

Тестирование гипотез. Домашнее задание, вариант 1.

В одном из крупных американских медицинских центров было проведено исследование уровня холестерина в крови пациентов в различные сроки после сердечного приступа и у пациентов, у которых сердечного приступа не было. Результаты этого эксперимента содержатся в файле `cholestg.csv`, в котором время после сердечного приступа закодировано в переменной `day`, а принадлежность пациента к опыту или контролю - в переменной `group` (1 - сердечный приступ, 2 - контроль, нет приступа).

Используя данные эксперимента, проверьте при помощи классического и пермутационного t-критерия, различается ли уровень холестерина у пациентов во второй день после сердечного приступа и в контрольной группе.

1. Для начала работы установите путь к рабочей директории, в которую вы сохранили файл с данными, одним из способов:

- либо через главное меню RStudio **Session -> Set Working Directory -> Choose Directory**
- либо выполните команду `setwd("путь-к-вашей-рабочей-директории")` вписав правильный путь

2. Откройте файл с данными

```
dat <- read.table(file = "cholestg.csv", header = T)
```

3. Рассчитайте объемы выборок в обеих группах, которые нужно сравнить, посчитав непропущенные значения. Эта информация пригодится вам для отчета.

```
filter <- dat$group == 1 & dat$day == 2
exp <- dat$cholest[filter] # вектор, холестерин в эксперименте
contr <- dat$cholest[dat$group == 2] # вектор, холестерин в контроле
sum(!is.na(exp)) # объем выборки в экспериментальной группе
sum(!is.na(contr)) # объем выборки в контроле
```

4. Сформулируйте нулевую гипотезу и протестируйте ее при помощи обычного двустороннего t-критерия используя функцию `t.test()`. Была ли отвергнута нулевая гипотеза, достоверны ли различия?
5. Сделайте пермутационный тест той же гипотезы, используя 1000 пермутаций. Вычислите величину уровня значимости пермутационной статистики. Была ли отвергнута нулевая гипотеза, достоверны ли различия? Сопоставьте результаты пермутационного и параметрического теста из предыдущего пункта.
6. Подумайте, какой биологический вывод можно сделать по результатам анализа?
7. Постройте график распределения пермутационной статистики Для этого загрузите пакет `ggplot2`, из которого вам понадобится функция `ggplot()`, геом `geom_histogram()`, и функции для подписывания осей `xlab()` и `ylab()`
8. Сохраните график в `png` файл одним из возможных способов:
 - либо сохраните получившийся график в переменную, а затем сохраните его при помощи при помощи функции `ggsave()`, например так:

```
dperm_pl <- код-вашего-графика
ggsave(filename = "distr-of-permutations.png", plot = dperm_pl)
```

- либо воспользовавшись меню на панели с графиками в RStudio: кнопка **Export** -> **Save Plot as Image ...**

9. Напишите отчет о ваших вычислениях по образцу, выложенному на сайте <https://yadi.sk/d/yMn70MO3baLby>. В отчете должен быть рисунок (либо в тексте, либо в отдельном файле) и достаточно подробная подпись к нему.
10. В отчете вам понадобится привести ссылки на использованные статистические методы и программное обеспечение.

Узнать, как правильно процитировать R можно, выполнив команду `citation()`

11. Проверьте, все ли вы сделали - посмотрите критерии оценок на сайте <https://sites.google.com/site/mathmethr/evaluation-criteria>.