

# **ESTATÍSTICA PARA SAÚDE COLETIVA**

## **Aula 12**

# Seminários de hoje

Nomes (Aluno ou Dupla)	Data seminário	Nome do artigo
CLAUDIA MARTINS	08/10/2020	Análise Espacial dos Casos de COVID-19 e leitos de terapia intensiva no estado do Ceará, Brasil
CLAUDINEIA SOARES TORRES	08/10/2020	Estimativas de impacto do Covid-19 na mortalidade de idosos institucionalizados no Brasil
PRISCILA PRATES	08/10/2020	Promoção comercial ilegal de produtos que competem com o aleitamento materno

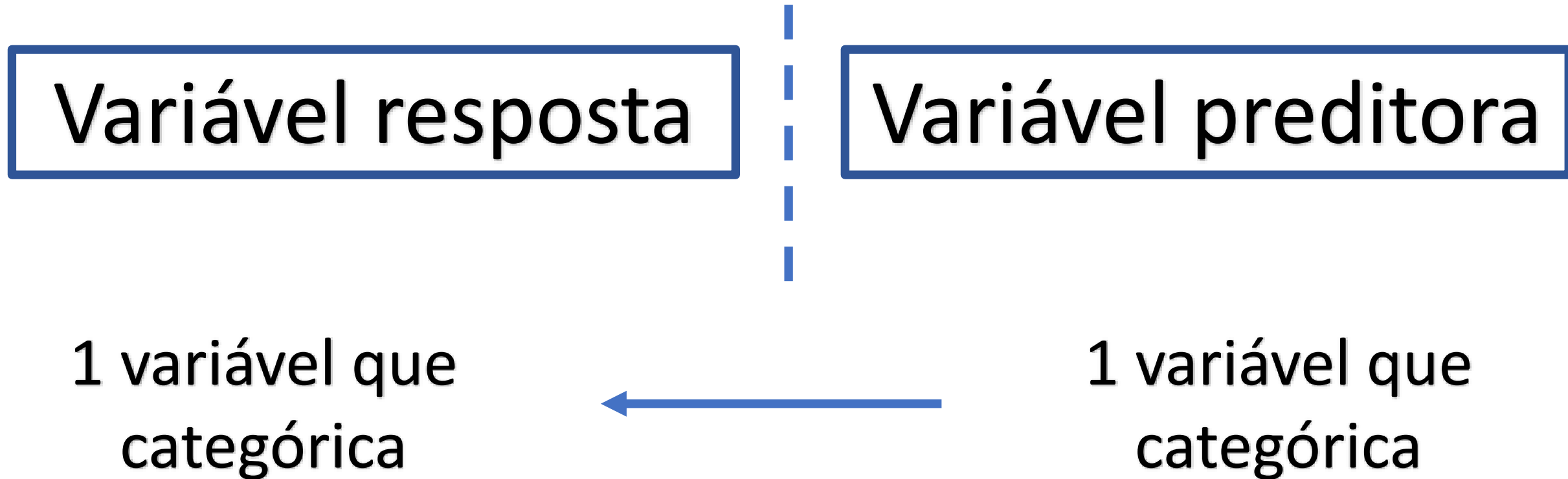
# Próxima semana

Nomes (Aluno ou Dupla)	Data seminário	Nome do artigo
EDNEIDE LOURENÇO + JOZIÊLDA ALVES	13/10/20	Qualidade de vida de mulheres com gravidez de alto risco durante o cuidado pré-natal
RAFAELA REIMBERG	13/10/20	Quedas em idosos institucionalizados: riscos, consequências e antecedentes
THAYLA NOLETO + RAFAELA PEREIRA	13/10/20	A saúde mental da enfermagem no enfrentamento da COVID-19 em um hospital universitário regional
ANDERSON SILVA	15/10/20	Características da epidemia de dengue em Pinhalzinho, Santa Catarina.
ROSA MARIA GARCIA + ANA LEITE	15/10/20	Consumo de alimentos fora do lar no Brasil segundo locais de aquisição
TAINARA PRADELLA + ADRIANA BARROS	15/10/20	Doenças Crônicas Não Transmissíveis e fatores de risco e proteção em adultos com ou sem plano de saúde
Pamela Monte Cruz	15/10/20	Prevalência de transtornos mentais comuns e fatores associados em moradores da área urbana de São Paulo, Brasil

Revisão

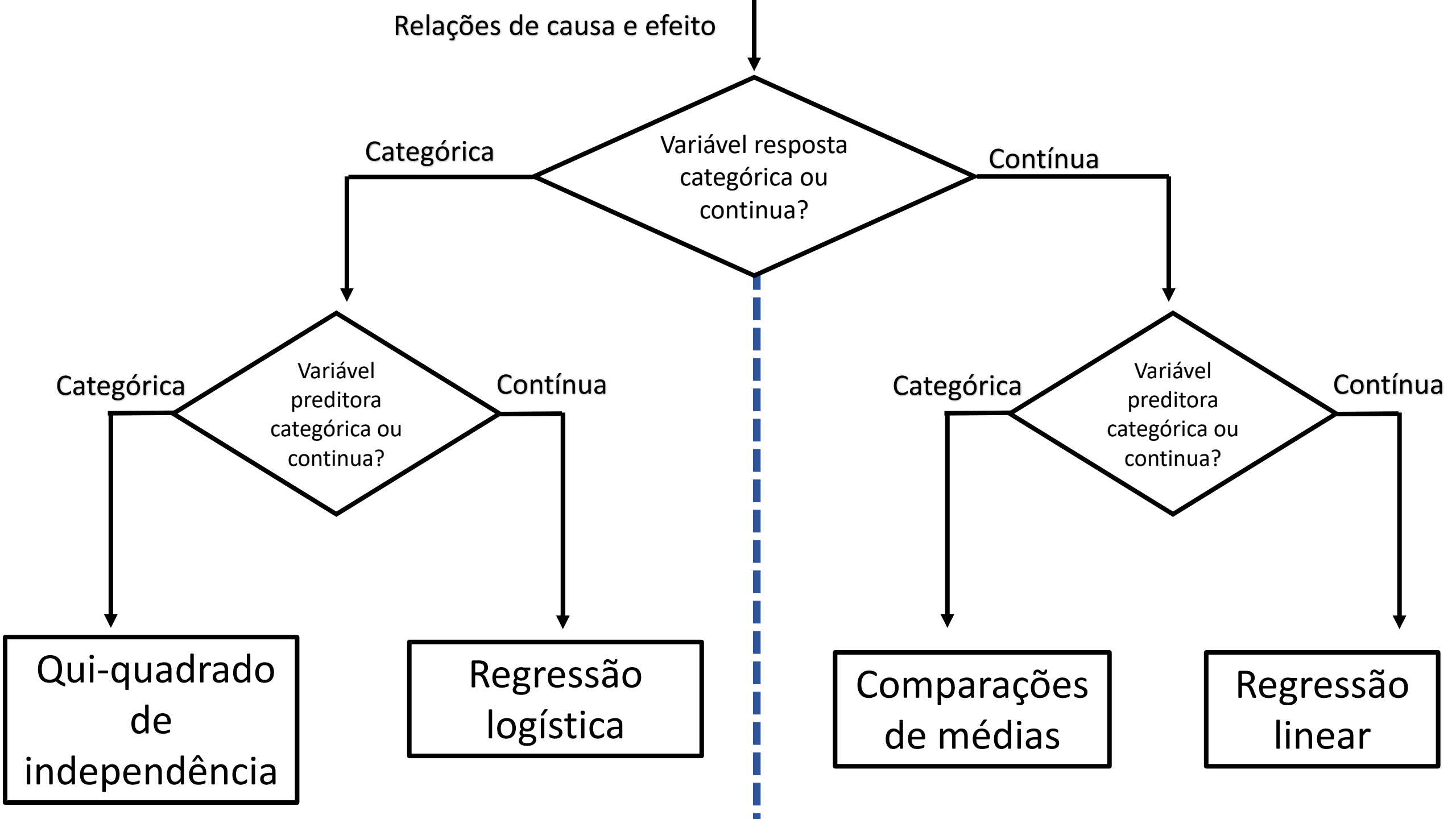
# Teste de qui-quadrado de independência

# 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%





Regressão logística

# Regressão logística

**Variável resposta**

**Variável preditora**

1 variável que  
categórica

1 variável que  
contínua



# Para que servem modelos preditivos?

- Porque as pessoas se interessam pela probabilidade de chuva do dia?

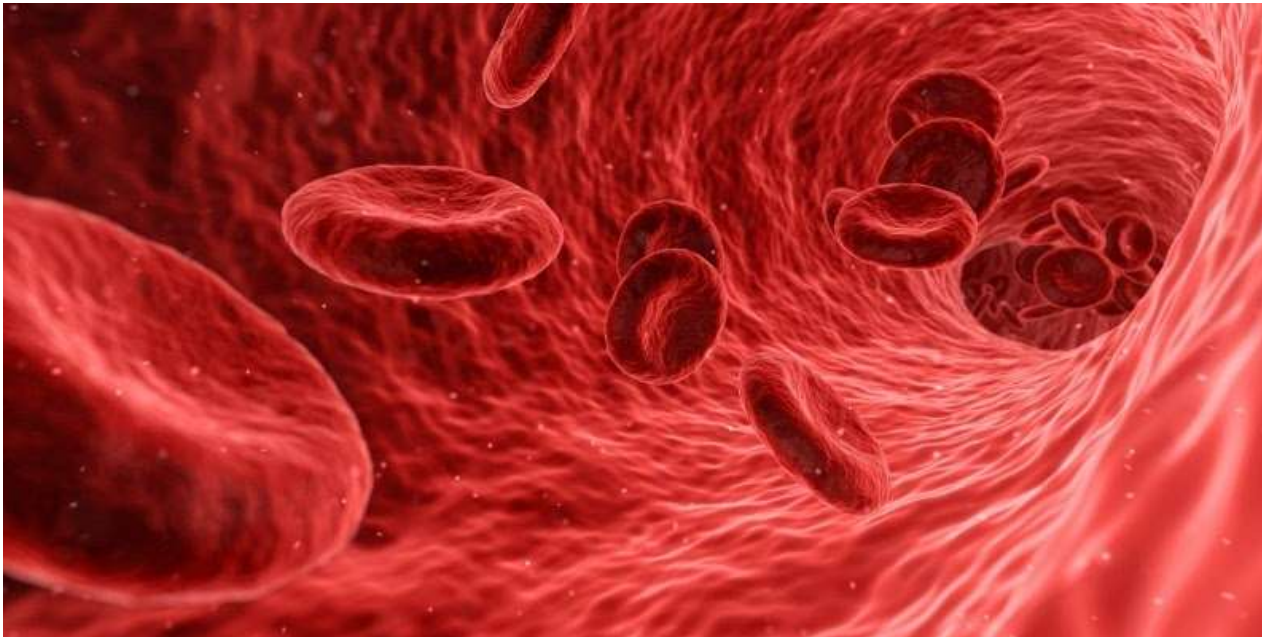


Para que servem modelos preditivos?



