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**

**Баранова Александра**

Вариант №2

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

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

**2024 г.**

**1. Описательные статистики**

*1.1. Выборочные характеристики*

Анализируемый признак 1 – B7 (Weight (lbs))

Анализируемый признак 2 – B8 (Height (inches))

Анализируемый признак 3 – B9 (Neck circumference (cm))

## 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 | 178.92 | 70.15 | 37.99 |
| Выборочная дисперсия | 860.30 | 13.36 | 5.89 |
| Выборочное среднеквадратическое отклонение | 29.33 | 3.66 | 2.43 |
| Выборочный коэффициент асимметрии | 1.20 | -5.35 | 0.55 |
| Выборочный эксцесс | 5.14 | 58.35 | 2.64 |

*1.2. Группировка и гистограммы частот*

Анализируемый признак – B7 (Weight (lbs))

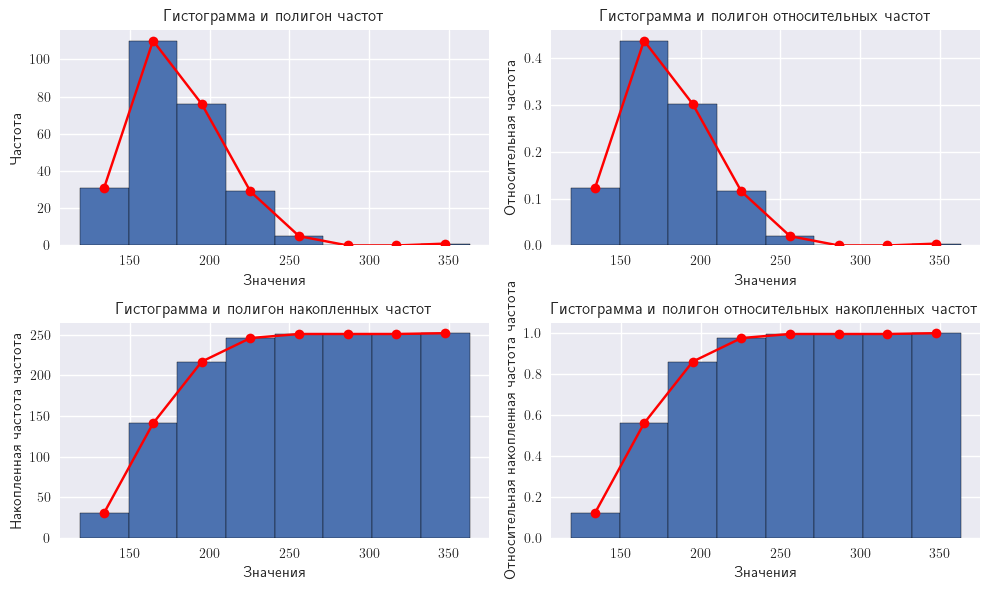
Объём выборки – 252

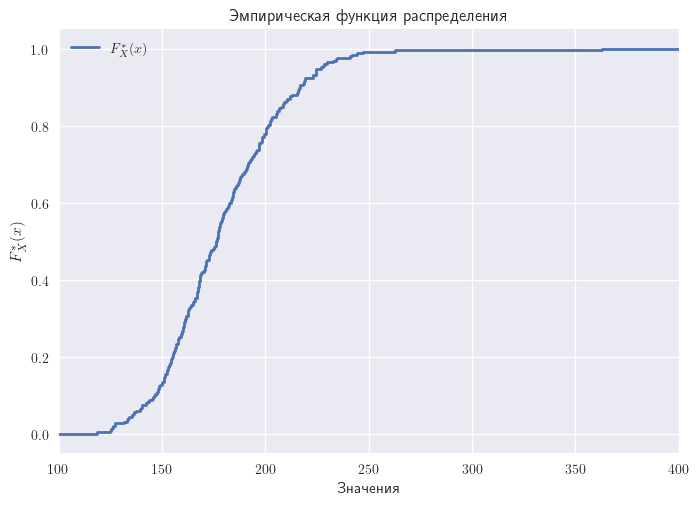
*а) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| 8 | Формула Стерджесса: | от 30.58 до 30.83 |

