Как работать с данными и не чувствовать беспомощность?

Никита Алексеев

Computational Biology Institute,
George Washington University

26 июля

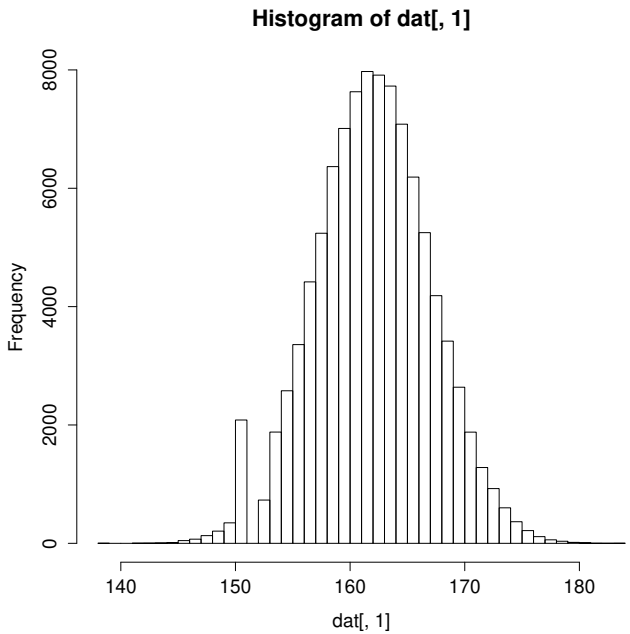


Рис. : Кеттле, 1844: распределение роста молодых людей призывного возраста во Франции

