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 Б17-511**

**Чудновца Ивана Владимировича**

#### Вариант № 15

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

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

2019 г.

ОТЧЕТ № 1

по теме «Проверка статистических гипотез»

Вариант №\_\_\_\_\_\_\_\_

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

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

Результаты статистических тестов:

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

Выводы:

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

ОТЧЕТ № 2

по теме «Анализ статистических взаимосвязей»

Вариант №\_\_\_\_\_\_\_\_

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

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

Результаты статистических тестов:

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

Выводы:

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

ОТЧЕТ № 3

по теме «Основы многомерного статистического анализа»

Вариант №\_\_\_\_\_\_\_\_

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

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

**Сводная таблица свойств различных регрессионных моделей:**

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

Выводы:

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

1. Описательные статистики

*1.1. Выборочные характеристики*

Анализируемый признак 1 – C7 (Grams of fat consumed per day)

Анализируемый признак 2 – C9 (Number of alcoholic drinks consumed per week)

Анализируемый признак 3 – C12 (Dietary retinol consumed (mcg per day))

## 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 | 77.0333 | 3.2794 | 832.7143 |
| Выборочная дисперсия | 1144.4312 | 151.8534 | 347261.5614 |
| Выборочное среднеквадратическое отклонение | 33.8294 | 12.3229 | 589.289 |
| Выборочный коэффициент асимметрии | 1.099 | 13.7571 | 4.4525 |
| Выборочный эксцесс | 1.9648 | 217.8151 | 37.4471 |

*1.2. Группировка и гистограммы частот*

Анализируемый признак – C7 (Grams of fat consumed per day)

Объём выборки – 315

*а) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| 9 | Формула Стерджесса | 24.6111 |

*б) Построить таблицу частот*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Накопл. частота | Относит. накопл. частота |
| 1 | 14.4 | 39.0111 | 33 | 0.1048 | 33 | 0.1048 |
| 2 | 39.0111 | 63.6222 | 100 | 0.3175 | 133 | 0.4222 |
| *3* | 63.6222 | 88.2334 | 84 | 0.2667 | 217 | 0.6889 |
| 4 | 88.2334 | 112.8445 | 54 | 0.1714 | 271 | 0.8603 |
| 5 | 112.8445 | 137.4556 | 29 | 0.0921 | 300 | 0.9524 |
| 6 | 137.4556 | 162.0667 | 6 | 0.019 | 306 | 0.9714 |
| 7 | 162.0667 | 186.6779 | 6 | 0.019 | 312 | 0.9905 |
| 8 | 186.6779 | 211.289 | 2 | 0.0063 | 314 | 0.9968 |
| 9 | 211.289 | 235.9 | 1 | 0.0032 | 315 | 1.0 |

*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 – C7 (Grams of fat consumed per day)

Объём выборки – 315

Оцениваемый параметр – мат. ожидание

*а) Привести формулы расчёта доверительных интервалов*

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

*б) Рассчитать доверительные интервалы*

|  |  |  |  |
| --- | --- | --- | --- |
| Граница доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| Нижняя граница | 72.0936 | 73.283 | 73.8888 |
| Верхняя граница | 81.9731 | 80.7836 | 80.1778 |

*2.2. Доверительные интервалы для дисперсии*

Анализируемый признак – C7 (Grams of fat consumed per day)

Объём выборки – 315

Оцениваемый параметр – дисперсия

*а) Привести формулы расчёта доверительных интервалов*

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

*б) Рассчитать доверительные интервалы*

|  |  |  |  |
| --- | --- | --- | --- |
| Граница доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| Нижняя граница | 939.9721 | 984.5714 | 1008.4927 |
| Верхняя граница | 1419.1945 | 1346.8726 | 1311.7503 |

*2.3. Доверительные интервалы для разности мат. ожиданий*

Анализируемый признак 1 – C13 (Plasma beta-carotene (ng/ml))

Анализируемый признак 2 – C14 (Plasma Retinol (ng/ml))

