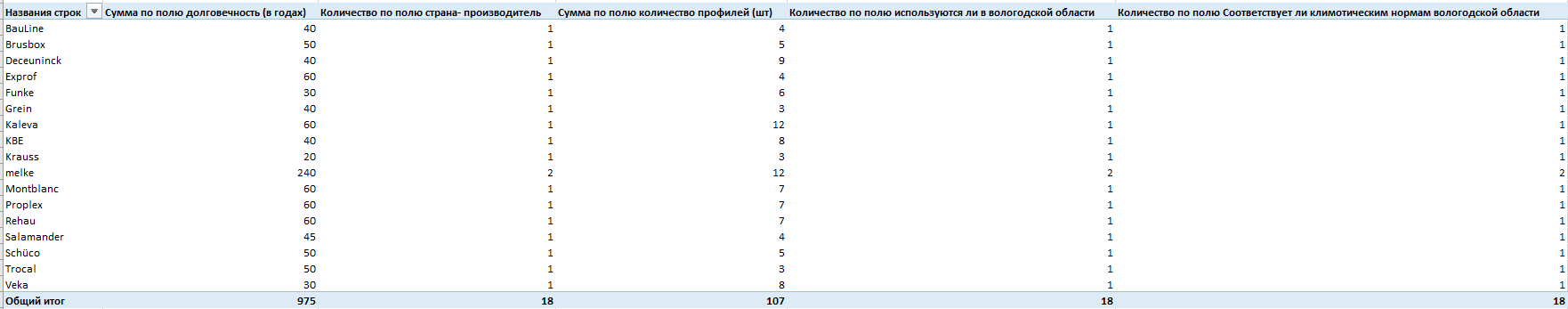


Рис. 2. Сводная таблица

Из сводной таблицы виден очень большой разброс в параметрах в каждой группе. Это связано с тем, что оконные профиля каждой из представленных компании имеют разные заявленные характеристики и большое разнообразие профилей.

Для более полного анализа рассмотрим графики, сделанные на основе сводной таблицы (рис. 3-8).

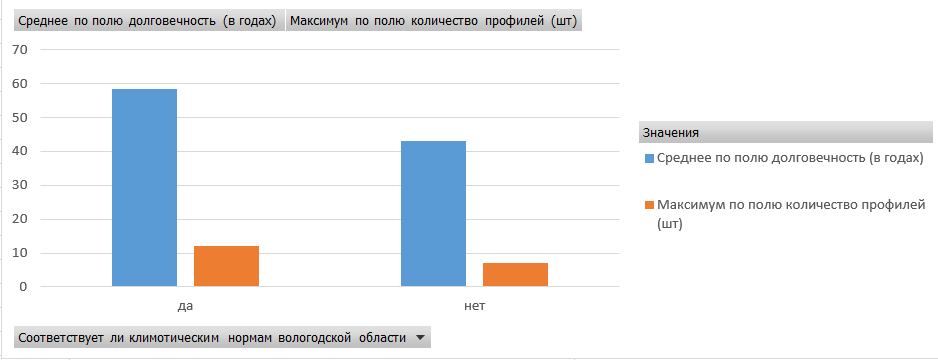


Рис. 3. Соответствие климатическим нормам.

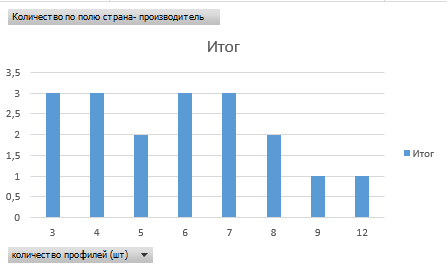


Рис. 4. Зависимость количества стран- производителей от разнообразия профилей.

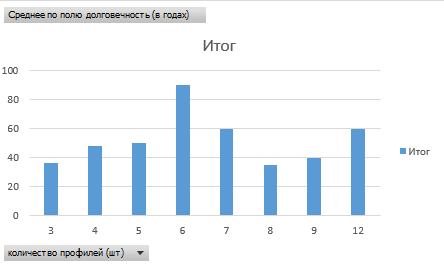


Рис. 5. Зависимость Средней долговечности от количества профилей.



Рис. 6. Зависимость долговечности от использования на территории вологодской области.



Рис. 7. Зависимость средней долговечности и разнообразия профилей от использования на территории вологодской области.