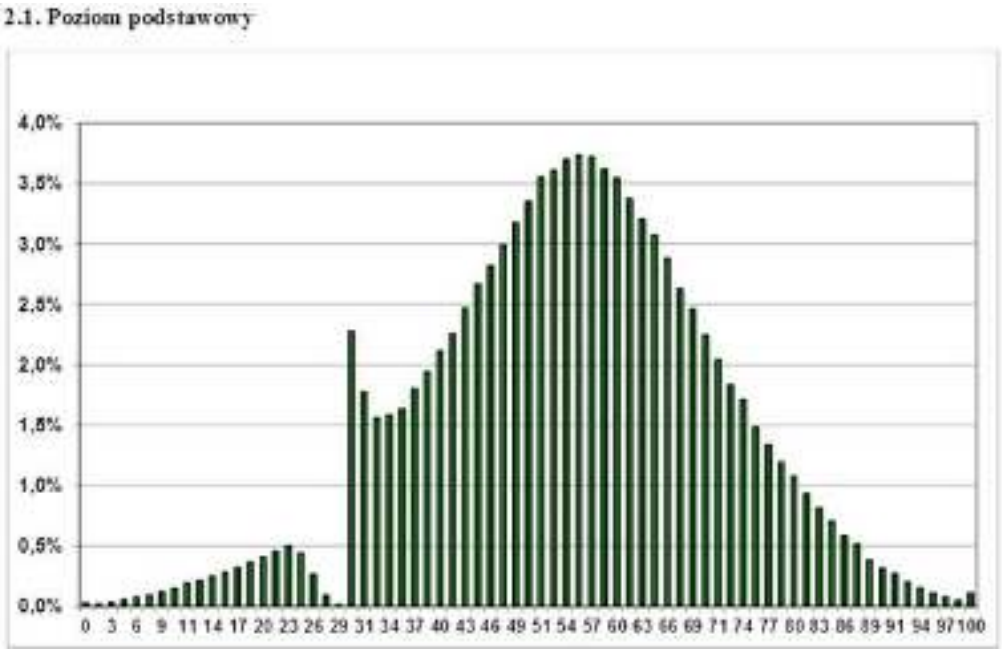


Рис. : Результаты государственного экзамена в Польше

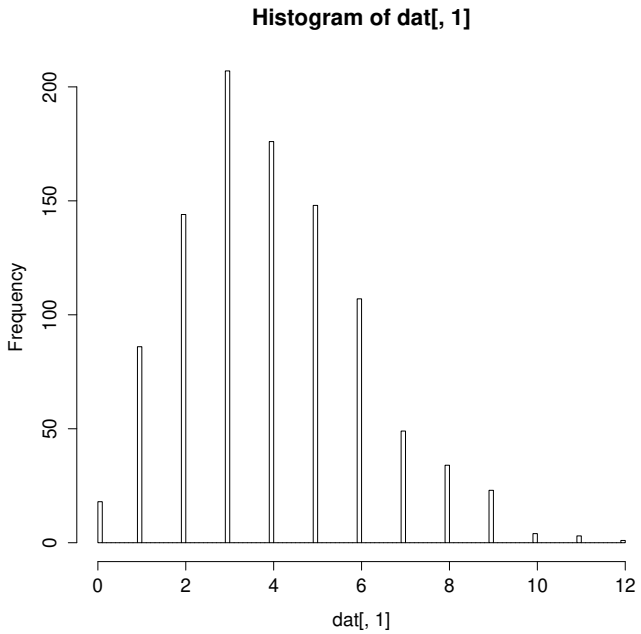


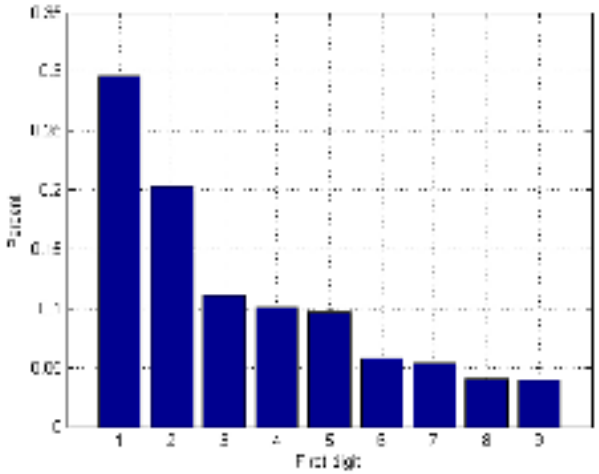
Рис. : Как много слов знал Шекспир? [Bradley Efron and Ronald Thisted, “Estimating the Number of Unseen Species: How Many Words Did Shakespeare Know?”, Biometrika (1976)]

Закон Бенфорда

Американский астроном Саймон Ньюком в 1881 году обнаружил, что книги, содержащие логарифмические таблицы, истрёпаны там, где содержатся логарифмы чисел, начинающихся с единицы, и целы для чисел, начинающихся на 9.

Это явление было повторно обнаружено физиком Фрэнком Бенфордом в 1938 году. Бенфорд проанализировал около 20 таблиц, среди которых были данные о площади бассейна 335 рек, удельной теплоёмкости и молекулярном весе тысяч химических соединений и, в том числе, номера домов первых 342 улиц, указанных в справочнике.

