

1 Условие

В городе Н. правительство решило начать борьбу с превышениями скорости автомобилей. Для выбора стратегии борьбы оно сначала решило провести исследования касательно того, влияет ли используемый водителем автомобиль на среднюю скорость передвижения.

Для этого было сформировано 3 выборки по 20 человек, в каждой из которой людям выдали одинаковые автомобили марок Mitsubishi, Audi и BMW, соответственно. В течение месяца замерялась средняя скорость каждого из автомобилей (см. файл).

Каждая из пар групп была проверена двувывборочным критерием на равенство распределений, также была проведена поправка на множественность гипотез.

Требуется:

1. Описать, в чём недостаток подхода правительства.
2. Предложить метод для более корректного решения задачи.
3. Записать формальное условие задачи.
4. Решить задачу аналитически (**все аналитические выкладки должны быть описаны**)

2 Недостаток подхода

Видно, что данные по Mitsubishi имеют довольно большой разброс в сравнении с другими автомобилями. Правительство использует не совсем релевантный критерий в данном случае, т.к. о распределении ничего сказать нельзя, а в него входят дисперсия и, возможно, другие параметры.

3 Метод

Воспользовавшись критерием ж, убедились, что данные распределены нормально. Наблюдения независимы, объемы выборок одинаковы и равны 20. Все это указывает на то, что верным методом является однофакторный дисперсионный анализ, в частности критерий Фишера.

4 Формальное условие

Выборки: $X^N = X_1^{20} \cup X_2^{20} \cup X_3^{20}$

Нулевая гипотеза: $H_0 : \alpha_1 = \dots = \alpha_3$

Альтернатива: $H_1 : H_0$ неверна

Статистика: $F(X^N) = \frac{SS_{bg}/(K-1)}{SS_{wg}/(N-K)}$, где

$SS_{bg} = \sum_{k=1}^K n_k(\bar{X}_k - \bar{X})^2$, $SS_{wg} = \sum_{k=1}^K \sum_{i=1}^{n_k} (X_{ki} - \bar{X}_k)^2$

Нулевое распределение: $F(K-1, N-1)$, $K=3$, $N=60$

5 Решение

Выше приведены статистики, давайте просто посчитаем их и посмотрим на p-value.

$$SS_{bg} = 1394,6; \quad SS_{wg} = 25126,4 \quad S(X^N) = 1.6$$

Значит,

$$p - value = 0.2.$$

Т.о., данные гипотезе не противоречат, а значит средняя скорость мало зависит от марки представленных автомобилей.