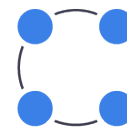


Домашнее задание по многофакторному анализу №1¹.



Максимальное количество баллов равно 7.

1.1 Соглашения

Задание стоит сдавать в виде R-markdown (если вы не знаете, как этим пользоваться, то можно посмотреть короткий туториал по [ссылке](#)). Не забывайте пояснять, что вы делаете.

1.2 Задачи

1. Известно, что креатинин k и мочевины m являются сильно коррелированными величинами, распределения которых близки к нормальным. При этом $\mu_k = 88.5, \sigma_k = 13.25, \mu_m = 5.4, \sigma_m = 1.45$, а коэффициент корреляции $\rho(k, m) = 0.6$.

Смоделируйте выборку S , содержащий информацию о 100 пациентах, у которых замеры эти два показателя. Предположим, что вы хотите построить линейную модель, связывающую их. (Воспользуйтесь пакетом *mvtnorm* чтобы смоделировать соответствующее многомерное нормальное распределение.)

- (а) (1 балл) Как выглядит модель линейной регрессии креатинина k на мочевины m ? Воспользуйтесь функцией *lm*, чтобы оценить параметры в этой модели. Как полученные оценки связаны с выборочными характеристиками (средним, дисперсией, корреляцией) выборки S ?
 - (б) (1 балл) Проверьте гипотезу о нормальном распределении остатков в вашей модели.
 - (с) (1 балл) Добавим в модель еще один признак W , никак не связанный с k и m . Смоделируйте одномерную выборку (с произвольным распределением), и, считая ее одним из признаков в выборке S , добавьте этот признак в вашу модель линейной регрессии. Как изменился коэффициент детерминации в новой модели? Как изменился модифицированный коэффициент детерминации?
2. (2 балла) Смоделируйте такую выборку объема 50 с 6 признаками, чтобы в модели линейной регрессии первого признака на остальные 5 каждый из коэффициентов регрессии в отдельности не значимо отличался от 0, но все параметры в совокупности — значимо. Воспользуйтесь пакетом *glmnet*, чтобы построить для такой модели лассо-оценку с параметром $r = 0.2$. Как изменились при этом оценки коэффициентов линейной регрессии? Придумайте из своей практики набор сильно зависимых показателей, которые могут быть связаны регрессионным соотношением.
 3. (2 балла) У пациента, попавшего в реанимацию с риском сепсиса, измеряют уровень нейтрофилов (Neu) и лимфоцитов (Ly). При определенном заболевании уровень лимфоцитов является нормальной случайной величиной с распределением $N(20; 5)$, а уровень нейтрофилов — $N(80; 5)$. Пусть при нейтрофильно-лимфоцитарном соотношении $NLR < 3$ угроза сепсиса отсутствует, $NLR > 9$ - сепсис неизбежен, при остальных значениях сепсис возникает с вероятностью $p = \frac{NLR-3}{6}$.

Смоделируйте трёхмерную выборку S объема 201, состоящую из наблюдений за (Neu, Ly, Sepsis). Воспользуйтесь пакетом *glmnet* и функцией *glm*, чтобы построить модель логистической регрессии признака Sepsis на Neu и Ly. Какая вероятность для случайной величины Sepsis быть равной 1, если Neu = 90, а Ly = 15? Какую вероятность предсказывает ваша модель (воспользуйтесь функцией *predict*)?

¹v. 1.0, от 8 декабря 2022 г.