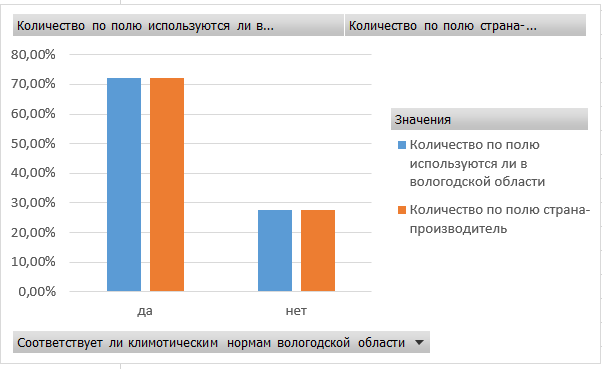


Рис. 8. Зависимость средней долговечности и разнообразия профилей от использования на территории вологодской области (процентное соотношение).

Графики, находящиеся на соответствующих листах рядом с таблицами, по которым строились, показывают точно то же самое, только значительно более наглядно.

Исходя из этих таблиц можно сказать, что наибольшее разнообразие профилей предоставляет компания Kaleva, при этом долговечность их продукции по заявленным гарантиям составляет 60 лет, что может составить очень уверенную конкуренцию всем другим представленным организациям, однако продукцию этой компании на территории вологодской области приобрести и установить нельзя, несмотря даже на соответствие климатическим нормам данной области. Компания melke, так же не доступная на территории вологодской области, наоборот, предоставляет не самый широкий спектр профилей, но дает гарантию на свою продукцию до 200 лет после монтажа. Компания Veka является худшим выбором по долговечности на территории области, однако предоставляет хороший спектр профилей на выбор (8 типов профиля). Однако рассматривая все компании, Funke уступает остальным по всем параметрам, ведь не предоставляет большого выбора профилей и не гарантирует долговечность своей продукции (гарантия на продукцию всего 30 лет с момента монтажа). Самым же лучшим выбором для жителей вологодской области неожиданно станут окна от компании Rehau, так как из всех доступных на территории области компании они могут предложить хороший выбор профилей и предоставляют лучшую гарантию (на 60 лет).

При помощи графиков зависимости долговечности от количества профилей, исследуем их взаимосвязь. Из графика в приложенном файле видно, что взаимосвязь слабая и частота мало зависит от принадлежности к группе. Такого результата удалось добиться тем, что использовалось именно среднее арифметическое долговечности, чтобы исключить влияние выбивающихся из общей выборки значении (к примеру, 200 лет гарантии встречается всего раз).

2. Изучение взаимосвязей

Взаимосвязи в статистике относятся к связям или отношениям между двумя или более переменными. Они позволяют нам понять, как изменение одной переменной может влиять на изменение другой переменной. Существуют различные методы для изучения взаимосвязей в статистике, включая: Корреляционный анализ: это метод, который позволяет определить меру взаимосвязи между двумя переменными. Он используется для измерения направления и степени связи между переменными. Коэффициент корреляции может быть положительным, отрицательным или равным нулю, что указывает на силу и направление связи.

* Регрессионный анализ: это метод, который используется для предсказания значения одной переменной на основе значения другой переменной. Регрессионный анализ используется для исследования возможной причинно-следственной связи между переменными.
* Анализ дисперсии: это метод, который используется для сравнения различий между группами или условиями на основе значений зависимой переменной. Анализ дисперсии позволяет определить, является ли разница между группами статистически значимой и влияет ли независимая переменная на зависимую переменную.
* Факторный анализ: это метод, который используется для исследования группы переменных с целью обнаружения скрытых факторов или конструктов, которые объясняют наблюдаемую взаимосвязь.

Взаимосвязи в статистике играют важную роль в анализе данных и помогают нам понять, какие переменные влияют на другие и как они взаимодействуют между собой.

Для данной курсовой работы были выбраны методы анализа дисперсии и регрессионный анализ. Была сформирована таблица на основе исходных данных популярных моделей майнеров для анализа дисперсии (рис. 9). Также были составлены графики для регрессионного анализа (рис. 8).

