Проект №4

Малина Лада

2025-06-01

# Введение

В первые месяцы жизни грудное молоко является основным источником питания для новорожденных. Объем потребляемого молока может зависеть от различных факторов, как связанных с ребенком (пол, вес), так и с матерью (вес, рост). В данном исследовании я анализирую влияние этих факторов на объем потребляемого молока.

# Материалы и методы

В проекте используются пакеты ggplot2 (Wickham 2016), dplyr (Wickham et al. 2023), ISwR (Dalgaard 2025), car (Fox and Weisberg 2019), broom (Robinson, Hayes, and Couch 2024). Также в исследовании использовался датасет “kfm”. Датасет содержит информацию о 50 кормящих матерях и их детях. Я исследовала связь между объемом потребляемого молока (dl.milk) и полом ребенка (sex), весом ребенка (weight), и весом матери (mat.weight).

Для анализа различий в объеме потребляемого молока между мальчиками и девочками использовался t-тест. Также были построены две линейные регрессионные модели:

1. dl.milk ~ weight (связь объема потребляемого молока с весом ребенка)
2. dl.milk ~ mat.weight (связь объема потребляемого молока с весом матери)

# Результаты

## Сравнение потребления молока по полу

##   
## Welch Two Sample t-test  
##   
## data: dl.milk by sex  
## t = 2.174, df = 47.89, p-value = 0.03468  
## alternative hypothesis: true difference in means between group boy and group girl is not equal to 0  
## 95 percent confidence interval:  
## 0.06729676 1.72470324  
## sample estimates:  
## mean in group boy mean in group girl   
## 7.9524 7.0564

Нулевая гипотеза: статистически значимых различий в объеме потребляемого молока между мальчиками и девочками нет. Если p.value < 0.05, то результаты t-теста показывают статистически значимые различия в объеме потребляемого молока между мальчиками и девочками. Если p.value > 0.05, то результаты t-теста не показывают статистически значимых различий в объеме потребляемого молока между мальчиками и девочками.В данном случае p < 0.05, поэтому различия есть.

## Регрессионные модели

1. Объём молока и вес ребёнка

##   
## Call:  
## lm(formula = dl.milk ~ weight, data = kfm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.9467 -0.8973 0.1148 0.8757 2.5186   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.821 1.641 -1.110 0.273   
## weight 1.753 0.307 5.711 6.92e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.178 on 48 degrees of freedom  
## Multiple R-squared: 0.4046, Adjusted R-squared: 0.3921   
## F-statistic: 32.61 on 1 and 48 DF, p-value: 6.915e-07

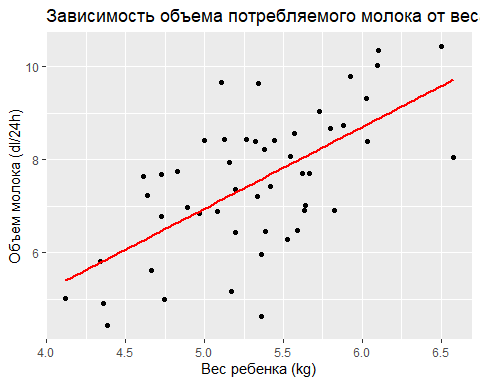
Результаты регрессионного анализа показывают, что вес ребенка является статистически значимым предиктором объема потребляемого молока (p < 0.05). Уравнение регрессии имеет вид:

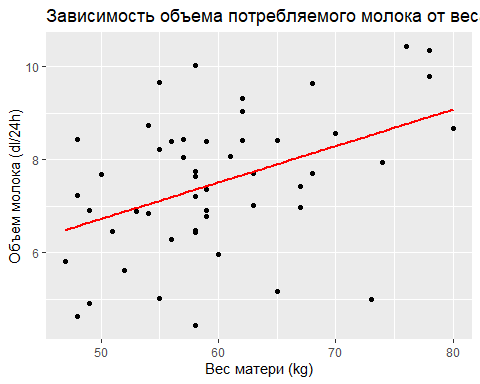
2. Объем молока и вес матери

##   
## Call:  
## lm(formula = dl.milk ~ mat.weight, data = kfm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5257 -0.7179 -0.0559 0.9595 2.6791   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.80837 1.41935 1.979 0.05361 .   
## mat.weight 0.07832 0.02345 3.340 0.00163 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.376 on 48 degrees of freedom  
## Multiple R-squared: 0.1886, Adjusted R-squared: 0.1717   
## F-statistic: 11.16 on 1 and 48 DF, p-value: 0.001627

