МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

**Национальный исследовательский ядерный университет «МИФИ»**

# 

**Институт интеллектуальных кибернетических систем**

**КАФЕДРА КИБЕРНЕТИКИ**

### БДЗ

**по курсу "Математическая статистика"**

**студента группы Б21-514**

**\_\_\_Шамаева Сергея Денисовича\_\_\_\_**

#### Вариант № \_19\_

**Оценка: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

##### Подпись: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2023 г.

ОТЧЕТ № 1

по теме «Проверка статистических гипотез»

Вариант № 19

ФИО студента Шамаева Сергея Денисовича группа Б21-514

Оценка: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Подпись:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Результаты статистических тестов:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № задания | Проверяемая гипотеза *H*0 | Критерий | Статистическое решение  (α = 0.1) | Вывод |
| 4.1 |  | Хи-квадрат |  |  |
| 4.2 |  | Харке-Бера |  |  |
| 5.1 |  | знаков |  |  |
| 5.2 |  | Хи-квадрат |  |  |

Выводы:

|  |
| --- |
| В результате проведённого в п.4 статистического анализа обнаружено, что |
| В результате проведённого в п.5 статистического анализа обнаружено, что |

ОТЧЕТ № 2

по теме «Анализ статистических взаимосвязей»

Вариант №\_\_\_\_\_\_\_\_

ФИО студента \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ группа \_\_\_\_\_\_\_\_

Оценка: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Подпись:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Результаты статистических тестов:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № задания | Проверяемая гипотеза *H*0 | Критерий | Статистическое решение  (α = 0.1) | Вывод |
| 6 |  | Хи-квадрат |  |  |
| 7 |  | ANOVA |  |  |

Выводы:

|  |
| --- |
| В результате проведённого в п.6 статистического анализа обнаружено, что |
| В результате проведённого в п.7 статистического анализа обнаружено, что |

ОТЧЕТ № 3

по теме «Основы регрессионного анализа»

Вариант №\_\_\_\_\_\_\_\_

ФИО студента \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ группа \_\_\_\_\_\_\_\_

Оценка: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Подпись:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Сводная таблица свойств различных регрессионных моделей:**

|  |  |  |  |
| --- | --- | --- | --- |
| Свойство | Простейшая линейная модель | Линейная модель с квадратичным членом | Множественная линейная модель |
| Точность |  |  |  |
| Значимость |  |  |  |
| Адекватность |  |  |  |
| Степень тесноты связи |  |  |  |

Выводы:

|  |
| --- |
| В результате проведённого в п.8 статистического анализа обнаружено, что |
| В результате проведённого в п.9 статистического анализа обнаружено, что |

1. Описательные статистики

*1.1. Выборочные характеристики*

Анализируемый признак 1 – C8

Анализируемый признак 2 – C10

Анализируемый признак 3 – C13

## а) Привести формулы расчёта выборочных характеристик

|  |  |
| --- | --- |
| Выборочная хар-ка | Формула расчета |
| Объём выборки | n |
| Среднее |  |
| Выборочная дисперсия |  |
| Выборочное среднеквадратическое отклонение |  |
| Выборочный коэффициент асимметрии |  |
| Выборочный эксцесс |  |

*б) Рассчитать выборочные характеристики*

|  |  |  |  |
| --- | --- | --- | --- |
| Выборочная хар-ка | Признак 1 | Признак 2 | Признак 3 |
| Среднее | 12.789 | 242.461 | 189.892 |
| Выборочная дисперсия | 28.411 | 17421.787 | 33489.294 |
| Выборочное среднеквадратическое отклонение | 5.330 | 131.992 | 183.001 |
| Выборочный коэффициент асимметрии | 1.153 | 1.480 | 3.564 |
| Выборочный эксцесс | 2.483 | 3.412 | 17.209 |

*1.2. Группировка и гистограммы частот*

Анализируемый признак – C8

Объём выборки – 315

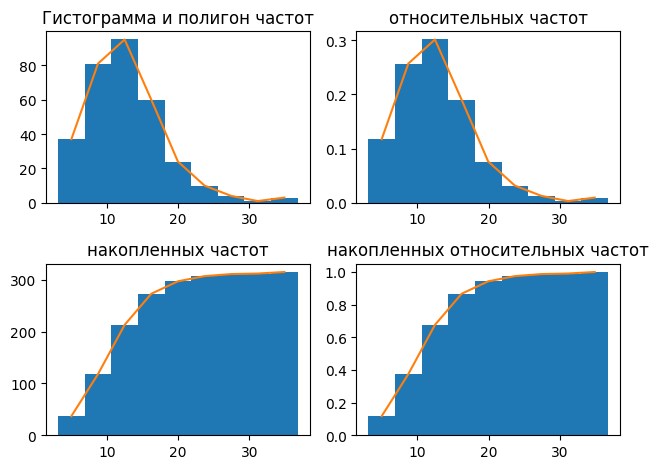
*а) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| 9 | Согласно формуле Стерджесса | 3.744444444444444 |

