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 Б22-534**

**Когановского Григория**

Вариант №7

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

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

**2024 г.**

**1. Описательные статистики**

*1.1. Выборочные характеристики*

Анализируемый признак 1 – C9 (Number of alcoholic drinks consumed per week)

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

Анализируемый признак 3 – C11 (Dietary beta-carotene consumed (mcg per day))

## а) Привести формулы расчёта выборочных характеристик

|  |  |
| --- | --- |
| Выборочная хар-ка | Формула расчета |
| Объём выборки |  |
| 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 | 3.28 | 242.46 | 2185.60 |
| Выборочная дисперсия | 151.37 | 17366.48 | 2165445.23 |
| Выборочное среднеквадратическое отклонение | 12.30 | 131.78 | 1471.55 |
| Выборочный коэффициент асимметрии | 13.76 | 1.47 | 1.61 |
| Выборочный эксцесс | 217.82 | 3.34 | 3.40 |

*1.2. Группировка и гистограммы частот*

Анализируемый признак – C9 (Number of alcoholic drinks consumed per week)

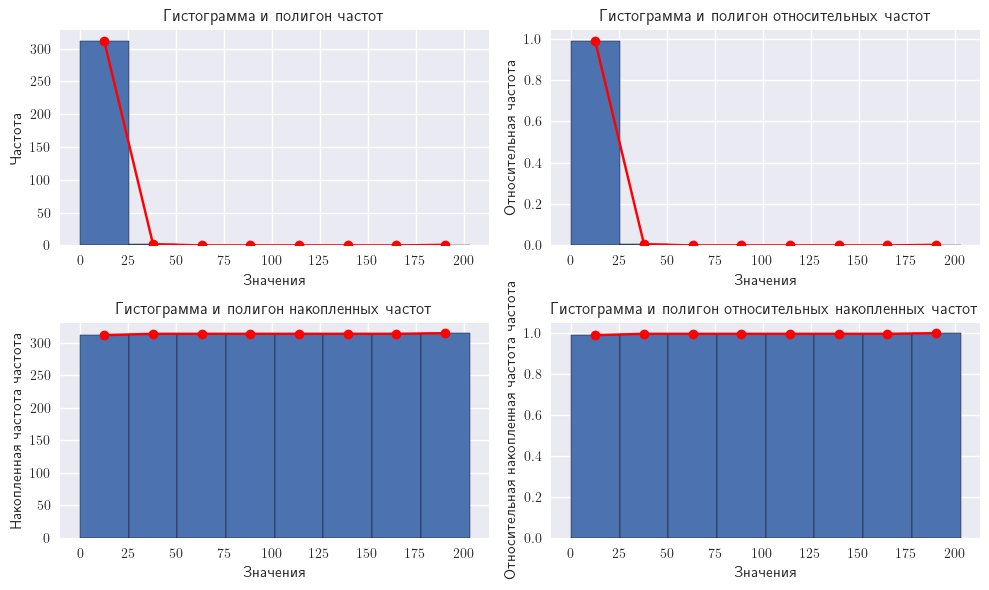
Объём выборки –

*а) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| 8 | Формула Стерджесса: | от 25.38 до 25.58 |

*б) Построить таблицу частот*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Накопл. частота | Относит. накопл. частота |
| 1 | -0.2 | 25.38 | 312 | 0.99 | 312 | 0.99 |
| 2 | 25.38 | 50.75 | 2 | 0.01 | 314 | 1.00 |
| 3 | 50.75 | 76.12 | 0 | 0.00 | 314 | 1.00 |
| 4 | 76.12 | 101.50 | 0 | 0.00 | 314 | 1.00 |
| 5 | 101.50 | 126.88 | 0 | 0.00 | 314 | 1.00 |
| 6 | 126.88 | 152.25 | 0 | 0.00 | 314 | 1.00 |
| 7 | 152.25 | 177.62 | 0 | 0.00 | 314 | 1.00 |
| 8 | 177.62 | 203.0 | 1 | 0.00 | 315 | 1.00 |

*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 – C9 (Number of alcoholic drinks consumed per week)

Объём выборки –

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 | 1.48 | 1.91 | 2.13 |
| Верхняя граница | 5.08 | 4.65 | 4.42 |

*2.2. Доверительные интервалы для дисперсии*

Анализируемый признак – C9 (Number of alcoholic drinks consumed per week)

Объём выборки –

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 | 124.72 | 130.64 | 133.82 |
| Верхняя граница | 188.31 | 178.72 | 174.05 |

