Проект

Сандра Алексеевна Мирошниченко

2025-06-01

# Введение

В данном проекте будут анализироваться ёмкости лёгких у мужчин и женщин разного роста и возраста. Сначала я определю есть ли различия в объёме лёгких у мужчин (2) и женщин (1). Затем я проанализирую связь объёма легких с ростом у мужчин и женщин.

# Материалы и методы

В проекте используются пакеты ggplot2 (Wickham 2016), dplyr (Wickham et al. 2023), car (Fox and Weisberg 2019), broom (Robinson, Hayes, and Couch 2024). Данные взяты из датасета tlc.csv.

# Анализ корреляции

##   
## Pearson's product-moment correlation  
##   
## data: lungs$tlc and lungs$sex  
## t = 3.6693, df = 30, p-value = 0.0009392  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.2579291 0.7581292  
## sample estimates:  
## cor   
## 0.5565664

Между признаками есть высокая, статистически значимая корреляция. Пороговое значение для t-критерия Стьюдента для отвержения нулевой гипотезы составляет 2

По результатам анализа: t в соr тесте больше чем t фактич, cледовательно нулевая гипотеза отвергается.

# Линейная модель

Мы хотим построить линейную модель следующего вида отдельно для мужчин и женщин.

Подбираем модель с помощью функции lm().

Для женщин:

lm(tlc ~ height, data = lungs[lungs$sex == "1",]) -> model\_1  
summary(model\_1)

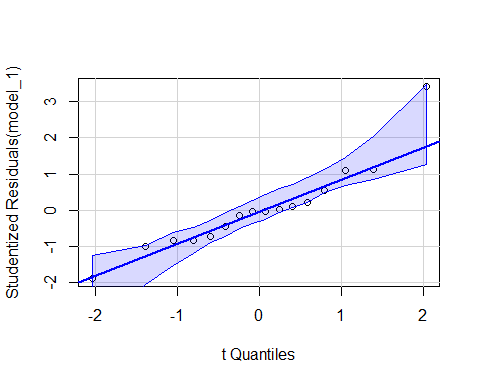
##   
## Call:  
## lm(formula = tlc ~ height, data = lungs[lungs$sex == "1", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.65193 -0.81215 -0.03884 0.31004 2.75770   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -7.55564 4.90645 -1.540 0.1459   
## height 0.07931 0.03046 2.604 0.0208 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.105 on 14 degrees of freedom  
## Multiple R-squared: 0.3262, Adjusted R-squared: 0.2781   
## F-statistic: 6.778 on 1 and 14 DF, p-value: 0.02083

Для мужчин:

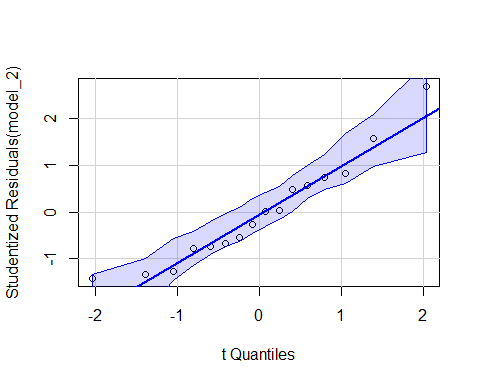
lm(tlc ~ height, data = lungs[lungs$sex == "2",]) -> model\_2  
summary(model\_2)

##   
## Call:  
## lm(formula = tlc ~ height, data = lungs[lungs$sex == "2", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.4965 -0.8306 -0.1637 0.7401 2.6984   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.82797 5.26859 -1.106 0.2873   
## height 0.07356 0.03022 2.435 0.0289 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.248 on 14 degrees of freedom  
## Multiple R-squared: 0.2975, Adjusted R-squared: 0.2473   
## F-statistic: 5.928 on 1 and 14 DF, p-value: 0.02888

## Анализ валидности модели



Квантильный график по женщинам



Квантильный график по мужчинам

Отклонений от нормальности распределения нет.

## Графики рассеивания

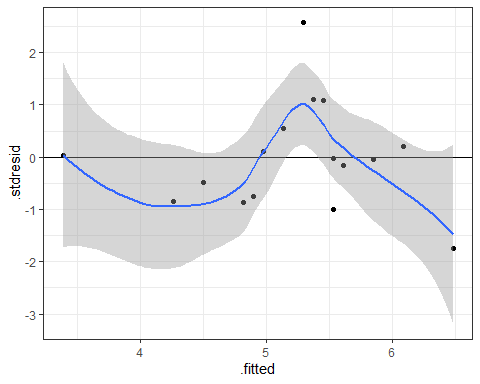


График рассеивания остатков у женщин

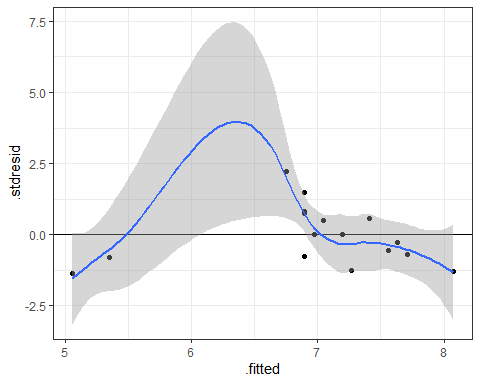


График рассеивания остатков у мужчин

Нарушений по обоим графикам нет.

# Итоги моделей

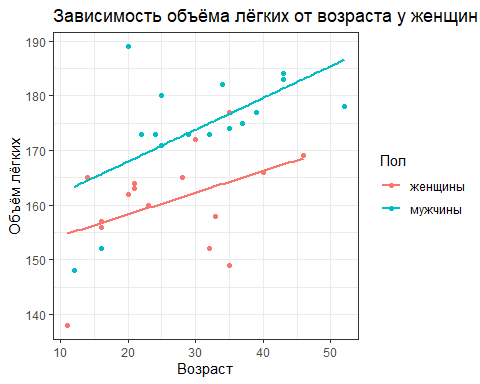
Таблица 1. Параметры регрессионной модели у женщин

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | -7.5556355 | 4.906455 | -1.539938 | 0.1458688 |
| height | 0.0793083 | 0.030462 | 2.603516 | 0.0208343 |

Таблица 2. Параметры регрессионной модели у мужчин

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | -5.8279713 | 5.2685883 | -1.106173 | 0.2872979 |
| height | 0.0735646 | 0.0302153 | 2.434683 | 0.0288819 |

# Регрессионный анализ



# Выводы

Между изученными параметрами существует сильная связь, которая описывается следующими моделями.

У женщин:

У мужчин:

# Библиография

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage. <https://www.john-fox.ca/Companion/>.

Robinson, David, Alex Hayes, and Simon Couch. 2024. *Broom: Convert Statistical Objects into Tidy Tibbles*. <https://CRAN.R-project.org/package=broom>.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.