

Как правильно и неправильно использовать статистику

Никита Алексеев,
Лаборатория им. П.Л. Чебышева

Эксперимент или наблюдательное исследование

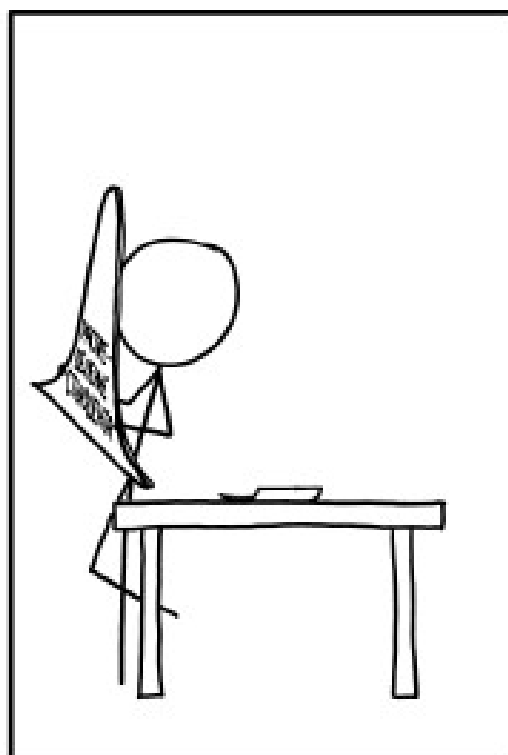
Наблюдательное исследование	Эксперимент
<ul style="list-style-type: none">• Исследователь не вмешивается в сбор данных• Можно понять, что переменные связаны, но не установить причинно-следственную связь• Исследование использует данные из прошлого (ретроспективное)	<ul style="list-style-type: none">• Исследователь может назначать контрольную и экспериментальную группу случайным образом• Можно установить причинно-следственные связи между переменными• Исследование проспективное

Репрезентативность выборки

- Сбор данных с тех, кого проще всего включить в исследование
- Метод сбора данных
- Сбор данных с людей, у которых есть ярко выраженное мнение по поводу исследуемого вопроса

