

Modelo Ordinal Final com Suposição de Proporcionalidade

Samuel Sobral Miller

2025-07-01

Carregamento e Tratamento dos Dados

```
dados <- read_csv("dados_limpo.csv", col_types = cols(  
  gestation = col_integer(),  
  parity = col_integer(),  
  race = col_factor(),  
  age = col_integer(),  
  ed = col_factor(),  
  ht = col_double(),  
  wt.1 = col_double(),  
  drace = col_factor(),  
  dage = col_integer(),  
  ded = col_factor(),  
  inc = col_factor(),  
  smoke = col_factor(),  
  time = col_factor(),  
  number = col_factor(),  
  low_birth_weigh = col_factor()  
)  
)  
  
dados = dados %>% filter(number != "fumo_mas_não_sei")  
  
# Recategorização de variáveis  
dados$ed <- fct_collapse(dados$ed,  
  Ensino_Medio_Incompleto = c("menor_8ª", "8ª-12ª", "Ensino_Medio_incompleto+curso_profissional"),  
  Ensino_Medio = c("Ensino_Medio", "Ensino_Medio+curso_profissional"),  
  Graduando = c("Ensino_Medio+curso_universitário"),  
  Ensino_Superior = c("Graduação_universitária")  
)  
  
dados$ded <- fct_collapse(dados$ded,  
  Ensino_Medio_Incompleto = c("menor_8ª", "8ª-12ª", "Ensino_Medio_incompleto+curso_profissional"),  
  Ensino_Medio = c("Ensino_Medio"),  
  Graduando = c("Ensino_Medio+curso_universitário"),  
  Ensino_Superior = c("Graduação_universitária")  
)  
  
dados$time <- fct_collapse(dados$time,  
  "Nunca" = "nunca",  
  "Fuma" = "ainda_fuma",  
  "-2_anos" = c("dentro_de_1_ano", "1_a_2_anos_atrás"),
```

```

"-5_anos" = c("2_a_3_anos_atrás", "3_a_4_anos_atrás"),
"+5_anos" = c("5_a_9_anos_atrás", "há_mais_de_10_anos"),
"Durante" = "durante_a_gravidez_atual",
"Nao_Sabe" = "desisti_e_não_sei"
)

dados$number <- fct_collapse(dados$number,
  "1-4" = "1-4",
  "5-9" = "5-9",
  "10-19" = c("10-14", "15-19"),
  "20-29" = "20-29",
  "30-39" = "30-39",
  "40+" = c("40-60", "60+"),
  "Nunca" = "nunca"
)

# Recodificação de fatores omitida para brevidade, assumimos que já foi feita.

# Subconjunto relevante
db = dados %>% select(gestation, parity, wt.1, number, race, low_birth_weigh)

# Variável resposta como ordinal
db$low_birth_weigh = factor(db$low_birth_weigh,
  levels = c("Baixo", "Médio", "Alto"),
  ordered = TRUE)

# Reordenação da referência
db$number = relevel(factor(db$number), ref = "Nunca")

```

Ajuste do Modelo Final (back2)

```

modelo_final <- glm(low_birth_weigh ~ gestation + parity + wt.1 + number + race,
  data = db, link = "logit")

```

```

## Warning: (3) Model is nearly unidentifiable: large eigenvalue ratio
## - Rescale variables?
## In addition: Absolute and relative convergence criteria were met

```

```
summary(modelo_final)
```

```

## formula: low_birth_weigh ~ gestation + parity + wt.1 + number + race
## data: db
##
## link threshold nobs logLik AIC niter max.grad cond.H
## logit flexible 1100 -504.87 1039.74 8(0) 4.91e-11 8.1e+07
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## gestation 0.0636945 0.0064668 9.849 < 2e-16 ***
## parity 0.1450312 0.0455066 3.187 0.001437 **

