Министерство науки и высшего образования Российской Федерации

Новосибирский Государственный технический университет

Кафедра автоматизированных систем управления



**Отчет по лабораторной работе № 2**

**по дисциплине «Методы анализа данных»**

**«Классификация»**

Выполнили

студенты группы АВТ-213

Андриевский В. В.

Боровинская Д. И.

Речкин Е. А.

Преподаватель:

к. т. н. Ганелина Н. Д.

г. Новосибирск 2025 г

**Содержание**

[Описательная статистика. Корреляционный анализ 3](#_Toc209367118)

[Регрессионный анализ 15](#_Toc209367119)

[Дисперсионный анализ 22](#_Toc209367120)

# Описательная статистика. Корреляционный анализ

Цель работы: изучение методов решения задачи классификации.

Среда выполнения: Python.

Описание исходных данных: датасет Bike Sharing - содержит информацию о почасовом и ежедневном количестве арендованных велосипедов в системе Capital bikeshare за период с 2011 по 2012 год, а также соответствующие данные о погоде и сезонах.

Независимая переменная: temp - нормализованная температура в градусах Цельсия (значения получены с помощью формулы (t-t\_min)/(t\_max-t\_min), где t\_min = -8, t\_max = +39).

Зависимая переменная: registered - количество зарегистрированных пользователей.

**Дескриптивный анализ**

|  |  |  |
| --- | --- | --- |
|  | temp | registered |
| Среднее | 0,496987 | 153,7869 |
| Стандартная ошибка | 0,001461 | 1,14813 |
| Медиана | 0,5 | 115 |
| Мода | 0,62 | 4 |
| Стандартное отклонение | 0,192556 | 151,3573 |
| Дисперсия выборки | 0,037078 | 22909,03 |
| Эксцесс | -0,94184 | 2,750018 |
| Асимметричность | -0,00602 | 1,557904 |
| Интервал | 0,98 | 886 |
| Минимум | 0,02 | 0 |
| Максимум | 1 | 886 |
| Сумма | 8637,14 | 2672662 |
| Счет | 17379 | 17379 |

Оценка близости выборок к нормальной: для оценки понадобятся такие описательные статистики, как мода, медиана, средн. значение, эксцесс и коэффициент асимметрии.

Если коэффициент асимметрии и эксцесс равны нулю, то гипотеза о нормальности не отклоняется.

Также для симметричного распределения характерно равенство моды, медианы и средн. значения (а так как нормальное распределение является симметричным, то это же условие характерно и для него).

По результатам дескриптивного анализа мы видим, что:

* для переменной temp значения моды (0,62), медианы (0,5) и средн. значения (0,496987) близки. Коэффициент асимметрии (-0,00602) довольно мал и близок к нулю, однако значение эксцесса (-0,94184) близко к -1. То есть, выборка близка к нормальной, но нельзя явно говорить о том, что она является таковой.
* для переменной registered значения моды (4), медианы (115) и средн. значения (153,7869) довольно различны. Коэффициент асимметрии 1,557904) больше нуля, то есть правое плечо распределения будет длиннее левого. Значение эксцесса (2,750018) также больше нуля, это говорит о том, что гистограмма будет более вытянутая. Значит, данная выборка не является нормальной.

Построение гистограмм: количество интервалов рассчитаем по правилу Стёрджесса: k = 1 + log2(N), где N – общее число наблюдений (17379).

Величина шага (длина интервала): h = (макс. значение – мин. значение) / k

h (temp) = 0,06496495

h (registered) = 58,73361788

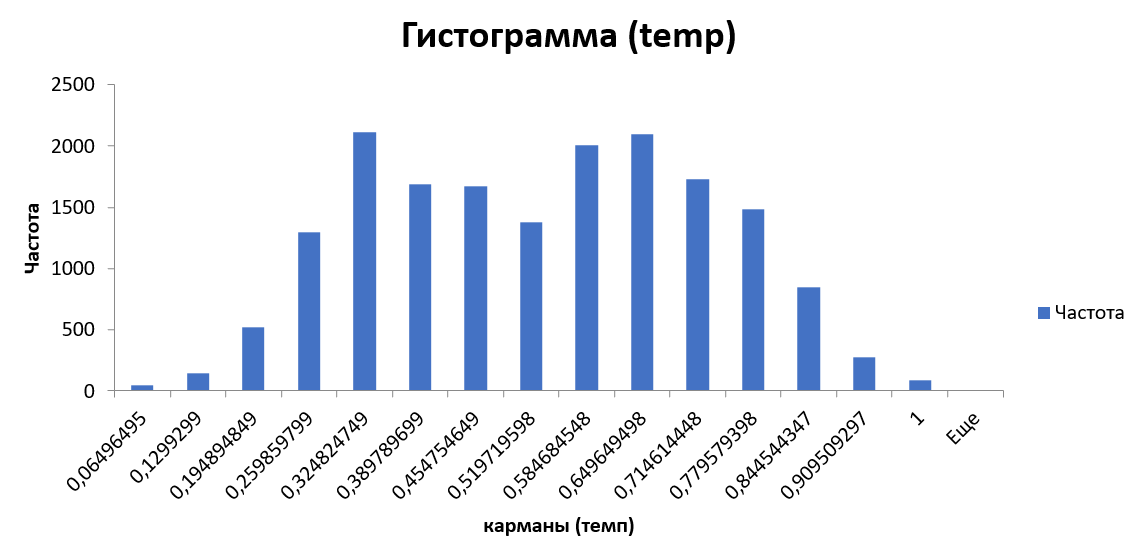


Рисунок 1 – гистограмма для выборки переменной temp

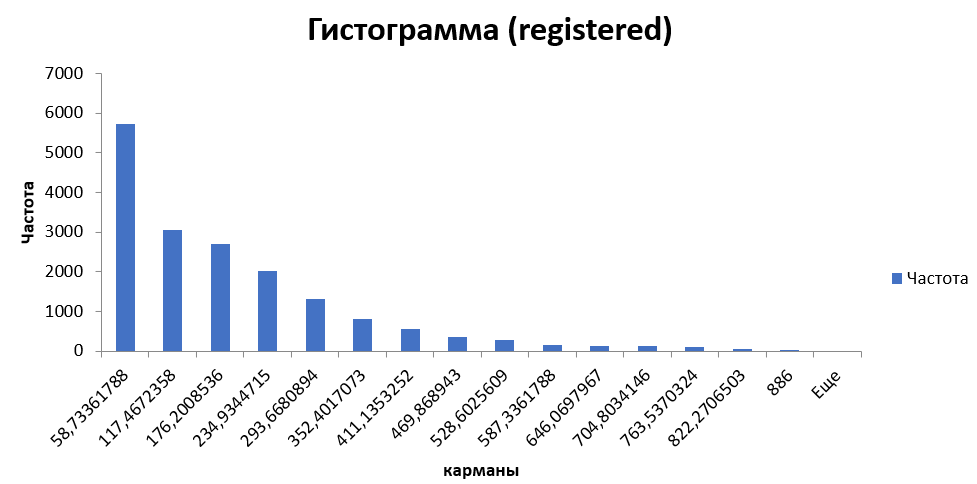


Рисунок 2 – гистограмма для выборки переменной registered

**Таблицы сопряженности**

Была сформулирована нулевая гипотеза: температура НЕ влияет на число регистраций.

