Пациентам, больным раком желудка, бронхов, кишечника, яичников или раком груди, давали в качестве поддерживающей терапии аскорбат (Cameron, Paulig, 1976, данные из кн.Hand et al. 1994). Сравните время выживания пациентов (Survival) с раком груди и другими видами рака (с каждым в отдельности).

Исходные данные для анализа находятся в файле "cancer.xlsx".

Публикуется с разрешения Taylor and Francis Group LLC Books из работы "A Handbook of Small Data Sets" Hand, D. J., F. Daly, A. D. Lunn, K, J. McConway, E. Ostrowski, 1994; разрешение получено через Copyright Clearance Center, Inc.

0. Внимательно прочитайте задание и исследуйте данные, чтобы корректно сформулировать вопрос/гипотезу, которую вы сможете проверить в этом проекте.

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

Визуализируйте характер распределения величины, находящейся в фокусе исследования (Среднее время выживания). Графики, визуализирующие распределения, должны быть включены в отчет.

Поможет ли трансформация данных (например, извлечение квадратного корня из зависимой переменной) преодолеть хотя бы некоторые сложности анализа?

2. Сравните среднее время выживания пациентов с раком груди и с другими видами рака при помощи t-критерия, сделав поправку на множественное тестирование.

3. Постройте график с доверительными интервалами (столбчатый или точечный).

4. В отчете опишите ход и результаты анализа. В отчет обязательно включите график, визуализирующий различия между группами.

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.

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

Источники

Cameron, E. and Pauling, L., 1978. Supplemental ascorbate in the supportive treatment of cancer: reevaluation of prolongation of survival times in terminal human cancer. Proceedings of the National Academy of Sciences, 75(9), pp.4538-4542.

Hand, D.J., Daly, F., McConway, K., Lunn, D. and Ostrowski, E., 1994. A handbook of small data sets. Springer. 458 p.

## 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).