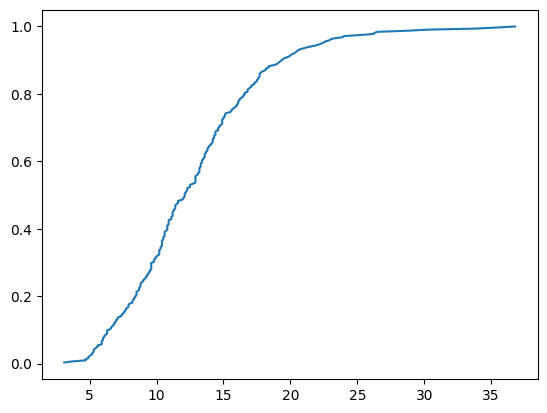
*б) Построить таблицу частот*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Накопл. частота | Относит. накопл. частота |
| 1 | 3.10 | 6.84 | 37 | 0.117 | 37 | 0.117 |
| 2 | 6.84 | 10.59 | 81 | 0.257 | 118 | 0.375 |
| *3* | 10.59 | 14.33 | 95 | 0.302 | 213 | 0.676 |
| 4 | 14.33 | 18.08 | 60 | 0.190 | 273 | 0.867 |
| 5 | 18.08 | 21.82 | 24 | 0.076 | 297 | 0.943 |
| 6 | 21.82 | 25.56 | 10 | 0.032 | 307 | 0.975 |
| 7 | 25.56 | 29.31 | 4 | 0.013 | 311 | 0.987 |
| 8 | 29.31 | 33.05 | 1 | 0.003 | 312 | 0.990 |
| 9 | 33.05 | 36.80 | 3 | 0.009 | 315 | 1 |

*в) Построить гистограммы частот и полигоны частот*



*г) Построить график эмпирической функции распределения*



**2. Интервальные оценки**

*2.1. Доверительные интервалы для мат. ожидания*

Анализируемый признак – С8

Объём выборки – 315

Оцениваемый параметр – мат. ожидание

*а) Привести формулы расчёта доверительных интервалов*

|  |  |
| --- | --- |
| Граница доверительного интервала | Формула расчета |
| Нижняя граница |  |
| Верхняя граница |  |

*б) Рассчитать доверительные интервалы*

|  |  |  |  |
| --- | --- | --- | --- |
| Граница доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| Нижняя граница | 11.790 | 12.197 | 12.293 |
| Верхняя граница | 13.786 | 13.379 | 13.284 |

*2.2. Доверительные интервалы для дисперсии*

Анализируемый признак – С8

Объём выборки – 315

Оцениваемый параметр – дисперсия

*а) Привести формулы расчёта доверительных интервалов*

|  |  |
| --- | --- |
| Граница доверительного интервала | Формула расчета |
| Нижняя граница |  |
| Верхняя граница |  |

*б) Рассчитать доверительные интервалы*

|  |  |  |  |
| --- | --- | --- | --- |
| Граница доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| Нижняя граница | 22.133 | 24.442 | 25.036 |
| Верхняя граница | 37.478 | 33.436 | 32.564 |

*2.3. Доверительные интервалы для разности мат. ожиданий*

Анализируемый признак 1 – С11

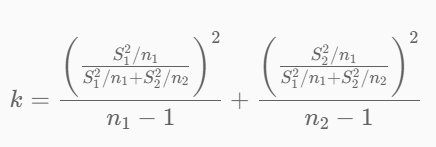
Анализируемый признак 2 – С12

Объёмы выборок – 315

Оцениваемый параметр – разность мат. ожиданий

*а) Привести формулы расчёта доверительных интервалов*

|  |  |
| --- | --- |
| Граница доверительного интервала | Формула расчета |
| Нижняя граница |  |
| Верхняя граница |  |



*б) Рассчитать доверительные интервалы*

|  |  |  |  |
| --- | --- | --- | --- |
| Граница доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| Нижняя граница | 1120.416 | 1176.586 | 1205.131 |
| Верхняя граница | 1585.361 | 1529.190 | 1500.646 |

*2.4. Доверительные интервалы для отношения дисперсий*

Анализируемый признак 1 – С11

Анализируемый признак 2 – С12

Объёмы выборок – 315

Оцениваемый параметр – отношение дисперсий

*а) Привести формулы расчёта доверительных интервалов*

|  |  |
| --- | --- |
| Граница доверительного интервала | Формула расчета |
| Нижняя граница |  |
| Верхняя граница |  |

*б) Рассчитать доверительные интервалы*

|  |  |  |  |
| --- | --- | --- | --- |
| Граница доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| Нижняя граница | 4.309 | 5.012 | 5.194 |
| Верхняя граница | 9.081 | 7.807 | 7.53 |