Фактические данные

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | temp | | | | | | | | |
|  | 0 - 0.25 | 0.26 - 0.32 | 0.33-0.38 | 0.39 - 0.58 | 0.59 - 0.77 | 0.78 - 0.84 | 0.85 - 0.90 | 0.91 - 1 |
| registered | 0 - 58 | 1269 | 951 | 580 | 1599 | 1299 | 30 | 0 | 0 |
| 59 - 117 | 425 | 501 | 394 | 881 | 704 | 102 | 40 | 14 |
| 118 - 293 | 266 | 529 | 557 | 1832 | 2146 | 489 | 156 | 52 |
| 294 - 411 | 23 | 71 | 102 | 404 | 598 | 107 | 43 | 11 |
| 412 - 886 | 30 | 60 | 55 | 340 | 560 | 117 | 35 | 7 |

Ожидаемые данные

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | temp | | | | | | | | |
|  | 0 - 0.25 | 0.26 - 0.32 | 0.33-0.38 | 0.39 - 0.58 | 0.59 - 0.77 | 0.78 - 0.84 | 0.85 - 0.90 | 0.91 - 1 |
| registered | 0 - 58 | 663,47 | 696,10 | 556,35 | 1666,423 | 1749,151 | 278,50 | 90,30 | 27,68 |
| 59 - 117 | 354,55 | 371,99 | 297,31 | 890,524 | 934,7331 | 148,83 | 48,26 | 14,79 |
| 118 - 293 | 698,10 | 732,43 | 585,39 | 1753,41 | 1840,456 | 293,04 | 95,02 | 29,13 |
| 294 - 411 | 157,41 | 165,15 | 131,99 | 395,3682 | 414,9959 | 66,07 | 21,42 | 6,56 |
| 412 - 886 | 139,45 | 146,31 | 116,94 | 350,2747 | 367,6637 | 58,54 | 18,98 | 5,81 |

Расчет ХИ2

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | temp | | | | | | | | |
|  | 0 - 0.25 | 0.26 - 0.32 | 0.33-0.38 | 0.39 - 0.58 | 0.59 - 0.77 | 0.78 - 0.84 | 0.85 - 0.90 | 0.91 - 1 |
| registered | 0 - 58 | 552,6469 | 93,33935 | 1,005056 | 2,727928 | 115,8482 | 221,7378 | 90,30853 | 27,68583 |
| 59 - 117 | 13,99687 | 44,74118 | 31,44438 | 0,101857 | 56,95505 | 14,73613 | 1,413813 | 0,042729 |
| 118 - 293 | 267,4586 | 56,50539 | 1,377299 | 3,522503 | 50,72492 | 131,0337 | 39,13007 | 17,95304 |
| 294 - 411 | 114,7728 | 53,67687 | 6,817347 | 0,188452 | 80,70084 | 25,34429 | 21,7224 | 2,989541 |
| 412 - 886 | 85,91218 | 50,92135 | 32,81029 | 0,30139 | 100,617 | 58,37781 | 13,51574 | 0,239495 |

Число степеней свободы: 28

Если вероятность, того что случайная величина, имеющая ХИ2 - распределение с (r-1)(c-1) степенями свободы,примет значение больше вычисленной статистики Х 20, т.е. P{Х 2(r-1)\*(c-1)>Х 20}, меньше уровня значимости , то нулевая гипотеза отклоняется.

Уровень значимости: 0,05

Значение p: 0

Итог: нулевая гипотеза отклоняется

**Корреляционный анализ**

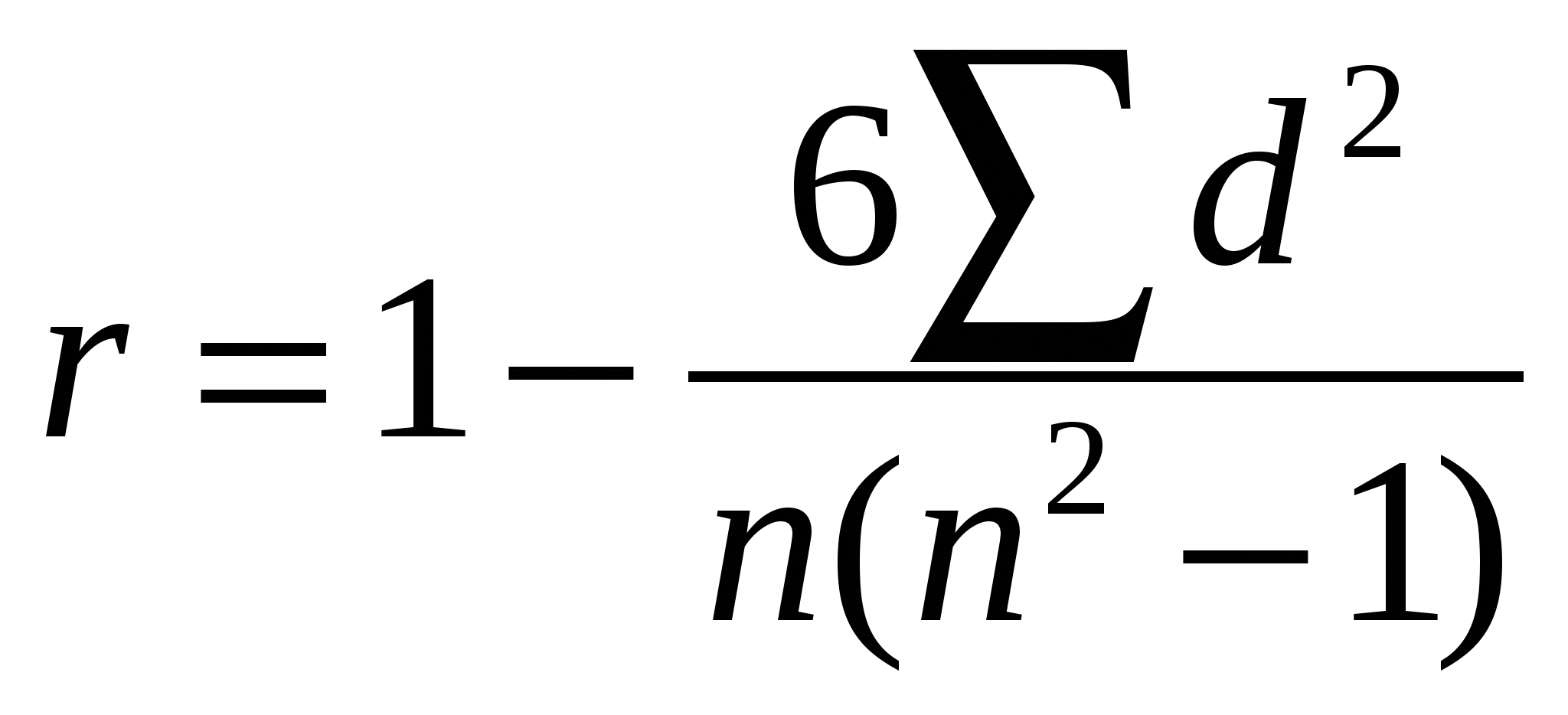
Для переменных, измеряемых в количественной (числовой) шкале, рассчитывают ковариацию или корреляционный момент, а на его основе линейный коэффициент корреляции (коэффициент корреляции Пирсона).