*2.3. Доверительные интервалы для разности мат. ожиданий*

Анализируемый признак 1 – C11 (Dietary beta-carotene consumed (mcg per day))

Анализируемый признак 2 – C12 (Dietary retinol consumed (mcg per day))

Объёмы выборок –

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 | 1121.82 | 1177.26 | 1205.56 |
| Верхняя граница | 1583.96 | 1528.52 | 1500.21 |

*2.4. Доверительные интервалы для отношения дисперсий*

Анализируемый признак 1 – C11 (Dietary beta-carotene consumed (mcg per day))

Анализируемый признак 2 – C12 (Dietary retinol consumed (mcg per day))

Объёмы выборок –

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 | 4.67 | 5.01 | 5.19 |
| Верхняя граница | 8.37 | 7.81 | 7.53 |

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

*3.1. Проверка статистических гипотез о математических ожиданиях*

Анализируемый признак – C9 (Number of alcoholic drinks consumed per week)

Объём выборки –

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 | 0.1 | 4.72 | *0.00* | отклоняется |  |
| 2 | 0.1 | 1.84 | *0.07* | отклоняется |  |
| 4 | 0.1 | -1.04 | *0.30* | принимается |  |

*3.2. Проверка статистических гипотез о дисперсиях*

Анализируемый признак – C9 (Number of alcoholic drinks consumed per week)

Объём выборки –

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 |
| 10 | 0.1 | 476.82 | *0.00* | отклоняется |  |
| 12.5 | 0.1 | 305.16 | *0.74* | принимается |  |
| 15 | 0.1 | 211.92 | *0.00* | отклоняется |  |

*3.3. Проверка статистических гипотез о равенстве математических ожиданий*

Анализируемый признак 1 – C11 (Dietary beta-carotene consumed (mcg per day))

Анализируемый признак 2 – C12 (Dietary retinol consumed (mcg per day))

Объёмы выборок –

Статистическая гипотеза –

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

|  |  |
| --- | --- |
|  | Выражение |
| Формула расчета статистики критерия | , где |
| 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 | 15.13 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

*3.4. Проверка статистических гипотез о равенстве дисперсий*

Анализируемый признак 1 – C11 (Dietary beta-carotene consumed (mcg per day))

Анализируемый признак 2 – C12 (Dietary retinol consumed (mcg per day))

Объёмы выборок –

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 | 6.26 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

**4. Критерии согласия**

Анализируемый признак – C9 (Number of alcoholic drinks consumed per week)

Объём выборки –

*4.1. Критерий хи-квадрат*

Теоретическое распределение – нормальное.

Статистическая гипотеза –

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

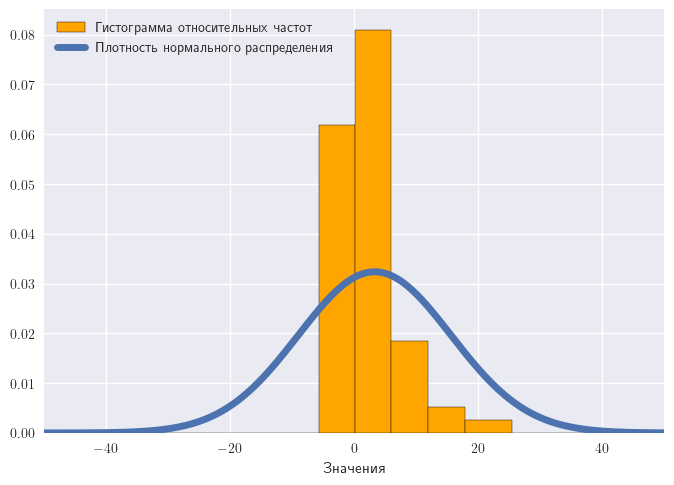
|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | - число интервалов в группированном статистическом ряду.  - частота попадания случайной величины в интервал .  - вероятность попадания случайной величины в интервал в условиях , то есть, если , то |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  | - количество оцениваемых параметров у предполагаемого распределения . |
| Формула расчета критической точки |  | Малые значения нам также подходят, поэтому критическая область выбирается правосторонней |
| Формула расчета *p-value* |  |  |

*б) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| 8 | - формула Стерджесса  - поправка на чувствительность критерия | От 7.65 до 152.91 |

