МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

**Национальный исследовательский ядерный университет «МИФИ»**

# Изображение выглядит как коллекция картинок Автоматически созданное описание

**Институт интеллектуальных кибернетических систем**

**КАФЕДРА КИБЕРНЕТИКИ**

### БДЗ

**по курсу "Математическая статистика"**

**студента группы Б19-504**

**Сидорова Александра**

#### Вариант № 19

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

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

2021 г.

1. Описательные статистики

*1.1. Выборочные характеристики*

Анализируемый признак 1 – Grams of fiber consumed per day (C8)

Анализируемый признак 2 – Cholesterol consumed (mg per day) (C10)

Анализируемый признак 3 – Plasma beta-carotene (ng/ml) (C13)

## а) Привести формулы расчёта выборочных характеристик

|  |  |
| --- | --- |
| Выборочная хар-ка | Формула расчета |
| Объём выборки | 315 |
| 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 |  |  |  |
| Выборочная дисперсия |  |  |  |
| Выборочное среднеквадратическое отклонение |  |  |  |
| Выборочный коэффициент асимметрии |  |  |  |
| Выборочный эксцесс |  |  |  |

*1.2. Группировка и гистограммы частот*

Анализируемый признак – Grams of fiber consumed per day (C8)

Объём выборки – 315

*а) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
|  | По формуле Стерджесса |  |

*б) Построить таблицу частот*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Накопл. частота | Относит. накопл. частота |
| 1 | 3,1 | 6,8 | 35 | 0,11 | 35 | 0,11 |
| 2 | 6,8 | 10,5 | 80 | 0,25 | 115 | 0,36 |
| 3 | 10,5 | 14,2 | 91 | 0,29 | 206 | 0,65 |
| 4 | 14,2 | 17,9 | 66 | 0,21 | 272 | 0,86 |
| 5 | 17,9 | 21,6 | 24 | 0,08 | 296 | 0,94 |
| 6 | 21,6 | 25,3 | 11 | 0,035 | 307 | 0,975 |
| 7 | 25,3 | 29 | 4 | 0,01 | 311 | 0,985 |
| 8 | 29 | 32,7 | 1 | 0,005 | 312 | 0,99 |
| 9 | 32,7 | 36,8 | 3 | 0,01 | 315 | 1 |

*в) Построить гистограммы частот и полигоны частот*

|  |
| --- |
| Гистограмма и полигон частот |
| Гистограмма и полигон относительных частот |
| Гистограмма и полигон накопленных частот |
| Гистограмма и полигон накопленных относительных частот |

*г) Построить график эмпирической функции распределения*

|  |
| --- |
| Эмпирическая функция распределения |

**2. Интервальные оценки**

*2.1. Доверительные интервалы для мат. ожидания*

Анализируемый признак – Grams of fiber consumed per day (С8)

Объём выборки – 315

Оцениваемый параметр – m

*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 | 12,01 | 12,2 | 12,29 |
| Верхняя граница | 13,57 | 13,38 | 13,28 |

*2.2. Доверительные интервалы для дисперсии*

Анализируемый признак – Grams of fiber consumed per day (С8)

Объём выборки – 315

Оцениваемый параметр –

*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 | 23,33 | 24,44 | 25,04 |
| Верхняя граница | 35,23 | 33,44 | 32.56 |

*2.3. Доверительные интервалы для разности мат. ожиданий*

Анализируемый признак 1 – Dietary beta-carotene consumed (mcg per day) (С11)

Анализируемый признак 2 – Dietary retinol consumed (mcg per day) (С12)

Объёмы выборок – 315

Оцениваемый параметр –

*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 | 1120,37 | 1176,36 | 1204,87 |
| Верхняя граница | 1585,41 | 1529,42 | 1500,9 |

*2.4. Доверительные интервалы для отношения дисперсий*

Анализируемый признак 1 – Dietary beta-carotene consumed (mcg per day) (С11)

Анализируемый признак 2 – Dietary retinol consumed (mcg per day) (С12)

Объёмы выборок – 315

Оцениваемый параметр –

*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 | 4,7 | 5 | 5,2 |
| Верхняя граница | 8,4 | 7,8 | 7,5 |

**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 |  |
| 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 |
|  |  |  |  |  |  |
|  |  |  | 0.482 |  |  |
|  |  |  | 0 |  |  |

*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 |
|  |  |  |  |  |  |
| 5 |  |  |  |  |  |
|  |  |  |  |  |  |

*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 |
|  |  |  |  |  |
|  |  |  |
|  |  |  |

*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 |
|  |  |  |  |  |
|  |  |  |
|  |  |  |

**4. Критерии согласия**

Анализируемый признак – C8

Объём выборки – 315

*4.1. Критерий хи-квадрат*

Теоретическое распределение – нормальное.

