

# **ESTATÍSTICA PARA SAÚDE COLETIVA**

## **Aula 13**

# Seminários de hoje

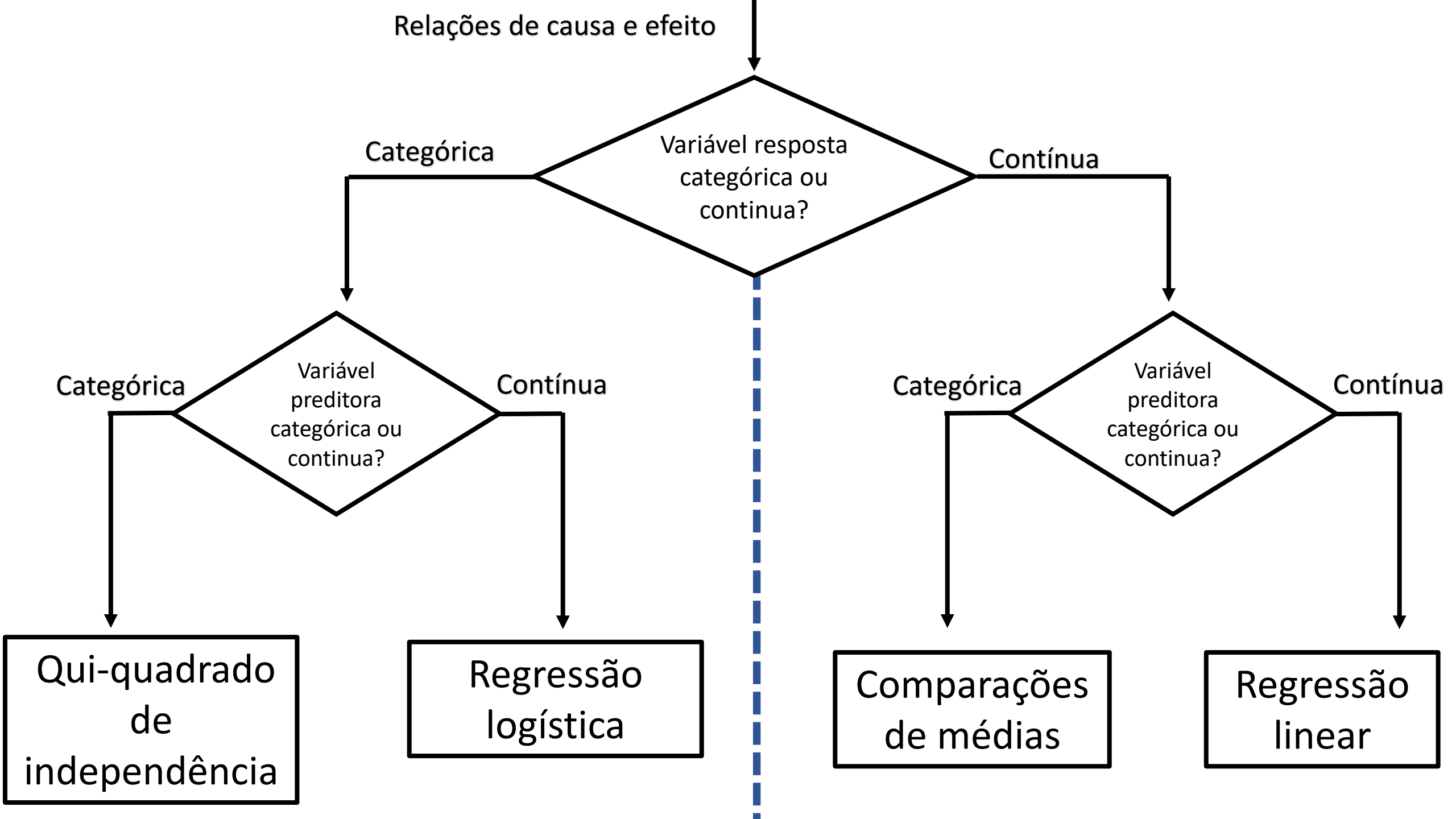
Nomes (Aluno ou Dupla)	Nome do artigo
EDNEIDE LOURENÇO + JOZIÊLDA ALVES	Qualidade de vida de mulheres com gravidez de alto risco durante o cuidado pré-natal
RAFAELA REIMBERG	Quedas em idosos institucionalizados: riscos, consequências e antecedentes
THAYLA NOLETO + RAFAELA PEREIRA	A saúde mental da enfermagem no enfrentamento da COVID-19 em um hospital universitário regional

# Próxima aula

Nomes (Aluno ou Dupla)	Data seminário	Nome do artigo
ANDERSON SILVA	15/10/2020	Características da epidemia de dengue em Pinhalzinho, Santa Catarina.
ROSA MARIA GARCIA + ANA LEITE	15/10/2020	Consumo de alimentos fora do lar no Brasil segundo locais de aquisição
TAINARA PRADELLA + ADRIANA BARROS	15/10/2020	Doenças Crônicas Não Transmissíveis e fatores de risco e proteção em adultos com ou sem plano de saúde
PAMELA MONTE CRUZ	15/10/2020	Prevalência de transtornos mentais comuns e fatores associados em moradores da área urbana de São Paulo, Brasil

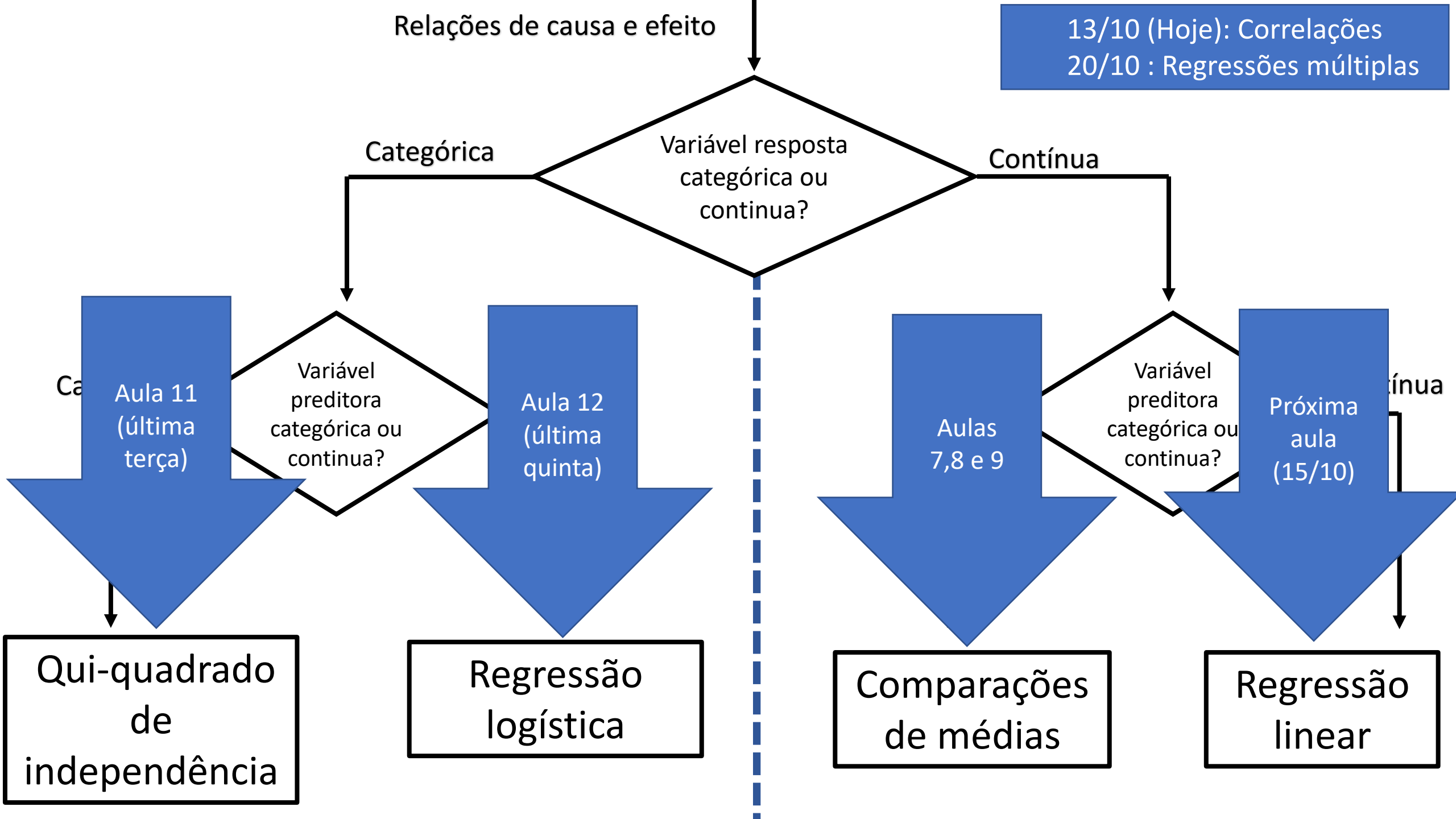
Revisão + feedback lista 11 e 12

# Teste de qui-quadrado de independência

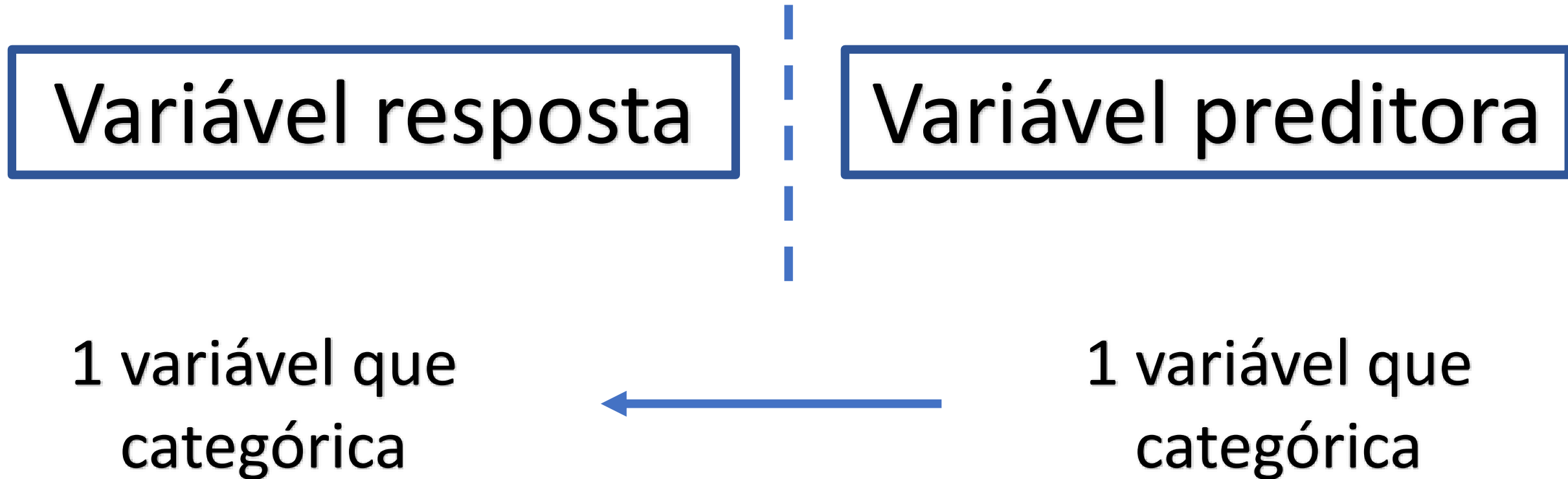


Relações de causa e efeito

13/10 (Hoje): Correlações  
20/10 : Regressões múltiplas



# Qui-quadrado de independência





Pergunta: A probabilidade de registro de obesidade depende do sexo do paciente?