# Para que servem modelos preditivos?

- E se fosse possível fazer um exame de sangue completo e por meio da observação de vários parâmetros e se estimasse a probabilidade da pessoa estar com alguma doença “silenciosa”?



# Regressões

- Regressões são usadas para como uma “variável varia” em função de outras
- Regressões são uma forma de você criar um modelo matemático (ou seja uma formula) que pode ser usada para prever alguma coisa, quando você só tiver a informação da outra

# Regressões



- Pense nos modelos matemáticos como manequins, onde você ajusta seus dados a uma estrutura previamente estabelecida

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})}}.$$

Equação que define probabilidade de observar um fenômeno, dado a observação de alguns parâmetros

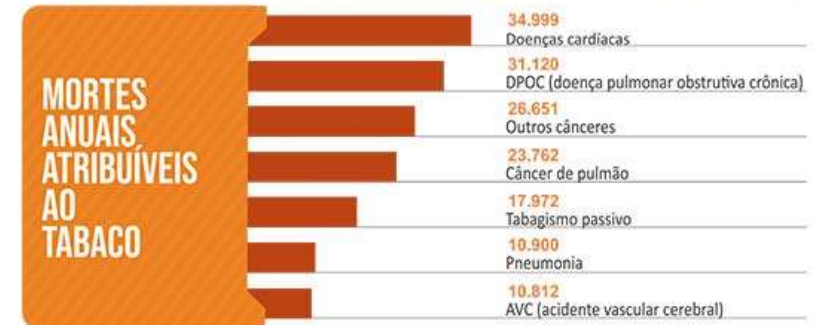


# Exemplo para ilustrar o teste

- O cigarro tem efeitos nocivos a em todos os sistemas do corpo humano
- 428 pessoas morrem por dia no brasil por causa do tabagismo.

[Link](#)

**12,6%** DE TODAS AS MORTES QUE OCORREM NO PAÍS  
PODEM SER ATRIBUÍDAS AO TABAGISMO



## O que acontece quando um fumante para de fumar

### 20 minutos

Os batimentos cardíacos e a pressão sanguínea voltam aos níveis normais

### 2 semanas

Melhora na circulação e função pulmonar

### 1 ano

O risco de contrair uma doença cardíaca coronária cai pela metade do que a de um fumante

### 10 anos

O risco de morrer de câncer de pulmão é reduzido pela metade se comparado com a um fumante. O risco de contrair câncer de laringe e pâncreas também diminui

### 12 horas

Os níveis de monóxido de carbono no sangue voltam ao normal

### 1 a 9 meses

Grande diminuição da tosse e da falta de ar. Os minúsculos pelos que revestem o pulmão (chamados de cílios) voltam a funcionar normalmente limpando o pulmão e reduzindo o risco de infecção.

### 5 anos

O risco de contrair câncer de boca, garganta, esôfago e bexiga é reduzido pela metade se comparando com um fumante. O risco de câncer cervical ou AVC também é reduzido

### 15 anos

O risco de doença cardíaca é equivalente aos do não fumantes





# Exemplo para ilustrar o teste

- Um pesquisador quer então fazer um modelo matemático que prevê o risco da pessoa ser diagnosticada com câncer de pulmão aos 50 anos de idade, dado a quantidade de cigarros fumados ao longo da vida.



# Regressão logística

**Variável resposta**

**Variável preditora**

Teve câncer ao  
atingir 50 anos?

Número de cigarros  
fumados ao longo da  
vida



# Exemplo para ilustrar o teste

- No modelo criado, cada cigarro fumado aumenta um pouquinho a probabilidade de câncer

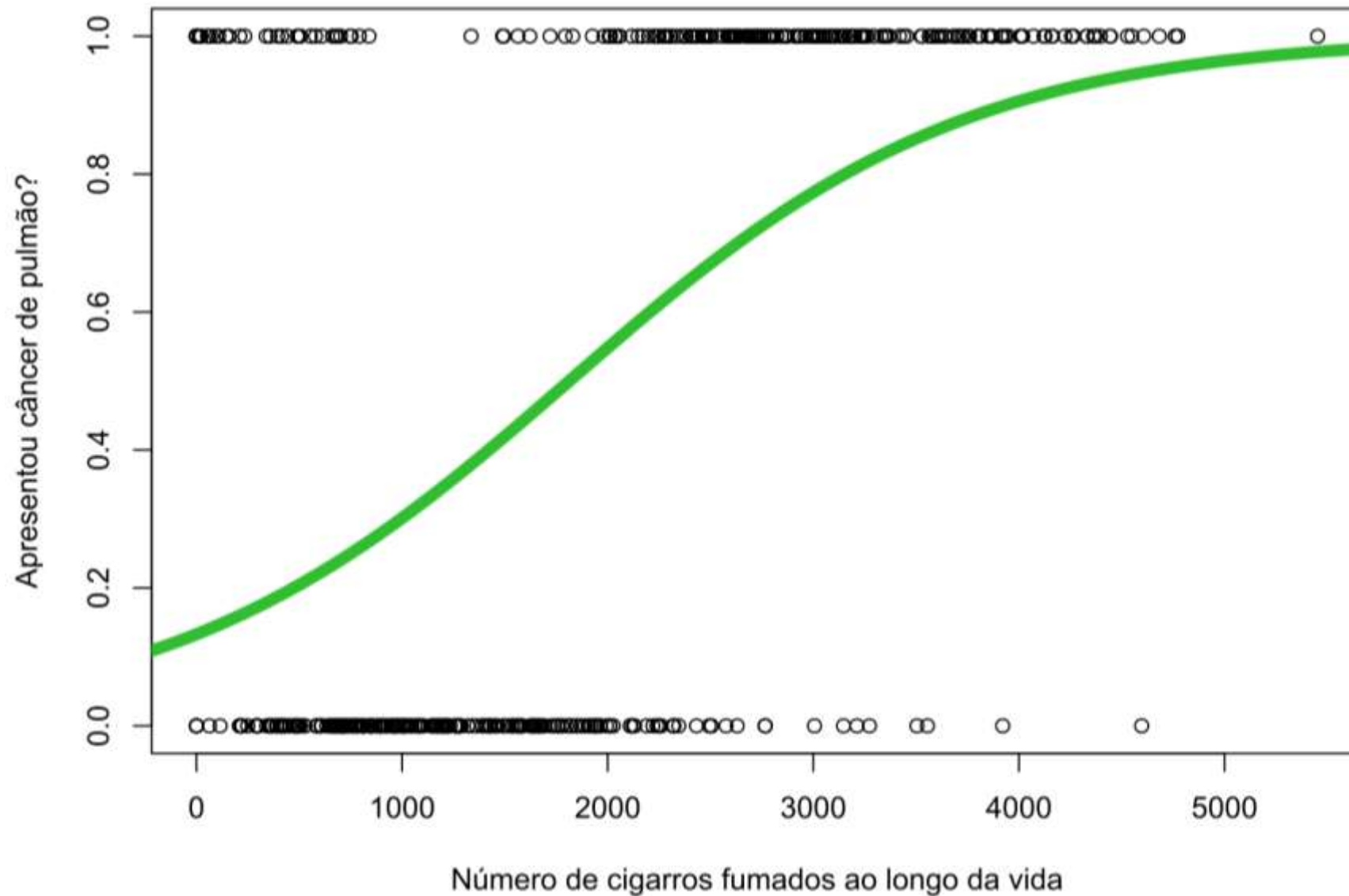
Número de cigarros fumados	Probabilidade de câncer de pulmão
0	16,05%
1	16,06%
2	16,07%
3	16,09%
50	16,82%
100	17,63%
1000	37,18%
10000	99,99%

