

The background is a light teal color. In the top left, there are orange brushstroke-like lines and a yellow star-like shape. In the top right, there is a teal cloud-like shape. In the bottom right, there is a yellow star-like shape and a teal cloud-like shape. The central text is on a white, rounded rectangular background. Surrounding the text are illustrations of a soccer ball, a basketball, a bowling pin, and a badminton shuttlecock.

# **LESÕES E INTENSIDADE DE TREINAMENTOS**

Lorenzo Lima e Mariana Coin



# Contextualização



**Banco de dados: “Athlete Injury and Performance Dataset”**

**O banco de dados escolhido analisa o impacto de algoritmos complexos de agendamento nas taxas de lesões e no desempenho atlético em um ambiente esportivo universitário.**

**Captura as informações demográficas, regimes de treinamento, cronogramas, níveis de fadiga e riscos de lesão dos atletas.**



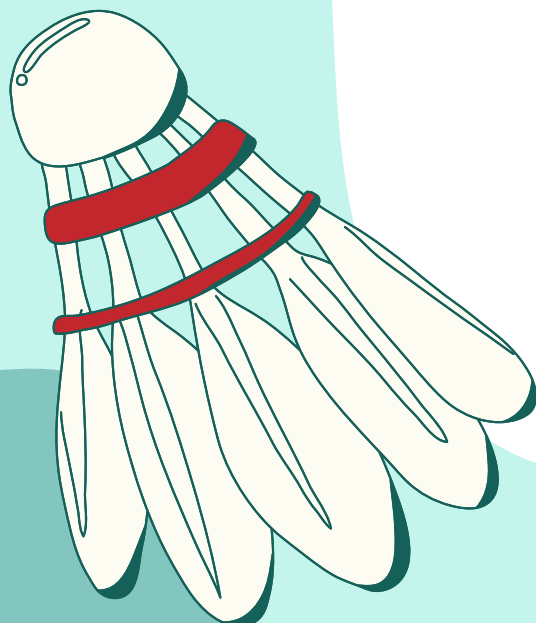
**Athlete Injury and Performance Dataset**

Dataset · 14d ago · by [Ziya](#)

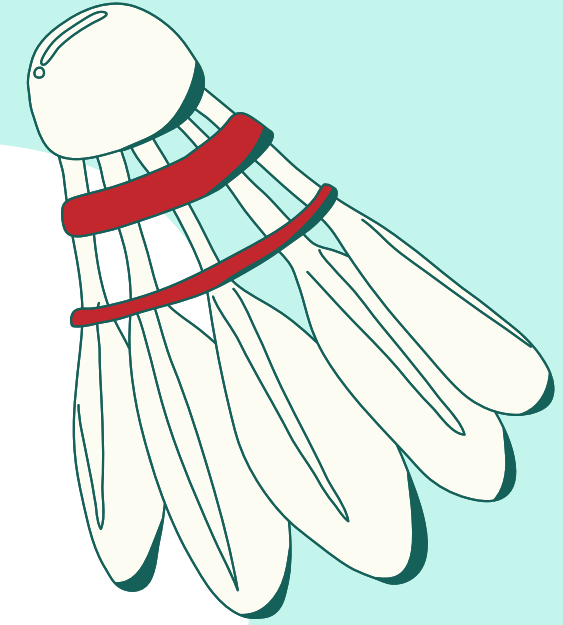
[Injury](#) Information InjuryIndicator: Target column indicating whether the [athlete](#) sustained an ACL [injury](#).