*б) Построить таблицу частот*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Накопл. частота | Относит. накопл. частота |
| 1 | 118.25 | 149.08 | 31 | 0.12 | 31 | 0.12 |
| 2 | 149.08 | 179.66 | 110 | 0.44 | 141 | 0.56 |
| 3 | 179.66 | 210.24 | 76 | 0.3 | 217 | 0.86 |
| 4 | 210.24 | 240.82 | 29 | 0.12 | 246 | 0.98 |
| 5 | 240.82 | 271.41 | 5 | 0.02 | 251 | 1.00 |
| 6 | 271.41 | 301.99 | 0 | 0.00 | 251 | 1.00 |
| 7 | 301.99 | 332.57 | 0 | 0.00 | 251 | 1.00 |
| 8 | 332.57 | 363.15 | 1 | 0.00 | 252 | 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 – B7 (Weight (lbs))

Объём выборки – 252

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 | 174.12 | 175.28 | 175.87 |
| Верхняя граница | 183.73 | 182.57 | 181.98 |

*2.2. Доверительные интервалы для дисперсии*

Анализируемый признак – B7 (Weight (lbs))

Объём выборки – 252

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 | 693.83 | 730.50 | 750.27 |
| Верхняя граница | 1100.22 | 1037.24 | 1006.86 |

*2.3. Доверительные интервалы для разности мат. ожиданий*

Анализируемый признак 1 – B11 (Abdomen circumference (cm))

Анализируемый признак 2 – B12 (Hip circumference (cm))

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 | -9.46 | -8.95 | -8.69 |
| Верхняя граница | -5.24 | -5.75 | -6.00 |

*2.4. Доверительные интервалы для отношения дисперсий*

Анализируемый признак 1 – B11 (Abdomen circumference (cm))

Анализируемый признак 2 – B12 (Hip circumference (cm))

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 | 1.63 | 1.77 | 1.84 |
| Верхняя граница | 3.14 | 2.90 | 2.79 |

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

*3.1. Проверка статистических гипотез о математических ожиданиях*

Анализируемый признак – B7 (Weight (lbs))

Объём выборки – 252

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 |
| 120 | 0.1 | 31.83 | *0.00* | отклоняется |  |
| 180 | 0.1 | -0.58 | *0.56* | принимается |  |
| 240 | 0.1 | -32.99 | *0.00* | отклоняется |  |

*3.2. Проверка статистических гипотез о дисперсиях*

Анализируемый признак – B7 (Weight (lbs))

Объём выборки – 252

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 | 2167.94 | *0.00* | отклоняется |  |
| 20 | 0.1 | 541.99 | *0.00* | отклоняется |  |
| 30 | 0.1 | 240.88 | *0.67* | принимается |  |

*3.3. Проверка статистических гипотез о равенстве математических ожиданий*

Анализируемый признак 1 – B11 (Abdomen circumference (cm))

Анализируемый признак 2 – B12 (Hip circumference (cm))

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 | -9.01 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

*3.4. Проверка статистических гипотез о равенстве дисперсий*

Анализируемый признак 1 – B11 (Abdomen circumference (cm))

Анализируемый признак 2 – B12 (Hip circumference (cm))

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 | 2.27 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

**4. Критерии согласия**

Анализируемый признак – B7 (Weight (lbs))

Объём выборки – 252

*4.1. Критерий хи-квадрат*

Теоретическое распределение – нормальное.

