

Обработка экспериментальных данных

Статистический инструментарий

МГСУ - 2020

Эксперимент

Постановка экспериментов – неотъемлемая часть исследований.

Результат эксперимента:

- качественное суждение;
- ранговый показатель;
- **численный (действительнозначный) показатель.**

Для последнего случая вопросы рациональной организации, формального анализа и предметной интерпретации результатов эксперимента являются наиболее проработанными. Дисциплина, задачей которой является разработка этих вопросов, носит название **математической теории эксперимента.**

Структура эксперимента

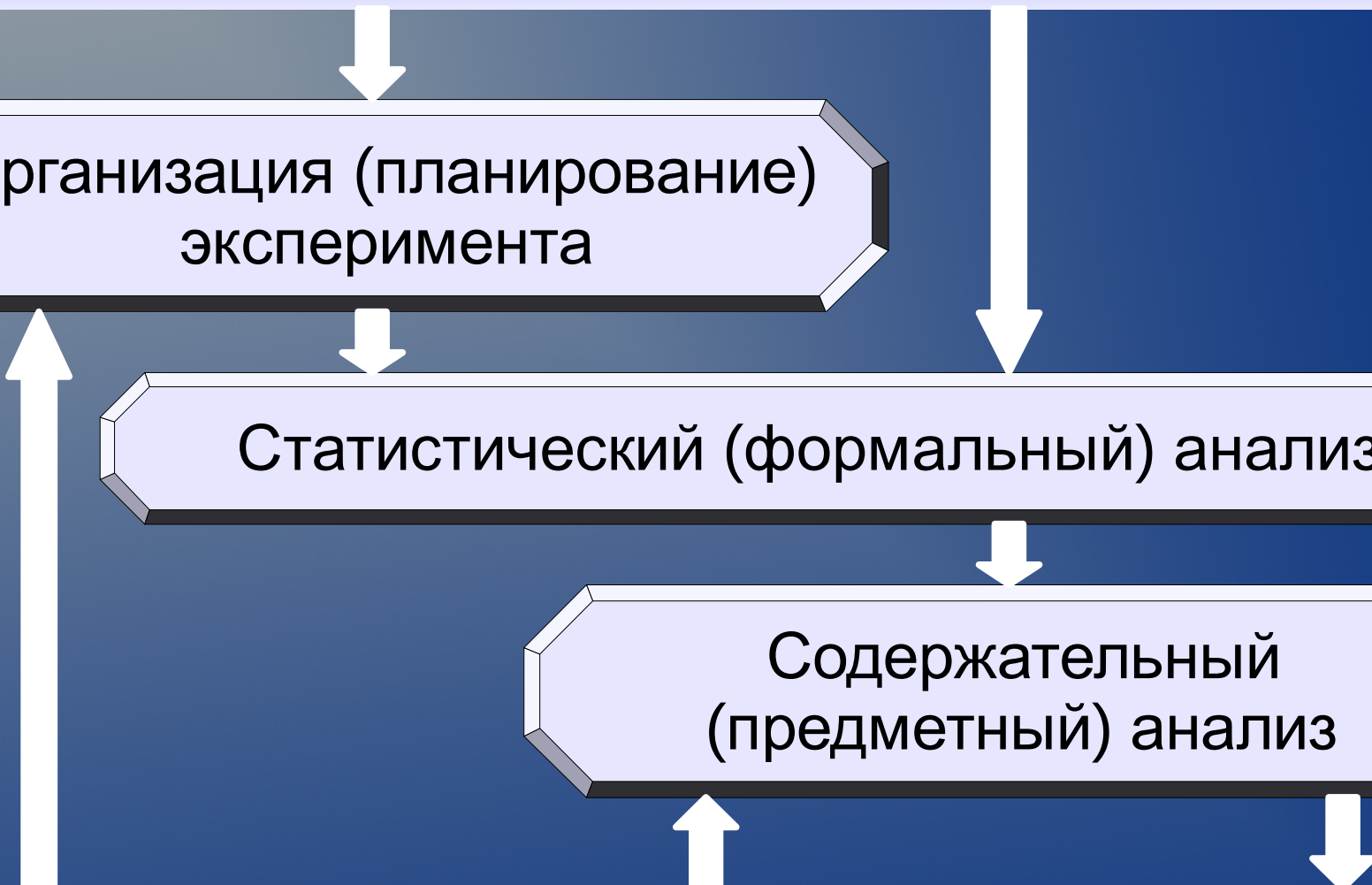
математическая теория эксперимента

Организация (планирование)
эксперимента

Статистический (формальный) анализ

Содержательный
(предметный) анализ

прикладная дисциплина



Терминология МТЭ

Входная переменная (варьируемый фактор) – независимая переменная, допускающая в эксперименте контролируемое изменение.

