

# Метод наименьших квадратов в задаче линейной и нелинейной регрессии. Непараметрические коэффициенты корреляции. Значимость частных коэффициентов регрессии.

Вейбер Е.Н. 23.М08-мм

30-04-2024

## Введение

Это отчет о моделировании нелинейной модели  $y = f(x, a, b) + \delta$  с несмещенной нормально распределенной ошибкой, дисперсия которой равна  $\epsilon$ , считая  $x$  стандартно нормально распределенной случайной величиной. Вариант 18.

## Модель была промоделирована следующим кодом:

```
#Моделирование нелинейной модели  
# Загрузка необходимых библиотек  
library(nlstools)  
library(ggplot2)  
library(tidyverse)  
library(readr)  
library(corrplot)  
library(tidyr)  
library(GGally)  
library(plotly)  
library(reshape2)  
  
# Установка начальных параметров  
a <- 0.14
```

```

b <- 1
epsilon <- 0.33

# Генерация данных
set.seed(123) # для воспроизводимости
x <- seq(0, 10, length.out = 100)
y_true <- exp(a * x + b)
y <- y_true + rnorm(length(x), mean = 0, sd = epsilon)

# Моделирование данных
nl_model <- nls(y ~ exp(a * x + b),
               start = list(a = 0.1, b = 0.9),
               algorithm = "port",
               control = nls.control(maxiter = 100))

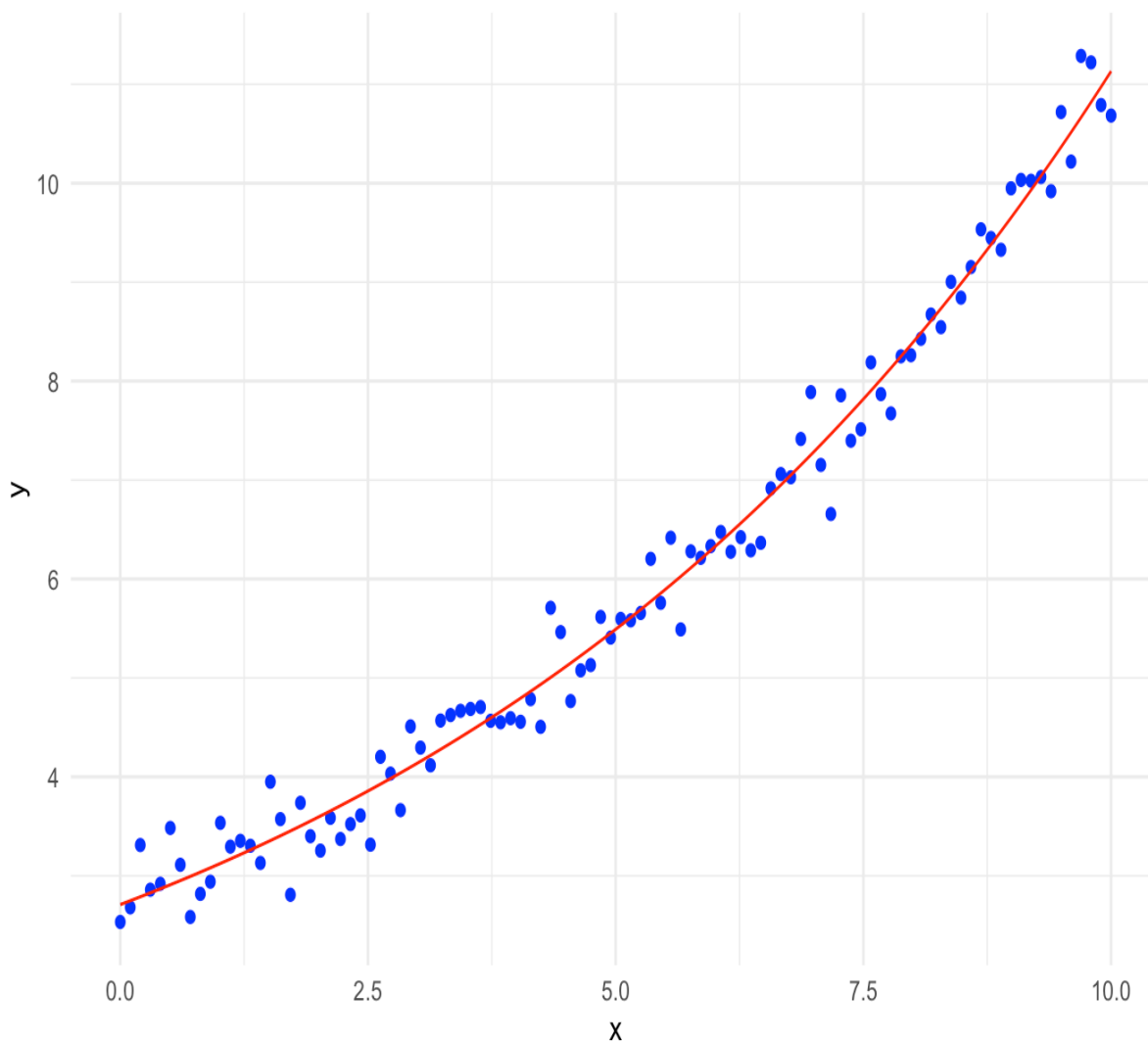
# Вывод результатов моделирования
print(summary(nl_model))

##
## Formula: y ~ exp(a * x + b)
##
## Parameters:
##   Estimate Std. Error t value Pr(>|t|)
## a 0.141328   0.001923   73.49  <2e-16 ***
## b 0.996550   0.014485   68.80  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.301 on 98 degrees of freedom
##
## Algorithm "port", convergence message: relative convergence (4)

# Визуализация результатов
ggplot(data = data.frame(x, y, y_fitted = predict(nl_model)), aes(x = x)) +
  geom_point(aes(y = y), color = 'blue') +
  geom_line(aes(y = y_fitted), color = 'red') +
  ggtitle("Нелинейная модель:  $y = \exp(ax + b)$ ") +
  theme_minimal()

