Школа лингвистики, 2020-21 уч. год
Теория вероятностей
Задачи по статистике. Выборки. (16.03.2021)
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. Отобрав 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 $\{x_1, x_2, \ldots, x_N\}$ 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 P 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) $\{x_1, x_2, \ldots, x_N\}$. Пусть ищется α -квантиль x_α. Обозначим через K верхнее целое от αN (то есть минимальное целое число, которое не меньше αN; обозначается $\lceil \alpha N \rceil$; ещё называется "потолок"от числа).

Тогда

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

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

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

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

(e) Построить график зависимости α-квантиля данной выборки от α.
(f) Построить «ящик с усами».

Задача 3. Пусть дана выборка $\{-5, -3, 0, 4, 4.1, 4.2, 3.9, 4.4, 3.7, 4\}$.
(a) найти медиану и первый и третий квартили;
(b) найти 0.3-квантиль и 0.9-квантиль;
(c) изобразить на графике выборочную функцию распределения.
(d) Найти среднее и дисперсию выборки;
(e) Построить «ящик с усами».

Задача 4. Для выборки $\{-20, -3, -2, -1, 0, 1, 2, 4, 12\}$ построить «ящик с усами» и найти выбросы и сильно отклоняющиеся значения.

Задача 5. 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 6. 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?