Статистическая гипотеза –

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

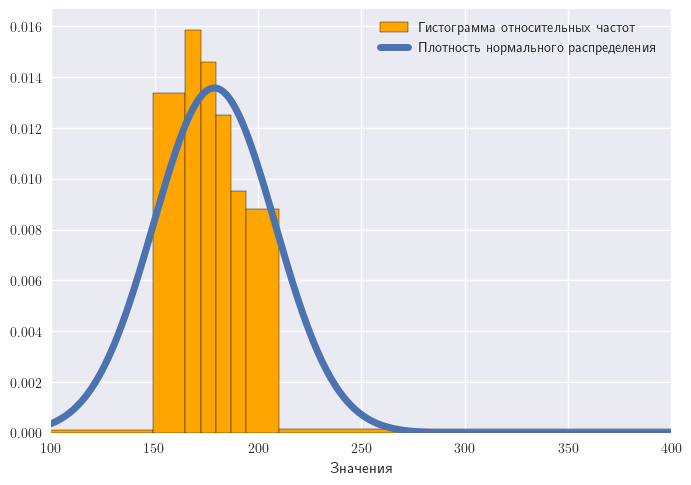
|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | - число интервалов в группированном статистическом ряду.  - частота попадания случайной величины в интервал .  - вероятность попадания случайной величины в интервал в условиях , 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 7.30 до |

*в) Построить таблицу частот*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Вероятность попадания в интервал при условии истинности основной гипотезы |
| 1 |  | 149.08 | 31 | 0.12 | 0.15 |
| 2 | 149.08 | 164.82 | 53 | 0.21 | 0.16 |
| 3 | 164.82 | 172.32 | 30 | 0.12 | 0.10 |
| 4 | 172.32 | 179.66 | 27 | 0.11 | 0.10 |
| 5 | 179.66 | 186.96 | 23 | 0.09 | 0.10 |
| 6 | 186.96 | 194.45 | 18 | 0.07 | 0.09 |
| 7 | 194.45 | 210.24 | 35 | 0.14 | 0.16 |
| 8 | 210.24 |  | 35 | 0.14 | 0.14 |

*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 | 9.07 | *0.11* | принимается |  |
| 0.05 | принимается |  |
| 0.1 | принимается |  |

*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 | 337.89 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.4 статистического анализа обнаружено, что критерий не отвергает гипотезу о нормальности распределения, но критерий Харке-Бера отвергает её.  Так как для этих данных критерий Харке-Бера более чувствителен, чем критерий , то можно сделать вывод, что выборка B7 не имеет нормального распределения. |

**5. Проверка однородности выборок**

Анализируемый признак 1 – B11 (Abdomen circumference (cm))

Анализируемый признак 2 – B12 (Hip circumference (cm))

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 | -12.72 | *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*

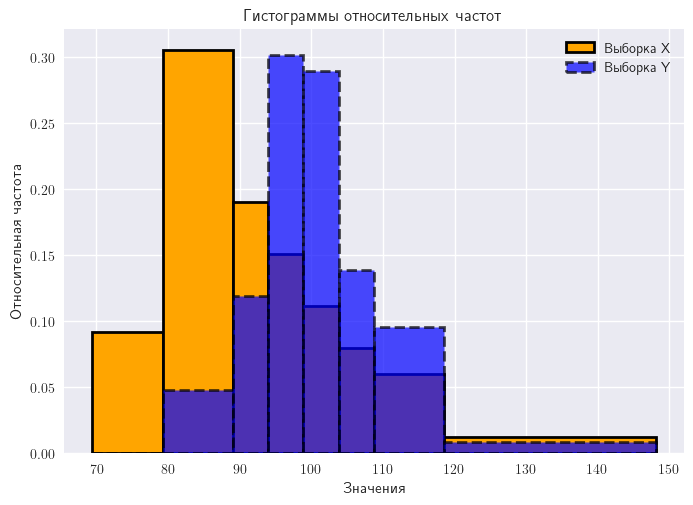
|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | - число интервалов в группированном статистическом ряду.  - частота попадания случайной величины в интервал .  - частота попадания случайной величины в интервал .  - объём 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 4.92 до 29.51 |