*в) Построить таблицу частот*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Вероятность попадания в интервал при условии истинности основной гипотезы |
| 1 |  | -11.83 | 0 | 0 | 0.11 |
| 2 | -11.83 | -5.65 | 0 | 0 | 0.12 |
| 3 | -5.65 | 0.04 | 111 | 0,35 | 0.16 |
| 4 | 0.04 | 5.92 | 150 | 0,48 | 0.19 |
| 5 | 5.92 | 11.91 | 35 | 0,11 | 0.17 |
| 6 | 11.91 | 17.91 | 10 | 0,03 | 0.12 |
| 7 | 17.91 | 25.38 | 6 | 0,02 | 0.08 |
| 8 | 25.38 |  | 3 | 0,01 | 0.04 |

*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 | 332.48 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

*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 | - выборочный коэффициент асимметрии.  - выборочный эксцесс.  - объём выборки. |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  | Правосторонняя критическая область |
| Формула расчета *p-value* |  |

*б) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 632630.74 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.4 статистического анализа обнаружено, что оба критерия согласия отвергают гипотезу о нормальности распределения величины C9, поэтому можно сделать вывод, что C9 (Number of alcoholic drinks consumed per week) не имеет нормального распределения. |

**5. Проверка однородности выборок**

Анализируемый признак 1 – C11 (Dietary beta-carotene consumed (mcg per day))

Анализируемый признак 2 – C12 (Dietary retinol consumed (mcg per day))

Объёмы выборок –

*5.1 Критерий знаков*

Статистическая гипотеза –

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

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | - число знаков `+` в выборке |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  | Двусторонняя критическая область |
| Формула расчета *p-value* |  |

*б) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | -14.03 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

*5.2. Критерий хи-квадрат*

Статистическая гипотеза –

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

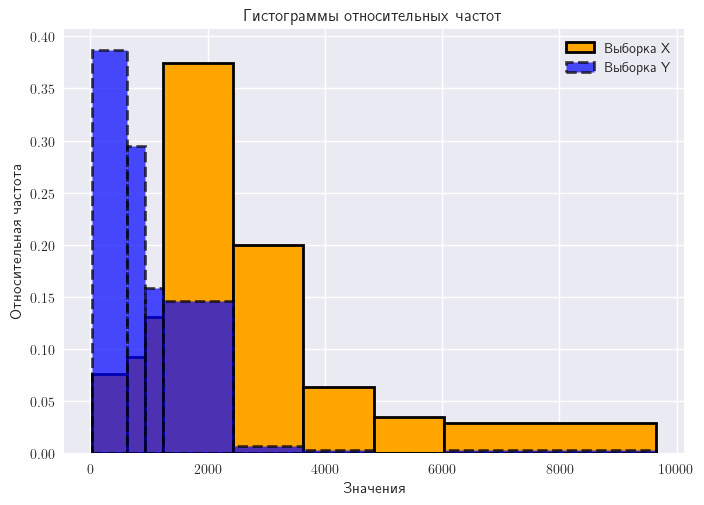
|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | - число интервалов в группированном статистическом ряду.  - частота попадания случайной величины в интервал .  - частота попадания случайной величины в интервал .  - объём выборки  - объём выборки |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  | Правосторонняя критическая область |
| Формула расчета *p-value* |  |

*б) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| 8 | - формула Стерджесса  - поправка на чувствительность критерия | от 300.38 до 3604.50 |

*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 | 30.00 | 630.75 | 24 | 122 | 0,08 | 0,39 |
| 2 | 630.75 | 931.12 | 29 | 93 | 0,09 | 0,3 |
| 3 | 931.12 | 1231.50 | 41 | 50 | 0,13 | 0,16 |
| 4 | 1231.50 | 2433.0 | 118 | 46 | 0,37 | 0,15 |
| 5 | 2433.0 | 3634.50 | 63 | 2 | 0,2 | 0,01 |
| 6 | 3634.50 | 4836.00 | 20 | 1 | 0,06 | 0 |
| 7 | 4836.00 | 6037.50 | 11 | 0 | 0,03 | 0 |
| 8 | 6037.50 | 9642.00 | 9 | 1 | 0,03 | 0 |

*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 | 223.69 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.5 статистического анализа обнаружено, что выборки С11 (Dietary beta-carotene consumed (mcg per day)) и C12 (Dietary retinol consumed (mcg per day)) неоднородны. |

**6. Таблицы сопряжённости**

Факторный признак *x* – B3 (Body fat)

Результативный признак *y* – B5 (Sex)