**3. Проверка статистических гипотез о математических ожиданиях и дисперсиях**

*3.1. Проверка статистических гипотез о математических ожиданиях*

Анализируемый признак – C8

Объём выборки – 315

Статистическая гипотеза – 

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

|  |  |
| --- | --- |
|  | Выражение |
| Формула расчета статистики критерия |  |
| Закон распределения статистики критерия при условии истинности основной гипотезы | T(n-1) |
| Формулы расчета критических точек |  |
| Формула расчета *p-value* |  |

*б) Выбрать произвольные значения m0 и проверить статистические гипотезы*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *m*0 | Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 5 | 0.1 | 25.934 | 4.7e-80 | H0 отвергается | m != 5 |
| 13 | 0.1 | -0.704 | 0.481 | H0 принимается | m можно принять 13 при данном уровне значимости |
| 20 | 0.1 | -24.012 | 4.6e-73 | H0 отвергается | m != 20 |

*3.2. Проверка статистических гипотез о дисперсиях*

Анализируемый признак – C8

Объём выборки – 315

Статистическая гипотеза – 

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

|  |  |
| --- | --- |
|  | Выражение |
| Формула расчета статистики критерия |  |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |
| Формулы расчета критических точек |  |
| Формула расчета *p-value* |  |

*б) Выбрать произвольные значения σ0 и проверить статистические гипотезы*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| σ0 | Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 2 | 0.1 | 2230.259 | 0.000 | H0 отвергается | σ != 2 |
| 5 | 0.1 | 356.841 | 0.096 | H0 принимается | σ можно принять 5 при данном уровне значимости |
| 20 | 0.1 | 22.302 | 7.0e-119 | H0 отвергается | σ != 20 |

*3.3. Проверка статистических гипотез о равенстве математических ожиданий*

Анализируемый признак 1 – С11

Анализируемый признак 2 – С12

Объёмы выборок – 315

Статистическая гипотеза – 

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

|  |  |
| --- | --- |
|  | Выражение |
| Формула расчета статистики критерия |  |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |
| Формулы расчета критических точек |  |
| Формула расчета *p-value* |  |

*б) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 15.126 | 1.9e-41 | H0 отвергается | Нельзя считать  m1 != m2 |
| 0.05 | H0 отвергается | Нельзя считать  m1 != m2 |
| 0.1 | H0 отвергается | Нельзя считать  m1 != m2 |

*3.4. Проверка статистических гипотез о равенстве дисперсий*

Анализируемый признак 1 – С11

Анализируемый признак 2 – С12

Объёмы выборок – 315

Статистическая гипотеза – 

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

|  |  |
| --- | --- |
|  | Выражение |
| Формула расчета статистики критерия |  |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |
| Формулы расчета критических точек |  |
| Формула расчета *p-value* |  |

*б) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 6.255 | 2.2e-16 | H0 отвергается | Нельзя считать  σ1 != σ2 |
| 0.05 | H0 отвергается | Нельзя считать  σ1 != σ2 |
| 0.1 | H0 отвергается | Нельзя считать  σ1 != σ2 |

**4. Критерии согласия**

Анализируемый признак – C8

Объём выборки – 315

*4.1. Критерий хи-квадрат*

Теоретическое распределение – нормальное.

Статистическая гипотеза – 

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  |  |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  |  |
| Формула расчета *p-value* |  |  |

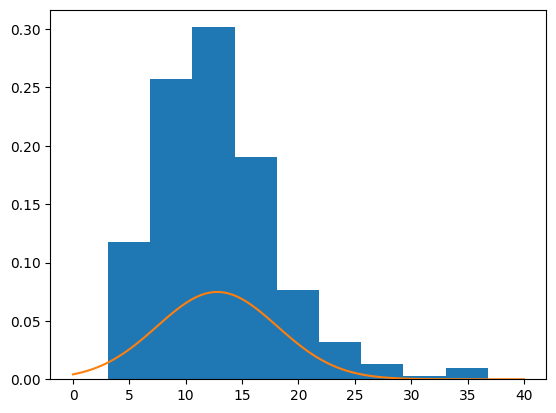
*б) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| 9 | Согласно формуле Стерджесса | 3.744444444444444 |

*в) Построить таблицу частот*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Вероятность попадания в интервал при условии истинности основной гипотезы |
| 1 | 3.10 | 6.84 | 37 | 0.117 | 0.132 |
| 2 | 6.84 | 10.59 | 81 | 0.257 | 0.207 |
| *3* | 10.59 | 14.33 | 95 | 0.302 | 0.274 |
| 4 | 14.33 | 18.08 | 60 | 0.190 | 0.225 |
| 5 | 18.08 | 21.82 | 24 | 0.076 | 0.115 |
| 6 | 21.82 | 25.56 | 10 | 0.032 | 0.036 |
| 7 | 25.56 | 29.31 | 4 | 0.013 | 0.007 |
| 8 | 29.31 | 33.05 | 1 | 0.003 | 0.0008 |
| 9 | 33.05 | 36.80 | 3 | 0.009 | 7.1e-05 |