Для ранговых переменных используются непараметрические ранговые коэффициенты корреляции: коэффициент ранговой корреляции Кендалла и коэффициент корреляции Спирмена.

Корреляция между категориальными переменными не может быть измерена с помощью коэффициентов Пирсона, Спирмена и Кендалла. Коэффициенты и выводы для категориальных данных обычно строятся на основании таблиц сопряжённости.

Так как рассматриваемые выборки (температура и число регистраций) далеки от нормального распределения, был проведен ранговый корреляционный анализ.

Для расчёта коэффициента корреляции воспользуемся формулой Спирмена:

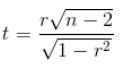
,

где n – число парных наблюдений,

– сумма квадратов разностей рангов

Коэффициент корреляции равен 0,373 – это средняя (умеренная) корреляция.

Определим статистическую значимость коэффициента при помощи t-критерия, рассчитанного по следующей формуле:



Значимость = 0,05

Число степеней свободы = 17377

t = 52,99

Табличное значение = 1,96

Рассчитанное значение больше табличного, значит корреляционная связь считается статистически значимой.

Также был произведен расчёт коэффициента корреляции для других параметров: часы, дни недели и время года. Их значения равны соответственно 0,51 (средняя (заметная) корреляция), 0,03 (слабая корреляция) и 0,18 (тоже слабая корреляция). То есть, среди всех признаков сильнокоррелирующим можно назвать часы. В ночное время не так часто берут в аренду велосипеды по сравнению с дневным временем.

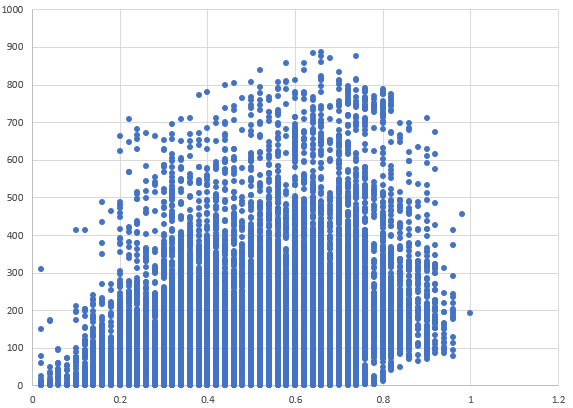


Рисунок 3 - диаграмма рассеивания для параметра температура

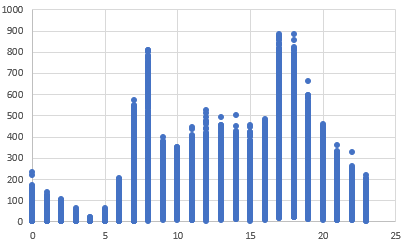


Рисунок 4 - диаграмма рассеивания для параметра часы

Анализ соотношения понятий «зависимость» и «корреляция»:

Статистическая зависимость — это связь двух случайных величин, при которой распределение вероятностей одной из них зависит от того, какие возможные значения приняла другая величина. Причинно-следственная связь обязательна.

Корреляция — взаимосвязь двух или нескольких случайных величин. При изменении значения одной переменной происходит закономерное изменение (уменьшению или увеличению) другой(-их) переменной(-ых). Корреляционная зависимость отражает только взаимосвязь между переменными и не говорит о причинно-следственных связях.

Независимость всегда эквивалентна некоррелированности. Обратное в целом неверно, однако при многомерном нормальном распределении понятия также эквивалентны.

**Ложная корреляция**

Рассматривалась корреляция выработки ветровой энергии на Фиджи от популярности мема «мальчик с воздушным шаром».

Гипотеза: корреляция выработки ветровой энергии на Фиджи от популярности мема «мальчик с воздушным шаром» является ложной.

Данные:

|  |  |
| --- | --- |
| Популярность мема «мальчик с воздушным шаром» | Выработка ветровой энергии на Фиджи |
| 0,02 | 37,6667 |
| 0,006 | 1,18182 |
| 0,005 | 1 |
| 0,007 | 1 |
| 0,005 | 1 |
| 0,004 | 1,6 |
| 0,006 | 1,83333 |
| 0,004 | 1 |
| 0,002 | 1 |
| 0,0026 | 1 |
| 0,0026 | 1 |
| 0,0026 | 1 |
| 0,0026 | 1,08333 |

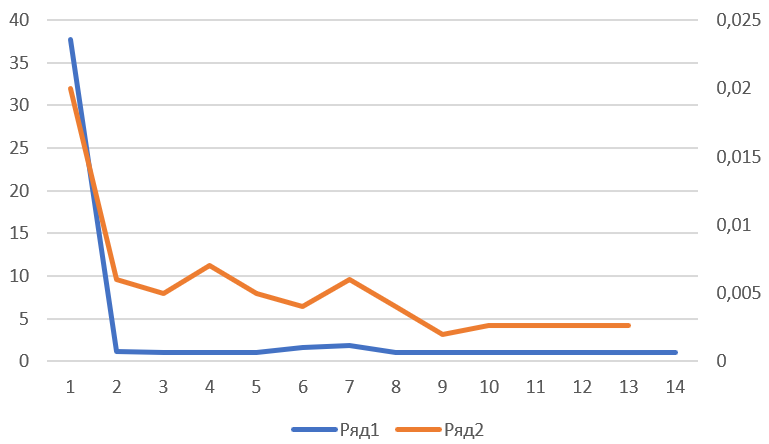


Рисунок 5 – график корреляции выработки ветровой энергии на Фиджи от популярности мема «мальчик с воздушным шаром»

Был произведен расчет значения коэффициента корреляции, он равен 0,49. Это средняя (умеренная) корреляция. То есть, даже несмотря на то, что логически две рассматриваемые переменные никак не связаны между собой, корреляция между числовыми значениями имеется.

