Каримова Альбина Миннуровна гр. 05-107 (var10\_Z2)

**Задание 4.1.** Проверка независимости по критерию хи-квадрат

1. В исследовательской группе измеряется коэффициент прочности металлов и их цена. Требуется выяснить, существует ли статистически значимая зависимость между прочностью металла и его ценой.
2. В эксперименте участвует набор металлов в количестве штук, они выбираются случайным образом.
3. Для проверки независимости случайных данных по критерию хи-квадрат необходимо построить таблицу сопряженности, в которой будут показаны частоты прочности металлов и их стоимости.
4. Ожидается, что прочность и цена металла взаимосвязаны (данные зависимы). Нулевая гипотеза : данные независимы, альтернативная гипотеза : данные зависимы.
5. Уровень значимости .
6. Для проверки независимости мы применим критерий хи-квадрат к таблице сопряженности. Для построения таблицы сопряженности, области значений признаков разбиваются на интервалов соответственно. Далее подсчитываются количества попаданий всех пар выборочных значений в каждую двумерную ячейку и заполняется таблица, где вычисляются суммы соответствующих строк и столбцов. Статистика критерия сопряженности вычисляется по формуле:

Ожидания будут подтверждены, если статистика примет достаточно большие значения, т.е. критическая область будет иметь вид .

1. Если случайные величины независимы, то значение представляет собой реализацию случайной величины с распределением, приближенно описываемым распределением хи-квадрат с

степенями свободы.

1. Критическая константа находится из уравнения

,

1. т.е. равна верхней 0.01-квантили распределения хи-квадрат с 6 степенями свободы.
2. Воспользовавшись пакетами pandas, numpy, scipy.stats в Python, нашли, что

Окончательный вид критической области .

1. a. По представленным данным найдено

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | 124.05 |  | Всего |
| > 87.05 | 0 | 6 | 10 | 16 |
| 87.05 | 0 | 37 | 0 | 37 |
| 84.05 | 0 | 24 | 0 | 24 |
|  | 10 | 4 | 0 | 14 |
| Всего | 10 | 71 | 10 |  |
| Статистика | | | 113.407 | |
| Степени свободы | | | 6 | |
| 1-% критическая область | | |  | |
| Гипотеза независимости | | | отвергается | |
| С критическим уровнем значимости | | |  | |

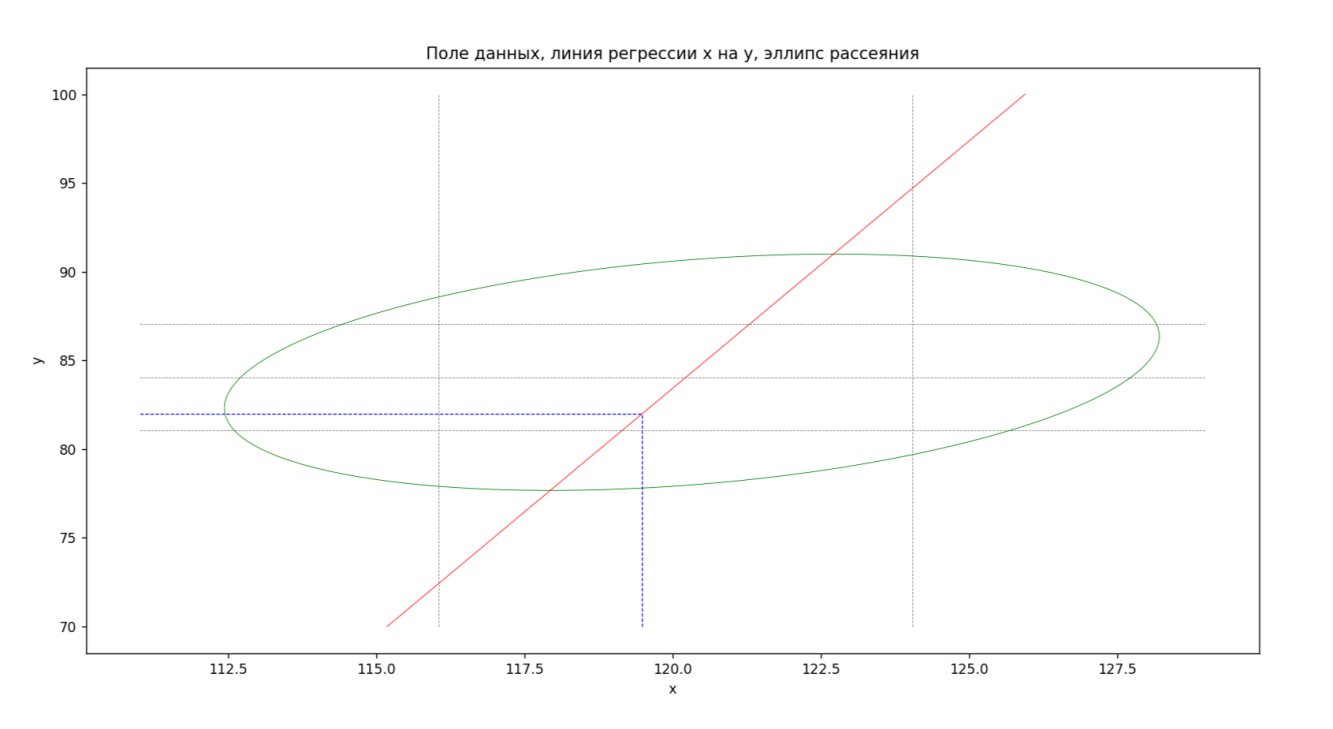
b. Критический уровень значимости вычисляется по формуле

