

ANÁLISE DE DADOS MULTIVARIADOS I - REGRESSÃO

(AULA 05)

Novembro e dezembro de 2018

Reinaldo Soares de Camargo

Modelos de Regressão

- **2ª Lista de exercícios para entregar em 19/11/2018.**

- 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 de acordo com o modelo abaixo:

```
mod2.ex <- lm(dados3$mort_infantil ~ dados3$renda_per_capita  
+ dados3$indice_gini  
+ dados3$salario_medio_mensal  
+ dados3$perc_crianças_extrem_pobres  
+ dados3$perc_crianças_pobres  
+ dados3$perc_pessoas_dom_agua_estogo_inadequados  
+ dados3$perc_pessoas_dom_paredes_inadequadas  
+ dados3$perc_pop_dom_com_coleta_lixo  
+ dados3$perc_pop_rural  
+ as.factor(dados3$Regiao))
```

- Questão 7: Com base nos resultados dessa nova equação, qual o efeito das regiões Norte, Sul, Nordeste e Sudeste, mesmo depois de “controlarmos” para as variáveis incluídas no modelo?

Modelos de Regressão

- Questão 7: Com base nos resultados dessa nova equação, qual o efeito das regiões Norte, Sul, Nordeste e Sudeste, mesmo depois de “controlarmos” para as variáveis incluídas no modelo?

```
call:
lm(formula = dados3$mort_infantil ~ dados3$renda_per_capita +
  dados3$indice_gini + dados3$salario_medio_mensal + dados3$perc_crianças_extrem_pobres +
  dados3$perc_crianças_pobres + dados3$perc_pessoas_dom_agua_estogo_inadequados +
  dados3$perc_pessoas_dom_paredes_inadequadas + dados3$perc_pop_dom_com_coleta lixo +
  dados3$perc_pop_rural + as.factor(dados3$Regiao))

Residuals:
    Min       1Q   Median       3Q      Max
-11.3942  -1.9967  -0.2933   1.6103  18.4841

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.103e+01  7.517e-01  27.979  < 2e-16 ***
dados3$renda_per_capita
-1.990e-03    5.326e-04   -3.737  0.000188 ***
dados3$indice_gini
-1.016e+01    1.194e+00   -8.510  < 2e-16 ***
dados3$salario_medio_mensal
-1.016e-01    8.282e-02   -1.227  0.219889
dados3$perc_crianças_extrem_pobres
2.265e-02    1.075e-02    2.107  0.035138 *
dados3$perc_crianças_pobres
1.026e-01    1.071e-02    9.579  < 2e-16 ***
dados3$perc_pessoas_dom_agua_estogo_inadequados
4.505e-02    5.605e-03    8.037  1.12e-15 ***
dados3$perc_pessoas_dom_paredes_inadequadas
5.500e-02    6.896e-03    7.976  1.83e-15 ***
dados3$perc_pop_dom_com_coleta_lixo
-1.490e-02    5.681e-03   -2.622  0.008762 **
dados3$perc_pop_rural
-5.808e-01    2.912e-01   -1.995  0.046114 *
as.factor(dados3$Regiao)Nordeste
5.928e+00    2.356e-01   25.156  < 2e-16 ***
as.factor(dados3$Regiao)Norte
1.249e+00    2.649e-01    4.716  2.46e-06 ***
as.factor(dados3$Regiao)Sudeste
-3.185e-01    1.899e-01   -1.677  0.093581 .
as.factor(dados3$Regiao)Sul
-1.948e+00    2.062e-01   -9.446  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.494 on 5550 degrees of freedom
Multiple R-squared:  0.7609,    Adjusted R-squared:  0.7604
F-statistic: 1359 on 13 and 5550 DF, p-value: < 2.2e-16
```

Resposta: Mantendo constantes os demais fatores, a região Nordeste possui em média 5,928 mortos para cada 1000 nascidos vivos a mais do que a região Centro-Oeste; a região Norte possui em média 1,249 mortos para cada 1000 nascidos vivo a mais que a região Centro-Oeste; a região Sudeste possui em média 0,3185 mortos para cada 1000 nascidos vivo menos que a região Centro-Oeste; a região Sul possui em média 1,948 mortos para cada 1000 nascidos vivo menos que a região Centro-Oeste;

Modelos de Regressão

- **2ª Lista de exercícios para entregar em 19/11/2018(continuação):**
 - Rode agora a regressão com efeitos das Regiões sobre a mortalidade infantil:

```
mod3.ex <- lm(dados3$mort_infantil ~ dados3$renda_per_capita
+ dados3$indice_gini
+ dados3$salario_medio_mensal
+ dados3$perc_crianças_extrem_pobres
+ dados3$perc_crianças_pobres
+ dados3$perc_pessoas_dom_agua_estogo_inadequados
+ dados3$perc_pessoas_dom_paredes_inadequadas
+ dados3$perc_pop_dom_com_coleta_lixo
+ dados3$perc_pop_rural
+ as.factor(dados3$Regiao)
+ as.factor(dados3$Regiao)*dados3$renda_per_capita)
```

- Questão 8: Com base nos resultados dessa nova equação, como o efeito da renda per capita, sobre mortalidade infantil, se altera de acordo com a macrorregião do município?

