

7. От Теории вероятностей к Статистике

Генеральная совокупность и выборка



- Даже те элементы ТВ, с которыми познакомились, позволяют решать многие практические задачи

умеем оценить надежность и риск отказа системы, если известна надежность ее элементов

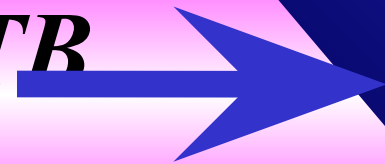
определить вероятность нарушить норматив по какому-то критерию, если знаем его $ZP \rightarrow$ риск отказа

- Главное – познакомились с основой, теоретической базой, которая позволит (при желании и / или необходимости) овладеть многообразием статистических методов решения практических задач

Методы – один из 3-х главных «смыслов» термина *статистика*
(см. «Статистика» Т.В. Ляшенко, стр. 5)

О СВЯЗИ

Статистики с ТВ



ТВ

Изучает «мысленный»
эксперимент

Предлагает
теоретические понятия
и модели –
теоретическая база С

С

Изучает реальный
эксперимент

Использует их для
извлечения из
статистических данных
информации для принятия
решений в условиях
неопределенности

Помогают в этом компьютеры с соответствующим
программным обеспечением – техническая база С
(изменили С)

ТВ

Аналогии

С

Изучает
СВ \equiv ЗР

Статистическая
совокупность – предмет
статистики

**Данные – сырьевая база
статистики**

Вероятность P

частота f (или w)

Пара
распреде

2 Проверяемая по статистическим
данным теоретическая модель
распределения

ФР и ПР,
 $F(x)$ и $f(x)$

Распределения частот,
 $F_n(x)$ и $f_n(x)$ **1**

Структура «Статистики»

(науки и учебного курса) МС

3 направления

- ▽ Эффективный сбор данных – подготовка сырья
- ▽ Собранные данные – их обобщение и представление
- ▽ Статистический анализ – исследование данных и выводы

извлечь полезную информацию
из доступных данных !

Цель статистических
средств и процедур

«Коснемся» этих
3-х разделов
Статистики

СЫРЬЕ

(см. версии
ЭТИХ ПОНЯТИЙ
на стр. 8)

В статистических исследованиях и
2 связанных друг с другом понятия:

генеральная совокупность (совокупность)

(массовые
однородные
испытания)

выборка (выборочная)

Г.С. – совокупность всех мыслимых
наблюдений за случайной величиной
(снятых с однородных объектов)

– значения x интересующего признака X ,
который варьирует от одного объекта к другому

В
ТВ

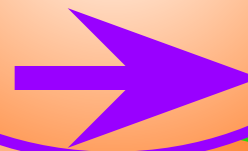
x_1, x_2, \dots, x_N – варианты
(значения СВ)

N – объем совокупности

Примеры:

- годовые доходы всех семей в определенном регионе, $N = 100000$
- все значения прочности одинаковых образцов

Поэтому
используют
выборки



Получить все элементы совокупности,
все возможные значения величины $X \rightarrow$
 \rightarrow трудоемко, дорого или невозможно

Выборка объема n из совокупности –
результаты ограниченного ряда наблюдений

x_1, x_2, \dots, x_n
за случайной величиной X

Примеры:

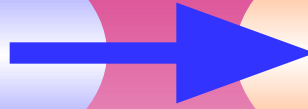
- годовые доходы 100 семей из
100000
- Результаты испытаний 6
образцов материала

**Возможность по выборочным
данным сделать заключение о
свойствах совокупности –
*суть статистического
метода***

**Его *назначение* – по
выборке получить
количественные основания
для решений относительно
всей совокупности**

**Выборочный \cong эмпирический \cong
статистический метод**

**Судить по части
о целом**

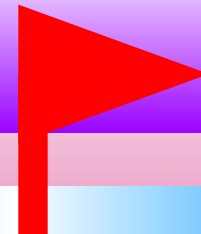
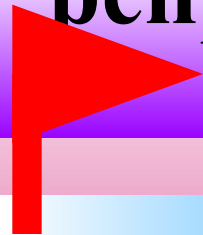


**Важно и нужно
судить ВЕРНО**

*Пример
?*




**Выборка должна быть
репрезентативной**



**равный шанс для
каждого элемента
попасть в выборку**


**достаточный
объем
выборки**

8. Описательная статистика



Четкое, яркое
представление
конфигурации
данных
(не «куча»)

См. «практику» и
пособие



Выборочные
числовые
характеристики
распределений –
статистики
(в частности, оценки
параметров)

Числовые характеристики эмпирических распределений

ЧХ распределения признака X – получаемые по статистическим данным числа, характеризующие наиболее существенные черты распределения

По всей совокупности – *генеральные*
(в частности, параметры θ : μ , σ , ...)

