

Факторный анализ

- Задачи однофакторного анализа
- Терминология
- Данные
- Статистические предположения, влияющие на постановку задачи
 - Характер изменчивости
 - Влияние измерительной шкалы
- Стратегия анализа
 - Выяснить наличие эффекта обработки
 - Оценка величины эффектов обработки и качества полученных оценок

• Задачи однофакторного анализа

- При исследовании зависимостей между случайными величинами одной из наиболее простых является ситуация, когда можно указать только один фактор (величину), влияющую на конечный результат, и этот фактор может принимать лишь конечное число значений (уровней).
- Такие задачи и являются задачами однофакторного анализа.
-

Пример.

Сравнение результатов нескольких различных способов действия, направленных на достижение одной цели:

Несколько школьных учебников, нескольких лекарств, нескольких способов обработки и т.д.

Терминология факторного анализа

- То, что, как мы считаем, должно оказывать влияние на конечный результат, называют **фактором** (или **факторами**, если их несколько).
- **Конкретную реализацию** фактора (определенный учебник, лекарство) называют **уровнем фактора**, или способом обработки.
- Значение измеряемого параметра (т.е. величину результата) называют **ОТКЛИКОМ**.

Экспериментальные данные.

Для сравнения влияния факторов на результат необходим статистический материал.

Его получают следующим образом: каждый из k способов обработки применяют несколько раз (не обязательно одно и то же) к исследуемому объекту и регистрируют результаты. В итоге имеем k выборок не обязательно одинакового объема.

Результаты представляются в таблице.

Таблица Результатов

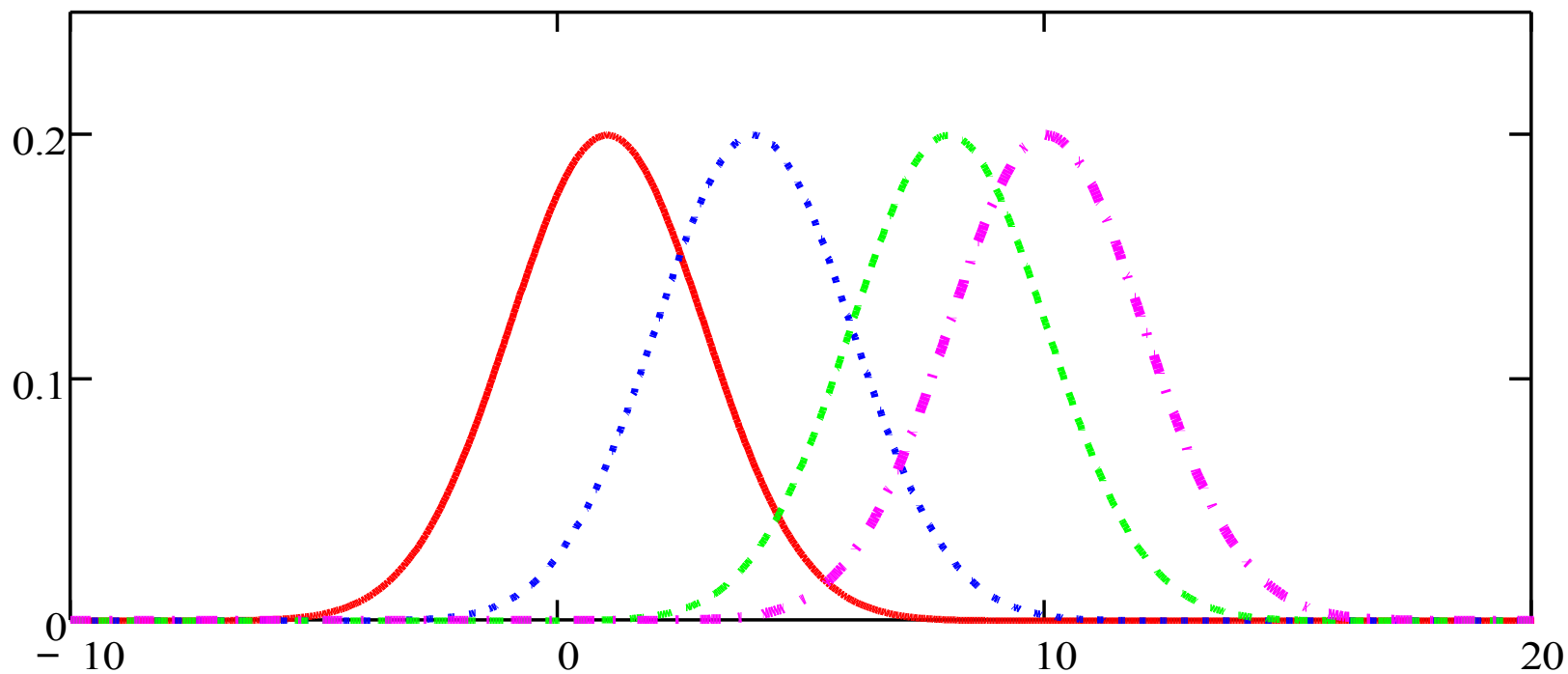
Уровни факторов	1	2	$\dots$	k
	$X_{1,1}$	$X_{1,2}$	$\dots$	$X_{1,k}$
Результаты	$X_{2,1}$	$X_{2,2}$	$\dots$	$X_{2,k}$
измерений	$\dots$	$\dots$	$\dots$	$\dots$
	$X_{n1,1}$	$X_{n2,2}$	$\dots$	$X_{nk,k}$

Статистические предположения, влияющие на постановку задачи (как фактор влияет на закон распределения)

- Опыт показывает, что изменчивости больше подвержено положение случайной величины — **медиана или среднее**.
- Т.е. предполагается часто, что распределение выборок одного вида (сдвиговое семейство выборок).
- Для описания данных, представленных в таблице в большинстве случаев приемлемой оказывается аддитивная линейная модель:

аддитивная линейная модель

$$x_{ij} = a_j + e_{ij}, j = 1 \dots k, i = 1 \dots n.$$



e_{ij} – неизвестные, независимые, одинаково непрерывно распределенные случайные величины, отражающие изменчивость, внутренне присущую наблюдениям.

Если распределение e_{ij} нормальное, то в дальнейшем используют методы *дисперсионного анализа*, в противном случае применяются непараметрические методы – например, *ранговые методы анализа*.

•Влияние измерительной шкалы

Различают следующие виды шкал:

количественные

порядковые (ординальные)

номинальные

Стратегия анализа

- Выяснение наличия влияния фактора.
- Количественная оценка эффекта обработки.

В начале следует ответить на вопрос: **существует ли влияние фактора на измеряемый признак?**

Для этого проводится **сравнение** изменчивости признака **внутри** каждой выборки и **изменчивость между выборками**.

На статистическом языке это формулируется в виде гипотезы

Но: все данные принадлежат одному и тому же распределению.

Если гипотеза принимается, то анализ на этом и заканчивается (нет влияния факторов), если ***Но*** отвергается, то возникает следующая задача

количественной оценки величины эффектов

обработки и выяснение качества полученных оценок.

Дисперсионный факторный анализ

$$\overline{X}_j = \frac{1}{n_j} \cdot \sum_{i=1}^{n_j} X_i$$

$$\overline{X}_0 = \frac{1}{N} \cdot \sum_{j=1}^{n_j} \overline{X}_j$$

$$Q_0 = \frac{\sum_{j=1}^k (\overline{X}_j - m_x)^2}{\sigma^2}$$