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 Б19-504**

**Кевролетина Алексея**

**Вариант № 9**

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

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

**2021 г**

**1. Описательные статистики**

*1.1. Выборочные характеристики*

Анализируемый признак 1 – A5 Average salary - full professors

Анализируемый признак 2 – A7 Average salary - assistant professors

Анализируемый признак 3 – A8 Average salary - all ranks

*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 | 526.4836905871389 | 355.08853681267476 | 428.0326188257223 |
| Выборочная дисперсия | 13855.93100147744 | 2901.5830279615984 | 8209.962589320685 |
| Выборочное среднеквадратическое отклонение | 117.71121867297713 | 53.86634411171412 | 90.60884388027851 |
| Выборочный коэффициент асимметрии | 0.6795990400916552 | 0.44799778382079375 | 0.8202883369737385 |
| Выборочный эксцесс | 0.5385360911658728 | 0.4395453194805077 | 0.9859226807733354 |

*1.2. Группировка и гистограммы частот*

Анализируемый признак – A7 Average salary - assistant professors

Объём выборки – 1073

*а) Выбрать число групп*

| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| --- | --- | --- |
| 12 | Формула Стерджесса: | 31.4 |

*б) Построить таблицу частот*

| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Накопл. частота | Относит. накопл. частота |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 198.623 | 230.417 | 3 | 0.002796 | 3 | 0.0027959 |
| 2 | 230.417 | 261.833 | 27 | 0.025163 | 30 | 0.02795899 |
| 3 | 261.833 | 293.25 | 98 | 0.091333 | 128 | 0.11929171 |
| 4 | 293.25 | 324.667 | 186 | 0.173346 | 314 | 0.29263747 |
| 5 | 324.667 | 356.083 | 267 | 0.248835 | 581 | 0.54147251 |
| 6 | 356.083 | 387.5 | 209 | 0.194781 | 790 | 0.73625349 |
| 7 | 387.5 | 418.917 | 162 | 0.150979 | 952 | 0.88723206 |
| 8 | 418.917 | 450.333 | 69 | 0.06430568 | 1021 | 0.95153774 |
| 9 | 450.333 | 481.75 | 28 | 0.02609506 | 1049 | 0.97763281 |
| 10 | 481.75 | 513.167 | 17 | 0.01584343 | 1066 | 0.99347623 |
| 11 | 513.167 | 544.583 | 5 | 0.00465983 | 1071 | 0.99813607 |
| 12 | 544.583 | 576.0 | 2 | 0.00186393 | 1073 | 1 |

*в) Построить гистограммы частот и полигоны частот*

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

*г) Построить график эмпирической функции распределения*

| Эмпирическая функция распределения |
| --- |

**2. Интервальные оценки**

*2.1. Доверительные интервалы для мат. ожидания*

Анализируемый признак – Average salary - assistant professors

Объём выборки – 1073

Оцениваемый параметр – m

*а) Привести формулы расчёта доверительных интервалов*

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

*б) Рассчитать доверительные интервалы*

| Граница доверительного интервала | α = 0.01 | α = 0.05 | α = 0.1 |
| --- | --- | --- | --- |
| Нижняя граница | 350.845 | 351.862 | 352.381 |
| Верхняя граница | 359.332 | 358.315 | 357.796 |

*2.2. Доверительные интервалы для дисперсии*

Анализируемый признак – Average salary - assistant professors

Объём выборки – 1073

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 | 2605.304 | 2673.289 | 2708.994 |
| Верхняя граница | 3255.03 | 3166.739 | 3122.793 |

*2.3. Доверительные интервалы для разности мат. ожиданий*

Анализируемый признак 1 – Average salary - assistant professors

Анализируемый признак 2 – Average salary - all ranks

Объёмы выборок – 1073

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 | -81.244 | -79.258 | -78.242 |
| Верхняя граница | -64.644 | -66.630 | -67.646 |

*2.4. Доверительные интервалы для отношения дисперсий*

Анализируемый признак 1 – Average salary - assistant professors

Анализируемый признак 2 – Average salary - all ranks

Объёмы выборок – 1073

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 | 0.302 | 0.314 | 0.320 |
| Верхняя граница | 0.414 | 0.398 | 0.391 |

