**Липецкий государственный технический университет**

Физико-технологический факультет

Кафедра высшей математики

Расчетно-графическое задание

по дисциплине «Теории вероятностей и математическая статистика»

Вариант 7

|  |  |  |
| --- | --- | --- |
| Студент |  | Нэйм А.А. |
| Группа МР-19-1 |  |  |
| Руководитель |  | Денисенко Ю.И. |

**Задание 1:** Простая линейная регрессия по сгруппированным данным

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | -4,23 | -3,31 | -2,39 | -1,47 | -0,55 | 0,37 | 1,29 | 2,21 | 3,13 | 4,05 | 4,97 | 5,89 | Ʃ |
| 0,00 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1,02 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 2,04 | 0 | 2 | 2 | 2 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 |
| 3,06 | 0 | 0 | 0 | 2 | 6 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 10 |
| 4,08 | 0 | 0 | 0 | 2 | 8 | 9 | 5 | 0 | 0 | 0 | 0 | 0 | 24 |
| 5,10 | 0 | 0 | 0 | 0 | 5 | 6 | 10 | 1 | 0 | 0 | 0 | 0 | 22 |
| 6,12 | 0 | 0 | 0 | 0 | 1 | 10 | 13 | 12 | 0 | 0 | 0 | 0 | 36 |
| 7,14 | 0 | 0 | 0 | 0 | 0 | 3 | 8 | 10 | 8 | 0 | 0 | 0 | 29 |
| 8,16 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 11 | 8 | 4 | 0 | 0 | 28 |
| 9,18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 12 | 3 | 1 | 0 | 21 |
| 10,20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 8 | 0 | 1 | 11 |
| 11,22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 4 |
| Ʃ | 1 | 2 | 3 | 6 | 24 | 29 | 42 | 39 | 30 | 16 | 2 | 3 | 197 |

Найти уравнение регрессии y на x и x на y.

Решение.

Выборочные средние – это величины, которые определяются по формулам:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Вычислим все необходимые суммы:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | -4,23 | -3,31 | -2,39 | -1,47 | -0,55 | 0,37 | 1,29 | 2,21 | 3,13 | 4,05 | 4,97 | 5,89 | Ʃ |
| Ʃ | 1 | 2 | 3 | 6 | 24 | 29 | 42 | 39 | 30 | 16 | 2 | 3 | 197 |
|  | -4,23 | -6,62 | -7,17 | -8,82 | -13,2 | 10,73 | 54,18 | 86,19 | 93,9 | 64,8 | 9,94 | 17,67 | 297,26 |
|  | 17,89 | 21,91 | 17,14 | 12,97 | 7,26 | 3,97 | 69,89 | 190,47 | 293,91 | 262,44 | 49,4 | 104,08 | 1051,33 |
|  | 0 | 2,04 | 1,7 | 3,06 | 3,74 | 5,28 | 6,00 | 7,32 | 8,43 | 9,56 | 10,2 | 10,88 | 0 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0,00 | 1,02 | 2,04 | 3,06 | 4,08 | 5,10 | 6,12 | 7,14 | 8,16 | 9,18 | 10,20 | 11,22 | Ʃ |
| Ʃ | 1 | 2 | 9 | 10 | 24 | 22 | 36 | 29 | 28 | 21 | 11 | 4 | 197 |
|  | 0 | 2,04 | 18,36 | 30,6 | 97,92 | 112,2 | 220,32 | 207,06 | 228,48 | 192,78 | 112,2 | 44,88 | 1263,84 |
|  | 0 | 2,08 | 37,45 | 93,63 | 399,51 | 572,22 | 1348,35 | 1478,40 | 1864,39 | 1763,72 | 1144,44 | 503,55 | 9207,74 |
|  | -4,23 | -1,47 | -1,78 | -0,46 | 0,1 | 0,66 | 1,29 | 2,02 | 2,57 | 3,13 | 4,05 | 5,2 |  |

В результате получаем выборочные средние:

Дисперсией называется средняя арифметическая квадратов отклонений вариантов от их средней арифметической.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

В результате получаем выборочные дисперсии:

Выборочный коэффициент корреляции – это показатель тесноты линейной связи между X и Y, который определяется по формуле:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Вычислим все необходимые суммы:

В результате получаем выборочный коэффициент корреляции:

Связь между переменными Y и X можно описать линейным уравнением, которое называется линейным уравнением регрессии. Оно определяется по формуле:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Аналогично определяется уравнение регрессии X по Y:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

Коэффициент регрессии показывает, на сколько единиц в среднем изменяется переменная Y при увеличении переменной X на одну единицу.

