В этом задании вам предстоит анализировать данные долгосрочного исследования национального образования США (US National Education Longitudinal Study, NELS). Случайную выборку из 1000 наблюдений из этого исследования можно найти в датасете NELS88 из пакета copulaData.

При помощи линейной модели найдите ответ на один из этих вопросов:

А - Что влияет на оценки школьников по математике (Math)?

Б - Что влияет на оценки по естествознанию (Science) учащихся городских (Urban) государственных (Public) школ?

В - Каким образом на оценки учащихся городских (Urban) государственных (Public) школ по чтению (Reading) влияет их пол (Female) и принадлежность к этническому меньшинству (Minority)?

1. Кратко опишите исходные данные. Важно указать объемы выборок. Проверьте, нет ли пропущенных значений. Оцените, подходят ли данные для анализа при помощи линейных моделей и какие могут быть ограничения и сложности.

2. Подберите линейную модель для ответа на выбранный вами вопрос задания.

Проверьте условия применимости. Диагностические графики можно привести в отчете, в разделе Материалы и методы.

Если с моделью что-то не так (условия применимости нарушены) - видоизмените модель, чтобы это исправить.

3. Упростите подобранную модель, если это возможно, и проведите ее диагностику. Все шаги упрощения модели должны быть отражены в коде.

4. Интерпретируйте результаты. Какие предикторы влияют на значение отклика? Насколько хорошо ваша модель описывает исходные данные? Если ваша модель сравнивала дискретные группы, то какие именно группы различаются? Постройте график предсказаний модели.

5. В отчете опишите ход и результаты анализа. Не забудьте включить график предсказаний модели. Какой вывод можно сделать из этой работы?

Отчет должен быть выполнен по форме, напоминающей миниатюрную научную статью (общий объем текста не более 2 страниц). Внимательно прочитайте требования к содержанию разделов отчета и анкету для оценивания - это поможет вам сделать задание качественно.

6. Файл с отчетом в формате .html загрузите на сайт.

Источники

Hofert M., Kojadinovic I., Maechler M., Yan J. (2017) copulaData: Data Sets for Copula Modeling. R package version 0.0-1.

## Требования к отчету

Анализ данных должен быть выполнен в R и оформлен в виде отчета. Отчет должен быть написан при помощи rmarkdown в формате .Rmd и скомпилирован при помощи knitr в файл html. Обратите внимание, отчеты в других форматах приниматься не будут! Пример оформления отчета в html и исходный Rmd файл вы можете найти в архиве report-html.zip.

Отчет должен быть кратким (общий объем текста не более 2 страниц) и должен включать разделы Введение, Материалы и методы, Расчеты, Результаты и обсуждение, Выводы и Список литературы. Содержание этих разделов кратко описано ниже:

* Введение. Поставьте себя на место авторов данных и объясните своими словами, зачем затевалось исследование и почему оно с вашей точки зрения может быть важно. Сформулируйте исследовательский вопрос и гипотезы, которые необходимо проверить. (Кратко, 2-5 предложений)
* Материалы и методы. Объясните, что и зачем делали во время исследования. Опишите дизайн исследования, особенности данных, порядок применения статистических методов. Например, здесь стоит сказать, какие условия применимости методов вы проверяли и каким способом, что показали результаты проверки и как они повлияли на дальнейший анализ. Здесь же можно привести диагностические графики. Обязательно должны быть ссылки на литературу с описанием метода, на использованные статистические пакеты (стиль оформления ссылок может быть любым). Этот пункт вполне может занять пару-тройку абзацев.
* Расчеты. Внутри этого раздела должны быть чанки c R кодом. Код чанков должен быть виден после компиляции отчета. Важно, чтобы в коде были все пункты, необходимые для воспроизведения хода анализа (включая чтение данных). Код должен содержать осмысленные комментарии.
* Результаты и обсуждение. Визуализируйте результаты, опишите и поясните их, по возможности, кратко. Описание результатов статистических тестов можно давать в тексте скобках или в виде таблиц. Смысл рисунков и таблиц должен быть понятен из текста подписей без обращения к основному тексту. В основном тексте должны быть ссылки на все таблицы и все рисунки, присутствующие в работе.
* Выводы. Сформулируйте выводы вашего исследования.
* Список литературы. Приведите список цитированных источников, включая интернет-ресурсы. Вы можете использовать любой стиль оформления списка источников, но список должен включать все необходимые библиографические поля (автор, год, название работы, издание или интернет-ресурс).