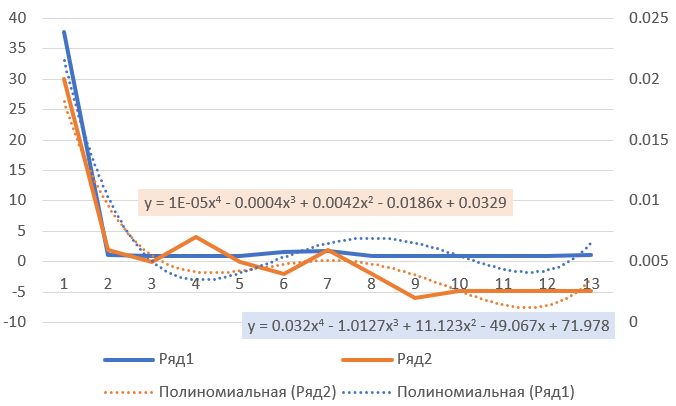


Рисунок 6 - график корреляции выработки ветровой энергии на Фиджи от популярности мема «мальчик с воздушным шаром» (с линиями тренда и уравнениями)

Уравнение тренда для графика популярности мема «мальчик с воздушным шаром»:

y = 0.032x4 - 1.0127x3 + 11.123x2 - 49.067x + 71.978

Уравнение тренда для графика выработки ветровой энергии на Фиджи:

y = 1E-05x4 - 0.0004x3 + 0.0042x2 - 0.0186x + 0.0329

Новые (предсказанные) значения зависимых переменных по построенным моделям:

|  |  |  |
| --- | --- | --- |
|  | Популярность мема «мальчик с воздушным шаром» | Выработка ветровой энергии на Фиджи |
| 1 | 33.0533 | 0.01811 |
| 2 | 10.7464 | 0.00946 |
| 3 | 0.1331 | 0.00491 |
| 4 | -2.9428 | 0.00266 |
| 5 | -1.8695 | 0.00115 |
| 6 | 0.7328 | -0.00094 |
| 7 | 3.0119 | -0.00469 |
| 8 | 3.8836 | -0.01094 |
| 9 | 3.0317 | -0.02029 |
| 10 | 0.908 | -0.0331 |
| 11 | -1.2677 | -0.04949 |
| 12 | -1.5076 | -0.06934 |
| 13 | 2.9441 | -0.09229 |

Остатки (исходное минус предсказанное):

|  |  |  |
| --- | --- | --- |
|  | Популярность мема «мальчик с воздушным шаром» | Выработка ветровой энергии на Фиджи |
| 1 | 4.6134 | 0.00189 |
| 2 | -9.56458 | -0.00346 |
| 3 | 0.8669 | 9E-05 |
| 4 | 3.9428 | 0.00434 |
| 5 | 2.8695 | 0.00385 |
| 6 | 0.8672 | 0.00494 |
| 7 | -1.17857 | 0.01069 |
| 8 | -2.8836 | 0.01494 |
| 9 | -2.0317 | 0.02229 |
| 10 | 0.092 | 0.0357 |
| 11 | 2.2677 | 0.05209 |
| 12 | 2.5076 | 0.07194 |
| 13 | -1.86077 | 0.09489 |

Стандартизированные остатки:

|  |  |  |
| --- | --- | --- |
|  | Популярность мема «мальчик с воздушным шаром» | Выработка ветровой энергии на Фиджи |
| 1 | 1.237553113 | 0.061100924 |
| 2 | -2.565716337 | -0.111856717 |
| 3 | 0.232547534 | 0.002909568 |
| 4 | 1.057663418 | 0.140305825 |
| 5 | 0.769748701 | 0.124464845 |
| 6 | 0.23262801 | 0.159702943 |
| 7 | -0.31615359 | 0.345591997 |
| 8 | -0.773531052 | 0.482988254 |
| 9 | -0.545007296 | 0.720602957 |
| 10 | 0.024679171 | 1.154128558 |
| 11 | 0.608314734 | 1.683993182 |
| 12 | 0.672668354 | 2.325714523 |
| 13 | -0.499155006 | 3.067654311 |



Рисунок 7 – график остатков значений популярности мема «мальчик с воздушным шаром»



Рисунок 8 – график остатков значений выработки ветровой энергии на Фиджи

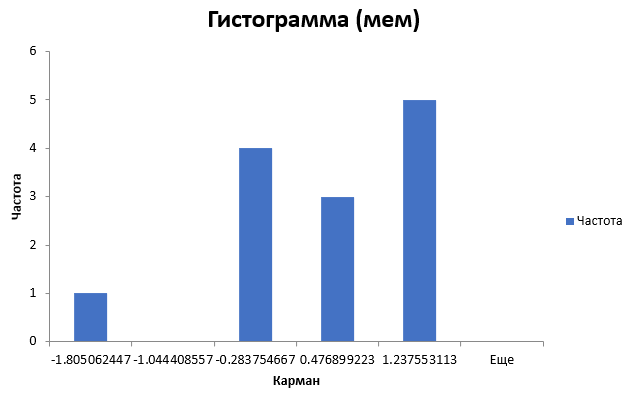


Рисунок 9 – гистограмма стандартизированных остатков значений популярности мема «мальчик с воздушным шаром»

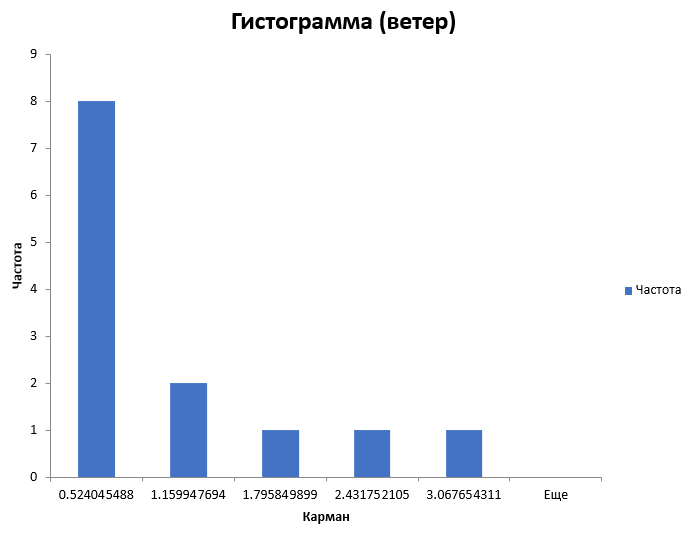


Рисунок 10 - гистограмма стандартизированных остатков значений выработки ветровой энергии на Фиджи

Исходный коэффициент корреляции между популярностью мема и выработкой ветровой энергии составил 0.49, что указывает на умеренную связь. Однако коэффициент корреляции между остатками составил 0.06, что говорит об отсутствии линейной связи между переменными. Это подтверждает гипотезу о ложной корреляции. Обе переменные меняются со временем, и обе трендируют в одном направлении, но при этом не связаны друг с другом. Иногда возможно влияние третьей переменной на оба параметра, но в данном случае это маловероятно.

