

Время выполнения заданий — 3 часа (11:00 — 14:00).

Экзамен включает 3 обязательных задания для получения максимальной оценки (100 баллов) и 30 бонусных баллов.

До 14:00 необходимо сдать задание на Github Classroom, <https://classroom.github.com/a/НKkCz-gC>, в формате .ipynb, указав имя, фамилию и группу в названии файла, и загрузить свой код по каждой задаче на Яндекс.Контеcт, <https://contest.yandex.ru/contest/18973/enter>, для проверки на плагиат.

Уточнения условия, если они будут, появятся по ссылке <https://www.overleaf.com/read/kbwtcskdtbbv>, в телеграм канале @room112, и в чате «Науки о данных».

1. Задача про блуждание (35 баллов)

Шерлок Холмс расследует дело о пропаже бюллетеней на избирательном участке. Ватсон заметил, что всегда во время своих размышлений Холмс ходит ровно по прямой. Предположим, что в начале движения Шерлок находится в точке с координатой 0. В каждый момент времени его координата изменяется на Δx метров, где Δx — случайная величина, имеющая стандартное нормальное распределение. Эта величина прибавляется к текущей координате после каждого шага и может принимать положительные и отрицательные значения.

- 1) Постройте график изменения положения Шерлока (его координаты на прямой) за 1000 шагов. По горизонтали — количество пройденных шагов, по вертикали — положение Шерлока. (5 баллов)
- 2) Постройте для траектории положения Шерлока из предыдущего пункта график изменения выборочного среднего и дисперсии для Δx . По горизонтали — количество шагов, по вертикали — значения выборочного среднего и выборочной дисперсии. (10 баллов)
- 3) Проведите 10^3 экспериментов по 500 шагов. Продемонстрируйте на одном графике 15 случайных траекторий из проведенных экспериментов. (10 баллов)
- 4) Добавьте на график из предыдущего пункта усредненную по всем экспериментам траекторию выборочной дисперсии для каждого шага. (10 баллов)

2. Парковки от мэра (40 баллов + 15 бонусных баллов)

Перед выборами в центре города открыли неразмеченную парковку длиной 100 метров. Каждый приезжающий водитель случайным образом выбирает место длиной 2 метра, и если место оказалось свободным, встает на парковку до конца дня. Если он не помещается в выбранный отрезок, то водитель в нервном состоянии уезжает с парковки. За один день на парковку пытается встать 1000 проезжающих мимо водителей. Для выполнения всех пунктов заданий проведите 2000 симуляций.

Hint: Для простоты будем считать, что выбор точки на парковке - это центр машины.

- 1) В среднем сколько машин оказываются припаркованы к концу дня? (20 баллов)
- 2) Постройте 95%-ый доверительный интервал для ожидаемого количества машин припаркованных к концу дня. (10 баллов)
- 3) Постройте 95%-ый доверительный интервал для ожидаемого количества свободного места на парковке к концу дня (в метрах). (10 баллов)
- 4) После возникшего хаоса на парковку добавили регулярную разметку для мест длиной 2 метра, но водители разделились на два лагеря — тех, кто соблюдает и не соблюдает правила разметки. С вероятностью p приезжает честный водитель и занимает случайное свободное место строго по разметке. Водители второго лагеря паркуются, как раньше. Нарисуйте график, где по оси ординат — процент честных водителей, а по оси абсцисс — процент свободного места на парковке к концу дня для вероятностей p равным от 0 до 1 с шагом в 5%. Для каждого порога вероятности проведите по 500 симуляций. (15 бонусных баллов)

3. Правдоподобие счастья (25 баллов + 15 бонусных баллов)

Загрузите набор данных по измерению уровня счастья в разных странах мира за 2019 год (**data.csv** в репозитории экзамена). Полное описание датасета доступно по ссылке на Kaggle: <https://www.kaggle.com/unsdsn/world-happiness?select=2019.csv>.

- 1) Постройте на графике гистограмму для уровня счастья (переменная Score) и ее функции плотности. Является ли распределение нормальным? (5 баллов)
- 2) Пусть уровень счастья – целевая вещественная переменная y , все y_i независимы и нормальны $N(\mu, \sigma^2)$, где $\mu = a + b \cdot x_i$. Выведите функцию правдоподобия и задачу для оптимизации поиска коэффициентов a , b и σ^2 (15 бонусных баллов).
- 3) Оцените параметры a , b и σ^2 с помощью метода максимального правдоподобия, если в роли x_i выступает ВВП на душу населения (GDP per capita). (10 баллов)
- 4) Используя оценённые параметры и медианное значение ВВП на душу населения, оцените вероятность того, что уровень счастья будет больше 5. (10 баллов)