Sexo	Obesidade		Total	Frequência obesidade
	Sim	Não		
Masculino	30	70	100	30%
Feminino	20	80	100	20%
Total	50	150	200	25%

# Lista 11 pergunta 4

Pergunta 4. O uso de álcool-gel (Coluna Alcool) afeta o risco de se contrair COVID (Coluna COVID)? Assuma um nível de significância de 5%.

1  pontos

- ☐ Não há associação entre as variáveis
- ☒ Há (SIM) associação entre as variáveis
- ☐ Não consegui responder essa pergunta porque não entendi como aplicar o teste.
- ☐ Não consegui responder essa pergunta porque não tenho computador disponível para executar o teste.
- ☐ Não sei/Não quero responder essa pergunta.

# Script

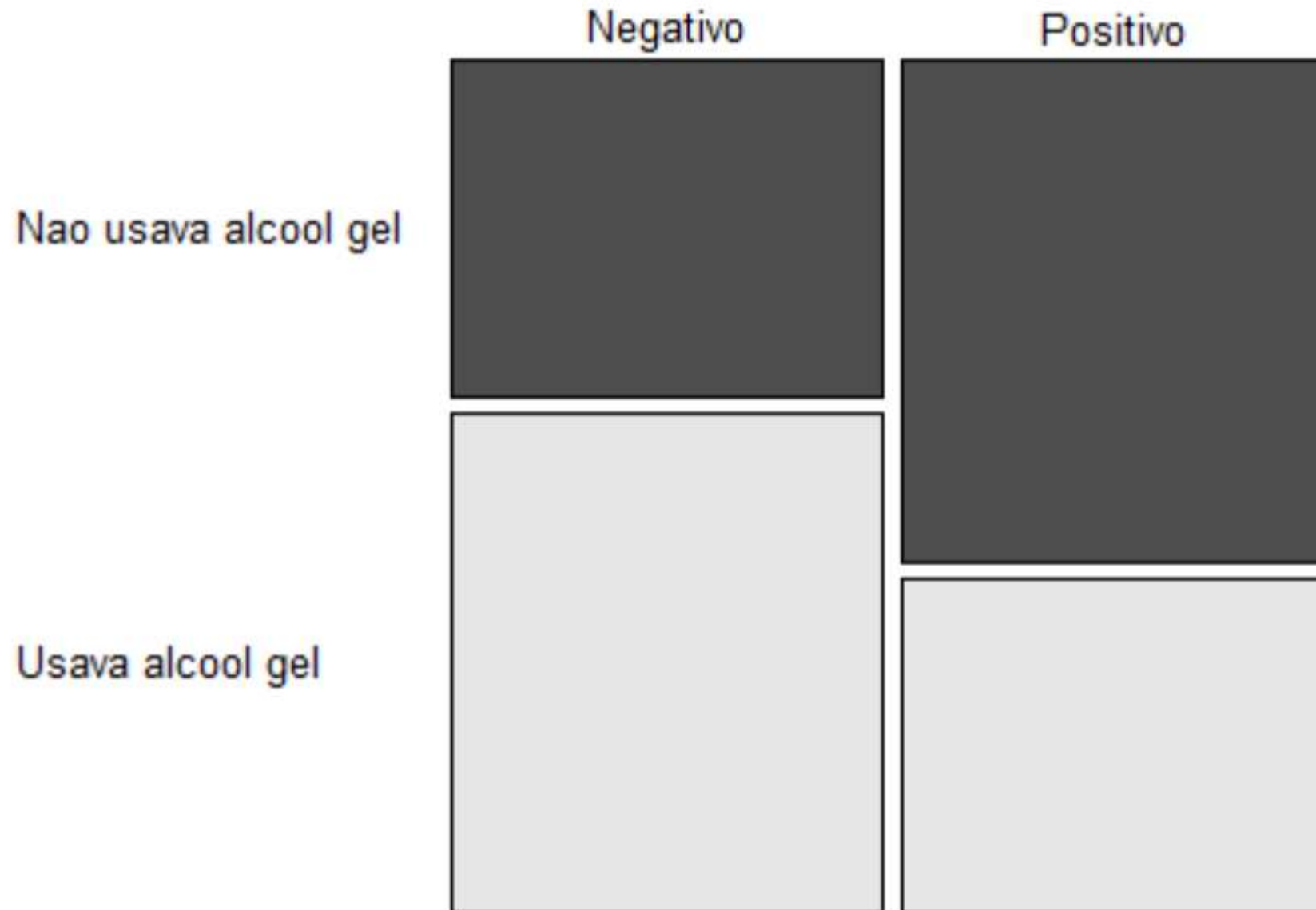
```
# Exercício 4
```

```
TabCont = table(tabela2$COVID,tabela2$Alcool)
```

```
mosaicplot(TabCont, color=TRUE, las = 1, cex.axis=0.8, main =NULL)
```

```
chisq.test(TabCont)
```

# Lista 11 pergunta 4



# Lista 11 pergunta 4

Pearson's Chi-squared test with Yates' continuity correction

data: TabCont

X-squared = 7.22, df = 1, p-value = 0.00721

Resultado: P menor que 0.05

Conclusão:  $H_1$  é verdadeira

Hipóteses testadas


$H_0$ : Não há associação entre as variáveis


$H_1$ : Há (SIM) associação entre as variáveis

Portanto:

Existe evidências de o uso de álcool gel modifica a probabilidade de se contrair COVID.

# Lista 11 pergunta 5

Pergunta 5. O risco de se contrair COVID (Coluna COVID), apresenta alguma associação com a profissão do paciente (Coluna Profissao)? Assuma um nível de significância de 5%. 1  pontos

- ☒ Não há associação entre as variáveis 
- ☐ Há (SIM) associação entre as variáveis
- ☐ Não consegui responder essa perguntar porque não entendi como aplicar o teste.
- ☐ Não consegui responder essa perguntar porque não tenho computador disponível para executar o teste.
- ☐ Não sei/Não quero responder essa pergunta.

# Script

```
# Exercício 5
```

```
TabCont = table(tabela2$COVID,tabela2$Profissao)
```

```
mosaicplot(TabCont, color=TRUE, las = 1, cex.axis=0.8, main =NULL)
```

```
chisq.test(TabCont)
```

# Lista 11 pergunta 5

	Negativo	Positivo
Comerciante		
Enfermeiro		
Professor		



# Lista 11 pergunta 5

Pearson's Chi-squared test

data: TabCont

X-squared = 0.56139, df = 2, p-value = 0.7553

Resultado: P maior que 0.05  
Conclusão:  $H_0$  é verdadeira

Hipóteses testadas

$H_0$ : Não há associação entre as variáveis

$H_1$ : Há (SIM) associação entre as variáveis

Portanto:

Não existe evidências de haja diferença na probabilidade de COVID em relação as profissões estudadas.

Regressão logística

# Regressão logística

**Variável resposta**

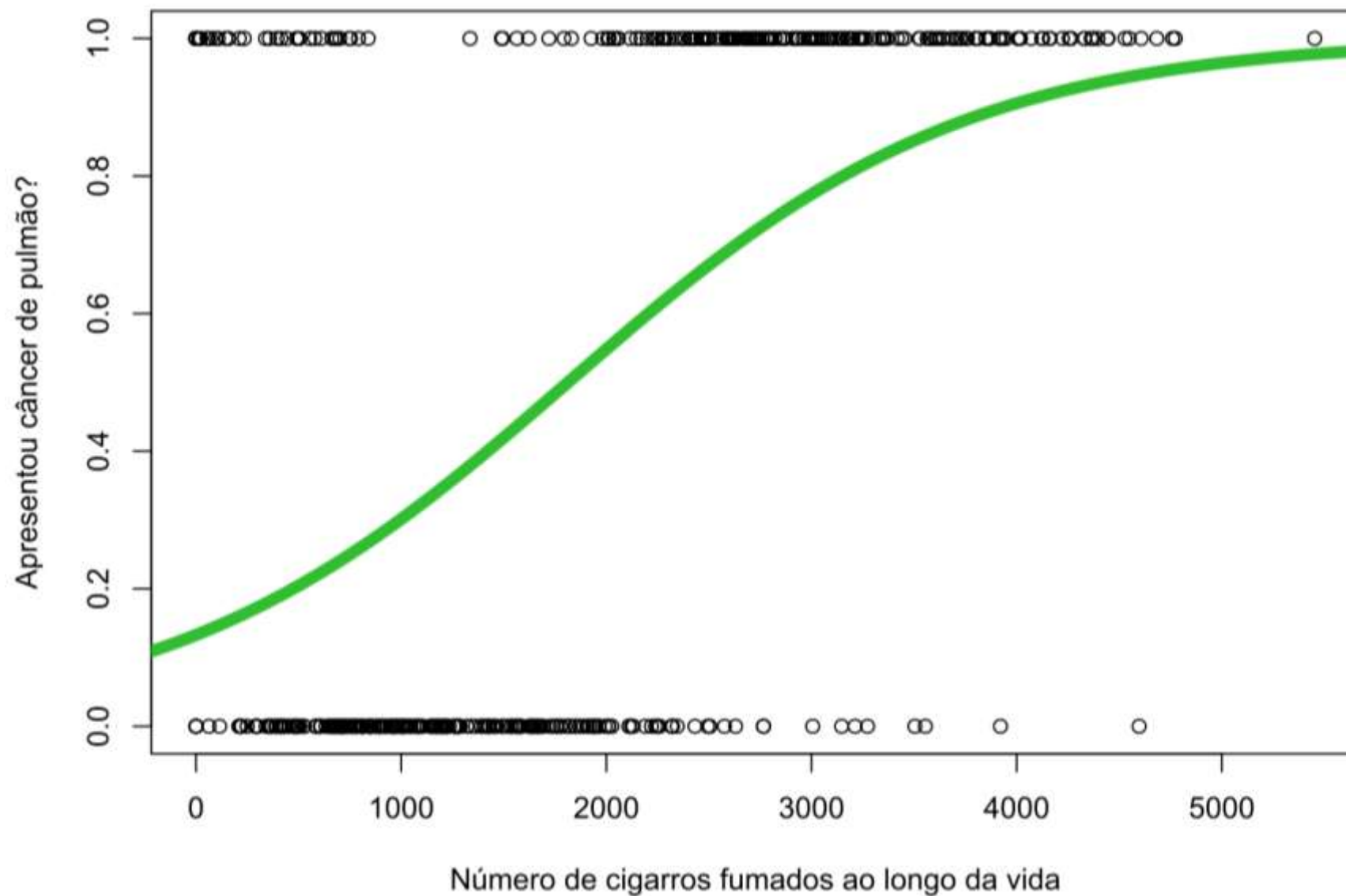
**Variável preditora**

1 variável que  
categórica

1 variável que  
contínua




# Regressão logística



# Lista 12 pergunta 5

Pergunta 5. Avalie se a quantidade de horas de atividades físicas praticadas antes da pandemia (Coluna AtividadePrevia), influência no risco de morte por COVID em pacientes internados em UTI (Coluna Morte). Para tal, aplique uma regressão logística para avaliar a relação de causa e efeito entre essas variáveis. Assumindo um nível de significância de 5%, é possível afirmar que:

1  pontos

☐ Não há associação entre as variáveis

☒ Há (SIM) associação entre as variáveis



☐ Não consegui responder essa pergunta porque não entendi como aplicar o teste.