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

*3.1. Проверка статистических гипотез о математических ожиданиях*

Анализируемый признак – A7 Average salary - assistant professors

Объём выборки – 1073

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  = 0.1 | Вывод |
| --- | --- | --- | --- | --- | --- |
| 355 | 0.1 | 0.054 | 0.957 | принимается |  |
| 50 | 0.1 | 185.441 | 0 | отвергается |  |
| 300 | 0.1 | 33.484 | 0 | отвергается |  |

*3.2. Проверка статистических гипотез о дисперсиях*

Анализируемый признак – A7 Average salary - assistant professors

Объём выборки – 1073

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  = 0.1 | Вывод |
| --- | --- | --- | --- | --- | --- |
| 53 | 0.1 | 1107.333 | 0.442 | принимается |  |
| 30 | 0.1 | 3456.108 | 0 | отвергается |  |
| 80 | 0.1 | 486.015 | 0 | отвергается |  |

*3.3. Проверка статистических гипотез о равенстве математических ожиданий*

Анализируемый признак 1 – A7 Average salary - assistant professors

Анализируемый признак 2 – A8 Average salary - all ranks

Объёмы выборок – 1073

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 | -22.657 | 0 | отвергается |  |
| 0.05 | отвергается |  |
| 0.1 | отвергается |  |

*3.4. Проверка статистических гипотез о равенстве дисперсий*

Анализируемый признак 1 – A7 Average salary - assistant professors

Анализируемый признак 2 – A8 Average salary - all ranks

Объёмы выборок – 1073

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 | 0.353 | 0 | отвергается |  |
| 0.05 | отвергается |  |
| 0.1 | отвергается |  |

**4. Критерии согласия**

Анализируемый признак – A7

Объём выборки – Average salary - assistant professors

*4.1. Критерий хи-квадрат*

Теоретическое распределение – нормальное.

Статистическая гипотеза –



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

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

*б) Выбрать число групп*

| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| --- | --- | --- |
| 11 | Формула Стерджесса: | 34.27 |

*в) Построить таблицу частот*

| Номер интервала | Нижняя граница | Верхняя граница | Частота | Относит. частота | Вероятность попадания в интервал при условии истинности основной гипотезы |
| --- | --- | --- | --- | --- | --- |
| 1 | 199 | 233 | 3 | 0.0028 | 0.0099 |
| 2 | 233 | 268 | 35 | 0.0326 | 0.0402 |
| *3* | 268 | 302 | 132 | 0.1230 | 0.1093 |
| 4 | 302 | 336 | 235 | 0.2190 | 0.2008 |
| 5 | 336 | 370 | 279 | 0.2600 | 0.2495 |
| 6 | 370 | 404 | 199 | 0.1854 | 0.2095 |
| 7 | 404 | 439 | 126 | 0.1174 | 0.1189 |
| 8 | 439 | 473 | 35 | 0.0326 | 0.0457 |
| 9 | 473 | 507 | 19 | 0.0177 | 0.0118 |
| 10 | 507 | 542 | 8 | 0.0075 | 0.0021 |
| 11 | 542 | 576 | 2 | 0.0019 | 0.0002 |

*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 | 47.799 | *0* |  |  |
| 0.05 | 47.799 | *0* |  |  |
| 0.1 | 47.799 | *0* |  |  |

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

Статистическая гипотеза –



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

|  | Выражение | Пояснение использованных обозначений |
| --- | --- | --- |
| Формула расчета статистики критерия | , где | - третий момент (ассиметрия),  - четвертый момент (эксцесс),  n - число наблюдений |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |  |
| Формула расчета критической точки |  | - уровень значимости |
| Формула расчета *p-value* |  |  |

*б) Проверить статистические гипотезы*

| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| --- | --- | --- | --- | --- |
| 0.01 | 44.132 | *0* |  |  |
| 0.05 |  |  |
| 0.1 |  |  |

*Вывод (в терминах предметной области)*

| В результате проведённого в п.4 статистического анализа обнаружено, что средняя заработная плата доцентов не является нормально распределенной величиной. |
| --- |