г) Построить гистограмму относительных частот и функцию плотности теоретического распределения на одном графике



*д) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 406.978 | 8.8e-85 | H0 отвергается | Нельзя считать распределение нормальным |
| 0.05 | 406.978 | 8.8e-85 | H0 отвергается | Нельзя считать распределение нормальным |
| 0.1 | 406.978 | 8.8e-85 | H0 отвергается | Нельзя считать распределение нормальным |

*4.2. Проверка гипотезы о нормальности на основе коэффициента асимметрии и эксцесса (критерий Харке-Бера)*

Статистическая гипотеза – 

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | S – коэфф асимметрии, K–коэфф эксцесса |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  |  |
| Формула расчета *p-value* |  |  |

*б) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 146.323 | 1.6e-32 | H0 отвергается | Нельзя считать распределение нормальным |
| 0.05 | H0 отвергается | Нельзя считать распределение нормальным |
| 0.1 | H0 отвергается | Нельзя считать распределение нормальным |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.4 статистического анализа обнаружено, что C8 (Граммы клетчатки, потребляемой в день) нельзя считать распределение нормальным. |

**5. Проверка однородности выборок**

Анализируемый признак 1 – С11

Анализируемый признак 2 – С12

Объёмы выборок – 315

*5.1 Критерий знаков*

Статистическая гипотеза – 

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  |  |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  |  |
| Формула расчета *p-value* |  |  |

*б) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 124.5 | 1.9e-50 | H0 отвергается | Признаки не однородны |
| 0.05 | H0 отвергается | Признаки не однородны |
| 0.1 | H0 отвергается | Признаки не однородны |

*5.2. Критерий хи-квадрат*

Статистическая гипотеза – 

а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  |  |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  |  |
| Формула расчета *p-value* |  |  |

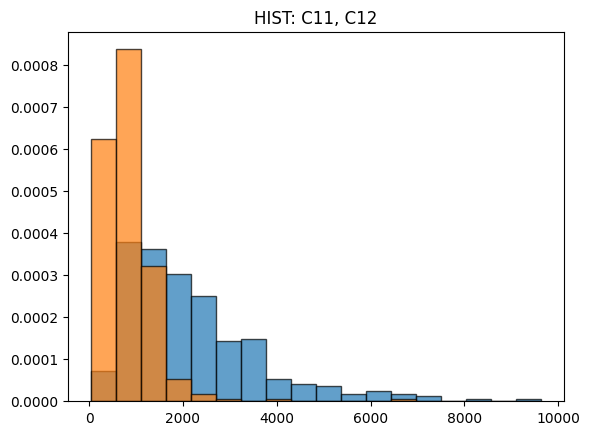
*б) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| 12 | По заполненности | 534 |

*в) Построить таблицу частот*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота признака 1 | Частота признака 2 | Относит. частота признака 1 | Относит. частота признака 2 |
| 1 | 30 | 564 | 12 | 105 | 0.038 | 0.333 |
| 2 | 564 | 1098 | 64 | 141 | 0.203 | 0.448 |
| 3 | 1098 | 1632 | 61 | 54 | 0.194 | 0.171 |
| 4 | 1632 | 2166 | 51 | 9 | 0.162 | 0.029 |
| 5 | 2166 | 2700 | 42 | 3 | 0.133 | 0.01 |
| 6 | 2700 | 3234 | 24 | 1 | 0.076 | 0.003 |
| 7 | 3234 | 3768 | 25 | 0 | 0.079 | 0.0 |
| 8 | 3768 | 4302 | 9 | 1 | 0.029 | 0.003 |
| 9 | 4302 | 4836 | 7 | 0 | 0.022 | 0.0 |
| 10 | 4836 | 5370 | 6 | 0 | 0.019 | 0.0 |
| 11 | 5370 | 5904 | 3 | 0 | 0.01 | 0.0 |
| 12 | 5904 | 9642 | 11 | 1 | 0.035 | 0.003 |

*г) Построить гистограммы относительных частот на одном графике*



*д) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 243.364 | 6.8e-46 | H0 отвергается | Признаки не однородны |
| 0.05 | H0 отвергается | Признаки не однородны |
| 0.1 | H0 отвергается | Признаки не однородны |

Вывод (в терминах предметной области)

|  |
| --- |
| В результате проведённого в п.5 статистического анализа обнаружено, что признаки С11(Потребляемый с пищей бета-каротин (мкг в день)) и С12 (Потребляемый с пищей ретинол (мкг в день)) не однородны. |

**6. Таблицы сопряжённости**

Факторный признак *x* – C2

Результативный признак *y* – C5

Объёмы выборок – 315

Статистическая гипотеза –

*а) Указать формулы расчёта показателей, используемых при проверке статистических гипотез*

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | n – наблюдаемые частоты, m - теоретические |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  |  |
| Формула расчета *p-value* |  |  |