Статистическая гипотеза – 

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

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | - число элементов выборки, принадлежащих интервалу  - вероятность попадания в каждый интервал |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  | k - число интервалов  r - число неизвестных параметров распределения |
| Формула расчета критической точки |  | Критическая область выбирается правосторонней |
| Формула расчета *p-value* |  |  |

*б) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
|  |  |  |

*в) Построить таблицу частот*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Вероятность попадания в интервал при условии истинности основной гипотезы |
| 1 | 3.1 | 6.8 |  |  |  |
| 2 | 6.8 | 10.6 |  |  |  |
| *3* | 10.6 | 14.3 |  |  |  |
| 4 | 14.3 | 18 |  |  |  |
| 5 | 18 | 21.8 |  |  |  |
| 6 | 21.8 | 25.6 |  |  |  |
| 7 | 25.6 | 29.3 |  |  |  |
| 8 | 29.3 | 33 |  |  |  |
| 9 | 33 | 36.8 |  |  |  |

*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 |
|  |  |  |  |  |
|  |  |  |
|  |  |  |

*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 | , где | - третий момент (ассиметрия),  - четвертый момент (эксцесс),  n - число наблюдений    - уровень значимости |
| 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 |  |  |  |  |
| 0.05 |  |  |
| 0.1 |  |  |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.4 статистического анализа обнаружено, что С8 является нормально распределенной величиной. |

**5. Проверка однородности выборок**

Анализируемый признак 1 – C11

Анализируемый признак 2 – C12

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 |  | n – объем выборок    – частота успеха    – число знаков «+» в последовательности знаков разностей  *x*1 – *y*1,…,*xn* – *yn* |
| 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 |  |  |  |  |
| 0.05 |  |  |
| 0.1 |  |  |

*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 |  | и – относительные частоты, где и объемы выборок X и Y соответственно.  k – число интервалов. |
| 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 |  | 1063 |

*в) Построить таблицу частот*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота признака 1 | Частота признака 2 | Относит. частота признака 1 | Относит. частота признака 2 |
|  |  |  |  |  |  | 0.78 |
|  |  |  |  |  |  | 0.2 |
|  |  |  |  |  |  | 0.01 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

*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 |  |  |  |  |
| 0.05 |  |  |
| 0.1 |  |  |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.5 статистического анализа удалось определить, что выборки C11 и C12 не имеют одинаковые распределения. Критерий хи-квадрат отверг гипотезу однородности. Критерий знаков также отклонил ее. Следовательно, по проведенному анализу, мы можем определить, что выборки C11 и C12 неоднородны. |

**6. Таблицы сопряжённости**

Факторный признак *x* – С2

Результативный признак *y* – С5

Объёмы выборок – 315

Статистическая гипотеза –

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

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | - частота пары в выборке,  - теоретические частоты  - число вариантов случайной величины ,  - число вариантов случайной величины  Критическая область для статистики критерия *Z* выбирается правосторонней.  - уровень значимости |
| 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 |  |  |  |  |
| Male |  |  |  |  |
| Σ |  |  |  |  |

*в) Построить теоретическую таблицу сопряжённости*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *x y* | No | Not often | Often | Σ |
| Female |  |  |  |  |
| Male |  |  |  |  |
| Σ |  |  |  |  |

*г) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 |  |  |  | C2 влияет на C5 |
| 0.05 |  | C2 влияет на C5 |
| 0.1 |  | C2 влияет на C5 |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.6 статистического анализа обнаружено, что пол (C2) оказывает влияние на употребление витаминов (C5) т.е. между этими признаками имеется значимая статистическая связь. |

**7. Дисперсионный анализ**

Факторный признак *x* – C3

Результативный признак *y* – C14

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 | Former |  |  |  |
| 2 | Never |  |  |  |
| 3 | Current Smoker |  |  |  |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак |  |  |  |
| Остаточные признаки |  |  |  |
| Все признаки |  |  |  |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак |  |  |  |
| Остаточные признаки |  |  |  |
| Все признаки |  |  |  |

*г) Проверить правило сложения дисперсий*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель | *Dмежгр* | *Dвнутригр* | *Dобщ* | *Dмежгр* + *Dвнутригр* |
| Значение |  |  |  |  |

*д) Рассчитать показатели тесноты связи между факторным и результативным признаками*

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

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

|  |
| --- |
| *Между выборками С3 и С14 имеется* ***слабая*** *связь (по шкале Чеддока).* |

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

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | К – кол-во вар-в факторного признака  n-объем выборки  - межгрупповая дисперсия  -общая дисперсия |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |
| 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 |  |  |  | C3 не влияет на C14 |
| 0.05 |  | C3 влияет на C14 |
| 0.1 |  | C3 влияет на C14 |