# Регрессионный анализ

Цель работы: изучение регрессионного анализа.

Среда выполнения: MS Excel.

Была выбрана другая независимая переменная: hr – часы.

Зависимая переменная осталась прежней.

**Дескриптивный анализ**

|  |  |
| --- | --- |
|  | hr |
| Среднее | 11,54675 |
| Стандартная ошибка | 0,05245 |
| Медиана | 12 |
| Мода | 16 |
| Стандартное отклонение | 6,914405 |
| Дисперсия выборки | 47,809 |
| Эксцесс | -1,19802 |
| Асимметричность | -0,01068 |
| Интервал | 23 |
| Минимум | 0 |
| Максимум | 23 |
| Сумма | 200671 |
| Счет | 17379 |

Оценка близости выборки к нормальной:

По результатам дескриптивного анализа мы видим, что:

* для переменной hr значения моды (16), медианы (12) и средн. значения (11,54675) близки. Коэффициент асимметрии (-0,01068) довольно мал и близок к нулю, однако значение эксцесса (-1,19802) далеко от 0. То есть, закон распределения не нормальный .

Построение гистограммы: количество интервалов рассчитаем по правилу Стёрджесса: k = 1 + log2(N), где N – общее число наблюдений (17379).

Величина шага (длина интервала): h = (макс. значение – мин. значение) / k

h (hr) = 1,524688

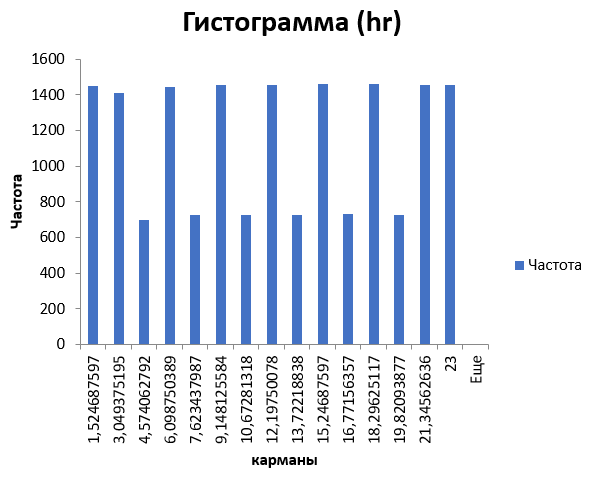


Рисунок 11 – гистограмма для выборки переменной hr

Построили линейную регрессионную модель для признаков с высоким коэффициентом корреляции (часы и количество зарегистрированных пользователей).

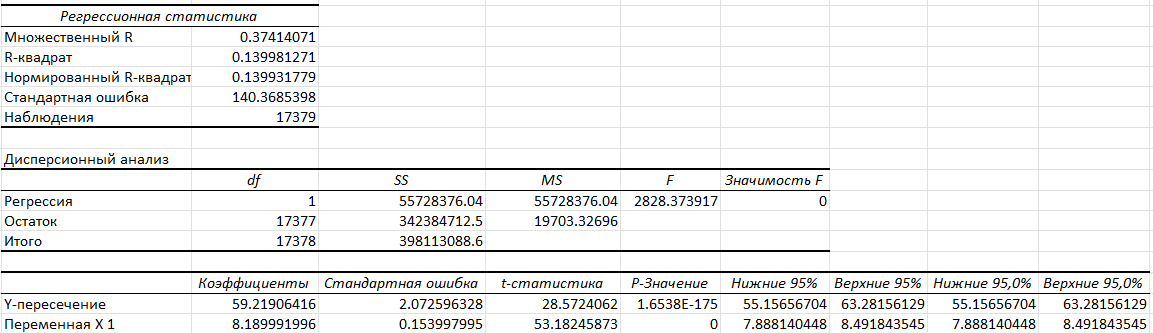


Рисунок 12 – регрессионный анализ

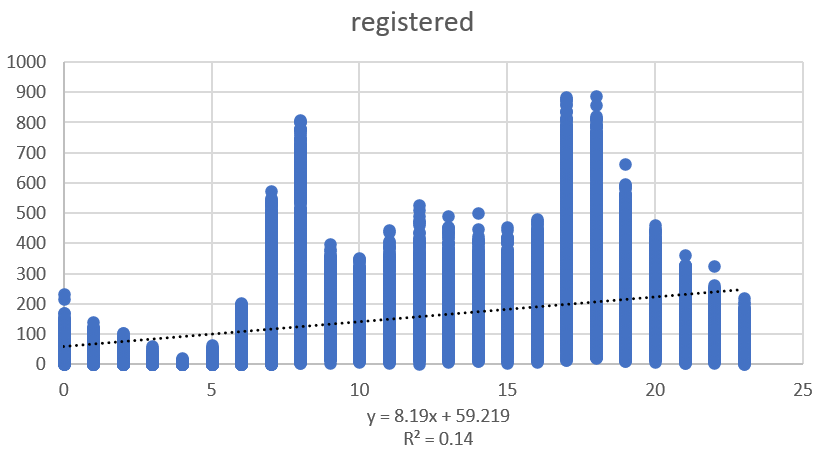


Рисунок 13 – линейная регрессионная модель

Уравнение регрессии: y = 8,19 \* x + 59,219

Коэффициент детерминации: R2=0,14



Рисунок 14 – диаграмма рассеяния остатков с полиномиальным трендом для линейной регрессионной модели

Модель нелинейна, так как имеется небольшой тренд на возрастание (постоянство среднего). Разброс непостоянен по всему графику, то есть дисперсия непостоянна.

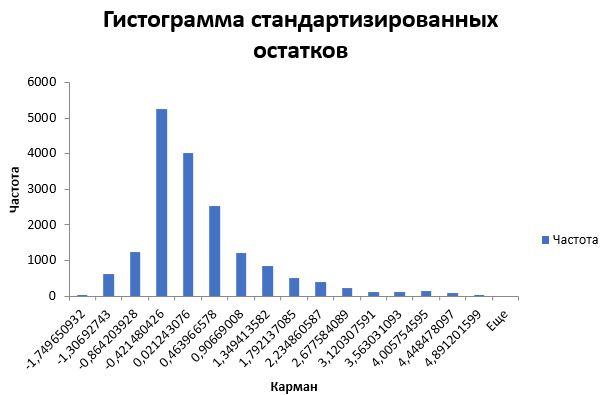


Рисунок 15 – гистограмма стандартизированных остатков

Оценка качества построенной модели: R-квадрат = 0,14 - модель плохо объясняет данные, потому что значение довольно далеко от 1. Гистограмма стандартизированных остатков ассиметрична и имеет выбросы, то есть модель не является качественной.