*б) Построить эмпирическую таблицу сопряжённости*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *x y* | No | Not often | Often | Σ |
| Female | 87 | 77 | 109 | 273 |
| Male | 24 | 5 | 13 | 42 |
| Σ | 111 | 82 | 122 | 315 |

*в) Построить теоретическую таблицу сопряжённости*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *x y* | No | Not often | Often | Σ |
| Female | 96.2 | 71.067 | 105.733 | 273 |
| Male | 14.8 | 10.933 | 16.267 | 42 |
| Σ | 111 | 82 | 122 | 315 |

*г) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 11.071 | 0.0039 | H0 принимается | Можно считать, что признаки независимы |
| 0.05 | H0 отклоняется | Нельзя считать, что признаки независимы |
| 0.1 | H0 отклоняется | Нельзя считать, что признаки независимы |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.6 статистического анализа обнаружено, что распределение признака С2 (пол) слабо зависит от признака С5 (употребление витаминов), при уровне значимости 0.01 можно считать что зависимости нет, а при уровнях 0.05 и 0.01 можно предполагать что статистическая связь имеется. |

**7. Дисперсионный анализ**

Факторный признак *x* – С3

Результативный признак *y* – С14

Число вариантов факторного признака – 3

Объёмы выборок – 315

Статистическая гипотеза –



*а) Рассчитать групповые выборочные характеристики*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Вариант факторного признака | Объём выборки | Групповые средние | Групповые дисперсии |
| 1 | Current Smoker | 43 | 563.069 | 42674.399 |
| 2 | Former | 115 | 644.243 | 53438.466 |
| 3 | Never | 157 | 583.305 | 35209.931 |

б) Привести формулы расчёта показателей вариации, используемых в дисперсионном анализе

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак |  | K – 1 |  |
| Остаточные признаки |  | N – K |  |
| Все признаки |  | N - 1 |  |

*в) Рассчитать показатели вариации, используемые в дисперсионном анализе*

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак | 1031.932 | 2 | 162529.436 |
| Остаточные признаки | 42883.752 | 312 | 43296.096 |
| Все признаки | 43637.319 | 314 | 44056.908 |

г) Проверить правило сложения дисперсий

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель | *Dмежгр* | *Dвнутригр* | *Dобщ* | *Dмежгр* + *Dвнутригр* |
| Значение | 1031.932 | 42883.752 | 43637.319 | 43915.684 |

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Эмпирический коэффициент детерминации |  | 0.024 |
| Эмпирическое корреляционное отношение |  | 0.155 |

*е) Охарактеризовать тип связи между факторным и результативным признаками*

|  |
| --- |
| практически отсутствует |

ж) Указать формулы расчёта показателей, используемых при проверке статистической гипотезы дисперсионного анализа

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | K – количество вариантов факторного признака, n–размер. Db – межгрупповая дисперсия, Dw – внутригрупповая дисперсия |
| Закон распределения статистики критерия при условии истинности основной гипотезы | F(K–1, n–K) |  |
| Формула расчета критической точки |  |  |
| Формула расчета *p-value* |  |  |

*з) Проверить статистическую гипотезу дисперсионного анализа*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 3.790 | 0.023 | H0 принимается | Можно считать, что признаки независимы |
| 0.05 | H0 отклоняется | Нельзя считать, что признаки независимы |
| 0.1 | H0 отклоняется | Нельзя считать, что признаки независимы |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.7 статистического анализа обнаружено, что распределение признака С3 (статус курения) слабо зависит от признака С14 (Ретинол плазмы (нг/мл)), при уровне значимости 0.01 можно считать что зависимости нет, а при уровнях 0.05 и 0.01 можно предполагать что статистическая связь имеется. |

**8. Корреляционный анализ**

*8.1. Расчёт парных коэффициентов корреляции*

Анализируемый признак 1 – C11

Анализируемый признак 2 –C12

Объёмы выборок – 315

*а) Рассчитать точечные оценки коэффициентов корреляции*

|  |  |  |
| --- | --- | --- |
|  | Формула расчета | Значение |
| Линейный коэффициент корреляции |  | 0.052 |
| Ранговый коэффициент корреляции по Спирмену |  | 0.196 |
| Ранговый коэффициент корреляции по Кендаллу |  | 0.132 |

б) Привести формулы расчёта доверительного интервала для линейного коэффициента корреляции

|  |  |
| --- | --- |
| Граница доверительного интервала | Формула расчета |
| Нижняя граница |  |
| Верхняя граница |  |

*в) Рассчитать доверительные интервалы для линейного коэффициента корреляции*

|  |  |  |  |
| --- | --- | --- | --- |
| Граница доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| Нижняя граница | -0.091 | -0.057 | -0.039 |
| Верхняя граница | 0.197 | 0.163 | 0.145 |

г) Указать формулы расчёта показателей, используемых при проверке значимости коэффициентов корреляции

|  |  |  |
| --- | --- | --- |
| Статистическая гипотеза | Формула расчета статистики критерия | Закон распределения статистики критерия при условии истинности основной гипотезы |
|  |  |  |
|  |  |  |
|  |  |  |