**5. Проверка однородности выборок**

Анализируемый признак 1 – A7 Average salary - assistant professors

Анализируемый признак 2 – A8 Average salary - all ranks

Объёмы выборок – 1073

*5.1 Критерий знаков*

Статистическая гипотеза –



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

|  | Выражение | Пояснение использованных обозначений |
| --- | --- | --- |
| Формула расчета статистики критерия |  | n – объем выборок    – частота успеха    – число знаков «+» в последовательности знаков разностей  *x*1 – *y*1,…,*xn* – *yn* |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |
| Формула расчета критической точки |  |
| Формула расчета *p-value* |  |

*б) Проверить статистические гипотезы*

| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| --- | --- | --- | --- | --- |
| 0.01 | 78.000 | *0* | отвергается |  |
| 0.05 | отвергается |  |
| 0.1 | отвергается |  |

*5.2. Критерий хи-квадрат*

Статистическая гипотеза –



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

|  | Выражение | Пояснение использованных обозначений |
| --- | --- | --- |
| Формула расчета статистики критерия |  | и – относительные частоты, где и объемы выборок X и Y соответственно.  k – число интервалов. |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |
| Формула расчета критической точки |  |
| Формула расчета *p-value* |  |

*б) Выбрать число групп*

| Число групп | Обоснование выбора числа групп | Ширина интервалов |
| --- | --- | --- |
| 11 |  | 60 |

*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 | 199 | 259 | 27 | 3 | 0.03 | 0.0 |
| 2 | 259 | 320 | 262 | 98 | 0.24 | 0.09 |
| 3 | 320 | 380 | 467 | 260 | 0.44 | 0.24 |
| 4 | 380 | 441 | 256 | 306 | 0.24 | 0.29 |
| 5 | 441 | 502 | 50 | 187 | 0.05 | 0.17 |
| 6 | 502 | 562 | 10 | 134 | 0.01 | 0.12 |
| 7 | 562 | 623 | 1 | 59 | 0.0 | 0.05 |
| 8 | 623 | 684 | 0 | 15 | 0.0 | 0.01 |
| 9 | 684 | 744 | 0 | 7 | 0.0 | 0.01 |
| 10 | 744 | 805 | 0 | 3 | 0.0 | 0.0 |
| 11 | 805 | 866 | 1 | 2 | 0.0 | 0.0 |

*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 | 181.294 | 0 | отвергается |  |
| 0.05 | отвергается |  |
| 0.1 | отвергается |  |

*Вывод (в терминах предметной области)*

| В результате проведённого в п.5 статистического анализа обнаружено, что выборки средней зарплаты доцентов (A7) и сотрудников всех званий (A8) имеют разное распределение, т.к. критерий знаков и критерий хи-квадрат отвергли гипотезу однородности. |
| --- |

**6. Таблицы сопряжённости**

Факторный признак *x* – A4 Type (I, IIA, or IIB)

Результативный признак *y* – A20 Number of faculty - all ranks > all-college average?

Объёмы выборок – 1073

Статистическая гипотеза –

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

|  | Выражение | Пояснение использованных обозначений |
| --- | --- | --- |
| Формула расчета статистики критерия |  | - частота пары в выборке,  - теоретические частоты |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  | - число вариантов случайной величины ,  - число вариантов случайной величины |
| Формула расчета критической точки |  | Критическая область для статистики критерия *Z* выбирается правосторонней.  - уровень значимости |
| Формула расчета *p-value* |  |  |

*б) Построить эмпирическую таблицу сопряжённости*

| *x y* | N | Y | Σ |
| --- | --- | --- | --- |
| I | 6 | 174 | 180 |
| IIA | 207 | 152 | 359 |
| IIB | 525 | 9 | 534 |
| Σ | 738 | 335 | 1073 |

*в) Построить теоретическую таблицу сопряжённости*

| *x y* | N | Y | Σ |
| --- | --- | --- | --- |
| I | 131 | 55 | 186 |
| IIA | 245 | 111 | 356 |
| IIB | 366 | 165 | 531 |
| Σ | 742 | 331 | 1073 |

*г) Проверить статистические гипотезы*

| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| --- | --- | --- | --- | --- |
| 0.01 | 596.636 | 0 |  | A4 влияет на A20 |
| 0.05 |  | A4 влияет на A20 |
| 0.1 |  | A4 влияет на A20 |

*Вывод (в терминах предметной области)*

| В результате проведённого в п.6 статистического анализа обнаружено, что тип A4 (I, IIA, or IIB) влияет на значение выражения (А20): количество преподавателей всех рангов > среднего показателя по всем колледжам. |
| --- |

