

Exercício extra

Lucas L Oliveira

2022-07-26

Desenvolvimento

A tabela que temos para o desenvolvimento da questão segue abaixo:

			nivel_da_doenca	1	2	3	4
##	poluicao_do_ar	poluicao_no_trabalho	status_fumo				
##	Baixa	Não	Não	158	9	5	0
##			Ex	167	19	5	3
##			Sim	307	102	83	68
##		Sim	Não	26	5	5	1
##			Ex	38	12	4	4
##			Sim	94	48	46	60
##	Alta	Não	Não	94	7	5	1
##			Ex	67	8	4	3
##			Sim	184	65	33	36
##		Sim	Não	32	3	6	1
##			Ex	39	11	4	2
##			Sim	77	48	39	51

Como variável resposta temos o nível da doença em que seus níveis correspondem a:

- **1:** Sem sintomas;
- **2:** Tosse por menos que 3 meses;
- **3:** Tosse por mais que 3 meses ao ano;
- **4:** Tosse e outros sintomas por mais que 3 meses ao ano.

Claramente estamos lidando com uma variável resposta ordinal e isso deve ser incorporado ao modelo que será o logit cumulativo. Abaixo temos os resultados do modelo estimado.

```
## Call:
## polr(formula = factor(nivel_da_doenca) ~ poluicao_do_ar + poluicao_no_trabalho +
##       status_fumo, data = dados, Hess = TRUE)
##
## Coefficients:
##      poluicao_do_arAlta poluicao_no_trabalhoSim      status_fumoEx
##             -0.03930433              0.86476854              0.40000636
##             status_fumoSim
##              1.85268636
##
## Intercepts:
```

```
##      1|2      2|3      3|4
## 2.088417 2.969625 3.893834
##
## Residual Deviance: 4178.511
## AIC: 4192.511
```

Podemos ver acima quais são os interceptos e quais são as variáveis para os nossos modelos. Lembrando que a definição do modelo se dá por

$$\log \left[\frac{P(Y \leq j)}{P(Y > j)} \right].$$

Dado isso, podemos tirar algumas conclusões.

Analisando os coeficientes podemos ver que:

- Poluição alta do ar tem coeficiente menor que zero e isso implica que ele reduz a chance da doença estar em nível mais avançado. Por outro lado, a chance da poluição baixa é de $\frac{1}{\exp\{-0.03930433\}} = 1.040087$ vezes mais que o fato da poluição ser alta e nesse caso temos uma diferença bem pequena de um fator para outro, indicando que pode não haver tanto impacto da poluição do ar no resultado do nível da doença;
- Poluição no trabalho tem coeficiente maior que zero indicando um possível impacto no nível da doença. Temos que $\exp\{0.86476854\} = 2.374456$ poluição no trabalho realmente tem um impacto no nível da doença respiratória crônica;
- No mais, temos o fato de ser ex fumante com $\exp\{0.40000636\} = 1.491834$ chances de ter complicações respiratórias mais graves que um não fumante e o fato de ser fumante ativo com $\exp\{1.85268636\} = 6.376927$ mais chances de ter complicações respiratórias que um não fumante;
- Por fim, analisando os interceptos podemos concluir que o perfil que temos é poluição do ar baixa, sem poluição no trabalho e não sendo fumante. Com isso, podemos concluir que esse perfil tem $\exp\{2.088417\} = 8.072127$ mais chances de ter doenças no primeiro nível que nos níveis 2, 3 ou 4; tem $\exp\{2.969625\} = 19.48461$ mais chances de ter a doença respiratória crônica nos níveis 1 ou 2 que no nível 3 ou 4; tem $\exp\{3.893834\} = 49.09877$ mais chances de ter doenças respiratória crônica nos níveis 1, 2 ou 3 que no nível 4.

Diagnóstico

Para diagnóstico, podemos analisar a existência de multicolineariedade do modelo. Para VIF podemos ver o resultado abaixo.

```
##              GVIF Df GVIF^(1/(2*Df))
## poluicao_do_ar    1.015781  1      1.007860
## poluicao_no_trabalho 1.017008  1      1.008468
## status_fumo      1.002841  2      1.000709
```

temos indicação da presença de multicolineariedade quando VIF maior que 10 que não é o caso do nosso modelo atual.