Объёмы выборок –

Статистическая гипотеза –

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

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | - наблюдения случайного вектора , где - случайные величины дискретного типа  - варианты признака  - варианты признака  - выборочная частота варианта в выборке  - теоретическая частота варианта в выборке при условии истинности |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  | Правосторонняя критическая область |
| Формула расчета *p-value* |  |

*б) Построить эмпирическую таблицу сопряжённости*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Вариант Вариант* |  |  |  |  |
|  | 87 | 77 | 109 | 273 |
|  | 24 | 5 | 13 | 42 |
|  | 111 | 82 | 122 | 315 |

*в) Построить теоретическую таблицу сопряжённости*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Вариант Вариант* |  |  |  |  |
|  | 96,20 | 71,07 | 105,73 | 273 |
|  | 14,80 | 10,93 | 16,27 | 42 |
|  | 111,00 | 82 | 122 | 315 |

*г) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 11.07 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.6 статистического анализа обнаружено, что между факторным признаком C2 и результативным признаком C5 присутствует статистическая связь. Под действием C3 оказывается влияние на распределение C5. |

**7. Дисперсионный анализ**

Факторный признак *x* – C5 (Vitamin Use)

Результативный признак *y* – C4 (Quetelet ())

Число вариантов факторного признака –

Объёмы выборок –

Статистическая гипотеза –

*а) Рассчитать групповые выборочные характеристики*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № п/п | Вариант факторного признака | Объём выборки | Групповые средние | Групповые дисперсии |
| 1 | No | 111 | 26.53 | 34.21 |
| 2 | Not often | 82 | 26.63 | 41.24 |
| 3 | Often | 122 | 25.51 | 33.53 |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак |  |  |  |
| Остаточные признаки | , где  — групповая дисперсия |  |  |
| 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 |  |  |  |  |
| Значение | 0.27 | 35.78 | 36.05 | 36.05 |

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

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Эмпирический коэффициент детерминации |  | 0.01 |
| Эмпирическое корреляционное отношение |  | 0.09 |

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

|  |
| --- |
| По шкале Чеддока наблюдается отсутствие статистической связи между факторным признаком С5 и результативным признаком С4. |

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

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | — число групп  — объём выборки |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  | Правосторонняя критическая область |
| Формула расчета *p-value* |  |

*з) Проверить статистическую гипотезу дисперсионного анализа*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 1.18 | *0.31* | принимается |  |
| 0.05 | принимается |  |
| 0.1 | принимается |  |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.7 статистического анализа обнаружено, что между факторным признаком C5 и результативным признаком C4 отсутствует статистическая связь. Под действием C5 не оказывается влияние на распределение C4. |

**8. Корреляционный анализ**

*8.1. Расчёт парных коэффициентов корреляции*

Анализируемый признак 1 –

Анализируемый признак 2 –

Объёмы выборок –

*а) Рассчитать точечные оценки коэффициентов корреляции*

|  |  |  |
| --- | --- | --- |
|  | Формула расчета | Значение |
| Линейный коэффициент корреляции | , |  |
| Ранговый коэффициент корреляции по Спирмену | , где |  |
| Ранговый коэффициент корреляции по Кендаллу | , где  - 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 |
| image14.pdf |  |  |
| image12.pdf |  |  |
| image13.pdf |  |  |

*д) Проверить значимость коэффициентов корреляции*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Статистическая гипотеза | Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| image14.pdf | 0.1 |  |  |  |  |
| image12.pdf | 0.1 |  |  |  |  |
| image13.pdf | 0.1 |  |  |  |  |

*8.2. Расчёт множественных коэффициентов корреляции*

Анализируемый признак 1 –

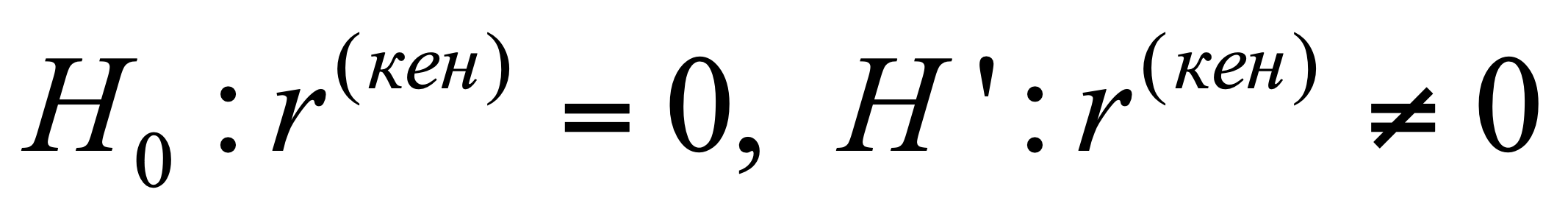
Анализируемый признак 2 –

Анализируемый признак 3 –

Объёмы выборок –

*а) Рассчитать матрицу ранговых коэффициентов корреляции по Кендаллу*

|  |  |  |  |
| --- | --- | --- | --- |
| Признак  Признак |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Признак  Признак |  |  |  |
|  | – |  |  |
|  |  | – |  |
|  |  |  | – |