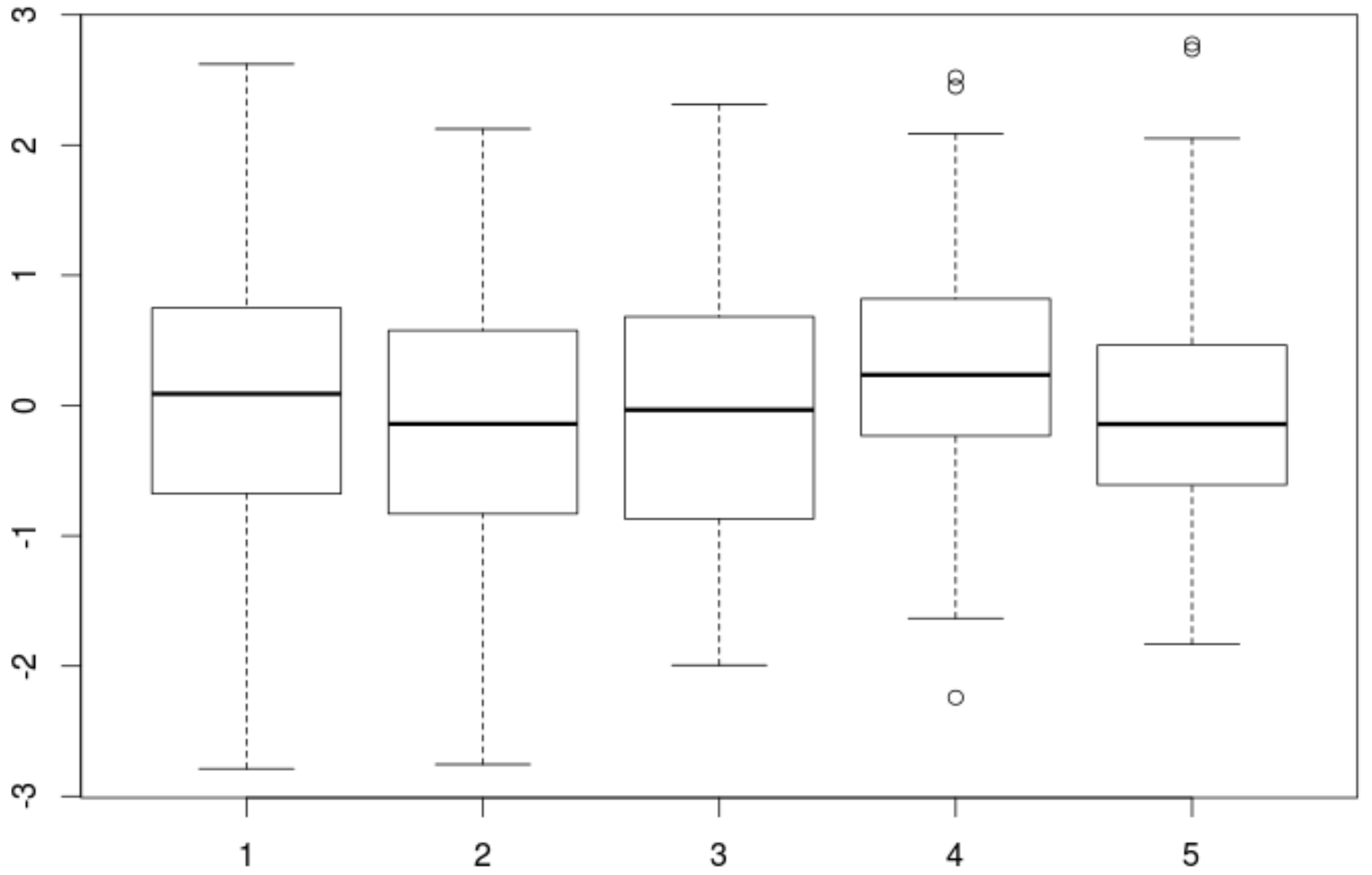
Тесты

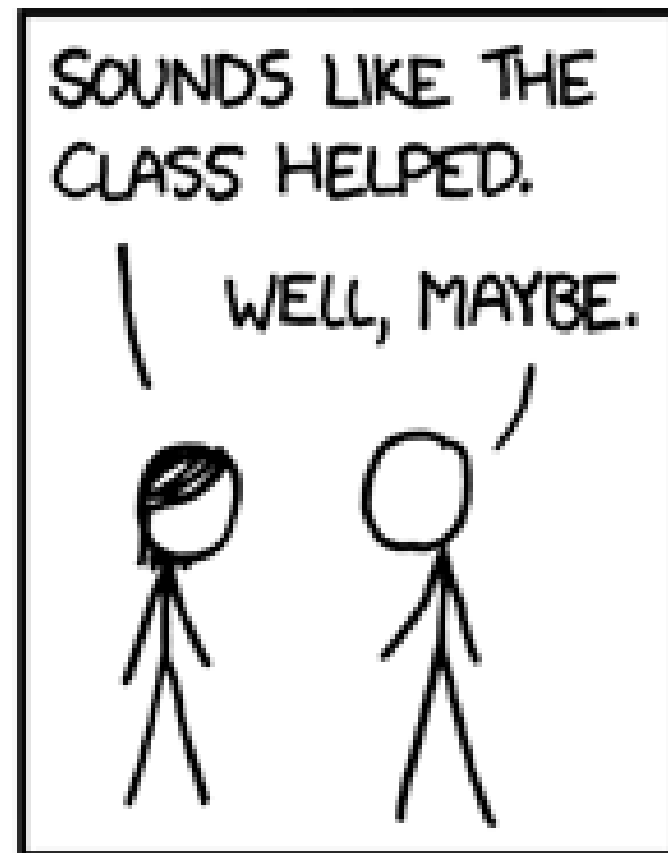
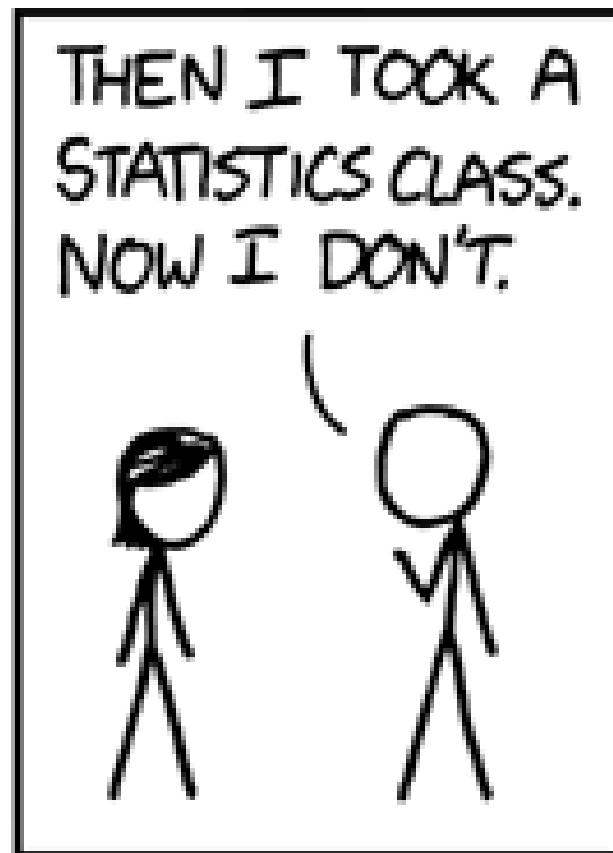
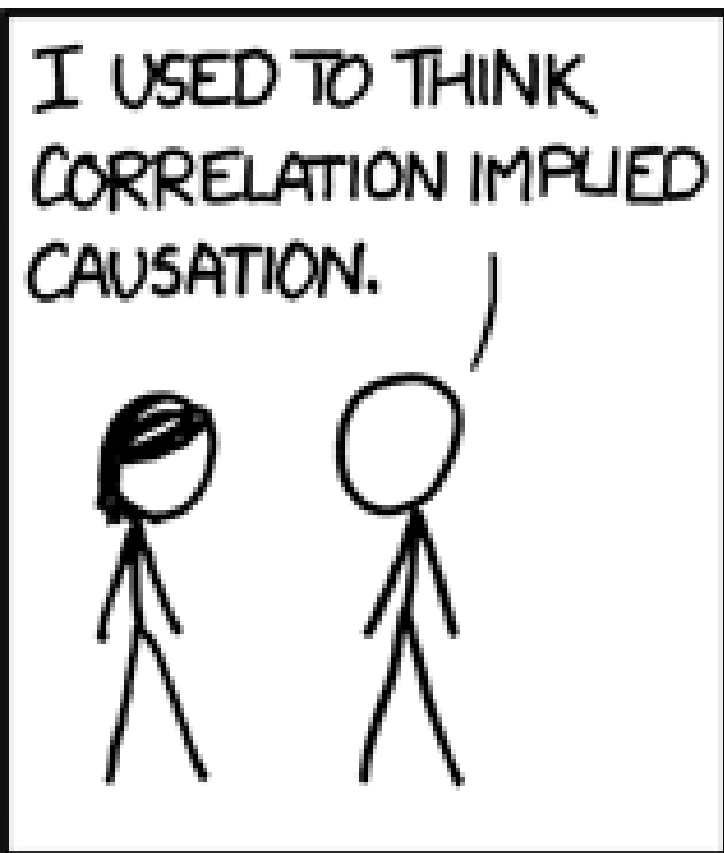
- Сравнение средних:
Т-критерий Стьюдента можно использовать только проверив данные на нормальность.
- Альтернативы: критерии Манна-Уитни и Уилкоксона



Множественные сравнения

- Сравнили контрольную группу с несколькими экспериментальными, в одном случае получили $p\text{-value} < 0.05$, объявили, что есть значимая разница.
- Поправка Бонферрони:
 $\alpha_{\text{real}} < k * \alpha_1$
 α_{real} – вероятность ошибки хотя бы в одном сравнении,
 k – количество сравнений
 α_1 – уровень значимости при одном сравнении





Корреляции

- Количество пожарных, прибывших на пожар, и ущерб от пожара
- Количество камер видеонаблюдения и количество карманных краж
- Ложная корреляция: количество аистиных гнезд и количество новорожденных