

Заповеди:

1. Be good, drink milk and **think of Russia** (с)
2. Обязательно фиксируйте зерно генератора случайных чисел в экспериментах. При перезапуске кода значения не должны меняться :)
3. Закомитьте задание в виде .ipynb файла на гитхаб <https://classroom.github.com/a/3KL2oSuX>.
4. Вверху файла подпишите фамилию, имя и группу.
5. Дедлайн: 7 июня 21:00 без штрафа, 8 июня 21:00 со штрафом 50%.
6. Ещё потребуется загрузить задачи в контекст для проверки на плагиат ;)

Задачи

1. Парадокс инспектора.

Автобусы отходят от автостанции с 8:00 до 20:00. Первый автобус отходит ровно в 8:00. Затем интервалы между автобусами случайны, независимы и равновероятно равны либо 5-и, либо 10-и минутам. Будем считать, что за минуту на автостанцию приходит ровно один пассажир, и все пассажиры едут ближайшим автобусом.

Проведите 10^4 экспериментов и с их помощью:

- а) [10] Постройте гистограмму количества автобусов, отошедших от автостанции за сутки. Похоже ли визуально распределение на нормальное?
- б) [10] Инспектор Тимон выбирает равновероятно один из всех автобусов отошедших от автостанции за сутки. Постройте гистограмму числа пассажиров на этом автобусе. Оцените математическое ожидание и дисперсию этого числа.
- в) [10] Инспектор Пумба приходит на автостанцию в случайный момент времени, равномерный от 8:00 до 20:00 и садится в первый пришедший автобус. Постройте гистограмму числа пассажиров на этом автобусе. Оцените математическое ожидание и дисперсию этого числа.
- г) [10] Как изменятся ответы на эти вопросы, если время между автобусами будет экспоненциально со средним в 10 минут?

2. Парадокс Хуана Мануэля Родригеса Паррондо.

У Атоса, Портоса и Арамиса по 1000 франков.

Атос постоянно ходит в казино А, где каждый раз выигрывает один франк с вероятностью 0.49 и проигрывает один франк с вероятностью 0.51.

Портос ходит в казино Б, где ситуация интереснее :) Если богатство посетителя кратно трём, то посетитель выигрывает франк с вероятностью 0.09 и проигрывает один франк с вероятностью 0.91. Если богатство посетителя не кратно трём, то посетитель выигрывает франк с вероятностью 0.74 и проигрывает один франк с вероятностью 0.26.

Арамис каждый раз выбирает казино А или казино Б равновероятно.

Проведите 10^4 симуляций эволюции благосостояния Атоса, Портоса и Арамиса. Каждая симуляция предусматривает 1000 посещений казино. Все три игрока все деньги носят с собой и ни на что не тратят :)

- [10] Постройте на графике 10 случайных траекторий изменений богатства Атоса, по горизонтали —номер посещения казино, по вертикали —богатство. На том же графике постройте усреднённую по всем экспериментам траекторию изменения богатства.
- [10] Постройте на графике 10 случайных траекторий изменений богатства Портоса, по горизонтали —номер посещения казино, по вертикали —богатство. На том же графике постройте усреднённую по всем экспериментам траекторию изменения богатства.
- [10] Постройте на графике 10 случайных траекторий изменений богатства Арамиса, по горизонтали —номер посещения казино, по вертикали —богатство. На том же графике постройте усреднённую по всем экспериментам траекторию изменения богатства.

3. Парадокс Берксона.

Предположим, что результаты ЕГЭ школьников по русскому и математике независимы и хорошо аппроксимируются нормальным распределением с ожиданием 60 и стандартным отклонением 10.

УШЭ (Урюпинская Школа Экономики) ныне престижна и забирает себе всех школьников набравших более n баллов в сумме по русскому и математике.

Для каждого n от 80 до 160 с шагом 5 случайно создайте 10^4 школьников и поделите их на прошедших и не прошедших в УШЭ.

- [10] Постройте график числа прошедших в УШЭ в зависимости от n .
- [10] Постройте график выборочной корреляции между результатами по русскому и математике в зависимости от n среди прошедших в УШЭ.
- [10] Постройте график выборочной корреляции между результатами по русскому и математике в зависимости от n среди не прошедших в УШЭ.

Тем, кто хочет получить бонусные балы и смыть ими тяжкие грехи прошлого...:)

4. Парадокс Штайна.

Ниф-Ниф, Наф-Наф и Нуф-Нуф качают пресс на карантине, чтобы подготовиться к встрече Волка :) Количества подъямов туловища в i -й день у поросят обозначим X_i , Y_i и Z_i . Эти величины независимы и хорошо аппроксимируются нормальным распределением $X_i \sim \mathcal{N}(60, 100)$, $Y_i \sim \mathcal{N}(70, 100)$, $Z_i \sim \mathcal{N}(80, 100)$.

Карантин длится 100 дней. Волк не знает математических ожиданий (60, 70, 80), но знает дисперсии. Волку удаётся подсмотреть, сколько раз поросята поднимают свои туловища.

Проведите 10^4 симуляций карантина :)

- [10] Для каждой симуляции помогите Волку посчитать оценки $\hat{\mu}_x$, $\hat{\mu}_y$, $\hat{\mu}_z$ методом максимального правдоподобия. Постройте гистограмму каждой из оценок и обозначьте на них истинные значения параметров.
- [3] Постройте гистограмму суммарной квадратичной ошибки, $S = (\hat{\mu}_x - \mu_x)^2 + (\hat{\mu}_y - \mu_y)^2 + (\hat{\mu}_z - \mu_z)^2$. Оцените математическое ожидание суммарной квадратичной ошибки.
- [5] Отложите оценки Волка $\hat{\mu}_x$ и $\hat{\mu}_y$ на диаграмме рассеяния. Найдите их выборочную корреляцию.

Обозначим вектор трёх оценок Волка одной буквой $\hat{\mu}$. Хитрый Лис тоже охотится на Трёх Поросят. Он берёт вектор оценок Волка, домножает его на хитрый множитель, и получает вектор оценок Хитрого Лиса:

$$\tilde{\mu} = \left(1 - \frac{1}{\|\hat{\mu}\|^2}\right) \hat{\mu}$$

- [10] Выполните предыдущие три пункта для оценок Хитрого Лиса.
- [2] Кто точнее оценивает накачанность Ниф-Нифа? Кто точнее оценивает вектор накачанности Трёх Поросят?

Почитать больше:

- <https://towardsdatascience.com/the-inspection-paradox-is-everywhere-2ef1c2e9d709>
- https://en.wikipedia.org/wiki/Parrondo%27s_paradox
- https://en.wikipedia.org/wiki/Berkson%27s_paradox
- <http://www.statslab.cam.ac.uk/~rjs57/SteinParadox.pdf>