☐ Não consegui responder essa pergunta porque não tenho computador disponível para executar o teste.

☐ Não sei/Não quero responder essa pergunta.

 [Adicionar feedback da resposta](#)

---

# Script

```
# Exercício 5
```

```
Form <- formula(Morte ~ AtividadePrevia)
```

```
B1 <- glm(Form, family = binomial, data = tabela1)
```

```
summary(B1)
```

# Lista 12 pergunta 5

```
> summary(B1)
```

Call:

```
glm(formula = Form, family = binomial, data = tabela1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3838	-0.6907	-0.3775	0.1869	2.8309

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.4731	0.2802	1.688	0.0913 .
AtividadePrevia	-0.8923	0.1549	-5.762	8.29e-09 ***

---

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

# Lista 12 pergunta 5

Portanto:

Existe evidências de que a prática de atividade antes da pandemia influencia no risco de morte, e essa relação pode ser prevista pelo modelo de regressão logística proposto.

Resultado: P maior que 0.05  
Conclusão:  $H_1$  é verdadeira

Hipóteses testadas

$H_0$ : Não há associação entre as variáveis

$H_1$ : Há (SIM) associação entre as variáveis

Coefficients:

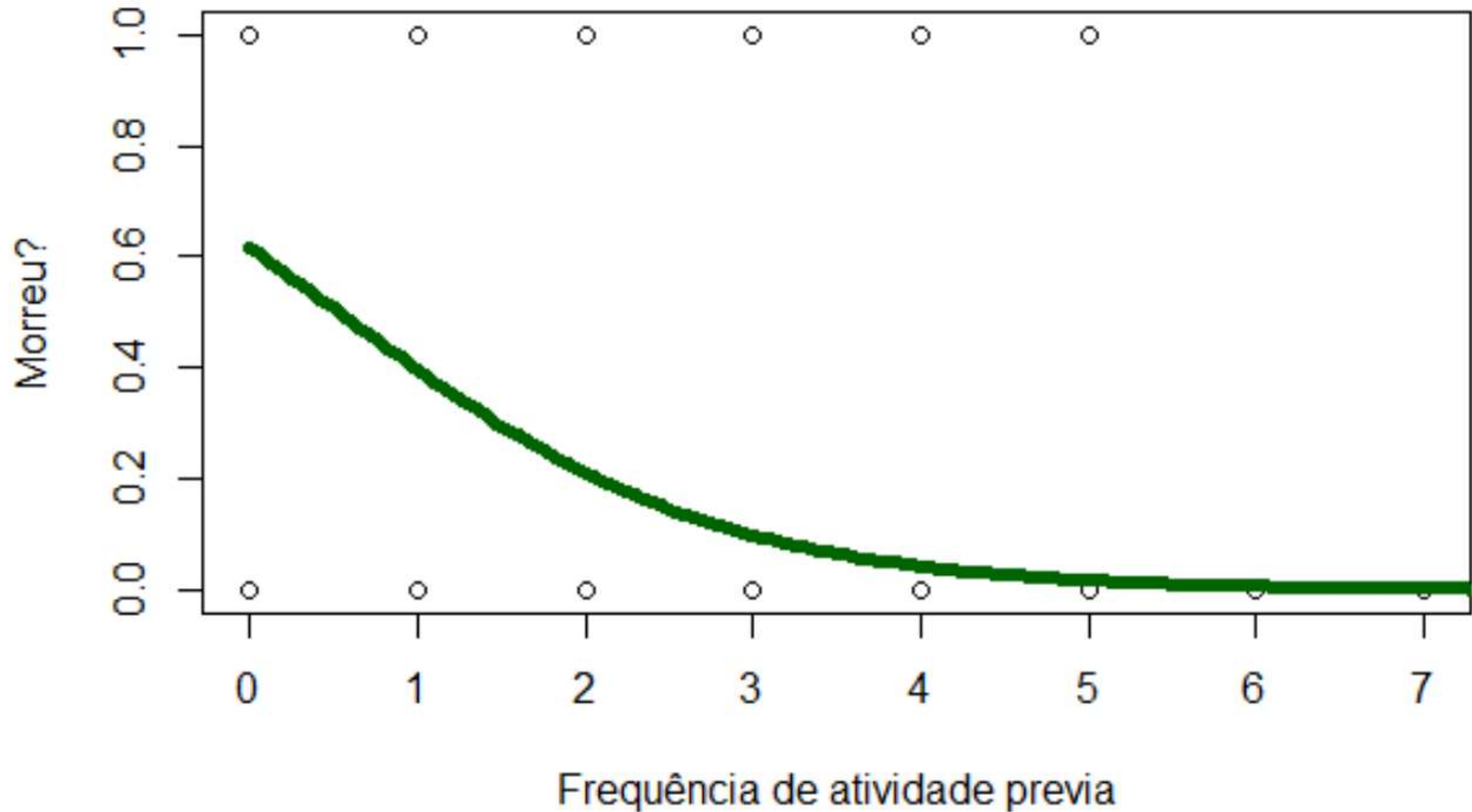
	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.4731	0.2802	1.688	0.0913 .
AtividadePrevia	-0.8923	0.1549	-5.762	8.29e-09 ***

---

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




## Lista 12 pergunta 5



# Lista 12 pergunta 6

Pergunta 6. Teste se o número de horas gastos em redes sociais (Coluna HorasRedeSocial), influência jovens e adultos a buscarem procedimentos estéticos (Coluna ProcedimentoEstatico). Para tal, aplique uma regressão logística para avaliar a relação de causa e efeito entre essas variáveis. Assumindo um nível de significância de 5%, é possível afirmar que:

1  pontos

- ☐ Não há associação entre as variáveis
- ☒ Há (SIM) associação entre as variáveis 
- ☐ Não consegui responder essa pergunta porque não entendi como aplicar o teste.
- ☐ Não consegui responder essa pergunta porque não tenho computador disponível para executar o teste.
- ☐ Não sei/Não quero responder essa pergunta.

 [Adicionar feedback da resposta](#)

# Script

```
# Exercício 6
```

```
Form <- formula(ProcedimentoEstetico ~ HorasRedeSocial)
```

```
B1 <- glm(Form, family = binomial, data = tabela1)
```

```
summary(B1)
```

# Lista 12 pergunta 6

```
> summary(B1)
```

Call:

```
glm(formula = Form, family = binomial, data = tabela1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.49720	-0.39407	-0.20267	-0.04141	2.82479

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-6.7216	0.9362	-7.180	6.97e-13	***
HorasRedeSocial	0.7374	0.1160	6.356	2.08e-10	***

---

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

# Lista 12 pergunta 6

Portanto:

Existe evidências de que o número de horas que uma pessoa passa nas redes sociais influencia na chance dessa pessoa realiza um procedimento estetico, e essa relação pode ser prevista pelo modelo de regressão logística proposto.

Resultado: P maior que 0.05  
Conclusão:  $H_1$  é verdadeira

Hipóteses testadas

$H_0$ : Não há associação entre as variáveis

$H_1$ : Há (SIM) associação entre as variáveis

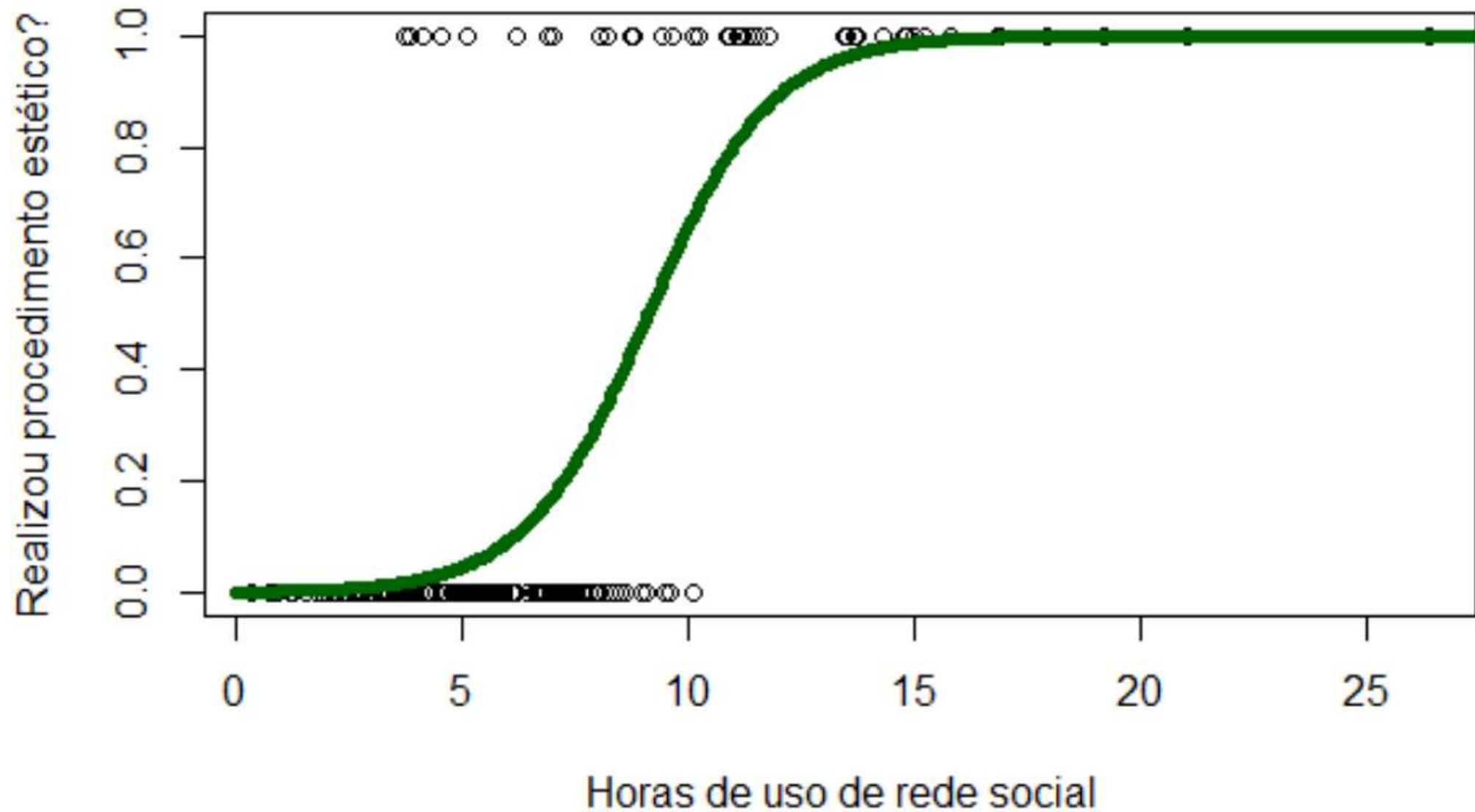
Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-6.7216	0.9362	-7.180	6.97e-13	***
HorasRedeSocial	0.7374	0.1160	6.356	2.08e-10	***

