

ОСНОВЫ СТАТИСТИКИ

Анатолий Карпов

1. Введение (первая неделя)

1.1 Информация о курсе

Этот курс посвящен введению в статистику. Этот курс будет состоять из трех недель:

- первая неделя - вводная, поговорим о базовых понятиях - выборка, генеральная совокупность, описательные статистики; разберемся с фразой Статистически значимые различия
- вторая неделя - регрессионный анализ, t-критерий, ANOVA
- третья неделя - корреляция и регрессия

1.2 Генеральная совокупность и выборка

1.2.1 План урока

- Понятие генеральной совокупности и выборки, репрезентативность выборки
- Простая случайная выборка
- Стратифицированная выборка
- Групповая выборка

1.2.2 Лекция

Генеральная совокупность - множество всех тех объектов, относительно которых мы будем делать выводы.

Например статистика по росту всех людей побывавших на луне; ГС будет все эти люди, их легко посчитать, измерить рост и т.д. Но в большинстве случаев объемы ГС очень велики (в некоторых биологических исследованиях объем популяции может насчитывать несколько тысяч представителей). Для того чтобы решить эту проблему исследователь выбирает только часть ГС т.е. формирует выборку:

Выборка - часть ГС, результаты исследования которой можно обобщить на ГС.

Очень важным условием является репрезентативность выборки, т.е. мы хотим чтобы наша выборка была моделью ГС, отражала свойства ГС.

Чтобы этого добиться существует несколько вариантов формирования репрезентативной выборки.

1.2.3 Задача

Если мы провели исследование с целью выявить уровень знаний в области биологии на 100 студентах биологического факультета СПбГУ, то на какую совокупность мы можем распространить наши выводы?

Варианты ответов:

- Студенты СПбГУ
- Студенты биологического факультета СПбГУ
- Мужчины и женщины в возрасте от 18 до 22 лет

1.2.4 Задача

Если объем выборки достаточно велик (больше 100), то такая выборка является репрезентативной?

- Да
- Нет
- Зависит от способа формирования выборки

1.2.5 - 1.2.7 Лекции

Виды выборок:

- Простая случайная выборка (simple random sample) - случайным образом выбираем элементы ГС; чем больше будет объем выбранных объектов, тем сильнее наша выборка будет походить на ГС по своим характеристикам.
- Стратифицированная выборка (stratified sample) - разбиваем ГС на несколько обособленных групп (strat) различных по своей природе, а уже после осуществляем выборку.
- Групповая выборка (cluster sample) - разбиваем ГС на кластеры очень похожие между собой и уже из них случайным образом осуществляем выборку.

1.2.8 Задача

Соотнесите способы формирования выборки с предложенными примерами.

А. Для того чтобы протестировать курс по введению в статистику и выяснить, насколько хорошо с ним справятся студенты Института биоинформатики, случайным образом было выбрано 50 студентов института.

Б. Для исследования взаимосвязи риска развития заболевания и группы крови человека потенциальных участников исследования разделили на четыре группы (в соответствии с группой крови). Затем из каждой группы случайным образом извлекли по 50 человек.

В. Чтобы проверить знания школьников Санкт-Петербурга по математике, было организовано исследование. Случайным образом было выбрано 10 школ, затем из каждой школы случайным образом было отобрано по 50 учащихся разных классов.

Ответы:

- А - Простая случайная выборка (random sample)
- Б - Стратифицированная выборка (stratified sample)
- В - Групповая выборка (cluster sample)

1.2.9 Задача

Предположим, после лекции по статистике, один студент решил выяснить, как хорошо студенты психологического факультета разбираются в этом предмете. Он подготовил серию заданий и пригласил 30 своих друзей с факультета принять участие в тестировании. Исследователь утверждает, что он сформировал простую случайную выборку. Так ли это?

- Это не простая случайная выборка. Не каждый член генеральной совокупности мог равновероятно принять участие в тестировании, т.к. исследователь позвал только своих друзей.
- Такая выборка должна рассматриваться, как стратифицированная выборка. Можно сказать, что исследователь разделил всех студентов на страты по 30 человек и выбрал только одну из них для исследования.
- Такая выборка может рассматриваться, как простая случайная выборка. Тридцать человек действительно являются студентами данного факультета.

1.3 Типы переменных

1.3.1 План урока

- Типы переменных
- Количественные переменные
- Номинативные переменные
- Ранговые переменные

1.3.3 Лекция

Глобально все переменные можно разделить на два типа - количественные и качественные (номинативные)

Количественные переменные - измеренные значения некоторого признака. Если переменная может принимать любое значение на промежутке, то такая переменная называется непрерывной.