



Regressão logística multinível

Ednaldo Ribeiro

Regressão Logística

- Adaptação dos procedimentos da regressão linear para problemas envolvendo variáveis dependentes categóricas
 - Binária, Multinomial, Probit...
- **Variáveis dependentes dicotômicas:** muitos problemas de pesquisa, sobretudo nas ciências sociais, envolvem variáveis dependentes que possuem apenas dois valores. Essas medidas são chamadas ainda de binárias ou dummies.
- **Codificação:** devido a sua natureza binária essas variáveis devem possuir a codificação 0 (representando a ausência de um atributo ou a não ocorrência de um evento) e 1 (representando a presença de um atributo ou a ocorrência de um evento).
- **Exemplos:**
 - participação em partidos (0=não participa/1=participa)
 - Democrata (0=não/1=sim)
 - Interessado em política (0=não/1=sim)

Previendo o protesto político

- Tomando um grupo de países latino-americanos para identificar quais condicionantes individuais e nacionais ajudam à prever o envolvimento em protesto político.
- **Variável dependente:**
 - "Participou de algum protesto ou manifestação pacífica nos últimos 12 meses? "
 - 0=Não
 - 1=Sim
 - O modelo pretende prever a probabilidade de 1 considerando preditores de nível 1 (indivíduo) e 2 (país).

Implementação no R

- **Preparação:** junção dos bancos de dados do Lapop de diferentes ondas com banco de dados nacionais (54 países).
 - Função *merge* a partir de uma variável identificadora comum "PA"=País Ano.

```
#JUNÇÃO DOS BANCOS  
ppar <- merge(alcompl, bd2, by = c("PA"))
```

- Banco final ppar
- Pacote *lme4*

```
install.packages("lme4")  
library(lme4)
```

Modelo Nulo

- Assim como nos modelos lineares, o primeiro passo é identificar se existe variabilidade na probabilidade de envolvimento em protestos que possa ser atribuída ao nível 2.
- Para isso rodamos o modelo nulo, sem preditores

```
Nulo.prot <- glmer(prot ~ (1 | PA), family = binomial("logit"), data = ppar)  
summary(Nulo.prot)
```

- Notem que temos a variável dependente "prot" e a variação pela variável de identificação "PA". Sem preditores!

```

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: prot ~ (1 | PA)
Data: ppar

      AIC      BIC   logLik deviance df.resid
50807.5 50826.2 -25401.7  50803.5    86747

Scaled residuals:
    Min       1Q   Median       3Q      Max
-0.4613 -0.3642 -0.2896 -0.2246  6.0354

Random effects:
 Groups Name      Variance Std.Dev.
 PA      (Intercept) 0.2588    0.5087
Number of obs: 86749, groups: PA, 54

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.46060    0.07007  -35.12   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

- As duas informações relevantes são a variância entre os países (0,26) e a probabilidade média de envolvimento em protesto (-2,46)

- É útil exponenciar essa última estimativa para tornar o efeito mais compreensível

- $\text{Exp}(-2,46)=0,085$
- A cada 100 indivíduos apenas aproximadamente 8 participam de protestos.

- Para sabermos se a variação nessa probabilidade entre os países é relevante o suficiente para a modelagem multinível precisamos comparar esse modelo nulo com um modelo alternativo que não permite essa variação de nível 2