Коэффициент регрессии показывает, на сколько единиц в среднем изменяется переменная X при увеличении переменной Y на одну единицу.

Найдем значения коэффициентов регрессии по формулам:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |
|  |  | (7) |

В результате получаем:

По формулам (4) - (5) находим уравнения линейной регрессии:

Для каждого значения т.е. для каждой строки корреляционной таблицы вычислим групповые средние

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

где частоты пар и число интервалов по переменной Y.

Вычисленные групповые средние поместим в последнем столбце корреляционной таблицы и изобразим графически в виде ломаной, называемой эмпирической линией регрессии Y по X.

Аналогично для каждого значения по формуле

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

вычислим групповые средние , где число интервалов по переменной X.

Вычисленные групповые средние поместим в последнюю строку корреляционной таблицы 1.

и т.д.

Аналогично и для X:

и т.д.

Поместим все необходимые данные в таблицу 1.

Изобразим полученную зависимость графически точками координатной плоскости. Такая зависимость называется полем корреляции.

Построим групповые средние и прямые линии регрессии на поле корреляции.

Рисунок 1. Поле корреляции.

Оценим статистическую значимость параметров регрессии с помощью критерия Стьюдента. Оценку статистической значимости параметров регрессии проведем с помощью t-статистики Стьюдента и путем расчета доверительного интервала каждого из показателей.

Выдвигаем гипотезу о статистической незначимости коэффициента корреляции.

Гипотеза отвергается, т.е. выборочный коэффициент корреляции значимо отличается от нуля, если

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

где табличное значение t-критерия Стьюдента, определяемое при уровне значимости при числе степеней свободы

Статистика критерия по формуле (10) составит:

Для уровня значимости и числа степеней свободы находим критическое значение статистики Поскольку коэффициент корреляции между Y и X значимо отличается от нуля.

Для значимого коэффициента корреляции целесообразно найти доверительный интервал (интервальную оценку), который с заданной надежностью содержит неизвестный генеральный коэффициент корреляции . Для построения такого интервала необходимо знать выборочное распределение коэффициента корреляции , которое при несимметрично и очень медленно (с ростом n) сходится к нормальному распределению. Поэтому прибегают к специально подобранным функциям от r, которые сходятся к хорошо изученным распределениям. Чаще всего для подбора функции применяют z-преобразование Фишера:

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

Так как коэффициент корреляции значим, то построим доверительный интервал для генерального коэффициента корреляции , применяя z-преобразование Фишера.

По формуле (11) получаем:

Распределение z является нормальным с математическим ожиданием

|  |  |  |
| --- | --- | --- |
|  |  | (12) |

и дисперсией

|  |  |  |
| --- | --- | --- |
|  |  | (13) |

Сначала строим доверительный интервал для

|  |  |  |
| --- | --- | --- |
|  |  | (14) |

где нормированное отклонение z, определяемое с помощью функции Лапласа:

|  |  |  |
| --- | --- | --- |
|  |  | (15) |

По формуле (15) из условия по таблице функции Лапласа находим

По формуле (14) построим доверительный интервал для M(z).

или

При определении границ доверительного интервала для т.е. для перехода от z к , существует специальная таблица. При ее отсутствии переход может быть осуществлен по формуле:

|  |  |  |
| --- | --- | --- |
|  |  | (16) |

где гиперболический тангенс z.

По формуле (16) находим границы доверительного интервала для

Если коэффициент корреляции значим, то коэффициенты регрессии и также значимо отличаются от нуля, а интервальные оценки для генеральных коэффициентов регрессии и могут быть получены по формулам, основанным на том, что статистики имеют t-распределение Стьюдента с n-2 степенями свободы:

|  |  |
| --- | --- |
|  | (17) |
|  | (18) |

По формулам (17) - (18) находим доверительные интервалы для и :

|  |
| --- |
|  |
|  |

Оценим статистическую значимость полученного уравнения регрессии . Выдвигаем гипотезу о статистически незначимом отличии показателей от нуля:

для числа степеней свободы и составит

Определим случайные ошибки

Находим фактические значения t-статистики:

и

поэтому гипотеза не принимается, т.е. параметры уравнения регрессии и статистически значимы.