Объёмы выборок – 315

Оцениваемый параметр – разность мат. ожиданий

*а) Привести формулы расчёта доверительных интервалов*

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

*б) Рассчитать доверительные интервалы*

|  |  |  |  |
| --- | --- | --- | --- |
| Граница доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| Нижняя граница | 372.4701 | 382.1705 | 387.1224 |
| Верхняя граница | 453.3267 | 443.6263 | 438.6744 |

*2.4. Доверительные интервалы для отношения дисперсий*

Анализируемый признак 1 – C13 (Plasma beta-carotene (ng/ml))

Анализируемый признак 2 – C14 (Plasma Retinol (ng/ml))

Объёмы выборок – 315

Оцениваемый параметр – отношение дисперсий

*а) Привести формулы расчёта доверительных интервалов*

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

*б) Рассчитать доверительные интервалы*

|  |  |  |  |
| --- | --- | --- | --- |
| Граница доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| Нижняя граница | 0.5734 | 0.6149 | 0.6372 |
| Верхняя граница | 1.0271 | 0.9578 | 0.9243 |

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

*3.1. Проверка статистических гипотез о математических ожиданиях*

Анализируемый признак – C7 (Grams of fat consumed per day)

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 |
| 73 | 0.1 | 2.116 | 0.0351 | H' | Ошибка 1-го рода |
| 75 | 0.1 | 1.0668 | 0.2869 | H0 | Ошибки нет |
| 77 | 0.1 | 0.0175 | 0.9861 | H0 | Ошибки нет |

*3.2. Проверка статистических гипотез о дисперсиях*

Анализируемый признак – C7 (Grams of fat consumed per day)

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 |
| 31.6 | 0.1 | 359.8696 | 0.076 | H' | Ошибка 1-го рода |
| 33.16 | 0.1 | 326.8062 | 0.5956 | H0 | Ошибки нет |
| 34.6 | 0.1 | 300.1699 | 0.5937 | H0 | Ошибки нет |

*3.3. Проверка статистических гипотез о равенстве математических ожиданий*

Анализируемый признак 1 – C13 (Plasma beta-carotene (ng/ml))

Анализируемый признак 2 – C14 (Plasma Retinol (ng/ml))

Объёмы выборок – 315

Статистическая гипотеза – 

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

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

*б) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 26.3874 | 0.0 | H' | Ошибки нет |
| 0.05 | H' | Ошибки нет |
| 0.1 | H' | Ошибки нет |

*3.4. Проверка статистических гипотез о равенстве дисперсий*

Анализируемый признак 1 – C13 (Plasma beta-carotene (ng/ml))

Анализируемый признак 2 – C14 (Plasma Retinol (ng/ml))

Объёмы выборок – 315

Статистическая гипотеза – 

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

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

*б) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 0.7674 | 0.0193 | H0 | Ошибки нет |
| 0.05 | H' | Ошибки нет |
| 0.1 | H' | Ошибки нет |

**4. Критерии согласия**

Анализируемый признак – C7 (Grams of fat consumed per day)

Объём выборки – 315

*4.1. Критерий хи-квадрат*

Теоретическое распределение – нормальное.

Статистическая гипотеза –

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

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

*б) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| 9 | Формула Стерджесса | 24.6111 |

*в) Построить таблицу частот*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Вероятность попадания в интервал при условии истинности основной гипотезы |
| 1 | 14.4 | 39.0111 | 33 | 0.1048 | 0.1301 |
| 2 | 39.0111 | 63.6222 | 100 | 0.3175 | 0.2155 |
| *3* | 63.6222 | 88.2334 | 84 | 0.2667 | 0.2842 |
| 4 | 88.2334 | 112.8445 | 54 | 0.1714 | 0.2256 |
| 5 | 112.8445 | 137.4556 | 29 | 0.0921 | 0.1077 |
| 6 | 137.4556 | 162.0667 | 6 | 0.019 | 0.0309 |
| 7 | 162.0667 | 186.6779 | 6 | 0.019 | 0.0053 |
| 8 | 186.6779 | 211.289 | 2 | 0.0063 | 0.0005 |
| 9 | 211.289 | 235.9001 | 1 | 0.0032 | 0.0 |