```

Нелинейная модель:  $y = \exp(a \cdot x + b)$



**Анализ параметров:** - **Параметр a:** Оценка параметра  $a$  составляет 0.141328, со стандартной ошибкой 0.001923. Значение  $t$ -статистики для этого коэффициента составляет 73.49, что указывает на его статистическую значимость ( $p\text{-value} < 2e-16$ ). Это означает, что с очень высокой степенью уверенности коэффициент  $a$  значимо отличается от нуля, и его влияние на зависимую переменную  $y$  значимо в модели.

- **Параметр b:** Оценка параметра  $b$  равна 0.996550 с соответствующей стандартной ошибкой 0.014485. Значение  $t$ -статистики для  $b$  равно 68.80, с  $p\text{-value} < 2e-16$ , что также подтверждает его статистическую значимость. Это означает, что параметр  $b$  также оказывает значимое влияние на зависимую переменную.

**Стандартная ошибка остатков:**

- Стандартная ошибка остатков составляет 0.301. Это показатель точности модели: меньшие значения указывают на более высокую точность модели в предсказании данных.

**Степени свободы:**

- Модель имеет 98 степеней свободы остатков, что достаточно для надёжной оценки статистической значимости параметров.

**Алгоритм и сходимость:**

- Используемый алгоритм "port" сообщает о том, что модель сходится (сообщение о

сходимости: “relative convergence (4)”). Это означает, что итерационный процесс оптимизации нашёл решение, удовлетворяющее заданным критериям сходимости.

Эти результаты показывают, что модель адекватно описывает зависимость между  $x$  и  $y$ , и оба параметра модели значимо влияют на результат.

