



# ДИСПЕРСИОННЫЙ АНАЛИЗ

Авторы: Л. Н. Большев

ДИСПЕРСИОННЫЙ АНАЛИЗ в математике, статистич. метод, предназначенный для выявления влияния отдельных факторов на результат эксперимента, а также для последующего планирования экспериментов.

Если значения неизвестных постоянных

$a_1, \dots, a_I$  могут быть измерены с помощью разл. методов или измерительных средств

$M_1, \dots, M_J$  и в каждом случае т. н. систематич. ошибка

$b_{ij}$  может, вообще говоря, зависеть как от выбранного метода

$M_j$ , так и от неизвестного измеряемого значения

$a_i$ , то результаты таких измерений представляют собой суммы вида

$$x_{ijk} = a_i + b_{ij} + y_{ijk}, i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, K,$$

где

$K$  – число независимых измерений неизвестной величины

$a_i$  методом

$M_j$ , а

$y_{ijk}$  – случайная ошибка

$k$ -го измерения величины

$a_i$  методом

$M_j$ . При этом предполагается, что все

$y_{ijk}$  – независимые одинаково распределённые случайные величины, имеющие нулевое математич. ожидание.

Такая линейная модель называется двухфакторной схемой Д. а.; первый фактор – истинное значение

измеряемой величины, второй – метод измерения, причём в данном случае для каждой возможной комбинации значений первого и второго факторов осуществляется одинаковое число

$K$  независимых измерений (в более сложных моделях Д. а. от последнего предположения иногда отказываются).

Напр., пусть в соревнованиях участвуют

$I$  спортсменов, мастерство которых оценивается

$J$  судьями, причём каждый участник соревнований выступает

$K$  раз (имеет

$K$  попыток). В этом случае

$a_i$  – истинное значение показателя мастерства спортсмена с номером

$i$ ,

$b_{ij}$  – систематич. ошибка, вносимая в оценку мастерства  $i$ -го спортсмена судьёй с номером

$j$ ,

$x_{ijk}$  – оценка, выставленная

$j$ -м судьёй

$i$ -му спортсмену за

$k$ -ю попытку, а

$y_{ijk}$  – случайная погрешность. Подобная схема типична для т. н. субъективной экспертизы качества нескольких объектов, осуществляемой группой экспертов. Др. пример – статистич. исследование урожайности с.-х. культуры в зависимости от одного из

$I$  сортов почвы и

$J$  методов её обработки, причём для каждого сорта почвы

$i$  и каждого метода обработки с номером

$j$  осуществляется

$K$  независимых экспериментов. В этом примере можно считать, что величины

$a_i$  равны нулю,

$b_{ij}$  – истинное значение урожайности для

$i$ -го сорта почвы при

$j$ -м методе обработки,

$x_{ijk}$  – соответствующая экспериментально наблюдаемая урожайность в

$k$ -м опыте, а

$y_{ijk}$  – её случайная ошибка, возникающая из-за тех или иных случайных причин.

Пусть

$$c_{ij} = a_i + b_{ij}$$

и

$$c_{i \cdot} = \frac{1}{J} \sum_j c_{ij}, c_{\cdot j} = \frac{1}{I} \sum_i c_{ij}, c_{\cdot \cdot} = \frac{1}{IJ} \sum_{ij} c_{ij} = \frac{1}{I} \sum_i c_{i \cdot} = \frac{1}{J} \sum_j c_{\cdot j}.$$

Пусть, кроме того,

$\alpha = c_{\cdot \cdot}, \beta_i = c_{i \cdot} - c_{\cdot \cdot}, \gamma_j = c_{\cdot j} - c_{\cdot \cdot}$  и

$\delta_{ij} = c_{ij} - c_{i \cdot} - c_{\cdot j} + c_{\cdot \cdot}$ . Идея Д. а. основана на тождествах

$$c_{ij} = \alpha + \beta_i + \gamma_j + \delta_{ij}, i = 1, \dots, I, j = 1, \dots, J.$$

В примере, связанном со спортивными соревнованиями, функция

$\delta_{ij}$  выражает отношение

$j$ -го судьи к

$i$ -му спортсмену (положительное значение

$\delta_{ij}$  означает подсуживание, т. е. систематич. завышение

$j$ -м судьёй оценки мастерства

$i$ -го спортсмена, а отрицательное значение

$\delta_{ij}$  – засуживание, т. е. систематич. снижение оценки). Равенство всех

$\delta_{ij}$  нулю – обычное требование, которое предъявляют к работе группы экспертов. В случае агрономич. опытов

такое равенство рассматривается как гипотеза, подлежащая проверке по результатам экспериментов. Если эта

гипотеза верна, то выявление наилучших почвы и обработки может быть осуществлено отдельно, что приводит

к существенному сокращению числа экспериментов.