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 | 141.9568 | 0.0 | H1 | Ошибки нет |
| 0.05 | 141.9568 | 0.0 | H1 | Ошибки нет |
| 0.1 | 141.9568 | 0.0 | H1 | Ошибки нет |

*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 |  |  |
| 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 | 77.4744 |  | 0.0 | H1 |
| 0.05 | 0.0 | H1 |
| 0.1 | 0.0 | H1 |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.4 статистического анализа обнаружено, что  Распределение C7 (Grams of fat consumed per day) отлично от нормального. |

**5. Проверка однородности выборок**

Анализируемый признак 1 – C13 (Plasma beta-carotene (ng/ml))

Анализируемый признак 2 – C14 (Plasma Retinol (ng/ml))

Объёмы выборок – 315

*5.1 Критерий знаков*

Статистическая гипотеза – 

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

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

*б) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | -16.5087 | 0.0 | H1 | Ошибки нет |
| 0.05 | H1 | Ошибки нет |
| 0.1 | H1 | Ошибки нет |

*5.2. Критерий хи-квадрат*

Статистическая гипотеза – 

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

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

*б) Выбрать число групп*

|  |  |  |
| --- | --- | --- |
| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| 9 | Формула Стерджесса | 191.9 |

*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 | 0.0 | 191.9 | 215 | 2 | 0.6825 | 0.0063 |
| 2 | 191.9 | 383.8 | 73 | 28 | 0.2317 | 0.0889 |
| *3* | 383.8 | 575.7 | 17 | 134 | 0.054 | 0.4254 |
| 4 | 575.7 | 767.6 | 2 | 92 | 0.0063 | 0.2921 |
| 5 | 767.6 | 959.5 | 3 | 45 | 0.0095 | 0.1429 |
| 6 | 959.5 | 1151.4 | 2 | 8 | 0.0063 | 0.0254 |
| 7 | 1151.4 | 1343.3 | 1 | 3 | 0.0032 | 0.0095 |
| 8 | 1343.3 | 1535.2 | 2 | 2 | 0.0063 | 0.0063 |
| 9 | 1535.2 | 1727.1 | 0 | 1 | 0.0 | 0.0032 |

*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 | 448.2991 | 0.0 | H1 | Ошибки нет |
| 0.05 | H1 |  |
| 0.1 | H1 |  |

Вывод (в терминах предметной области)

|  |
| --- |
| В результате проведённого в п.5 статистического анализа обнаружено, что  Случайные величины C13 (Plasma beta-carotene (ng/ml)) и C14 (Plasma Retinol (ng/ml)) имеют различные законы распределения. |

**6. Таблицы сопряжённости**

Факторный признак *x* – C2(Sex)

Результативный признак *y* – C3(Smoking status)

Объёмы выборок – 315

Статистическая гипотеза –



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

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

*б) Построить эмпирическую таблицу сопряжённости*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *x y* | Current Smoker | Former | Never | Σ |
| Female | 36 | 93 | 144 | 273 |
| Male | 7 | 22 | 13 | 42 |
| Σ | 43 | 115 | 157 | 315 |

*в) Построить теоретическую таблицу сопряжённости*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *x y* | Current Smoker | Former | Never | Σ |
| Female | 37.2667 | 99.6667 | 136.0667 | 273.0001 |
| Male | 5.7333 | 15.3333 | 20.9333 | 41.9999 |
| Σ | 43.0 | 115.0 | 157.0 | 315.0 |

*г) Проверить статистические гипотезы*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 7.1365 | 0.0282 | H0 | Ошибки нет |
| 0.05 | H1 | Ошибки нет |
| 0.1 | H1 | Ошибки нет |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.6 статистического анализа обнаружено, что  Курение зависит от пола на уровне значимости большем чем 0.01. |