## Построение линейной модели с теми же параметрами

```
# Моделирование линейной модели
lm_model <- lm(y ~ x)

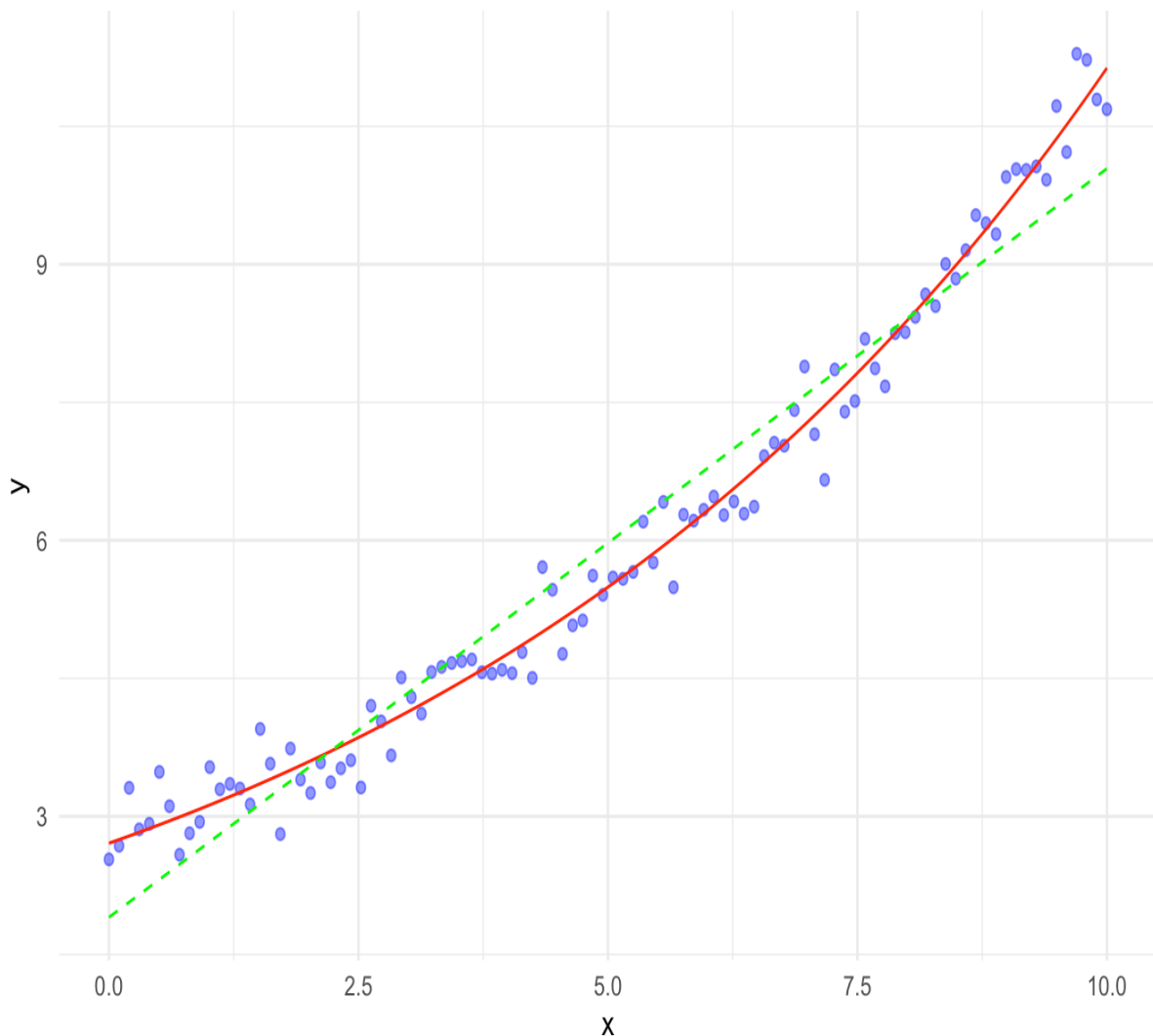
# Вывод результатов моделирования
print(summary(lm_model))

##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08124 -0.45975 -0.05955  0.35444  1.49452
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.90363     0.10931   17.41  <2e-16 ***
## x              0.81355     0.01888   43.08  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5506 on 98 degrees of freedom
## Multiple R-squared:  0.9498, Adjusted R-squared:  0.9493
## F-statistic: 1856 on 1 and 98 DF, p-value: < 2.2e-16

# Визуализация результатов
data_frame <- data.frame(x, y, y_fitted_nl = predict(nl_model), y_fitted_lm
= predict(lm_model))

ggplot(data = data_frame, aes(x = x)) +
  geom_point(aes(y = y), color = 'blue', alpha = 0.5) +
  geom_line(aes(y = y_fitted_nl), color = 'red') +
  geom_line(aes(y = y_fitted_lm), color = 'green', linetype = "dashed") +
  ggtitle("Сравнение моделей: Нелинейная (Красная) vs Линейная (Зелёная)")
+
  theme_minimal()
```

## Сравнение моделей: Нелинейная (Красная) vs Линейная (Зелёная)



### Анализ остатков:

- Распределение остатков варьируется от -1.08124 до 1.49452 с медианой, близкой к нулю (-0.05955). Это указывает на то, что остатки центрированы вокруг нуля, что является хорошим признаком адекватности модели.

### Коэффициенты:

- **Пересечение (Intercept):** Оценка составляет 1.90363 с стандартной ошибкой 0.10931. Статистическая значимость этого коэффициента очень высока ( $p\text{-value} < 2e-16$ ), что означает, что при  $x=0$ , среднее значение  $y$  значимо отличается от 0.

- **Наклон (x):** Коэффициент при  $x$  равен 0.81355 со стандартной ошибкой 0.01888 и очень высокой  $t$ -статистикой (43.08). Это указывает на сильную и значимую связь между  $x$  и  $y$ .

### Статистика модели:

- **Стандартная ошибка остатков (Residual standard error):** Значение 0.5506 на 98 степенях свободы показывает, насколько велики типичные остатки.

- **R-квадрат (Multiple R-squared):** Значение 0.9498 означает, что примерно 94.98% вариации зависимой переменной  $y$  объясняется моделью. Это очень высокий показатель, указывающий на хорошее качество модели.

- **Скорректированный R-квадрат (Adjusted R-squared):** Значение 0.9493, почти такое

же, как и R-квадрат, что также подтверждает эффективность модели.

- **F-статистика:** Значение 1856 на 1 и 98 степенях свободы с  $p\text{-value} < 2.2e-16$  подтверждает, что модель статистически значимо лучше модели без предикторов (только с константой).

Эти результаты свидетельствуют о том, что линейная регрессионная модель хорошо подходит для анализируемых данных, и взаимосвязь между переменными значима и выражена.