---

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

# Lista 12 pergunta 6



Aula de hoje

# Correlações



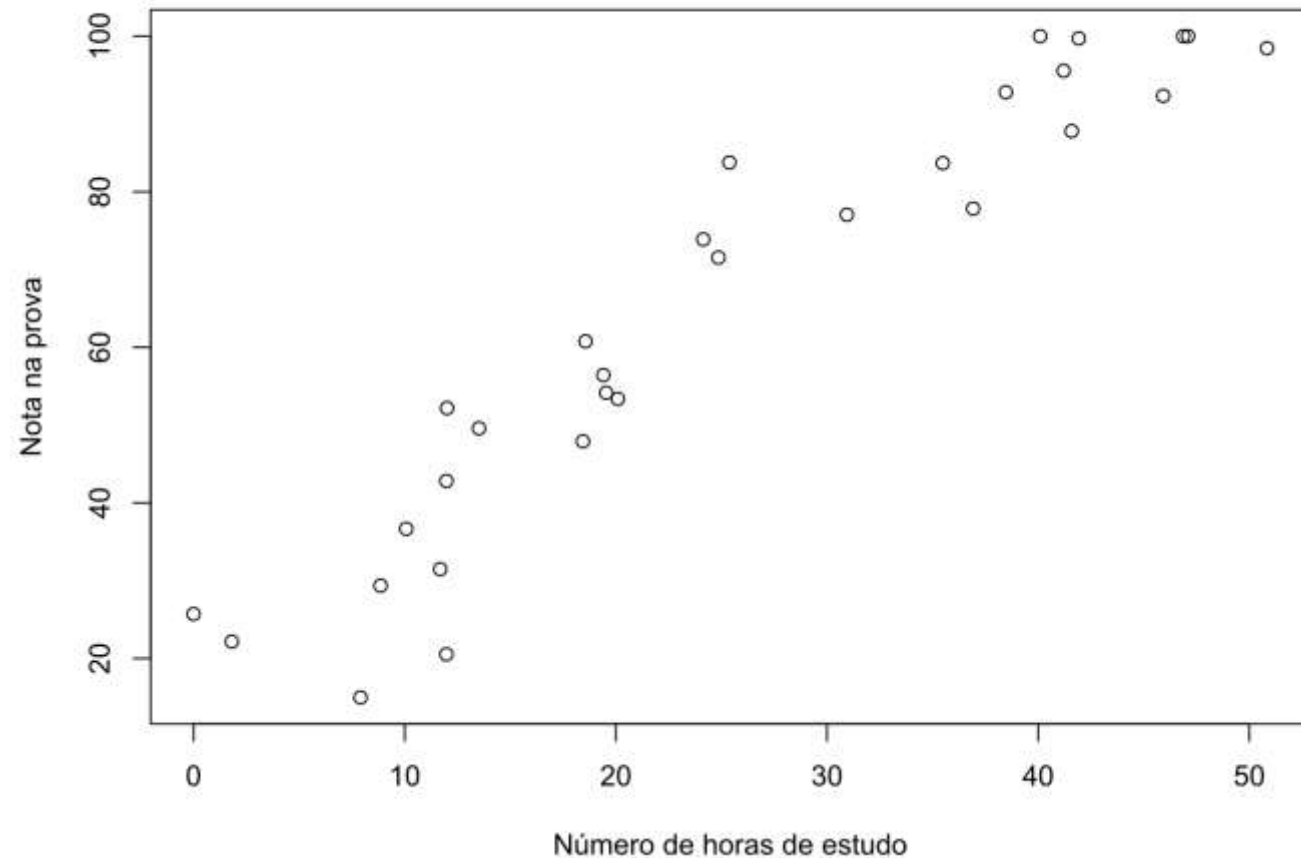
O que são correlações?

# O que são correlações?

- Correlação significa uma semelhança ou relação entre duas coisas.
- Correlação pode ser traduzida como o “andar junto” das variáveis
- A quantidade de sorvete vendido tem correlação com a temperatura do dia?

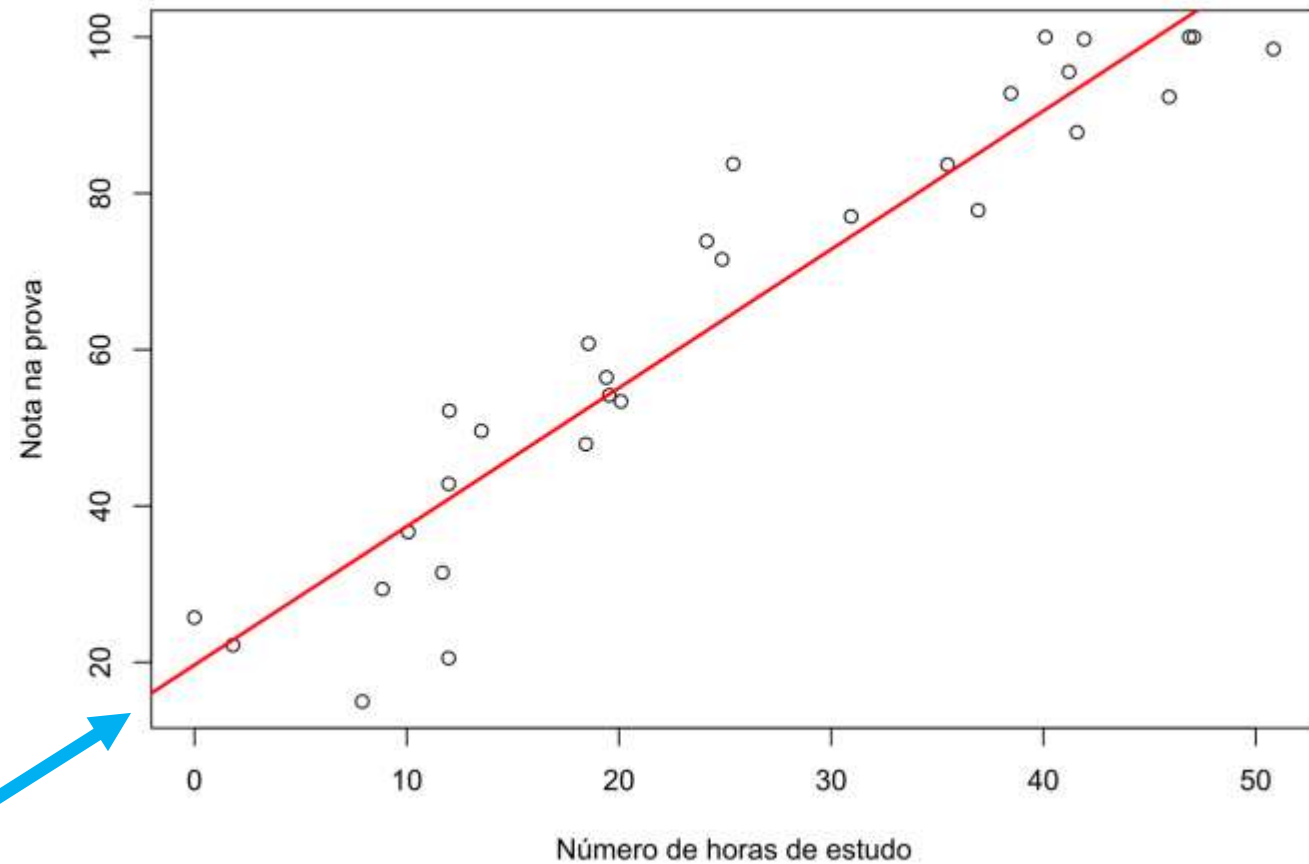
# Exemplo

- Compare o numero de horas de estudo vs. nota em um teste
- Os pontos indicam a nota e numero de horas de estudo para cada aluno



# Exemplo

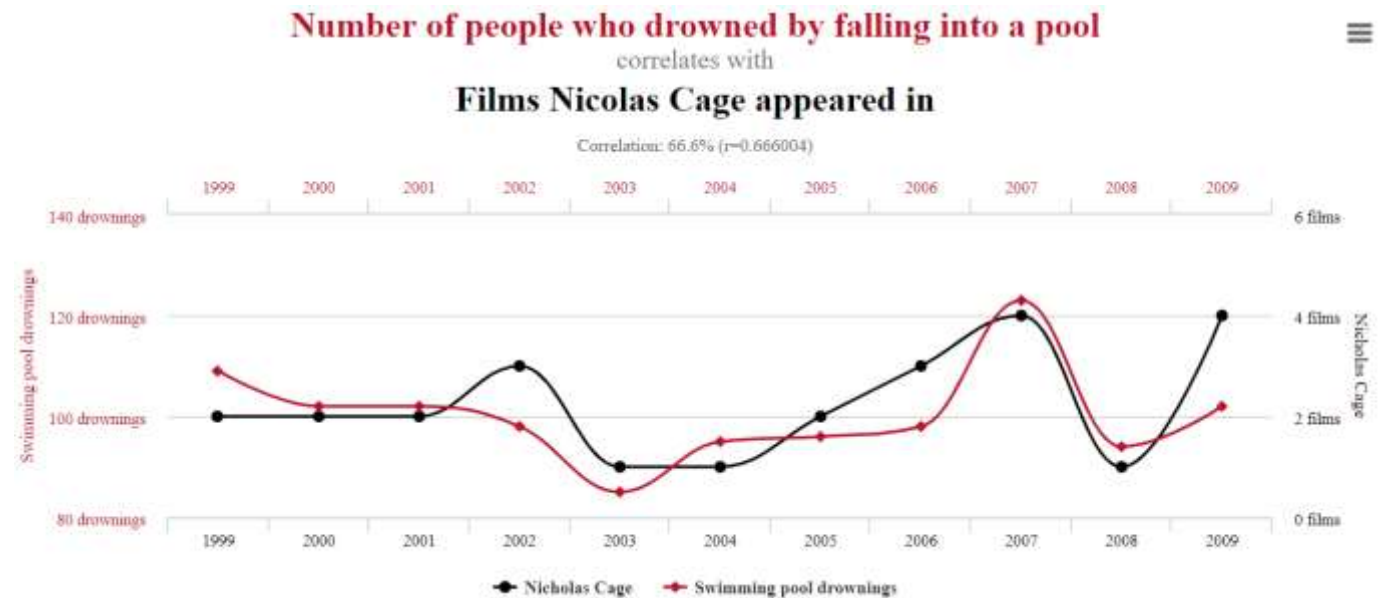
- Compare o numero de horas de estudo vs. nota em um teste
- Os pontos indicam a nota e numero de horas de estudo para cada aluno



Quanto mais próximo os dados estiverem  
dessa reta imaginaria, maior será a correlação

# Correlação não implica em causa e efeito

- Correlação  $\neq$  Casualidade
  - O fato de duas variáveis serem correlacionadas não implica que uma variável está afetando diretamente a outra. Pode ser que exista um 3 elemento que afeta as duas variáveis. Exemplos:
    - Número de afogamentos em piscina e numero de lançamentos de filmes que o Nicolas Cage é parte do elenco