Факторное пространство – арифметическое пространство, координатами в котором являются варьируемые факторы.

Число варьируемых факторов называют *размерностью факторного пространства*.

Отклик – измеряемая в эксперименте величина.

Уровень фактора – его значение.

Множество уровней называют *планом эксперимента*.

Прикладная статистика

Статистическими данными называют данные, связанные с массовыми случайными явлениями.

Предметом *прикладной статистики* является разработка методов систематизации и анализа статистических данных.

Статистическими оценками называют функции от наблюдаемых значений. Скалярные оценки называют *точечными*.

Статистической моделью называют модель, полученную на основе статистических данных. Если статистическая модель включает набор параметров, то такую модель называют *параметрической*.

Однокритериальная задача

Отклик системы – скалярная величина

Число независимых управляющих переменных
(варьируемых факторов)

Требуемая полнота исследования: от выбора
направления в пространстве управляющих
переменных до координат целевой точки

Требуемая строгость исследования: от
параметров экспериментально-статистической
модели до вероятностных суждений

Дескриптивная статистика

Описательная (дескриптивная) статистика – это совокупность методов обработки данных, не включающая построение параметрических моделей. Центральное положение в описательной статистике занимают методы первичного анализа выборки значений одного признака.

Оценки математического ожидания и медианы:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad Me = \begin{cases} x_{(n+1)/2}, & n = 2k + 1 \\ \frac{x_{n/2} + x_{n/2+1}}{2}, & n = 2k \end{cases}, k \in N$$

Оценки дисперсии и стандартного отклонения:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad s = \sqrt{s^2}$$

Проверка статистических гипотез

Статистическая гипотеза – предположение о виде неизвестного распределения или о параметрах известного распределения.

Проверка гипотезы – сравнение вероятности связанного с ней *критического события* с вероятностью события, которое принимают практически невозможным (сравнение с *уровнем значимости*).

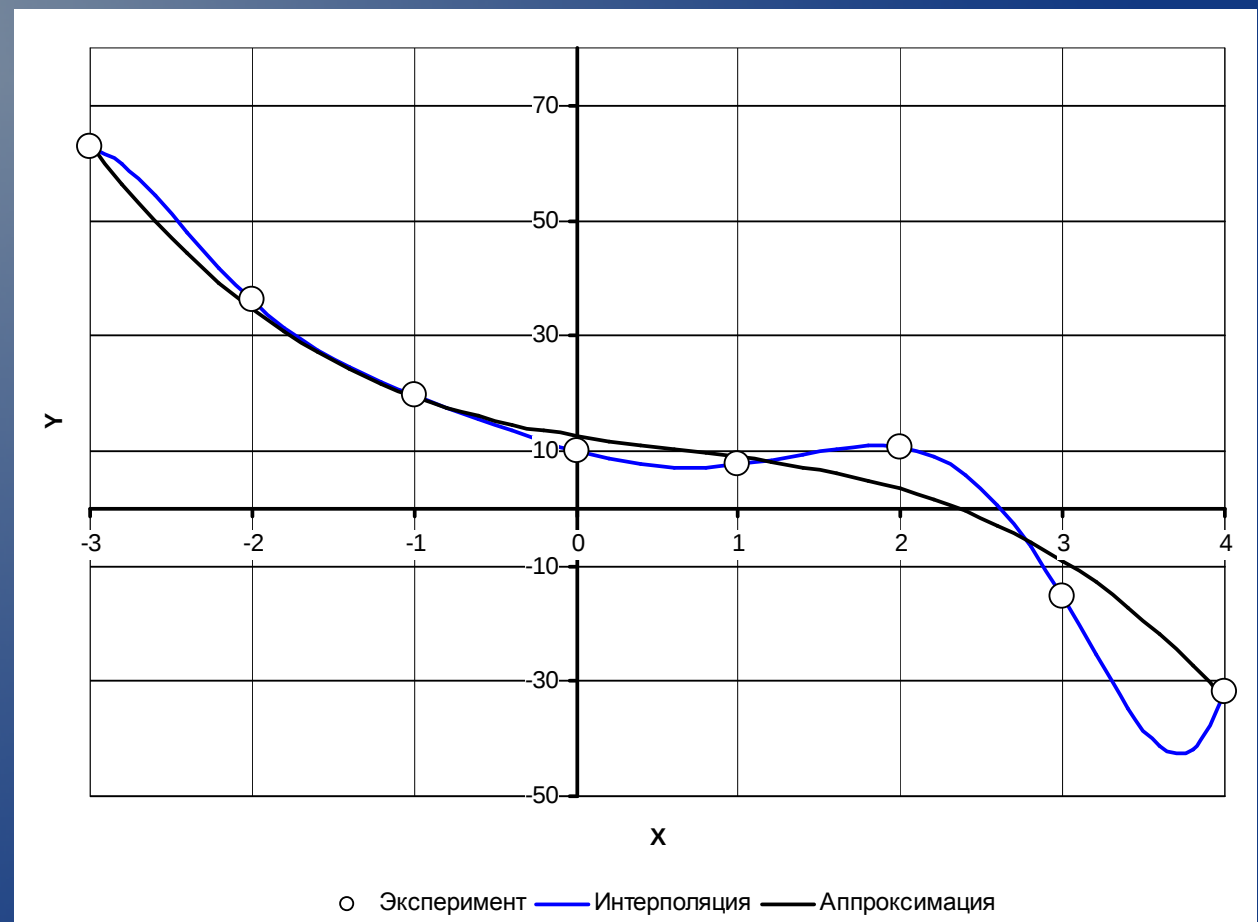
Основной одновыборочной задачей является задача проверки генеральной совокупности на нормальный закон распределения. Основными двухвыборочными задачами являются задачи проверки равенства математических ожиданий и дисперсий.