*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 | 69.4 | 79.24 | 23 | 0 | 0,09 | 0 |
| 2 | 79.24 | 89.08 | 77 | 12 | 0,31 | 0,05 |
| 3 | 89.08 | 93.99 | 48 | 30 | 0,19 | 0,12 |
| 4 | 93.99 | 98.91 | 38 | 76 | 0,15 | 0,3 |
| 5 | 98.91 | 103.83 | 28 | 73 | 0,11 | 0,29 |
| 6 | 103.83 | 108.75 | 20 | 35 | 0,08 | 0,14 |
| 7 | 108.75 | 118.59 | 15 | 24 | 0,06 | 0,1 |
| 8 | 118.59 | 148.10 | 3 | 2 | 0,01 | 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 | 113.71 | *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 B11 и B12 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* |  |  |  |
|  | 65 | 55 | 120 |
|  | 8 | 31 | 39 |
|  | 20 | 73 | 93 |
|  | 93 | 159 | 252 |

*в) Построить теоретическую таблицу сопряжённости*

|  |  |  |  |
| --- | --- | --- | --- |
| *Вариант*  *Вариант* |  |  |  |
|  | 44,29 | 75,71 | 120 |
|  | 14,39 | 24,61 | 39 |
|  | 34,32 | 58,68 | 93 |
|  | 93 | 159 | 252 |

*г) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 29.33 | *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 B3 и результативным признаком B5 присутствует статистическая связь. Под действием B3 оказывается влияние на распределение B5. |

**7. Дисперсионный анализ**

Факторный признак *x* – B6 (Town)

Результативный признак *y* – B1 (Body density determined from underwater weighing)

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 | Arlington | 73 | 1.05 | 0.00 |
| 2 | Norwood | 41 | 1.05 | 0.00 |
| 3 | Revere | 82 | 1.05 | 0.00 |
| 4 | Somerville | 56 | 1.06 | 0.00 |

*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 | 0.00 | 0.00 | 0.00 | 0.00 |

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

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Эмпирический коэффициент детерминации |  | 0.02 |
| Эмпирическое корреляционное отношение |  | 0.13 |

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

|  |
| --- |
| По шкале Чеддока наблюдается слабая степень статистической связи между факторным признаком B6 и результативным признаком B1. |

*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 | 1.34 | *0.26* | 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 B6 и результативным признаком B1 отсутствует статистическая связь. Под действием B6 не оказывается влияние на распределение B1. |

**8. Корреляционный анализ**

*8.1. Расчёт парных коэффициентов корреляции*

Анализируемый признак 1 – B11 (Abdomen circumference (cm))

Анализируемый признак 2 – B12 (Hip circumference (cm))

Объёмы выборок –

*а) Рассчитать точечные оценки коэффициентов корреляции*

|  |  |  |
| --- | --- | --- |
|  | Формула расчета | Значение |
| Линейный коэффициент корреляции |  | 0.87 |
| Ранговый коэффициент корреляции по Спирмену | , где и — ранги для выборок и соответственно.  Можно показать, что , где | 0.85 |
| Ранговый коэффициент корреляции по Кендаллу | , где  — количество пар точек таких, что соответственно.  Другой вариант расчёта: , где — число инверсий в выборке , предварительно отсортированной по возрастанию . | 0.66 |

*б) Привести формулы расчёта доверительного интервала для линейного коэффициента корреляции*

Для небольших объёмов выборок:

|  |  |
| --- | --- |
| 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 | 0.83 | 0.84 | 0.85 |
| Верхняя граница | 0.91 | 0.90 | 0.90 |

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Статистическая гипотеза | Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
|  | 0.1 | 28.45 | *0.00* | отклоняется |  |
|  | 0.1 | 25.00 | *0.00* | отклоняется |  |
|  | 0.1 | 15.71 | *0.00* | отклоняется |  |

*д) Проверить значимость коэффициентов корреляции*

*8.2. Расчёт множественных коэффициентов корреляции*

Анализируемый признак 1 – B7 (Weight (lbs))

Анализируемый признак 2 – B8 (Height (inches))

Анализируемый признак 3 – B9 (Neck circumference (cm))