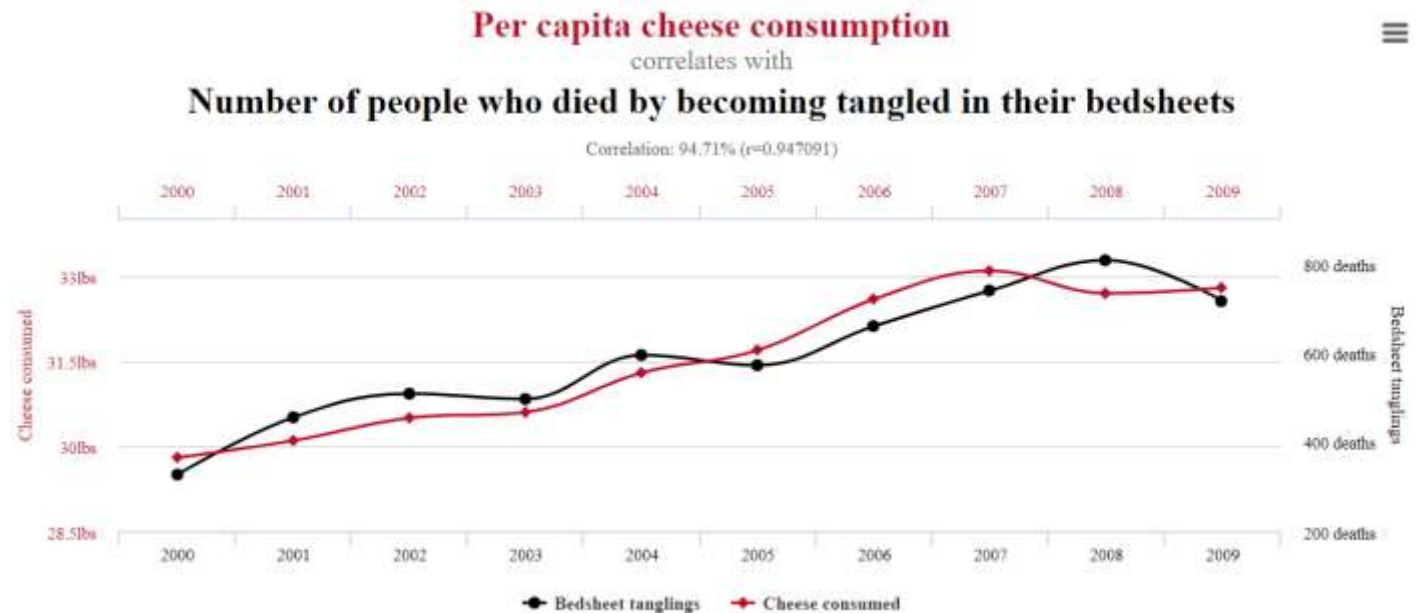
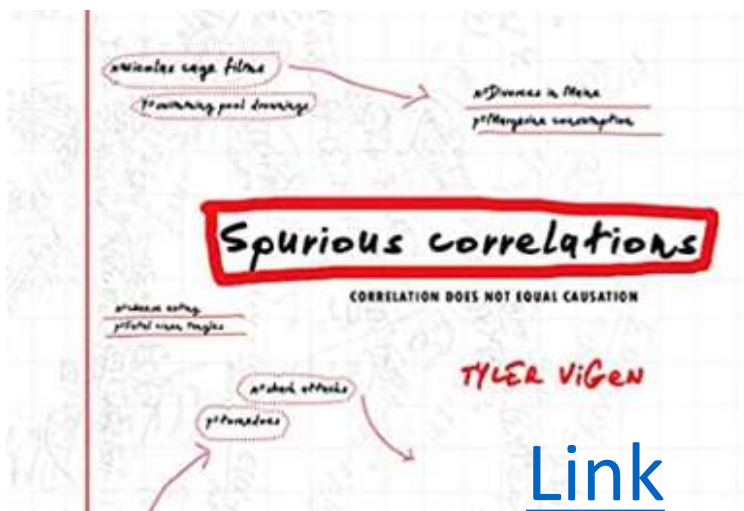


# Correlação não implica em causa e efeito

- Correlação  $\neq$  Casualidade

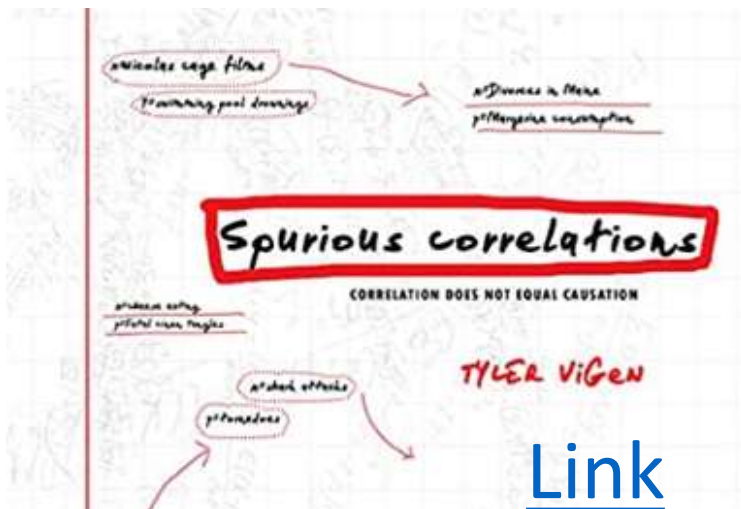
- O fato de duas variáveis serem correlacionadas não implica que uma variável está afetando diretamente a outra. Pode ser que exista um 3 elemento que afeta as duas variáveis. Exemplos:

- Consumo de queijo per-capto e número de pessoas que morrem sufocadas por estarem enroladas em seu lençol



# Correlação não implica em causa e efeito

- Correlação  $\neq$  Casualidade
  - O fato de duas variáveis serem correlacionadas não implica que uma variável está afetando diretamente a outra. Pode ser que exista um 3 elemento que afeta as duas variáveis. Exemplos:
    - Consumo per-capto de margarina e taxa de divórcios



# Correlação de Pearson



# Medida de correlação

- Coeficiente de correlação de Pearson (só usado quando você tem duas variáveis quantitativas contínuas)
- Pode assumir qualquer valor entre -1 e +1
- Interpretação do valor observado, segundo Vieira (2008)
  - De 0 até 0,25: correlação pequena
  - De 0,25 até 0,5: correlação fraca
  - De 0,5 até 0,75: correlação moderada
  - Acima de 0,75: correlação forte
- Obs: essa classificação é subjetiva e outros autores assumem outros níveis para considerar as intensidades das relações

# Pré-requisitos

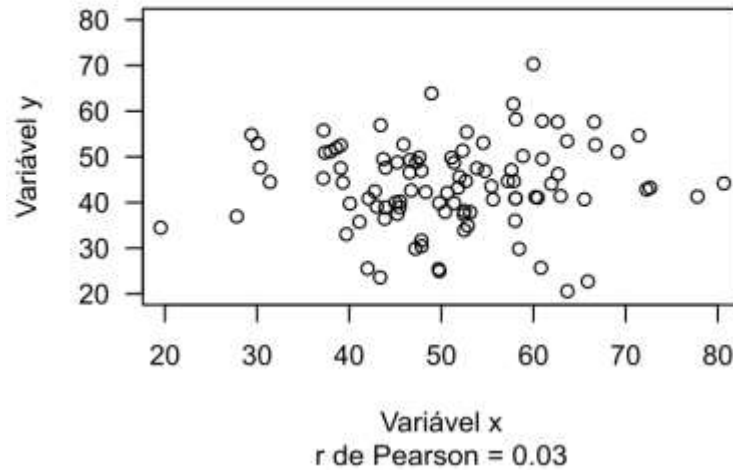
- Normalidade
- Verificar se não existem outliers

# Hipótese testada

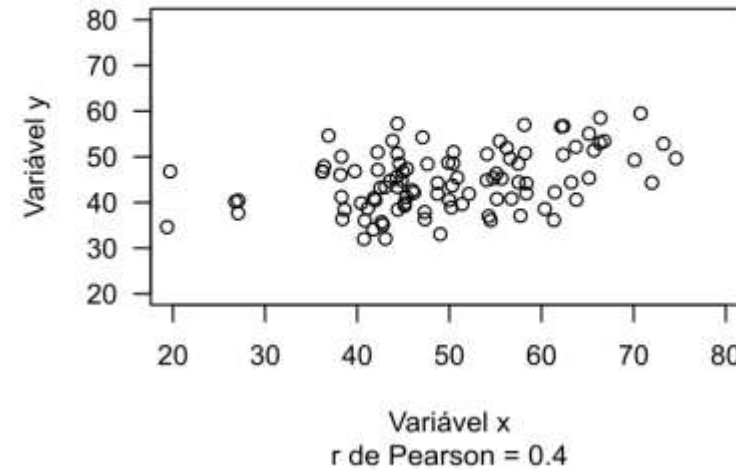
- $H_0$ : O coeficiente de correlação não é diferente de zero
- $H_1$ : O coeficiente de correlação é diferente de zero