Modelos de Regressão

- 2ª Lista de exercícios para entregar em 19/11/2018(continuação):
 - Questão 8: Com base nos resultados dessa nova equação, como o efeito da renda per capita, sobre mortalidade infantil, se altera de acordo com a macrorregião do município?

```
Residuals:
    Min       1Q   Median       3Q      Max
-10.8311  -1.9372  -0.2724   1.5532  17.6192

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.653e+01  9.316e-01  17.741  < 2e-16 ***
dados3$renda_per_capita 5.669e-04  1.110e-03   0.511  0.60968
dados3$indice_gini    -3.239e+00  1.299e+00  -2.493  0.01270 *
dados3$salario_medio_mensal 8.966e-05  8.239e-02   0.001  0.99913
dados3$perc_crianças_extrem_pobres -3.272e-02  1.141e-02  -2.866  0.00417 **
dados3$perc_crianças_pobres 8.827e-02  1.077e-02   8.200  2.96e-16 ***
dados3$perc_pessoas_dom_agua_estogo_inadequados 3.581e-02  5.641e-03   6.348  2.36e-10 ***
dados3$perc_pessoas_dom_paredes_inadequadas 4.559e-02  6.818e-03   6.687  2.51e-11 ***
dados3$perc_pop_dom_com_coleta lixo -1.306e-02  5.583e-03  -2.340  0.01934 *
dados3$perc_pop_rural -6.161e-01  2.862e-01  -2.152  0.03141 *
as.factor(dados3$Regiao)Nordeste 1.395e+01  7.037e-01  19.829  < 2e-16 ***
as.factor(dados3$Regiao)Norte 6.285e+00  7.835e-01   8.022  1.26e-15 ***
as.factor(dados3$Regiao)Sudeste 2.561e+00  6.520e-01   3.928  8.67e-05 ***
as.factor(dados3$Regiao)Sul 2.552e-02  7.079e-01   0.036  0.97125
dados3$renda_per_capita:as.factor(dados3$Regiao)Nordeste -1.937e-02  1.401e-03 -13.828  < 2e-16 ***
dados3$renda_per_capita:as.factor(dados3$Regiao)Norte -9.111e-03  1.578e-03  -5.772  8.23e-09 ***
dados3$renda_per_capita:as.factor(dados3$Regiao)Sudeste -4.733e-03  1.062e-03  -4.458  8.43e-06 ***
dados3$renda_per_capita:as.factor(dados3$Regiao)Sul -3.271e-03  1.120e-03  -2.919  0.00352 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.426 on 5546 degrees of freedom
Multiple R-squared:  0.7703,    Adjusted R-squared:  0.7696
F-statistic: 1094 on 17 and 5546 DF,  p-value: < 2.2e-16
```

Resposta: Mantendo constantes os demais fatores, na região Nordeste o aumento de uma unidade na renda per capita reduz a mortalidade infantil em 0,01937 mortos para cada 1000 nascido vivos; na região Norte o aumento de uma unidade na renda per capita reduz a mortalidade infantil em 0,00911 mortos para cada 1000 nascido vivos; na região Sudeste o aumento de uma unidade na renda per capita reduz a mortalidade infantil em 0,00911 mortos para cada 1000 nascido vivos; na região Sul o aumento de uma unidade na renda per capita reduz a mortalidade infantil em 0,003271 mortos para cada 1000 nascido vivos.

Modelos de Regressão

- 2ª Lista de exercícios para entregar em 19/11/2018(continuação):
 - Questão 9: Houve uma melhora no R^2 ajustado quando adicionamos os efeitos das macrorregiões sobre o coeficiente da renda per capita (mod3 versus mod2)?

Residual standard error: 3.509 on 5554 degrees of freedom
Multiple R-squared: 0.7587, Adjusted R-squared: 0.7583
F-statistic: 1941 on 9 and 5554 DF, p-value: < 2.2e-16

Residual standard error: 3.426 on 5546 degrees of freedom
Multiple R-squared: 0.7703, Adjusted R-squared: 0.7696
F-statistic: 1094 on 17 and 5546 DF, p-value: < 2.2e-16

Resposta: o R^2 passou de 75,83% para 76,96%, ou seja, o modelo quando adicionamos os efeitos das macrorregiões sobre o coeficiente da renda per capita parece mais adequado para explicar mortalidade infantil.

