

ANÁLISE DE DADOS MULTIVARIADOS I - REGRESSÃO

(Lista de exercício 4)

Novembro e dezembro de 2018

Reinaldo Soares de Camargo

Modelos de Regressão Logística

- **4ª Lista de exercícios para entregar em 03/12/2018.**
 - Os exercícios podem ser entregues em grupos de 2 alunos, e o grupo deve submeter o código em R utilizado para responder ao exercício, juntamente com a discussão dos resultados.
 - Utilize a base de dados do IDH brasil 2010 (IDH_Brasil_2010.csv)
 - Rode a regressão logística abaixo:

```
mod5.reduzido <- glm(formula = alta_mort_infantil ~ renda_per_capita  
  + indice_gini  
  + salario_medio_mensal  
  + perc_crianças_extrem_pobres  
  + perc_crianças_pobres  
  + perc_pessoas_dom_agua_estogo_inadequados  
  + perc_pessoas_dom_paredes_inadequadas  
  + perc_pop_dom_com_coleta_lixo  
  + perc_pop_rural  
  + as.factor(Regiao),  
  family = binomial(link = "logit"), data = dados3)  
summary(mod5.reduzido)
```

- Questão 1: Interprete os coeficientes da regressão que apresentem significância estatística;

Modelos de Regressão Logística

- Questão 1: Interprete os coeficientes da regressão que apresentem significância estatística;

```
> odds.ratio(mod5.reduzido)
waiting for profiling to be done...
```

	OR	2.5 %	97.5 %	p
(Intercept)	2.6954e+00	3.5765e-01	19.8171	0.3320800
renda_per_capita	9.8618e-01	9.8213e-01	0.9901	1.610e-11 ***
indice_gini	2.1191e+01	5.2598e-01	877.6071	0.1065654
salario_medio_mensal	9.3111e-01	7.8001e-01	1.1070	0.4185865
perc_crianças_extrem_pobres	9.5890e-01	9.3577e-01	0.9824	0.0007045 ***
perc_crianças_pobres	1.0214e+00	9.9191e-01	1.0518	0.1570583
perc_pessoas_dom_agua_estogo_inadequados	1.0010e+00	9.9271e-01	1.0094	0.8214049
perc_pessoas_dom_paredes_inadequadas	1.0224e+00	1.0109e+00	1.0343	0.0001470 ***
perc_pop_dom_com_coleta_lixo	9.9770e-01	9.8858e-01	1.0066	0.6156627
perc_pop_rural	6.8686e-01	3.5231e-01	1.3351	0.2688692
as.factor(Regiao)Nordeste	1.0837e+01	4.8904e+00	28.8175	8.538e-08 ***
as.factor(Regiao)Norte	2.4183e+00	1.0457e+00	6.6080	0.0563536 .
as.factor(Regiao)Sudeste	2.8510e-01	1.0580e-01	0.8532	0.0165362 *
as.factor(Regiao)Sul	1.1618e-06	1.6150e-09	0.0000	0.9566043

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

Uma razão de chances (odds-ratio) de 1 indica que a condição ou evento sob estudo é igualmente provável de ocorrer nos dois grupos. Uma razão de chances maior do que 1 indica que a condição ou evento tem maior probabilidade de ocorrer no primeiro grupo. Finalmente, uma razão de chances menor do que 1 indica que a probabilidade é menor no primeiro grupo do que no segundo.

Modelos de Regressão Logística

- Questão 1: Interprete os coeficientes da regressão que apresentem significância estatística;

```
> odds.ratio(mod5.reduzido)
waiting for profiling to be done...
```

	OR	2.5 %	97.5 %	p
(Intercept)	2.6954e+00	3.5765e-01	19.8171	0.3320800
renda_per_capita	9.8618e-01	9.8213e-01	0.9901	1.610e-11 ***
indice_gini	2.1191e+01	5.2598e-01	877.6071	0.1065654
salario_medio_mensal	9.3111e-01	7.8001e-01	1.1070	0.4185865
perc_crianças_extrem_pobres	9.5890e-01	9.3577e-01	0.9824	0.0007045 ***
perc_crianças_pobres	1.0214e+00	9.9191e-01	1.0518	0.1570583
perc_pessoas_dom_agua_estogo_inadequados	1.0010e+00	9.9271e-01	1.0094	0.8214049
perc_pessoas_dom_paredes_inadequadas	1.0224e+00	1.0109e+00	1.0343	0.0001470 ***
perc_pop_dom_com_coleta_lixo	9.9770e-01	9.8858e-01	1.0066	0.6156627
perc_pop_rural	6.8686e-01	3.5231e-01	1.3351	0.2688692
as.factor(Regiao)Nordeste	1.0837e+01	4.8904e+00	28.8175	8.538e-08 ***
as.factor(Regiao)Norte	2.4183e+00	1.0457e+00	6.6080	0.0563536 .
as.factor(Regiao)Sudeste	2.8510e-01	1.0580e-01	0.8532	0.0165362 *
as.factor(Regiao)Sul	1.1618e-06	1.6150e-09	0.0000	0.9566043