*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 |  |

*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 |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  | - уровень значимости |
| Формула расчета *p-value* |  |  |

*д) Проверить значимость коэффициента конкордации*

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

*Вывод (в терминах предметной области)*

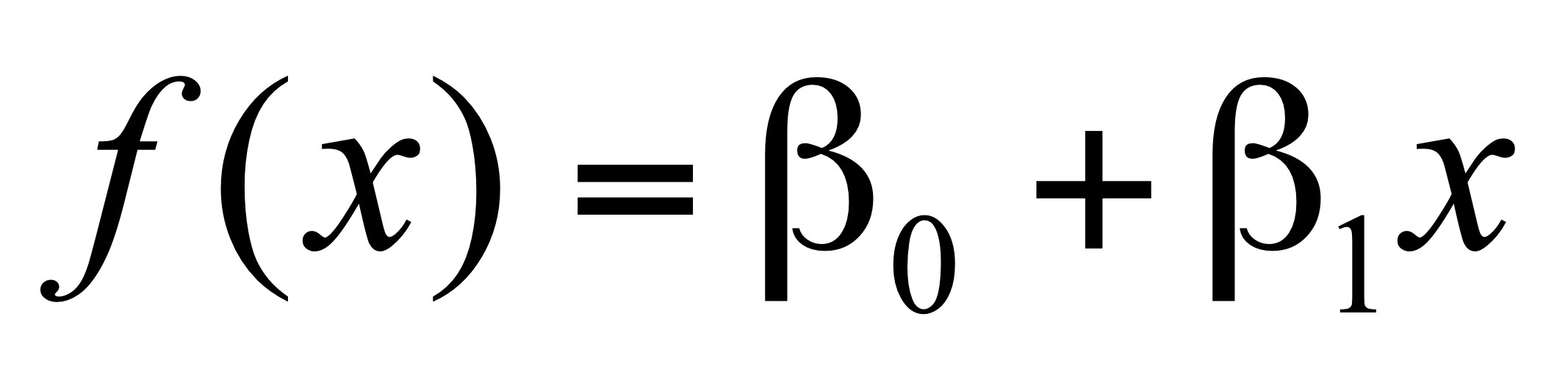
|  |
| --- |
| В результате проведённого в п.8 статистического анализа обнаружено, что |

**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*

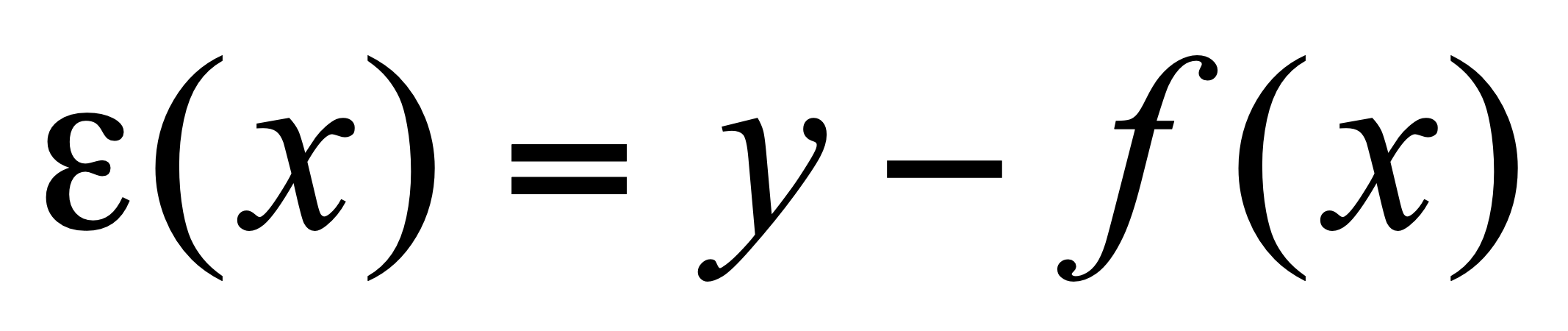
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Параметр | Границы доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| β0 | Нижняя граница |  |  |  |
| Верхняя граница |  |  |  |
| β1 | Нижняя граница |  |  |  |
| Верхняя граница |  |  |  |

