

## Уровень каких белков различается в мышинной модели синдрома Дауна

*Проект должен быть представлен в виде отчета в формате .rmd (достаточно загрузить в отдельную ветку вашего приватного репозитория этот файл). Технические требования к компиляции и оформлению отчета сохраняются такими же как и для предыдущих двух.*  
*Дедлайн сдачи задания 24 февраля 23:59. С 20 января и до этого времени можно свободно задавать вопросы и обсуждать возможные варианты решения*

В 2015 году на мышинной модели были проведены исследования как синдром Дауна влияет на изменения уровней различных белков. Данные для проекта можно скачать по следующей ссылке -- <https://archive.ics.uci.edu/ml/datasets/Mice+Protein+Expression#>  
Doi статьи на которую можно ориентироваться -- 10.1371/journal.pone.0119491.

Для этих данных вам необходимо:

1. Сделать описание датасета (5 баллов)
  - сколько всего мышей было в эксперименте
  - какие группы вы можете выделить
  - насколько эти группы сбалансированы
  - какое количество полных наблюдений (речь про NA)
2. Есть ли различия в уровне продукции BDNF\_N в зависимости от класса в эксперименте (10 баллов)
3. Попробовать построить линейную модель, способную предсказать уровень продукции белка ERBB4\_N на основании данных о других белках в эксперименте (15 баллов)
  - провести диагностику полученной линейной модели
  - объяснить, почему это является хорошим/не хорошим решением
4. Сделайте PCA (15 баллов)
  - ординацию
  - постройте графики факторных нагрузок
  - определите, какой процент объясняет каждая компонента
  - постройте трехмерный график для первых 3-х компонент
5. Поиск дифференциальных белков -- творческая часть задания (15 баллов)
  - можно сделать реанализ из статьи, но предупреждаю сразу, что там машинное обучение
  - один из вариантов решения - использование методов направленной

ординации

-- можно использовать limma/DeSeq2 (limma проще с осознания логики, также в limma и DeSeq2 немного по разному работают статистические тесты)  
Дополнительные баллы по накопительной системе за каждую адекватную идею и её реализацию (до 15 баллов).