**7. Дисперсионный анализ**

Факторный признак *x* – A4 Type (I, IIA, or IIB)

Результативный признак *y* –A7 Average salary - assistant professors

Число вариантов факторного признака – 3

Объёмы выборок – 1073

Статистическая гипотеза –

*а) Рассчитать групповые выборочные характеристики*

| № п/п | Вариант факторного признака | Объём выборки | Групповые средние | Групповые дисперсии |
| --- | --- | --- | --- | --- |
| 1 | I | 180 | 412.711 | 1815.261 |
| 2 | IIA | 359 | 369.050 | 1783.017 |
| 3 | IIB | 534 | 326.279 | 1939.497 |

*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 | 1035.282 | 2 | 555428.544 |
| Остаточные признаки | 1866.301 | 1070 | 1871.534 |
| Все признаки | 2901.583 | 1072 | 2904.290 |

*г) Проверить правило сложения дисперсий*

| Показатель | *Dмежгр* | *Dвнутригр* | *Dобщ* | *Dмежгр* + *Dвнутригр* |
| --- | --- | --- | --- | --- |
| Значение | 1035.282 | 1866.301 | 2901.583 | 2901.583 |

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

| Показатель | Формула расчета | Значение |
| --- | --- | --- |
| Эмпирический коэффициент детерминации |  | 0.357 |
| Эмпирическое корреляционное отношение |  | 0.597 |

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

| По шкале Чеддока степень тесноты связи между выборками A4 и A7 **заметная**. |
| --- |

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

|  | Выражение | Пояснение использованных обозначений |
| --- | --- | --- |
| Формула расчета статистики критерия |  | К – кол-во вар-в факторного признака  n-объем выборки  - межгрупповая дисперсия  -общая дисперсия |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  |
| Формула расчета критической точки |  |
| Формула расчета *p-value* |  |

*з) Проверить статистическую гипотезу дисперсионного анализа*

| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| --- | --- | --- | --- | --- |
| 0.01 | 296.777 | 0 | отвергается | А4 влияет на А7 |
| 0.05 | отвергается | А4 влияет на А7 |
| 0.1 | отвергается | А4 влияет на А7 |

*Вывод (в терминах предметной области)*

| В результате проведённого в п.7 статистического анализа обнаружено, что тип (I, IIA, IIB) заметно влияет среднюю заработную плату доцентов. |
| --- |

**8. Корреляционный анализ**

*8.1. Расчёт парных коэффициентов корреляции*

Анализируемый признак 1 – A7 Average salary - assistant professors

Анализируемый признак 2 –A8 Average salary - all ranks

Объёмы выборок – 1073

*а) Рассчитать точечные оценки коэффициентов корреляции*

|  | Формула расчета | Значение |
| --- | --- | --- |
| Линейный коэффициент корреляции | , | 0.928 |
| Ранговый коэффициент корреляции по Спирмену | , где | 0.938 |
| Ранговый коэффициент корреляции по Кендаллу | , где  - число пар наблюдений , для которых выполнено условие  - число пар наблюдений , для которых выполнено условие | 0.787 |

*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 | 0.916 | 0.919 | 0.920 |
| Верхняя граница | 0.938 | 0.936 | 0.934 |

*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 | 81.095 | 0 |  |  |
|  | 0.1 | 88.288 | 0 |  |  |
|  | 0.1 | 38.636 | 0 |  |  |

*8.2. Расчёт множественных коэффициентов корреляции*

Анализируемый признак 1 – A5 Average salary - full professors

Анализируемый признак 2 – A7 Average salary - assistant professors

Анализируемый признак 3 – A8 Average salary - all ranks

Объёмы выборок – 1073

*а) Рассчитать матрицу ранговых коэффициентов корреляции по Кендаллу*

| Признак  Признак | A5 | A7 | A8 |
| --- | --- | --- | --- |
| A5 | 1 | 0.765 | 0.841 |
| A7 | 0.765 | 1 | 0.787 |
| A8 | 0.841 | 0.787 | 1 |

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



| Признак  Признак | A5 | A7 | A8 |
| --- | --- | --- | --- |
| A5 | – | 0 | 0 |
| A7 | 0 | – | 0 |
| A8 | 0 | 0 | – |