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

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

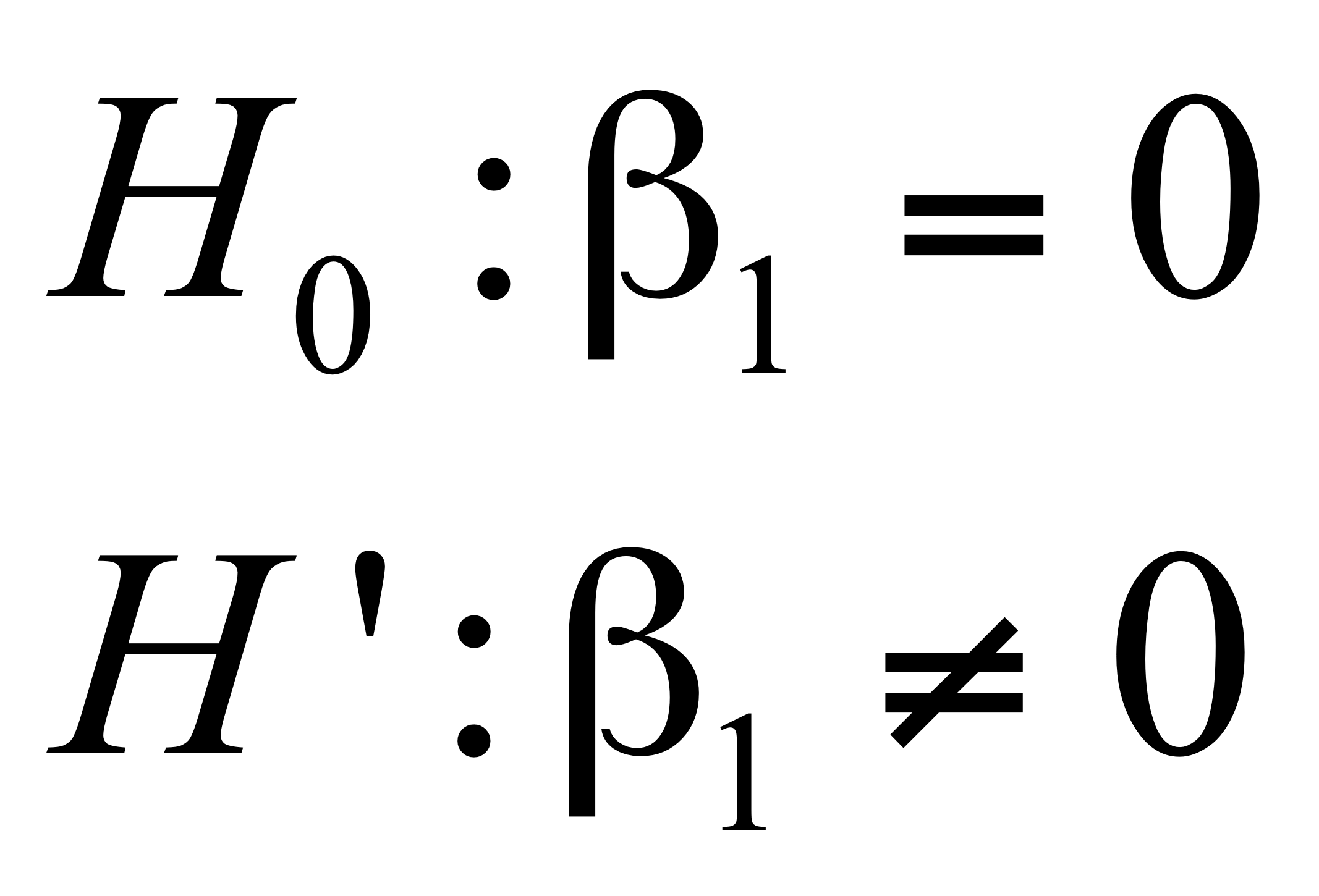
*г) Построить диаграмму рассеяния признаков x и y*. *Нанести на диаграмму функцию регрессии f(x), а также нижние и верхние границы линии регрессии flow(x) и fhigh(x) на уровне значимости α = 0.1*