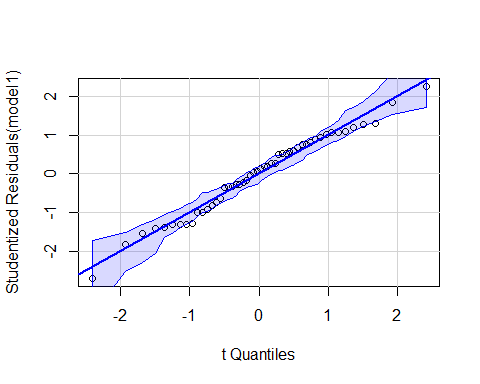
Результаты регрессионного анализа показывают, что вес матери также является статистически значимым предиктором объема потребляемого молока (p < 0.05).

## Графики





## Анализ валидности модели 1



Квантильный график 1

Отклонений от нормальности распределения нет.

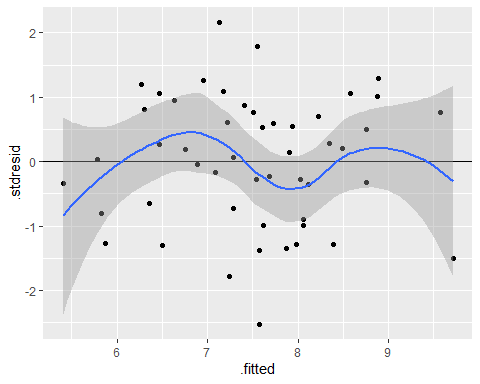
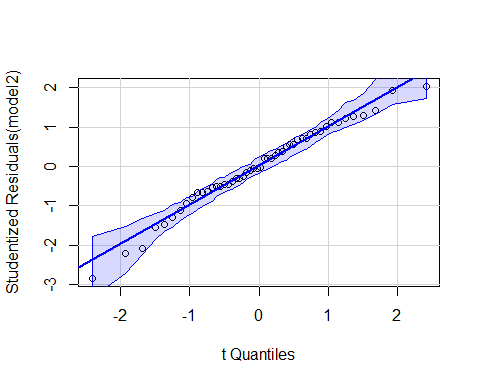


График рассеивания остатков 1

Нарушений нет.

## Анализ валидности модели 2



Квантильный график 2

Отклонений от нормальности распределения нет.

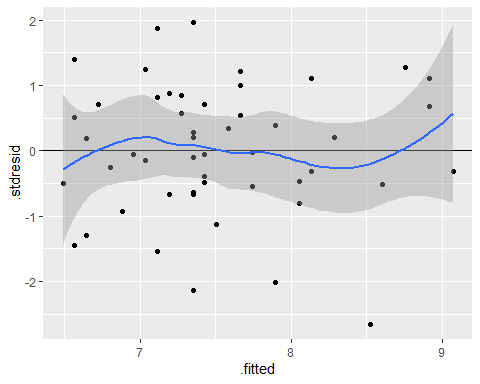


График рассеивания остатков 2

Нарушений нет.

## Таблица

Таблица 1. Параметры регрессионной модели.

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | -1.821428 | 1.6415361 | -1.109587 | 0.2727051 |
| weight | 1.753390 | 0.3070378 | 5.710666 | 0.0000007 |

Таблица 2. Параметры регрессионной модели.

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | 2.8083742 | 1.4193469 | 1.978638 | 0.0536103 |
| mat.weight | 0.0783193 | 0.0234481 | 3.340107 | 0.0016269 |

# Обсуждение

Проведенный анализ показывает, что есть статистически значимые различия в объеме потребляемого молока между мальчиками и девочками. Вес ребенка а также вес матери оказывают значимое влияние на объем потребляемого молока.

Связи между изученными параметрами описываются следующими моделями:

* Необходимо отметить, что исследование проводилось на небольшой выборке, что ограничивает возможность обобщения результатов.

# Библиография

Dalgaard, Peter. 2025. *ISwR: Introductory Statistics with r*. <https://CRAN.R-project.org/package=ISwR>.

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>.