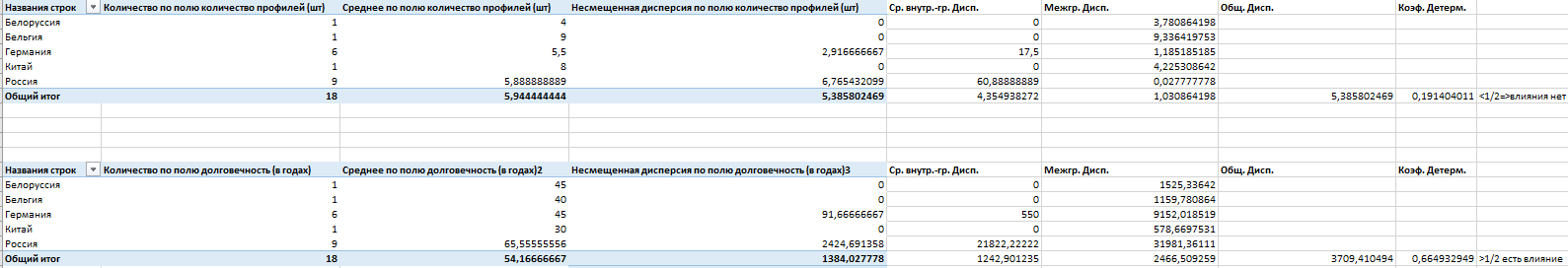


Рис. 9. Анализ дисперсии

Коэффициент детерминации, также известный как коэффициент R-квадрат, является статистической мерой, которая оценивает, насколько хорошо линейная регрессионная модель соответствует данным. Он показывает долю вариации в зависимой переменной, которая может быть объяснена регрессионной моделью. Более конкретно, коэффициент детерминации выражает, какую часть изменчивости (дисперсии) значений зависимой переменной можно объяснить с помощью линейной регрессионной модели. Этот коэффициент принимает значения от 0 до 1, где 0 означает, что модель не объясняет никакой вариации, а 1 означает, что модель объясняет всю вариацию. Если число близко к 0,5, то это означает что у модели низкая вариация, а если больше 0,5 - высокая вариация.

Исходя из результатов данной таблицы можно сделать вывод, что количество профилей от страны производителя не зависит, так как коэффициент детерминации составляет всего 0,19. Но в свою очередь страна- производитель сильно влияет на долговечность продукции, ведь коэффициент детерминации составляет 0,65.

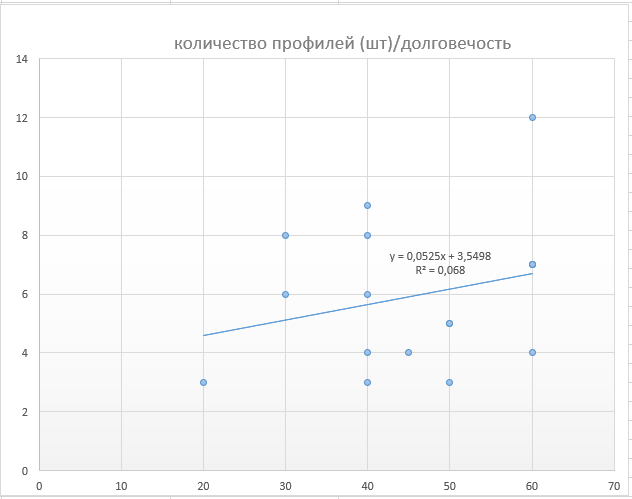


Рис. 10. Регрессионный анализ

Регрессия заключается в определении аналитического выражения, в котором изменение одной величины Y обусловлено влиянием одной или нескольких независимых величин X1, X2, X3… Регрессия может быть однофакторной (парной) и многофакторной (множественной). Для парной регрессии можно использовать графическое изображение. При множественной регрессии трудно чётко разграничить одни причинные явления от других. В этом случае наиболее приемлемый способ – метод перебора уравнений, реализуемый при помощи компьютера. После выбора регрессионной модели, используя результаты наблюдений зависимой переменной и факторов, нужно вычислить приближённые значения параметров регрессии, затем проверить значимость и адекватность модели. Коэффициент корреляции - это статистическая мера, которая вычисляет силу связи между относительными движениями двух переменных. Значения коэффициента корреляции находятся в диапазоне от -1.0 до 1.0.

На рис. 10 представлен график корреляционно-регрессионного анализа отношений количества профилей к долговечности. На графике с помощью встроенных функций Excel были найдены значения коэффициента детерминации и линейное уравнение регрессии.

Исходя из результатов корреляционно-регрессионного анализа можно сделать вывод, что данные имеют слабую тесноту взаимосвязи друг с другом.