Закон Бенфорда



1	2	3	4	5	6	7	8	9
30,1%	17,6 %	12,5%	9,7%	7,9%	6,7%	5,8%	5,1%	4,6%

Корреляции

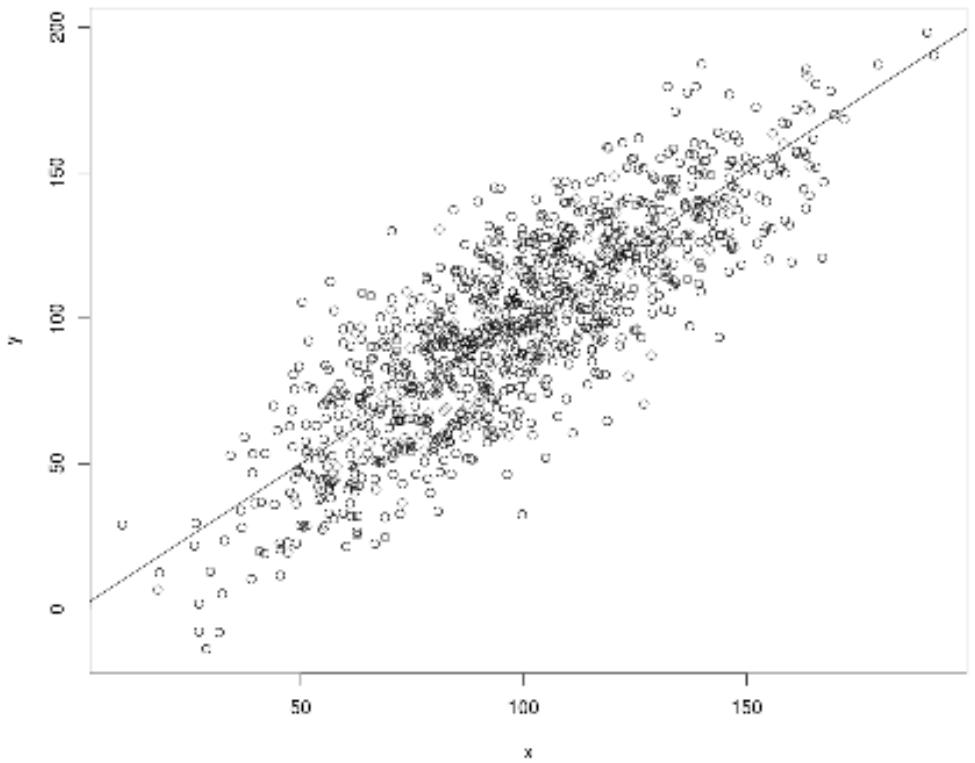


Рис. : Положительная корреляция

Корреляции

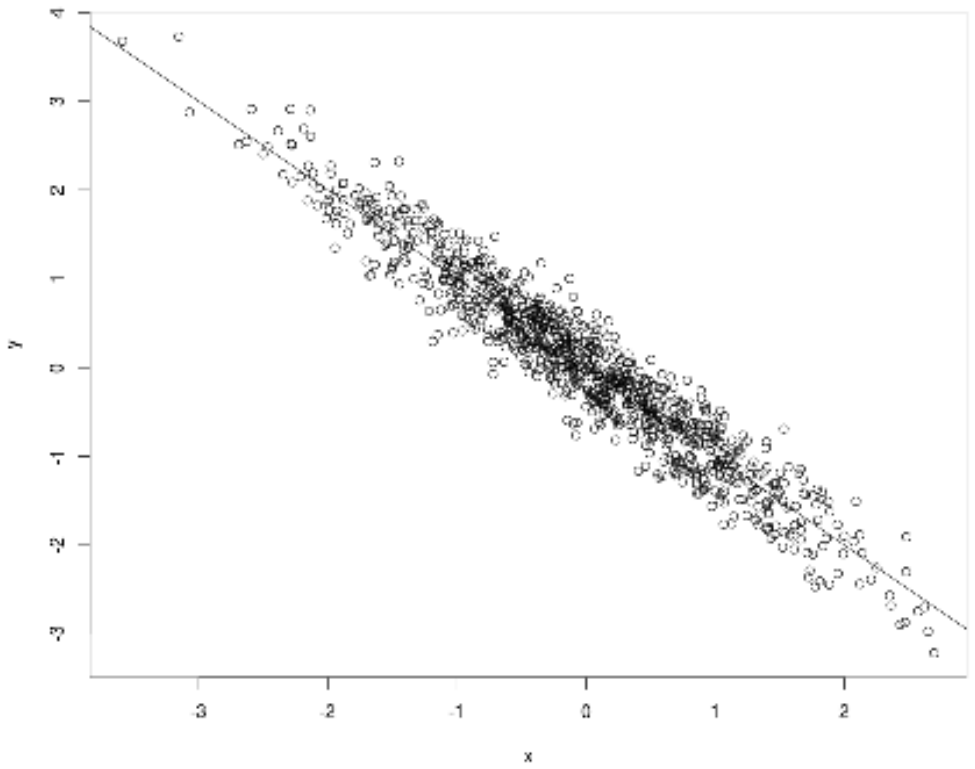


Рис. : Отрицательная корреляция

Корреляции

- Как коррелируют уровень “плохого” холестерина (LDL) и риск сердечно-сосудистых осложнений?
- Как коррелируют уровень “хорошего” холестерина (HDL) и риск сердечно-сосудистых осложнений?
- В Дубаи устанавливают камеры видеонаблюдения для профилактики уличных преступлений (например, карманных краж). Какова корреляция между количеством видеокамер и количеством краж (по документам полиции)?
- После каждого пожара страховая служба сохраняет данные: количество пожарных, прибывших на пожар, и ущерб от пожара. Какова корреляция между двумя этими величинами?

False Positivity

Таблица : HIV test

	Test is negative	Test is positive
Patient is HIV-negative	98%	1.5%
Patient is HIV-positive	0%	0.5%

False-positive rate – conditional probability that patient's test is positive IF he is negative.