По выборочным данным – *выборочные*
(в частности, оценки $\hat{\theta}$ параметров θ :
 $\hat{\mu} = \bar{X}$, $\hat{\sigma} = s$, ...)

Характеристики центра - средние

**Среднее
арифметическое
совокупности
(вариантов)**
- *генеральное
среднее*

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

*Выборочное
среднее
- оценка
генерального
среднего*

Свойства среднего

Плохой пример

*Медиана (Me) - срединный элемент
вариационного ряда
Номер медианного элемента - $(n + 1) / 2$*

В данном примере

*Процентили - на 100, децили на - 10,
квантили - на равные*

порядковые статистики

Me делит на 2 секции.

Квартиль Q - на 4: $Q_1, Q_2 = Me, Q_3$

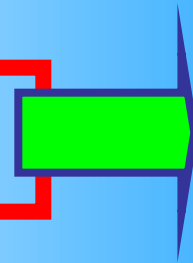
Другие характеристики центра

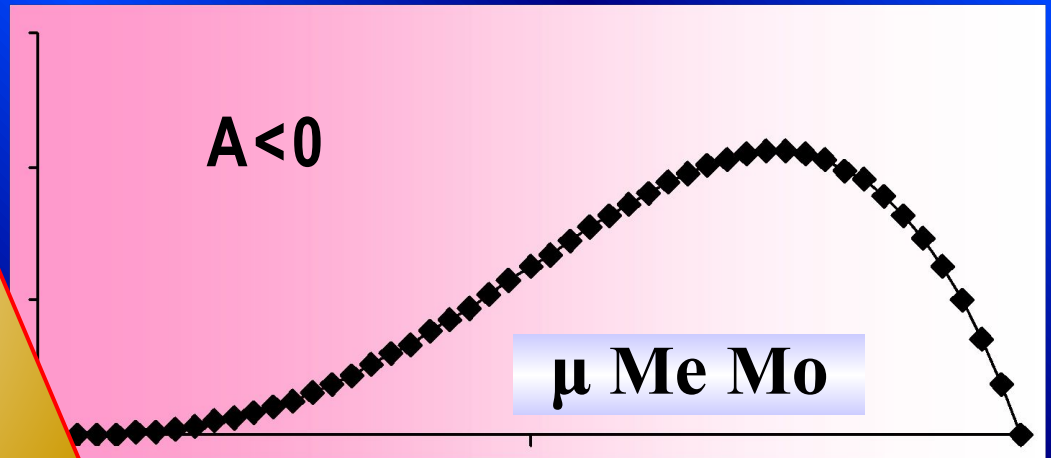
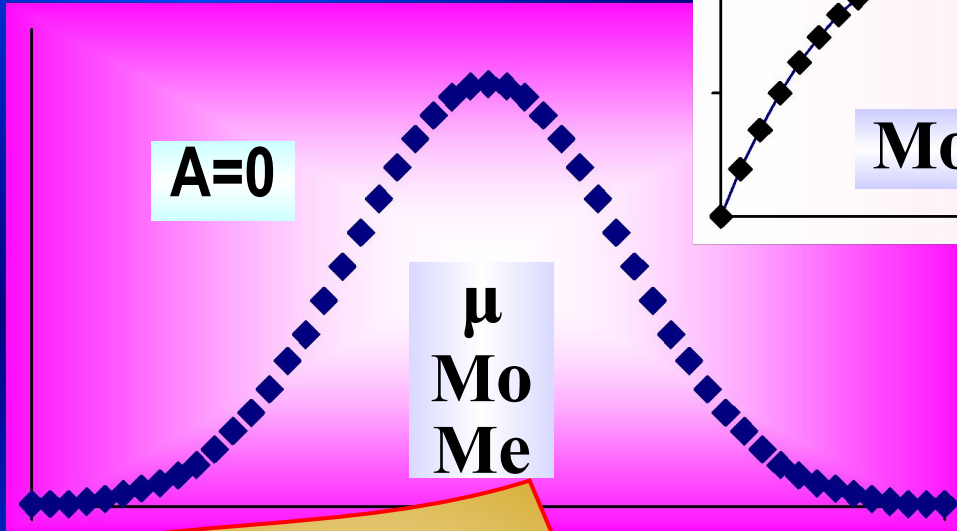
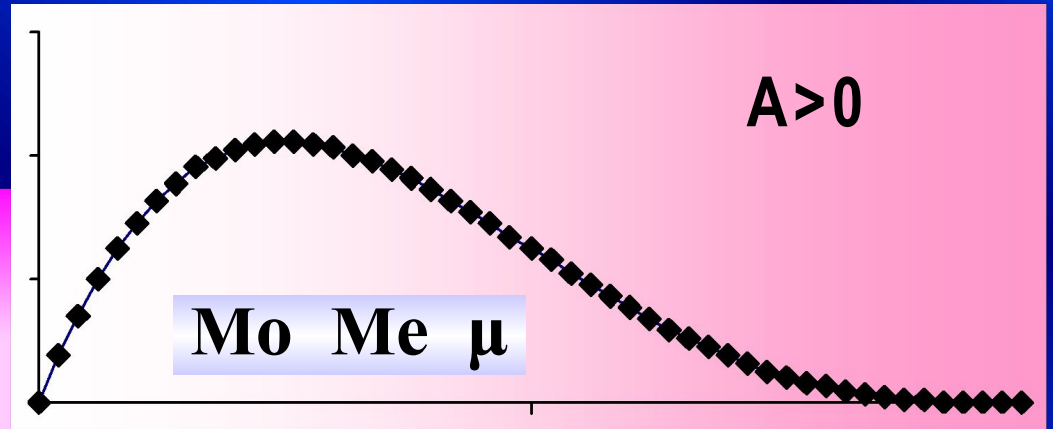
Взвешенное, групповое, интегральное, общее