## Ряды динамики

Ряды динамики (или временные ряды) в статистике представляют собой последовательность данных, собранных в разные моменты времени. Они позволяют анализировать изменения во времени и исследовать динамику различных явлений или переменных. Ряды динамики широко используются в различных областях, таких как экономика, финансы, социология, маркетинг и др. Они являются помощником в прогнозировании будущих значений и разработке стратегических решений. В статистике анализ рядов динамики включает в себя различные методы, такие как:

* Визуализация: построение графиков временных рядов, чтобы исследовать их основные характеристики. Это позволяет выявить тренды, сезонность, цикличность и случайную составляющую.
* Декомпозиция: разложение временного ряда на его компоненты, такие как тренд, сезонность и остаточную (случайную) составляющую. Это позволяет более детально изучить структуру ряда.
* Прогнозирование: использование статистических методов для предсказания будущих значений временного ряда. Это позволяет принимать обоснованные решения на основе имеющихся данных.
* Анализ стационарности: проверка условия стационарности ряда, то есть отсутствия систематических изменений характеристик во времени. Это важно для применения некоторых статистических методов анализа.
* Автокорреляция: проверка наличия зависимостей между значениями ряда в разные моменты времени. Это может помочь идентифицировать структуру ряда и выбрать соответствующий модельный подход.

Использование рядов динамики в статистике позволяет извлечь полезную информацию из данных и принимать обоснованные решения на основе анализа различных временных шкал.

Для данной курсовой работы был выбран метод визуализации временных рядов в виде графика. Для его создания были проведены дополнительные поиски информации о средней заработной плате населения вологодской области на конкретный промежуток времени, также были проведены дополнительные вычисления среднего уровня ряда, среднего абсолютного прироста, среднего темпа роста и прироста (рис. 11).

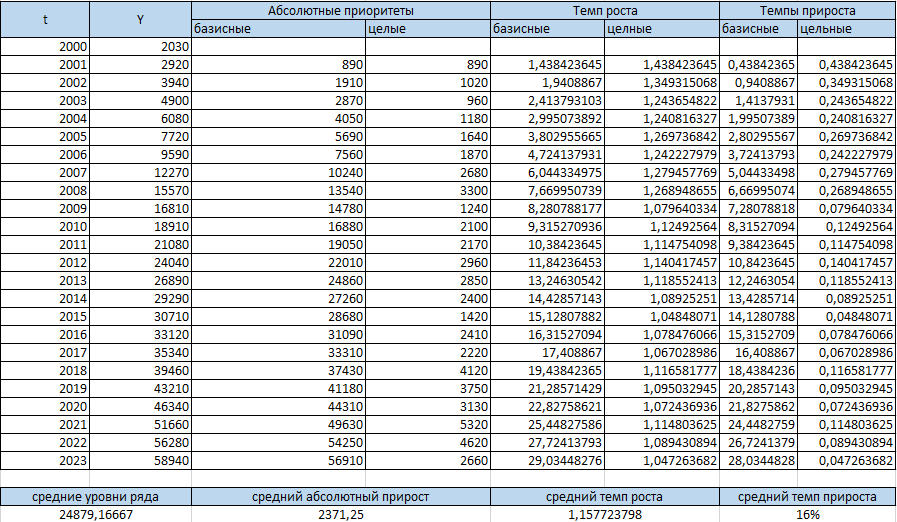


Рис. 11. Ряды динамики

На основе данных полученных при поиске информации, мы можем построить график временных рядов, чтобы исследовать их основные характеристики (рис. 12).



Рис. 12. График временных рядов

Исходя из результатов данной таблицы и графика можно сделать вывод, что период резкого увеличения средней заработной платы на территории вологодской области не было. Также можно сказать, что среднее ряда «Y» (средняя зарплата по вологодской области) будет равняться 24879,16. Средний абсолютный прирост составит 2371.25. Средний темп роста 1,1, а средний темп прироста будет около 16%. Это означает, что будет наблюдаться восходящий тренд.

Заключение

В ходе выполнения курсовой работы была выявлена слабая зависимость долговечности оконных конструкции от разнообразия предоставляемых компанией профилей. Данный фактор играет решающую роль, следовательно, гипотеза подтверждена.

Список литературы

1. Ершов, Е.В. Методика и организация самостоятельной работы: учебное пособие. [Текст] / Ершов Е.В., Виноградова Л.Н., Селивановских В.В. // Череповец: ЧГУ, 2015. – 243 с.

2. Шанченко, Н.И. Эконометрика: лабораторный практикум: учебное пособие. [Текст] / Шанченко Н.И. // Ульяновск: УлГТУ, 2004. – 80 с.

3. Microsoft Excel – Википедия [электронный ресурс]. https://ru.wikipedia.org/wiki/Microsoft\_Excel Дата обращения: 16.11.2023.