*д) Проверить значимость коэффициентов корреляции*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Статистическая гипотеза | Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
|  | 0.1 | 0.936 | 0.349 | H0 принимается | Можно считать, что признаки независимы |
|  | 0.1 | 3.541 | 0.000 | H0 отклоняется | Нельзя считать, что признаки независимы |
|  | 0.1 | 3.517 | 0.000 | H0 отклоняется | Нельзя считать, что признаки независимы |

*8.2. Расчёт множественных коэффициентов корреляции*

Анализируемый признак 1 – C8

Анализируемый признак 2 – C10

Анализируемый признак 3 – C13

Объёмы выборок – 315

*а) Рассчитать матрицу ранговых коэффициентов корреляции по Кендаллу*

|  |  |  |  |
| --- | --- | --- | --- |
| Признак  Признак | C8 | C10 | C13 |
| C8 | 1 | 0.148 | 0.128 |
| C10 | 0.148 | 1 | -0.092 |
| C13 | 0.128 | -0.092 | 1 |

*б) Рассчитать матрицу значений p-value для ранговых коэффициентов корреляции по Кендаллу (статистическая гипотеза* *)*

|  |  |  |  |
| --- | --- | --- | --- |
| Признак  Признак | C8 | C10 | C13 |
| C8 | – | 8.7e-05 | 0.0007 |
| C10 | 8.7e-05 | – | 0.014 |
| C13 | 0.0007 | 0.014 | – |

*в) Рассчитать точечную оценку коэффициента конкордации*

|  |  |  |
| --- | --- | --- |
|  | Формула расчета | Значение |
| Коэффициент конкордации |  | 0.393 |

г) Указать формулы расчёта показателей, используемых при проверке значимости коэффициента конкордации

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | N – размер выборок, k–количество выборок, W–выборочный коэффициент конкордации |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  |  |
| Формула расчета *p-value* |  |  |

*д) Проверить значимость коэффициента конкордации*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 248.04 | 0 | H0 отклоняется | Нельзя считать, что признаки независимы в совокупности |
| 0.05 | H0 отклоняется | Нельзя считать, что признаки независимы в совокупности |
| 0.1 | H0 отклоняется | Нельзя считать, что признаки независимы в совокупности |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.8 статистического анализа обнаружено, что признаки С8 (Граммы клетчатки, потребляемой в день), С10 (Потребляемый холестерин (мг в день)), С13(Бета-каротин плазмы (нг/мл)) имеют какую-то взаимосвязь в совокупности. |

**9. Регрессионный анализ**

*9.1 Простейшая линейная регрессионная модель*

Факторный признак *x* – С6

Результативный признак *y* – С14

Уравнение регрессии – 

*9.1.1. Точечные оценки линейной регрессионной модели*

*а) Рассчитать точечные оценки параметров линейной регрессионной модели*

|  |  |  |
| --- | --- | --- |
| Параметр | Формула расчета | Значение |
| β0 |  | 643.242 |
| β1 |  | -0.022 |

*б) Записать точечную оценку уравнения регрессии*

|  |
| --- |
| f(x) = 643.2421313446065 \* x - 0.02251498706688208 |

в) Привести формулы расчёта показателей вариации, используемых в регрессионном анализе

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак |  | k–1 |  |
| Остаточные признаки |  | n–k |  |
| Все признаки |  | n–1 |  |

*г) Рассчитать показатели вариации, используемые в регрессионном анализе*

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак | 233.897 | 1 | 73677.444 |
| Остаточные признаки | 43264.891 | 313 | 43541.344 |
| Все признаки | 43637.319 | 314 | 43776.291 |

д) Проверить правило сложения дисперсий

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель | *Dрегр* | *Dост* | *Dобщ* | *Dрегр* + *Dост* |
| Значение | 233.897 | 43264.891 | 43637.319 | 43498.787 |

е) Рассчитать показатели тесноты связи между факторным и результативным признаками

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Коэффициент детерминации |  | 0.005 |
| Корреляционное отношение |  | 0.073 |

ж) Охарактеризовать тип связи между факторным и результативным признаками, определяемой рассчитанной линейной регрессией

|  |
| --- |
| практически отсутствует |

*9.1.2. Интервальные оценки линейной регрессионной модели*

а) Привести формулы расчёта доверительных интервалов для параметров линейной регрессионной модели

|  |  |  |
| --- | --- | --- |
| Параметр | Границы доверительного интервала | Формула расчета |
| β0 | Нижняя граница |  |
| Верхняя граница |  |
| β1 | Нижняя граница |  |
| Верхняя граница |  |

