

# Estimação

## Carregando base de dados

```
data = utils::read.csv(file = "pense2019_microdados.csv")
```

## Limpendo os dados

```
library(janitor)

# Selecionando variáveis importantes e filtrando registros válidos.

clean_data = janitor::clean_names(data) |>
  dplyr::rename(
    tabagismo = b04003,
    sexo = b01001a,
    idade = b01003,
    mes_aniversario = b01004,
    ano_nascimento = b01005,
    cor = b01002,
    celular = b01014,
    computador = b01015b,
    internet = b01016,
    carro = b01017,
    moto = b01018a,
    banheiro = b01019a,
    escolaridade_mae = b01008b,
    ano_escolar = b01021a,
    atividade_fisica = b03006b,
    bebida_alcoolica = b05004a,
    dias_tristes = b12005,
    sentimento_corpo = b11007,
    frequencia_pais_problemas = b07004,
    amigos_fumaram = b04016,
    pais_fumam = b04006b,
    escola_frequencia_violencia_colegas = b07012,
    escola_acesso_internet = e01p09a,
    escola_atividades_fisicas = e01p74,
    escola_prevencao_uso_tabaco = e01p8506,
    escola_consumo_cigarro_professores = e01p26a,
  ) |>
  dplyr::select(
    "regiao",
```

```

"uf",
"municipio_cap",
"escola",
"aluno",
"tabagismo",
"sexo",
"idade",
"mes_aniversario",
"ano_nascimento",
"cor",
"celular",
"computador",
"internet",
"carro",
"moto",
"banheiro",
"escolaridade_mae",
"ano_escolar",
"atividade_fisica",
"bebida_alcoolica",
"dias_tristes",
"sentimento_corpo",
"frequencia_pais_problemas",
"amigos_fumaram",
"pais_fumam",
"escola_frequencia_violencia_colegas",
"escola_acesso_internet",
"escola_atividades_fisicas",
"escola_prevencao_uso_tabaco",
"escola_consumo_cigarro_professores"
) |>
dplyr::filter(
  if_all(
    c(
      "tabagismo",
      "sexo",
      "idade",
      "cor",
      "celular",
      "computador",
      "internet",
      "mes_aniversario",
      "ano_nascimento",
      "carro",
      "moto",
      "banheiro",
      "escolaridade_mae",
      "ano_escolar",
      "atividade_fisica",
      "bebida_alcoolica",
      "dias_tristes",
      "sentimento_corpo",
      "frequencia_pais_problemas",

```

```

    "amigos_fumaram",
    "pais_fumam",
    "escola_frequencia_violencia_colegas",
    "escola_acesso_internet",
    "escola_atividades_fisicas",
    "escola_prevencao_uso_tabaco",
    "escola_consumo_cigarro_professores"
  ),
  ~ !.x %in% c(-2, 9, 99, -1, 9999))
) |>
tidyr::drop_na()

```

## Ajustando variáveis

```

# Sexo.
# 1, se homem.
# 0, se mulher.
clean_data$sexo <- ifelse(clean_data$sexo == 1, 1, 0)

# Idade (Ano da pesquisa - ano de nascimento)
clean_data$idade <- 2019 - clean_data$ano_nascimento

# Cor
# 1, se Branco/Amarelo.
# 0, caso contrário.sudo apt -y install libfontconfig1-dev
clean_data$cor <- ifelse(clean_data$cor %in% c(1, 3), 1, 0)

# Celular
# 1, se tem celular.
# 0, caso contrário.
clean_data$celular <- ifelse(clean_data$celular == 1, 1, 0)

# Computador
# 1, se tem computador
# 0, caso contrário.
clean_data$computador <- ifelse(clean_data$computador == 1, 1, 0)

# Internet
# 1, se tem internet.
# 0, caso contrário.
clean_data$internet <- ifelse(clean_data$internet == 1, 1, 0)

# Carro
# 1, se tem carro
# 0, caso contrário.
clean_data$carro <- ifelse(clean_data$carro == 1, 1, 0)

# Moto
# 1, se tem moto.
# 0, caso contrário.