```

```
## wt.1          0.0279980  0.0093745   2.987 0.002821 **
## number1-4     0.2211188  0.2654194   0.833 0.404793
## number20-29  -0.9371200  0.2738527  -3.422 0.000622 ***
## number5-9     -0.0353997  0.2602269  -0.136 0.891794
## number10-19  -0.7652249  0.3551732  -2.155 0.031200 *
## number30-39  -0.9270431  0.5787187  -1.602 0.109180
## number40+     0.0004863  0.6005419   0.001 0.999354
## raceBranco    1.0325001  0.4985900   2.071 0.038374 *
## racePreto     0.0317376  0.5238676   0.061 0.951691
## raceMexicano  1.2507060  0.6725287   1.860 0.062927 .
## raceMisto     0.4328195  0.8048105   0.538 0.590722
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##           Estimate Std. Error z value
## Baixo|Médio   16.693      1.838   9.084
## Médio|Alto    22.781      1.964  11.598
```

Teste da Suposição de Proporcionalidade

```
nominal_test(modelo_final)
```

```
## Tests of nominal effects
##
## formula: low_birth_weigh ~ gestation + parity + wt.1 + number + race
##           Df logLik   AIC   LRT Pr(>Chi)
## <none>      -504.87 1039.7
## gestation   1 -503.90 1039.8 1.9338 0.16435
## parity      1 -503.36 1038.7 3.0256 0.08196 .
## wt.1        1 -503.92 1039.8 1.9077 0.16721
## number      6 -500.24 1042.5 9.2515 0.15992
## race        4 -502.28 1042.6 5.1812 0.26921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretação: Os resultados do `nominal_test` indicam que **não há violação significativa da suposição de proporcionalidade** com as variáveis selecionadas, justificando o uso do modelo ordinal.

Conclusão

O modelo ordinal proporcional ajustado indica que as variáveis gestação, paridade, peso pré-gestacional (wt.1), número de cigarros e raça estão associadas com o peso ao nascer. O modelo respeita a suposição de proporcionalidade e apresentou melhores resíduos e ajuste do que os modelos anteriores com a variável altura.

Ajuste do Modelo Ordinal (Back2)

```
modelo_final <- clm(low_birth_weigh ~ gestation + parity + wt.1 + number + race,
  data = db, link = "logit")
```

```
## Warning: (3) Model is nearly unidentifiable: large eigenvalue ratio
## - Rescale variables?
## In addition: Absolute and relative convergence criteria were met
```

```
summary(modelo_final)
```

```
## formula: low_birth_weigh ~ gestation + parity + wt.1 + number + race
## data:      db
##
## link threshold nobs logLik AIC      niter max.grad cond.H
## logit flexible 1100 -504.87 1039.74 8(0) 4.91e-11 8.1e+07
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## gestation      0.0636945  0.0064668   9.849 < 2e-16 ***
## parity         0.1450312  0.0455066   3.187 0.001437 **
## wt.1           0.0279980  0.0093745   2.987 0.002821 **
## number1-4      0.2211188  0.2654194   0.833 0.404793
## number20-29    -0.9371200  0.2738527  -3.422 0.000622 ***
## number5-9      -0.0353997  0.2602269  -0.136 0.891794
## number10-19    -0.7652249  0.3551732  -2.155 0.031200 *
## number30-39    -0.9270431  0.5787187  -1.602 0.109180
## number40+       0.0004863  0.6005419   0.001 0.999354
## raceBranco      1.0325001  0.4985900   2.071 0.038374 *
## racePreto       0.0317376  0.5238676   0.061 0.951691
## raceMexicano    1.2507060  0.6725287   1.860 0.062927 .
## raceMisto       0.4328195  0.8048105   0.538 0.590722
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##              Estimate Std. Error z value
## Baixo|Médio    16.693      1.838   9.084
## Médio|Alto     22.781      1.964  11.598
```

Tabela com Coeficientes e ORs

```
# Carregar pacotes
library(tidyverse)
library(broom)
library(ordinal)
library(knitr)

# Ajuste do modelo (exemplo)
modelo <- clm(low_birth_weigh ~ gestation + parity + wt.1 + number + race, data = db, link = "logit")
```

