

Школа лингвистики, 2019-20 уч. год**Теория вероятностей****Задачи по статистике. Выборки. (06.04.2020)***И. В. Щуров, Д. А. Филимонов, Р. Я. Будылин*

Задача 1. Американский журнал «Литературное обозрение» провел опрос относительно исхода президентских выборов в США в 1936 году. Кандидатами на этих выборах были Ф.Д.Рузвельт и А.М.Ландон. В качестве генеральной совокупности редакция журнала использовала телефонные книги. Отбрав случайно 4 миллиона адресов, она разослала по всей стране открытки с вопросом об отношении к кандидатам в президенты. Затратив большую сумму на рассылку и обработку открыток, журнал объявил, что на предстоящих выборах президентом США с большим перевесом будет избран Ландон. Результат выборов оказался противоположным этому прогнозу. В чем они ошиблись?

Определение 1. Эмпирическое распределение выборки $\{x_1, x_2, \dots, x_N\}$ это функция $F(t)$, такая что, $F(t)$ равно доле элементов выборки, не превосходящих t .

Нам потребуется определить α -квантиль — это число, отсекающее часть выборки с самыми маленькими значениями, составляющими долю α от всей выборки.

Определение 2. α -квантилью функции распределения F называется число x , такое что

$$P(X \leq x) \geq \alpha,$$

$$P(X > x) \geq 1 - \alpha.$$

Квантиль для выборки - это оценка квантили истинного распределения. Существует множество вариантов её нахождения, мы будем использовать следующий (квантиль эмпирического распределения с усреднением в точках разрыва):

Определение 3. Рассмотрим вариационный ряд (упорядоченную по возрастанию выборку) $\{x_1, x_2, \dots, x_N\}$. Пусть ищется α -квантиль x_α . Обозначим через K нижнее целое от αN (то есть максимальное целое число, не превосходящее αN , обозначается $\lfloor \alpha N \rfloor$).

Тогда

- при αN — не целом, $x_\alpha = x_{K+1}$
- при αN — целом, $x_\alpha = (x_K + x_{K+1})/2$

Выборочная дисперсия и выборочное среднее - это математическое ожидание и дисперсия эмпирического распределения.

Задача 2. Пусть дана выборка $\{2, 4, 1, 3, 2, 4\}$.

- найти медиану и первый и третий квартили;
- найти 0.3-квантиль и 0.9-квантиль;
- изобразить на графике выборочную функцию распределения.
- Найти среднее и дисперсию выборки;
- Построить график зависимости α -квантиля данной выборки от α .

Задача 3. Пусть дана выборка $\{-5, -3, 0, 4, 4.1, 4.2, 3.9, 4.4, 3.7, 4\}$.

- (а) найти медиану и первый и третий квартили;
- (б) найти 0.3-квантиль и 0.9-квантиль;
- (с) изобразить на графике выборочную функцию распределения.
- (d) Найти среднее и дисперсию выборки;

Задача 4. Из двух школ — школы A и школы B — отобрали лучших учеников (учеников с более высоким баллом, т.е. суммой отметок по всем предметом) и собрали в школу 1, остальных учеников отправили в школу 2. Назовём рейтингом школы средний балл его учеников. Правда ли, что средний балл учеников школы 1 больше среднего балла учеников школы номер 2? Если да, то докажите, если нет — приведите пример.

Задача 5. Роберт Мулдун (Robert Muldoon), премьер-министр Новой Зеландии в 1975-1984 годах, как-то ответил на вопрос журналиста о людях, эмигрирующих из Новой Зеландии в Австралию: «эти эмигранты повышают средний уровень IQ в обеих странах». В чём смысл высказывания?