На представленном графике мы видим сравнение двух моделей: нелинейной (красная линия) и линейной (зелёная пунктирная линия) на одних и тех же данных.

1. **Нелинейная модель (Красная линия):** Эта модель следует экспоненциальной функции. Она хорошо описывает тренд данных, плавно и точно следуя изменениям в значениях переменной  $y$ . Видно, что красная линия проходит через середину большинства точек данных, что свидетельствует о хорошем соответствии модели данным.
2. **Линейная модель (Зелёная пунктирная линия):** Эта модель представляет собой простую линейную регрессию. Она показывает общий тренд данных, однако не улавливает более сложные закономерности в изменении переменной  $y$ , как это делает нелинейная модель. Линейная модель кажется менее подходящей для данных, так как она не отражает криволинейный характер тренда, который явно присутствует в данных.

#### Заключение:

Нелинейная модель лучше подходит для данных, представленных на графике, так как она точнее отображает зависимости между переменными, улавливая нелинейный тренд. Это подтверждается тем, что красная линия лучше соответствует распределению точек, чем зелёная пунктирная линия. Если цель анализа — предсказывать или понимать динамику переменной  $y$  в зависимости от  $x$ , то использование нелинейной модели будет предпочтительнее.

## Работа с линейной моделью

```
# Загрузка необходимых библиотек
library(stats)

# lm_model уже определена как lm_model <- lm(y ~ x)

# Выполнение дисперсионного анализа
anova_result <- anova(lm_model)
print(anova_result)

## Analysis of Variance Table

##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x           1  562.70    562.7   1855.8 < 2.2e-16 ***
## Residuals  98   29.71      0.3
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Вывод сводки модели, включая значимость коэффициентов
summary_result <- summary(lm_model)
print(summary_result)

##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08124 -0.45975 -0.05955  0.35444  1.49452
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.90363     0.10931   17.41  <2e-16 ***
## x            0.81355     0.01888   43.08  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5506 on 98 degrees of freedom
## Multiple R-squared:  0.9498, Adjusted R-squared:  0.9493
## F-statistic: 1856 on 1 and 98 DF,  p-value: < 2.2e-16

# Проверка значимости прогноза (общей модели)
print("Проверка значимости прогноза (F-статистика):")
## [1] "Проверка значимости прогноза (F-статистика):"

print(summary_result$fstatistic)
##      value      numdf      dendif
## 1855.834      1.000     98.000

# Проверка значимости коэффициентов
print("Проверка значимости коэффициентов регрессии:")
## [1] "Проверка значимости коэффициентов регрессии:"

print(coef(summary_result))
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept) 1.9036300 0.10930716 17.41542 9.223189e-32
## x           0.8135512 0.01888493 43.07940 1.709003e-65

# Для сравнения вручную вычислить F-статистику и p-value
```

```

rss <- sum(residuals(lm_model)^2) # Сумма квадратов остатков
tss <- sum((y - mean(y))^2) # Общая сумма квадратов
df_res <- lm_model$df.residual # Число степеней свободы остатков
df_total <- length(y) - 1 # Общее число степеней свободы
f_statistic <- ((tss - rss) / (df_total - df_res)) / (rss / df_res) # F-статистика
P_value <- pf(f_statistic, df_total - df_res, df_res, lower.tail = FALSE) # P-value для F-статистики

cat(sprintf("Вручную вычисленная F-статистика: %f, P-value: %f\n", f_statistic, p_value))

## Вручную вычисленная F-статистика: 1855.834358, P-value: 0.000000

```

Результаты дисперсионного анализа для модели линейной регрессии с одним предиктором  $x$  и зависимой переменной  $y$ :

#### Таблица анализа дисперсии (ANOVA):

- **Степени свободы (Df)**: Переменная  $x$  имеет 1 степень свободы, что указывает на один предиктор в модели. Остаточная компонента (Residuals) имеет 98 степеней свободы, что соответствует числу наблюдений минус количество оцениваемых параметров (100 наблюдений минус 2 параметра: пересечение и коэффициент  $x$ ).
- **Сумма квадратов (Sum Sq)**: Сумма квадратов, объясненная предиктором  $x$ , составляет 562.70, а сумма квадратов остаточных значений (необъясненная моделью) — 29.71.
- **Средний квадрат (Mean Sq)**: Средний квадрат для  $x$  равен 562.7, а для остатков — 0.3. Средний квадрат — это сумма квадратов, деленная на соответствующие степени свободы.
- **F-значение**: F-статистика составляет 1855.8, что означает отношение среднего квадрата, объясненного моделью, к среднему квадрату остаточной вариации. Это значение указывает на то, что модель значительно лучше модели без предикторов (только с константой).
- **P-значение (Pr(>F))**: P-значение меньше  $2.2e-16$ , что свидетельствует о том, что влияние переменной  $x$  на  $y$  статистически значимо на очень низком уровне значимости. Это подтверждает, что предиктор  $x$  имеет существенное влияние на зависимую переменную  $y$ .