5

195 downloads



# Objetivo de pesquisa



**Aplicaremos a técnica de regressão linear para investigar os fatores que influenciam as taxas de lesões e o desempenho atlético em um ambiente de esportes universitários.**



# Variáveis utilizadas

**“ACL\_Risk\_Score”**

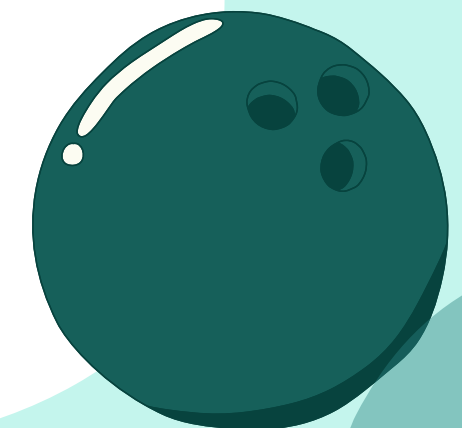
**Pontuação de risco prevista para lesões no ligamento cruzado anterior (LCA).**

**“Training\_Intensity”**

**Intensidade média das sessões de treinamento.**

**“Training\_Hours\_Per\_Week”**

**Total de horas de treinamento por semana.**



# Análise Gráfica

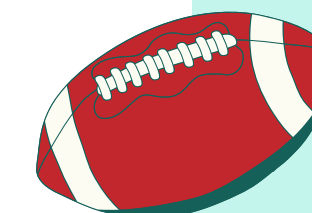
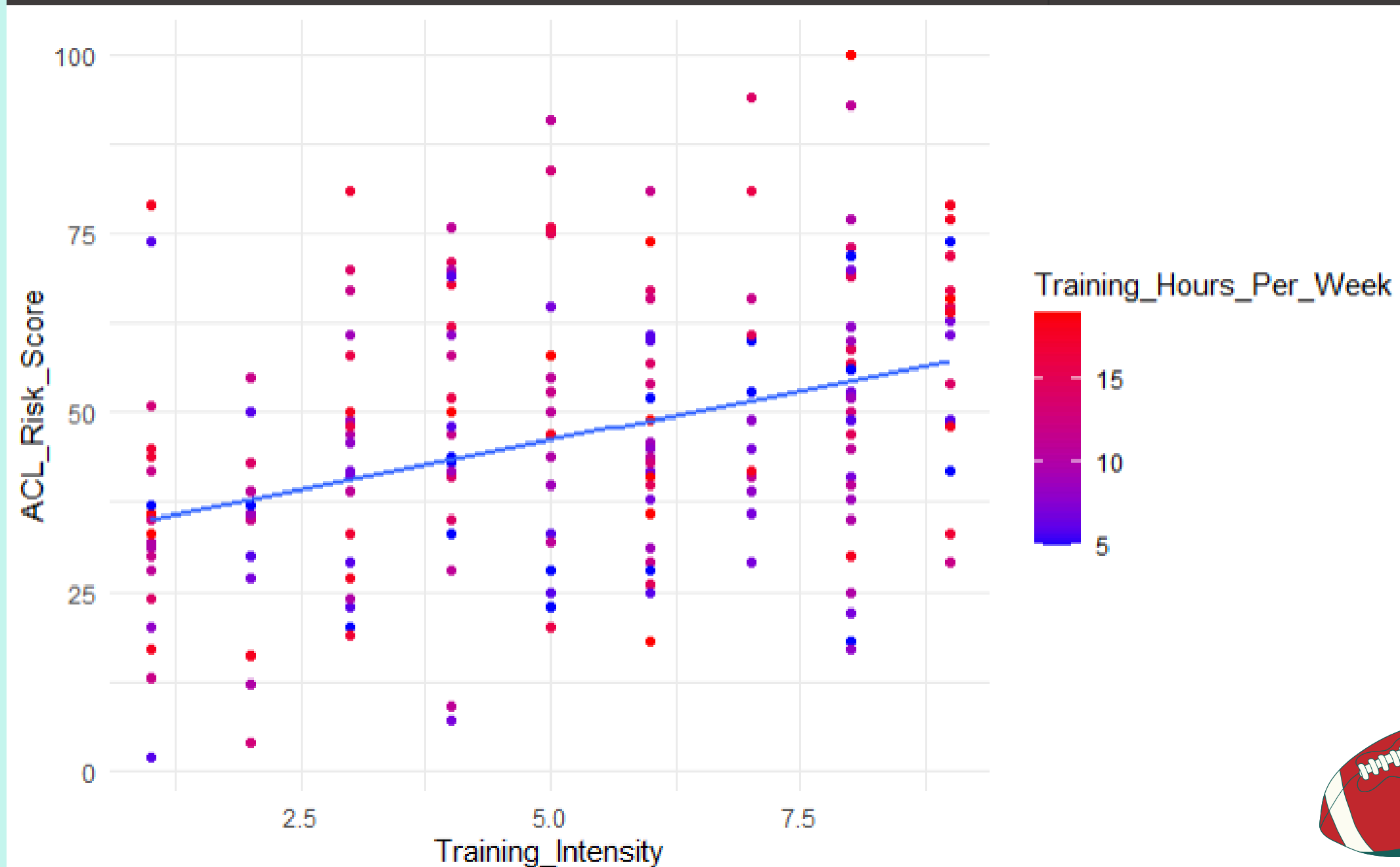


```
```{r}

