

# **МНОГОМЕРНЫЕ МЕТОДЫ АНАЛИЗА ДАННЫХ**

**Что нужно вспомнить из  
предыдущего курса  
«БИОСТАТИСТИКА (= БИОМЕТРИЯ)**

# Общие понятия

- **1. Выборка и генеральная совокупность**
- репрезентативность (равновероятность), случайность (нетенденциозность)
- Случайные и типические (“послойные”) выборки
- **2. “Уровни” вероятности при суждениях**
- Вероятности 0.05 – 0.01 – 0.001

- **3. “Классы” признаков**
- (На самом деле это группы не самих признаков, а особенностей их выражения!!!)
- Качественные=номинальные (цвет, форма, принадлежность к определенной группе, таксону), в том числе – альтернативные (есть-нет = 0 – 1 или “+” - “-” = «видовые списки»),
- Порядковые (больше-меньше) = ранговые или балловые.  
 (Ранжировать можно значения признака, выраженные количественно! Балловые оценки "назначаются" заранее),
- Количественные = интервальные (мерные=измеряемые=метрические и счетные=меристические).

*Attributes – ranked variables – measurement variables*

# Статистики и критерии

- Описательные статистики (descriptive statistics)

а) показатели *положения или центральной тенденции* (средние)

$X = M$ ; arithmetic mean

медиана ( $Me$ )

мода ( $Mo$ ) и др.

- б) показатели разброса  
(варьирования)

пределы (минимум-максимум),

среднее квадратическое отклонение =  
стандартное отклонение (standart  
deviation –  $SD$ ),

дисперсия (варианса – variance –  $SD^2 =$   
 $S^2$ ),

коэффициент вариации ( $CV$  или  $V$ );  
квантили (квартили, процентиля)

- в) показатели **формы распределения**  
(distribution)

**сравнение с эталонными  
распределениями;**

**коэффициенты асимметрии и  
эксцесса (skewness, kurtosis)**

- 2. Показатели **связи - зависимости**  
коэффициенты корреляции  
(Браве-Пирсона; ранговые;  
коэффициенты сопряженности для  
качественных признаков);  
  
коэффициенты и уравнения  
регрессии.

**В ЧЕМ РАЗЛИЧИЕ ЗАДАЧ  
(и интерпретации результатов)?**

- **Корреляция** = согласованность в изменчивости разных признаков
- **Регрессия** = зависимость изменений одного признака от значений другого (других)



### 3. Характеристики отношений выборки и совокупности

- а) ошибки (= показатели разброса выборочных характеристик при повторных выборках из той же совокупности) – standart errors (of statistics) –  $SE$  ( $s_{st}$ )
- б) доверительные интервалы – confidence limits –  $SE * (coefficient)$

# 4. Методы и критерии сравнения

- Одна выборка

- а) отдельного значения и выборки («выбраковка» - ???) ; выборки и «эталона»;

- Выборок больше одной

- б) разных выборок по показателям... (см. пункты 1-2);

для средних арифметических – t-критерий Стьюдента,

для дисперсий – F-критерий Фишера,

для формы распределения –  $\chi^2$  Пирсона и др.

# Две выборки или - больше двух...

- Почему столько «сложностей» и чем принципиально отличается
- ДИСПЕРСИОННЫЙ АНАЛИЗ?
  - **для числа выборок больше двух**
- Заметим: этот анализ не заменяет, но и не исключает попарное сравнение
- Изменяется «идеология» оценки различий между конкретными выборками (об этом – позже)

## Цели курса:

- 1) ознакомление с разнообразными методами анализа данных; соответствие задач исследования и методов,
- 2) сильные и слабые стороны методов,
- 3) связи между методами: использование результатов, полученных в рамках одного из методов, как исходных данных для другого,
- 4) ознакомление с терминами и понятиями, необходимыми при чтении публикаций.
- 5) приемы реализации анализа данных по основным методам с помощью пакета STATISTICA for WINDOWS (практика),

Предполагаемая  
последовательность работы  
исследователя

- **1. Постановка и формулирование задачи исследования в биологических понятиях**
- **2. Постановка и формулирование задачи исследования в математических понятиях,**
- **3. Подбор адекватного метода анализа ожидаемых результатов**
- **4. Планирование эксперимента (наблюдения) в соответствии с пп.2-3.**
- **5. Анализ полученных результатов, представление их в наглядном виде, интерпретация**
- **6. Корректировка плана дальнейшего исследования.**

- По 6 моллюсков из двух популяций
- (Взяли из двух мешков)
- RAPD
- Один из признаков (банды) присутствовал только у особей из популяции А
- Остальные – полиморфны в каждой из выборок

**Как можно было бы сделать?**

## *ОСНОВНЫЕ ЗАПОВЕДИ:*

- **1. Планируя исследование, сразу определите наиболее подходящий/ие методы анализа его результатов.**
- **2. Не откладывайте анализ получаемых результатов по полного завершения исследования. Промежуточные результаты могут быть использованы при корректировке плана работы, а также помогут освоить особенности выбранного метода анализа.**
- **3. Обязательно проверяйте правильность введенных в компьютер данных. Ошибки ввода могут оказать существенное влияние на результаты!**



- **4. Проверьте соответствие результатов, полученных с применением компьютерной программы и результатов “ручных” вычислений (для более сложных алгоритмов – введите простую “тестовую” задачу: например, из учебника). Не всякий результат, выданный компьютером, - безупречен!**
- **5. Первый результат не следует рассматривать как единственный и окончательный, он дает лишь материал для обдумывания плана последующих действий**