Отчет о домашнем задании

Сандра Алексеевна Мирошниченко

2025

# Точечный график

Задача этого раздела - построить точечный график. Мы это сделаем с помощью пакета ggplot2 [1]. Данные были упорядочены с помощью пакета dplyr [2]

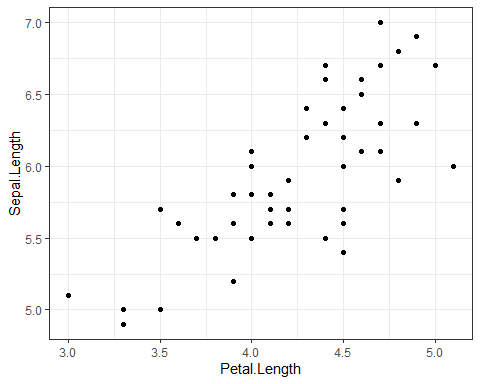


Рисунок 1. Связь между изученными признаками.

Мы видим прекрасную связь!!!

## Анализ корреляции

##   
## Pearson's product-moment correlation  
##   
## data: df\_versicolor$Petal.Length and df\_versicolor$Sepal.Length  
## t = 7.9538, df = 48, p-value = 2.586e-10  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.6020680 0.8532995  
## sample estimates:  
## cor   
## 0.754049

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

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

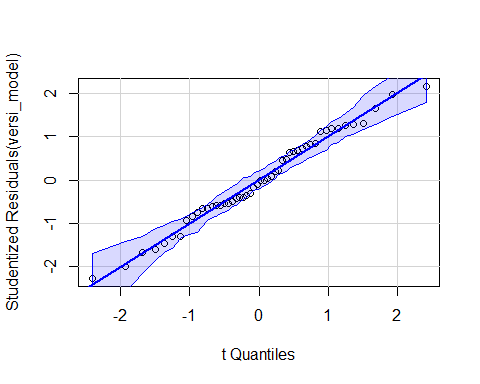
# Линейная модель

Мы хотим построить линейную модель следующего вида.

Подбираем модель с помощью функции lm().

versi\_model <- lm(formula = Sepal.Length ~ Petal.Length, data = df\_versicolor)

## Анализ валидности модели



Квантильный график

Отклонения от нормальности распределеиня нет.

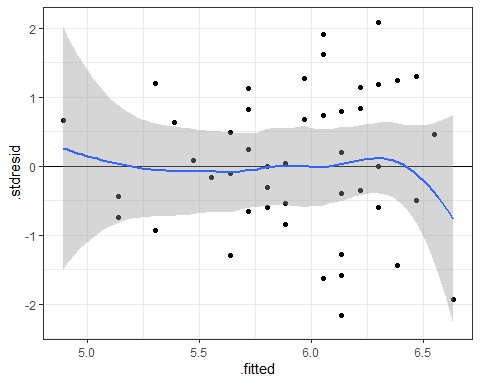


График рассеяния остатков

Нарушений нет.

## Итоги модели

Таблица 1. Паармтеры регрессионной модели.

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | 2.407523 | 0.4462583 | 5.394909 | 2.1e-06 |
| Petal.Length | 0.828281 | 0.1041364 | 7.953806 | 0.0e+00 |

# Выводы

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

# Библиография

1. Wickham H. [ggplot2: Elegant graphics for data analysis](https://ggplot2.tidyverse.org). Springer-Verlag New York, 2016.

2. Wickham H. et al. [Dplyr: A grammar of data manipulation](https://CRAN.R-project.org/package=dplyr). 2023.