```

```

clean_data$moto <- ifelse(clean_data$moto == 1, 1, 0)

# Banheiro
# 0, não tem banheiro.
# 1, caso contrário (1 ou mais banheiros).
clean_data$banheiro <- ifelse(clean_data$banheiro == 1, 0, 1)

# Dummies escolaridade da mãe:

## mae_nao_estudou:
## 1, se não estudou.
## 0, caso contrário.
clean_data$mae_nao_estudou <- ifelse(clean_data$escolaridade_mae == 1, 1, 0)

## mae_concluiu_ensino_medio.
## 1, se concluiu pelo menos ensino médio.
## 0, caso contrário.
clean_data$mae_concluiu_ensino_medio <- ifelse(clean_data$escolaridade_mae %in% c(5,6,7), 1, 0)

## mae_concluiu_ensino_fundamental.
## 1, se concluiu o ensino fundamental.
## 0, caso contrário.
clean_data$mae_concluiu_ensino_fundamental <- ifelse(clean_data$escolaridade_mae == 3, 1, 0)

## mae_concluiu_ensino_superior
## 1, se concluiu o ensino superior.
## 0, caso contrário.
clean_data$mae_concluiu_ensino_superior <- ifelse(clean_data$escolaridade_mae == 7, 1, 0)

# Distorção idade-série:
# Distorção idade-série = Idade - Anos de Estudo - 7.
# Anos de estudo := ano escolar + 5 (Nesta base de dados).
# Se distorção idade-serie < 0, 0.
clean_data$distorcao_idade_serie <- ifelse(
  (clean_data$idade - (clean_data$ano_escolar + 5) - 7) < 0,
  0,
  (clean_data$idade - (clean_data$ano_escolar + 5) - 7)
)

# Dummies atividade física:
## af_nenhum_dia:
## 1, se não fez atividade física nenhum dia.
## 0, caso contrário.
clean_data$af_nenhum_dia <- ifelse(clean_data$atividade_fisica == 1, 1, 0)

## af_1_a_3_dias
## 1, se fez atividade física de 1 à 3 dias.
## 0, caso contrário.
clean_data$af_1_a_3_dias <- ifelse(
  clean_data$atividade_fisica %in% c(2, 3, 4), 1, 0
)

```

```

## af_4_a_5_dias
## 1, se fez atividade física de 4 à 5 dias.
## 0, caso contrário.
clean_data$af_4_a_5_dias <- ifelse(
  clean_data$atividade_fisica %in% c(5, 6), 1, 0
)

## af_6_a_7_dias
## 1, se fez atividade física de 6 à 7 dias.
## 0, caso contrário.
clean_data$af_6_a_7_dias <- ifelse(
  clean_data$atividade_fisica %in% c(7, 8), 1, 0
)

# Dummies Bebida Alcoólica
## ba_nenhum_dia
## 1, se não ingeriu bebida alcoólica nos últimos 30 dias.
## 0, caso contrário.
clean_data$ba_nenhum_dia <- ifelse(clean_data$bebida_alcoolica == 1, 1, 0)

## ba_1_a_9_dias
## 1, se ingeriu bebida alcoólica de 1 a 9 dias, nos últimos 30 dias.
## 0, caso contrário.
clean_data$ba_1_a_9_dias <- ifelse(
  clean_data$bebida_alcoolica %in% c(2, 3, 4), 1, 0
)

## ba_10_a_19_dias
## 1, se ingeriu bebida alcoólica de 10 a 19 dias, nos últimos 30 dias.
## 0, caso contrário.
clean_data$ba_10_a_19_dias <- ifelse(clean_data$bebida_alcoolica == 5, 1, 0)

## ba_20_a_todos_dias
## 1, se ingeriu bebida alcoólica de 20 a todos dias, nos últimos 30 dias.
## 0, caso contrário.
clean_data$ba_20_a_todos_dias <- ifelse(
  clean_data$bebida_alcoolica %in% c(6, 7), 1, 0
)

# Dummy Saúde Mental
# 1, caso tenha se sentido triste "na maioria das vezes" ou "sempre".
# 0, caso contrário.
clean_data$triste_na_maioria_das_vezes <- ifelse(
  clean_data$dias_tristes %in% c(1,2,3), 0, 1
)

# Dummies Sentimento em relação ao corpo

## corpo_satisfeito
## 1, caso esteja satisfeito/muito satisfeito com o corpo.
## 0, caso contrário.
clean_data$corpo_satisfeito <- ifelse(

