Итоговая работа

Василькова С.В.

2025-06-01

# Введение

В рамках курса по математическим методам в биологии была поставлена задача изучить зависимость между объемом легких от возраста у курящих и некурящих людей с помощью полученных знаний.

# Материал и методика

В данной работе будут использоваться данные из следующих исследований (в библиографии), а также использоваться пакеты ggplot2 (Wickham 2016), dplyr (Wickham et al. 2023), car (Fox and Weisberg 2019), readxl (Wickham and Bryan 2025), broom (Robinson, Hayes, and Couch 2024), knitr (в библоиграфии). Что касается методики, то будут использоваться линейный регрессионный анализ и t-test.

# Результаты (ответы на 2 заданных вопроса)

Различается ли объем легких у курящих и некурящих людей?

Нулевая гипотеза: статистически значимых отличий нет (объем легких не отличается) (т.е. средние значения равны) Альтернативная гипотеза: статистически значимые отличия есть (наоборот, объем легких у курящих и некурящих разный) (средние значения будут отличаться)

##   
## Welch Two Sample t-test  
##   
## data: FEV by Smoker  
## t = 7.1902, df = 81.364, p-value = 2.833e-10  
## alternative hypothesis: true difference in means between group Current and group Non is not equal to 0  
## 95 percent confidence interval:  
## 0.5224013 0.9221010  
## sample estimates:  
## mean in group Current mean in group Non   
## 3.284391 2.562139

## [1] Различия есть

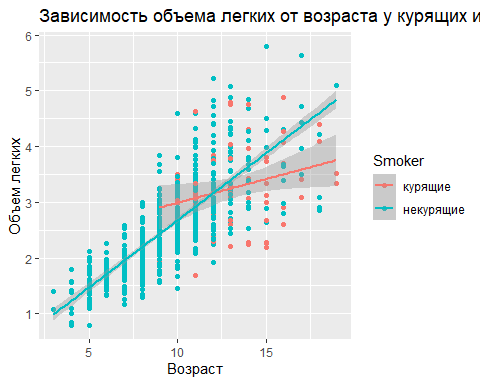
Так как значение p-value > 0,05, отвергаем H0 и принимаем H1.

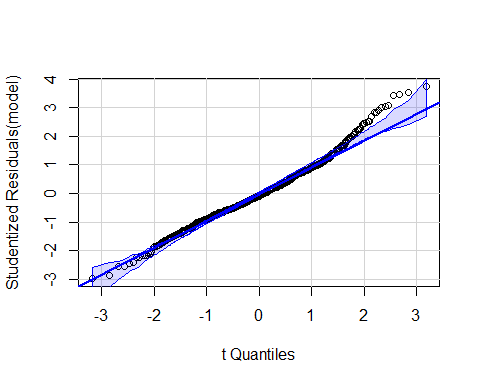
Как зависит объем легких от возраста у курящих и некурящих людей?

Линейная модель будет следующего вида:

##   
## Call:  
## lm(formula = FEV ~ Age + Smoker, data = fev)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.65435 -0.35325 -0.04928 0.34708 2.09387   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.375675 0.081756 4.595 5.2e-06 \*\*\*  
## Age 0.229538 0.008226 27.904 < 2e-16 \*\*\*  
## Smokerкурящие -0.186458 0.081280 -2.294 0.0221 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5658 on 649 degrees of freedom  
## Multiple R-squared: 0.5735, Adjusted R-squared: 0.5722   
## F-statistic: 436.4 on 2 and 649 DF, p-value: < 2.2e-16

##   
## Call:  
## lm(formula = FEV ~ Age + Smoker, data = fev)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.65435 -0.35325 -0.04928 0.34708 2.09387   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.189217 0.131548 1.438 0.1508   
## Age 0.229538 0.008226 27.904 <2e-16 \*\*\*  
## Smokerнекурящие 0.186458 0.081280 2.294 0.0221 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5658 on 649 degrees of freedom  
## Multiple R-squared: 0.5735, Adjusted R-squared: 0.5722   
## F-statistic: 436.4 on 2 and 649 DF, p-value: < 2.2e-16





Квантильный график

Отклонений от нормальности распределения почти нет.

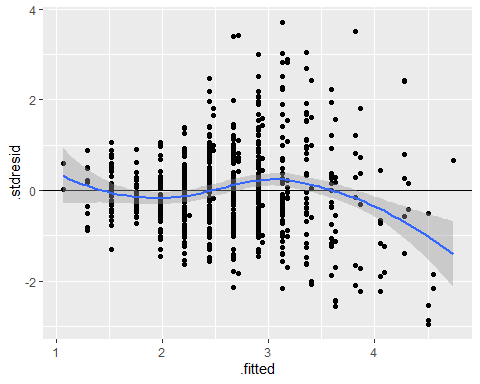


График рассеивания остатков

Нарушений нет.

Таблица 1. Параметры регрессионной модели.

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | 0.1892172 | 0.1315481 | 1.438388 | 0.1508060 |
| Age | 0.2295378 | 0.0082259 | 27.904381 | 0.0000000 |
| Smokerнекурящие | 0.1864579 | 0.0812795 | 2.294032 | 0.0221072 |

Между изученными параметрами существует сильная связь, которая описывается следующей моделью.

# Обсуждение

Подводя итог, на графике мы явно видим, что у некурящих объем легких с возрастом беспрепятственно увеличивается, не замедляется, в отличие от курящих.

# Библиография

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.

Xie Y (2025). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.50, <https://yihui.org/knitr/>.

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, and Jennifer Bryan. 2025. *Readxl: Read Excel Files*. <https://CRAN.R-project.org/package=readxl>.

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>.