среднее

Мода

Соотношения $\square X, M_o, M_e$





Соотношени
е
 μ Me Mo
указывает на
форму
распределен
ия

Характеристики рассеяния

Генеральные

Генеральная дисперсия

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

Генеральное
стандартное отклонение

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

$$z = \frac{(x - \mu)}{\sigma}$$

Нормализованное отклонение

Выборочные



Выборочная дисперсия

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{(n - 1)} = \hat{\sigma}^2$$

Если $n \rightarrow$
смещенная
оценка,
можно при
больших n
($n \approx n - 1$)

ис
пр

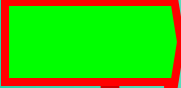
степеней

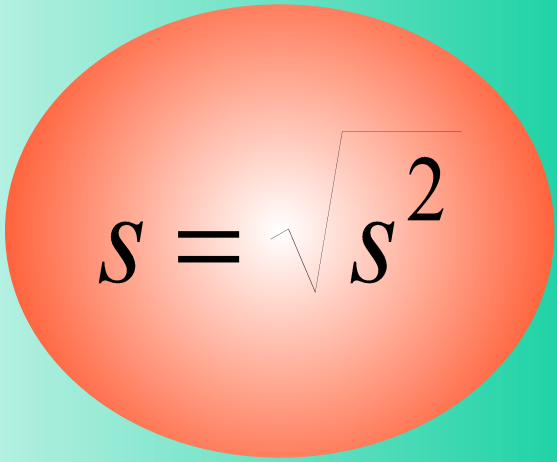
1 с.с. «потеряна»
наложена 1 связь

Число степеней свободы (f, k, v, \dots) – число «свободных друг от друга», независимых, не связанных элементов данных (несущих уникальную информацию)

**Выборочное
стандартное
отклонение**

$$s = \sqrt{\frac{\sum (x_i - \bar{X})^2}{(n-1)}} = \hat{\sigma}$$


$$s^2 = \frac{n \sum x_i^2 - (\sum x_i)^2}{n(n-1)}$$


$$s = \sqrt{s^2}$$

$$z = (x - \bar{X}) / s$$

**Выборочное
нормализованное
отклонение**

Генеральный
коэффициент
вариации

$$v = \sigma / \mu$$

Выборочный
коэффициент
вариации

$$\hat{v} = s / \bar{X}$$

И другие статистики

**The
End**