```
Nulo.prot.A <- glm(prot ~ 1, data = ppar, family = binomial("logit"))
```

- Na sequência calculamos a diferença no LogLik dos dois modelos

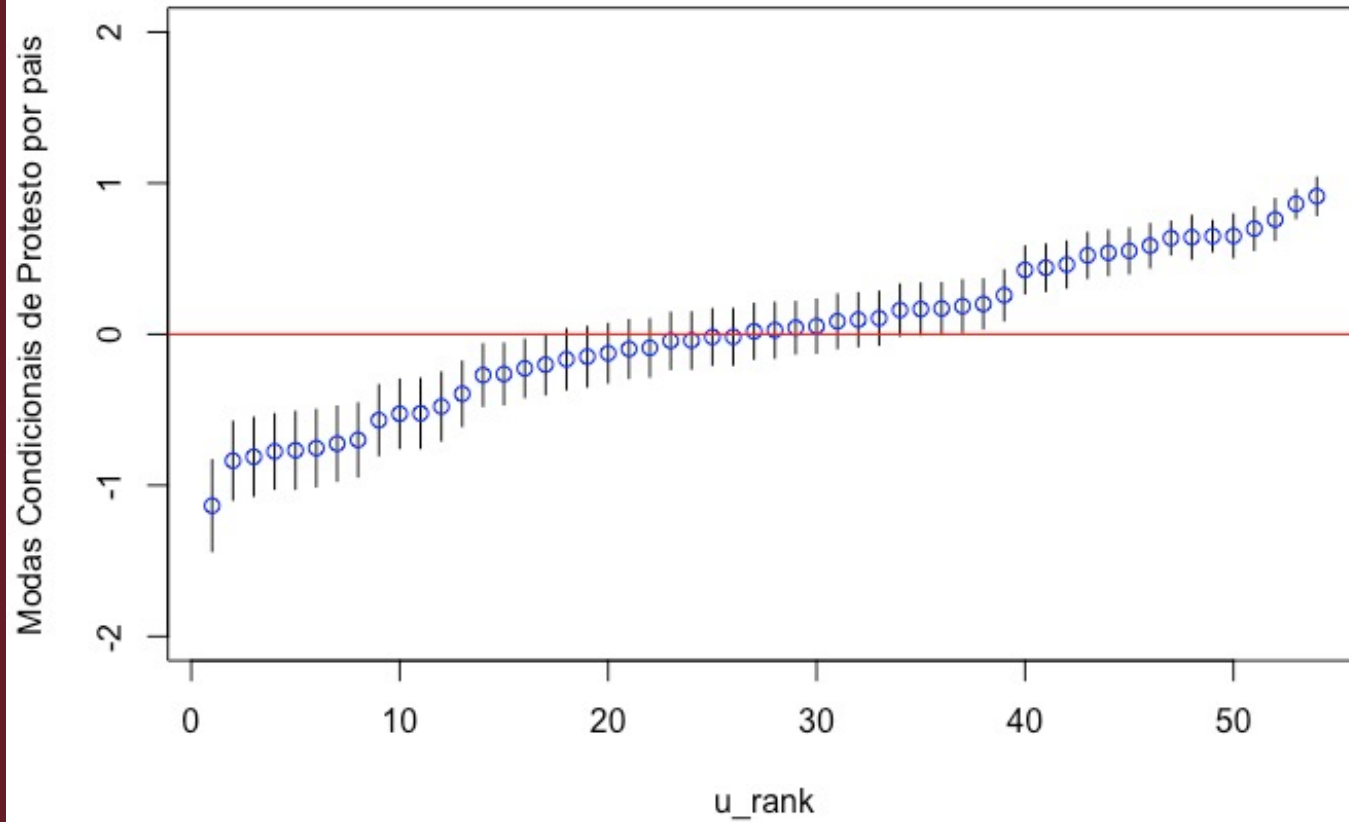
```
logLik(Nulo.prot.A)-logLik(Nulo.prot)
```

```
'log Lik.' -751.838 (df=1)
```

- O resultado indica diferença significativa entre os dois modelos, com vantagem para o Nulo, já que é o segundo termo na subtração.

- Podemos visualizar a variabilidade entre os países por meio de um gráfico Caterpillar ("Efeito País")

```
u0 <- ranef(Nulo.prot, postVar = TRUE)
u0se <- sqrt(attr(u0[[1]], "postVar")[1, , ])
commid <- as.numeric(rownames(u0[[1]]))
u0tab <- cbind("commid" = commid, "u0" = u0[[1]], "u0se" = u0se)
colnames(u0tab)[2] <- "u0"
u0tab <- u0tab[order(u0tab$u0), ]
u0tab <- cbind(u0tab, c(1:dim(u0tab)[1]))
u0tab <- u0tab[order(u0tab$commid), ]
colnames(u0tab)[4] <- "u0rank"
plot(u0tab$u0rank, u0tab$u0, type = "n", xlab = "u_rank", ylab = "Modas Condicionais de Protesto por pais", ylim = c(-2, 2))
segments(u0tab$u0rank, u0tab$u0 - 1.96*u0tab$u0se, u0tab$u0rank, u0tab$u0 + 1.96*u0tab$u0se)
points(u0tab$u0rank, u0tab$u0, col = "blue")
abline(h = 0, col = "red")
```

- No eixo Y temos as modas de probabilidade de protesto por país
- No X o ranking dos países por essas médias
- Para cada país temos o intervalo de confiança
- Devemos observar o quanto as modas se distinguem, pois isso indica variabilidade

Modelo micro

- Confirmada a variabilidade entre os países, podemos começar a modelagem inserindo as variáveis de nível individual
 - Denominação religiosa (ref=Católico)
 - Associativismo
 - Educação
 - Interesse por política
 - Satisfação com a democracia
 - Satisfação com a vida
 - Confiança nas Instituições
 - Ideologia