Приложение 1

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное

Образовательное учреждение высшего образования

«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Институт информационных технологий

*наименование института (факультета)*

Математическое и программное обеспечение ЭВМ

*наименование кафедры*

Прикладная статистика

*наименование дисциплины в соответствии с учебным планом*

УТВЕРЖДАЮ

Зав. кафедрой                  ,

д. т. н., профессор                    Ершов Е. В.

«      »                       2023 г.

Изучение связи между уровнем жизни, денежных доходов в расчете на душу населения и ожидаемой продолжительностью жизни в регионах Российской Федерации

Техническое задание на курсовой проект

Листов 5

|  |  |
| --- | --- |
| Руководитель | Гонтарева И. Б. |
|  | *ФИО преподавателя* |
| Исполнитель |  |
| студент | 1ПИб-02-03оп-22 |
|  | *группа* |
|  | Бурмистров Дмитрий Сергеевич |
|  | *Фамилия, имя, отчество* |

2023 год

Введение

Данное техническое задание распространяется на исследование производства автомобилей по типам двигателей.

1. Основания для разработки

Основанием для разработки является задание на курсовую работу по дисциплине «Прикладная статистика», выданное на кафедре МПО ЭВМ ИИТ ЧГУ.

Дата утверждения: 30 сентября 2023 года.

Наименование темы разработки: «является изучение долговечности и разнообразия выбора пластиковых оконных конструкции от разных производителей».

1. Назначение разработки

При помощи знаний за курс «Прикладная статистика» провести статистический анализ связи между принадлежностью к определённой группе процессоров и их производительностью.

1. Требования к программе

3.1. Требования к функциональным характеристикам

Таблицы должны быть понятными логически, данные должны быть взяты с официального сайта intel.inc, пользователь сможет создавать диаграммы и редактировать данные.

3.2. Требования к надёжности

Чтобы разработка была надёжной, необходимо хранить копии файла Excel на стабильном носителе.

3.3. Условия эксплуатации

Файл необходимо открывать в Microsoft Office Excel, в которой она разрабатывалась.

3.4. Требования к составу и параметрам технических средств

Состав технических средств: персональный компьютер с установленной на нём программой Excel, стабильный носитель (на нем хранится разработка).

3.5. Требования к информационной и программной совместимости

Операционная система – Microsoft Windows 7, 8, 10. Процессор – тактовая частота 1,6 ГГц и больше, два или более ядер, минимум 1 ГБ ОЗУ, минимум 1 ГБ доступного пространства.

3.6. Требования к маркировке и упаковке

Хранение на стабильном носителе.

3.7. Требования к транспортированию и хранению

Разработка должна храниться на переносном устройстве (USB-носитель) и стабильном носителе, а также её копии

3.8. Специальные требования

Требования не предъявляются.

1. Требования к программной документации

4.1 Содержание расчётно-пояснительной записки:

Программная документация должна содержать расчётно-пояснительную записку с приложением «техническое задание».

4.2. Требования к оформлению

Расчётно-пояснительная записка должна быть оформлена в соответствии с ГОСТ 2.105-79.

5. Стадии и этапы разработки

Таблица П1.1

Стадии и этапы разработки

|  |  |  |  |
| --- | --- | --- | --- |
| Наименование этапа разработки ПО | Сроки разработки | Результат выполнения | Отметка о выполнении |
| Определение темы  разработки | 30.09.2023 | Определена тема разработки |  |
| Оформление технического задания | 15.10.2023 | Создано техническое задание |  |
| Сбор статистических  данных | 09.11.2023 | Собраны статистические данные |  |
| Проведение  статистического  анализа | 11.11.2023 – 29.11.2023 | Проведён статистический анализ |  |
| Оформление расчётно-  пояснительной записки | 01.12.2023 – 24.12.2023 | Оформлена расчётно-пояснительная записка |  |

6. Порядок контроля и приёмки

Таблица П1.2

Порядок контроля и приёма

|  |  |  |  |
| --- | --- | --- | --- |
| Наименование контрольного этапа выполнения курсовой работы | Сроки контроля | Результат выполнения | Отметка о приёмке результата контрольного этапа |
| Определение  темы курсовой  работы | 30.09.2023 | Определена тема курсовой работы |  |
| Оформление  технического  задания | 15.10.2023 | Оформлено техническое задание |  |
| Проведение статистического анализа | 30.11.2023 | Проведён статистический анализ |  |
| Оформление РПЗ | 16.12.2023 | Оформленная РПЗ |  |
| Сдача РПЗ, оценка | 20.12.2023 | Получение итоговой оценки за курсовую работу |  |