Dados simulados para o exemplo usado nos próximos slides



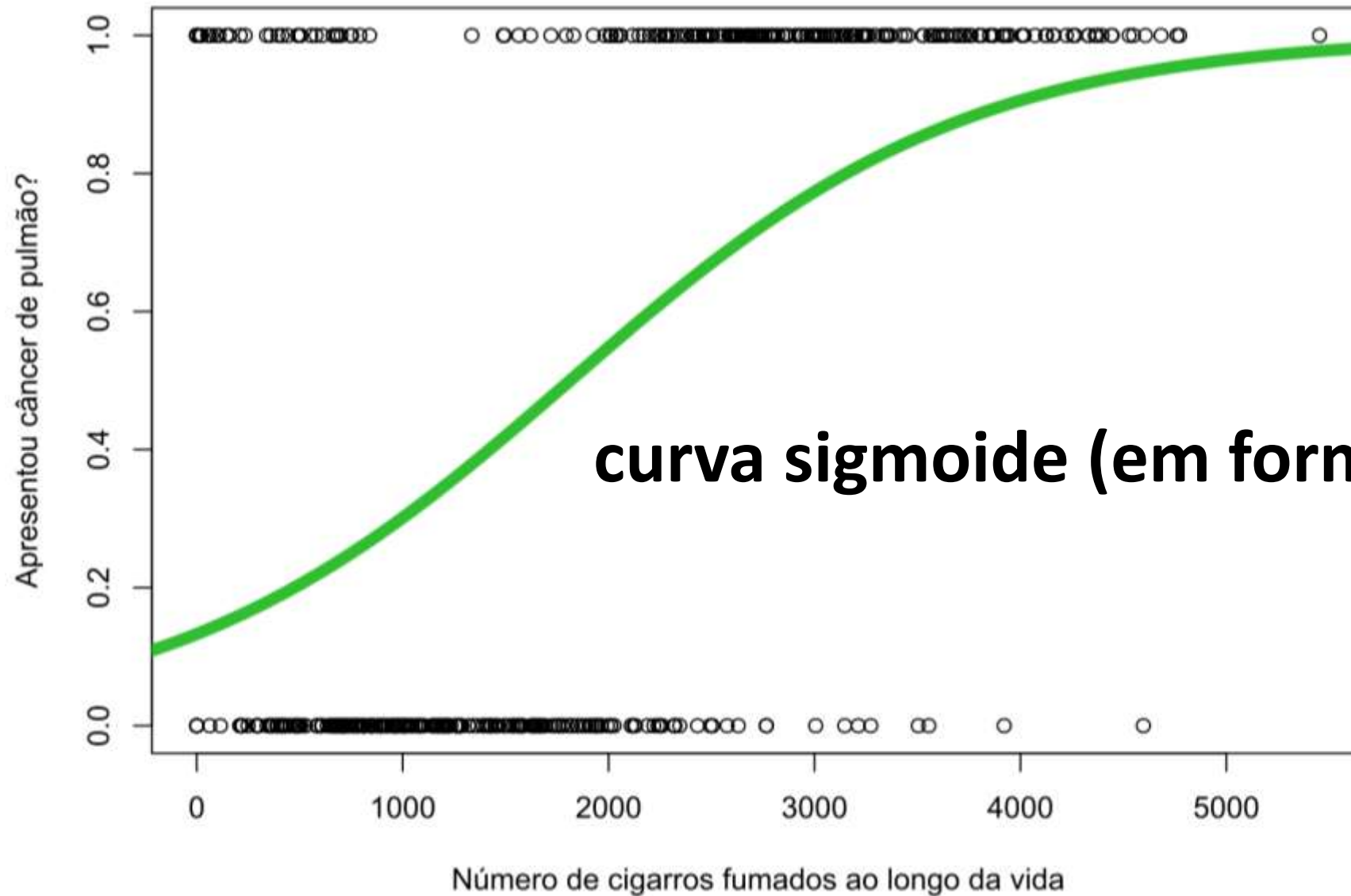
# Ideia geral do teste

Curva de distribuição de probabilidades



# Ideia geral do teste

Curva de distribuição de probabilidades



**curva sigmoide (em formato de S)**

# Hipóteses

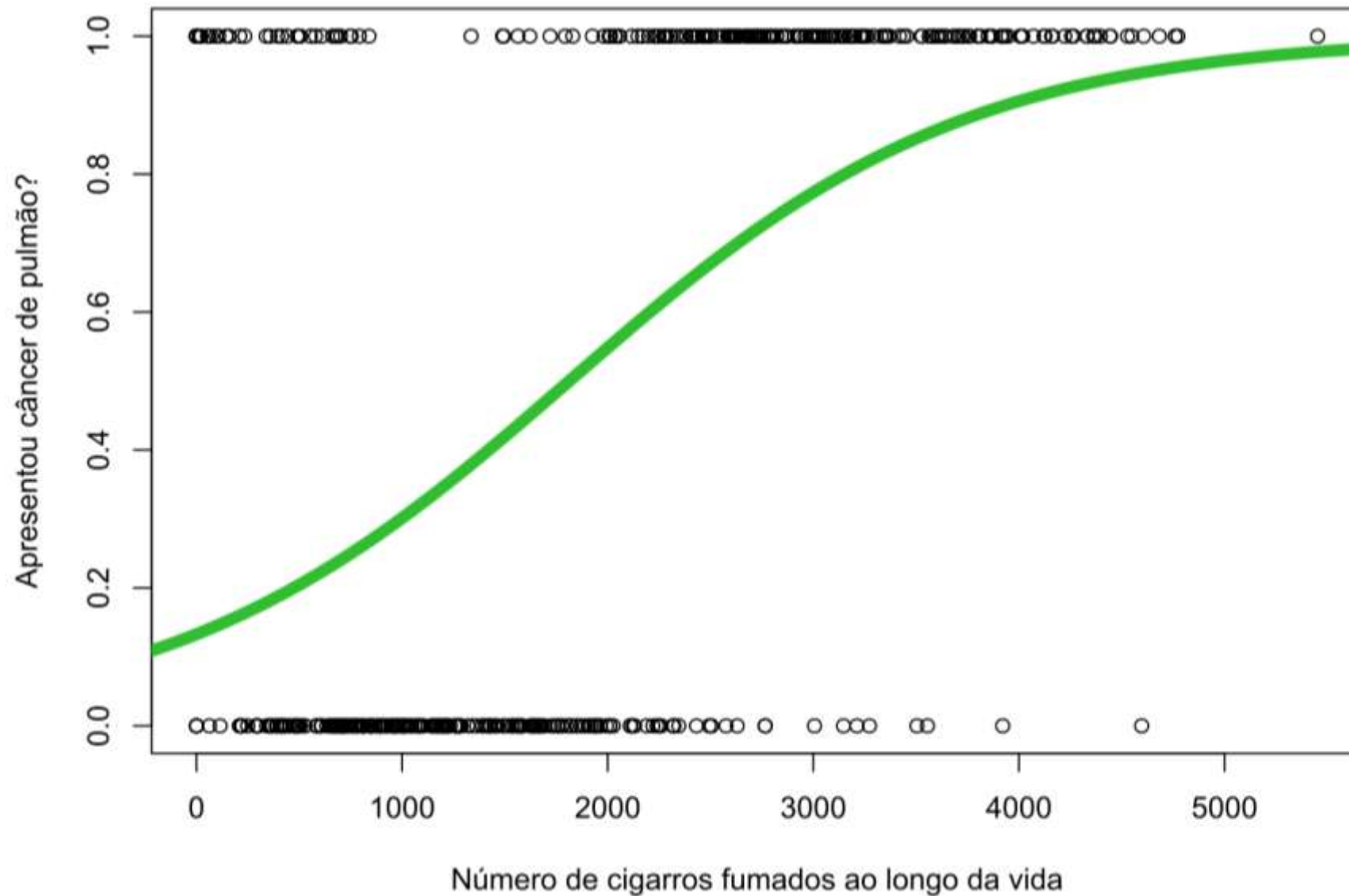
- Hipóteses testadas
  - $H_0$ : Não há diferenças entre os resultados previstos pelo modelo e os valores observados
  - $H_1$ : Há (SIM) diferenças entre os resultados previstos pelo modelo e os valores observados
- Em outras palavras
  - $H_0$ : Não é possível prever a variável respostas com base no modelo ajustado
  - $H_1$ : É possível prever a variável respostas com base no modelo ajustado

# Logica por trás de uma comparação com modelo nulo

- Imagine que no exemplo, 30% das pessoas tinham câncer e 70% não tiveram câncer.
- O modelo nulo ( $H_0$ ) vai dizer que todo mundo tem 30% de chance de ter câncer. Portanto individualmente, é mais provável que ninguém tenha câncer.
  - Dessa forma, o modelo nulo vai acertar 70% das vezes
- Em oposição, o modelo de regressão logística, calcula a probabilidade de cada indivíduo ter câncer. Em seguida faz uma previsão individual se a pessoa terá ou não a doença.
  - Nessa caso, o modelo vai acertar algumas previsões e errar outras.
- Se o modelo acerta mais vezes que quando você simplesmente assume que ninguém vai ter câncer, isso indica que o modelo é bom!

