Санкт-Петербургский Государственный Университет

Поляков Иван Михайлович

Отчёт по Лабораторной Работе № 8_3

Двуфакторный дисперсионный анализ

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

Следующая таблица содержит данные о товарообороте для магазинов трёх типов в течение четырёх лет после открытия магазина. В клетках стоят средние значения по группе магазинов.

```
Исходные данные
s_lit: [7.45 8.23 8.61 7.12]
g_lit: [6.73 6.85 7.55 6.58]
m_lit: [5.41 6.13 5.57 3.73]
```

На уровне значимости 5% необходимо проверить нулевые гипотезы об отсутствии эффектов столбцов и строк.

2 Ход работы

В отличие от однофакторного дисперсионного анализа, в двухфакторном анализе значения наблюдений как в строках, так и в столбцах являются неоднородным.

Каждое наблюдение может быть представлено в виде

$$x_{ij} = \mu + b_i + t_j + \epsilon_{ij}, i = \overline{1,n}; j = \overline{1,k} \tag{1}$$

$$\sum_{i=1}^{n} b_i = 0; \sum_{j=1}^{k} t_j = 0. \tag{2}$$

ϵ_{ij} предполагаются независимыми, одинаково распределёнными, подчиняющиеся нормальному распределению $N(0, \sigma^2)$.

Можно показать, что общая вариация равна сумме вариаций строк, вариаций столбцов и остаточной вариации:

$$\sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij} - \overline{x}_{..})^2 = k\sum_{i=1}^{n}(\overline{x}_{i.} - \overline{x}_{..})^2 + n\sum_{j=1}^{k}(\overline{x}_{.j} - \overline{x}_{..})^2 + \sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij} - \overline{x}_{i.} - \overline{x}_{.j} + \overline{x}_{..})^2 \tag{3}$$

где

$$\overline{x}_{..} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij},$$

$$\overline{x}_{i.} = \frac{1}{k}\sum_{j=1}^{k} x_{ij}$$

$$\overline{x}_{.j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

В рассматриваемой модели наблюдений предполагается, что взаимодействие между эффектами строк и столбцов отсутствует. Последнее слагаемое в (3)

$$S_1 = \sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij} - \overline{x}_{i.} - \overline{x}_{.j} + \overline{x}_{..})^2 \tag{4}$$

будет иметь $(n-1)(k-1)$ степеней свободы.

Сформулируем две нулевые гипотезы, которые заключаются в том, что влияние каждого из факторов отсутствует.

$H_0 : b_1 = b_2 = \ldots = b_n = 0$ - влияние фактора строк отсутствует; альтернативная - H_1 : - не все b_i равны нулю.

$H_0 : t_1 = t_2 = \ldots = t_k = 0$ - влияние фактора столбцов отсутствует; альтернативная - H_1 : - не все t_i равны нулю.

- Если верна первая нулевая гипотеза, то модель наблюдений принимает вид

$$x_{ij} = \mu + t_j + \epsilon_{ij} \tag{5}$$

$$S_R = k \sum_{i=1}^{n} (\overline{x}_{i.} - \overline{x}_{..})^2$$

будет иметь $(n-1)$ степень свободы.

При справедливости гипотезы $H_0 : b_i = 0$ статистика

$$F_1 = \frac{\frac{S_R}{n-1}}{\frac{S_1}{(n-1)(k-1)}} \tag{6}$$

подчиняется распределению Фишера со степенями свободы $(n-1)$ и $(n-1)(k-1)$.

- Если верна вторая нулевая гипотеза, то модель наблюдений принимает вид

$$x_{ij} = \mu + b_i + \epsilon_{ij} \tag{7}$$

$$S_G = n \sum_{j=1}^{k} (\overline{x}_{.j} - \overline{x}_{..})^2$$

будет иметь $(k-1)$ степень свободы.

При справедливости гипотезы $H_0 : t_j = 0$, то статистика

$$F_2 = \frac{\frac{S_G}{k-1}}{\frac{S_1}{(n-1)(k-1)}} \tag{8}$$

подчиняется распределению Фишера со степенями свободы $(k-1)$ и $(n-1)(k-1)$.

- При справедливости альтернативной гипотезы в обоих случаях статистика имеет тенденцию принимать большие значения, так как числитель в среднем оказывается больше, чем при справедливости нулевой гипотезы. Применяя статистики F_1 и F_2 можно проверить каждую из нулевых гипотез.

Результаты работы двухфакторного дисперсионного анализа приведена ниже.

```
Средние по столбцам: [6.53  7.07  7.243 5.81 ]
Средние по строкам: [7.852 6.928 5.21 ]

Двухфакторный дисперсионный анализ
1ая гипотеза: H0: b_i = 0
2ая гипотеза: H0: t_i = 0

Вторая нулевая гипотеза H0 отвергается
на уровне значимости 0.05 при условии принятия первой нулевой гипотезы,
```

так как 22.611 >= 5.143

Первая нулевая гипотеза H0 отвергается
на уровне значимости 0.05 при условии принятия второй нулевой гипотезы,
так как 6.974 >= 4.757