```

```

    clean_data$sentimento_corpo %in% c(1,2), 1, 0
  )

  ## corpo_insatisfeito
  ## 1, caso esteja insatisfeito/muito insatisfeito com o corpo.
  ## 0, caso contrário.
  clean_data$corpo_insatisfeito <- ifelse(
    clean_data$sentimento_corpo %in% c(4,5), 1, 0
  )

  ## corpo_indiferente
  ## 1, caso esteja indiferente em relação ao corpo.
  ## 0, caso contrário.
  clean_data$corpo_indiferente <- ifelse(clean_data$sentimento_corpo == 3, 1, 0)

# Dummies Compreensão familiar

## compreensao_nunca
## 1, se os pais não compreendem nunca.
## 0, caso contrário.
clean_data$compreensao_nunca <- ifelse(
  clean_data$frequencia_pais_problemas == 1, 1, 0
)

## compreensao_as_vezes
## 1, se os pais compreendem as vezes
## 0, caso contrário.
clean_data$compreensao_as_vezes <- ifelse(
  clean_data$frequencia_pais_problemas %in% c(2,3), 1, 0
)

## compreensao_sempre
## 1, se os pais compreendem sempre
## 0, caso contrário.
clean_data$compreensao_sempre <- ifelse(
  clean_data$frequencia_pais_problemas %in% c(4,5), 1, 0
)

# Dummy Amigos fumantes
# 1, caso algum amigo tenha fumado na sua presença nos últimos 30 dias.
# 0, caso contrário.
clean_data$amigos_fumaram <- ifelse(clean_data$amigos_fumaram == 1, 1, 0)

# Dummy Pais fumantes
# 1, se algum deles fuma.
# 0, caso contrário.
clean_data$pais_fumam <- ifelse(clean_data$pais_fumam %in% c(2,3,4), 1, 0)

# Dummy Internet Escola
# 1, se tem internet para os alunos.
# 0, caso contrário.
clean_data$escola_acesso_internet <- ifelse(
  clean_data$escola_acesso_internet == 1, 1, 0
)

```

```

)

# Dummy Atividade Física Escola
# 1, se tem prática de atividade física para os alunos.
# 0, caso contrário.
clean_data$escola_atividades_fisicas <- ifelse(
  clean_data$escola_atividades_fisicas == 1, 1, 0
)

# Dummy Violência na escola
# 1, se o aluno foi agredido nos últimos 30 dias.
# 0, caso contrário.
clean_data$escola_agredido <- ifelse(
  clean_data$escola_frequencia_violencia_colegas == 1, 0, 1
)

# Dummy Prevenção uso de Tabaco
# 1, se a escola fez alguma campanha de prevenção do
# uso de tabaco nos últimos 12 meses.
# 0, caso contrário.
clean_data$escola_prevencao_uso_tabaco <- ifelse(
  clean_data$escola_prevencao_uso_tabaco == 1, 1, 0
)

# Dummy consumo de cigarro professores/funcionarios
# 1, se professores e/ou funcionários fumaram nas dependências da escola.
# 0, caso contrário.
clean_data$escola_consumo_cigarro_professores_e_ou_funcionarios <- ifelse(
  clean_data$escola_consumo_cigarro_professores %in% c(1,2,3), 1, 0
)

# Transformando a variável dependente em um fator:

clean_data$tabagismo <- factor(as.character(clean_data$tabagismo))

```

## Estimando o modelo

A amostra restante (alunos e escolas Não Sem resposta) é a abordagem mais utilizada para analisar dados ordinais hierárquicos. Para um resultado de nível  $k$ , a probabilidade cumulativa de sucesso nas divisões cumulativas  $K - 1$  é baseado em um modelo logit cumulativo para as respostas,  $R_{ij}$ , para o aluno  $i$  na escola  $j$ . Utilizando a terminologia de RAUDENBUSH; BRYK (2002), o modelo é caracterizado por nível, da seguinte forma:

$$\text{Nível 1:} \quad \eta_{kij} = \ln \left( \frac{P(R_{ij} \leq k)}{P(R_{ij} \geq k)} \right) = \beta_{0j} + \sum_{q=1}^Q \beta_{qij} + \sum_{k=2}^{K-1} D_{kij} \rho_k$$

$$\text{Nível 2:} \quad \beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + \mu_{qj}$$