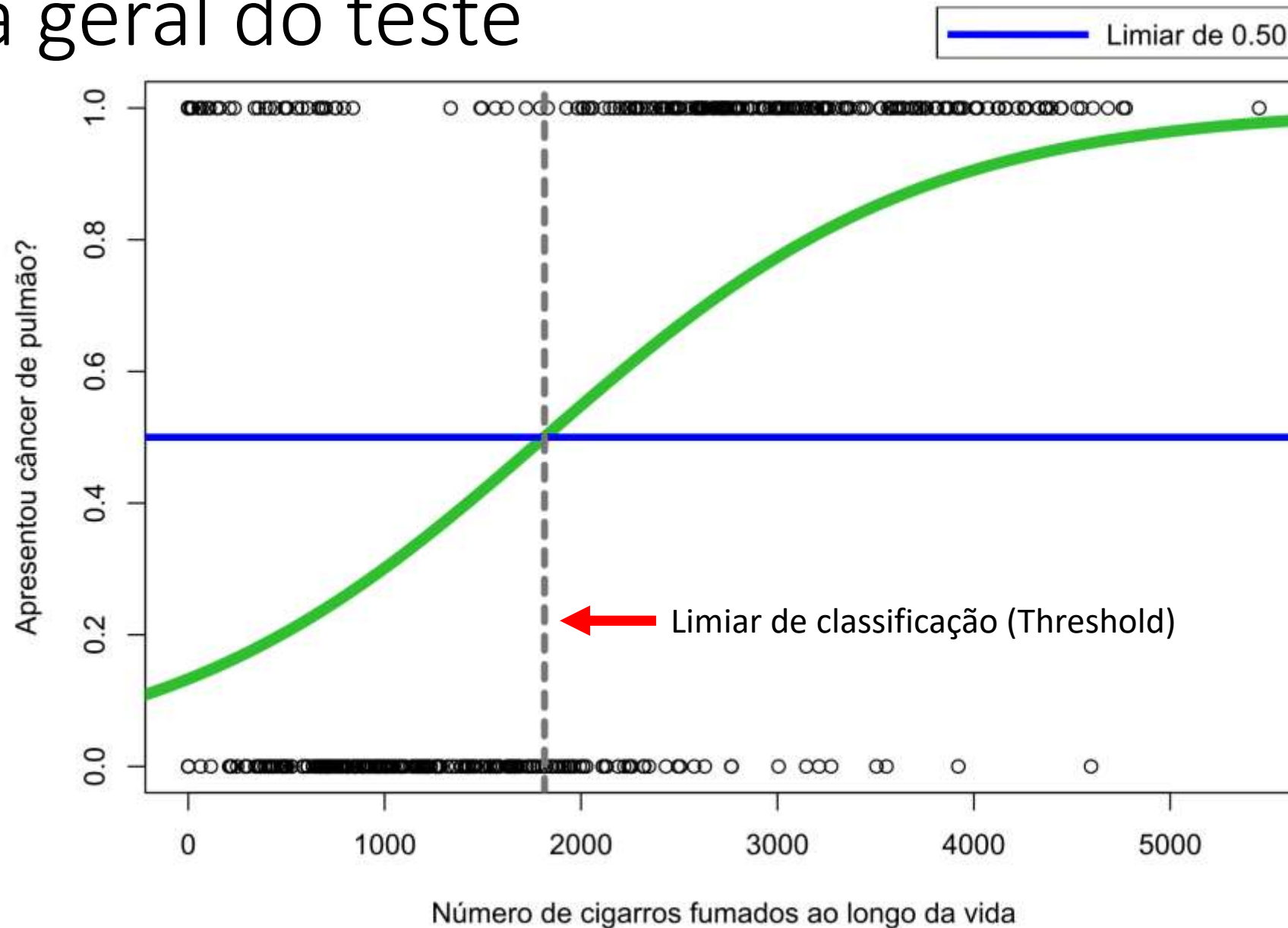
# Ideia geral do teste

Curva de distribuição de probabilidades





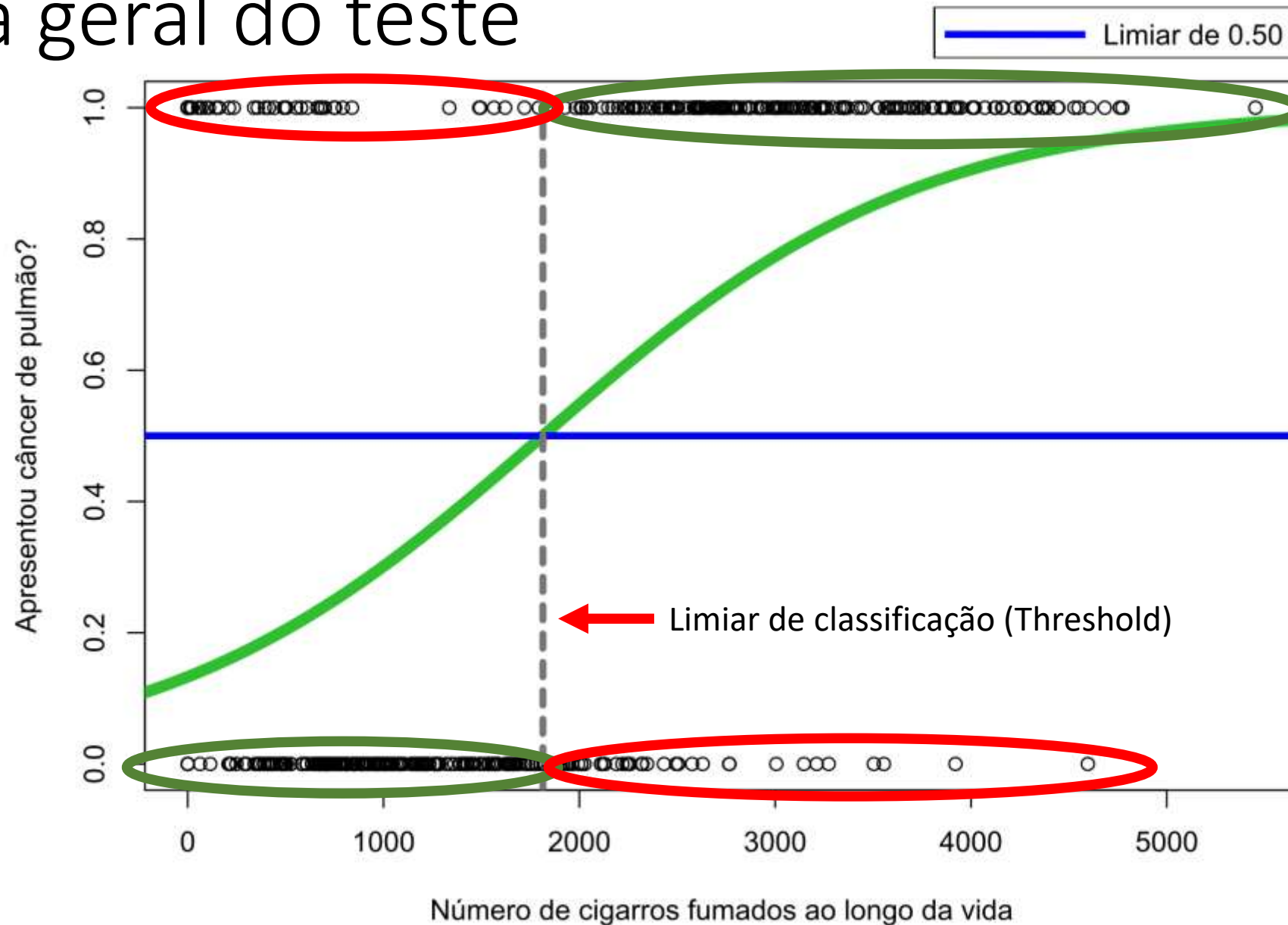
# Ideia geral do teste



Acertou nesses casos

Errou nesses casos

# Ideia geral do teste



# Interpretação dos resultados

```
> summary(B1)
```

```
Call:
```

```
glm(formula = Cancer ~ Cigarros, family = binomial)
```

```
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-2.6728	-0.6850	-0.2146	0.6547	1.8891

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.601e+00	1.668e-01	-9.597	<2e-16 ***
Cigarros	1.129e-03	9.529e-05	11.844	<2e-16 ***

```
---
```

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

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 693.15  on 499  degrees of freedom  
Residual deviance: 481.38  on 498  degrees of freedom  
AIC: 485.38
```

```
Number of Fisher Scoring iterations: 4
```

# Interpretação dos resultados

```
> summary(B1)
```

Significância do seu preditor

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-1.601e+00	1.668e-01	-9.597	<2e-16	***
Cigarros	1.129e-03	9.529e-05	11.844	<2e-16	***

# Interpretação dos resultados

- Odds: razão entre chances
  - Valor maior que 1 indica aumento de probabilidade de pertencer ao grupo da variável resposta
  - Valor menos que 1 indica redução de probabilidade de pertencer ao grupo da variável resposta
- Pode ser calculado tanto para variáveis contínuas como categóricas
  - Variável contínua. Valor de 1.001, indica que cada incremento na variável preditora (cada cigarro fumado) aumenta a chance de ter câncer em 1.001x

```
> # Odds
> exp(coef(B1))
(Intercept)      Cigarros
0.1911631      1.0011306
```

- Variável categórica. Suponha que você incluiu o sexo do participante da pesquisa, e o cálculo deu de *odds ratio* de F x M teve resultado de 2.5 isso indicaria que indivíduos do sexo feminino tem 2,5 x mais chance de ter câncer que indivíduos do sexo masculino

# Pré-requisitos

*Residual deviance* tem que ser menor que os graus de liberdade, isso indica que não existe *overdispersion* (superdispersão)

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 693.15 on 499 degrees of freedom  
Residual deviance: 481.38 on 498 degrees of freedom  
AIC: 485.38

Number of Fisher Scoring iterations: 4

# Pré-requisitos

- Não existe multicolinearidade: em caso de regressões múltiplas (assunto da última aula)
- Não devem existir outliers
- N mínimo. Você pode calcular n mínimo por meio de formulas, mas um n mínimo sugerido é pelo menos 10 observações para cada variável estudada

# Como fica uma tabela desse tipo de dado?

Nome	Numero de cigarros fumados ao longo da vida	Teve câncer?
Lorena	0	Não
Livia	1303	Sim
Maria Luiza	4	Não
Cecilia	100	Sim
Eloa	0	Não
Giovanna	1235	Não
Maria Clara	516	Não
Maria Eduarda	20	Sim
Mariana	500	Não
Lara	3506	Sim
Beatriz	322156	Sim
Antonella	26358	Sim



