Низкое соотношение CD4 + и CD8 + Т-лимфоцитов (CD4 / CD8) в периферической крови (даже на антиретровирусной терапии) может быть связано с хроническим воспалением при ВИЧ. Также, возможно, отношение CD4/CD8 может использоваться в качестве маркера хронического заболевания легких ВИЧ-позитивных пациентов. Для проверки возможной взаимосвязи используется датасет по 190 испытуемым, включенным в исследование EXHALE.

Описание переменных приведено в дополнительном файле.

Для анализа отношение CD4 / CD8 предлагается разбить на три категории: меньше 0.4; 0.4 - 1; больше 1 *(Подумайте, как от этого поменяется тип переменной, как это повлияет на выбор статистических тестов. В каком случае (с или без преобразования) сохранится больше информации?)*

показатель тяжести эмфиземы на две категории: больше 10% или меньше или равно 10%. *(Подумайте, изменится ли в этом случае тип переменной?)*

**Задание 1** (рекомендуется выполнять в группе от 2х человек).

Эксплораторный анализ данных по группам CD4 / CD8.

Определить связь между соотношением CD4 / CD8 и рентгенографической эмфиземой.

*(Можно посмотреть корреляции этих параметров, с предложенными ранее преобразованиями и без)*

Проверить предположение, что низкое соотношение CD4 / CD8 может быть связано с эмфиземой. (*Попробуйте провести статистические тесты для корреляций, держите в уме тип ваших данных. Обратите внимание на вывод для непараметрического теста, если будете его проводить. Подумайте, каким еще статистическим тестом можно воспользоваться, держа в уме количество ваших групп и тип данных)*

Определить другие факторы, влияющие на тяжесть эмфиземы (age, sex, ethnicity, BMI, tobacco, chronic disease, inhalation, iv drug use, CD4 cell count, viral load, nadir CD4, soluble CD14). *(Возможно, есть смысл начать с общего визуального обзора вклада этих переменных посредством PCA? Как поступить с qualitative data?)*

(\*) Построить и проанализировать множественную логистическую регрессионную модель тяжести эмфиземы. *(Подумайте, почему модель должна быть логистической)*

Задание 2 (рекомендуется выполнять в группе от 2х человек).

Эксплораторный графический анализ данных по группам CD4 / CD8.

Определить связь между соотношением CD4 / CD8 и другими маркерами ХОБЛ и легочной функции - обструкцией воздушного потока (Airflow Obstruction),, объемом форсированного выдоха за одну секунду (FEV1) и диффузионной способностью (DLCO). *(Попробуйте начать с корреляций)*

(\*) Проверить предположение, что низкое соотношение CD4 / CD8 может быть связано с функцией легких при ХОБЛ, независимо от других параметров, т.е. построить для каждого варианта легочной дисфункции множественную регрессионную модель. Дополнительные параметры: age, sex, ethnicity, BMI, tobacco, chronic disease, inhalation, iv drug use, CD4 cell count, viral load, nadir CD4, soluble CD14. *(Почему на этот раз мы берем обычную, а не логистическую модель? Попробуйте определить, каких дополнительных параметров (предикторов) более-менее достаточно для построения моделей)*

*Mode*

*Geometric mean*