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

Resposta:

O aumento de uma unidade na renda per capita, aumenta a chance de um município ser classificado como de baixa mortalidade infantil (não alta mortalidade infantil) em 1,38% (1-0,986), *ceteris paribus*.

Modelos de Regressão Logística

- Questão 1: Interprete os coeficientes da regressão que apresentem significância estatística;

```
> odds.ratio(mod5.reduzido)
waiting for profiling to be done...
```

	OR	2.5 %	97.5 %	p
(Intercept)	2.6954e+00	3.5765e-01	19.8171	0.3320800
renda_per_capita	9.8618e-01	9.8213e-01	0.9901	1.610e-11 ***
indice_gini	2.1191e+01	5.2598e-01	877.6071	0.1065654
salario_medio_mensal	9.3111e-01	7.8001e-01	1.1070	0.4185865
perc_crianças_extrem_pobres	9.5890e-01	9.3577e-01	0.9824	0.0007045 ***
perc_crianças_pobres	1.0214e+00	9.9191e-01	1.0518	0.1570583
perc_pessoas_dom_agua_estogo_inadequados	1.0010e+00	9.9271e-01	1.0094	0.8214049
perc_pessoas_dom_paredes_inadequadas	1.0224e+00	1.0109e+00	1.0343	0.0001470 ***
perc_pop_dom_com_coleta_lixo	9.9770e-01	9.8858e-01	1.0066	0.6156627
perc_pop_rural	6.8686e-01	3.5231e-01	1.3351	0.2688692
as.factor(Regiao)Nordeste	1.0837e+01	4.8904e+00	28.8175	8.538e-08 ***
as.factor(Regiao)Norte	2.4183e+00	1.0457e+00	6.6080	0.0563536 .
as.factor(Regiao)Sudeste	2.8510e-01	1.0580e-01	0.8532	0.0165362 *
as.factor(Regiao)Sul	1.1618e-06	1.6150e-09	0.0000	0.9566043

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

Resposta:

O aumento de uma unidade no percentual de crianças extremamente pobres, aumenta a chance de um município ser classificado como de baixa mortalidade infantil (não alta mortalidade infantil) em 4,11% (1-0,959), *ceteris paribus*.

Modelos de Regressão Logística

- Questão 1: Interprete os coeficientes da regressão que apresentem significância estatística;

```
> odds.ratio(mod5.reduzido)
waiting for profiling to be done...
```

	OR	2.5 %	97.5 %	p
(Intercept)	2.6954e+00	3.5765e-01	19.8171	0.3320800
renda_per_capita	9.8618e-01	9.8213e-01	0.9901	1.610e-11 ***
indice_gini	2.1191e+01	5.2598e-01	877.6071	0.1065654
salario_medio_mensal	9.3111e-01	7.8001e-01	1.1070	0.4185865
perc_crianças_extrem_pobres	9.5890e-01	9.3577e-01	0.9824	0.0007045 ***
perc_crianças_pobres	1.0214e+00	9.9191e-01	1.0518	0.1570583
perc_pessoas_dom_agua_estogo_inadequados	1.0010e+00	9.9271e-01	1.0094	0.8214049
perc_pessoas_dom_paredes_inadequadas	1.0224e+00	1.0109e+00	1.0343	0.0001470 ***
perc_pop_dom_com_coleta_lixo	9.9770e-01	9.8858e-01	1.0066	0.6156627
perc_pop_rural	6.8686e-01	3.5231e-01	1.3351	0.2688692
as.factor(Regiao)Nordeste	1.0837e+01	4.8904e+00	28.8175	8.538e-08 ***
as.factor(Regiao)Norte	2.4183e+00	1.0457e+00	6.6080	0.0563536 .
as.factor(Regiao)Sudeste	2.8510e-01	1.0580e-01	0.8532	0.0165362 *
as.factor(Regiao)Sul	1.1618e-06	1.6150e-09	0.0000	0.9566043

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

Resposta:

O aumento de uma unidade no percentual de pessoas vivendo em domicílios com paredes inadequadas, aumenta a chance de um município ser classificado como de alta mortalidade infantil em 2,24% ($1,0224 - 1$), *ceteris paribus*.

