

Семинар 4: основы статистики!

Задача 1

Коллекционер Настя собрала целых 10 наблюдений и записала их в табличку. Теперь Настя хочет стать аналитиком и проанализировать таблицу. Помогите ей.

имя	пол	возраст	вес
Кхал	м	14	80
Санса	ж	16	40
Мелисандра	ж	20	40
Эддард	м	20	80
Сандор	м	14	80
Миссаедея	ж	25	40
Якен	м	30	80
Теон	ж	23	40
Тирион	м	22	80
Станис	м	16	440

- а) Что такое непрерывная переменная? Что такое категориальная переменная? Какие переменные в табличке относятся к непрерывным? Какие к категориальным? Приведите ещё примеров непрерывных и категориальных переменных!
- б) Найдите долю мужчин и женщин в выборке. Постройте для пола гистограмму.
- в) Найдите средний возраст и медианный возраст. Что означают эти числа. В чём они измеряются?
- г) Найдите дисперсию возраста. В чём измеряются эта величина? Зачем обычно ищут среднее квадратическое отклонение? Найдите его.
- д) Постройте гистограмму для возраста. Считайте, что ширина одного столбца — 5 лет. Если человек попадает на правую границу отрезка, он попадает в текущий столбец. Изобразите на гистограмме среднее, медиану. Как бы вы нарисовали на гистограмме стандартное отклонение?
- е) Что такое выброс? Есть ли выбросы в возрасте? Есть ли выбросы в весе? Как выглядит выброс на гистограмме? Найдите средний вес и медианный вес. Чем медиана в данном случае лучше, чем среднее?
- ж) Чувствительна ли дисперсия к выбросам?
- з) Что такое мода? Почему использовать её для непрерывных переменных не очень хорошая идея? Найдите моду для имени, пола и возраста. (уточнить что это дискретная мода и тп)
- и) Что такое квантиль? Предложите способ, борьбы с выбросами, основанный на знании того что такое квантиль...

Ещё задачи!

Тут находится несколько задачек, о которых вам нужно подумать самостоятельно, в домашних условиях, за чашкой чая. Возможно, что похожие задачи попадутся вам на самостоятельной работе. Рискнете проверить?

Задача 2

Имеется пять чисел: x , 9, 5, 4, 7. При каком значении x медиана будет равна среднему? А можно ли поставить такие цифры в условии задачи, чтобы x не существовал?

Задача 3

Измерен рост 25 человек. Средний рост оказался равным 160 см. Медиана оказалась равной 155 см. Машин рост в 163 см был ошибочно внесен как 173 см. Как изменится медиана и среднее после исправления ошибки? А как могут измениться медиана и среднее, если рост Маши равен 153?

Задача 4

Деканат утверждает, что если студента N перевести из группы A в группу B , то средний рейтинг каждой группы возрастет. Возможно ли такое?

Задача 5

Иногда в качестве меры разброса используют размах. Находят максимальное значение в выборке, минимальное значение выборке, а после вычитают из максимума минимум. Как думаете, такая мера чувствительна к выбросам? Предложите способ сделать её устойчивой к ним.