# Compare diferentes valores de correlações

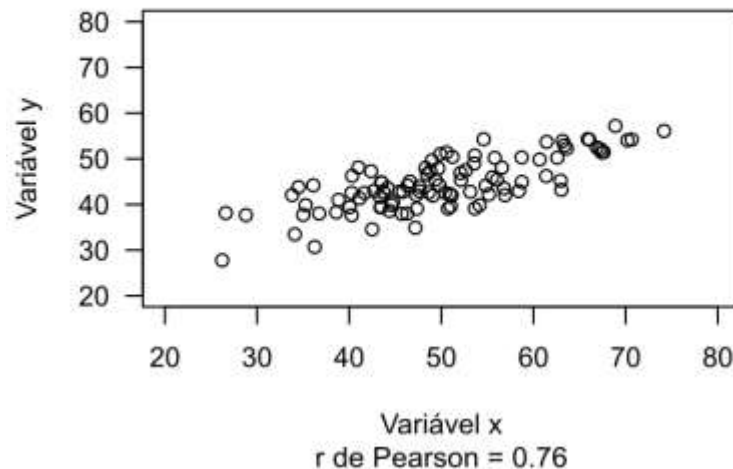
**Correlação pequena**



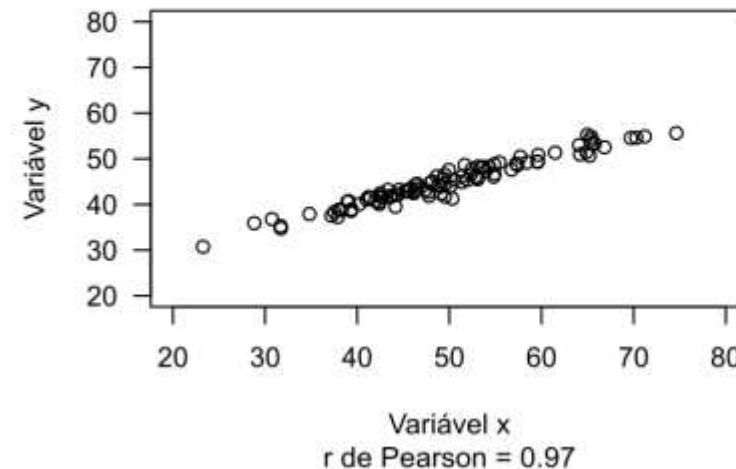
**Correlação fraca**



**Correlação moderada**

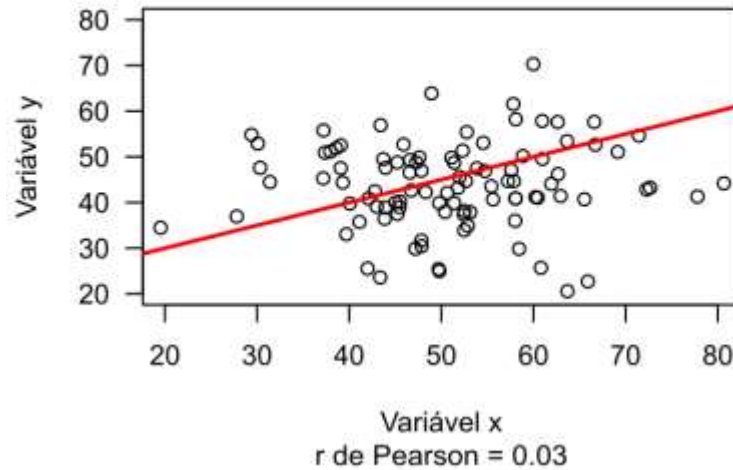


**Correlação forte**

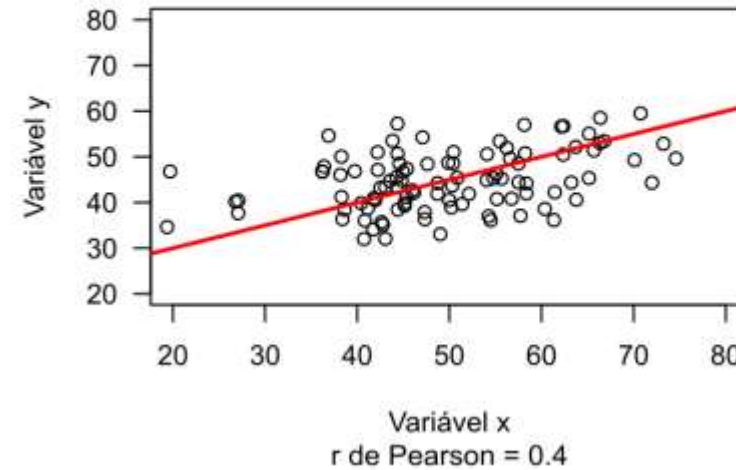


# Compare diferentes valores de correlações

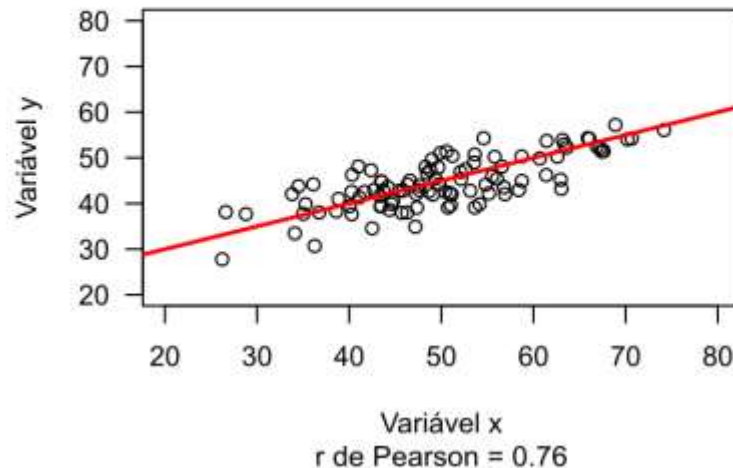
**Correlação pequena**



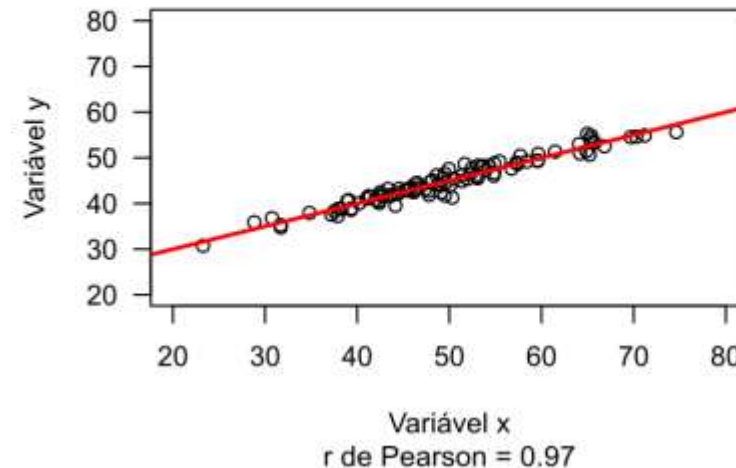
**Correlação fraca**



**Correlação moderada**

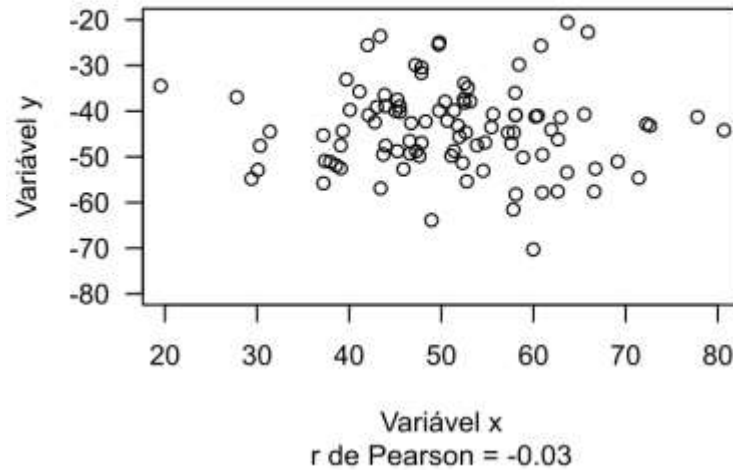


**Correlação forte**

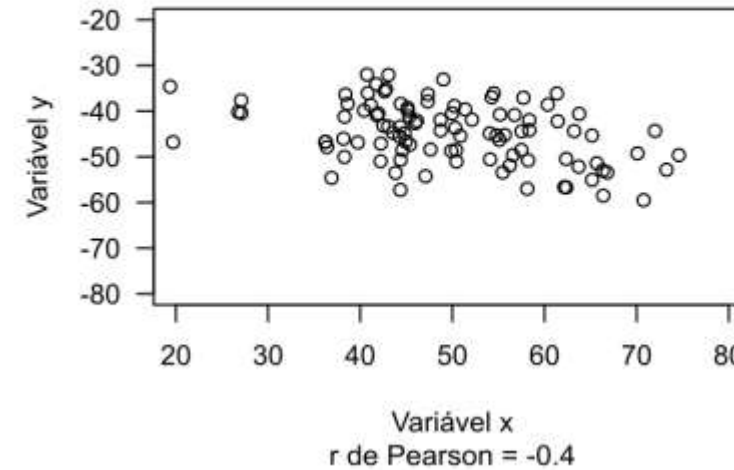


# Compare diferentes valores de correlações

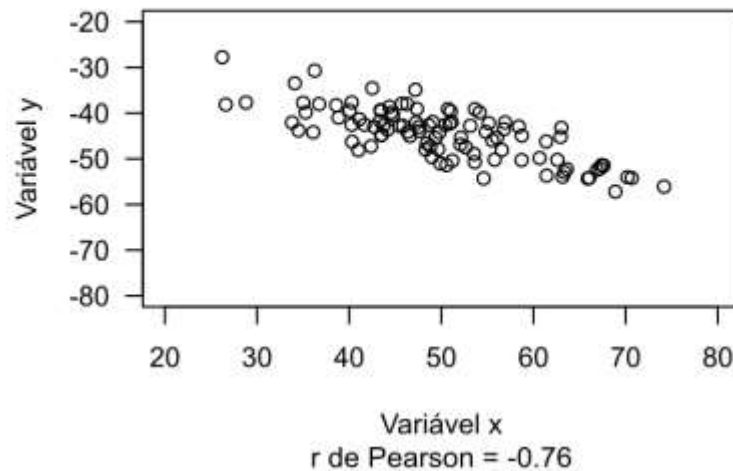
**Correlação pequena**



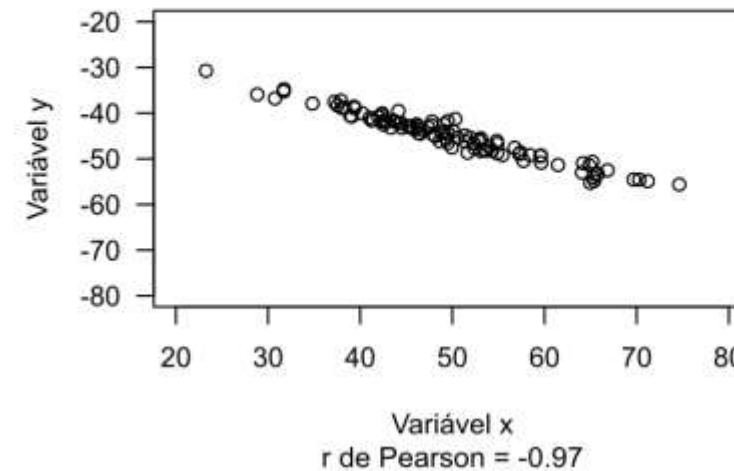
**Correlação fraca**



**Correlação moderada**

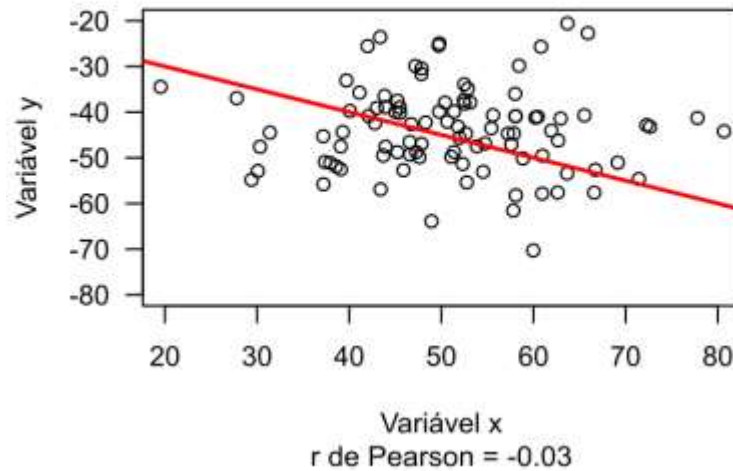


**Correlação forte**

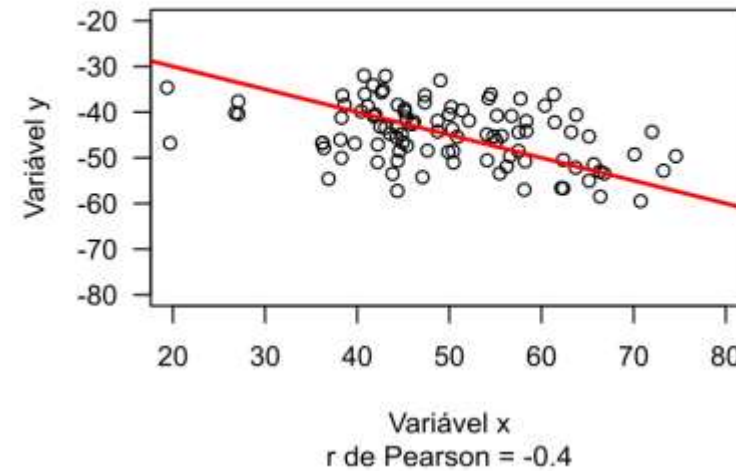


# Compare diferentes valores de correlações

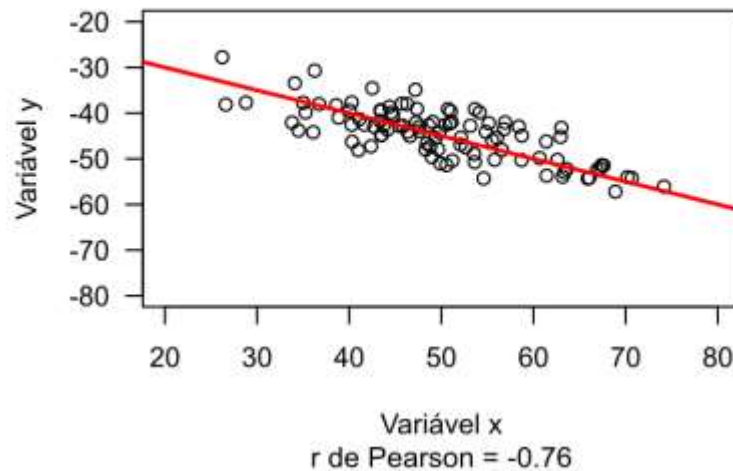
**Correlação pequena**



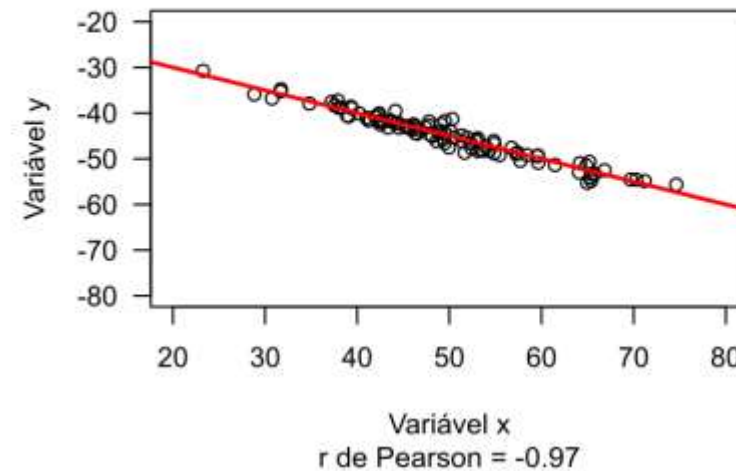
**Correlação fraca**



**Correlação moderada**



**Correlação forte**



# Correlação de Spearman



# Medida de correlação

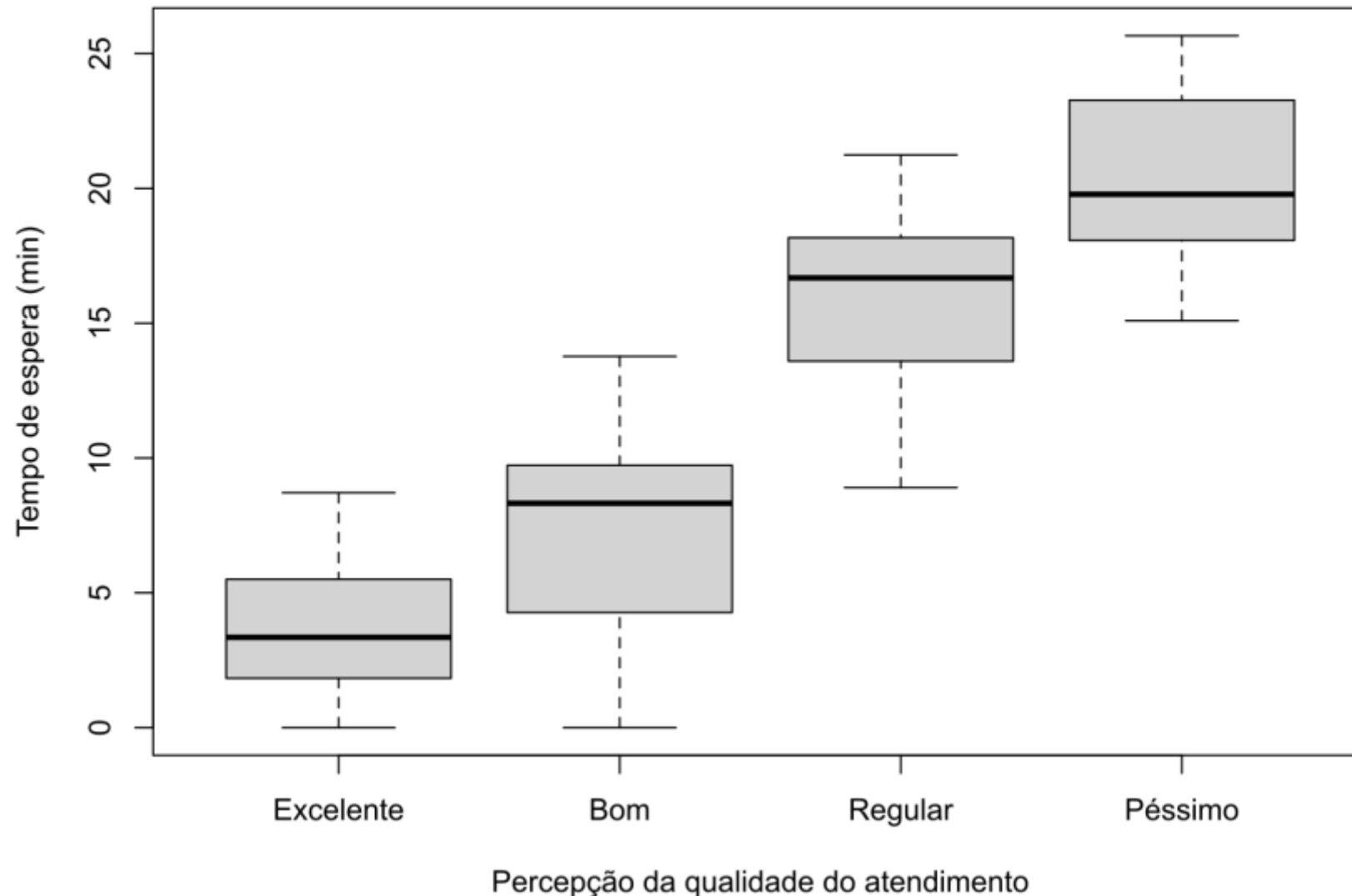
- Coeficiente de correlação de Spearman (usado quando você tem uma das variáveis categóricas ordinais)
- Pode assumir qualquer valor entre -1 e +1
- Interpretação do valor observado
  - De 0 até 0,25: correlação pequena
  - De 0,25 até 0,5: correlação fraca
  - De 0,5 até 0,75: correlação moderada
  - Acima de 0,75: correlação forte
- Obs: essa classificação é subjetiva e outros autores assumem outros níveis para considerar as intensidades das relações

# Correlação de Kendall

# Medida de correlação

- Coeficiente de correlação de Kendell (usado quando você tem uma das variáveis categóricas ordinais, e que existem empates no seu conjunto de dados)