Identificação do dado

Variável preditora  
contínua

Variável resposta  
categórica

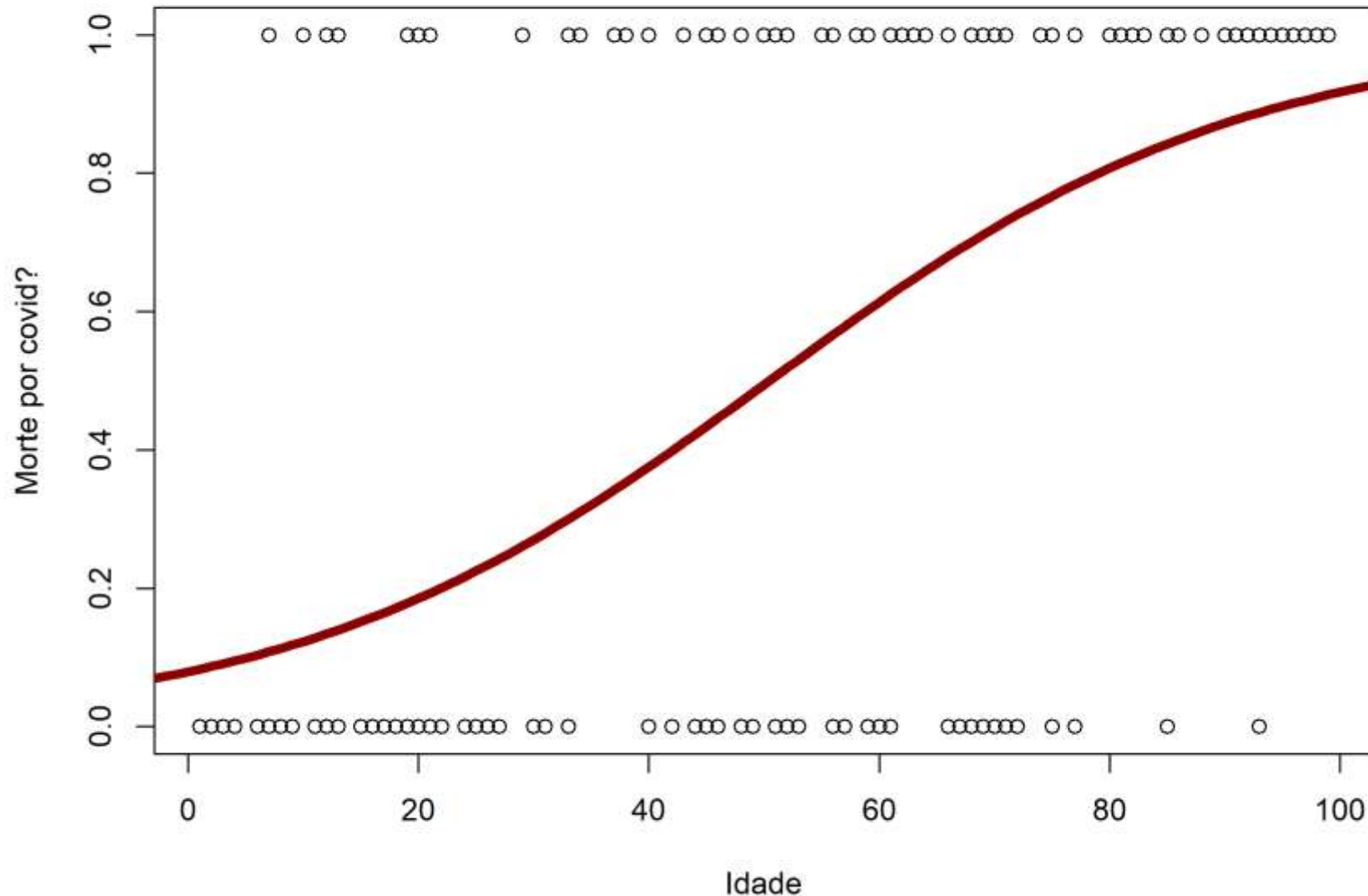
a um  
esse tipo de

Nome	Numero de cigarros fumados ao longo da vida	Teve câncer?
Lorena	0	Não
Livia	1303	Sim
Maria Luiza	4	Não
Cecilia	100	Sim
Eloa	0	Não
Giovanna	1235	Não
Maria Clara	516	Não
Maria Eduarda	20	Sim
Mariana	500	Não
Lara	3506	Sim
Beatriz	322156	Sim
Antonella	26358	Sim

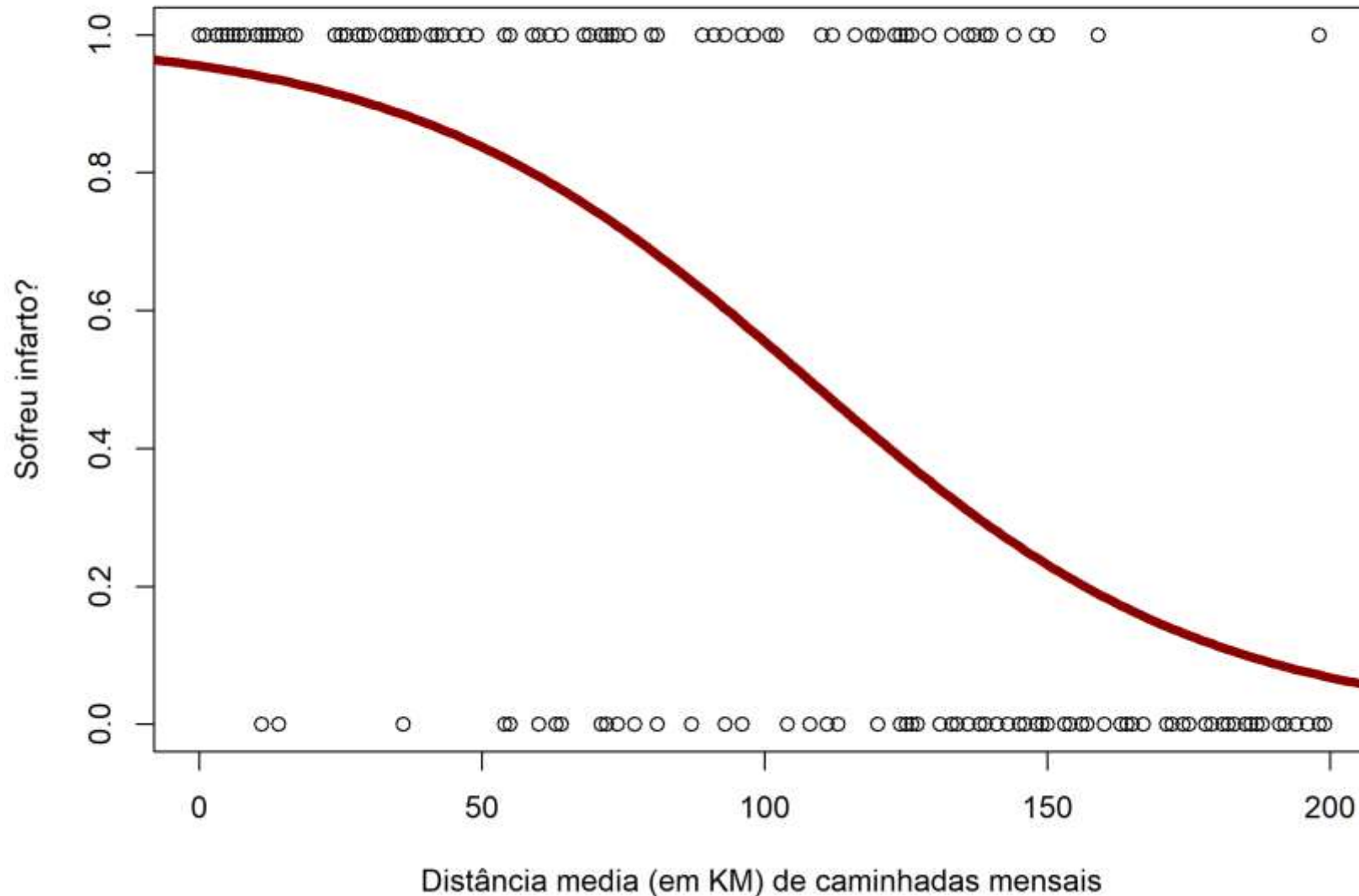
# Exemplos de adaptações de perguntas da Lista um que podem ser respondidas com essa metodologia

1. A mortalidade por COVI-19 é influenciada pela idade do paciente? (Adriana Barros)
2. A prática de atividade física realmente reduz risco de doenças cardíacas? (Edneide)
3. Comer doces em excesso causa diabetes? (Edneide)
4. Consumir álcool em excesso pode prejudicar o fígado?(Pamela)
5. O consumo de alimentos ricos em açúcar aumenta o risco de cárie?(Rafaela Reimberg)
6. O risco de morte de um recém nascido é influenciado pela renda dos seus pais? (Rosa Garcia)
7. Pessoas se automedicam mais frequentemente quando médicos de uma cidade cobram um valor elevado por uma consulta médica? (Elza)

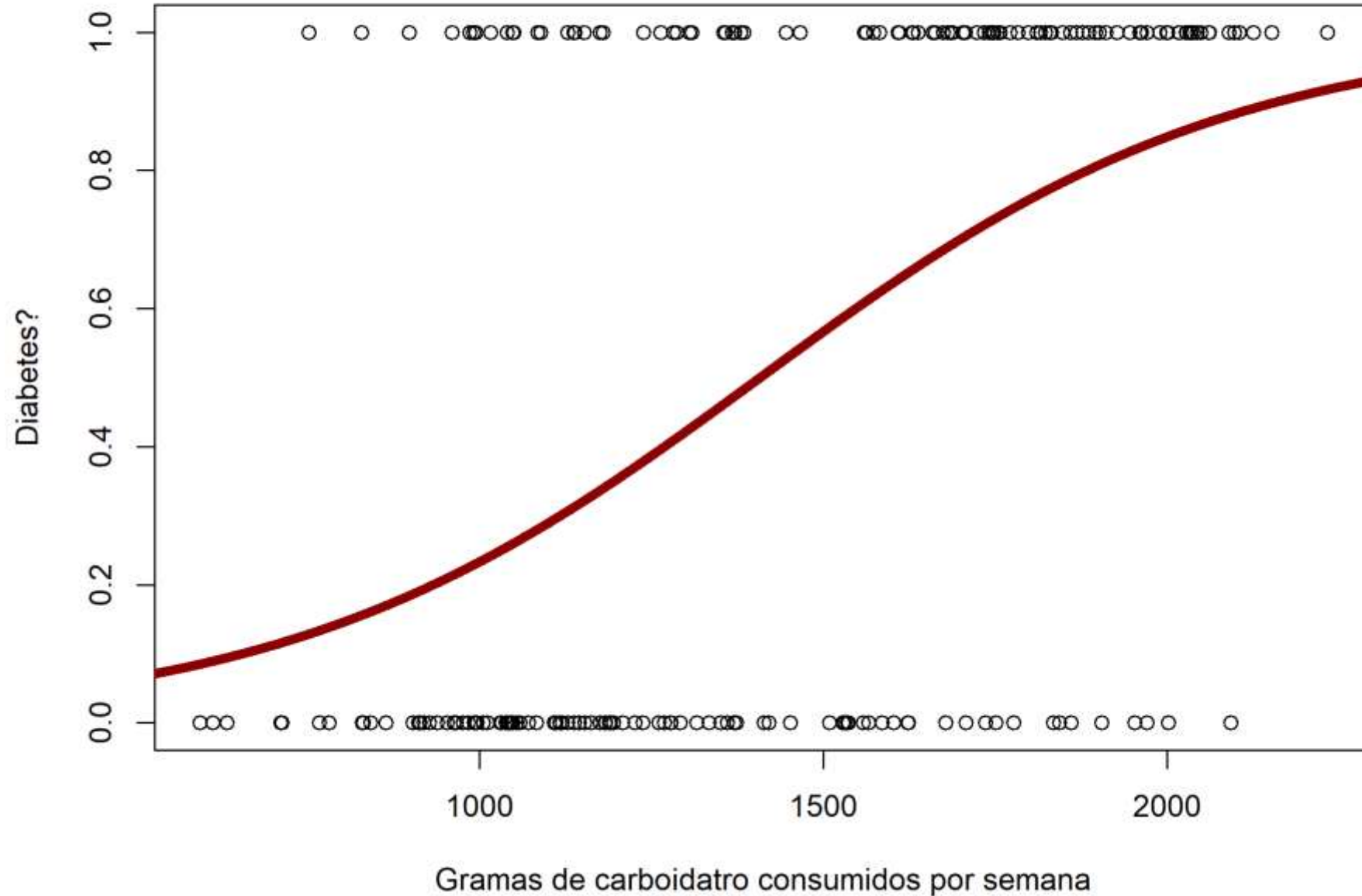
# A mortalidade por COVI-19 é influenciada pela idade do paciente?