*в) Рассчитать точечную оценку коэффициента конкордации*

|  | Формула расчета | Значение |
| --- | --- | --- |
| Коэффициент конкордации | , где  n - объем выборок  k - число выборок  - ранг i-го элемента выборки в j-ом признаке | 0.972 |

*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 | 3126.793 | *0.000* |  | A5,A7,A8 - зависимы |
| 0.05 |  | A5,A7,A8 - зависимы |
| 0.1 |  | A5,A7,A8 - зависимы |

*Вывод (в терминах предметной области)*

| В результате проведённого в п.8 статистического анализа обнаружено, что средняя заработная плата преподавателей (А5), доцентов (A7) и средняя заработная плата сотрудников всех званий (А8) зависимы, в том числе и попарно. |
| --- |

**9. Регрессионный анализ**

*9.1 Простейшая линейная регрессионная модель*

Факторный признак *x* – A17 Number of assistant professors

Результативный признак *y* – A12 Average compensation - assistant professors

Уравнение регрессии –



*9.1.1. Точечные оценки линейной регрессионной модели*

*а) Рассчитать точечные оценки параметров линейной регрессионной модели*

| Параметр | Формула расчета | Значение |
| --- | --- | --- |
| β0 |  | 411.462 |
| β1 |  | 0.478 |

*б) Записать точечную оценку уравнения регрессии*

| *f*(*x*) = 411.462 + 0.478x |
| --- |

*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 | 1246.056 | 1 | 1337017.724 |
| Остаточные признаки | 4260.423 | 1071 | 4268.389 |
| Все признаки | 5506.479 | 1072 | 5511.616 |

*д) Проверить правило сложения дисперсий*

| Показатель | *Dрегр* | *Dост* | *Dобщ* | *Dрегр* + *Dост* |
| --- | --- | --- | --- | --- |
| Значение | 1246.056 | 4260.423 | 5506.479 | 5506.479 |

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

| Показатель | Формула расчета | Значение |
| --- | --- | --- |
| Коэффициент детерминации |  | 0.226 |
| Корреляционное отношение |  | 0.476 |

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

| Тип связи между факторным и результативным признаками, определяемой рассчитанной простейшей линейной регрессией - **умеренная**. |
| --- |

*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 | 404.212 | 405.949 | 406.837 |
| Верхняя граница | 418.711 | 416.975 | 416.087 |
| β1 | Нижняя граница | 0.409 | 0.425 | 0.434 |
| Верхняя граница | 0.548 | 0.531 | 0.523 |

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

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

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

|  |
| --- |

*д) Построить график остатков*



|  |
| --- |

*9.1.3. Проверка значимости линейной регрессионной модели*

Статистическая гипотеза –



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

|  | Выражение | Пояснение использованных обозначений |
| --- | --- | --- |
| Формула расчета статистики критерия |  | *-* статистика, для точечной оценки коэффициента детерминации |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  | Распределение Фишера с 1 и *n*–2 степенями свободы |
| Формула расчета критической точки |  | - уровень значимости |
| Формула расчета *p-value* |  |  |

*б) Проверить значимость линейной регрессионной модели*

| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| --- | --- | --- | --- | --- |
| 0.01 | 313.238 | 0 | отклоняется | Регрессионная модель значима |
| 0.05 | отклоняется | Регрессионная модель значима |
| 0.1 | отклоняется | Регрессионная модель значима |

*9.2 Линейная регрессионная модель общего вида*

Факторный признак *x* – A17 Number of assistant professors

Результативный признак *y* – A12 Average compensation - assistant professors

Уравнение регрессии – квадратичное по *x*:



*9.2.1. Точечные оценки линейной регрессионной модели*

*а) Рассчитать точечные оценки параметров линейной регрессионной модели*

| Параметр | Формула расчета | Значение |
| --- | --- | --- |
| β0 | —вектор параметров модели— вектор откликов модели | 389.18325 |
| β1 | 1.07249 |
| β2 | -0.0019638 |

*б) Записать точечную оценку уравнения регрессии*

| *f*(*x*) = 389.18 + 1.1 x - 0.00196x^2 |
| --- |

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

| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| --- | --- | --- | --- |
| Факторный признак | 1599.07 | 2 | 857901.17 |
| Остаточные признаки | 3907.41 | 1070 | 3918.36 |
| Все признаки | 5506.48 | 1072 | 5511.62 |