Аналогично оценим статистическую значимость полученного уравнения регрессии . Выдвигаем гипотезу о статистически незначимом отличии показателей от нуля:

для числа степеней свободы и составит

Определим случайные ошибки

Находим фактические значения t-статистики:

и

поэтому гипотеза не принимается, т.е. параметры уравнения регрессии и статистически значимы.

Для построения гистограммы по оси абсцисс указывают значения границ интервалов и на их основании строят прямоугольники, высота которых пропорциональна частотам.

Для направляющих X и Y построим гистограммы.

Рисунок 2. Гистограмма распределения X.

Рисунок 3. Гистограмма распределения Y.

По виду гистограммы распределения (рис. 2) можно предположить нормальный

закон распределения признака. Итак, выдвигаем гипотезу случайная величина X распределена нормально с параметрами

Проверку гипотезы осуществляем с помощью критерия Пирсона.

Для расчета вероятностей попадания величины X в интервал используем формулу Лапласа в соответствии со свойством нормального распределения:

Для определения статистики удобно составить таблицу:

Таблица 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| i |  | Эмпирические частоты, | Вероятности, | Теоретические частоты, |  |  |
| 1 | -0,15 | 1 |  |  |  |  |
| 2 | 0,67 | 6 | 0,016 | 3,07 | 8,563 | 2,786 |
| 3 | 1,49 | 12 | 0,039 | 7,72 | 18,344 | 2,377 |
| 4 | 2,31 | 19 | 0,079 | 15,65 | 11,204 | 0,716 |
| 5 | 3,13 | 35 | 0,130 | 25,65 | 87,385 | 3,407 |
| 6 | 3,95 | 34 | 0,172 | 33,97 | 0,001 | 0,000 |
| 7 | 4,77 | 35 | 0,184 | 36,34 | 1,802 | 0,050 |
| 8 | 5,59 | 24 | 0,159 | 31,42 | 55,037 | 1,752 |
| 9 | 6,41 | 16 | 0,111 | 21,95 | 35,367 | 1,611 |
| 10 | 7,23 | 11 | 0,063 | 12,39 | 1,924 | 0,155 |
| 11 | 8,05 | 4 | 0,029 | 5,65 | 2,718 | 0,481 |
| 12 | 8,87 | 1 | 0,011 | 2,08 | 1,168 | 0,561 |
|  |  | 198 | 0,99 | 195,89 | 223,513 | 13,896 |

Итак, наблюдаемое значение критерия равно 13.896.

Число интервалов равно , а нормальный закон распределения определяется параметрами, то число степеней свободы равно . Соответствующее критическое значение по таблице распределения равно

Так как , то гипотеза о выбранном теоретическом нормальном законе согласуется с опытными данными.

Аналогично, выдвигаем гипотезу о нормальном распределении величины Y. Итак, выдвигаем гипотезу случайная величина Y распределена нормально с параметрами

Для определения статистики удобно составить таблицу:

Таблица 2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| j |  | Эмпирические частоты, | Вероятности, | Теоретические частоты, |  |  |
| 1 | -7,09 | 3 |  |  |  |  |
| 2 | -5,91 | 4 | 0,015 | 3,018 | 0,965 | 0,320 |
| 3 | -4,73 | 11 | 0,033 | 6,610 | 19,276 | 2,916 |
| 4 | -3,55 | 16 | 0,062 | 12,349 | 13,332 | 1,080 |
| 5 | -2,37 | 24 | 0,099 | 19,680 | 18,660 | 0,948 |
| 6 | -1,19 | 23 | 0,135 | 26,755 | 14,102 | 0,527 |
| 7 | -0,01 | 43 | 0,157 | 31,028 | 143,326 | 4,619 |
| 8 | 1,17 | 19 | 0,155 | 30,696 | 136,787 | 4,456 |
| 9 | 2,35 | 19 | 0,131 | 25,904 | 47,667 | 1,840 |
| 10 | 3,53 | 20 | 0,094 | 18,648 | 1,828 | 0,098 |
| 11 | 4,71 | 10 | 0,068 | 13,464 | 11,999 | 0,891 |
| 12 | 5,89 | 6 | 0,040 | 7,920 | 3,686 | 0,465 |
|  |  | 198 | 1,00 | 196,071 | 411,627 | 18,161 |

Итак, наблюдаемое значение критерия равно 18.161

Число интервалов равно , а нормальный закон распределения определяется параметрами, то число степеней свободы равно . Соответствующее критическое значение по таблице распределения равно

Так как , то гипотеза о выбранном теоретическом нормальном законе согласуется с опытными данными.