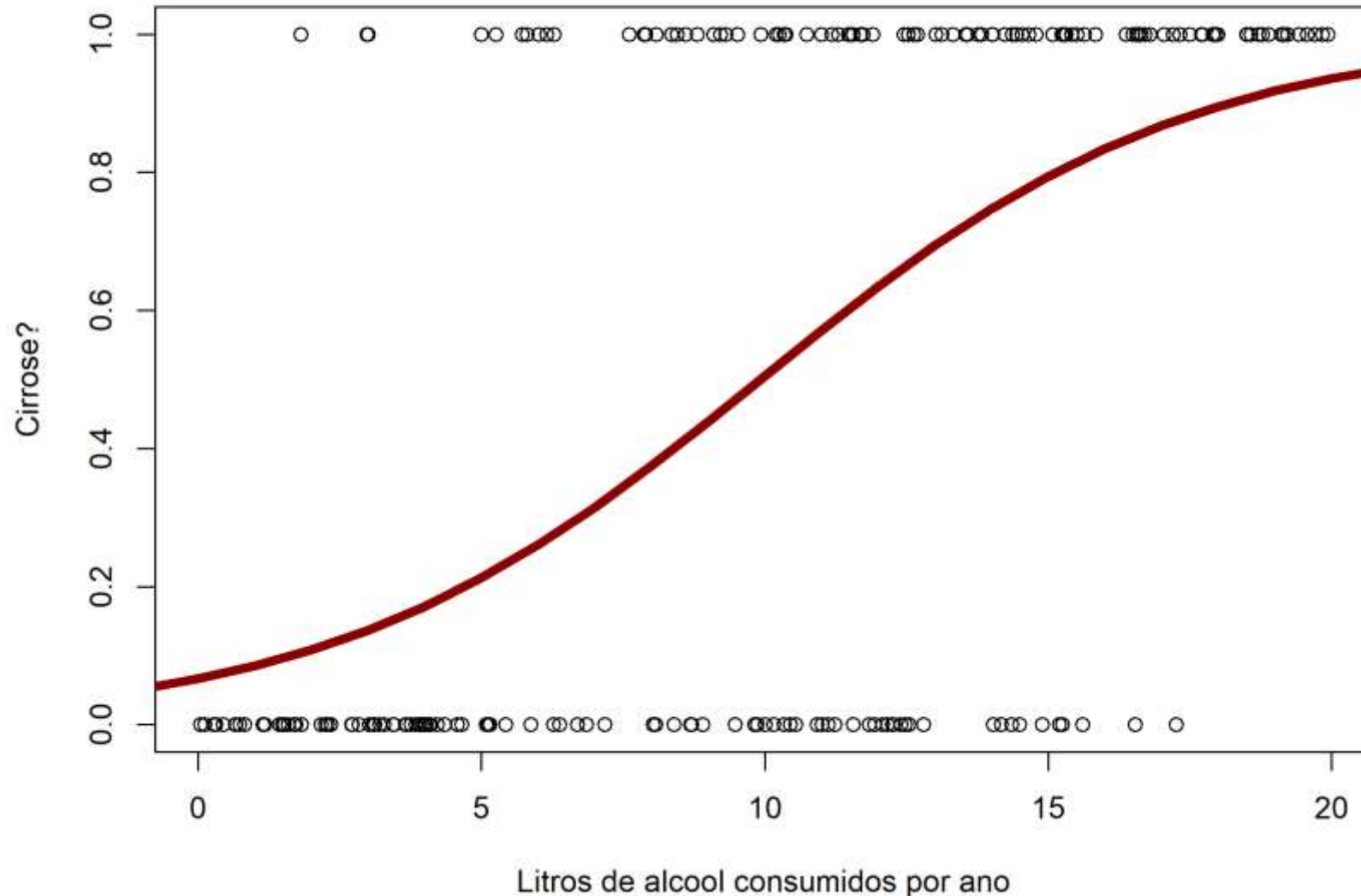
# A prática de atividade física realmente reduz risco de doenças cardíacas?



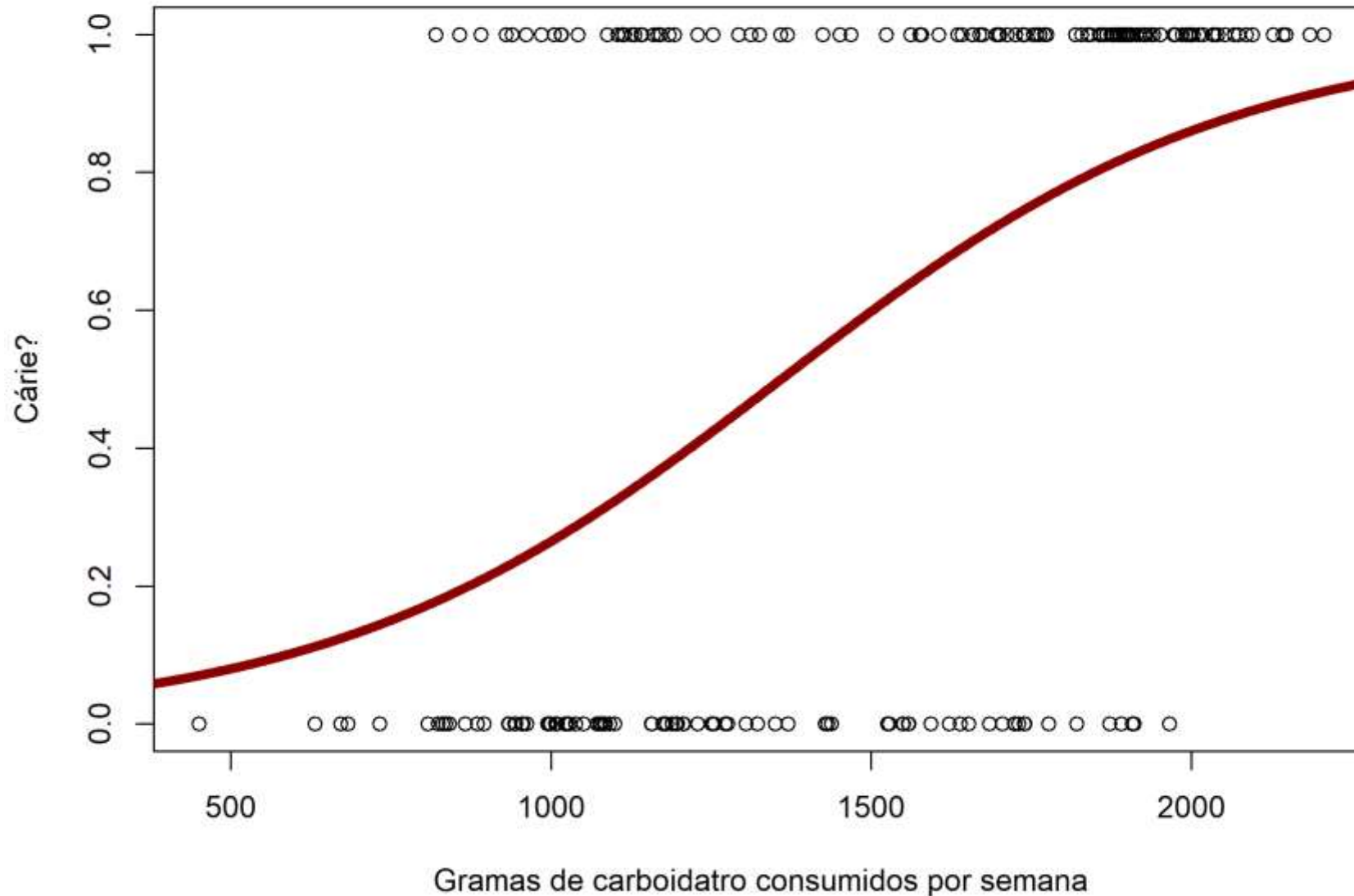
# Comer doces em excesso causa diabetes?