Neste modelo,  $\eta_{kij}$  é o preditor logístico para a comparação cumulativa  $k$  e para o aluno  $i$  na escola  $j$ . Como o logit é o logaritmo natural da razão de chances (*log-odds*) da probabilidade de sucesso, podemos utilizar a

relação abaixo para obter a razão de chances (*odds*),  $\pi_{kij}$ :

$$\pi_{kij}(x) = \frac{\exp(\eta_{kij})}{1 + \exp(\eta_{kij})} = \frac{odds_{(kij)}}{1 + odds_{(kij)}}$$

Onde  $x$  é um vetor de variáveis explicativas de nível 1 e nível 2.

Para cada aluno, uma série de probabilidades  $K-1$  é determinada a partir do modelo, cada uma representando a probabilidade da resposta estar em ou abaixo de uma determinada categoria, condicionada ao conjunto de preditores (O'CONNELL, 2010).

A  $K$ -ésima probabilidade será sempre igual a 1, uma vez que todas as respostas devem estar em ou abaixo do nível  $K$  nos dados. Para cada unidade ou grupo de nível 2, a equação de regressão no nível 1, fornece um conjunto único de interceptos e coeficientes de regressão. A suposição de probabilidades proporcionais sustenta que, em todas as divisões cumulativas  $K-1$  dos dados, os coeficientes de inclinações são constantes, embora variem de grupo para grupo. No nível 2, a variabilidade nos interceptos e coeficientes de inclinação entre os grupos, é capturada pelos resíduos do nível 2,  $\mu_{qj}$  (O'CONNELL, 2010).

A variação nas estimativas de parâmetros de regressão aleatória pode ser modelada usando preditores do nível 2,  $W_{sj}$ , que não precisam ser os mesmos para cada coeficiente de regressão do nível 1. Os  $\gamma_{qs}$  do nível 2 são os coeficientes de regressão fixos. A medida em que a explicação dos coeficientes aleatórios de nível 1 melhora com a adição de preditores apropriados de nível 2, os resíduos no nível 2 tornam-se menores. Esses resíduos são considerados normalmente distribuídos com a matriz var-cov  $T$ :  $\mu_{1j} \sim N(0, T)$  (O'CONNELL, 2010).

Para ajustar o modelo, podemos utilizar o pacote '*ordinal*' (CHRISTENSEN, 2015) do R que contém a função `clmm()` (*Cumulative Link Mixed-Effects*), utilizado e recomendado por BOUSQUET et al. (2017) e AGRETI (2002).

## Modelo de interceptos

Estimando o *Cumulative Link Mixed-Effects* nulo (*intercept-only*):

```
library(ordinal)

logit_nulo <- clmm(tabagismo ~ 1 + (1 | escola), data=clean_data)
summary(logit_nulo)
```

Cumulative Link Mixed Model fitted with the Laplace approximation

```
formula: tabagismo ~ 1 + (1 | escola)
data:      clean_data
```

```
link threshold nobs logLik AIC niter max.grad cond.H
logit flexible 23648 -24176.38 48366.76 754(1400) 1.13e-03 1.1e+05
```

Random effects:

```
Groups Name      Variance Std.Dev.
escola (Intercept) 9.235e-09 9.61e-05
Number of groups: escola 69
```

No Coefficients

Threshold coefficients:

```
Estimate Std. Error z value
```



1 2	0.91310	0.01439	63.48
2 3	1.77333	0.01846	96.06
3 4	2.19525	0.02166	101.36
4 5	2.51145	0.02468	101.76
5 6	2.97001	0.03018	98.41
6 7	3.35590	0.03603	93.13

