

Санкт-Петербургский Государственный Университет

Поляков Иван Михайлович

Отчёт по Лабораторной Работе № 9_1

Проверка наличия зависимости между номинальными признаками (таблицы сопряжённости)

Направление 01.04.02: «Прикладная математика и информатика»
Образовательная программа ВМ.5505.2021: «Математическое и информационное обеспечение экономической деятельности»

Преподаватель:
доктор технических наук,
профессор Буре Владимир Мансурович

Санкт-Петербург
2022 г.

1 Постановка задачи

По смоделированным выборкам из нормального распределения с произвольными математическими ожиданиями a_x , a_y и произвольными дисперсиями σ_x^2 , σ_y^2 необходимо проверить гипотезу о независимости номинальных признаков, используя 2 статистики

$$X^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{\left(n_{ij} - \frac{n_i m_j}{n}\right)^2}{\frac{n_i m_j}{n}} \quad (1)$$

$$Y^2 = 2 \sum_{i=1}^r \sum_{j=1}^s n_{ij} \ln \left(\frac{n_{ij}}{\frac{n_i m_j}{n}} \right), \quad (2)$$

где

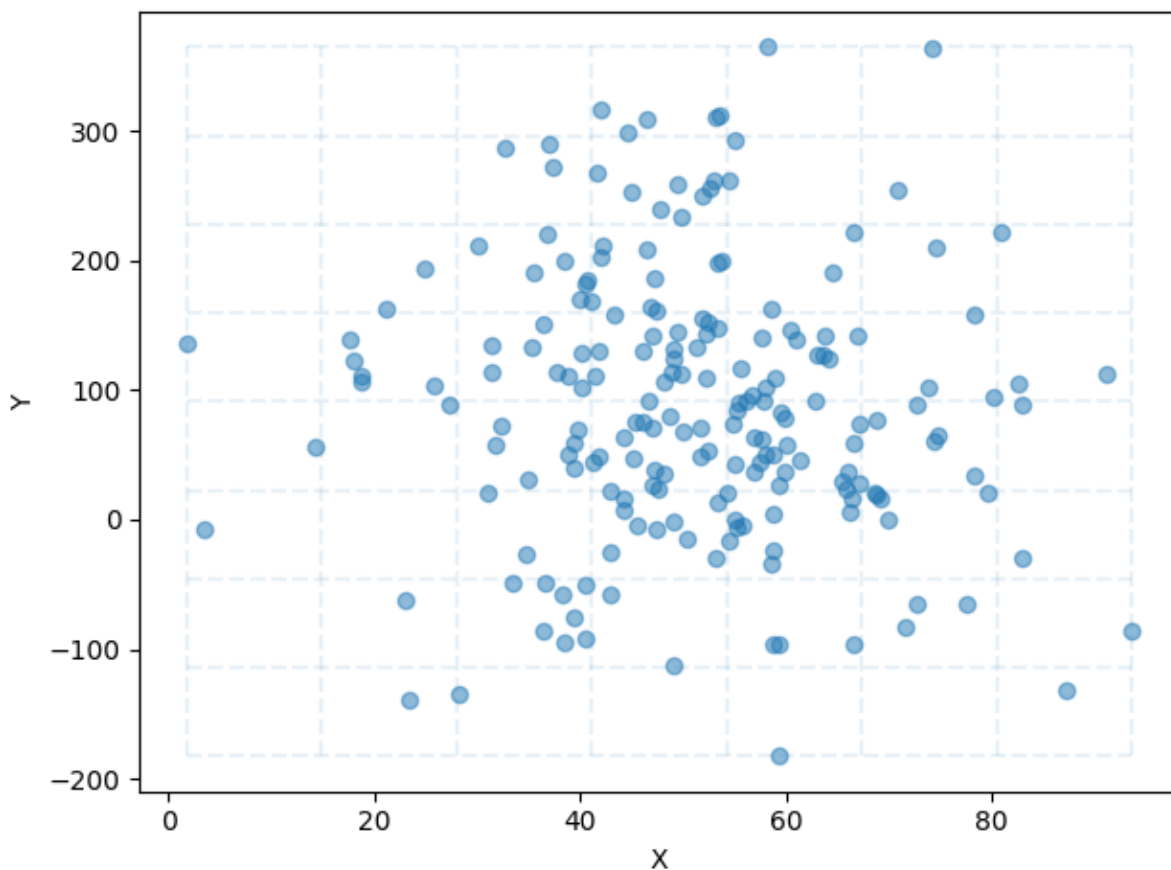
$$n_i = \sum_{j=1}^s n_{ij};$$

$$m_j = \sum_{i=1}^r n_{ij};$$

$$n = \sum_{i=1}^r \sum_{j=1}^s n_{ij};$$

2 Ход работы

Были сгенерированы выборки объёмом $n = 200$ из нормального распределения (параметры и сами выборки находятся в конце файла). Поле наблюдений представлено на графике снизу.



Можно разбить ось x на r непересекающихся промежутков, а ось y – на s промежутков. В результате можно считать, что имеется два номинальных признака – признак A имеет r уровней A_1, \dots, A_r ; признак B имеет s уровней B_1, \dots, B_s . Таким образом, имеется выборка случайно отобранных n объектов из генеральной совокупности, по которой нужно найти частоты совместной встречаемости событий $A_i \cap B_j$.

