СЕВЕРНЫЙ (АРКТИЧЕСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В. ЛОМОНОСОВА

Дисперсионный анализ

Лебедева Антонина, Евстратова Екатерина, Дуркина Галина

Назначение метода

- Дисперсионный анализ – анализ изменчивости признака под влиянием каких-либо контролируемых переменных факторов.
- Методами дисперсионного анализа устанавливается наличие влияния заданного фактора на изучаемый процесс (на выходную переменную процесса) за счёт статистической обработки наблюдаемой совокупности выборочных данных.

Назначение метода

- Дисперсионный анализ есть совокупность статистических методов, предназначенных для проверки гипотез о связи между определенными признаками и исследуемыми факторами, которые не имеют количественного описания, а также для установления степени влияния факторов и их взаимодействия. В специальной литературе его часто называют ANOVA (от англоязычного названия Analysis of Variations). Впервые этот метод был разработан Р. Фишером в 1925 г.

Типы решаемых задач

- Основной целью дисперсионного анализа является исследование значимости различия между средними. Установить различаются ли три группы или более по какому-либо одному количественному признаку
- Обобщенно задача дисперсионного анализа состоит в том, чтобы из общей вариативности признака выделить три частные вариативности:

- Вариативность, обусловленную действием каждой из исследуемых независимых переменных.

- Вариативность, обусловленную взаимодействием исследуемых независмых переменных.

- Вариативность случайную, обусловленную всеми неучтенными обстоятельствами.

Цель и задачи

- **Цель:** выявить дифференциацию представлений о нормативном сексуальном поведении среди представителей различных поколений в современной России и определяющие её факторы.
- **Задачи:**

1. Проанализировать теоретико-методологические подходы к проблеме сексуальной нормы в научной литературе.
2. Обобщить результаты эмпирических исследований о дифференциации представлений о сексуальной норме среди населения современной России.
3. Определить факторы, обусловливающие различия в представлениях о сексуальной норме среди населения России в целом
4. Выявить специфику межпоколенных различий в представлениях о сексуальной норме в Архангельской области по сравнению с общероссийской ситуацией

Возможности применения методов в социологических исследования

- Применительно к социологическим данным таковыми являются частости различных вариантов ответов респондентов на вопросы анкет массовых опросов, измеряемые в процентах от общего числа ответов и интерпретируемые как вероятности отношения изучаемых групп населения к исследуемым социальным явлениям или процессам.

Этапы

- Построение дисперсионного комплекса.
- Вычисление средних квадратов отклонений.
- Вычисление дисперсии
- Сравнение факторной и остаточной дисперсий
- Оценка результатов с помощью теоретических значений распределения Фишера-Снедекор

Методы (методики, разновидности) данного вида анализа

- 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).

Суть метода, формулы

- Однофакторный дисперсионный анализ основан на том, что сумму квадратов отклонений статистического комплекса возможно разделить на компоненты:
- SS = SSa + SSe,

Где:

SS - общая сумма квадратов отклонений,

SSa - объяснённая влиянием фактора a сумма квадратов отклонений,

SSe - необъяснённая сумма квадратов отклонений или сумма квадратов отклонений ошибки.

- Если через ni обозначить число вариантов в каждом классе градации (группе) и a - общее число градаций фактора (групп), то $\sum_{i=1}^{n} n_i = n$ - общее число наблюдений и можно получить следующие формулы:
- общее число квадратов отклонений: $SS = \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{\overline{X}}\right)^2$
- объяснённая влиянием фактора a сумма квадратов отклонений: $SS_a = \sum_{i=1}^{a} n_i\left(\overline{X}_i - \overline{\overline{X}}\right)^2$
- необъяснённая сумма квадратов отклонений или сумма квадратов отклонений ошибки: $SS_e = \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}_i\right)^2$

Где

$\overline{\overline{X}} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{n_i} X_{ij}$ - общее среднее наблюдений,

$\overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_j$ - среднее наблюдений в каждой градации фактора (группе).

- Кроме того,

$$SS_e = \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X_i}\right)^2 = \sum_{i=1}^{a}(n_i - 1)s_i^2 =$$
$$= (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_a - 1)s_a^2,$$

где - дисперсия градации фактора (группы).

- Чтобы провести однофакторный дисперсионный анализ данных статистического комплекса, нужно найти фактическое отношение Фишера - отношение дисперсии, объяснённой влиянием фактора (межрупповой), и необъяснённой дисперсии (внутригрупповой): $F = \frac{MS_a}{MS_e}$
- и сравнить его с критическим значением Фишера: $F_{\alpha;\nu_a;\nu_e}$

- Дисперсии рассчитываются следующим образом:

$MS_a = \frac{SS_a}{a-1}$ - объяснённая дисперсия,

$MS_e = \frac{SS_e}{n-a}$ - необъяснённая дисперсия,

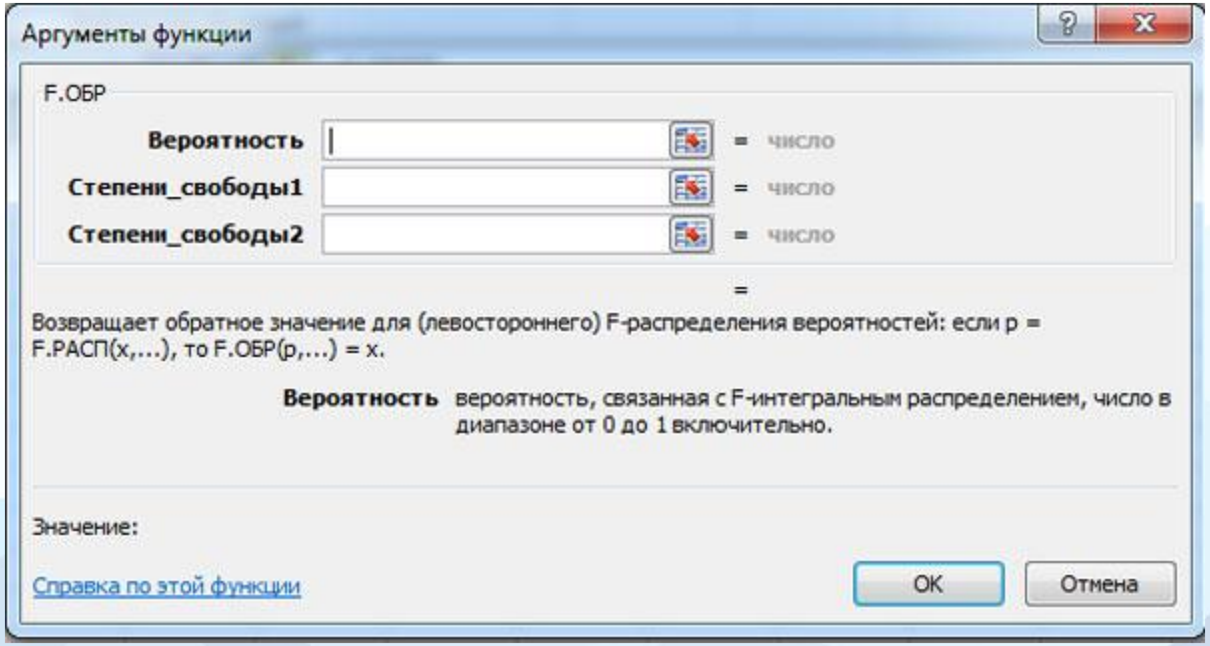
при этом

$v_a = a - 1$ - число степеней свободы объяснённой дисперсии,

$v_e = n - a$ - число степеней свободы необъяснённой дисперсии,

$v = n - 1$ - общее число степеней свободы.

- Критическое значение отношения Фишера с определёнными значениями уровня значимости и степеней свободы можно найти в статистических таблицах или рассчитать с помощью функции MS Excel F.ОБР



Двухфакторный анализ

- Двухфакторный дисперсионный анализ применяется для того, чтобы проверить возможную зависимость результативного признака от двух факторов - А и В. Тогда a - число градаций фактора А и b - число градаций фактора В. В статистическом комплексе сумма квадратов остатков разделяется на три компоненты:
- SS = SSa + SSb + SSe,

$SS = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(X_{ij} - \overline{\overline{X}}\right)^2$ - общая сумма квадратов отклонений,

$SS_a = b\sum_{i=1}^{a}\left(\overline{X}_i - \overline{\overline{X}}\right)^2$ - объяснённая влиянием фактора А сумма квадратов отклонений,

$SS_b = a\sum_{j=1}^{b}\left(\overline{X}_j - \overline{\overline{X}}\right)^2$ -объяснённая влиянием фактора В сумма квадратов отклонений,

$$SS_e = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(X_{ij} - \overline{X}_i - \overline{X}_j + \overline{\overline{X}}\right)^2$$ - необъяснённая сумма квадратов отклонений или сумма квадратов отклонений ошибки,

$$\overline{\overline{X}} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} X_{ij}$$ - общее среднее наблюдений,

$$\overline{X}_i = \frac{1}{b}\sum_{j=1}^{b} X_{ij}$$ - среднее наблюдений в каждой градации фактора А,

$$\overline{X}_j = \frac{1}{a}\sum_{i=1}^{a} X_{ij}$$ - среднее число наблюдений в каждой градации фактора В.

- Дисперсии вычисляются следующим образом:

$$MS_a = \frac{SS_a}{a-1}$$ - дисперсия, объяснённая влиянием фактора А,

$$MS_b = \frac{SS_b}{b-1}$$ - дисперсия, объяснённая влиянием фактора В,

$$MS_e = \frac{SS_e}{(a-1)(b-1)}$$ - необъяснённая дисперсия или дисперсия ошибки,

Где:
- $v_a = a - 1$ - число степеней свободы дисперсии, объяснённой влиянием фактора A,
- $v_b = b - 1$ - число степеней свободы дисперсии, объяснённой влиянием фактора B,
- $v_e = (a - 1)(b - 1)$ - число степеней свободы необъяснённой дисперсии или дисперсии ошибки,
- $v = ab - 1$ - общее число степеней свободы.
- Если факторы не зависят друг от друга, то для определения существенности факторов выдвигаются две нулевые гипотезы и соответствующие альтернативные гипотезы:

для фактора A:

H_0: $\mu_{1A} = \mu_{2A} = ... = \mu_{aA}$,

H_1: не все μ_{iA} равны;

для фактора B:

H_0: $\mu_{1B} = \mu_{2B} = ... = \mu_{aB}$,

H_1: не все μ_{iB} равны.

- Если фактическое отношение Фишера больше критического отношения Фишера, то следует отклонить нулевую гипотезу с уровнем значимости α. Это означает, что фактор существенно влияет на данные: данные зависят от фактора с вероятностью $P = 1 - \alpha$.
- Если фактическое отношение Фишера меньше критического отношения Фишера, то следует принять нулевую гипотезу с уровнем значимости α. Это означает, что фактор не оказывает существенного влияния на данные с вероятностью $P = 1 - \alpha$.

СЕВЕРНЫЙ (АРКТИЧЕСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В. ЛОМОНОСОВА

Спасибо за внимание!