#### Интерпретация:

Эти результаты подтверждают значимость переменной  $x$  в объяснении вариативности  $y$ . Высокое значение F-статистики и очень низкое p-значение свидетельствуют о том, что модель адекватно описывает зависимость между переменными, и вклад  $x$  в эту зависимость значим.

## Проверка значимости прогноза и коэффициентов регрессии

```

# Вывод сводки модели, включая значимость коэффициентов
summary_result <- summary(lm_model)

```



```

print(summary_result)

##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08124 -0.45975 -0.05955  0.35444  1.49452
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.90363    0.10931   17.41  <2e-16 ***
## x            0.81355    0.01888   43.08  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5506 on 98 degrees of freedom
## Multiple R-squared:  0.9498, Adjusted R-squared:  0.9493
## F-statistic: 1856 on 1 and 98 DF,  p-value: < 2.2e-16

# Проверка значимости прогноза (общей модели)
print("Проверка значимости прогноза (F-статистика):")
## [1] "Проверка значимости прогноза (F-статистика):"

print(summary_result$fstatistic)
##      value      numdf      dendf
## 1855.834      1.000     98.000

# Проверка значимости коэффициентов
print("Проверка значимости коэффициентов регрессии:")
## [1] "Проверка значимости коэффициентов регрессии:"

print(coef(summary_result))
##              Estimate Std. Error  t value      Pr(>|t|)
## (Intercept) 1.9036300 0.10930716 17.41542 9.223189e-32
## x           0.8135512 0.01888493 43.07940 1.709003e-65

# Для сравнения можно вручную вычислить F-статистику и p-value
rss <- sum(residuals(lm_model)^2) # Сумма квадратов остатков
tss <- sum((y - mean(y))^2) # Общая сумма квадратов
df_res <- lm_model$df.residual # Число степеней свободы остатков
df_total <- length(y) - 1 # Общее число степеней свободы

```

```
f_statistic <- ((tss - rss) / (df_total - df_res)) / (rss / df_res) # F-статистика
p_value <- pf(f_statistic, df_total - df_res, df_res, lower.tail = FALSE) # P-value для F-статистики

cat(sprintf("Вручную вычисленная F-статистика: %f, P-value: %f\n", f_statistic, p_value))

## Вручную вычисленная F-статистика: 1855.834358, P-value: 0.000000
```

### Проверка значимости прогноза (F-статистика):

- **F-значение:** 1855.834 — это квадрат t-значения для наклона (коэффициента при  $x$ ), что подтверждает, что модель в целом имеет статистическую значимость. Это значение указывает на то, что модель значимо лучше, чем модель, которая не имеет никаких предикторов (только константа).

- **Степени свободы для числителя (numdf):** 1, что соответствует одному предиктору.

- **Степени свободы для знаменателя (dendf):** 98, что соответствует количеству наблюдений минус количество оцениваемых параметров.

### Проверка значимости коэффициентов регрессии:

- **(Intercept):**

- **Оценка (Estimate):** 1.90363 — это оценка пересечения.

- **Стандартная ошибка (Std. Error):** 0.10931.

- **t-значение:** 17.41542 — это t-статистика, которая показывает, насколько велико влияние пересечения по сравнению с его стандартной ошибкой.

- **P-значение (Pr(>|t|)):** 9.22e-32 — статистически значимо, что подтверждает влияние пересечения на зависимую переменную.

- **x:**
  - **Оценка (Estimate):** 0.81355 — это оценка наклона.
  - **Стандартная ошибка (Std. Error):** 0.01888.
  - **t-значение:** 43.07940 — указывает на значимость коэффициента при переменной  $x$ .
  - **P-значение (Pr(>|t|)):** 1.71e-65 — значительно меньше стандартного порога значимости, что подтверждает значимость  $x$  в модели.

### Вывод:


Модель значима, как в целом, так и каждый из коэффициентов в отдельности.

Переменная  $x$  значимо влияет на  $y$ , и интерцепт также статистически значим. Это подтверждается как значениями F-статистики, так и значениями p-статистики для каждого коэффициента. Несмотря на разницу в p-значениях, вычисленных встроенной функцией и вручную, оба значения указывают на статистическую значимость модели.