Частоты собираются в таблицу, называемую таблицей сопряжённости. Ниже такая таблица сформирована по моделированным выше выборкам

Таблица сопряжённости

[0	0	1	1	1	0	0	0]
[1	1	0	1	5	2	0	0]
[1	8	2	7	8	8	3	0]
[0	2	10	16	18	8	8	5]
[1	3	10	23	14	3	2	1]
[0	3	5	5	3	1	1	1]
[1	1	1	1	2	1	0	0]]

Таблица численно показывает распределение наблюдений, которые изображены на графике.

Пусть $p_{ij} = P(A_i \cup B_j)$, $p_i = P(A_i)$, $q_j = P(B_j)$. Сформулируем гипотезу о независимости признаков A и B : $H_0 : p_{ij} = p_i q_j$, $i = \overline{1, r}$, $j = \overline{1, s}$. При этом $p_i = \sum_{j=1}^s p_{ij}$, $q_i = \sum_{i=1}^r p_{ij}$

Метод проверки гипотезы H_0 основан на статистиках (1) и (2). Если признаки A и B независимы, то статистики X^2 и Y^2 имеют распределение χ^2 с $(r-1)(s-1)$ степенями свободы.

Таким образом,

- Для независимых признаков статистика X^2 асимптотически распределена по закону χ^2 с $(r-1)(s-1)$ степенями свободы.
- Для зависимых признаков X^2 неограниченно возрастает при увеличении n

Для проверки гипотезы необходимо вычислить одну из статистик (1) и (2) и сравнить их с соответствующими критическими значениями χ_α^2 – верхняя квантиль уровня α из распределения χ^2 с $(r-1)(s-1)$ степенями свободы.

Гипотеза H_0

- Принимается, если $X^2 < \chi_\alpha^2$ ($Y^2 < \chi_\alpha^2$)
- Отвергаем, если $X^2 \geq \chi_\alpha^2$ ($Y^2 \geq \chi_\alpha^2$)

Результаты и исходные данные приведены ниже.

Параметры для формирования выборок

```
a_x : 51
a_y : 92
sigma_x: 15
sigma_y: 107
```

```
r: 9
s: 8
```

Гипотеза о независимости номинальных признаков

Гипотеза H_0 о независимости признаков принимается на уровне значимости 0.05, так как значение $X^2 < \chi^2$ ($55.78 < 74.47$)

Гипотеза H_0 о независимости признаков принимается на уровне значимости 0.05, так как значение $Y^2 < \chi^2$ ($56.14 < 74.47$)