```
Mod.prot.n1 <- glmer(prot ~ denr.evan + assb.n + educ + intpol.n + sdem.n + svida.n  
                    + confi.in + ideo + (1 | PA), family = binomial("logit"), data = ppar)  
summary(Mod.prot.n1)
```

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-3.310535	0.110816	-29.874	< 2e-16	***
denr.evan1	0.057055	0.039186	1.456	0.1454	
denr.evan2	0.431768	0.045825	9.422	< 2e-16	***
denr.evan3	0.071625	0.070845	1.011	0.3120	
assb.n	0.416696	0.014971	27.833	< 2e-16	***
educ1	0.468648	0.032197	14.556	< 2e-16	***
intpol.n	0.461858	0.014767	31.276	< 2e-16	***
sdem.n	-0.045942	0.021699	-2.117	0.0342	*
svida.n	-0.117670	0.018970	-6.203	5.54e-10	***
confi.in	-0.735588	0.068878	-10.680	< 2e-16	***
ideo	-0.038663	0.005471	-7.067	1.59e-12	***

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

- Os resultados para denominação religioso não indicam diferenças significativas na probabilidade de envolvimento em protestos entre católicos (referência) e evangélicos (1) e ateus (3)
- Diferença no grupo dos "Outras religiões" (2), com vantagem em relação aos católicos.
- Efeitos positivos de associativismo, educação e interesse por política.
- Efeitos negativos de satisfação com a democracia, satisfação com a vida, confiança nas instituições e ideologia.

- Para calcular os exponenciais das estimativas

```
exp(fixef(Mod.prot.n1))
```

```
(Intercept)  denr.evan1  denr.evan2  denr.evan3  assb.n  educ1  intpol.n  sdem.n  svida.n
0.03649665  1.05871405  1.53997845  1.07425239  1.51694105  1.59783207  1.58702049  0.95509747  0.88898972
confi.in      ideo
0.47922353  0.96207529
```

- Para saber as estimativas por país

```
coef(Mod.prot.n1)
```

```
$PA
(Intercept)  denr.evan1  denr.evan2  denr.evan3  assb.n  educ1  intpol.n  sdem.n  svida.n
112  -3.441524  0.05705501  0.4317684  0.07162496  0.4166958  0.4686478  0.4618584 -0.04594188 -0.1176696
114  -3.850407  0.05705501  0.4317684  0.07162496  0.4166958  0.4686478  0.4618584 -0.04594188 -0.1176696
117  -3.302409  0.05705501  0.4317684  0.07162496  0.4166958  0.4686478  0.4618584 -0.04594188 -0.1176696
212  -3.451094  0.05705501  0.4317684  0.07162496  0.4166958  0.4686478  0.4618584 -0.04594188 -0.1176696
214  -4.621457  0.05705501  0.4317684  0.07162496  0.4166958  0.4686478  0.4618584 -0.04594188 -0.1176696
217  -3.281446  0.05705501  0.4317684  0.07162496  0.4166958  0.4686478  0.4618584 -0.04594188 -0.1176696
312  -3.943453  0.05705501  0.4317684  0.07162496  0.4166958  0.4686478  0.4618584 -0.04594188 -0.1176696
314  -4.065419  0.05705501  0.4317684  0.07162496  0.4166958  0.4686478  0.4618584 -0.04594188 -0.1176696
317  -4.187224  0.05705501  0.4317684  0.07162496  0.4166958  0.4686478  0.4618584 -0.04594188 -0.1176696
412  -3.490106  0.05705501  0.4317684  0.07162496  0.4166958  0.4686478  0.4618584 -0.04594188 -0.1176696
414  -3.588190  0.05705501  0.4317684  0.07162496  0.4166958  0.4686478  0.4618584 -0.04594188 -0.1176696
```

Modelo completo

- Com a adição de
 - V-Dem (escala 0-10)
 - Restrição governamental à religiões (escala de 0-10)
 - Suporte oficial à religião específica (Católica; escala de 0-10)

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-4.015237	0.511033	-7.857	3.93e-15	***
denr.evan1	0.057946	0.039186	1.479	0.1392	
denr.evan2	0.433628	0.045874	9.452	< 2e-16	***
denr.evan3	0.071814	0.070851	1.014	0.3108	
assb.n	0.417106	0.014983	27.839	< 2e-16	***
educ1	0.468669	0.032201	14.554	< 2e-16	***
intpol.n	0.462080	0.014770	31.286	< 2e-16	***
sdem.n	-0.045558	0.021704	-2.099	0.0358	*
svida.n	-0.118060	0.018973	-6.223	4.89e-10	***
confi.in	-0.733364	0.068898	-10.644	< 2e-16	***
ideo	-0.038682	0.005473	-7.068	1.57e-12	***
vdem10	0.033078	0.039543	0.837	0.4029	
Shipew10	0.066494	0.035238	1.887	0.0592	.
ARDA_SUPORTOFICIAL	0.041567	0.039231	1.060	0.2893	

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

- A única variável com efeito estatisticamente significativo (ainda assim com 0,059) foi a medida de restrição governamental à liberdade religiosa, elevando a probabilidade de protesto.
- Como se trata de modelo incremental, pessoas de outras religiões que tem maior chance de protestar tem essa probabilidade ampliada nos contextos com restrições.

Modelo com interações cross-level

- Nessa etapa podemos interagir os efeitos da medida de restrição governamental com a denominação religiosa, para testar a hipótese de que nesses contextos os filiados à religiões não dominantes são mais propensos ao protesto.