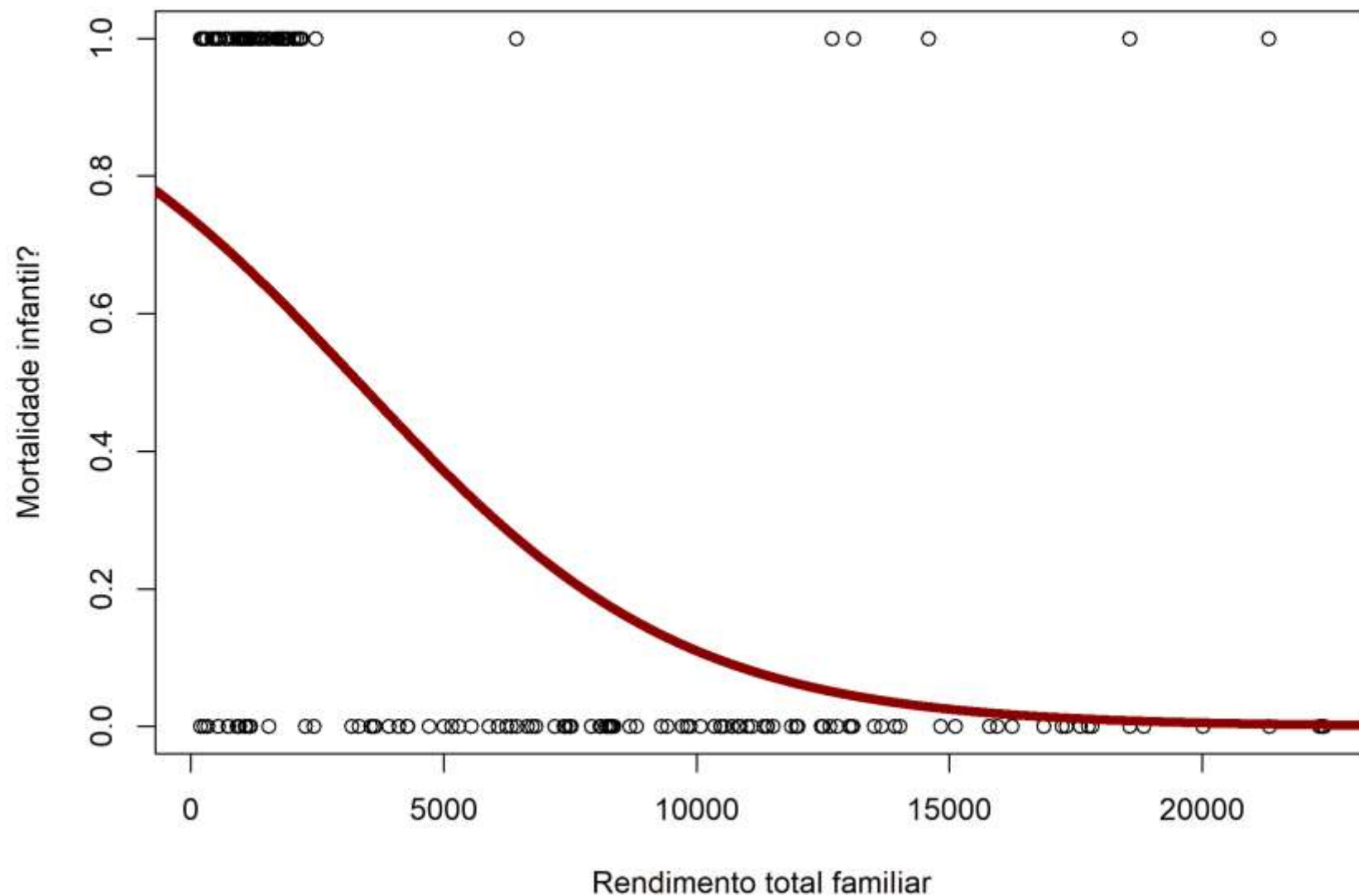
# Consumir álcool em excesso pode prejudicar o fígado?



# O consumo de alimentos ricos em açúcar aumenta o risco de cárie?

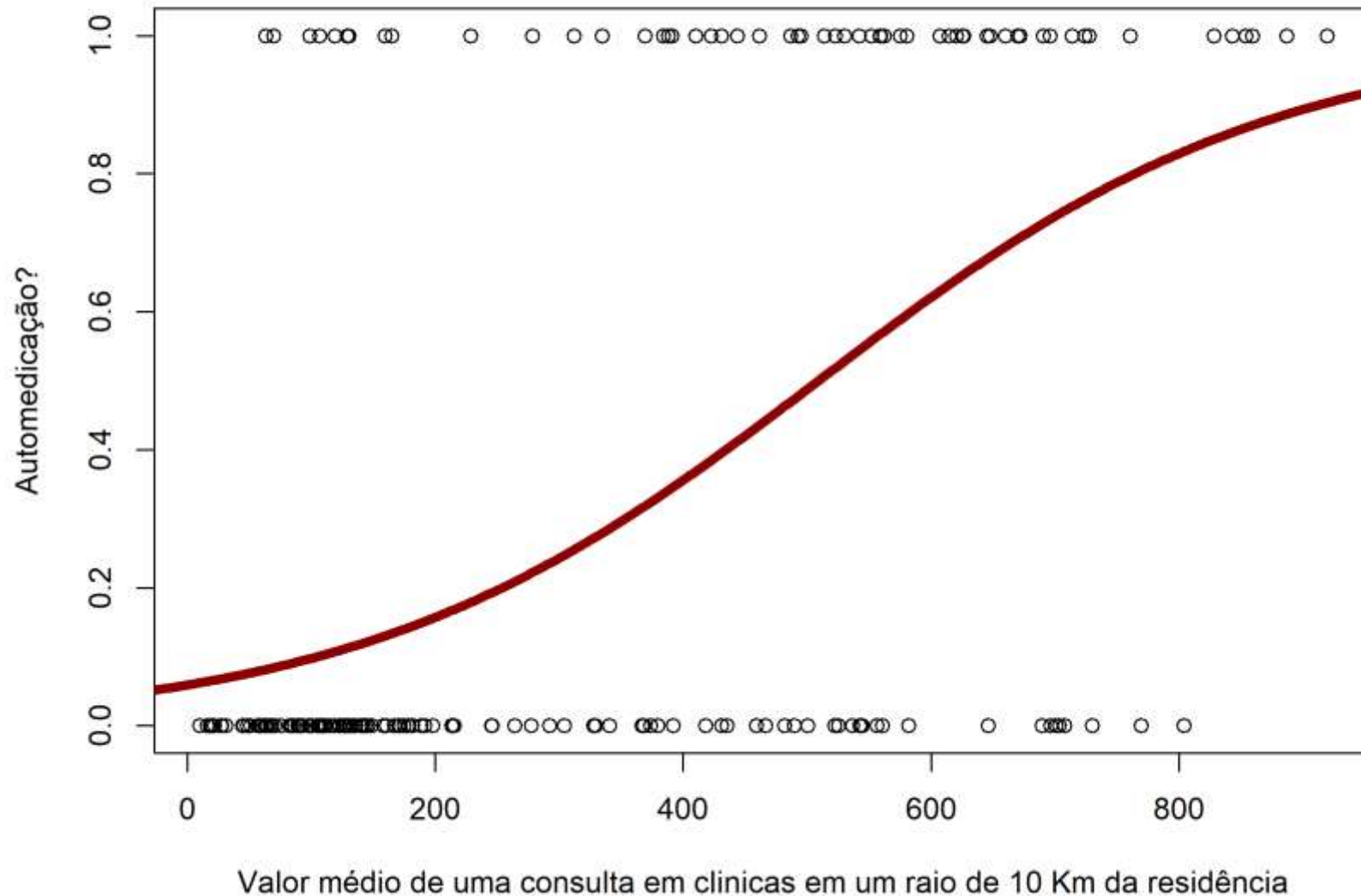


O risco de morte de um recém nascido é influenciado pela renda dos seus pais





Pessoas se automedicam mais frequentemente quando médicos de uma cidade cobram um valor elevado por uma consulta médica?