|  |
| --- |
|  |

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

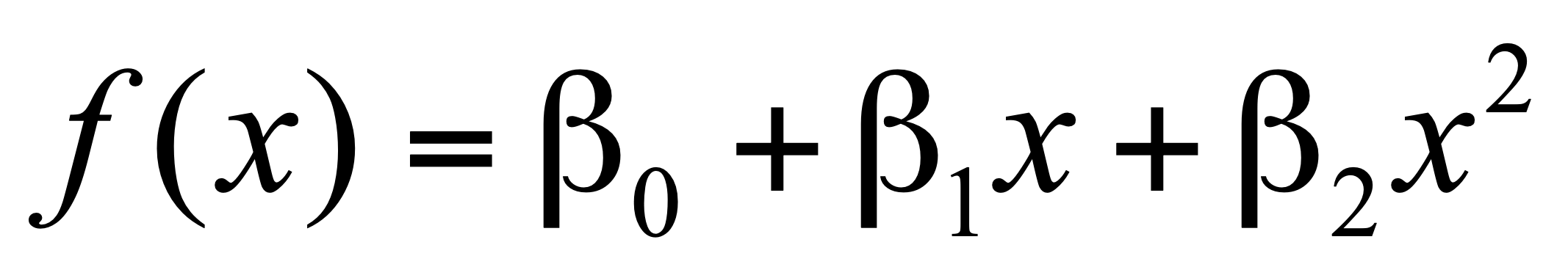
## б) Проверить значимость линейной регрессионной модели

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |  |  |
| 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*

|  |
| --- |
|  |

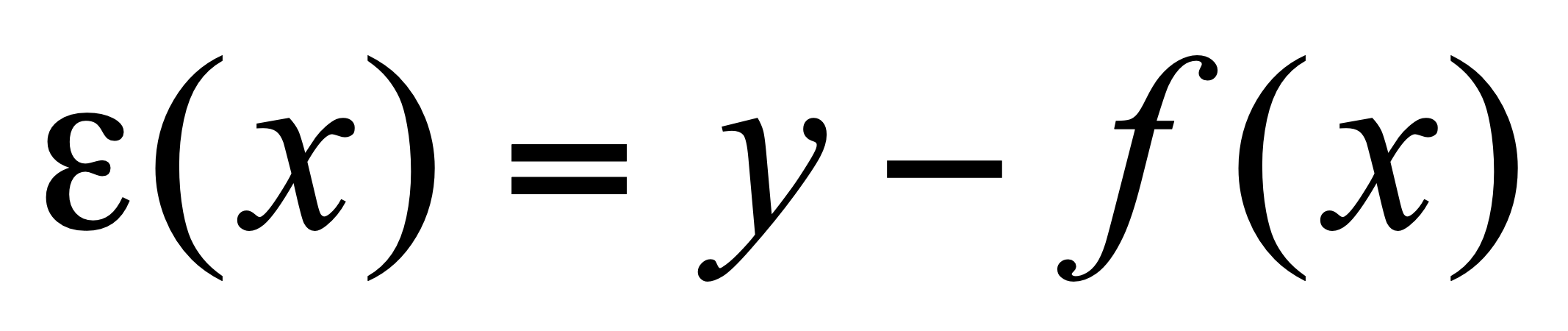
*9.2.2. Интервальные оценки линейной регрессионной модели*