Modelos de Regressão Logística

- Questão 1: Interprete os coeficientes da regressão que apresentem significância estatística;

```
> odds.ratio(mod5.reduzido)
waiting for profiling to be done...
```

	OR	2.5 %	97.5 %	p
(Intercept)	2.6954e+00	3.5765e-01	19.8171	0.3320800
renda_per_capita	9.8618e-01	9.8213e-01	0.9901	1.610e-11 ***
indice_gini	2.1191e+01	5.2598e-01	877.6071	0.1065654
salario_medio_mensal	9.3111e-01	7.8001e-01	1.1070	0.4185865
perc_crianças_extrem_pobres	9.5890e-01	9.3577e-01	0.9824	0.0007045 ***
perc_crianças_pobres	1.0214e+00	9.9191e-01	1.0518	0.1570583
perc_pessoas_dom_agua_estogo_inadequados	1.0010e+00	9.9271e-01	1.0094	0.8214049
perc_pessoas_dom_paredes_inadequadas	1.0224e+00	1.0109e+00	1.0343	0.0001470 ***
perc_pop_dom_com_coleta_lixo	9.9770e-01	9.8858e-01	1.0066	0.6156627
perc_pop_rural	6.8686e-01	3.5231e-01	1.3351	0.2688692
as.factor(Regiao)Nordeste	1.0837e+01	4.8904e+00	28.8175	8.538e-08 ***
as.factor(Regiao)Norte	2.4183e+00	1.0457e+00	6.6080	0.0563536 .
as.factor(Regiao)Sudeste	2.8510e-01	1.0580e-01	0.8532	0.0165362 *
as.factor(Regiao)Sul	1.1618e-06	1.6150e-09	0.0000	0.9566043

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

Resposta:

Municípios da região nordeste possui uma chance 9,837 (10,837-1) vezes maior de ser classificado como de alta mortalidade infantil do que municípios da região Centro-Oeste, *ceteris paribus*.

Modelos de Regressão Logística

- Questão 1: Interprete os coeficientes da regressão que apresentem significância estatística;

```
> odds.ratio(mod5.reduzido)
waiting for profiling to be done...
```

	OR	2.5 %	97.5 %	p
(Intercept)	2.6954e+00	3.5765e-01	19.8171	0.3320800
renda_per_capita	9.8618e-01	9.8213e-01	0.9901	1.610e-11 ***
indice_gini	2.1191e+01	5.2598e-01	877.6071	0.1065654
salario_medio_mensal	9.3111e-01	7.8001e-01	1.1070	0.4185865
perc_crianças_extrem_pobres	9.5890e-01	9.3577e-01	0.9824	0.0007045 ***
perc_crianças_pobres	1.0214e+00	9.9191e-01	1.0518	0.1570583
perc_pessoas_dom_agua_estogo_inadequados	1.0010e+00	9.9271e-01	1.0094	0.8214049
perc_pessoas_dom_paredes_inadequadas	1.0224e+00	1.0109e+00	1.0343	0.0001470 ***
perc_pop_dom_com_coleta_lixo	9.9770e-01	9.8858e-01	1.0066	0.6156627
perc_pop_rural	6.8686e-01	3.5231e-01	1.3351	0.2688692
as.factor(Regiao)Nordeste	1.0837e+01	4.8904e+00	28.8175	8.538e-08 ***
as.factor(Regiao)Norte	2.4183e+00	1.0457e+00	6.6080	0.0563536 .
as.factor(Regiao)Sudeste	2.8510e-01	1.0580e-01	0.8532	0.0165362 *
as.factor(Regiao)Sul	1.1618e-06	1.6150e-09	0.0000	0.9566043

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

Reposta:

Municípios da região sul possui uma chance 0,715 (1-0,285) vezes maior de ser classificado como de baixa mortalidade infantil (não alta mortalidade infantil) do que municípios da região Centro-Oeste, *ceteris paribus*.