Объёмы выборок –

*а) Рассчитать матрицу ранговых коэффициентов корреляции по Кендаллу*

|  |  |  |  |
| --- | --- | --- | --- |
| Признак  Признак | B7 | B8 | B9 |
| B7 | 1.00 | 0.37 | 0.62 |
| B8 | 0.37 | 1.00 | 0.22 |
| B9 | 0.62 | 0.22 | 1.00 |

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

Статистическая гипотеза:

|  |  |  |  |
| --- | --- | --- | --- |
| Признак  Признак | B7 | B8 | B9 |
| B7 | — | 0.00 | 0.00 |
| B8 | 0.00 | — | 0.00 |
| B9 | 0.00 | 0.00 | — |

*в) Рассчитать точечную оценку коэффициента конкордации*

|  |  |  |
| --- | --- | --- |
|  | Формула расчета | Значение |
| Коэффициент конкордации | , где  — ранг -ого объекта в -ой выборке.  — количество выборок.  —объём выборок. | 0.70 |

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

*Статистический критерий:*

|  |  |  |
| --- | --- | --- |
|  | Выражение | Пояснение использованных обозначений |
| Формула расчета статистики критерия |  | — точечная оценка коэффициента конкордации.  — количество выборок.  — объём каждой выборки. |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  | Правосторонняя критическая область |
| Формула расчета *p-value* |  |

*д) Проверить значимость коэффициента конкордации*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 351.65 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.8 статистического анализа обнаружено, что между признаками B11 и B12 существует сильная положительная, в первую очередь линейная, связь. Между признаками B7, B8 и B9 также наблюдается положительная монотонная корреляционная связь. B7 и B9 имеют наиболее сильную положительную монотонную связь, а B8 с остальными имеет более слабую связь. |

**9. Регрессионный анализ**

*9.1 Простейшая линейная регрессионная модель*

Факторный признак *x* – B1 (Body density determined from underwater weighing)

Результативный признак *y* – B2 (Percent body fat from Siri's equation)

Уравнение регрессии –

*9.1.1. Точечные оценки линейной регрессионной модели*

*а) Рассчитать точечные оценки параметров линейной регрессионной модели*

|  |  |  |
| --- | --- | --- |
| Параметр | Формула расчета | Значение |
| β0 |  | 477.65 |
| β1 |  | -434.36 |

*б) Записать точечную оценку уравнения регрессии*

|  |
| --- |
|  |

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

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

— число оцениваемых параметров функции регрессии .

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

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Регрессия | 68.06 | 1 | 17152.07 |
| Остаточные признаки | 1.69 | 250 | 1.71 |
| Все признаки | 69.76 | 251 | 70.04 |

*д) Проверить правило сложения дисперсий*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель |  |  |  |  |
| Значение | 68.06 | 1.69 | 69.76 | 69.76 |

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

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Коэффициент детерминации |  | 0.98 |
| Корреляционное отношение |  | 0.99 |

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

|  |
| --- |
| Наблюдается очень сильная (почти функциональная) корреляционная связь между факторным признаком B1 и результативным признаком B2. |

*9.1.2. Интервальные оценки линейной регрессионной модели*

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

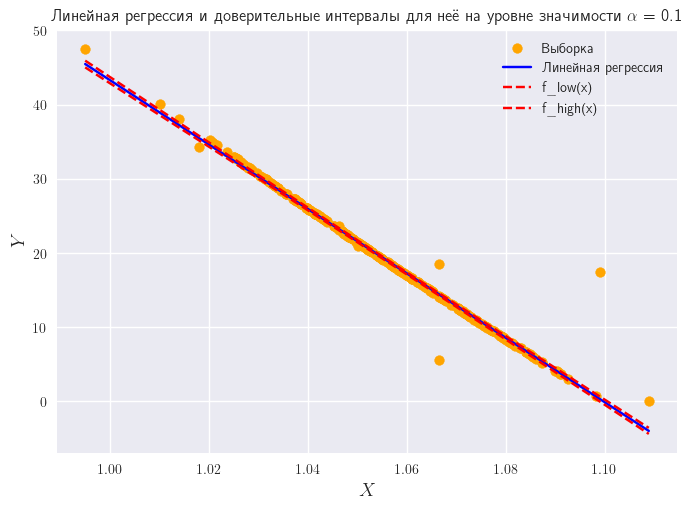
|  |  |  |
| --- | --- | --- |
| Параметр | Границы доверительного интервала | Формула расчета |
| β0 | Нижняя граница |  |
| Верхняя граница |  |
| β1 | Нижняя граница |  |
| Верхняя граница |  |