- Pode assumir qualquer valor entre -1 e +1
- Interpretação do valor observado
  - De 0 até 0,25: correlação pequena
  - De 0,25 até 0,5: correlação fraca
  - De 0,5 até 0,75: correlação moderada
  - Acima de 0,75: correlação forte
- Obs: essa classificação é subjetiva e outros autores assumem outros níveis para considerar as intensidades das relações

# Exemplo: Percepção da qualidade do atendimento vs. tempo de espera



# Resumo

- Pearson: variáveis quantitativas que apresentam distribuição normal
- Spearman: variáveis quantitativas que não apresentam distribuição normal
- Kendall: variáveis que não apresentam distribuição normal e apresentem empates (valores repetidos)

Prática

# Prática – 1 Existe correlação entre número de horas que uma pessoa fica sentado e o nível de estresse?

- Tabela “DadosAula12.xlsx”
- Colunas da tabela:
  - HorasSentado: Variável quantitativa contínua
  - NivelEstresse: Variável quantitativa contínua

# Interpretação dos resultados

Pearson's product-moment correlation

```
data:  tabela1$Horassentado and tabela1$NivelEstresse  
t = 44.881, df = 98, p-value < 2.2e-16  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 0.9652515 0.9841733  
sample estimates:
```

```
cor  
0.976527
```

**Resultado: P menor que 0.05**  
**Conclusão:  $H_1$  é verdadeira**

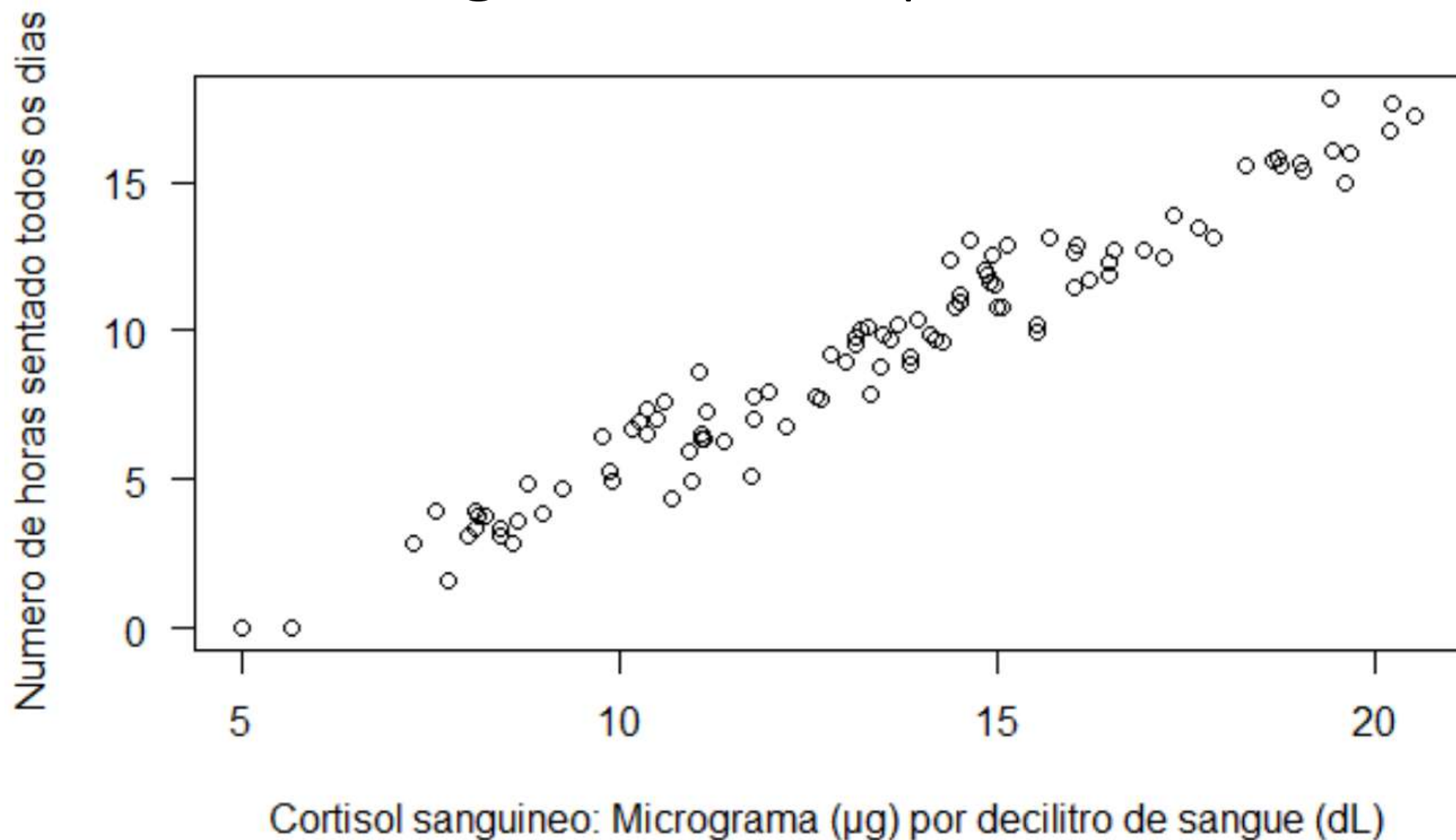
Hipóteses testadas

$H_0$ : Não há correlação entre as variáveis

$H_1$ : Há (SIM) correlação entre as variáveis



# Gráfico ou diagrama de dispersão



# Prática – 2 Existe correlação entre o nível de estresse e a pratica de atividade física?

- Tabela “DadosAula12.xlsx”
- Colunas da tabela:
  - HorasSentado: Variável quantitativa contínua
  - AtividadeFisica: Variável quantitativa contínua