Интерполяция и аппроксимация

Одной из центральных прикладных задач является восстановление аналитических зависимостей по опытным (сеточным, табличным) данным.

Интерполяция — восстановление аналитической зависимости проходящей через все узлы сетки.

Аппроксимация — восстановление зависимости, «близкой» к исходной сетке.



Регрессионный анализ

В рамках *регрессионного анализа* исследуются зависимости между одной или несколькими неслучайными величинами (варьируемыми факторами) и одной случайной величиной (откликом).

Регрессионный анализ – основное средство концентрации, «свертки» эмпирической информации, отсева «информационного шума».

Конечная цель – получение регрессионной (экспериментально-статистической) модели: «наименее зашумленной» аппроксимирующей зависимости, которая была бы еще пригодна для адекватного представления опытных данных.

Выбор общего вида регрессионной модели – неформальная операция, основанная на интуиции исследователя. Интуиция, в свою очередь, основана на опыте решения аналогичных задач.

Поиск параметров регрессионной модели – формальная операция, в основе которой лежит принцип максимального правдоподобия: **наилучшим описанием исследуемой системы является такое, при котором максимальна вероятность получить измеренные на опыте значения отклика.**

Планирование эксперимента

Первый этап: выбор вида регрессионной модели.

Второй этап: выбор плана эксперимента, позволяющего построить выбранную регрессионную модель.

Третий этап: натурный эксперимент (получение выборки).

Четвертый этап: построение регрессионной модели.

Пятый этап: интерпретация регрессионной модели.

Корреляционный анализ

Корреляционный анализ – один из методов исследования взаимных зависимостей между двумя или более случайными величинами.

Линейный корреляционный анализ – исследование наличия или отсутствия линейной зависимости между случайными величинами.

Если по результатам выполнения линейного корреляционного анализа есть основания считать, что зависимость между случайными величинами имеется, то такие случайные величины называют коррелированными.

Из коррелированности следует взаимная зависимость, а из независимости следует некоррелированность, однако обратные утверждения в общем случае не справедливы.

Многокритериальные задачи

Отклик системы – векторная величина

Скаляризация (сведение к единому показателю)

Малое число индивидуальных скалярных критериев (скалярных откликов): неформальные методы скаляризации

Большое число индивидуальных скалярных критериев: формальные методы скаляризации (метод главных компонент, кластерный анализ)

Инструментальные средства

От специализированных до универсальных

Коммерческие

SPSS (www.ibm.com/spss) – одна из старейших и наиболее полных реализаций методов прикладной статистики.

Statistica (<http://www.statsoft.com/Products/STATISTICA-Features>) – подробная документация (<http://www.statsoft.com/textbook>).

Пакет анализа в составе MS Excel – неочевидный доступ, отдельные ошибки в реализации, сравнительно малый объем реализации, странная документация. Входит в состав пакета MS Office, но разработчик постепенно ограничивает доступ к пакету.

Свободные

Полнота реализации и качество документации варьируются в широком диапазоне.

OpenOffice Calc, Gnumeric – реализованы базовые и матричная алгебра. Реализации отдельных методов прикладной статистики включены в универсальные пакеты численной математики (SciLab, Octave). Особого внимания заслуживает проблемно-ориентированный (статистический) язык R.

Спасибо за внимание!



НОЦ НТ НИУ МГСУ