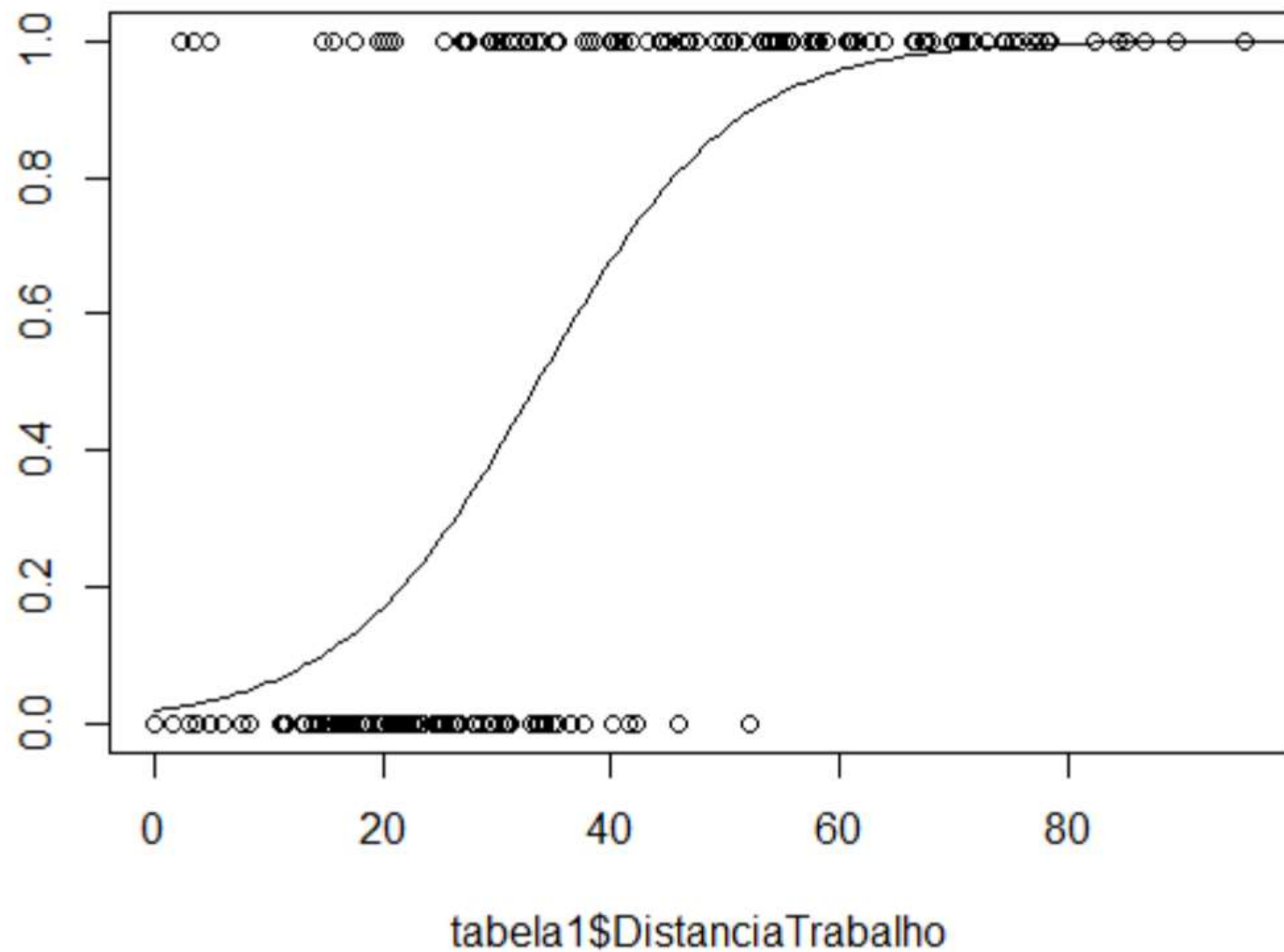


Prática

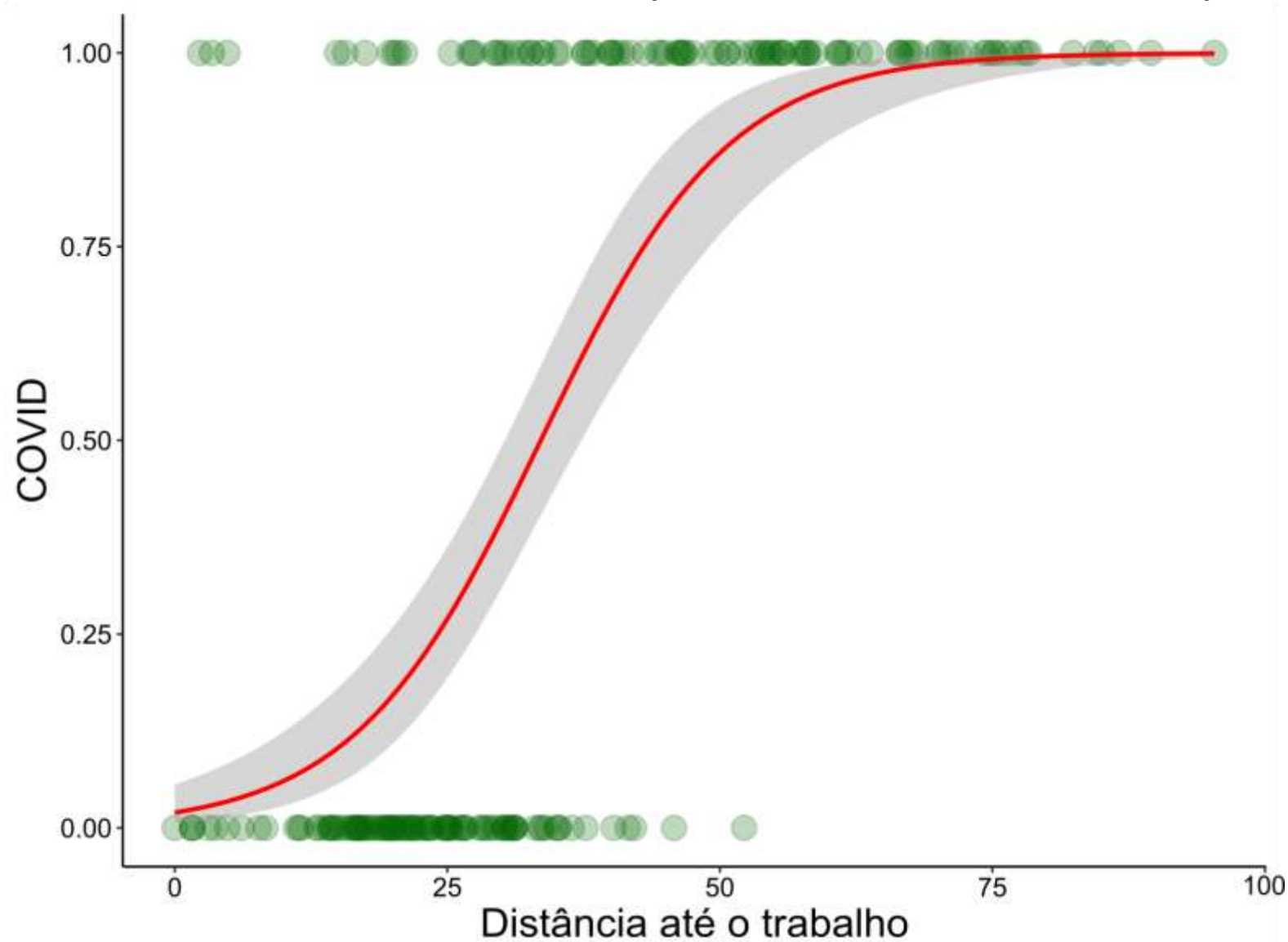
# Prática – 1 A distância ao trabalho influencia no risco de contrair COVID?

- Tabela “DadosAula12.xlsx”
  - Dados hipotéticos
- COVID: representa o diagnóstico de COVID.
  - Respostas possíveis: 1 = positivo; 0 = negativo
- Distancia: distância em Km da residência do paciente até o seu local de trabalho.
  - Respostas possíveis: variável quantitativa contínua

tabela1\$COVID



# Figura mais rebuscada (material extra)



# Interpretação dos resultados

```
> summary(B1)
```

```
Call:
```

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

```
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-2.1339	-0.6702	-0.0804	0.4846	2.7064

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-3.90484	0.55359	-7.054	1.74e-12	***
DistanciaTrabalho	0.11630	0.01665	6.987	2.81e-12	***

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

Hipóteses testadas

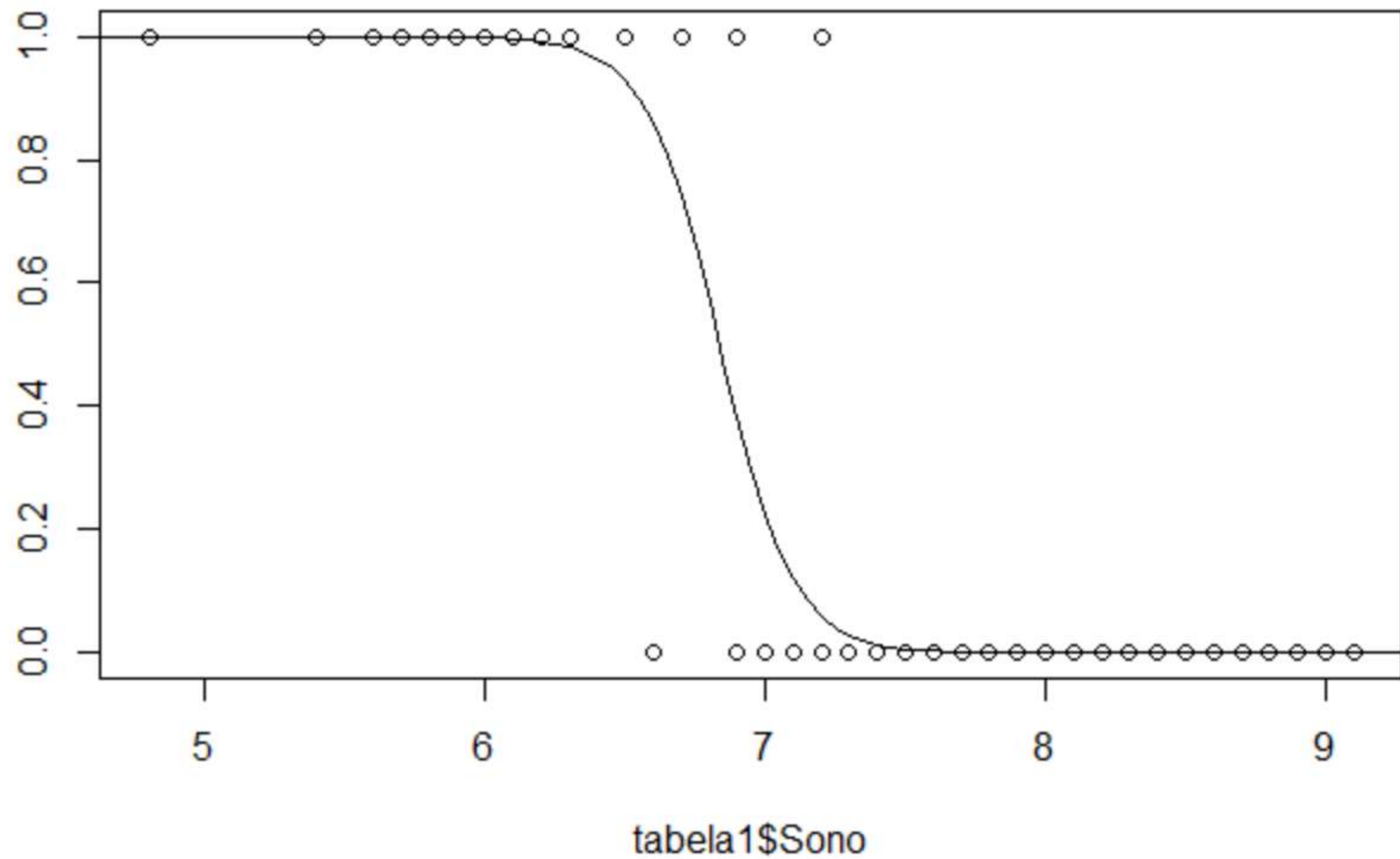
$H_0$ : Não há diferenças entre os resultados previstos e os valores observados

$H_1$ : Há (SIM) diferenças entre os resultados previstos e os valores observados

# Prática – 2 Acidentes de transito estão relacionado ao numero de horas dormidas na noite anterior?

- Tabela “DadosAula12.xlsx”
  - Dados hipotéticos
- AcidenteTransito : representa o registrou ou não de acidente.
  - Respostas possíveis: 1 = sofreu acidente; 0 = não sofreu acidente
- Sono: Numero de horas de sono do motorista 1 dia antes de pegar estrada.
  - Respostas possíveis: variável quantitativa contínua

tabela1\$AcidenteTransito





# Interpretação dos resultados

```
> summary(B1)
```

```
Call:
```

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

```
Deviance Residuals:
```

Min	1Q	Median	3Q	Max
-1.98022	-0.03580	-0.01139	-0.00169	2.38009

```
Coefficients:
```

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	52.196	15.060	3.466	0.000528	***
Sono	-7.634	2.175	-3.510	0.000449	***

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

Hipóteses testadas

$H_0$ : Não há diferenças entre os resultados previstos e os valores observados

$H_1$ : Há (SIM) diferenças entre os resultados previstos e os valores observados