Modelos de Regressão

- **2ª Lista de exercícios para entregar em 19/11/2018 (Continuação).**
 - Vamos incluir agora uma interação entre a macrorregião e a variável “perc_pessoas_dom_agua_estogo_inadequados”:

```
mod3.ex <- lm(dados3$mort_infantil ~ dados3$renda_per_capita
+ dados3$indice_gini
+ dados3$salario_medio_mensal
+ dados3$perc_crianças_extrem_pobres
+ dados3$perc_crianças_pobres
+ dados3$perc_pessoas_dom_agua_estogo_inadequados
+ dados3$perc_pessoas_dom_paredes_inadequadas
+ dados3$perc_pop_dom_com_coleta_lixo
+ dados3$perc_pop_rural
+ as.factor(dados3$Regiao)
+ as.factor(dados3$Regiao)*dados3$renda_per_capita
+ as.factor(dados3$Regiao)*dados3$perc_pessoas_dom_agua_estogo_inadequados)
```

Modelos de Regressão

- 2ª Lista de exercícios para entregar em 19/11/2018 (Continuação).
- Questão 10: Vamos assumir que a variável “perc_pessoas_dom_agua_estogo_inadequados” seja uma variável direta de política pública. De acordo com os resultados da regressão acima, em qual região políticas de melhoria do acesso a água e esgoto seriam mais eficazes para reduzir a mortalidade infantil?

```
Residuals:
    Min       1Q   Median       3Q      Max
-11.2091  -1.9438  -0.2821   1.5426   17.7637

Coefficients:
(Intercept)                                16.0274766    0.9838203    16.291    < 2e-16 ***
dados3$renda_per_capita                    0.0009083    0.0011885     0.764    0.444740
dados3$indice_gini                         -2.9977815    1.3090116    -2.290    0.022052 *
dados3$salario_medio_mensal                0.0147787    0.0825317     0.179    0.857892
dados3$perc_crianças_extrem_pobres         -0.0341171    0.0114397    -2.982    0.002873 **
dados3$perc_crianças_pobres                0.0889045    0.0110488     8.047    1.03e-15 ***
dados3$perc_pessoas_dom_agua_estogo_inadequados 0.0644592    0.0305935     2.107    0.035166 *
dados3$perc_pessoas_dom_paredes_inadequadas 0.0436858    0.0068381     6.389    1.81e-10 ***
dados3$perc_pop_dom_com_coleta_lixo        -0.0129498    0.0055889    -2.317    0.020536 *
dados3$perc_pop_rural                     -0.5909319    0.2872558    -2.057    0.039717 *
as.factor(dados3$Regiao)Nordeste          14.0063587    0.8435092    16.605    < 2e-16 ***
as.factor(dados3$Regiao)Norte              8.0960669    1.0230192     7.914    2.99e-15 ***
as.factor(dados3$Regiao)Sudeste            2.8233097    0.7723477     3.655    0.000259 ***
as.factor(dados3$Regiao)Sul                0.7150179    0.8216248     0.870    0.384202
dados3$renda_per_capita:as.factor(dados3$Regiao)Nordeste -0.0193202    0.0015049   -12.838    < 2e-16 ***
dados3$renda_per_capita:as.factor(dados3$Regiao)Norte -0.0115962    0.0017951    -6.460    1.14e-10 ***
dados3$renda_per_capita:as.factor(dados3$Regiao)Sudeste -0.0049989    0.0011548    -4.329    1.53e-05 ***
dados3$renda_per_capita:as.factor(dados3$Regiao)Sul -0.0039455    0.0012045    -3.276    0.001061 **
dados3$perc_pessoas_dom_agua_estogo_inadequados:as.factor(dados3$Regiao)Nordeste -0.0176081    0.0312381    -0.564    0.572998
dados3$perc_pessoas_dom_agua_estogo_inadequados:as.factor(dados3$Regiao)Norte -0.0566676    0.0322490    -1.757    0.078941 .
dados3$perc_pessoas_dom_agua_estogo_inadequados:as.factor(dados3$Regiao)Sudeste -0.0158194    0.0384886    -0.411    0.681077
dados3$perc_pessoas_dom_agua_estogo_inadequados:as.factor(dados3$Regiao)Sul -0.1053997    0.0479871    -2.196    0.028103 *
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.422 on 5542 degrees of freedom
Multiple R-squared:  0.7709,    Adjusted R-squared:  0.7701
F-statistic: 888.2 on 21 and 5542 DF,  p-value: < 2.2e-16
```


Obrigado.