Полученные выборки

X Y

48.96, 113.79
31.77, 57.14
49.47, 258.48
44.55, 299.32
34.99, 30.88
56.82, 37.53
27.22, 88.14
59.83, 78.13
74.14, 363.78
3.54, -7.62
87.28, -132.19
55.48, 90.36
55.25, -6.43
14.19, 55.49
21.09, 162.22
51.93, 154.66
47.19, 38.36
25.83, 103.43
80.09, 94.36
40.05, 129.31
30.94, 19.99
55.52, 116.31
46.51, 309.36
82.88, 88.44
49.75, 232.93
67.20, 74.32
58.03, 50.79
49.13, 131.82
17.72, 138.75
74.60, 209.62
53.39, 13.01
46.63, 91.78
18.82, 111.42
39.96, 170.55
40.45, -91.00
58.15, 364.41
55.19, 83.93
65.50, 29.50
36.89, 219.61
46.85, 163.40
37.40, 271.27
61.37, 45.15
48.17, 106.19
38.80, 111.50
53.56, 312.49
23.07, -61.77
66.06, 37.63
49.11, -1.58
47.01, 25.93
72.77, 88.05

60.07, 57.25
44.26, 63.42
53.66, 200.28
51.78, 249.92
77.47, -65.58
66.20, 5.31
60.99, 139.41
52.49, 53.26
54.29, 21.23
78.23, 34.43
55.07, 42.72
42.08, 202.29
42.97, -25.23
36.40, 150.87
18.70, 106.21
48.10, 34.81
47.00, 70.34
42.85, 21.91
18.07, 122.31
38.31, -57.20
59.36, -182.27
63.11, 126.91
55.03, -0.26
72.73, -65.16
56.82, 64.00
71.58, -83.21
66.55, 221.33
49.09, -113.12
31.38, 113.29
46.05, 75.43
57.55, 139.67
57.43, 44.38
54.46, -16.61
42.02, 315.99
53.31, 197.94
46.53, 207.79
79.60, 19.91
66.48, -96.79
59.82, 36.89
40.21, 101.96
41.17, 43.94
24.81, 193.20
59.33, 26.93
32.67, 286.22
48.71, 80.52
38.47, 199.30
69.87, -0.22
50.07, 68.03
53.39, 147.96
58.67, -24.39
34.75, -27.15
54.97, 292.52

52.50, 152.65
38.54, -94.01
74.69, 64.57
54.76, 74.09
58.64, 162.18
68.66, 20.12
55.75, -4.19
42.94, -57.72
32.24, 72.67
49.36, 144.60
39.34, -76.00
47.84, 238.97
68.87, 76.48
44.17, 16.35
33.53, -48.95
45.47, -4.78
70.82, 253.68
41.90, 129.86
56.69, 96.48
1.81, 135.29
45.32, 74.67
43.34, 157.58
36.37, -86.41
52.28, 109.79
50.43, -15.00
58.54, -33.64
41.78, 48.41
40.60, -50.28
36.59, -48.77
39.35, 39.37
53.19, 310.68
47.35, -7.86
28.28, -134.44
41.07, 168.74
40.73, 184.72
58.71, 49.70
59.40, 82.51
35.33, 133.16
93.59, -86.05
41.66, 267.12
80.83, 221.37
47.63, 23.46
23.43, -139.35
45.02, 252.70
56.15, 92.28
58.05, 102.58
52.16, 143.23
49.03, 123.52
52.99, 261.95
54.52, 262.29
30.01, 211.12
46.01, 129.33

40.51, 182.42
63.79, 141.20
49.88, 111.79
82.95, -29.29
47.15, 186.41
64.53, 190.21
63.54, 126.36
35.50, 190.12
51.31, 133.66
58.69, 4.48
67.09, 27.71
44.23, 7.41
39.43, 58.54
59.35, -96.62
64.14, 124.20
45.17, 46.86
31.45, 134.89
74.39, 60.83
47.11, 142.53
51.67, 48.05
65.78, 23.39
57.84, 90.94
60.43, 145.93
38.86, 50.29
52.56, 255.41
58.68, -96.34
39.85, 69.83
36.98, 289.56
82.51, 104.90
51.63, 71.35
41.39, 110.47
91.11, 112.49
78.26, 158.11
53.20, -29.70
68.82, 19.84
37.76, 113.75
42.21, 211.08
58.88, 108.66
66.33, 15.94
66.96, 142.00
47.37, 160.81
69.22, 15.82
66.61, 58.67
73.79, 101.22
62.86, 91.08
57.56, 62.01