Вам предоставлен файл «trauma.xlsx», который содержит следующие данные:

• id – номер пациента в списке

• Name – имя фамилия пациента

• Sex – пол пациента

• Age – возраст пациента (лет)

• Height – рост пациента (дюймы1)

• Weight – масса тела пациента (фунты2)

• SBP – систолическое артериальное давление при поступлении (мм рт.ст.) • DBP – диастолическое артериальное давление при поступлении (мм рт.ст.) • FOUR – балл по шкале комы FOUR при поступлении

• GSC – балл по шкале комы Глазго при поступлении

• Hb – уровень гемоглобина при поступлении (г/дл)

• Death – летальный исход в течение 24 часов (1 – наступил, 0 – нет)

Это полностью сгенерированный учебный датасет, но давайте представим, что это данные пациентов, госпитализированных в больницу после получения черепно-мозговой травмы вследствие падения с электросамоката.

Загрузите датасет из файла «trauma.xlsx» и ответьте на следующие вопросы:

1. Дайте описательную статистику для переменных, включённых в датасет. Дополнительно рассчитайте, у какого количества пациентов и в каком проценте случаев у пациентов был снижен уровень гемоглобина? Используйте следующие референтные значения (Мужчины: 13.5–16 г/дл, Женщины: 12–14 г/дл).

2. Рассчитайте индекс массы тела у пациентов (кг / м2). Каков был средний (M (SD)) уровень ИМТ у пациентов, включённых в исследование? Какая доля пациентов имела ожирение (ИМТ > 30)?

3. Как выглядит ROC-кривая для предсказания летального исхода в течение 24 часов по переменной, характеризующей уровень гемоглобина? Постройте график. Чем может быть обусловлена такая форма кривой?

4. Чему равна площадь под ROC-кривой, которую вы построили в вопросе 3? Чему равен 95% двусторонний ДИ для площади под ROC-кривой, которую вы построили в вопросе 3? 5. Проведите ROC-анализ и определите, какое пороговое значение является оптимальным для предсказания летального исхода в течение 24 часов по шкале комы Глазго. Какой чувствительностью и специфичностью обладает данный порог?

6. Какая из количественных переменных в датасете (включая рассчитанный вами ранее ИМТ) обладает наибольшей площадью под ROC-кривой? Как вы можете интерпретировать это знание? Какая количественная переменная имеет наименьшую площадь?

Результаты нужно оформить в виде документа word, pdf или html. Отчёт необходимо сгенерировать с помощью rmarkdown или quarto (аргумент echo = TRUE). Результаты расчётов должны сопровождаться короткими комментариями (1-2 предложения), которые будут понятны врачу.

Уделите внимание оформлению отчёта. Перебарщивать не нужно, но если отчёт будет трудно читаемым (например, если в генерируемый отчёт будут целиком выводиться таблицы на 1000 строк) оценка будет снижена.

1 1 дюйм – 2.54 см

2 1 кг – 2.2 фунта