

Основы анализа и визуализации данных для медиа

Семинар 1
3, 7 сентября 2019

Рогович Татьяна
rogovich@gmail.com

Введение в статистику. Основные понятия

Структура семинара

- Обсудим, что такое статистика и научный процесс: как мы движемся от исследовательского вопроса к выводам
- Выясним, чем отличается наблюдение от эксперимента
- Разберемся, что такое переменные и какие они бывают
- Поговорим про измерения и ошибки, операционализацию и концептуализацию

Статистика –

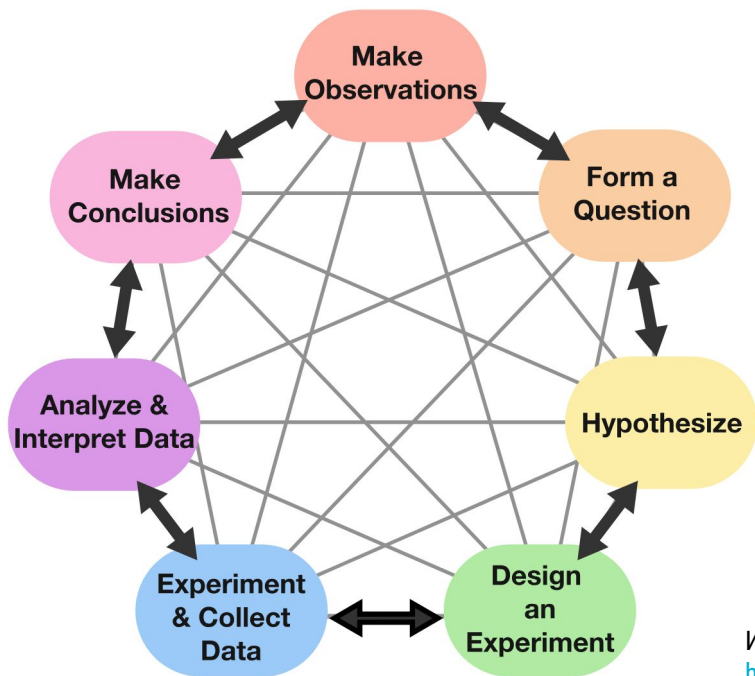
а) числовые данные, такие как уровень безработицы, число людей, умирающих ежегодно от пчелиных укусов.

б) определенные методы анализа данных и результаты такого анализа, такие как t -статистика или статистика хи-квадрат.

в) числа, использованные для описания выборок, в противоположность параметрам (числам, характеризующим генеральную совокупность).

г) область науки, которая разрабатывает и использует математические методы для описания данных и формирования суждений о них.

Диаграмма научного процесса



- Вместо эксперимента может быть проведено исследование по данным наблюдений или опроса
- В социальных науках есть дополнительные сложности с точным определением переменных, их измерениями и проверкой.

Что проще: узнать скорость ветра или уровень счастья человека?

Источник картинки и упражнения:

<https://brilliant.org/practice/observations-questions-hypotheses/?chapter=the-scientific-process>

Методы исследования

Метод эксперимента Experimental Study

это метод научного познания, при котором объекты погружаются в искусственно созданную среду, а их поведение управляется экспериментатором. Главная цель такого действия – проверка гипотезы, поиск новых фактов, которые могут ответить на важные для науки вопросы.