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

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

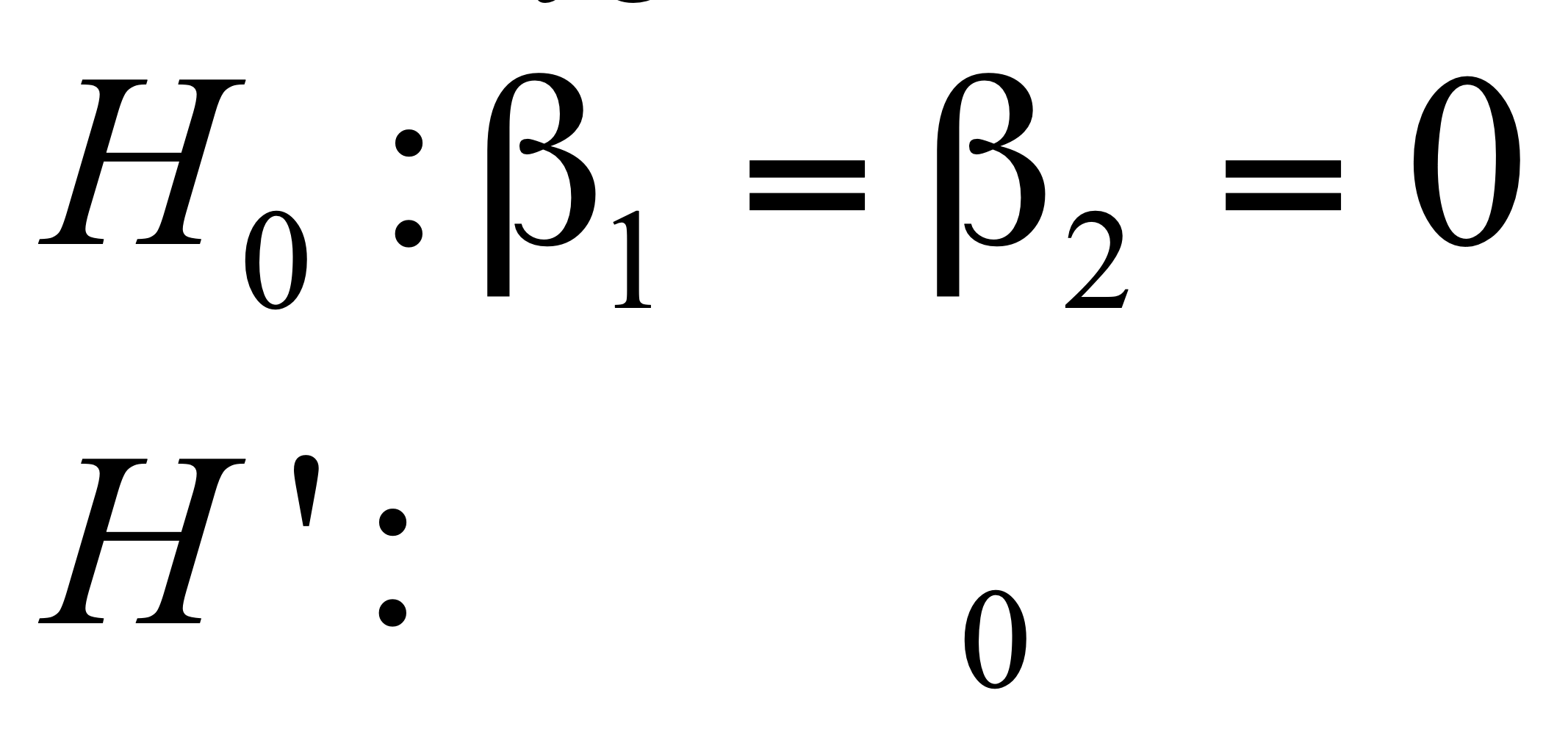
*б) Построить диаграмму рассеяния признаков x и y. Нанести на диаграмму функцию регрессии f(x), а также нижние и верхние границы линии регрессии flow(x) и fhigh(x) на уровне значимости α = 0.1*

|  |
| --- |
|  |

*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*

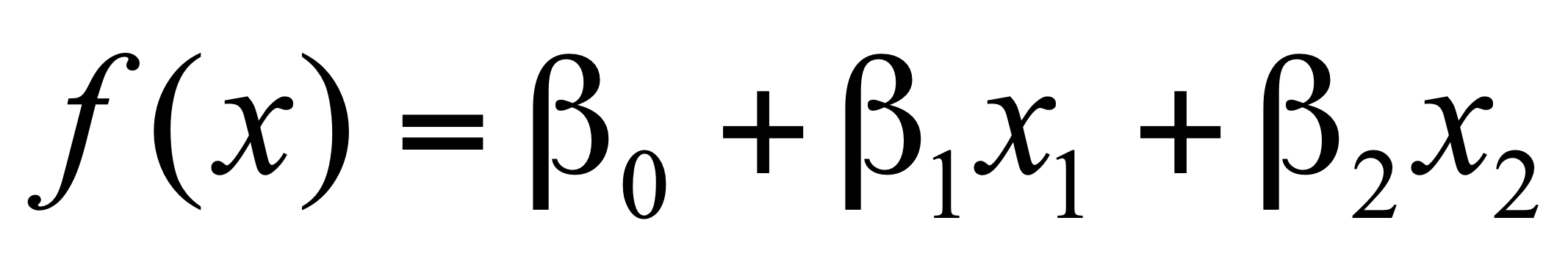
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 |  |  |  |  |
| 0.05 |  |  |
| 0.1 |  |  |

*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 |  |  |
| 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 |