«Применение новой терапии у пациентов»

**Доступно:**

1. Файл (БД с моно\_full) с информацией о пациентах со среднетяжелым и тяжелым течением болезни, которым вводились препараты:

- номер истории болезни (CaseID),

- даты госпитализации (Start, End)

- пол,

- возраст,

- тяжесть ((крайне тяжелое или тяжёлое) / среднетяжелое)

- исходы: умер / выписан / переведён

- статус вакцинации от ковид.

2. Файл (БД без моно\_full) с информацией о пациентах с среднетяжелым и тяжелым течением болезни, которым не вводились новые препараты - аналогичные данные (см выше).

3. Файл с результатами исследований показатель\_D

4. Файл с результатами исследований показатель\_F

**Необходимый минимум (+ любой дополнительный анализ по желанию ☺ ):**

**1. Провести предварительную обработку данных:**

* Загрузить данные в Dataframe, если у какой-то записи из файлов с информацией о пациентах отсутствует CaseID, присвоить (учесть, что он не должен встречаться в других файлах)
* Проверить уникальность записей. Если есть повторяющиеся, оставлять последнее пребывание с стационаре.
* Объединить данные из таблиц
* Проверить, есть ли пустые значения в данных по показателям
* Создать dataframe только с данными пациентов, у которых известен исход (жив/ мертв).

2. Привести данные описательной статистики для группы пациентов, которым вводились и не вводились препараты (% м и ж, средний возраст, % исходов в группах). Визуализировать результаты.

3. Создать пары пациентов из 2ух групп по правилу

- один пол

- возраст +/- 3 года

- уровень показатель\_D +/- 10%

- уровень показатель\_F +/- 10%

- схожая тяжесть заболевания (крайне тяжелые и тяжелые – одна группа, среднетяжелые – вторая группа)

Как изменятся группы, если добавить дополнительный критерий - одинаковый статус вакцинации

4. Оценить сопоставимость групп по указанным факторам с приведением показателя отсутствия статистической значимости различий

5. Сравнить полученные группы по исходам, визуализировать:

- жив/умер

- для живых - быстрее ли выздоравливали (=сроки госпитализации)

5. Оценить влияние отдельных факторов на исходы, визуализировать:

-возраст

-факт вакцинации

-пол

- тяжесть

- уровень показатель\_D

- уровень показатель\_F