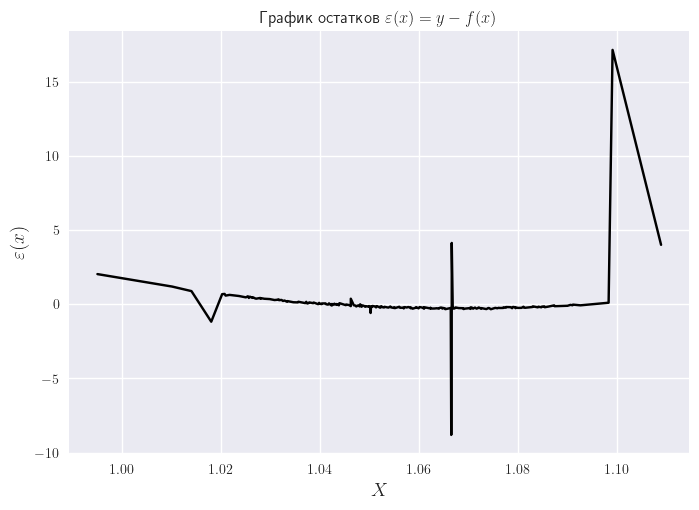
*б) Рассчитать доверительные интервалы для параметров линейной регрессионной модели*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Параметр | Границы доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| β0 | Нижняя граница | 465.77 | 468.65 | 470.1 |
| Верхняя граница | 489.53 | 486.66 | 485.2 |
| β1 | Нижняя граница | -445.61 | -442.9 | -441.52 |
| Верхняя граница | -423.11 | -425.82 | -427.20 |

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

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

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

## б) Проверить значимость линейной регрессионной модели

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 10044.03 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

*9.2 Линейная регрессионная модель общего вида*

Факторный признак *x* – B1 (Body density determined from underwater weighing)

Результативный признак *y* – B2 (Percent body fat from Siri's equation)

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 |
|  | , где — регрессионная матрица, — вектор значений результативного признака. |  |

*б) Записать точечную оценку уравнения регрессии*

|  |
| --- |
|  |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Регрессия | 68.30 | 2 | 8606.14 |
| Остаточные признаки | 1.46 | 249 | 1.47 |
| Все признаки | 69.76 | 251 | 70.04 |

*— число оцениваемых параметров функции регрессии .*

*г) Проверить правило сложения дисперсий*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель |  |  |  |  |
| Значение | 68.30 | 1.46 | 69.76 | 69.76 |

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

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Коэффициент детерминации |  | 0.98 |
| Корреляционное отношение |  | 0.99 |