**7. Дисперсионный анализ**

Факторный признак *x* – C5(Vitamin Use)

Результативный признак *y* – C13(Plasma beta-carotene (ng/ml))

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 | No | 111 | 136.8919 | 8417.1415 |
| 2 | Not often | 82 | 185.6585 | 20522.0297 |
| 3 | Often | 122 | 240.959 | 59566.4164 |

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 | 2004.5274 | 2 | 315713.0595 |
| Остаточные признаки | 31378.4506 | 312 | 31680.1665 |
| Все признаки | 33382.9788 | 314 | 33489.294 |

г) Проверить правило сложения дисперсий

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель | *Dмежгр* | *Dвнутригр* | *Dобщ* | *Dмежгр* + *Dвнутригр* |
| Значение | 2004.5274 | 31378.4506 | 33382.9788 | 33382.978 |

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Эмпирический коэффициент детерминации |  | 0.06 |
| Эмпирическое корреляционное отношение |  | 0.245 |

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

|  |
| --- |
| Согласно шкале Чеддока степень тесноты связи слабая. |

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

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

*з) Проверить статистическую гипотезу дисперсионного анализа*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 9.9656 | 6e-05 | H1 | Ошибки нет |
| 0.05 | H1 | Ошибки нет |
| 0.1 | H1 | Ошибки нет |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.7 статистического анализа обнаружено, что  Внутригрупповые мат. ожидания неравны друг другу. Значит, есть вероятность, что существует статистическая связь между факторными и результативными признаками. |

**8. Корреляционный анализ**

*8.1. Расчёт парных коэффициентов корреляции*

Анализируемый признак 1 – C13 (Plasma beta-carotene (ng/ml))

Анализируемый признак 2 – C14 (Plasma Retinol (ng/ml))

Объёмы выборок – 315

*а) Рассчитать точечные оценки коэффициентов корреляции*

|  |  |  |
| --- | --- | --- |
|  | Формула расчета | Значение |
| Линейный коэффициент корреляции |  | 0.0716 |
| Ранговый коэффициент корреляции по Спирмену |  | 0.1306 |
| Ранговый коэффициент корреляции по Кендаллу |  | 0.0858 |

б) Привести формулы расчёта доверительного интервала для линейного коэффициента корреляции

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

*в) Рассчитать доверительные интервалы для линейного коэффициента корреляции*

|  |  |  |  |
| --- | --- | --- | --- |
| Граница доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| Нижняя граница | -0.0739 | -0.0391 | -0.0213 |
| Верхняя граница | 0.2143 | 0.1808 | 0.1635 |

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 | 1.2696 | 0.2052 | H0 |  |
|  | 0.1 | 2.3309 | 0.0204 | H' |  |
|  | 0.1 | 2.2786 | 0.0234 | H' |  |

*8.2. Расчёт множественных коэффициентов корреляции*

Анализируемый признак 1 – C7(Grams of fat consumed per day)

Анализируемый признак 2 – C9(Number of alcoholic drinks consumed per week)

Анализируемый признак 3 – C12(Dietary retinol consumed (mcg per day))

Объёмы выборок – 315

*а) Рассчитать матрицу ранговых коэффициентов корреляции по Кендаллу*

|  |  |  |  |
| --- | --- | --- | --- |
| Признак  Признак | C7 | C9 | C12 |
| C7 | 1.0 | 0.0401 | 0.3519 |
| C9 | 0.0401 | 1.0 | -0.0298 |
| C12 | 0.3519 | -0.0298 | 1.0 |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Признак  Признак | C7 | C9 | C12 |
| C7 | – | 0.314 | 0.0 |
| C9 | 0.314 | – | 0.4536 |
| C12 | 0.0 | 0.4536 | – |

*в) Рассчитать точечную оценку коэффициента конкордации*

|  |  |  |
| --- | --- | --- |
|  | Формула расчета | Значение |
| Коэффициент конкордации | n – объем выборок, k – количество выборок, - ранг i показателя определённый j экспертом | 0.4442 |

