ЛЕКЦИЯ 6

ТЕМА: Дисперсионный анализ

Вопросы темы. Однофакторный дисперсионный анализ для несвязанных выборок. Дисперсионный анализ для связанных выборок.

Дисперсионный анализ, предложенный **Р. Фишером**, является статистическим методом, предназначенным для выявления влияния ряда отдельных факторов на результаты экспериментов.

В основе дисперсионного анализа лежит предположение о том, что одни переменные могут рассматриваться как причины (**факторы, независимые переменные**), а другие как следствия (**зависимые переменные**). Независимые переменные называют иногда регулируемыми факторами именно потому, что в эксперименте исследователь имеет возможность варьировать ими и анализировать получающийся результат.

Сущность дисперсионного анализа заключается в расчленении общей дисперсии изучаемого признака на отдельные компоненты, обусловленные влиянием конкретных факторов, и проверке гипотез о значимости влияния этих факторов на исследуемый признак. Сравнивая компоненты дисперсии друг с другом посредством **F — критерия Фишера**, можно определить, какая доля общей вариативности результативного признака обусловлена действием регулируемых факторов.

Исходным материалом для дисперсионного анализа служат данные исследования трех и более выборок, которые могут быть как **равными**, так и **неравными** по численности, как **связными**, так и **несвязными**. По количеству выявляемых регулируемых факторов дисперсионный анализ может быть **однофакторным** (при этом изучается влияние одного фактора на результаты эксперимента), **двухфакторным** (при изучении влияния двух факторов) и **многофакторным** (позволяет оценить не только влияние каждого из факторов в отдельности, но и их взаимодействие)**.**

Дисперсионный анализ относится к группе параметрических методов и поэтому его следует применять только тогда, когда доказано, что распределение является нормальным..

Однофакторный дисперсионный анализ для несвязанных выборок

Изучается действие только одной переменной (фактора) на исследуемый признак. Исследователя интересует вопрос, как изменяется определенный

признак в разных условиях действия переменной (фактора). Например, как изменяется время решения задачи при разных условиях мотивации испытуемых (низкой, средней, высокой мотивации) или при разных способах предъявления задачи (устно, письменно или в виде текста с графиками и иллюстрациями), в разных условиях работы с задачей (в одиночестве, в комнате с преподавателем, в классе). В первом случае фактором является мотивация, во втором – степень наглядности, в третьем – фактор публичности.

В данном варианте метода влиянию каждой из градаций подвергаются разные выборки испытуемых. Градаций фактора должно быть не менее **трех**.

Пример 1. Три различные группы из шести испытуемых получили списки из десяти слов. Первой группе слова предъявлялись с низкой скоростью -1 слово в 5 секунд, второй группе со средней скоростью - 1 слово в 2 секунды, и третьей группе с большой скоростью - 1 слово в секунду. Было предсказано, что показатели воспроизведения будут зависеть от скорости предъявления слов. Результаты представлены в табл. 1.

Таблица 1. Количество воспроизведенных слов (по J. Greene, M D'Olivera, 1989, p. 99)

№ испытуемого	Группа 1 низкая скорость	Группа 2 средняя скорость	Группа 3 высокая скорость
1	8	7	4
2	7	8	5
3	9	5	3
4	5	4	6
5	6	6	2
6	8	7	4
суммы	43	37	24
средние	7,17	6,17	4,00
Общая сумма	104		

Дисперсионный однофакторный анализ позволяет проверить гипотезы:

H_0: различия в объеме воспроизведения слов *между* группами являются не более выраженными, чем случайные различия *внутри* каждой группы

H_1: Различия в объеме воспроизведения слов *между* группами являются более выраженными, чем случайные различия *внутри* каждой группы.

Последовательность операций в однофакторном дисперсионном анализе для несвязанных выборок:

1. подсчитаем **$SS_{факт}$** - вариативность признака, обусловленную действием исследуемого фактора. [обозначение SS - сокращение от "суммы квадратов" (sum of squares)].

$$SS_{факт} = \frac{\sum T_c^2}{n} - \frac{(\sum x_i)^2}{N}, \qquad (1)$$

где T_c – сумма индивидуальных значений по каждому из условий. Для нашего примера 43, 37, 24;

c – количество условий (градаций) фактора (=3);

n – количество испытуемых в каждой группе (=6);

N – общее количество индивидуальных значений (=18);

$(\sum x_i)^2$ - квадрат общей суммы индивидуальных значений ($=104^2=10816$)

Отметим разницу между $\sum (x_i^2)$, в которой все индивидуальные значения сначала возводятся в квадрат, а потом суммируются, и $(\sum x_i)^2$, где индивидуальные значения сначала суммируются для получения общей суммы, а потом уже эта сумма возводится в квадрат.