```
## Warning: (3) Model is nearly unidentifiable: large eigenvalue ratio
```

```
## - Rescale variables?
## In addition: Absolute and relative convergence criteria were met

# Gerar os coeficientes com tidy
result <- tidy(modelo) %>%
  filter(str_detect(term, "[^\\|]")) %>% # Exclui os interceptos threshold (|)
  mutate(
    OR = exp(estimate),
    p_value = round(p.value, 4)
  )

# Mostrar a tabela com kable
result %>%
  select(term, estimate, std.error, OR, p_value) %>%
  knitr::kable(digits = 3, caption = "Coeficientes do Modelo Ordinal com ORs")
```

Table 1: Coeficientes do Modelo Ordinal com ORs

term	estimate	std.error	OR	p_value
Baixo Médio	16.693	1.838	1.776790e+07	0.000
Médio Alto	22.781	1.964	7.825919e+09	0.000
gestation	0.064	0.006	1.066000e+00	0.000
parity	0.145	0.046	1.156000e+00	0.001
wt.1	0.028	0.009	1.028000e+00	0.003
number1-4	0.221	0.265	1.247000e+00	0.405
number20-29	-0.937	0.274	3.920000e-01	0.001
number5-9	-0.035	0.260	9.650000e-01	0.892
number10-19	-0.765	0.355	4.650000e-01	0.031
number30-39	-0.927	0.579	3.960000e-01	0.109
number40+	0.000	0.601	1.000000e+00	0.999
raceBranco	1.033	0.499	2.808000e+00	0.038
racePreto	0.032	0.524	1.032000e+00	0.952
raceMexicano	1.251	0.673	3.493000e+00	0.063
raceMisto	0.433	0.805	1.542000e+00	0.591

```
# Carregue os pacotes necessários
library(tidyverse)
library(nnet) # para multinom
library(broom) # para tidy()

# Ajuste do modelo multinomial
modelo_multinom <- multinom(low_birth_weigh ~ smoke + parity + ed + age, data = dados)
```

```
## # weights: 30 (18 variable)
## initial value 1208.473518
## iter 10 value 578.185638
## iter 20 value 568.893079
## final value 568.885589
## converged
```

```

# Extração das estimativas com intervalo de confiança e odds ratios
resumo <- tidy(modelo_multinom, exponentiate = TRUE, conf.int = TRUE)

# Tabela formatada com ORs (odds ratios)
odds_table <- resumo |>
  select(y.level, term, estimate) |>
  pivot_wider(names_from = y.level, values_from = estimate) |>
  rename("Variável" = term,
         "Baixo vs Médio" = "Baixo",
         "Alto vs Médio" = "Alto")

# Visualizar tabela
odds_table

```

```

## # A tibble: 9 x 3
##   Variável                'Alto vs Médio' 'Baixo vs Médio'
##   <chr>                  <dbl>         <dbl>
## 1 (Intercept)           0.0318         0.00566
## 2 smokeainda_fuma       0.688          2.96
## 3 smokeate_a_gravidez_atual 1.74          1.33
## 4 smokeuma_vez_não_agora  1.99          1.14
## 5 parity                1.06          0.990
## 6 edEnsino_Medio        1.42          1.57
## 7 edEnsino_Medio_Incompleto 1.66          3.68
## 8 edGraduando           1.22          2.18
## 9 age                   1.04          1.05

```

Análise de Correspondência Múltipla (ACM)

A Análise de Correspondência Múltipla (ACM) é uma técnica exploratória ideal para examinar associações entre múltiplas variáveis categóricas. No contexto deste estudo, a ACM pode ser usada para complementar a análise descritiva e visualizar as relações entre o peso ao nascer (`low_birth_weight`) e variáveis categóricas como `smoke`, `number`, `ed`, `ded`, `race` e `inc`.

1. Pré-processamento