modelo = lm(ACL_Risk_Score ~ Training_Intensity + Training_Hours_Per_Week, data = dados)

ggplot(dados, aes(x = Training_Intensity, y = ACL_Risk_Score, color = Training_Hours_Per_Week)) +
  geom_point() +
  geom_smooth(method = lm, se = FALSE)+
  scale_color_gradient(low = "blue", high = "red") +
  theme_minimal()

```
```



# Análise Descritiva

Utilizando a função “summary” no R, podemos obter:

**Intercept = 21.8521; Training\_Intensity = 2.777; Training\_Hours\_Per\_Week = 0.9228**

**P-valores individuais:**

**Intercept = 0.0000005461; Training\_Intensity = 0.0000000505; Training\_Hours\_Per\_Week = 0.000981**

**P-valor da estatística F:      R<sup>2</sup> ajustado:**  
**0.000000004242                      0.1694**



Coefficients:

|                         | Estimate | Std. Error | t value | Pr(> t )     |     |
|-------------------------|----------|------------|---------|--------------|-----|
| (Intercept)             | 21.8521  | 4.2185     | 5.180   | 0.0000005461 | *** |
| Training_Intensity      | 2.7770   | 0.4898     | 5.669   | 0.0000000505 | *** |
| Training_Hours_Per_Week | 0.9228   | 0.2758     | 3.346   | 0.000981     | *** |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.26 on 197 degrees of freedom

Multiple R-squared: 0.1778, Adjusted R-squared: 0.1694

F-statistic: 21.29 on 2 and 197 DF, p-value: 0.000000004242

# Análise Descritiva

## Interpretações:

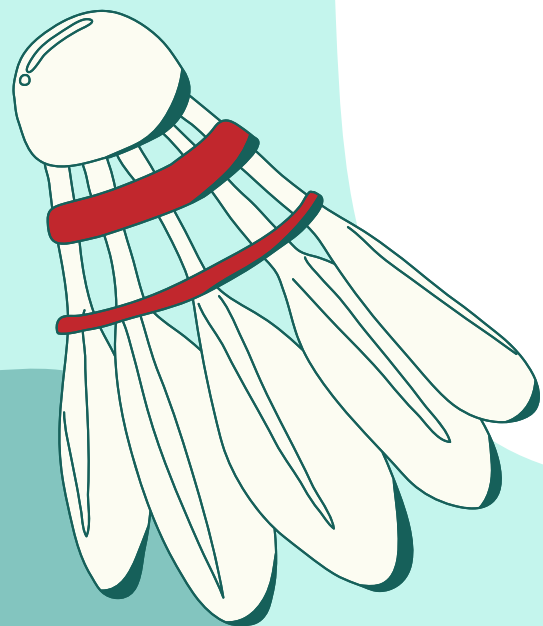
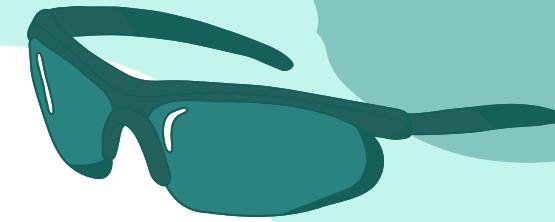
O valor esperado da variável “ACL\_Risk\_Score” (escore de risco de lesão no LCA) quando todas as outras variáveis são zero é 21.8521.