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

|  |  |  |
| --- | --- | --- |
|  | 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* | 1- |  |

*д) Проверить значимость коэффициента конкордации*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 418.4116 | 7.1302e-05 | H0 | Ошибки нет |
| 0.05 | H0 | Ошибки нет |
| 0.1 | H0 | Ошибки нет |

*Вывод (в терминах предметной области)*

|  |
| --- |
| В результате проведённого в п.8 статистического анализа обнаружено, что  Нет оснований полагать, что существует статистическая связь между факторными и результативными признаками. |

**9. Регрессионный анализ**

*9.1 Простейшая линейная регрессионная модель*

Факторный признак *x* – C6(Number of calories consumed per day)

Результативный признак *y* – C13(Plasma beta-carotene (ng/ml))

Уравнение регрессии – 

*9.1.1. Точечные оценки линейной регрессионной модели*

*а) Рассчитать точечные оценки параметров линейной регрессионной модели*

|  |  |  |
| --- | --- | --- |
| Параметр | Формула расчета | Значение |
| β0 |  | 200.6239 |
| β1 |  | -0.00597 |

*б) Записать точечную оценку уравнения регрессии*

|  |
| --- |
| *f*(*x*) = 200.6239 -0.00597 \* 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 | 16.4628 | 1 | 5185.7758 |
| Остаточные признаки | 33366.516 | 313 | 33579.7206 |
| Все признаки | 33382.9788 | 314 | 33489.294 |

д) Проверить правило сложения дисперсий

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель | *Dрегр* | *Dост* | *Dобщ* | *Dрегр* + *Dост* |
| Значение | 16.4628 | 33366.516 | 33383 | 33382.9788 |

е) Рассчитать показатели тесноты связи между факторным и результативным признаками

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Коэффициент детерминации |  | 0.0005 |
| Корреляционное отношение |  | 0.0222 |

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

|  |
| --- |
| Согласно таблице Чеддока, связь слабая. |

*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 | 124.9595 | 143.1792 | 152.4587 |
| Верхняя граница | 276.2883 | 258.0686 | 248.7892 |
| β1 | Нижняя граница | -0.0454 | -0.0359 | -0.0310 |
| Верхняя граница | 0.0334 | 0.0239 | 0.0191 |

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

## б) Проверить значимость линейной регрессионной модели

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| 0.01 | 0.1544 | 0.6946 | H0 | Ошибка отсутствует |
| 0.05 | H0 | Ошибка отсутствует |
| 0.1 | H0 | Ошибка отсутствует |

*9.2 Линейная регрессионная модель общего вида*

Факторный признак *x* – C6(Number of calories consumed per day)

Результативный признак *y* – C13(Plasma beta-carotene (ng/ml))

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 |  | 189.8920 |
| β1 |  | 0.0917 |
| β2 |  | 2.6312e-05 |

*б) Записать точечную оценку уравнения регрессии*

|  |
| --- |
| *f*(*x*) = 189.8920 + 0.0917 \* x + 2.6312e-05 \* x2 |

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

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| Факторный признак | 136.4610 | 2 | 21492.6107 |
| Остаточные признаки | 33246.5178 | 312 | 33566.1958 |
| Все признаки | 33382.9788 | 314 | 33489.2940 |

г) Проверить правило сложения дисперсий

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатель | *Dрегр* | *Dост* | *Dобщ* | *Dрегр* + *Dост* |
| Значение | 136.4610 | 33246.5178 | 33382.9788 | 33382.9788 |

д) Рассчитать показатели тесноты связи между факторным и результативным признаками

|  |  |  |
| --- | --- | --- |
| Показатель | Формула расчета | Значение |
| Коэффициент детерминации |  | 0.0041 |
| Корреляционное отношение |  | 0.0639 |

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

|  |
| --- |
| Согласно шкале Чеддока статистическая связь слабая. |

*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 |