Санкт-Петербургский Государственный Университет

Поляков Иван Михайлович

Отчёт по Лабораторной Работе № 8_1

Однофакторный 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

Необходимо провести однофакторный дисперсионный анализ приведённых ниже данных об условии хранения продуктов и содержании влаги. При необходимости применить метод линейных контрастов.

```
Исходные данные
1: [10.1, 7.3, 5.6, 6.2, 8.4, 8.1, 8.0, 7.6, 5.3, 7.2]
2: [11.7, 12.2, 11.8, 7.8, 8.9, 9.9, 12.4, 11.0, 10.3, 13.8, 10.5, 9.8, 9.1]
3: [10.2, 12.0, 8.8, 8.7, 10.5, 11.0, 9.1]
```

На уровне значимости $\alpha = 0.05$ проверить гипотезу о том, что условия хранения продукта не оказывают влияния на содержание влаги.

2 Ход работы

Цель дисперсионного анализа – исследование наличия или отсутствия существенного влияния какого-либо качественного или количественного фактора на изменения исследуемого результативного признака. Для этого фактор разделяют на классы градации и выясняют, одинаково ли влияние фактора путём исследования значимости между средними в наборах данных.

Однофакторный дисперсионный анализ основан на том, что сумму квадратов отклонений статистического комплекса возможно разделить на состовляющие компоненты:

$$SS = SS_a + SS_e \tag{1}$$

или

$$\sum_{r=1}^{k}\sum_{i=1}^{n_i}\left(X_{ir} - \overline{X}\right)^2 = \sum_{r=1}^{k} n_r \cdot (\overline{X_r} - \overline{X})^2 + \sum_{r=1}^{k}\sum_{i=1}^{n_i}\left(X_{ir} - \overline{X_r}\right) = Q_1 + Q_2 = Q, \tag{2}$$

где

$$\overline{X_r} = \frac{1}{n_i} \cdot \sum_{i=1}^{n_i} X_{ir} \tag{3}$$

– выборочное среднее;

$$\overline{X} = \frac{1}{n} \cdot \sum_{r=1}^{k}\sum_{i=1}^{n_i} X_{ir} \tag{4}$$

– общее выборочное среднее.

Результаты исходных данных представляют собой k независимых выборок с допущением, что они получены из k нормально распределенных генеральных совокупностей, где $k = 3$:

$$\begin{array}{cc} X_{11}, \ldots, X_{1n_1} & N(m_1; \sigma^2) \\ X_{21}, \ldots, X_{2n_2} & N(m_2; \sigma^2) \\ \ldots & \ldots \\ X_{k1}, \ldots, X_{kn_k} & N(m_k; \sigma^2) \end{array} \tag{5}$$

Сформулируем нулевую гипотезу $H_0 : m_1 = m_2 = \ldots = m_k$ и альтернативную H_1 : *отрицание нулевой гипотезы*. По формуле (2) можно найти сумму квадратов отклонений, объяснённых влиянием фактора k, сумму квадратов отклонений ошибки и общую сумму квадратов.

Статистика $\frac{Q_2}{\sigma^2}$ распределена по закону χ^2 с $n-k$ степенями свободы. При условии справедливости нулевой гипотезы статистика $\frac{Q_1}{\sigma^2}$ распределена по закону χ^2 с $k-1$ степенями свободы. Следовательно, статистика

$$\frac{\frac{Q_1}{k-1}}{\frac{Q_2}{n-k}} \tag{6}$$

распределена по закону Фишера с $k-1$, $n-k$ степенями свободы. При справедливости альтернативной гипотезы статистика имеет тенденцию принимать большие значения, так как числитель в среднем оказывается больше, чем при справедливости нулевой гипотезы.

Было найдено наблюдаемое и критическое значения F критерия. Нулевая гипотеза H_0 отвергается при уровне значимости $\alpha = 0.05$, так как $F > F_c rit$ (см. результаты в конце файла).

В случае отклонения нулевых гипотез, необходимо проверить следующие при помощи *метода линейных контрастов*:

$$H_0^{(1)} : m_1 = m_2; H_0^{(2)} : m_1 = m_3; H_0^{(3)} : m_2 = m_3; H_0^{(4)} : \frac{1}{2}(m_1 + m_3) = m_2 \tag{7}$$

После проведения однофакторного дисперсионного анализа и по его итогу принятия альтернативной гипотезы важно узнать, какие именно математические ожидания значимо отличались, а какие равны. Линейный контраст Lk определяется по формуле

$$Lk = \sum_{r=1}^{k} c_r m_r, \tag{8}$$

где c_r - задаваемые константы, причём $\sum\limits_{r=1}^{k} c_r = 0$

Оценка линейного контраста имеет следующий вид:

$$\hat{Lk} = \sum_{r=1}^{k} c_r \overline{X_r} \tag{9}$$

Оценка дисперсии $\hat{Lk}$ вычисляется по формуле

$$S_{Lk}^2 = \sum_{r=1}^{k} \frac{c_r^2}{n_r} \cdot \hat{\sigma}^2 = \frac{Q_2}{n-k} \cdot \sum_{r=1}^{k} \frac{c_r^2}{n_r} \tag{10}$$

Для сформулированных нулевых гипотез (7) о равенстве математических ожиданий применим соответствующие линейные контрасты:

$$Lk_1 = m_1 - m_2; Lk_2 = m_1 - m_3; Lk_3 = m_2 - m_3; Lk_4 = \frac{1}{2}(m_1 + m_3) - m_2 \tag{11}$$

с коэффициентами $c_1 = 1, c_2 = -1, c_3 = 0$

Доверительный интервал для оценок контраста вычисляется по формуле

$$\hat{Lk} \mp S_{lk} \cdot \sqrt{(k-1) \cdot F_{1-\alpha}(k-1, n-k)} \tag{12}$$

Проведя вычисления, все 4 нулевые гипотезы отвергаются на уровне значимости 5%, так как 0 не принадлежит ни одному из полученных доверительных интервалов.

Результаты работы приведены ниже

```
Q1 = 65.871
Q2 = 60.442
Q = Q1 + Q2 = 126.314

Критерий Фишера
Гипотеза H0 о равенстве средних отклоняется,
так как F_наблюдаемое > F_критическое (14.71 > 3.354)
```

```
Гипотеза H0 о равенстве двух средних отвергается,
так как ноль не содержится в доверительном интервале (3.286; 3.369)

Гипотеза H0 о равенстве двух средних отвергается,
так как ноль не содержится в доверительном интервале (2.62; 2.705)

Гипотеза H0 о равенстве двух средних отвергается,
так как ноль не содержится в доверительном интервале (-0.7104; -0.6192)

Гипотеза H0 о равенстве средних отвергается,
так как ноль не содержится в доверительном интервале (2.426; 2.899)
```