

# Прикладная статистика и анализ данных.

## Задание 10.

- Дедлайн **15 мая 02:00**. После дедлайна работы не принимаются кроме случаев наличия уважительной причины.
- Выполненную работу нужно отправить на почту `mipt.stats@yandex.ru`, указав тему письма "[asda] Фамилия Имя - задание 10". Квадратные скобки обязательны. Если письмо дошло, придет ответ от автоответчика.
- Прислать нужно ноутбук и его pdf-версию (без архивов). Названия файлов должны быть такими: `10.N.ipynb` и `10.N.pdf`, где  $N$  — ваш номер из таблицы с оценками.
- Решения, размещенные на каких-либо интернет-ресурсах не принимаются. Кроме того, публикация решения в открытом доступе может быть приравнена к предоставлении возможности списать.
- В каждой задаче не забывайте делать **пояснения и выводы**.
- Все численные вычисления можно производить в Питоне или в R.

1. (1 балл) Случайные величины  $X$ ,  $Y$  и  $Z$  имеют следующее распределение.

	$Y = 0$ $Z = 0$	$Y = 1$ $Z = 0$	$Y = 0$ $Z = 1$	$Y = 1$ $Z = 1$
$X = 0$	0.405	0.045	0.125	0.125
$X = 1$	0.045	0.005	0.125	0.125

Являются ли величины  $X$  и  $Y$

- (a) независимыми;
- (b) условно независимыми по  $Z$ ?

2. (2 балла) При проведении испытания лекарства получены следующие данные.

Количество выздоровевших	Принимали плацебо	Принимали лекарство
Низкое давление	81 из 87	234 из 270
Высокое давление	192 из 263	55 из 80

С помощью метода интервенции оцените причинно-следственный эффект влияния лекарства на выздоровление, если

- (a) Кровяное давление пациентов измерялось до начала испытания, после чего принималось решение о приеме лекарства.
- (b) Кровяное давление пациентов измерялось в конце испытания. При этом известно, что лекарство может оказывать эффект на кровяное давление, которое в свою очередь может оказывать эффект на выздоровление.

3. **(3 балла)** Имеется  $n$  пациентов, которым соответствуют данные  $(X_1, Y_1), \dots, (X_n, Y_n)$ , где  $X_i \in \{0, 1\}$  — факт проведения лечения  $i$ -го пациента, а  $Y_i \in \{0, 1\}$  — исход заболевания. Как известно, в общем случае нельзя оценить величину причинно-следственного эффекта  $\theta$ . Тем не менее, можно получить некоторый интервал на  $\theta$ . Докажите, что существуют числа  $a, b$ , для которых выполнено  $\theta \in [a, b]$ ,  $b - a = 1$ , причем для  $a$  и  $b$  существуют состоятельные оценки. Найдите эти оценки и покажите их состоятельность.

*Указание.* Воспользуйтесь формулой полной вероятности.

4. **(3 балла)** Распределение случайного вектора  $V = (X, Y, Z)$  задается следующим образом. Случайная величина  $X$  имеет распределение  $\mathcal{N}(0, 1)$ . Условное распределение случайной величины  $Y$  при условии  $X = x$  есть  $\mathcal{N}(\alpha x, 1)$ . Условное распределение случайной величины  $Z$  при условии  $X = x, Y = y$  есть  $\mathcal{N}(\beta y + \gamma x, 1)$ .

(a) Нарисуйте граф причинно-следственных связей, соответствующий данной модели. Выпишите плотность вектора  $V$ .

(b) Посчитайте  $\rho$  — корреляцию между  $Y$  и  $Z$ .

*Указание.* Воспользуйтесь аддитивностью параметра сдвига.

(c) Предположим, что величина  $X$  не наблюдаема. Можно ли утверждать, что  $Y$  влечет  $Z$ , если  $\rho \neq 0$ ?

(d) Ответьте на предыдущий вопрос, если известно, что  $\alpha = 0$ .

5. **(4 балла)** В условиях предыдущей задачи посчитайте  $E(Z|Y = y)$  и  $E(Z|Y := y)$ . Сравните результаты между собой.

*Внимание.* Теоретические вычисления могут быть довольно затруднительными. В таком случае можно свести ответ к интегралам, которые посчитать методом Монте-Карло для некоторых значений  $y$  и параметров  $\alpha, \beta, \gamma$ .