По формуле (1) рассчитав фактическую вариативность признака, получаем:

$$SS_{факт} = \frac{(43^2 + 37^2 + 24^2)}{6} - \frac{104^2}{18} = 31,44$$

2. подсчитаем **$SS_{общ}$** – общую вариативность признака:

$$SS_{общ} = \frac{\sum x_i^2 - (\sum x_i)^2}{N} = \\ = \frac{8^2 + 7^2 + 9^2 + 5^2 + 6^2 + 8^2 + 7^2 + 8^2 ... + 2^2 + 4^2 - 104^2}{18} = 63.11 \qquad (2)$$

3. подсчитаем случайную (остаточную) величину **$SS_{сл}$**, обусловленную неучтенными факторами:

$$SS_{сл} = SS_{общ} - SS_{факт} = 63,11 - 31,44 = 31,67 \qquad (3)$$

4. **число степеней свободы** равно:

$k_{факт} = k_1 = c - 1$=3-1=2

$k_{общ} = N - 1 = 18 - 1 = 17 \quad k_{сл} = k_2 = k_{общ} - k_{факт} = 17 - 2 = 15$

5. **«средний квадрат»** или **математическое ожидание суммы квадратов,** усредненная величина соответствующих сумм квадратов SS равна:

$$MS_{факт} = \frac{SS_{факт}}{k_{факт}} = \frac{31,44}{2} = 15,72 \qquad (5)$$

$$MS_{сл} = \frac{SS_{сл}}{k_{сл}} = \frac{31,67}{15} = 2,11$$

6. значение **статистики критерия $F_{эмп}$** рассчитаем по формуле:

$$F_{эмп} = \frac{MS_{факт}}{MS_{случ}} \qquad (6)$$

Для нашего примера имеем*: $F_{эмп}=15,72/2,11=7,45$*

7. определим $\mathbf{F_{крит}}$ по статистическим таблицам **Приложения 3** для $df_1=k_1=2$ и $df_2=k_2=15$ табличное значение статистики равно 3,68

8. если $\mathbf{F_{эмп}} < F_{крит}$, то нулевая гипотеза принимается, в противном случае принимается альтернативная гипотеза. Для нашего примера $\mathbf{F_{эмп}} > F_{крит}$ (7.45>3.68), следовательно принимается альтернативная гипотеза.

Вывод: различия в объеме воспроизведения слов между группами являются более выраженными, чем случайные различия внутри каждой группы ($p<0,05$). Т.о. скорость предъявления слов влияет на объем их воспроизведения.

Дисперсионный анализ для связанных выборок

Метод дисперсионного анализа для связанных выборок применяется в тех случаях, когда исследуется влияние разных градаций фактора или разных условий на **одну и ту же выборку испытуемых.** Градаций фактора должно быть не менее **трех.**

В данном случае различия между испытуемыми - возможный самостоятельный источник различий. Однофакторный дисперсионный анализ для связанных выборок позволит определить, что перевешивает - тенденция, выраженная кривой изменения фактора, или индивидуальные различия между

испытуемыми. Фактор индивидуальных различий может оказаться более значимым, чем фактор изменения экспериментальных условий.

Пример 2. Группа из 5 испытуемых была обследована с помощью трех экспериментальных заданий, направленных на изучение интеллектуальной, настойчивости (Сидоренко Е. В., 1984). Каждому испытуемому индивидуально предъявлялись последовательно три одинаковые анаграммы: четырехбуквенная, пятибуквенная и шестибуквенная. Можно ли считать, что фактор длины анаграммы влияет на длительность попыток ее решения?

Таблица 2. Длительность решения анаграмм (сек)

Код испытуемого	Условие 1. четырехбуквенная анаграмма	Условие 2. Пятибуквенная анаграмма	Условие 3. шестибуквенная анаграмма	Суммы по испытуемым
1	5	235	7	247
2	7	604	20	631
3	2	93	5	100
4	2	171	8	181
5	35	141	7	183
суммы	51	1244	47	1342

Сформулируем гипотезы. Наборов гипотез в данном случае два.

Набор А.

H_0(А): Различия в длительности попыток решения анаграмм разной длины являются не более выраженными, чем различия, обусловленные случайными причинами.

H_1(А): Различия в длительности попыток решения анаграмм разной длины являются более выраженными, чем различия, обусловленные случайными причинами.

Набор Б.

