

Школа лингвистики, 2016-17 уч. год**Теория вероятностей****Задачи по статистике. Выборки. (17.03.2017)***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 $\{x_0, x_2, \dots, x_{N-1}\}$ 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 P называется число x , такое что

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

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

Квантиль для выборки - это квантиль эмпирического (выборочного) распределения. Существует следующая процедура для ее нахождения:

Определение 3. Рассмотрим выборку $\{x_0, x_2, \dots, x_{N-1}\}$. (Мы начинаем нумерацию с нуля, в выборке N элементов.) Перенумераем элементы в выборке по возрастанию. Обозначим то, что получится, через $\{v_0, v_1, \dots, v_{N-1}\}$. Пусть α — доля (число от 0 до 1). Обозначим через K нижнее целое от $\alpha \cdot (N - 1)$ (то есть максимальное целое число, не превосходящее $\alpha \cdot (N - 1)$).

Тогда

- при $K + 1 < \alpha N$, α -квантиль = v_{K+1}
- при $K + 1 = \alpha N$, α -квантиль = $(v_k + v_{K+1})/2$
- при $K + 1 > \alpha N$, α -квантиль = v_K .

Выборочная дисперсия и выборочное среднее - это мат.ожидание и дисперсия эмпирического распределения. Выборочное среднее - это просто среднее арифметическое всех чисел выборки.

Определение 4. Средним (средним арифметическим) k чисел называется сумма этих чисел, делённая на k .

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

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

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

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

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?