Задание 2

Найти: уравнение регрессии

Исходные данные:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 0,35 | 0,45 | 0,55 | 0,65 | 0,75 | 0,85 | 0,95 | 1,05 | 1,15 | 1,25 |
| Y | 6,5 | 4,58 | 7,05 | 5,29 | 7,26 | 5,47 | 7,54 | 5,99 | 8,04 | 6,84 |
| X | 1,35 | 1,45 | 1,55 | 1,65 | 1,75 | 1,85 | 1,95 | 2,05 | 2,15 | 2,25 |
| Y | 8,86 | 7,54 | 9,46 | 8,14 | 9,87 | 8,47 | 11,01 | 9,35 | 11,29 | 9,96 |

Решение:

Рисунок 1 – поле корреляции

Для заданной зависимости вычисления проводим по следующим формулам:

*  *

   

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | y | ln(yi) | xi\*ln(yi) |  |  |  |  |  |
| 0,35 | 6,5 | 0,81291 | 0,28452 | 0,12 | 42,25 | 2,275 | 8,278 | 0,551 |
| 0,45 | 4,58 | 0,66087 | 0,29739 | 0,2 | 20,98 | 2,061 | 8,473 | 1,261 |
| 0,55 | 7,05 | 0,84819 | 0,4665 | 0,3 | 49,7 | 3,8775 | 8,673 | 1,225 |
| 0,65 | 5,29 | 0,72346 | 0,47025 | 0,42 | 27,98 | 3,4385 | 8,787 | 0,669 |
| 0,75 | 7,26 | 0,86094 | 0,6457 | 0,56 | 52,71 | 5,445 | 9,088 | 0,612 |
| 0,85 | 5,47 | 0,73799 | 0,62729 | 0,72 | 29,92 | 4,6495 | 9,302 | 0,693 |
| 0,95 | 7,54 | 0,87737 | 0,8335 | 0,9 | 56,85 | 7,163 | 9,522 | 1,272 |
| 1,05 | 5,99 | 0,77743 | 0,8163 | 1,1 | 35,88 | 6,2895 | 9,747 | 1,014 |
| 1,15 | 8,04 | 0,90526 | 1,04104 | 1,32 | 64,64 | 9,246 | 9,977 | 0,452 |
| 1,25 | 6,84 | 0,83506 | 1,04382 | 1,56 | 46,79 | 8,55 | 10,213 | 0,87 |
| 1,35 | 8,86 | 0,94743 | 1,27904 | 1,82 | 78,5 | 11,961 | 10,454 | 0,839 |
| 1,45 | 7,54 | 0,87737 | 1,27219 | 2,1 | 56,85 | 10,933 | 10,701 | 0,361 |
| 1,55 | 9,46 | 0,97589 | 1,51263 | 2,4 | 89,49 | 14,663 | 10,954 | 0,972 |
| 1,65 | 8,14 | 0,91062 | 1,50253 | 2,72 | 66,26 | 13,431 | 11,212 | 0,984 |
| 1,75 | 9,87 | 0,99432 | 1,74006 | 3,06 | 97,42 | 17,2725 | 11,477 | 0,888 |
| 1,85 | 8,47 | 0,92788 | 1,71658 | 3,42 | 71,74 | 15,6695 | 11,748 | 1,017 |
| 1,95 | 11,01 | 1,04179 | 2,03149 | 2,8 | 121,22 | 21,4695 | 12,026 | 1,45 |
| 2,05 | 9,35 | 0,97081 | 1,99016 | 4,2 | 87,42 | 19,1675 | 12,310 | 0,792 |
| 2,15 | 11,29 | 1,05269 | 2,26329 | 4,62 | 127,46 | 24,2735 | 12,601 | 1,298 |
| 2,25 | 9,96 | 0,99826 | 2,24608 | 5,06 | 99,2 | 22,41 | 12,898 | 0,312 |
| Ʃ 26 | 158,51 | 17,74 | 24,08 | 40,45 | 1323,27 | 224,25 |  | 17,536 |

Подставляя числовые значения в формулы выше, полечаем следующие значения:

; 

; 

; ; ; 

; 

В соответствии с проведенными расчетами искомая зависимость имеет вид 