В ситуации спортивных соревнований функция

$\gamma_j$  может трактоваться как систематич. ошибка, допускаемая

$j$ -м судьёй по отношению ко всем спортсменам, т. е.

$\gamma_j$  – характеристика строгости или либеральности

$j$ -го судьи. В реальных условиях

$\gamma_j$  могут иметь ненулевые значения, что приходится учитывать при подведении итогов экспертизы. Сумма двух оставшихся функций

$\alpha + \beta_i$  зависит лишь от

$i$  и поэтому может быть использована для оценки мастерства

$i$ -го спортсмена. Однако следует учитывать, что величина

$\alpha + \beta_i \neq a_i$  оценивает не только мастерство

$i$ -го спортсмена, но в той или иной мере отношение экспертов к его мастерству.

Истинные значения функций

$\alpha, \beta_i, \gamma_j$  и

$\delta_{ij}$  неизвестны и выражаются в терминах неизвестных функций

$c_{ij}$ . Поэтому первый этап Д. а. заключается в отыскании статистич. оценок для

$c_{ij}$  по результатам наблюдений

$x_{ijk}$ . Несмещённая и имеющая минимальную дисперсию оценка для

$c_{ij}$  выражается формулой

$$\hat{c}_{ij} = x_{ij} \cdot = \frac{1}{K} \sum_k x_{ijk}.$$

Несмещённые оценки

$\hat{\alpha}, \hat{\beta}_i, \hat{\gamma}_j, \hat{\delta}_{ij}$  для функций

$\alpha, \beta_i, \gamma_j, \delta_{ij}$ , имеющие минимальные дисперсии, получаются в результате замены аргументов

$c_{ij}$  соответствующими оценками

$\hat{c}_{ij}$ .

На основе этих оценок строится второй этап Д. а., посвящённый выявлению влияния первого и второго факторов на результаты эксперимента (в агрономич. опытах первый фактор – сорт почвы, второй – метод обработки).

Статистич. методами проверяется гипотеза отсутствия взаимодействия факторов, которая соответствует равенству

$\sum_{ij} \delta_{ij}^2 = 0$  а также гипотезы

$\sum_j \gamma_j^2 = 0$  и

$\sum_i \beta_i^2 = 0$

Дальнейшие этапы Д. а. существенно зависят не только от реального содержания конкретной задачи, но также и от результатов статистич. проверки гипотез на втором этапе. Напр., в условиях агрономич. опытов

справедливость гипотезы

$\sum_{ij} \delta_{ij}^2 = 0$  позволяет более экономично спланировать дальнейшие эксперименты (если помимо гипотезы

$\sum_{ij} \delta_{ij}^2 = 0$  справедлива также и гипотеза

$\sum_j \gamma_j^2 = 0$ , то это означает, что урожайность зависит лишь от сорта почвы). Справедливость гипотезы

$\sum_{ij} \delta_{ij}^2 = 0$  даёт основание для упорядочивания сравниваемых объектов (напр., спортсменов) по значениям величин

$$\hat{\alpha} + \hat{\beta}_i, i = 1, \dots, I.$$

Первоначально Д. а. был предложен Р. [Фишером](#) (1925) для обработки результатов агрономич. опытов по выявлению условий, при которых испытываемый сорт с.-х. культуры даёт максимальный урожай. Совр.

приложения Д. а. охватывают широкий круг задач экономики, социологии, биологии и техники и трактуются обычно в терминах статистич. теории выявления систематич. различий между результатами измерений, выполненных при тех или иных меняющихся условиях. См. также [Математическая статистика](#).

## Литература

Лит.: Хальд А. Математическая статистика с техническими приложениями. М., 1956; Снедекор Дж. У.

Статистические методы в применении к исследованиям в сельском хозяйстве и биологии. М., 1961; Шеффе Г.

Дисперсионный анализ. 2-е изд. М., 1980.

Processing math: 100%