

Школа лингвистики, 2017-18 уч. год**Теория вероятностей****ANOVA и мощность (15.05.2018)***И. В. Щуров, Д. А. Филимонов, Р. Я. Будылин*

Задача 1. Как отличается длина "р" у носителей русского языка из разных городов? Для ответа на этот вопрос были опрошены люди в нескольких городах и подсчитаны средние значения (в условных единицах) и оценки S . Вычислить значение F -статистики и, воспользовавшись дисперсионным анализом (ANOVA), выяснить, имеются ли различия в длине произношения в этих группах. В качестве уровня значимости взять $\alpha = 0.15$. Произвести попарное сравнение с помощью критерия Стьюдента (не забыть про поправку Бонферрони!)

	группа 1	группа 2	группа 3
(a) \bar{x}_j	5.2	6.1	5.7
S_j	0.6	0.4	0.7
N_j	10	10	10
	группа 1	группа 2	группа 3
(b) \bar{x}_j	5.2	8	5.7
S_j	0.6	0.4	0.7
N_j	20	20	20

Задача 2. Вы планируете эксперимент, в котором собираетесь измерять время реакции (его можно считать нормально распределённым). Вы пользуетесь предыдущими исследованиями, которые показали, что для разных случаев можно считать, что среднее отклонение оказывается всегда одинаковым, порядка 50мс. В конкретной работе вы проверяете, совпадает ли время реакции с 200мс или превышает его. Рассмотрите случаи уровня значимости 0.01 и 0.05.

- Какую разницу в значениях вы сможете уверенно отличать с вероятностью 80%, при условии, что у вас 25 человек в выборке? А с вероятностью 90%? А что если у вас 100 человек?
- Какая мощность у критерия если вы хотите, чтобы он видел отличия на 20мс, а у вас всего 25 человек? А если у вас 100 человек? А что если достаточно видеть различия в 50мс?
- Сколько нужно опросить для того, чтобы с вероятностью 80% отличать разницу в 25мс? А разницу в 10мс? А что если нужна вероятность 90%?