```

library(FactoMineR)
library(factoextra)

```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

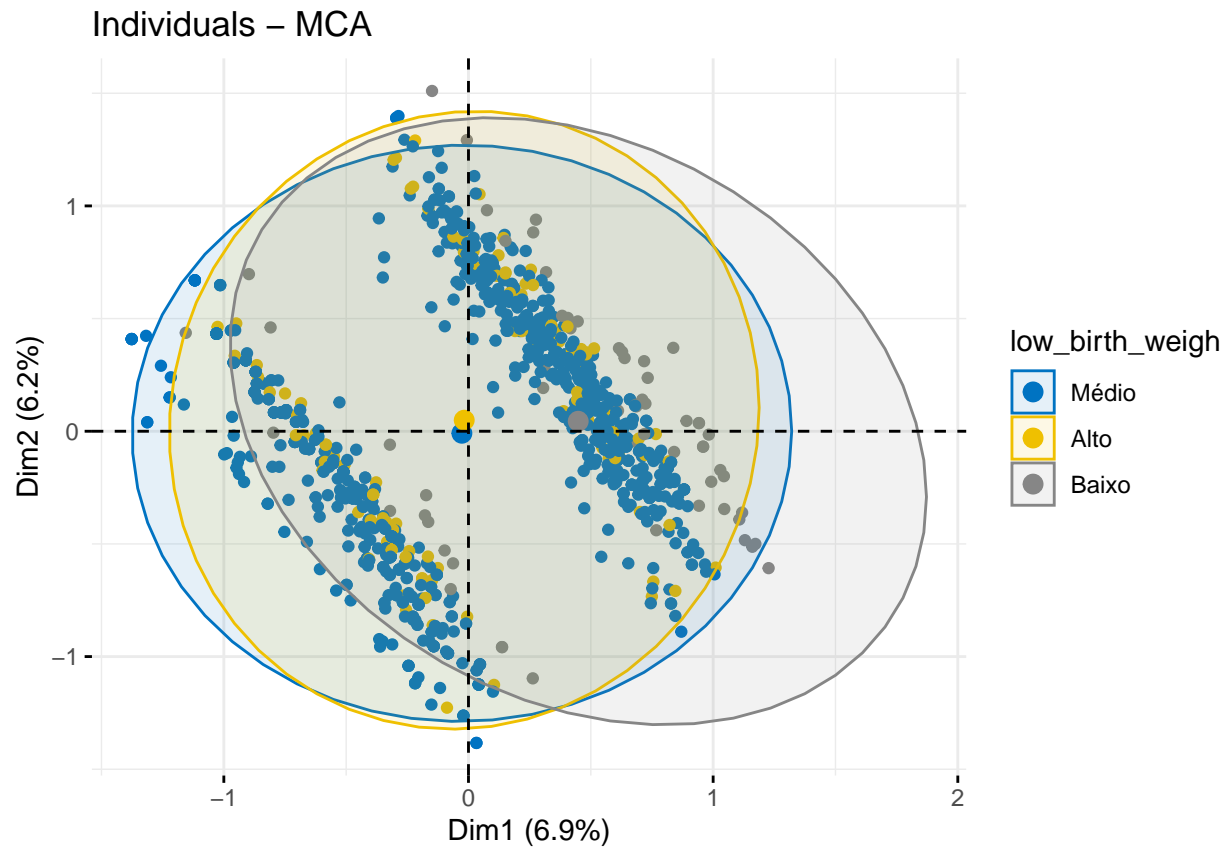
```

# Subconjunto de dados categóricos
dados_cat <- dados %>%
  select(low_birth_weight, smoke, number, ed, ded, race, inc) %>%
  na.omit() # remover NAs para análise

# Executar ACM
acm_result <- MCA(dados_cat, graph = FALSE)

```

```
# Visualizar os indivíduos
fviz_mca_ind(acm_result,
  label = "none",
  habillage = "low_birth_weigh",
  palette = "jco",
  addEllipses = TRUE)
```



```
# Visualizar categorias das variáveis
fviz_mca_var(acm_result,
  repel = TRUE,
  col.var = "contrib")
```

Variable categories – MCA