Modelos de Regressão Logística

- 4ª Lista de exercícios para entregar em 03/12/2018 (continuação).
- Questão 2: Refaça a questão 1, considerando o modelo de regressão logística abaixo, interprete os odds-ratio dos coeficientes que apresentem significância estatística.

```
mod5.reduzido <- glm(formula = alta_mort_infantil ~ renda_per_capita  
+ indice_gini  
+ salario_medio_mensal  
+ perc_crianças_extrem_pobres  
+ perc_crianças_pobres  
+ perc_pessoas_dom_agua_estogo_inadequados  
+ perc_pessoas_dom_paredes_inadequadas  
+ perc_pop_dom_com_coleta_lixo  
+ perc_pop_rural  
+ as.factor(Regiao)  
+ as.factor(Regiao)*renda_per_capita,  
family = binomial(link = "logit"), data = dados3)  
summary(mod5.reduzido)
```

Questão 3: Com base nos critérios AIC e BIC qual desses modelos seriam selecionados?

Modelos de Regressão Logística

```
> odds.ratio(mod5.reduzido)
waiting for profiling to be done...
```

	OR	2.5 %	97.5 %	p
(Intercept)	9.1124e-02	2.1009e-03	3.7033	0.2069807
renda_per_capita	9.9188e-01	9.8421e-01	0.9986	0.0271426 *
indice_gini	3.0964e+02	6.4213e+00	15422.9430	0.0038579 **
salario_medio_mensal	9.4218e-01	7.8933e-01	1.1209	0.5003198
perc_crianças_extrem_pobres	9.4939e-01	9.2567e-01	0.9735	5.262e-05 ***
perc_crianças_pobres	1.0133e+00	9.8377e-01	1.0438	0.3806137
perc_pessoas_dom_agua_estogo_inadequados	1.0033e+00	9.9492e-01	1.0119	0.4396356
perc_pessoas_dom_paredes_inadequadas	1.0219e+00	1.0103e+00	1.0341	0.0002552 ***
perc_pop_dom_com_coleta lixo	9.9829e-01	9.8905e-01	1.0073	0.7141883
perc_pop_rural	5.4580e-01	2.7737e-01	1.0698	0.0785543 .
as.factor(Regiao)Nordeste	6.5088e+02	2.3354e+01	21994.5819	0.0001851 ***
as.factor(Regiao)Norte	1.1457e+01	4.1256e-01	383.9418	0.1587160
as.factor(Regiao)Sudeste	8.7418e+01	1.3491e+00	7297.8842	0.0406332 *
as.factor(Regiao)Sul	5.9506e-09	5.2144e-11	0.0000	0.9863875
renda_per_capita:as.factor(Regiao)Nordeste	9.8937e-01	9.8242e-01	0.9971	0.0043106 **
renda_per_capita:as.factor(Regiao)Norte	9.9781e-01	9.9083e-01	1.0056	0.5568390
renda_per_capita:as.factor(Regiao)Sudeste	9.8427e-01	9.7220e-01	0.9954	0.0077343 **
renda_per_capita:as.factor(Regiao)Sul	1.0077e+00	1.0021e+00	1.0130	0.9959732

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

Reposta:

O aumento de um unidade da renda percapita na região Nordeste, aumenta a possibilidade do município ser classificado como de baixa mortalidade infantil em 1,06% (1-0,9894) em relação à região Centro-Oeste, *ceteris paribus*.

O aumento de um unidade na região Sul, aumenta a possibilidade do município ser classificado como de alta mortalidade infantil em 0,77% (1,0077-1) em relação à região Centro-Oeste, *ceteris paribus*.

Modelos de Regressão Logística

```
as.factor(Regiao)Nordeste      2.383e+00  4.450e-01   5.355 8.54e-08 ***
as.factor(Regiao)Norte         8.831e-01  4.627e-01   1.908 0.056354 .
as.factor(Regiao)Sudeste      -1.255e+00  5.236e-01  -2.397 0.016536 *
as.factor(Regiao)Sul          -1.367e+01  2.511e+02  -0.054 0.956604
```

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

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 6163.7  on 5563  degrees of freedom
Residual deviance: 2522.6  on 5550  degrees of freedom
AIC: 2550.6
```