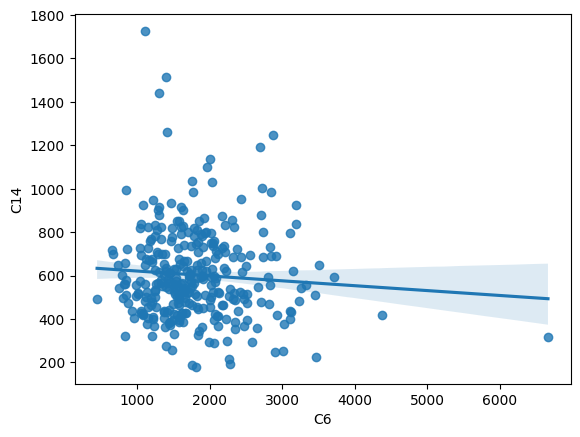
б) Рассчитать доверительные интервалы для параметров линейной регрессионной модели

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Параметр | Границы доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| β0 | Нижняя граница | 557.219 | 577.933 | 588.483 |
| Верхняя граница | 729.264 | 708.550 | 698.001 |
| β1 | Нижняя граница | 643.197 | 643.208 | 643.213 |
| Верхняя граница | 643.286 | 643.276 | 643.270 |

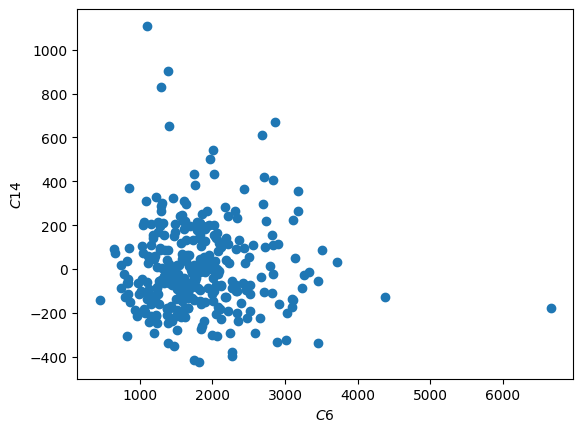
*в) Привести формулы расчёта доверительного интервала для значений регрессии f(x)*

|  |  |
| --- | --- |
| Границы доверительного интервала | Формула расчета |
| Нижняя граница *flow*(*x*) |  |
| Верхняя граница *fhigh*(*x*) |  |

*г) Построить диаграмму рассеяния признаков x и y*. *Нанести на диаграмму функцию регрессии f(x), а также нижние и верхние границы линии регрессии flow(x) и fhigh(x) на уровне значимости α = 0.1*



*д) Построить график остатков *



*9.1.3. Проверка значимости линейной регрессионной модели*

Статистическая гипотеза – 

а) Указать формулы расчёта показателей, используемых при проверке значимости линейной регрессионной модели

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | RY|X – коэффициент детерминации |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  |  |
| Формула расчета *p-value* |  |  |

## б) Проверить значимость линейной регрессионной модели

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 1.692 | 0.194 | H0 принимается | Можно считать, что  b1 != 0 |
| 0.05 | H0 принимается | Можно считать, что  b1 != 0 |
| 0.1 | H0 принимается | Можно считать, что  b1 != 0 |

*9.2 Линейная регрессионная модель общего вида*

Факторный признак *x* – C6

Результативный признак *y* – C14

Уравнение регрессии – квадратичное по *x*: 

*9.2.1. Точечные оценки линейной регрессионной модели*

*а) Рассчитать точечные оценки параметров линейной регрессионной модели*

|  |  |  |
| --- | --- | --- |
| Параметр | Формула расчета | Значение |
| β0 |  | 5.911 e+02 |
| β1 | β = | 2.838 e-02 |
| β2 |  | -1.065 e-05 |

*б) Записать точечную оценку уравнения регрессии*

|  |
| --- |
| *f*(*x*) = 5.91101645e+02 + 2.83858107e-02 \* x - 1.06551135e-05 \* x \*\* 2 |

*в) Рассчитать показатели вариации, используемые в регрессионном анализе*

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак | 401.697 | 2 | 253069.181 |
| Остаточные признаки | 43097.090 | 312 | 43511.485 |
| Все признаки | 43637.319 | 314 | 43776.291 |

г) Проверить правило сложения дисперсий

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель | *Dрегр* | *Dост* | *Dобщ* | *Dрегр* + *Dост* |
| Значение | 401.697 | 43097.0901 | 43637.319 | 43498.787 |

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Коэффициент детерминации |  | 0.009 |
| Корреляционное отношение |  | 0.095 |

е) Охарактеризовать тип связи между факторным и результативным признаками, определяемой рассчитанной линейной регрессией

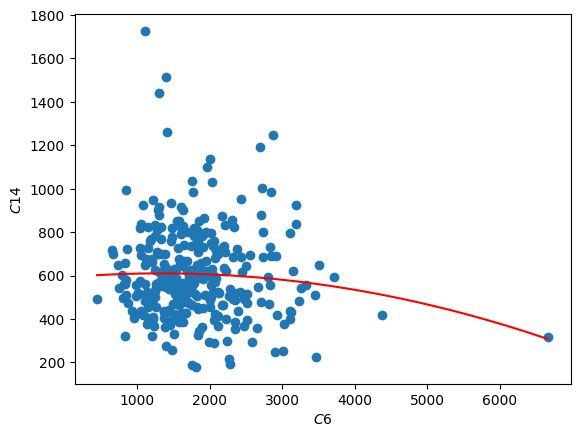
|  |
| --- |
| практически отсутствует |

