**Курс «Теория вероятностей и статистика в Машинном Обучении»**

**Лабораторная работа №3:** Меры связи случайных величин. Интервальные оценки параметров

1. Найти на сайте РосСтат <https://rosstat.gov.ru/statistic> (или в других источниках) предположительно зависимые данные (например: «доля людей в возрасте 70+» и «приверженность жителей тех к ЗОЖ» по регионам), посчитать коэффициент корреляции, оценить меру связи.

Построить гистограммы рассеяния, линию аппроксимации и посчитать величину достоверности аппроксимации R^2.

1. Найти датасет с объемом ~100 наблюдений (и более), сделать выборку =~ 1/3 наблюдений 1) случайным образом,

2) сделать стратифицированную выборку (разбив выборку на группы по смысловому содержанию, например, для регионов стратами могут быть федеральные округа).

1. Посчитать

1) среднее значение по выборкам из п.2 (2.1)случайной и 2.2)стратифицированной),

2) доверительный интервал для среднего (на уровне доверия 90%, 95%, 99%).

Сравнить среднее генеральной выборки п.2 с полученными в выборках 2.1) и 2.2) и с границами доверительных интервалов из п.3.2).