```
as.factor(Regiao)Sul          -1.894e+01  1.110e+03  -0.017 0.986387
renda_per_capita:as.factor(Regiao)Nordeste -1.069e-02  3.745e-03  -2.854 0.004311 **
renda_per_capita:as.factor(Regiao)Norte    -2.195e-03  3.737e-03  -0.588 0.556839
renda_per_capita:as.factor(Regiao)Sudeste  -1.586e-02  5.953e-03  -2.663 0.007734 **
renda_per_capita:as.factor(Regiao)Sul       7.632e-03  1.512e+00   0.005 0.995973
```

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

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 6163.7  on 5563  degrees of freedom
Residual deviance: 2490.5  on 5546  degrees of freedom
AIC: 2526.5
```

Questão 3: Com base nos critérios AIC e BIC qual desses modelos seriam selecionados?

Resposta:

O modelo com a interação renda per capita região seria selecionado, pois possui AIC 2526,5 contra AIC 2550,6 do modelo sem estas variáveis.

- **4ª Lista de exercícios para entregar em 03/12/2018 (continuação).**
- Considere o modelo completo abaixo. Usando os diversos métodos aprendidos em sala de aula, encontre um modelo, subconjunto do modelo abaixo, que apresente o menor AIC. No resultado entregue, você deverá incluir o código em R para obter o melhor modelo, e deverá incluir também a fórmula em R para essa “melhor” regressão

```
mod.full <- lm(mort_infantil ~ renda_per_capita
+ I(renda_per_capita^2)
+ I(renda_per_capita^3)
+ I(renda_per_capita^4)
+ I(renda_per_capita^5)
+ indice_gini
+ I(indice_gini^2)
+ I(indice_gini^3)
+ I(indice_gini^4)
+ I(indice_gini^5)
+ salario_medio_mensal
+ I(salario_medio_mensal^2)
+ I(salario_medio_mensal^3)
+ I(salario_medio_mensal^4)
+ I(salario_medio_mensal^5)
+ perc_crianças_extrem_pobres
+ perc_crianças_pobres
+ perc_pessoas_dom_agua_estogo_inadequados
+ perc_pessoas_dom_paredes_inadequadas
+ perc_pop_dom_com_coleta_lixo
+ perc_pop_rural
+ as.factor(Regiao)
+ as.factor(Regiao)*renda_per_capita, data = dados3)
```

Resposta: AIC: 29437,32

- **4ª Lista de exercícios para entregar em 03/12/2018 (continuação).**
- Considere o modelo completo abaixo. Usando os diversos métodos aprendidos em sala de aula, encontre um modelo, subconjunto do modelo abaixo, que apresente o menor AIC. No resultado entregue, você deverá incluir o código em R para obter o melhor modelo, e deverá incluir também a fórmula em R para essa “melhor” regressão

```
mod.full <- lm(mort_infantil ~ renda_per_capita
  + I(renda_per_capita^2)
  + I(renda_per_capita^3)
  + I(renda_per_capita^4)
  + I(renda_per_capita^5)
  + indice_gini
  + I(indice_gini^2)
  + I(indice_gini^3)
  + I(indice_gini^4)
  + I(indice_gini^5)
  + salario_medio_mensal
  + I(salario_medio_mensal^2)
  + I(salario_medio_mensal^3)
  + I(salario_medio_mensal^4)
  + I(salario_medio_mensal^5)
  + perc_crianças_extrem_pobres
  + perc_crianças_pobres
  + perc_pessoas_dom_agua_estogo_inadequados
  + perc_pessoas_dom_paredes_inadequadas
  + perc_pop_dom_com_coleta_lixo
  + perc_pop_rural
  + as.factor(Regiao)
  + as.factor(Regiao)*renda_per_capita, data = dados3)
```

Modelo selecionado

```
> formula(step3)
mort_infantil ~ renda_per_capita + I(renda_per_capita^3) + I(renda_per_capita^4) +
  I(renda_per_capita^5) + I(indice_gini^2) + I(indice_gini^3) +
  I(indice_gini^4) + I(salario_medio_mensal^3) + I(salario_medio_mensal^4) +
  perc_crianças_extrem_pobres + perc_crianças_pobres + perc_pessoas_dom_agua_estogo_inadequados +
  perc_pessoas_dom_paredes_inadequadas + perc_pop_dom_com_coleta_lixo +
  perc_pop_rural + as.factor(Regiao) + renda_per_capita:as.factor(Regiao)
```

Obrigado!