Оценка значимости построенной модели: p-значения для коэффициентов равны нулю. Вывод: нулевая гипотеза отклоняется. Коэффициенты статистически значимы. Множественный R = 0,374; R-квадрат = 0,14 - модель плохо объясняет данные. Значимость F = 0 - нулевая гипотеза отклоняется, уравнение значимо.

Оценка значимости коэффициентов:



Рисунок 16 - Оценка значимости коэффициентов линейной регрессионной модели

Значения t-статистики были взяты из линейной регрессионной модели. Критическое значение t-распределения при уровне значимости 0,05 равно 1,96. Так как критическое значение меньше t-статистики, то коэффициенты значимы.

**Сравнение регрессионных моделей**

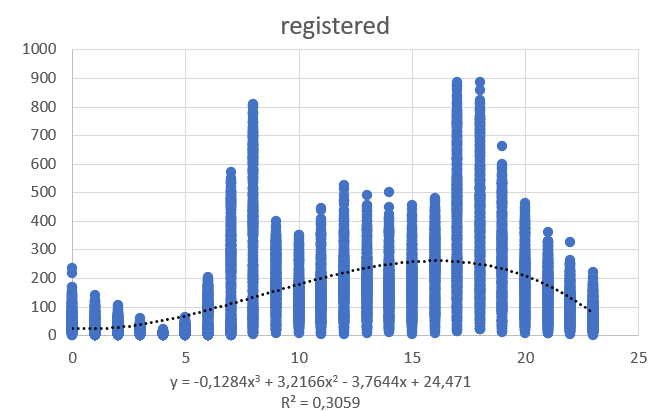


Рисунок 17 – диаграмма рассеяния для переменной числа зарегистрированных пользователей с полиномиальным трендом

Уравнение полиномиального тренда: y=-0,1284x3 + 3,2166x2 – 3,7644x + 24,471

Коэффициент детерминации: R2 = 0,3059



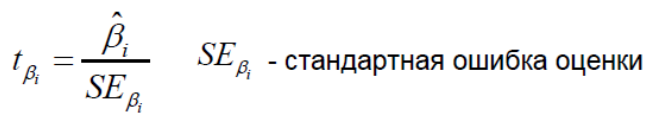
Рисунок 18 – диаграмма рассеяния остатков для переменной числа зарегистрированных пользователей с полиномиальным трендом

Были построены и сравнены линейная и полиномиальная регрессионные модели. Значения коэффициента детерминации R² для обеих моделей являются низкими 0,30 и 0,14, что указывает на слабую связь между временем и количеством регистраций. Время суток объясняет лишь часть вариации данных.

Вывод: Несмотря на низкий коэффициент R², полиномиальная модель является предпочтительной. Низкие значения R² обеих моделей позволяют сделать основной вывод: фактор времени суток не является значимым или сильным предиктором количества регистраций.

Проверка значимости коэффициентов регрессии:

Гипотеза H0​: βi​=0 (коэффициент незначим)



Для проведения регрессионного анализа в MS Excel были отдельно посчитаны значения x, x2 и x3.

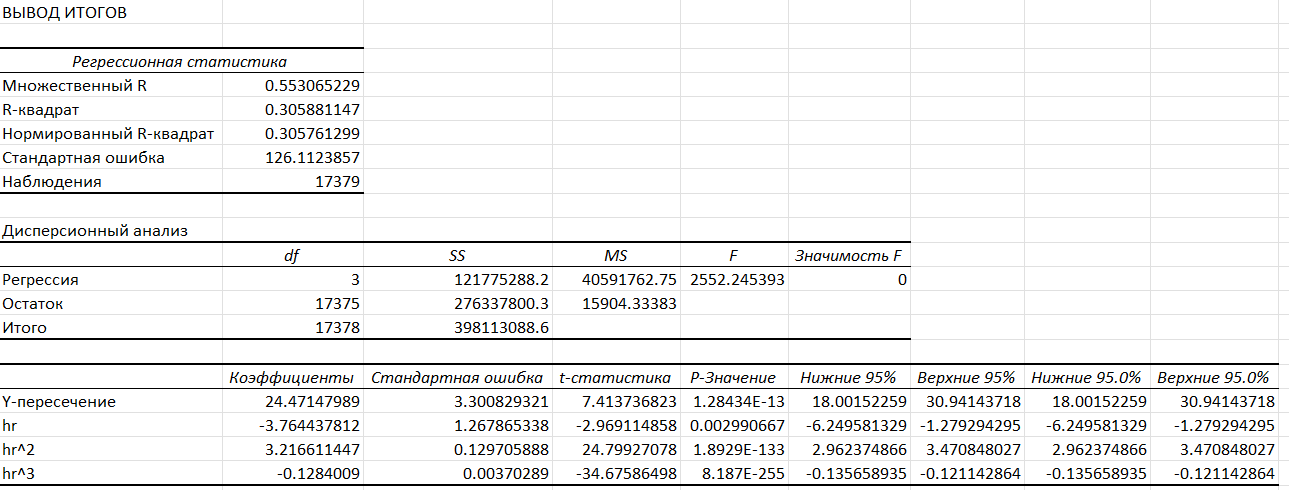


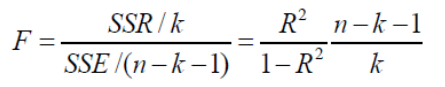
Рисунок 19 – регрессионный анализ для полиномиальной модели

У каждого коэффициента p-value сильно меньше уровня значимости (0,05). Отсюда мы можем сделать вывод, что гипотеза отвергается.

Проверка значимости уравнения регрессии:

Гипотеза H0​: β1​=β2​=β3​=0 — все коэффициенты при x равны нулю → модель не имеет значения

k = 3 (число независимых переменных в нашем уравнении)



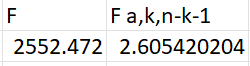


Рисунок 20 – Результаты расчёта F-критерия

Так как F > F a,k,n-k-1, гипотеза о незначимости регрессии отвергается.

# Дисперсионный анализ

Однофакторный дисперсионный анализ:

Данные остались прежними (см. пункт Регрессионный анализ).

Часы были разбиты на несколько интервалов (подкатегорий): 0<=y<8, 8<=y<10, 10<=y<13, 13<=y<16, 16<=y<=23.