Например,

<https://nplus1.ru/news/2018/10/02/choice-overload>

<https://nplus1.ru/news/2017/09/26/marshmallow-test>

Методы исследования

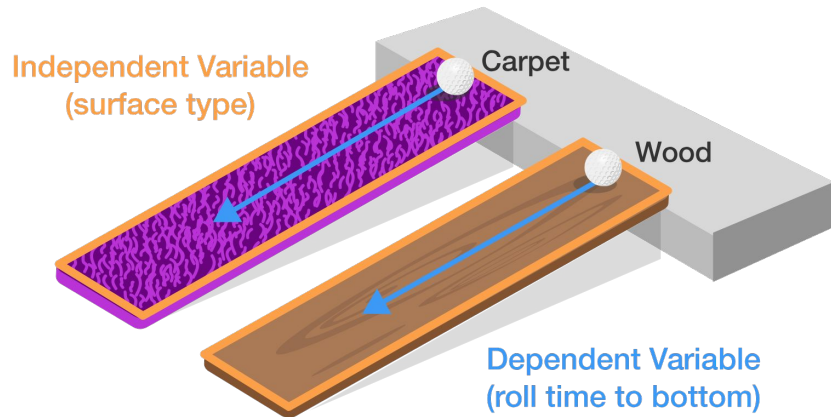
Метод наблюдения Observational Study

это метод познания, при котором наблюдатель изучает свойства исследуемого объекта и фиксирует их. Вмешательство в естественную среду минимально, а проводить данные мероприятия может любое лицо, даже при отсутствии оборудования и техники, а также специальных познаний.

Например,

<https://nplus1.ru/news/2015/11/18/laresistance>

Пример эксперимента



Исследовательский вопрос:

Влияет ли тип поверхности на скорость мяча, катящегося по ней?

Гипотеза:

При прочих равных по гладкой поверхности, мяч будет катиться быстрее.

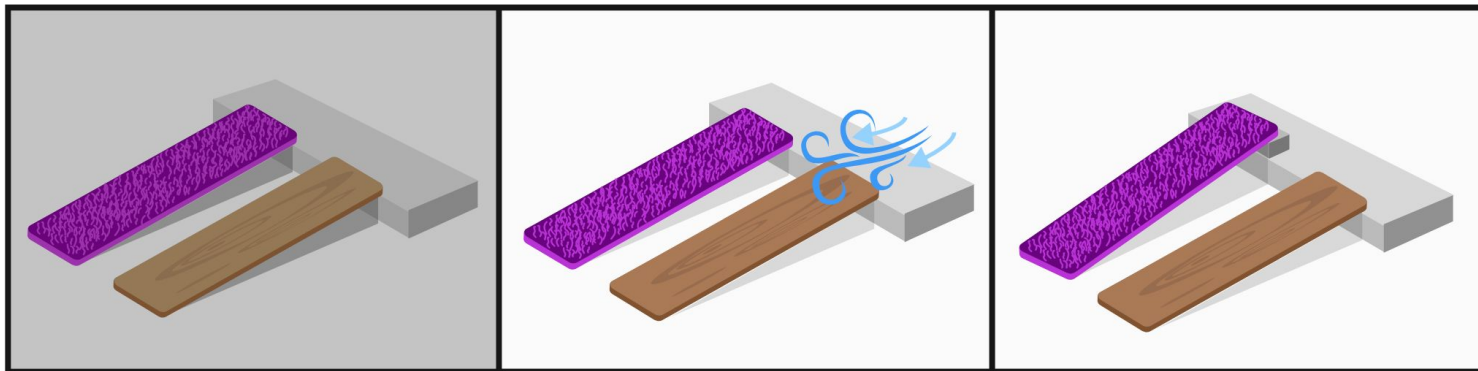
Пример эксперимента

Что еще может влиять на результат эксперимента?

а) яркость

б) ветер

в) наклон



Переменные

Переменная

это эмпирически наблюдаемое свойство
некоторого явления, которое может
принимать более одного значения.

Пример эксперимента

Зависимая переменная:

Независимая переменная:

Переменные для контроля:

Пример эксперимента

Зависимая переменная: скорость мяча
(то, что изучаем)

Независимая переменная: тип поверхности
(то, что может влиять согласно гипотезе)

Переменные для контроля: тип меча, ветер, влажность и т.д.
(то, что тоже может влиять,
но чье влияние хотим исключать)

Найдите переменные в гипотезах

- Чем выше интерес человека к материалам на политические темы в СМИ, тем выше его политическая активность.
- Политические предпочтения горожан на выборах в местные органы власти определяются возрастом, полом, образованием, но в наибольшей степени – родом занятий.
- Чем ниже уровень правового просвещения и уровень правосознания населения местного сообщества, тем больше склонность некоторых его членов к совершению противоправных поступков и опасность девиантного поведения.