Para cada unidade adicional na intensidade do treinamento, em média, Y aumenta em 2.777.

Para cada hora adicional de treinamento por semana, em média, Y aumenta em 0.9228.

Aproximadamente 16,94% da variabilidade dos dados é explicada pelo modelo.

A partir do p-valor de 0.000000004242, podemos dizer que há evidências de que essas variáveis explicam variações no escore de risco de lesão no LCA



# Avaliação de Multicolinearidade

Para observar se existe ou não multicolinearidade entre as componentes, podemos utilizar o cálculo do VIF.

Se o VIF é próximo de 1, temos indicações de que não existe multicolinearidade.

No R:

```
library(car)
kable(vif(modelo))
```

|                         | x        |
|-------------------------|----------|
| Training_Intensity      | 1.000409 |
| Training_Hours_Per_Week | 1.000409 |

Percebemos que o VIF de ambas as variáveis é dado pelo valor 1.000409, muito próximo de 1, portanto, as indicações dizem que não existe multicolinearidade, como comentado acima.





# Testes de normalidade

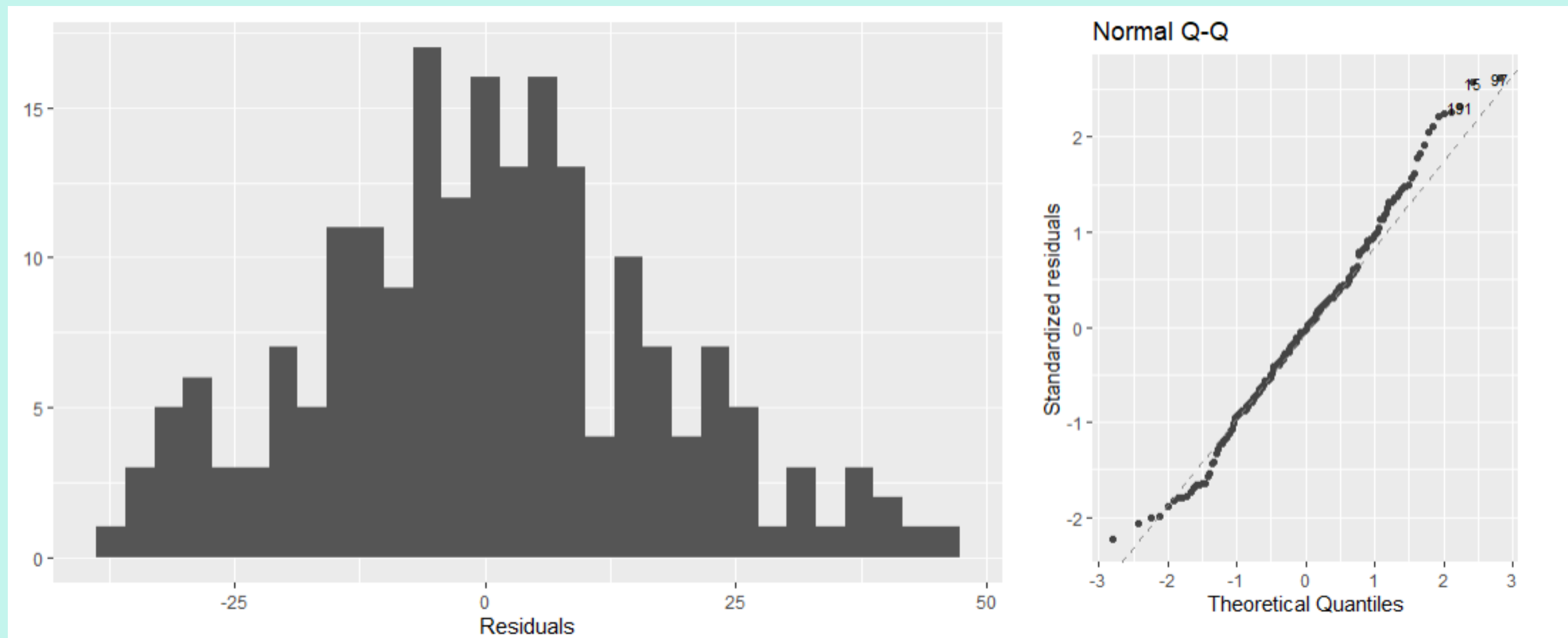
Os seguintes testes de normalidade foram realizados para avaliar os resíduos de forma quantitativa:

- Shapiro-Wilk
- Kolmogorov-Smirnov
- Anderson-Darling
- Cramer-Von-Mises

| Testes de normalidade |           |           |
|-----------------------|-----------|-----------|
|                       | Statistic | P.value   |
| Shapiro-Wilk          | 0.9906444 | 0.2214996 |
| Kolmogorov-Smirnov    | 0.4420274 | 0.0000000 |
| Anderson-Darling      | 0.3586942 | 0.4484754 |
| Cramer-Von-Mises      | 0.0483848 | 0.5319471 |

Os testes de Shapiro-Wilk, Anderson-Darling e Cramer-Von-Mises nos trouxeram um p-valor relativamente alto. Portanto, aceitamos a hipótese nula de que os dados possuem uma distribuição normal.

# Testes de normalidade

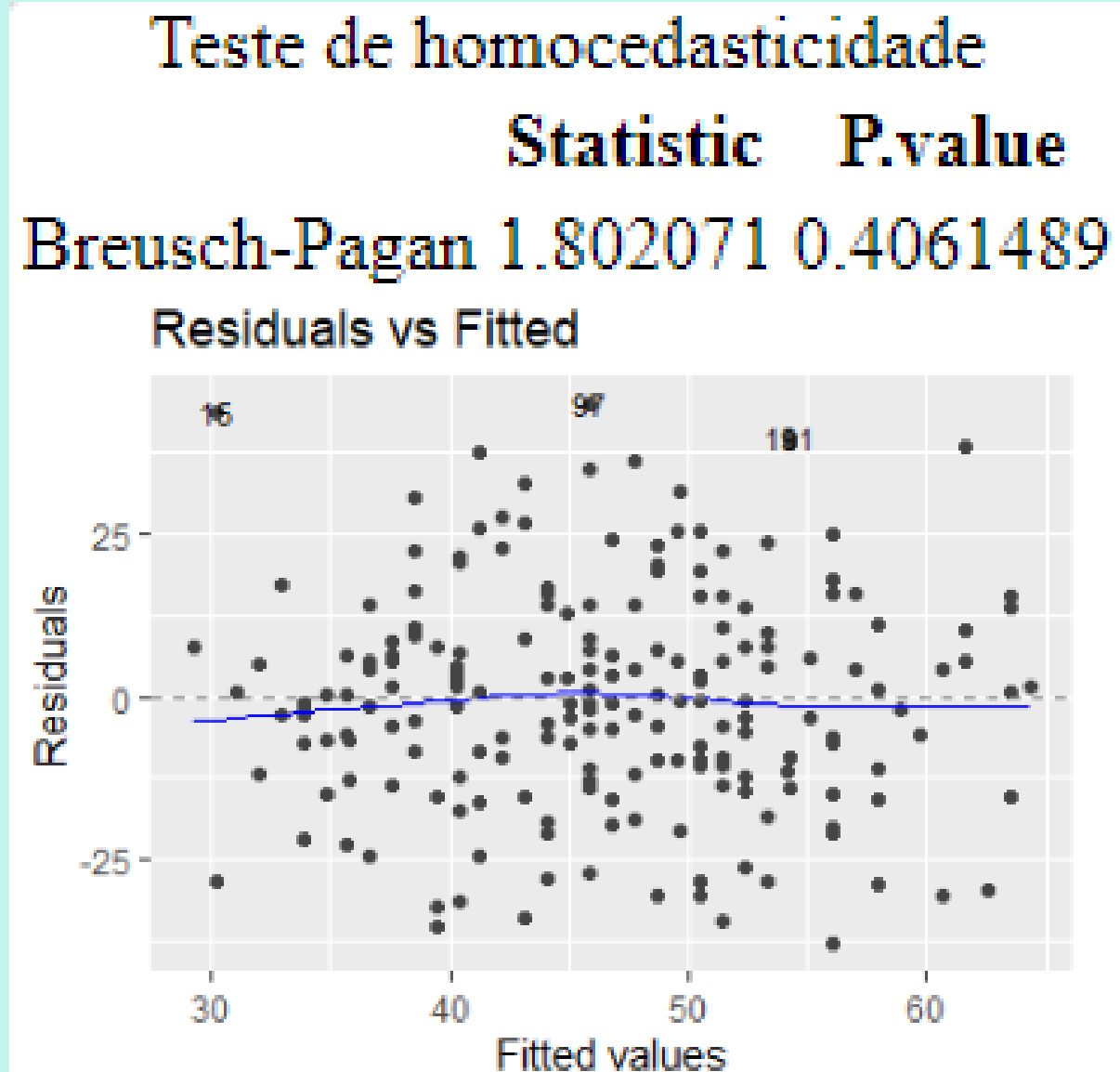


**Graficamente, também podemos assumir que a premissa de normalidade foi atendida.**

# Testes de homocedasticidade



Para avaliar a homocedasticidade, utilizamos os testes de Breusch-Pagan e analisando o gráfico de resíduos versus valores ajustados.



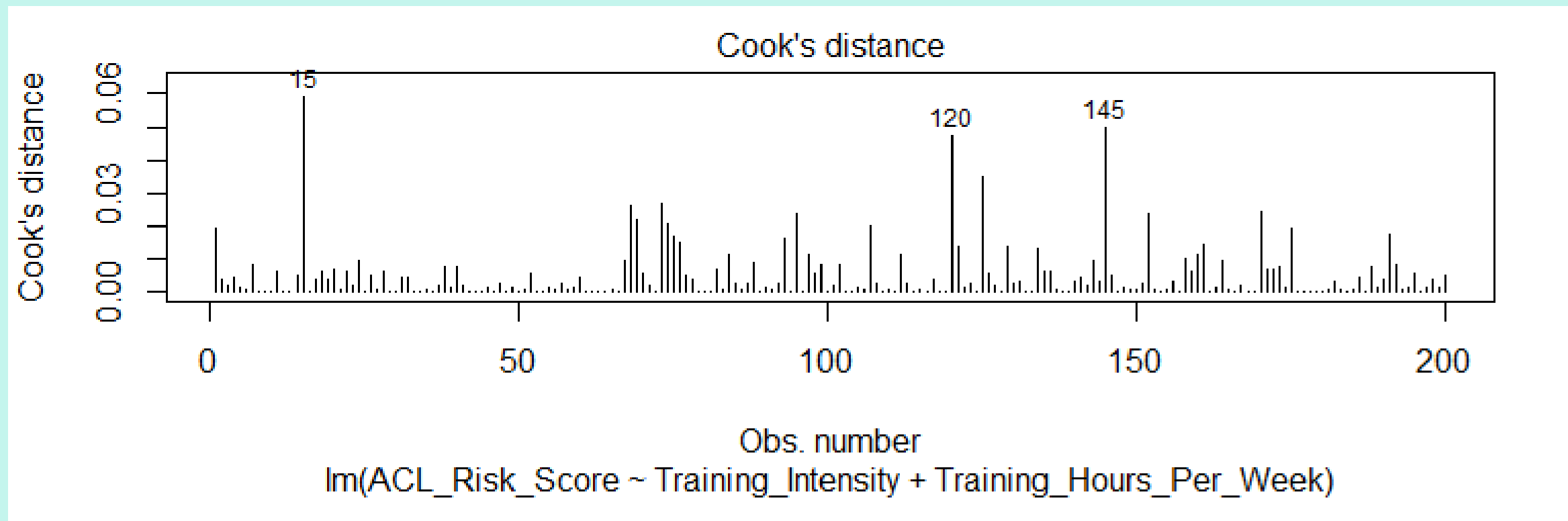
O p-valor obtido no teste de Breusch-Pagan foi de 0.4061, indicando que não há evidências suficientes para rejeitar a hipótese nula de homocedasticidade.