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

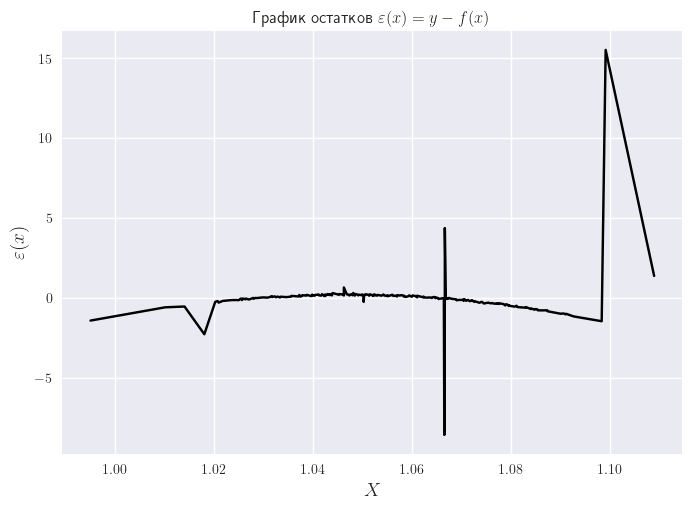
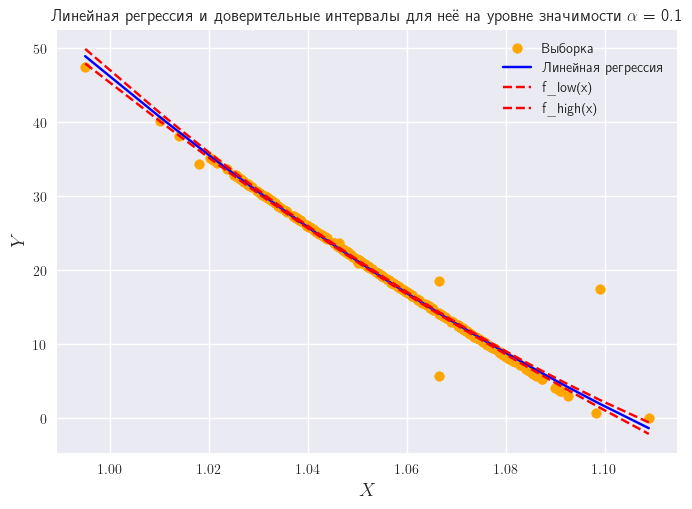
|  |
| --- |
| Наблюдается очень сильная (почти функциональная) корреляционная связь между факторным признаком B1 и результативным признаком B2. |

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

*б)* *Проверить значимость линейной регрессионной модели*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 5843.56 | *0.00* | отклоняется |  |
| 0.05 | отклоняется |  |
| 0.1 | отклоняется |  |

*9.3 Множественная линейная регрессионная модель*

Факторный признак 1 *x*1 – B1 (Body density determined from underwater weighing)

Факторный признак 2 *x*2 – B4 (Age (years))

Результативный признак *y* – B2 (Percent body fat from Siri's equation)

Уравнение регрессии –

*а) Рассчитать точечные оценки параметров линейной регрессионной модели*

|  |  |  |
| --- | --- | --- |
| Параметр | Формула расчета | Значение |
|  | , где — регрессионная матрица, — вектор значений результативного признака. |  |

*б) Записать точечную оценку уравнения регрессии*

|  |
| --- |
|  |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Регрессия | 68.09 | 2 | 8578.86 |
| Остаточные признаки | 1.67 | 249 | 1.69 |
| Все признаки | 69.75 | 251 | 70.04 |

*— число оцениваемых параметров функции регрессии .*

*г) Проверить правило сложения дисперсий*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель |  |  |  |  |
| Значение | 68.09 | 1.67 | 69.76 | 69.76 |

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

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Коэффициент детерминации |  | 0.98 |
| Корреляционное отношение |  | 0.99 |

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

|  |
| --- |
| Наблюдается очень сильная (почти функциональная) корреляционная связь между факторными признаками B1 и B4 и результативным признаком B2. |

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

*б) Сводная таблица свойств различных регрессионных моделей*

|  |  |  |  |
| --- | --- | --- | --- |
| Свойство | Простейшая линейная модель | Линейная модель с квадратичным членом | Множественная линейная модель |
| Точность | Точная | Точная | Точная |
| Значимость | Значимая | Значимая | Значимая |
| Адекватность | Адекватная | Неадекватная | Неадекватная |
| Степень тесноты связи | Очень тесная | Очень тесная | Очень тесная |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.9 статистического анализа обнаружено, что результативный признак B2 имеет очень тесную связь с факторным признаком B1. При добавлении в модель факторного признака B4 связь между результативным признаком и факторными признаками не ухудшилась. |