### Logit ordenado (nível aluno)

Estimando o *Cumulative Link Mixed-Effects* apenas com preditores de nível 1 (aluno):

```
logit_ordenado <- clm(
  tabagismo ~ sexo + idade + cor + celular + computador +
  internet + carro + moto + banheiro +
  mae_nao_estudou +
  mae_concluiu_ensino_fundamental +
  distorcao_idade_serie + af_1_a_3_dias + af_4_a_5_dias +
  af_6_a_7_dias + ba_1_a_9_dias + ba_10_a_19_dias +
  ba_20_a_todos_dias +
  triste_na_maioria_das_vezes + corpo_indiferente +
  corpo_insatisfeito + compreensao_sempre +
  compreensao_as_vezes + pais_fumam + amigos_fumaram +
  escola_acesso_internet +
  escola_agredido +
  escola_atividades_fisicas +
  escola_prevencao_uso_tabaco +
  escola_consumo_cigarro_professores_e_ou_funcionarios,
  data = clean_data
)
summary(logit_ordenado)
```

formula:

```
tabagismo ~ sexo + idade + cor + celular + computador + internet + carro + moto + banheiro + mae_nao_
data:      clean_data
```

	link	threshold	nobs	logLik	AIC	niter	max.grad	cond.H
logit	flexible	23648	-21506.25	43084.50	6(1)	4.80e-09	3.4e+05	

Coefficients:

	Estimate	Std. Error
sexo	0.3652457	0.0352317
idade	-0.0217067	0.0107812
cor	0.0114568	0.0326914
celular	-0.2445899	0.0468830
computador	-0.0223769	0.0367245
internet	-0.2026650	0.0541060
carro	0.0101726	0.0357166
moto	-0.0478443	0.0317083
banheiro	-0.2741629	0.1007273
mae_nao_estudou	0.2024537	0.0769587
mae_concluiu_ensino_fundamental	0.0778524	0.0662472
distorcao_idade_serie	0.2388156	0.0320012
af_1_a_3_dias	-0.0274834	0.0373735

af_4_a_5_dias	-0.1066656	0.0471301
af_6_a_7_dias	-0.0950756	0.0504653
ba_1_a_9_dias	1.0155527	0.0388598
ba_10_a_19_dias	1.8140162	0.0617313
ba_20_a_todos_dias	2.1673053	0.0754734
triste_na_maioria_das_vezes	0.1782484	0.0352458
corpo_indiferente	0.0173862	0.0503217
corpo_insatisfeito	0.0039832	0.0377649
compreensao_sempre	-0.2329625	0.0430942
compreensao_as_vezes	-0.1797089	0.0394198
pais_fumam	0.2007963	0.0329810
amigos_fumaram	1.7040565	0.0391858
escola_acesso_internet	-0.0150385	0.0325721
escola_agredido	0.1299688	0.0411689
escola_atividades_fisicas	-0.0100325	0.0794902
escola_prevencao_uso_tabaco	0.0289559	0.0306596
escola_consumo_cigarro_professores_e_ou_funcionarios	0.0006604	0.0363691
z value Pr(> z )		
sexo	10.367	< 2e-16 ***
idade	-2.013	0.04407 *
cor	0.350	0.72600
celular	-5.217	1.82e-07 ***
computador	-0.609	0.54231
internet	-3.746	0.00018 ***
carro	0.285	0.77579
moto	-1.509	0.13133
banheiro	-2.722	0.00649 **
mae_nao_estudou	2.631	0.00852 **
mae_concluiu_ensino_fundamental	1.175	0.23992
distorcao_idade_serie	7.463	8.48e-14 ***
af_1_a_3_dias	-0.735	0.46211
af_4_a_5_dias	-2.263	0.02362 *
af_6_a_7_dias	-1.884	0.05957 .
ba_1_a_9_dias	26.134	< 2e-16 ***
ba_10_a_19_dias	29.386	< 2e-16 ***
ba_20_a_todos_dias	28.716	< 2e-16 ***
triste_na_maioria_das_vezes	5.057	4.25e-07 ***
corpo_indiferente	0.346	0.72972
corpo_insatisfeito	0.105	0.91600
compreensao_sempre	-5.406	6.45e-08 ***
compreensao_as_vezes	-4.559	5.14e-06 ***
pais_fumam	6.088	1.14e-09 ***
amigos_fumaram	43.487	< 2e-16 ***
escola_acesso_internet	-0.462	0.64430
escola_agredido	3.157	0.00159 **
escola_atividades_fisicas	-0.126	0.89957
escola_prevencao_uso_tabaco	0.944	0.34495
escola_consumo_cigarro_professores_e_ou_funcionarios	0.018	0.98551

---

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

Threshold coefficients:

	Estimate	Std. Error	z value
1 2	2.0516	0.2180	9.41

2 3	3.0805	0.2186	14.09
3 4	3.5598	0.2190	16.25
4 5	3.9104	0.2194	17.82
5 6	4.4073	0.2202	20.01
6 7	4.8174	0.2211	21.79

### Logit ordenado multinível

Estimando o *Cumulative Link Mixed-Effects* completo, com variáveis do nível 1 e 2 (aluno e escola), captando também, efeitos aleatórios do nível 2:

```
logit_ordenado_multinivel <- clmm(
  tabagismo ~ sexo + idade + cor + celular + computador +
  internet + carro + moto + banheiro +
  mae_nao_estudou +
  mae_concluiu_ensino_fundamental +
  distorcao_idade_serie + af_1_a_3_dias + af_4_a_5_dias +
  af_6_a_7_dias + ba_1_a_9_dias + ba_10_a_19_dias +
  ba_20_a_todos_dias +
  triste_na_maioria_das_vezes + corpo_indiferente +
  corpo_insatisfeito + compreensao_sempre +
  compreensao_as_vezes + pais_fumam + amigos_fumaram +
  escola_agredido +
  (1|escola) + escola_acesso_internet +
  escola_atividades_fisicas +
  escola_prevencao_uso_tabaco +
  escola_consumo_cigarro_professores_e_ou_funcionarios,
  data = clean_data
)
summary(logit_ordenado_multinivel)
```

Cumulative Link Mixed Model fitted with the Laplace approximation

```
formula: tabagismo ~ sexo + idade + cor + celular + computador + internet +
  carro + moto + banheiro + mae_nao_estudou + mae_concluiu_ensino_fundamental +
  distorcao_idade_serie + af_1_a_3_dias + af_4_a_5_dias + af_6_a_7_dias +
  ba_1_a_9_dias + ba_10_a_19_dias + ba_20_a_todos_dias + triste_na_maioria_das_vezes +
  corpo_indiferente + corpo_insatisfeito + compreensao_sempre +
  compreensao_as_vezes + pais_fumam + amigos_fumaram + escola_agredido +
  (1 | escola) + escola_acesso_internet + escola_atividades_fisicas +
  escola_prevencao_uso_tabaco + escola_consumo_cigarro_professores_e_ou_funcionarios
data:      clean_data
```

link	threshold	nobs	logLik	AIC	niter	max.grad	cond.H
logit	flexible	23648	-21507.53	43089.06	8618(18516)	3.37e+01	2.0e+05

Random effects:

Groups	Name	Variance	Std.Dev.
escola	(Intercept)	0.002896	0.05381

Number of groups: escola 69

Coefficients:

Estimate	Std. Error
----------	------------

sexo	0.3608998	0.0352385
idade	-0.0222743	0.0108207
cor	0.0125847	0.0327007
celular	-0.2440478	0.0469131
computador	-0.0215480	0.0367504
internet	-0.2060432	0.0541111
carro	0.0092998	0.0357528
moto	-0.0507181	0.0317235
banheiro	-0.2728795	0.1007717
mae_nao_estudou	0.2002650	0.0769854
mae_concluiu_ensino_fundamental	0.0690443	0.0663651
distorcao_idade_serie	0.2377168	0.0320640
af_1_a_3_dias	-0.0308876	0.0373943
af_4_a_5_dias	-0.1074839	0.0471429
af_6_a_7_dias	-0.0884940	0.0504392
ba_1_a_9_dias	1.0170581	0.0388704
ba_10_a_19_dias	1.8121156	0.0617725
ba_20_a_todos_dias	2.1752193	0.0754643
triste_na_maioria_das_vezes	0.1818055	0.0352497
corpo_indiferente	0.0236241	0.0502799
corpo_insatisfeito	-0.0009193	0.0377853
compreensao_sempre	-0.2321853	0.0431044
compreensao_as_vezes	-0.1798098	0.0394318
pais_fumam	0.1995478	0.0329885
amigos_fumaram	1.6993421	0.0391570
escola_agredido	0.1310252	0.0411740
escola_acesso_internet	-0.0154915	0.0326489
escola_atividades_fisicas	-0.0077546	0.0796662
escola_prevencao_uso_tabaco	0.0288576	0.0307731
escola_consumo_cigarro_professores_e_ou_funcionarios	0.0015054	0.0364590
z value Pr(> z )		
sexo	10.242	< 2e-16 ***
idade	-2.058	0.03954 *
cor	0.385	0.70035
celular	-5.202	1.97e-07 ***
computador	-0.586	0.55765
internet	-3.808	0.00014 ***
carro	0.260	0.79478
moto	-1.599	0.10987
banheiro	-2.708	0.00677 **
mae_nao_estudou	2.601	0.00929 **
mae_concluiu_ensino_fundamental	1.040	0.29817
distorcao_idade_serie	7.414	1.23e-13 ***
af_1_a_3_dias	-0.826	0.40881
af_4_a_5_dias	-2.280	0.02261 *
af_6_a_7_dias	-1.754	0.07935 .
ba_1_a_9_dias	26.165	< 2e-16 ***
ba_10_a_19_dias	29.335	< 2e-16 ***
ba_20_a_todos_dias	28.824	< 2e-16 ***
triste_na_maioria_das_vezes	5.158	2.50e-07 ***
corpo_indiferente	0.470	0.63846
corpo_insatisfeito	-0.024	0.98059
compreensao_sempre	-5.387	7.18e-08 ***
compreensao_as_vezes	-4.560	5.11e-06 ***

```

pais_fumam                6.049 1.46e-09 ***
amigos_fumaram            43.398 < 2e-16 ***
escola_agredido           3.182 0.00146 **
escola_acesso_internet    -0.474 0.63515
escola_atividades_fisicas -0.097 0.92246
escola_prevencao_uso_tabaco 0.938 0.34837
escola_consumo_cigarro_professores_e_ou_funcionarios 0.041 0.96706
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Threshold coefficients:

	Estimate	Std. Error	z value
1 2	2.0345	0.2185	9.312
2 3	3.0632	0.2190	13.986
3 4	3.5427	0.2195	16.143
4 5	3.8934	0.2199	17.707
5 6	4.3906	0.2207	19.898
6 7	4.8009	0.2215	21.670

## Testes

### ICC

O Coeficiente de Correlação Intraclass (ICC) fornece uma avaliação de quanta variabilidade nas respostas está no nível do grupo. Quando os dados são dicotômicos, a variabilidade dentro do grupo é definida pela distribuição amostral dos dados, normalmente a distribuição de Bernoulli. Quando o modelo de distribuição logística é aplicado, assume-se que os resíduos de nível um seguem a distribuição logística padrão, que tem uma média de 0 e uma variância de  $\pi^2/3 = 3,29$  (O'CONNELL, 2010):

$$ICC = \frac{Var(\mu_{qj})}{Var(\mu_{qj}) + 3,29}$$

Ou seja, o ICC é a proporção entre a variância a nível de escola  $Var(\mu_{qj})$  e a variância total, dos dois níveis.

```

ICC = 0.001769/(0.001769 + 3.29)
ICC

```

```
[1] 0.000537401
```

O ICC pode assumir valores entre 0 e 1. Quanto mais próximo de 0, quer dizer que menos ocorre variância dos alunos entre as escolas, ou seja, a probabilidade de ter algum nível de tabagismo independe da escola em que o aluno está inserido. Nessa situação, deve-se utilizar modelos convencionais, considerando apenas um nível. Como  $ICC = 0,0005$ , quer dizer que apenas 0,05% das variações na probabilidade de ter algum nível de tabagismo deve-se às heterogeneidades da escola. O que indica que deve-se utilizar modelos convencionais.

### Likelihood-ratio

O LR test (Likelihood-ratio test ou Teste da razão de verossimilhanças) avalia a qualidade do ajuste de dois modelos estatísticos concorrentes com base na razão de suas verossimilhanças. Se a hipótese nula não for rejeitada, significa que ambos os modelos não se diferem de forma significativa.