# Interpretação dos resultados

Kendall's rank correlation tau

```
data: tabela1$NivelEstresse and tabela1$AtividadeFisica  
z = 2.0103, p-value = 0.0444  
alternative hypothesis: true tau is not equal to 0  
sample estimates:
```

```
tau  
0.1364877
```

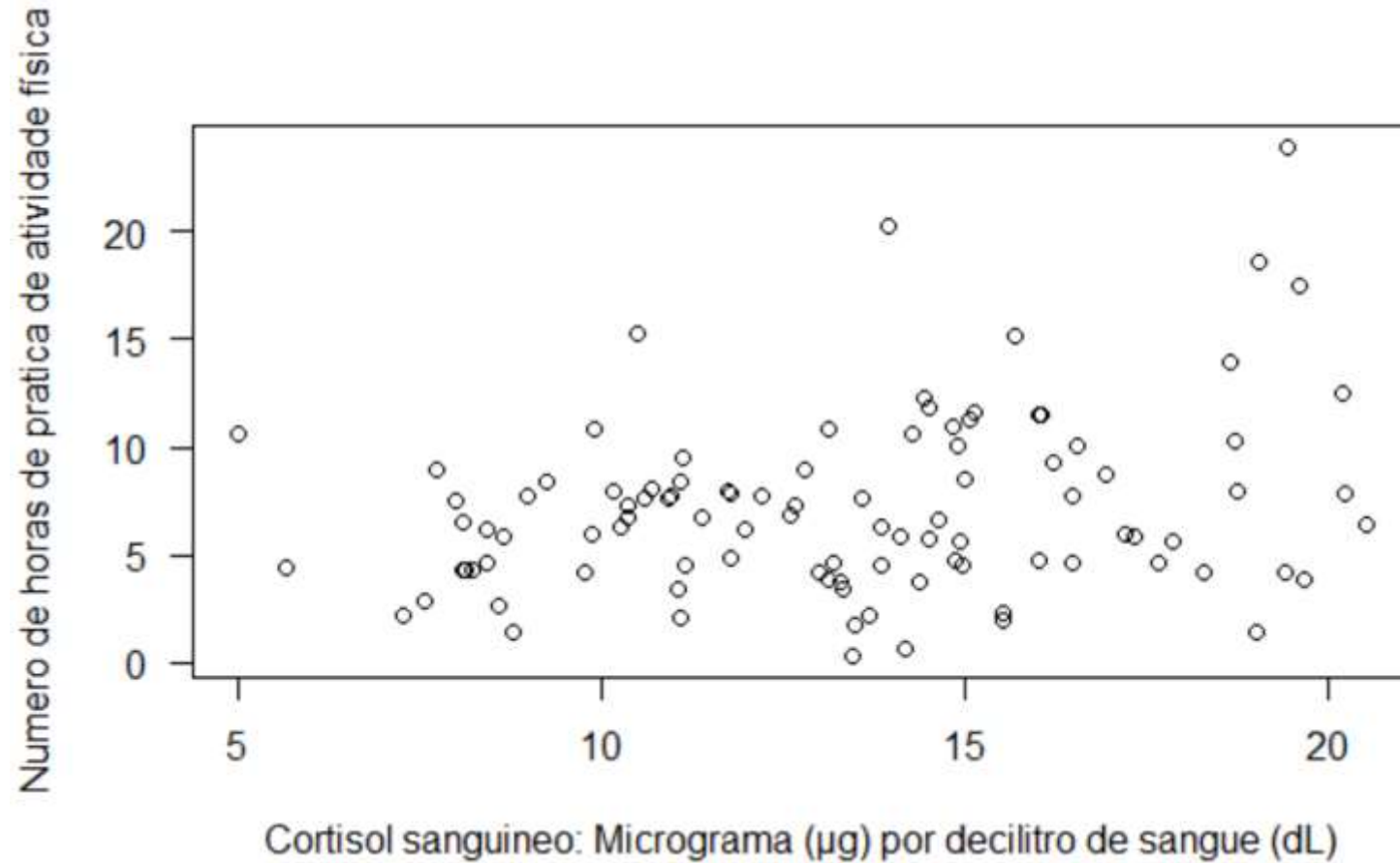
**Resultado: P menor que 0.05**  
**Conclusão:  $H_1$  é verdadeira**

Hipóteses testadas

$H_0$ : Não há correlação entre as variáveis

$H_1$ : Há (SIM) correlação entre as variáveis

# Gráfico ou diagrama de dispersão



# Prática – 3 Existe correlação entre numero de horas sentado e relato de dor nas costas?

- Tabela “DadosAula12.xlsx”
- Colunas da tabela:
  - HorasSentado: Variável quantitativa contínua
  - DoresCostas : Variável qualitativa ordinal

# Interpretação dos resultados

Kendall's rank correlation tau

```
data: rank(c(tabela1$DoresCostas)) and tabela1$HorasSentado  
z = 9.5387, p-value < 2.2e-16  
alternative hypothesis: true tau is not equal to 0  
sample estimates:
```

```
tau  
0.7188973
```

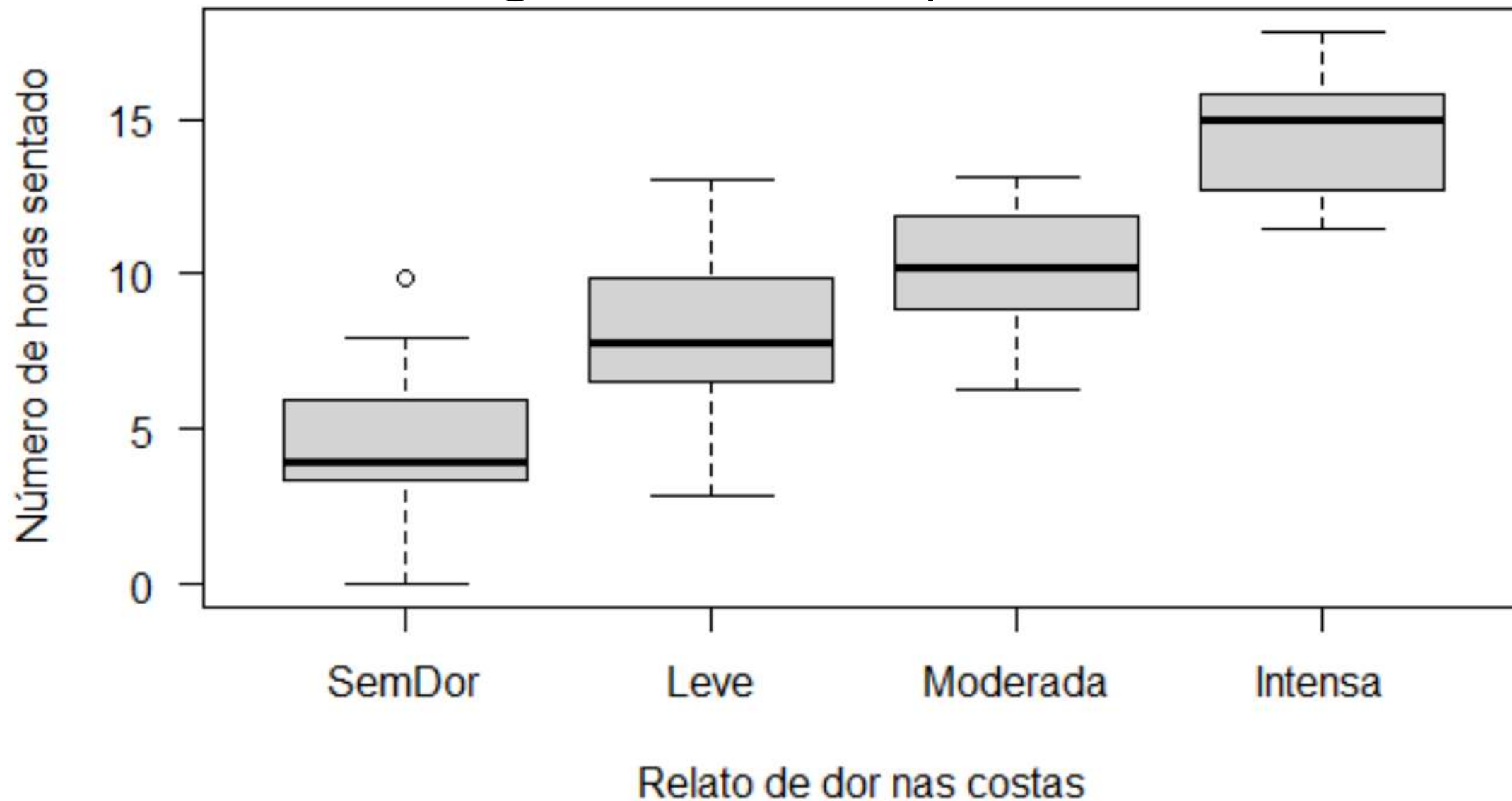
**Resultado: P menor que 0.05**  
**Conclusão:  $H_1$  é verdadeira**

Hipóteses testadas

$H_0$ : Não há correlação entre as variáveis

$H_1$ : Há (SIM) correlação entre as variáveis

# Gráfico ou diagrama de dispersão



# Prática – 4 Existe correlação entre a glicose sanguínea e insulina?

- Tabela “Dados Fleury.xlsx”
- Colunas da tabela:
  - Glicose
  - Insulina



# Interpretação dos resultados

Pearson's product-moment correlation

data: tabela2\$Glicose and tabela2\$Insulina  
t = 7.3691, df = 579, p-value = 5.957e-13  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:

0.2166411 0.3654630

sample estimates:

cor  
0.2928245

Resultado: P menor que 0.05  
Conclusão:  $H_1$  é verdadeira

Hipóteses testadas

$H_0$ : Não há correlação entre as variáveis

$H_1$ : Há (SIM) correlação entre as variáveis

# Gráfico ou diagrama de dispersão