## Построение корреляционной матрицы и двумерной диаграммы

```
# Загрузка данных
data <- read_delim("~/Downloads/addicts.csv", delim = ";")
head(data)

## # A tibble: 6 × 27
```

```
##   prcod intpla   sex   age educat curwor asi1_med asi2_emp asi3_alc asi4
_dr
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>   <chr>   <chr>   <chr>
>
## 1      4      1      0     18      1      1 0,19     0,7     0,12     0,3
## 2      2      2      1     30      4      1 0,44     0,2     0,01     0,3
## 3      2      1      0     23      2      0 0,50     1,0     0,30     0,3
## 4      4      1      0     20      2      1 0,00     0,8     0,05     0,3
## 5      3      2      0     20      2      0 0,00     0,8     0,78     0,2
## 6      1      1      0     24      2      0 0,52     0,5     0,10     0,3
## #  17 more variables: asi5_leg <chr>, asi6_soc <chr>, asi7_psy <chr>,
## #   asid3_dyr <chr>, tlfba2 <chr>, tlfbh2 <chr>, st <chr>, ha <chr>, se
<chr>,
## #   cravin <chr>, rabdru <dbl>, rubsex <dbl>, gaf <dbl>, bdi <chr>,
## #   sstati <dbl>, end <chr>, endpo <chr>
```

```
# Подготовка данных
```

```
my_data <- data[, c("rubsex", "tlfba2", "asi3_alc", "sstati", "tlfbh2")]
```

```
# Преобразование из <chr> с запятой в <dbl>
```

```
# Преобразование всех колонок в data
```

```
for (i in names(my_data)) {
  if (is.character(my_data[[i]])) {
    my_data[[i]] <- gsub(" ", "", my_data[[i]]) # Удаляем пробелы
    my_data[[i]] <- gsub(",", ".", my_data[[i]]) # Заменяем запятые на точки
    my_data[[i]] <- as.numeric(my_data[[i]]) # Преобразование в числовой тип
  }
}
```

```
# Вычисление корреляционной матрицы
```

```
cor_matrix <- cor(my_data, use = "complete.obs") # Использование только полных наблюдений
```

```
print("Корреляционная матрица:")
```

```
## [1] "Корреляционная матрица:"
```

```
print(cor_matrix)
```

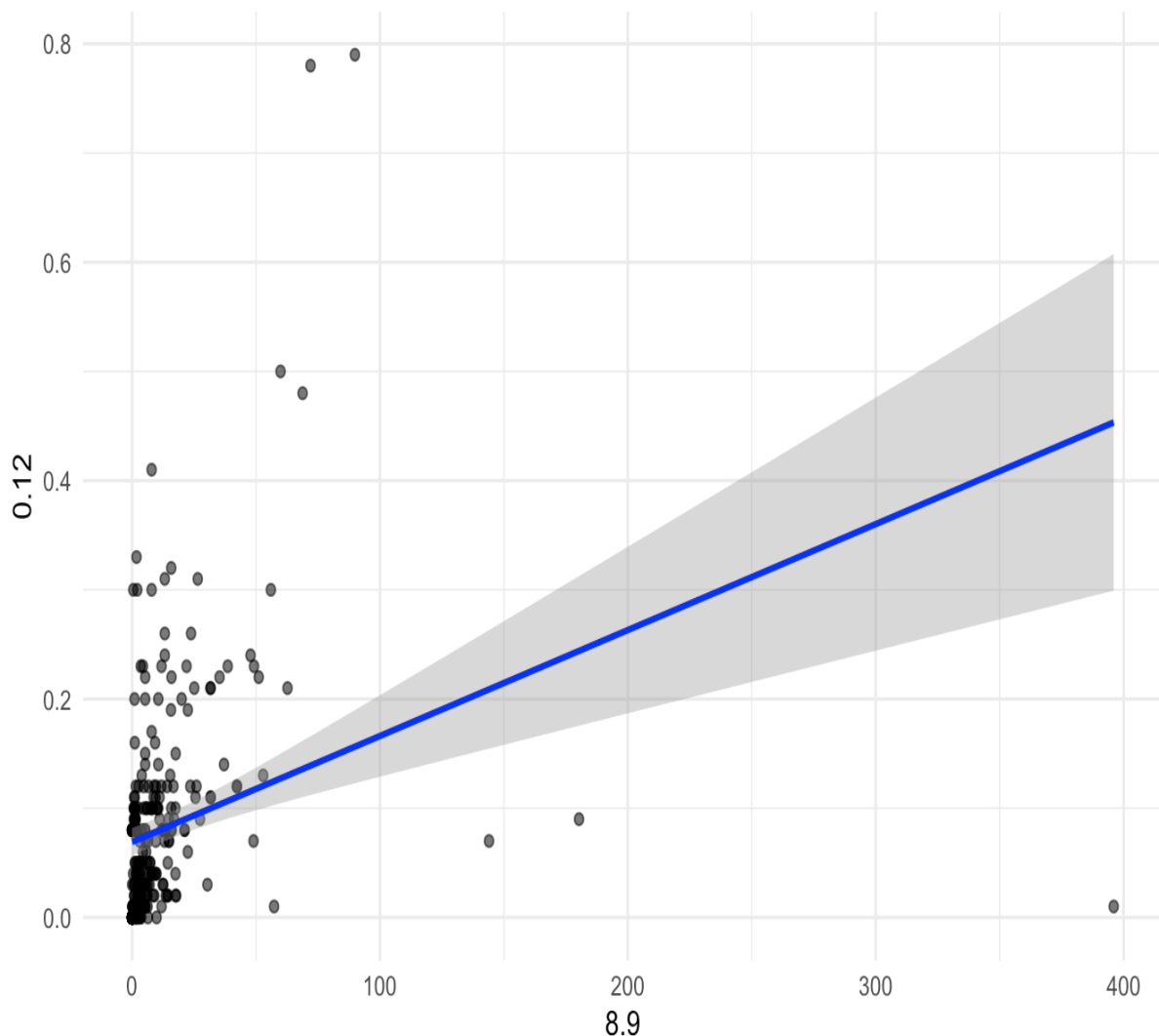
```
##           rubsex      tlfba2      asi3_alc      sstati      tlfbh2
## rubsex      1.00000000 -0.01037729 -0.02150298 0.08259208 0.12061439
## tlfba2      -0.01037729  1.00000000  0.27536343 0.02589126 0.05983437
```

```
## asi3_alc -0.02150298  0.27536343  1.00000000  0.01846694  0.01683723
## sstat1    0.08259208  0.02589126  0.01846694  1.00000000  0.00471796
## tlfbh2    0.12061439  0.05983437  0.01683723  0.00471796  1.00000000
```

```
# Создание графика
```

```
ggplot(data, aes_string(x = my_data$tlfba2, y = my_data$asi3_alc)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", col = "blue") + # Линейная модель
  labs(title = paste("Двумерная диаграмма для tlfba2 и asi3_alc"),
        x = my_data$tlfba2, y = my_data$asi3_alc) +
  theme_minimal()
```