```
Mod.prot.R.D <- glmer (prot ~ denr.evan + denr.evan:Shipew10 + assb.n + educ + intpol.n + sdem.n +  
  svida.n + confi.in + ideo + vdem10 + Shipew10 + ARDA_SUPORTOFICIAL + (1 | PA), family =  
  binomial("logit"), data = ppar)
```

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-3.980757	0.510578	-7.797	6.36e-15	***
denr.evan1	0.040017	0.051649	0.775	0.4385	
denr.evan2	0.354823	0.061441	5.775	7.70e-09	***
denr.evan3	0.086321	0.096884	0.891	0.3729	
assb.n	0.417099	0.014983	27.837	< 2e-16	***
educ1	0.468240	0.032205	14.539	< 2e-16	***
intpol.n	0.461442	0.014774	31.234	< 2e-16	***
sdem.n	-0.045366	0.021707	-2.090	0.0366	*
svida.n	-0.118102	0.018976	-6.224	4.85e-10	***
confi.in	-0.730965	0.068925	-10.605	< 2e-16	***
ideo	-0.038644	0.005473	-7.061	1.65e-12	***
vdem10	0.033194	0.039489	0.841	0.4006	
Shipew10	0.058534	0.035615	1.644	0.1003	
ARDA_SUPORTOFICIAL	0.039338	0.039175	1.004	0.3153	
denr.evan1:Shipew10	0.008608	0.017442	0.494	0.6216	
denr.evan2:Shipew10	0.038078	0.019435	1.959	0.0501	.
denr.evan3:Shipew10	-0.005241	0.028456	-0.184	0.8539	

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

- A interação com “outras religiões” com nível de significância de 0,05 indica que em contextos com restrições à liberdade religiosa esse grupo tem maior probabilidade de protestar!
- Cada ponto adicional na escala de restrições eleva o efeito dessa denominação em 0,038.

- Para os exponenciais

```
exp(fixef(Mod.prot.R.D))
```

(Intercept)	denr.evan1	denr.evan2	denr.evan3	assb.n
0.01867151	1.04082822	1.42592859	1.09015657	1.51755261
educ1	intpol.n	sdem.n	svida.n	confi.in
1.59717991	1.58635985	0.95564757	0.88860514	0.48144440
ideo	vdem10	Shipew10	ARDA_SUPORTOFICIAL	denr.evan1:Shipew10
0.96209315	1.03375116	1.06028120	1.04012194	1.00864509
denr.evan2:Shipew10	denr.evan3:Shipew10			
1.03881190	0.99477234			

- Cada avanço na escala de restrição eleva em 3,8% o efeito da denominação, que sozinha já elevava a chance de protesto em 42,59%.