Градации фактора были выбраны по гистограмме (см. рисунок 12).

Правила и условия применения дисперсионного анализа: согласованность выборок с нормальным распределением, одинаковая дисперсия, примерно равное количество значений в каждой подкатегории.

Распределение выборок не соответствует нормальному распределению.

Значения дисперсии на всех пяти интервалах сильно отличаются.

В каждой подкатегории количество значений сильно различается.

Гипотеза H0: средние значения числа регистраций во всех группах равны (фактор часы не оказывает влияния на число регистраций).

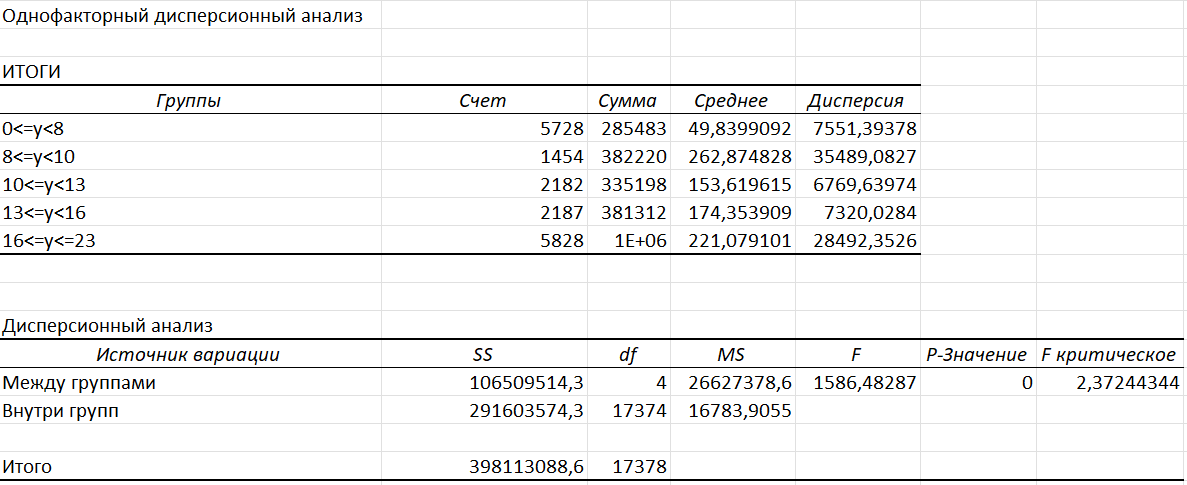


Рисунок 21 - Однофакторный дисперсионный анализ

Гипотеза H0 отклоняется, так как F(1586,5) > F(2,4) критическое. Значит, значение часов влияет на число регистраций.

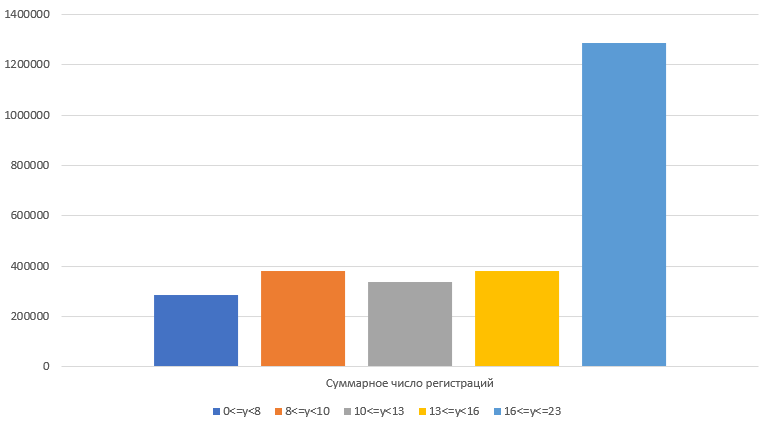


Рисунок 22 – диаграмма зависимости суммарного числа регистраций от градации фактора часы

По диаграмме мы видим, что в вечерние часы (с 16 до 23 включительно) больше всего людей предпочитают кататься на велосипедах, а в ночное время не так много желающих, несмотря на то, что в этих двух градациях фактора (от 16 до 23 включительно и от 0 до 8) примерно одинаковое количество значений (5828 и 5728 соответственно), а в остальных группах значений в несколько раз меньше.

Анализ результатов регрессионного анализа (таблица «Дисперсионный анализ»):

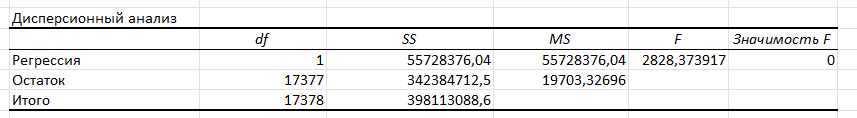


Рисунок 23 - результаты регрессионного анализа (таблица «Дисперсионный анализ»)

Значения из однофакторного дисперсионного анализа и из результатов регрессионного анализа (таблица «Дисперсионный анализ») различаются.

В регрессионном анализе (таблица «Дисперсионный анализ») проверяется значимость коэффициентов. Например, коэффициент детерминации, который используется для оценки качества построенной модели, вычисляется следующим образом:

R² = 1 – (Σ(y\_i – ŷ\_i)² / Σ(y\_i – ȳ)²),

где

* y\_i — фактические значения;
* ŷ\_i — значения, предсказанные моделью;
* ȳ — среднее значение зависимой переменной.

В самом дисперсионном анализе данные сравниваются, будучи предварительно разбитыми на подкатегории (категориальная переменная). Также проверяется гипотеза о равенстве средних значений в группах.

Двухфакторный дисперсионный анализ: были взяты два фактора – temp (температура) и hr(часы).

Значение температуры было разбито на следующие подкатегории: 0-0.25, 0.26-0.32, 0.33-0.58, 0.59-0.77, 0.78-1.0.

Значение часов было разбито на те же категории (см. подпункт Однофакторный дисперсионный анализ).

Условие применения дисперсионного анализа было нарушено, так как в каждой подкатегории количество значений сильно различается.

Проверяем три нулевых гипотезы:

1. Об отсутствии эффекта фактора temp (температура). Средние значения выборок, относящиеся к различным уровням фактора temp (температура), не отличаются статистически значимо.
2. Об отсутствии эффекта фактора hr(часы). Средние значения выборок, относящиеся к различным уровням фактора hr(часы), не отличаются статистически значимо.
3. Об отсутствии эффекта взаимодействия факторов temp (температура) и hr(часы). Факторы temp (температура) и hr(часы) считаются взаимодействующими, если эффект фактора temp (температура) зависит от уровня фактора hr(часы).

Двухфакторный дисперсионный анализ имеет две разновидности: без повторений и с повторениями. В первом случае каждому уровню фактора соответствует только одна выборка данных, во втором – определенным уровням факторов соответствует n выборок.