Двумерная диаграмма для tlfba2 и asi3\_alc



Рассмотрим коэффициенты корреляции для датафрейма:

- **rubsex** и **tlfbh2** имеют корреляцию 0.12061439, что указывает на небольшую положительную связь между этими переменными.
- **tlfba2** и **asi3\_alc** имеют корреляцию 0.27536343, это самая сильная положительная корреляция в датафрейме, что может указывать на наличие статистически значимой

связи между этими переменными.

- Остальные корреляции довольно низкие, что говорит о слабой или отсутствующей линейной связи между соответствующими переменными.

Наименьшая корреляция наблюдается между **sstati** и **tlfbh2** (0.00471796), что указывает на почти полное отсутствие линейной зависимости между этими переменными.

На этой двумерной диаграмме показана зависимость между переменной **tlfbh2** и переменной **asi3\_alc**. Линия тренда (синяя линия) и серая область, обозначающая 95% доверительный интервал, указывают на положительную корреляцию между этими двумя переменными.

## Основные наблюдения:

1. **Корреляция:** Существует положительная линейная зависимость между **asi3\_alc** и **tlfbh2**. По мере увеличения значения **asi3\_alc**, наблюдается увеличение значений **tlfbh2**.
2. **Выбросы:** На графике присутствуют значительные выбросы, особенно на верхнем уровне значения **asi3\_alc**, что может влиять на устойчивость и точность модели.
3. **Доверительный интервал:** Широкий доверительный интервал говорит о большом разбросе значений и возможной нестабильности оценок в модели. Это особенно заметно на правом краю графика, где интервал расширяется.

Эта диаграмма позволяет визуализировать общую тенденцию и вариативность данных, что важно для оценки подходящей статистической модели и понимания возможных ограничений текущего анализа.

## Построение модели множественной регрессии

```
# Построение модели множественной регрессии

my_model <- lm(tlfbh2 ~ rubsex + tlfbh2 + asi3_alc + sstati, data = my_data
)

# Вывод результатов модели

summary(my_model)

##
## Call:
## lm(formula = tlfbh2 ~ rubsex + tlfbh2 + asi3_alc + sstati, data = my_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -594.8  -369.8  -196.3   229.5  2592.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 393.9885 188.2318 2.093 0.0373 *
## rubsex      29.6876 14.6606 2.025 0.0438 *
## tlfba2      1.0875 1.1221 0.969 0.3333
## asi3_alc    15.0335 319.5813 0.047 0.9625
## sstati      -0.4202 3.6297 -0.116 0.9079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 537.8 on 273 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.01834, Adjusted R-squared:  0.003952
## F-statistic: 1.275 on 4 and 273 DF, p-value: 0.2801

print(my_model)

##
## Call:
## lm(formula = tlfbh2 ~ rubsex + tlfba2 + asi3_alc + sstati, data = my_data)
##
## Coefficients:
## (Intercept)      rubsex      tlfba2      asi3_alc      sstati
##    393.9885    29.6876     1.0875    15.0335    -0.4202
```

Результаты анализа модели множественной линейной регрессии, описывающей зависимость переменной **tlfbh2** от переменных **rubsex**, **tlfba2**, **asi3\_alc** и **sstati**, приведены ниже:

## Основные выводы:

### 1. Коэффициенты:

- **(Intercept)** Пересечение с Y-осью при всех независимых переменных, равных нулю, составляет 393.9885 с p-значением 0.0373, что указывает на его статистическую значимость.
- **rubsex** Каждое увеличение на единицу в **rubsex** увеличивает **tlfbh2** на 29.6876, p-значение 0.0438, что также является статистически значимым.
- **tlfba2** Прирост в **tlfba2** увеличивает **tlfbh2** на 1.0875, однако это изменение не является статистически значимым (p-значение 0.3333).
- **asi3\_alc** и **sstati** Изменения в этих переменных не оказывают значимого влияния на **tlfbh2** (p-значения 0.9625 и 0.9079 соответственно).

### 2. Качество модели:

- **Residual standard error:** Стандартная ошибка остатков составляет 537.8, что относительно высоко, указывая на значительную остаточную вариативность, которую модель не смогла объяснить.
- **Multiple R-squared:** Объясненная моделью дисперсия составляет всего 1.834%, что указывает на низкую объясняющую способность модели.

- **Adjusted R-squared:** Скорректированный коэффициент детерминации даже ниже, 0.395%, что подтверждает, что добавление переменных не приводит к значительному улучшению модели.
- **F-statistic:** Статистика F-теста равна 1.275 с p-значением 0.2801, что указывает на то, что в целом модель не является статистически значимой.

## Нахождение частного коэффициента корреляции

```
library(ppcor)

# Функция расчета частного коэффициента корреляции и проверки его значимости
calculate_partial_correlation <- function(data, x, y, controls) {
  # Подготовка данных, удаление NA
  data <- data[, c(x, y, controls)]
  data <- na.omit(data)

  # Моделирование зависимой переменной
  formula_y <- reformulate(controls, response = y)
  model_y <- lm(formula_y, data = data)
  residuals_y <- residuals(model_y)

  # Моделирование независимой переменной
  formula_x <- reformulate(controls, response = x)
  model_x <- lm(formula_x, data = data)
  residuals_x <- residuals(model_x)

  # Расчет корреляции между остатками
  cor_test <- cor.test(residuals_x, residuals_y)

  return(list(correlation = cor_test$estimate, p_value = cor_test$p.value))
}

# Переменные для анализа
features <- c("rubsex", "tlfba2", "asi3_alc", "sstati") # Пример переменных
target <- "tlfbh2" # Зависимая переменная

# Расчет частных корреляций
partial_corrs <- list()
for (var in features) {
```

```

controls <- setdiff(features, var)
result <- calculate_partial_correlation(my_data, var, target, controls)
partial_corrs[[var]] <- result
}

# Вывод результатов
for (var in names(partial_corrs)) {
  coef <- partial_corrs[[var]]$correlation
  p_value <- partial_corrs[[var]]$p_value
  cat(sprintf("Признак: %s, \t коэффициент корреляции: %.4f, \t p-value (значимость): %.4f\n", var, coef, p_value))
}

## Признак: rubsex, коэффициент корреляции: 0.1216, p-value (значимость): 0.0427
## Признак: tlfba2, коэффициент корреляции: 0.0586, p-value (значимость): 0.3307
## Признак: asi3_alc, коэффициент корреляции: 0.0028, p-value (значимость): 0.9623
## Признак: sstati, коэффициент корреляции: -0.0070, p-value (значимость): 0.9074

```

Эти результаты указывают на следующее:

1. **rubsex**: С коэффициентом корреляции 0.1216 и р-значением 0.0427, эта переменная показывает статистически значимую слабую положительную связь с зависимой переменной `tlfbh2`. Это означает, что при увеличении `rubsex` на единицу, `tlfbh2` также имеет тенденцию увеличиваться, хотя и не очень сильно.
2. **tlfba2**: С коэффициентом корреляции 0.0586 и р-значением 0.3307, связь между `tlfba2` и `tlfbh2` слабая и статистически не значима. Это говорит о том, что изменения в `tlfba2` не влияют на `tlfbh2` в статистически значимой степени.
3. **asi3\_alc**: С очень низким коэффициентом корреляции 0.0028 и р-значением 0.9623, влияние `asi3_alc` на `tlfbh2` практически нулевое и статистически не значимо.
4. **sstati**: Отрицательный коэффициент корреляции -0.0070 и высокое р-значение 0.9074 указывают на отсутствие значимой связи между `sstati` и `tlfbh2`.

Из этих результатов можно сделать вывод, что только переменная `rubsex` имеет значимое влияние на `tlfbh2`. Это важно учитывать при моделировании и анализе данных, где `rubsex` может быть рассмотрен как потенциальный фактор, влияющий на изменения в `tlfbh2`. Остальные переменные, вероятно, не имеют значимого вклада в изменения этой зависимой переменной в данном наборе данных.