Так как , следует считать наблюдения зависимыми.

**Задание 4.2.** Наилучший прогноз, эллипс рассеяния

1. Инвестиционная компания хочет спрогнозировать цены акций двух различных компаний для принятия решений по инвестициям. Данные включают исторические цены акций за последние месяцев. Требуется найти оценки коэффициентов линейной среднеквадратической регрессии стоимости акций второй компании (η) на стоимость акций первой компании (ξ); представить график линии регрессии в поле всех данных; найти прогноз ξ при значении η=82; дать оценку точности прогноза и изобразить эллипс рассеяния.
2. Были собраны данные о стоимости акций двух различных компаний за последние месяцов. Случайной величиной будет являться цена акций за какой-либо месяц.
3. По представленным данным найдено

|  |  |
| --- | --- |
| Коэффициент линейной регрессии |  |
| Уравнение регрессии η на ξ |  |
| Прогноз при y=82 |  |
| Коэфф.корреляции |  |
| Станд.отклонение наблюдений прочности |  |
| Оценка станд.ошибки прогноза | 36.247 |



Вывод: при таком невысоком значении коэффициента корреляции и высокой оценки стандартной ошибки прогноза прогностические качества линии регрессии очень низкие.

**Задание 4.3.** Проверка независимости по выборочному коэффициенту корреляции

1. Исследователи хотят выяснить, существует ли взаимосвязь между уровнем физической активности и уровнем холестерина у взрослых пациентов. Они проводят исследование, чтобы определить, влияет ли регулярная физическая активность на уровень холестерина в крови.
2. Измеряется уровень физической активности и холестерина у взрослых пациентов, которые будут выбираться случайным образом.
3. Используется выборочный коэффициент корреляции, который определяется как

,

где - соответствующие выборочные средние, - выборочные стандартные отклонения (на основе смещенной выборочной дисперсии).

1. Ожидается, что уровень физической активности и холестерина взаимосвязаны. Нулевая гипотеза : , альтернативная гипотеза : .
2. Уровень значимости .
3. Для данной задачи применим критерий независимости компонент двумерного случайного вектора. Тестовой статистикой является преобразование Стьюдента для выборочного коэффициента корреляции

Ожидания будут подтверждены, если абсолютное значение статистики примет значение не больше единицы, т.е. критическая область будет иметь вид {.

1. Если выборка получена из двумерного нормального закона, то значение представляет собой реализацию случайной величины с распределением Стьюдента с степенями свободы.
2. Критическая константа находится из уравнения

,

1. т.е. равна верхней 0.05-квантили распределения Стьюдента с 89 степенями свободы.
2. Воспользовавшись пакетами pandas, numpy, scipy.stats в Python, нашли, что

Окончательный вид критической области .

1. a. По представленным данным найдено

|  |  |  |  |
| --- | --- | --- | --- |
|  | x | | y |
| Среднее | 120.32 | | 84.34 |
| Дисперсия | 15.57 | | 11.09 |
| Станд.отклонение s | 3.95 | | 3.33 |
| Объем выборки, n | 91 | | 91 |
| Коэффициент корреляции R | | 0.303 | |
| Преобразование Стьюдента t | | 3.00014 | |
| 10-% критическая область | |  | |
| Гипотеза независимости | | отвергается | |
| С критическим уровнем значимости | |  | |

b. Критический уровень значимости вычислялся по формуле

Так как , следует считать отклонение выборочного коэффициента корреляции от нуля статистически значимым.