```
anova(logit_ordenado_multinivel, logit_ordenado)
```

Likelihood ratio tests of cumulative link models:

```

                                formula:
logit_ordenado                 tabagismo ~ sexo + idade + cor + celular + computador + internet + carro +
logit_ordenado_multinivel      tabagismo ~ sexo + idade + cor + celular + computador + internet + carro +
                                link: threshold:
logit_ordenado                 logit flexible
logit_ordenado_multinivel      logit flexible

                                no.par   AIC logLik LR.stat df Pr(>Chisq)
logit_ordenado                 36 43085 -21506
logit_ordenado_multinivel      37 43089 -21508 -2.5571  1          1

```

Como  $p > 0,05$ , aceitamos a hipótese nula e optamos pelo modelo mais simples, já que o ICC também foi inferior à 0,1.

## Resultados

### Odds Ratio (Razão de chances)

Razão de chance para cada preditor com intervalo de confiança de 5%:

```

ci <- confint(logit_ordenado)

exp(
  cbind(
    OR = tail(
      logit_ordenado$coefficients,
      n=(length(logit_ordenado$coefficients) - 6)
    ),
    ci
  )
)

```

	OR	2.5 %
sexo	1.4408679	1.3447912
idade	0.9785272	0.9580702
cor	1.0115227	0.9487001
celular	0.7830256	0.7144470
computador	0.9778716	0.9100078
internet	0.8165518	0.7346154
carro	1.0102245	0.9419649
moto	0.9532822	0.8957873
banheiro	0.7602082	0.6247365
mae_nao_estudou	1.2244033	1.0520104
mae_concluiu_ensino_fundamental	1.0809631	0.9484921
distorcao_idade_serie	1.2697444	1.1923805
af_1_a_3_dias	0.9728909	0.9041463
af_4_a_5_dias	0.8988262	0.8193427
af_6_a_7_dias	0.9093042	0.8234658
ba_1_a_9_dias	2.7608888	2.5591303
ba_10_a_19_dias	6.1350375	5.4355627
ba_20_a_todos_dias	8.7347145	7.5327667

triste_na_maioria_das_vezes	1.1951221	1.1153411
corpo_indiferente	1.0175382	0.9216125
corpo_insatisfeito	1.0039911	0.9323025
compreensao_sempre	0.7921833	0.7280422
compreensao_as_vezes	0.8355134	0.7734514
pais_fumam	1.2223758	1.1457704
amigos_fumaram	5.4961973	5.0920297
escola_acesso_internet	0.9850740	0.9242167
escola_agredido	1.1387928	1.0502871
escola_atividades_fisicas	0.9900176	0.8482065
escola_prevencao_uso_tabaco	1.0293792	0.9693316
escola_consumo_cigarro_professores_e_ou_funcionarios	1.0006606	0.9316519
	97.5 %	
sexo	1.5439578	
idade	0.9994292	
cor	1.0784120	
celular	0.8585971	
computador	1.0509142	
internet	0.9081928	
carro	1.0835302	
moto	1.0143478	
banheiro	0.9273239	
mae_nao_estudou	1.4225406	
mae_concluiu_ensino_fundamental	1.2298058	
distorcao_idade_serie	1.3517791	
af_1_a_3_dias	1.0468051	
af_4_a_5_dias	0.9856096	
af_6_a_7_dias	1.0036066	
ba_1_a_9_dias	2.9802457	
ba_10_a_19_dias	6.9238005	
ba_20_a_todos_dias	10.1264332	
triste_na_maioria_das_vezes	1.2805963	
corpo_indiferente	1.1225982	
corpo_insatisfeito	1.0810615	
compreensao_sempre	0.8620319	
compreensao_as_vezes	0.9027018	
pais_fumam	1.3039067	
amigos_fumaram	5.9375466	
escola_acesso_internet	1.0500899	
escola_agredido	1.2342362	
escola_atividades_fisicas	1.1584208	
escola_prevencao_uso_tabaco	1.0931225	
escola_consumo_cigarro_professores_e_ou_funcionarios	1.0744138	

## Referências

- AGRESTI, A. **Categorical data analysis, Second Edition**. New York, New York: Wiley, 2002.
- BOUSQUET, C. A. H. et al. [Determinants of leadership in groups of female mallards](#). **Behaviour**, v. 154, n. 4, p. 467–507, 2017.
- CHRISTENSEN, R. H. B. **ordinal — Regression models for ordinal data**. Disponível em: <<http://www.cran.r-project.org/package=ordinal/>>.
- O'CONNELL, A. A. **An illustration of multilevel models for ordinal response data**. Data and

context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8). **Anais...**2010.

RAUDENBUSH, S. W.; BRYK, A. S. **Hierarchical linear models: Applications and data analysis methods.** [s.l.] sage, 2002. v. 1