*г) Проверить правило сложения дисперсий*

| Показатель | *Dрегр* | *Dост* | *Dобщ* | *Dрегр* + *Dост* |
| --- | --- | --- | --- | --- |
| Значение | 1599.07 | 3907.41 | 5506.48 | 5506.48 |

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

| Показатель | Формула расчета | Значение |
| --- | --- | --- |
| Коэффициент детерминации |  | 0.290 |
| Корреляционное отношение |  | 0.539 |

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

| Тип связи между факторным и результативным признаками, определяемой рассчитанной простейшей линейной регрессией - **заметная**. |
| --- |

*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 |  | *-* статистика, для точечной оценки коэффициента детерминации |
| Закон распределения статистики критерия при условии истинности основной гипотезы |  | Распределение Фишера с *k*–1 и *n*–*k* степенями свободы |
| Формула расчета критической точки |  | - уровень значимости |
| Формула расчета *p-value* |  |  |

*б)* *Проверить значимость линейной регрессионной модели*

| Уровень значимости | Выборочное значение статистики критерия | *p-value* | Статистическое решение | Вывод |
| --- | --- | --- | --- | --- |
| 0.01 | 218.944 | 0 | отклоняется | Регрессионная модель значима |
| 0.05 | отклоняется | Регрессионная модель значима |
| 0.1 | отклоняется | Регрессионная модель значима |

*9.3 Множественная линейная регрессионная модель*

Факторный признак 1 *x*1 – A17 Number of assistant professors

Факторный признак 2 *x*2 – A7 Average salary - assistant professors

Результативный признак *y* – A12 Average compensation - assistant professors

Уравнение регрессии –



*а) Рассчитать точечные оценки параметров линейной регрессионной модели*

| Параметр | Формула расчета | Значение |
| --- | --- | --- |
| β0 | —вектор параметров модели— вектор откликов модели | -30.703 |
| β1 | -0.00987 |
| β2 | 1.3459 |

*б) Записать точечную оценку уравнения регрессии*

| *f*(*x*) = -30.703 - 0.00987+ 1.3459 |
| --- |

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

| Источник вариации | Показатель вариации | Число степеней свободы | Несмещенная оценка |
| --- | --- | --- | --- |
| Факторный признак | 5203.97 | 2 | 2791927.63 |
| Остаточные признаки | 302.51 | 1070 | 303.36 |
| Все признаки | 5506.48 | 1072 | 5511.62 |

*г) Проверить правило сложения дисперсий*

| Показатель | *Dрегр* | *Dост* | *Dобщ* | *Dрегр* + *Dост* |
| --- | --- | --- | --- | --- |
| Значение | 5203.97 | 302.519 | 5506.48 | 5506.48 |

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

| Показатель | Формула расчета | Значение |
| --- | --- | --- |
| Множественный коэффициент детерминации |  | 0.945 |
| Множественное корреляционное отношение |  | 0.972 |

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

| Тип связи между факторным и результативными признаками, определяемой рассчитанной линейной регрессией - **сильная**. |
| --- |

*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 | 1246.056 | 1599.07 | 5203.97 |
| Остаточные признаки | 4260.423 | 3907.41 | 302.51 |
| Все признаки | 5506.48 | 5506.48 | 5506.48 |

*б) Сводная таблица свойств различных регрессионных моделей*

| Свойство | Простейшая линейная модель | Линейная модель с квадратичным членом | Множественная линейная модель |
| --- | --- | --- | --- |
| Точность | 22.6% | 29.04% | 94.5% |
| Значимость | Да | Да | Да |
| Адекватность | Да | Да | Да |
| Степень тесноты связи | умеренная | заметная | сильная |

*Вывод (в терминах предметной области)*

| В результате проведённого в п.9 статистического анализа обнаружено, что все предложенные регрессионные модели адекватно отражают реальную зависимость средней компенсации доцентам (А12) от их количества (А17) и от их средней заработной платы (А12), при этом в данном случае точность простейшей линейной модели и линейной модели с квадратичным членом достаточно низкая, в отличии от множественной линейной модели. |
| --- |