Portanto, os resíduos podem ser considerados homocedásticos.

O gráfico de resíduos vs. valores ajustados foi gerado apresentando os resíduos distribuídos de maneira relativamente aleatória em torno de uma linha de tendência próxima de zero. Também não foram observadas evidências significativas de heterocedasticidade no modelo.

# Outliers

Podemos utilizar o gráfico de distância de Cook para observar a presença ou não de outliers.



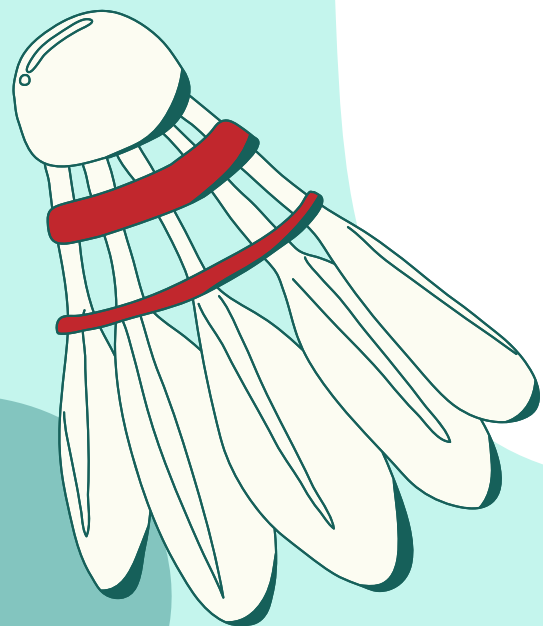
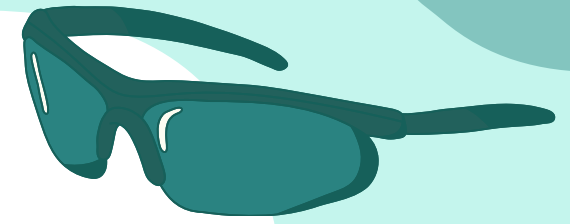
Neste âmbito, podemos perceber que as observações 15, 120 e 145 são outliers.

# Conclusões finais

**A intensidade do treinamento e as horas semanais de treinamento foram identificadas como variáveis significativamente associadas ao risco de lesão no ligamento cruzado anterior.**

**Testes de normalidade e homocedasticidade indicaram que o modelo atende, em geral, às premissas da regressão múltipla.**

**Os resultados ressaltam a importância de balancear intensidade e duração do treinamento para minimizar riscos e melhorar o desempenho dos atletas.**



**Obrigado!**

