Теория вероятностей и математическая статистика

Лектор А.А. Лобузов

Семестр 6 **Лекция 1**

Введение в математическую статистику

Статистика – наука (область знаний) о сборе

(и методах сбора), обработке, анализе,

и интерпретации данных.

Задачи статистики:

•Организация сбора данных.

•Обработка собранных первичных данных (группировка, обобщение, оформление в виде списков, таблицах и т.п., графическое представление).

•Анализ данных (расчёт производных показателей, характеристик,).

•Интерпретация данных.

•Извлечение новой информации из данных.

В статистике, как научной дисциплине,

выделяют следующие разделы:

• **описательная статистика** (descriptive statistics) – часть статистики, которая описывает данные, т.е. обобщает данные и выделяет их характерные свойства.

• **статистики вывода** (inferential statistics) – часть статистики, которая делает выводы из данных.

Содержание математической статистики – разработка приёмов статистического наблюдения и анализа статистических данных.

Задачи математической статистики:

• разработка методов сбора и описания данных;

• разработка методов оценивания (построение точечных и интервальных оценок параметров распределений);

• разработка методов проверки статистических гипотез;

• разработка методов планирования экспериментов;

• разработка методов статистического последовательного анализа.

Основные понятия математической статистики

Статистическая модель $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathcal{P})$

$\mathcal{X}$ - выборочное пространство (совокупность всевозможных выборок размера N) (случайный вектор) (элементы вектора независимые одинаково распределены)

$\mathcal{B}(\mathcal{X})$ - некоторая σ-алгебра подмножеств $\mathcal{X}$

$\mathcal{P} = \{P\}$ - некоторое семейство распределений вероятностей, заданное на $\mathcal{B}(\mathcal{X})$

Если данные числовые, то $\mathcal{X} \in \mathbb{R}^N$, а $\mathcal{B}(\mathcal{X})$ - борелевская σ-алгебра

Статистическая модель $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathcal{P})$ называется параметрической если $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$, где Θ множество параметров, $\Theta \subset \mathbb{R}^r$ (r - размерность параметров)

В противном случае статистическая модель - непараметрическая

$x \in \mathcal{X}$, $x = (x_1, x_2, \dots, x_N)$ - выборка объёма N -

Случайный вектор $X = (X_1, X_2, \dots, X_N)$ со значениями в $\mathcal{X}$ называется случайной выборкой объёма N

Рассмотрим случайную величину

$\xi : \Omega \to \mathfrak{X}_1$, со значениями в $\mathfrak{X}_1 \subseteq R$

с функцией распределения

$F(x) = F_\xi(x) = P_\Omega(\xi \leq x)$

Пусть случайные величины $X_1, X_2, \ldots, X_n$

как и ξ заданы на вероятностном

пространстве (Ω, A, P_Ω), независимы

и имеют такое же ~~расп~~ распределение

$X_i \sim \xi$. Тогда $F_{\mathbf{X}}(x) = F_\xi(x_1) F_\xi(x_2) \ldots F_\xi(x_n)$

$x = (x_1, x_2, \ldots, x_n)$ - реализация случайной

выборки X.

Получаем статистическую модель, которая

называется моделью повторных независимых

наблюдений (испытаний), при этом $\mathfrak{X}_1$ называется

генеральной совокупностью, $\mathbf{X}$ называется

выборкой из генеральной совокупности, а также

выборкой из распределения с.в. ξ.

Модель повторных независимых наблюдений

кратко обозначается $\mathcal{F} = \{F_\xi\}$, или класс

допустимых распределений с.в. ξ определяется параметром $\theta \in \Theta$, то $\mathcal{F} = \{F_\xi(x, \theta) \mid \theta \in \Theta\}$

Модель $\mathcal{F} = \{F_\xi\}$ называется дискретной, если все допустимые распределения случайной величины ξ дискретны.

Модель $\mathcal{F} = \{F_\xi\}$ называется абсолютно непрерывной, если все допустимые распределения случайной величины абсолютно непрерывны.