*9.2.2. Интервальные оценки линейной регрессионной модели*

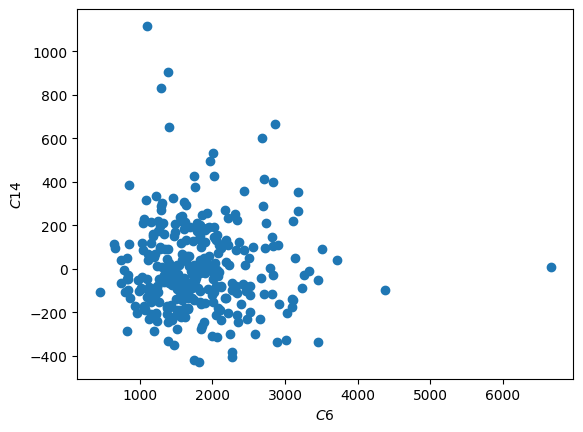
*а) Привести формулы расчёта доверительного интервала для значений регрессии f(x)*

|  |  |
| --- | --- |
| Границы доверительного интервала | Формула расчета |
| Нижняя граница *flow*(*x*) |  |
| Верхняя граница *fhigh*(*x*) |  |

*б) Построить диаграмму рассеяния признаков x и y. Нанести на диаграмму функцию регрессии f(x), а также нижние и верхние границы линии регрессии flow(x) и fhigh(x) на уровне значимости α = 0.1*



*в) Построить график остатков *



*9.2.3. Проверка значимости линейной регрессионной модели*

Статистическая гипотеза – 

а) Указать формулы расчёта показателей, используемых при проверке значимости линейной регрессионной модели

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  |  |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  |  |
| Формула расчета *p-value* |  |  |

*б)* *Проверить значимость линейной регрессионной модели*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 1.449 | 0.236 | H0 принимается | Можно считать, что  коэфф != 0 |
| 0.05 | H0 принимается | Можно считать, что  коэфф != 0 |
| 0.1 | H0 принимается | Можно считать, что  коэфф != 0 |

*9.3 Множественная линейная регрессионная модель*

Факторный признак 1 *x*1 – С6

Факторный признак 2 *x*2 – С12

Результативный признак *y* – С14

Уравнение регрессии – 

*а) Рассчитать точечные оценки параметров линейной регрессионной модели*

|  |  |  |
| --- | --- | --- |
| Параметр | Формула расчета | Значение |
| β0 |  | 6.461e+02 |
| β1 |  | -1.760-02 |
| β2 |  | -1.408e-02 |

*б) Записать точечную оценку уравнения регрессии*

|  |
| --- |
| f(x1, x2) = 6.46147981e+02 + -1.76059169e-02 \* x1 + -1.40813648e-02 \* x2 |

*в) Рассчитать показатели вариации, используемые в регрессионном анализе*

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак | 291.415 | 2 | 183591.712 |
| Остаточные признаки | 43207.372 | 312 | 43622.827 |
| Все признаки | 43637.319 | 314 | 43776.291 |

г) Проверить правило сложения дисперсий

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель | *Dрегр* | *Dост* | *Dобщ* | *Dрегр* + *Dост* |
| Значение | 291.415 | 43207.372 | 43637.319 | 43498.787 |

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Множественный коэффициент детерминации |  | 0.006 |
| Множественное корреляционное отношение |  | 0.081 |

е) Охарактеризовать тип связи между факторным и результативным признаками, определяемой рассчитанной линейной регрессией

|  |
| --- |
| практически отсутствует |

*9.4. Выводы*

*а) Сводная таблица показателей вариации для различных регрессионных моделей*

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Простейшая линейная модель | Линейная модель с квадратичным членом | Множественная линейная модель |
| Факторный признак | 233.897 | 401.697 | 401.697 |
| Остаточные признаки | 43264.891 | 43097.090 | 43097.090 |
| Все признаки | 43637.319 | 43637.319 | 43637.319 |

*б) Сводная таблица свойств различных регрессионных моделей*

|  |  |  |  |
| --- | --- | --- | --- |
| Свойство | Простейшая линейная модель | Линейная модель с квадратичным членом | Множественная линейная модель |
| Точность | не точная | не точная | не точная |
| Значимость | значима | значима | значима |
| Адекватность | адекватна | адекватна | адекватна |
| Степень тесноты связи | практически отсутствует | практически отсутствует | практически отсутствует |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.9 статистического анализа обнаружено, что для С6 и С14 нет подходящей линейной модели (так как признаки практически не имеют зависимости). |