Project №2

Elizaveta Demidenko

2025

# Введение

Объём выдоха (FEV, forced expiratory volume) - это количество воздуха, которое человек может выдохнуть с максимальным усилием после глубокого вдоха. Данный показатель используется для оценки функции дыхательной системы. На значения FEV могут оказывать влияние такие факторы, как пол, возраст и наличие вредных привычек, в частности курение.

В рамках данного исследования были поставлены две основные задачи:

1. Оценить различия в FEV между мужчинами и женщинами
2. Проанализировать FEV у представителей разных возрастных групп (≤ 21 год (Youth), 21 год > (Adult)) в зависимости от наличия привычки к курению

# Материалы и методы

Пакеты ggplot2 (Wickham 2024), dplyr (Wickham et al. 2024), car (Fox and Weisberg 2024), broom (Robinson et al. 2024).

Данные взяты из работы Tager, I. B., Weiss, S. T., Rosner, B., and Speizer, F. E. (1979). Effect of parental cigarette smoking on pulmonary function in children. American Journal of Epidemiology, 110, 15-26. Rosner, B. (1990). Fundamentals of Biostatistics, 3rd Edition. PWS-Kent, Boston, Massachusetts.

Анализ проводился на основе датасета fev.xls, sheet 3: tidy\_data

## Структура данных

В датасете представлены такие переменные, как:

Age - Возраст

FEV - Объем легких при выдохе (литры)

Height - Рост (дюймы)

Sex - пол (Male or Female)

Smoker - некурящие (Non), курящие (Current)

## Методика

Для анализа различий FEV между мужчинами и женщинами применялся двухвыборочный t-тест Уэлча. Этот тест был выбран потому как группы имеют разные размеры, независимы и могут иметь неодинаковую дисперсию.

##   
## Female Male   
## 318 336

Формула для статистики t-теста:

### Ограничения дисперсионного анализа

При подготовке данных и построении графиков обнаружилось,а затем было подтверждено проверкой, что в датасете отсутствуют данные по Adult, что является примером несбалансированных данных. Из-за чего поставленная задача (оценить влияние двух факторов (возрастной группы и курения) и их взаимодействия на FEV) не может быть выполнена. В связи с этим был проведён однофакторный анализ среди возрастной группы Youth по статусу курения.

# Результаты

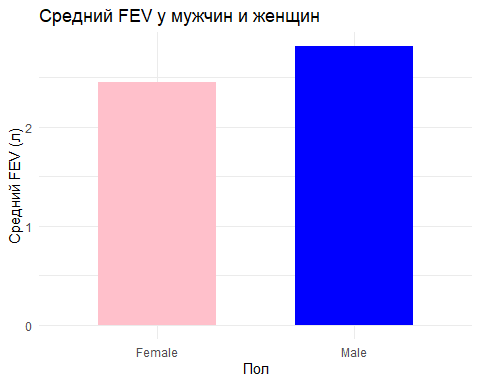
Результаты t-теста (FEV у мужчин и женщин)

| estimate | estimate1 | estimate2 | statistic | p.value | parameter | conf.low | conf.high | method | alternative |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| -0.3612766 | 2.45117 | 2.812446 | -5.503733 | 1e-07 | 575.7531 | -0.4902038 | -0.2323494 | Welch Two Sample t-test | two.sided |

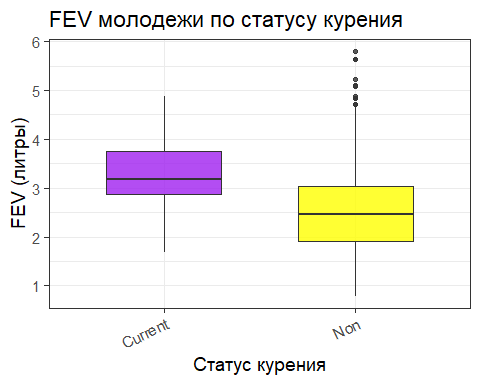
## p-value: 5.6e-08

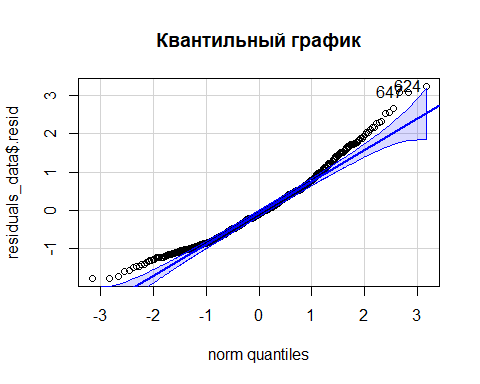
## p < 0.0001

## p-value < 0.0001, отвергаем нулевую гипотезу, статистически значимые различия между мужчинами и женщинами по FEV



Средний объём FEV у женщин составил 2.45 л, у мужчин 2.81 л. Разница между группами статистически значима: p < 0.0001. Таким образом, мужчины имеют более высокий FEV, чем женщины.





## [1] 624 647

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = FEV ~ Smoker, data = fev\_youth\_clean)  
##   
## $Smoker  
## diff lwr upr p adj  
## Non-Current -0.7222512 -0.9389704 -0.5055319 0

Средний FEV у курящих оказывается выше, чем у некурящих.

# Обсуждение результатов

По результатам анализа было выявлено различие FEV между мужчинами и женщинами, что подтверждает альтернативную гипотезу. Мужчины имеют больший FEV, что обусловлено различиями в физиологии.

Анализ данных по статусу курения среди молодёжи неоднозначен, данные по молодежи включают широкий диапазон возрастов (5- 16 лет), что влияет на результаты. Очевидно, молодежь младшего возраста скорее всего не курит и имеет обусловленный возрастом более низкий FEV, чем подростки, которые чаще курят. Это влечет за собой потенциальное смещение среднего значения FEV в группе некурящих, которое может казаться ниже, чем у курящих подростков. Отсутствие данных по взрослой группе ограничивает возможность анализа влияния курения и возраста на FEV в полном объёме.

# Библиография

Fox, John, and Sanford Weisberg. 2024. *Car: Companion to Applied Regression*. <https://cran.r-project.org/package=car>.

Robinson, David et al. 2024. *Broom: Convert Statistical Analysis Objects into Tidy Tibbles*. <https://broom.tidymodels.org>.

Wickham, Hadley. 2024. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://ggplot2.tidyverse.org>.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2024. *Dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.