Виды переменных

Количественные



Дискретные

Непрерывные

Категориальные



Номинальные

Порядковые

Бинарные

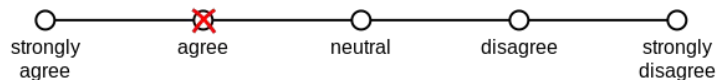
Виды переменных

Переменная	Определение	Ключ
survival	Выжил ли в крушении	0 = Нет, 1 = Да
pclass	Класс билета	1 = 1-ый, 2 = 2-ой, 3 = 3-ий
sex	Пол	М = Мужской, Ж = Женский
age	Возраст в годах	-
fare	Стоимость билета	-
cabin	Номер каюты	-
embarked	Порт посадки	C = Cherbourg, Q = Queenstown, S = Southampton

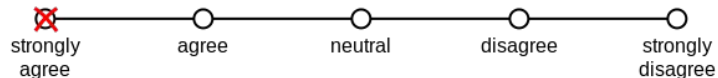
Виды переменных

Website User Survey

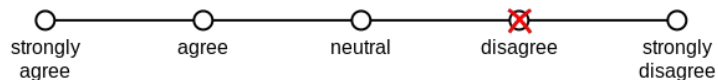
1. The website has a user friendly interface.



2. The website is easy to navigate.



3. The website's pages generally have good images.



Шкала
Ликерта
Likert scale
(порядковые
переменные)

Измерение Measurement

Процесс систематического присвоения чисел объектам и их свойствам для облегчения использования математического аппарата при изучении и описании объектов и их взаимосвязей.

Измерение и ошибка

$$X = T + E,$$

где X – наблюдаемое значение измерения, T – истинное значение, а E – ошибка.



Случайная

Систематическая

(Закономерность можно обнаружить! -> история)

Корректность измерений

Концептуализация Conceptualization

процесс наделения исходного понятия конкретным смыслом. В науке это наделение понятия теоретическими признаками, которые согласуются с общепризнанной научной теорией, а не с чьими-то субъективными взглядами

Например, дать определение “среднего класса”.

Пример

Проведите концептуализацию понятия “доверие СМИ”.

Корректность измерений

Операционализация Operationalization

процесс определения способа описания и измерения признаков.
Перевод абстрактных понятий в эмпирические величины.

Проблема операционализации более существенна в социальных науках, где свойства интересующего нас объекта часто нельзя измерить напрямую.

Пример

Например, команда приобрела баскетболиста, обладающего *спортивным мастерством*. Это абстрактное утверждение. Как на самом деле оценивает этого игрока тренер?

Корректность измерений

Опосредованное измерение

Proxu variable

процесс замены одного измерения другим
(например, для снижения стоимости исследования или из-за чувствительности вопроса).

Корректность измерений

Надежность

Reliability

согласованность или воспроизводимость наблюдений.

Например, надежность внутренней непротиворечивости (разные вопросы измеряют один и тот же объект, критерий - альфа Кронбаха), надежность вариантов (разные варианты вопроса стабильно измеряют одно и то же), надежность шкалы измерения (измерения последовательны).

Корректность измерений

Валидность

Validity

насколько хорошо тест или балльная шкала измеряют то, что планировалось измерить.

Типы: содержательная валидность, конструктивная валидность, совокупная валидность, предсказательная валидность и т.д.

Корректность измерений

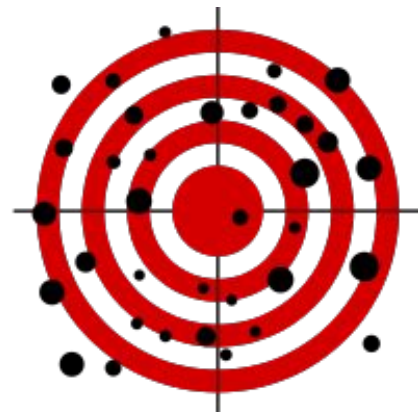
Содержательная валидность характеризует, насколько хорошо измерения характеризуют ключевое содержание объекта исследований (например, тест на умение программировать при приеме на работу - насколько он хорошо тестирует все ключевые навыки).

Предсказательная валидность – способность делать предсказания касательно некоторого события в будущем. Например, предсказывает ли тест, проведенный в начале года оценки в конце года.

Надежность и валидность



Unreliable & Invalid



Unreliable, But Valid



Reliable, Not Valid



Both Reliable & Valid