H_o(Б): Индивидуальные различия между испытуемыми являются не более выраженными, чем различия, обусловленные случайными причинами.

H_1(Б): Индивидуальные различия между испытуемыми являются более выраженными, чем различия, обусловленные случайными причинами.

Последовательность операций в однофакторном дисперсионном анализе для связанных выборок:

1. подсчитаем $\mathbf{SS_{факт}}$ - вариативность признака, обусловленную действием исследуемого фактора по формуле (1).

$$SS_{факт} = \frac{\sum T_c^2}{n} - \frac{(\sum x_i)^2}{N} = \frac{(51^2 + 1244^2 + 47^2)}{5} - \frac{1342^2}{15} = 190405,$$

где T_c – сумма индивидуальных значений по каждому из условий (столбцов). Для нашего примера 51, 1244, 47 (см. табл. 2); c – количество условий (градаций) фактора (=3); n – количество испытуемых в каждой группе (=5); N – общее количество индивидуальных значений (=15); $(\sum x_i)^2$ - квадрат общей суммы индивидуальных значений ($=1342^2$)

2. подсчитаем $\mathbf{SS_{исп}}$ - вариативность признака, обусловленную индивидуальными значения испытуемых.

$$SS_{факт} = \frac{\sum T_u^2}{c} - \frac{(\sum x_i)^2}{N} = \frac{(247^2 + 631^2 + 100^2 + 181^2 + 183^2)}{3} - \frac{1342^2}{15} = 58409$$

где $T_и$ – сумма индивидуальных значений по каждому испытуемому. Для нашего примера 247, 631, 100, 181, 183 (см. табл. 2); c – количество условий (градаций) фактора (=3); N – общее количество индивидуальных значений (=15);

3. подсчитаем $\mathbf{SS_{общ}}$ – общую вариативность признака по формуле (2):

$$SS_{общ} = \frac{\sum x_i^2 - (\sum x_i)^2}{N} =$$

$$= \frac{5^2 + 7^2 + 2^2 + 2^2 + 35^2 + 235^2 + 604^2 + 93^2 ... + 8^2 + 7^2 - 1342^2}{15} = 359642$$

4. подсчитаем случайную (остаточную) величину $\mathbf{SS_{сл}}$, обусловленную неучтенными факторами по формуле (3):

$$SS_{сл} = SS_{общ} - SS_{факт} - SS_{исп} = 359642 - 190405 - 58409 = 110828$$

5. **число степеней свободы** равно (4):

$$k_{факт} = k_1 = c - 1 = 3 - 1 = 2; \quad k_{исп} = k_3 = n - 1 = 5 - 1 = 4; \quad k_{общ} = N - 1 = 15 - 1 = 14;$$

$$k_{сл} = k_2 = k_{общ} - k_{факт} - k_{исп} = 14 - 2 - 4 = 8$$

6. **«средний квадрат»** или **математическое ожидание суммы квадратов,** усредненная величина соответствующих сумм квадратов SS равна (5):

$$MS_{факт} = \frac{SS_{факт}}{k_{факт}} = \frac{190405}{2} = 95202,2; \; MS_{исп} = \frac{SS_{исп}}{k_{исп}} = \frac{58409}{4} = 14602,2$$

$$MS_{сл} = \frac{SS_{сл}}{k_{сл}} = \frac{110827}{8} = 13853,4$$

7. значение статистики критерия $\mathbf{F_{эмп}}$ рассчитаем по формуле (6):

$$F_{эмп\_факт} = \frac{MS_{факт}}{MS_{случ}} = \frac{95202,5}{13853,4} = 6,872 \; ; \; F_{эмп\_исп} = \frac{MS_{исп}}{MS_{случ}} = \frac{14602,5}{13853,4} = 1,054$$

8. определим $F_{крит}$ по статистическим таблицам Приложения 3 для $df_1=k_1=2$ и $df_2=k_2=8$ табличное значение статистики $F_{крит\_факт}=4,46$, и для $df_3=k_3=4$ и $df_2=k_2=8$ $F_{крит\_исп}=3,84$

9. $\mathbf{F_{эмп\_факт}} > F_{крит\_факт}$ (6,872>4,46), следовательно принимается альтернативная гипотеза.

10. $\mathbf{F_{эмп\_исп}} < F_{крит\_исп}$ (1,054<3,84), следовательно принимается нулевая гипотеза.

Вывод: различия в объеме воспроизведения слов в разных условиях являются более выраженными, чем различия, обусловленные случайными причинами ($p<0,05$). Индивидуальные различия между испытуемыми являются не более выраженными, чем различия, обусловленные случайными причинами.