В данном случае был выбран тип двухфакторного анализа с повторениями.

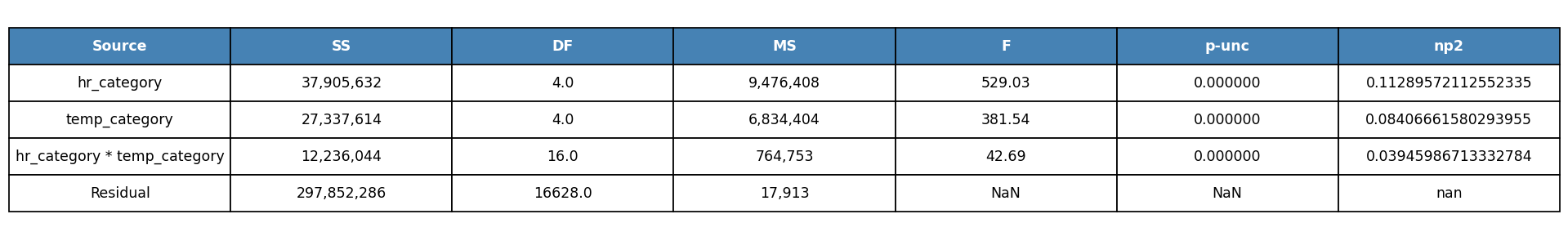


Рисунок 25 – Двухфакторный дисперсионный анализ с повторениями

Критерий значимости: 0.05

Все p-значения < 0.000001, что означает, что все три эффекта статистически значимы — мы отвергаем нулевую гипотезу для каждого из них.

То есть, количество регистраций зависит от времени суток и от температуры, также эффекта фактора температура зависит от времени суток и наоборот.

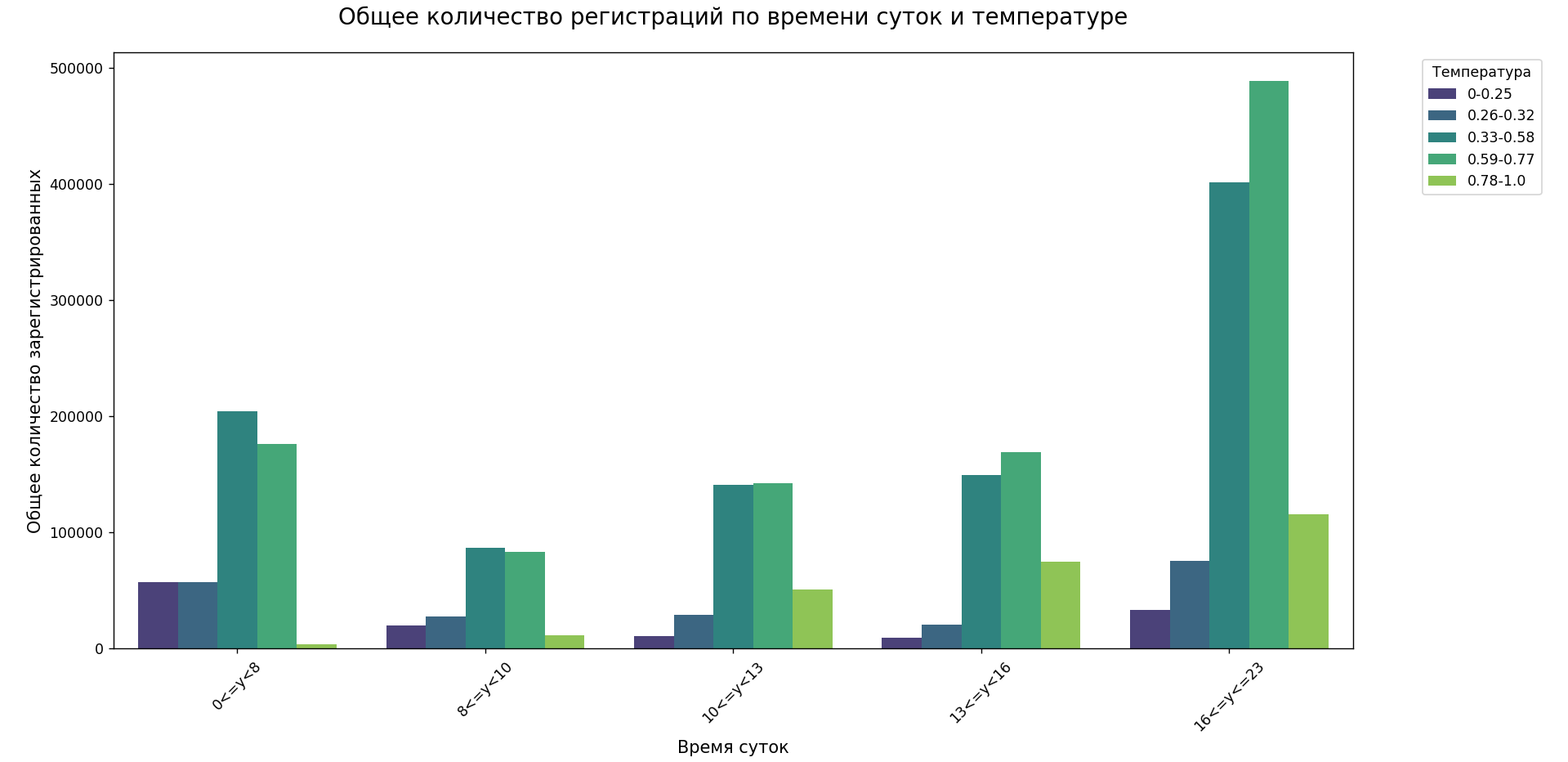


Рисунок 26 – диаграмма зависимости числа регистраций от градации факторов часы и температура

По диаграмме мы видим, что при двух градациях фактора температуры: 0,33-0,58(от 7,51ᵒС до 19,26ᵒС) и 0,59-0,77 (от 19,73ᵒС до 28,19ᵒС) - самое большое число регистраций. Значит, самая благоприятная температура для катания на велосипеде это от 7,51ᵒС до 28,19ᵒС. В сильно жаркую погоду (больше 29ᵒС) и в сильно прохладную погоду (меньше 7ᵒС) не очень комфортно кататься.

Анализируя распределение значений относительно градации фактора часы, мы делаем почти такой же вывод, как и в пункте однофакторный дисперсионный анализ: в вечерние часы (с 16 до 23 включительно) больше всего людей предпочитают кататься на велосипедах. Поэтому практически все столбцы в данном блоке (все градации температуры с 16 до 23 часов включительно) выше, чем в другое время суток (кроме столбца с температурой 0-0,25(от -8ᵒС до 3,75ᵒС)).