*Вывод (в терминах предметной области)*

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

**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 | , где |  |
| Ранговый коэффициент корреляции по Спирмену | где , - средние значения рангов |  |
| Ранговый коэффициент корреляции по Кендаллу | - число пар наблюдений , для которых выполнено условие  - число пар наблюдений , для которых выполнено условие |  |

*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 |  |  |  |
| Верхняя граница |  |  |  |

*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 |  |  |  |  |
|  | 0.1 |  |  |  |  |
|  | 0.1 |  |  |  |  |

*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 | С8 | С10 | С13 |
| С8 |  |  |  |
| С10 |  |  |  |
| С13 |  |  |  |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Признак  Признак | С8 | С10 | С13 |
| С8 |  |  |  |
| С10 |  |  |  |
| С13 |  |  |  |

*в) Рассчитать точечную оценку коэффициента конкордации*

|  |  |  |
| --- | --- | --- |
|  | Формула расчета | Значение |
| Коэффициент конкордации | , где  n - объем выборок  k - число выборок  - ранг i-го элемента выборки в j-ом признаке | 0.394 |

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

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | W - коэффициент конкордации  n - размер выборки  k - число выборок  - уровень значимости |
| 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 |  |  | + | С8, С10, С13 – не зависимы |
| 0.05 | - | С8, С10, С13 – зависимы |
| 0.1 | - | С8, С10, С13 – зависимы |

*Вывод (в терминах предметной области)*

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

**9. Регрессионный анализ**

*9.1 Простейшая линейная регрессионная модель*

Факторный признак *x* –

Результативный признак *y* –

Уравнение регрессии – 

*9.1.1. Точечные оценки линейной регрессионной модели*

*а) Рассчитать точечные оценки параметров линейной регрессионной модели*

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

*б) Записать точечную оценку уравнения регрессии*

|  |
| --- |
| *f*(*x*) = |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак |  |  |  |
| Остаточные признаки |  |  |  |
| Все признаки |  |  |  |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак |  |  |  |
| Остаточные признаки |  |  |  |
| Все признаки |  |  |  |

*д) Проверить правило сложения дисперсий*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель | *Dрегр* | *Dост* | *Dобщ* | *Dрегр* + *Dост* |
| Значение |  |  |  |  |

*е) Рассчитать показатели тесноты связи между факторным и результативным признаками*

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

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

|  |
| --- |
|  |

*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 |  |  |  |
| Верхняя граница |  |  |  |
| β1 | Нижняя граница |  |  |  |
| Верхняя граница |  |  |  |

*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 |  | *- статистика, для точечной оценки коэффициента 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 |  |  |  |  |
| 0.05 |  |  |
| 0.1 |  |  |

*9.2 Линейная регрессионная модель общего вида*

Факторный признак *x* –

Результативный признак *y* –

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 | —вектор параметров модели— вектор откликов модели |  |
| β1 |  |
| β2 |  |

*б) Записать точечную оценку уравнения регрессии*

|  |
| --- |
| *f*(*x*) = |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак |  |  |  |
| Остаточные признаки |  |  |  |
| Все признаки |  |  |  |

*г) Проверить правило сложения дисперсий*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель | *Dрегр* | *Dост* | *Dобщ* | *Dрегр* + *Dост* |
| Значение |  |  |  |  |

*д) Рассчитать показатели тесноты связи между факторным и результативным признаками*

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

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

|  |
| --- |
|  |

*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 |  |  |  |  |
| 0.05 |  |  |
| 0.1 |  |  |

*9.3 Множественная линейная регрессионная модель*

Факторный признак 1 *x*1 –

Факторный признак 2 *x*2 –

Результативный признак *y* –

Уравнение регрессии – 

*а) Рассчитать точечные оценки параметров линейной регрессионной модели*

|  |  |  |
| --- | --- | --- |
| Параметр | Формула расчета | Значение |
| β0 | —вектор параметров модели— вектор откликов модели |  |
| β1 |  |
| β2 |  |

*б) Записать точечную оценку уравнения регрессии*

|  |
| --- |
| *f*(*x*) = |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак |  |  |  |
| Остаточные признаки |  |  |  |
| Все признаки |  |  |  |

*г) Проверить правило сложения дисперсий*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель | *Dрегр* | *Dост* | *Dобщ* | *Dрегр* + *Dост* |
| Значение |  |  |  |  |

*д) Рассчитать показатели тесноты связи между факторным и результативным признаками*

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

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

|  |
| --- |
|  |

*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 |  |  |  |
| Остаточные признаки |  |  |  |
| Все признаки |  |  |  |

*б) Сводная таблица свойств различных регрессионных моделей*

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

*Вывод (в терминах предметной области)*

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