Lista 3 estatística - Regressão

Diego dos Santos Fernandes e Humberto Nunes

21 de junho de 2019

# Exercício regressão questão 1 a 6

obeservação: os códigos estão abaixo das respostas

## Questão 1: No modelo anterior, quais as variáveis explicativas e qual a variável dependente?

Essa regressão incial tenta revelar alguma relaçao entre mortalidade infantil e alguma caraterísticas socieconômicas.

Neste caso a variável dependente é a mortalidade infantil e as variáveis explicativas são: renda per capita; índice de gini; salário médio mensal; percentual de crianças em extrema probreza, percentual de crianças pobres; ércentual de domicílio com agua e esgoto inadequados; percentual de domicílios com paredes inadequadas e por fim, o percentual de domicílios com coleta de lixo

## Questão 2: Os coeficientes encontrados estão com os sinais de acordo com o esperado?

Todos os coeficientes apresentaram os sinais de acordo com o esperado. Talvez o que possa ser mencionado é a sgnificancia de cada um. Destaca-se o nível de significancia de renda per capita de 5% e de salário médio de 10%. Os outros apresentam siginificancia abaixo de 1%

## Questão 3: Qual o percentual da variabilidade da mortalidade infantil que é explicada pelas variáveis explicativas?

O R- quadrado ajustado da regressão apresenta o valor de 0.6799, ou seja, cerca de 68% da variabilidade da mortalidade infantil é explicada pelas variáveis explicativas neste modelo

## Questão 4: Utilizando o comando abaixo, crie a variável ‘perc\_pop\_rural’, indicando o percentual do município que vive em domicílios na zona rural. Adicione essa variável ao modelo de regressão. Com base no coeficiente estimado, “controlando-se” para as variáveis já presentes no modelo, qual o efeito da localização na zona rural sobre a taxa de mortalidade infantil?

dados3//perc\_pop\_rural <- dados3//populacao\_rural / dados3$populacao\_total

## Questão 5: Com a inclusão da nova variável, o que aconteceu com o coeficiente de determinação e com o R2 ajustado?

O R quadrado ajustado aumentou para 0.6936, ou seja, o percentual de variabilidade de mortalidade infantil que é explicada pelas variáveis explicativas aumentou para quase 70% mesmo com a penalização que a formulado Rquarado ajustado carrega. ##Questão 6: Os dados utilizados para essa regressão são dados do tipo cross-section, do tipo séries de tempo ou do tipo dados de painel?

Esses dados são do tipo cross-section, uma vez que são uma “fotografia” de um instante específico no tempo

options(scipen = 10)  
mod1 <- lm(dados3$mort\_infantil ~ dados3$renda\_per\_capita   
 + dados3$indice\_gini  
 + dados3$salario\_medio\_mensal  
 + dados3$perc\_criancas\_extrem\_pobres  
 + dados3$perc\_criancas\_pobres  
 + dados3$perc\_pessoas\_dom\_agua\_estogo\_inadequados  
 + dados3$perc\_pessoas\_dom\_paredes\_inadequadas  
 + dados3$perc\_pop\_dom\_com\_coleta\_lixo)  
summary(mod1)

##   
## Call:  
## lm(formula = dados3$mort\_infantil ~ dados3$renda\_per\_capita +   
## dados3$indice\_gini + dados3$salario\_medio\_mensal + dados3$perc\_criancas\_extrem\_pobres +   
## dados3$perc\_criancas\_pobres + dados3$perc\_pessoas\_dom\_agua\_estogo\_inadequados +   
## dados3$perc\_pessoas\_dom\_paredes\_inadequadas + dados3$perc\_pop\_dom\_com\_coleta\_lixo)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.5530 -2.4952 -0.3666 1.9344 20.8067   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) 19.3648493 0.8196173  
## dados3$renda\_per\_capita -0.0012777 0.0005784  
## dados3$indice\_gini -14.2977241 1.2465495  
## dados3$salario\_medio\_mensal -0.1775215 0.0951454  
## dados3$perc\_criancas\_extrem\_pobres 0.0385401 0.0121614  
## dados3$perc\_criancas\_pobres 0.2158968 0.0114768  
## dados3$perc\_pessoas\_dom\_agua\_estogo\_inadequados 0.0505541 0.0060208  
## dados3$perc\_pessoas\_dom\_paredes\_inadequadas 0.0429694 0.0079240  
## dados3$perc\_pop\_dom\_com\_coleta\_lixo -0.0070450 0.0065204  
## t value Pr(>|t|)   
## (Intercept) 23.627 < 2e-16 \*\*\*  
## dados3$renda\_per\_capita -2.209 0.02721 \*   
## dados3$indice\_gini -11.470 < 2e-16 \*\*\*  
## dados3$salario\_medio\_mensal -1.866 0.06212 .   
## dados3$perc\_criancas\_extrem\_pobres 3.169 0.00154 \*\*   
## dados3$perc\_criancas\_pobres 18.812 < 2e-16 \*\*\*  
## dados3$perc\_pessoas\_dom\_agua\_estogo\_inadequados 8.397 < 2e-16 \*\*\*  
## dados3$perc\_pessoas\_dom\_paredes\_inadequadas 5.423 0.0000000612 \*\*\*  
## dados3$perc\_pop\_dom\_com\_coleta\_lixo -1.080 0.27999   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.038 on 5555 degrees of freedom  
## Multiple R-squared: 0.6804, Adjusted R-squared: 0.6799   
## F-statistic: 1478 on 8 and 5555 DF, p-value: < 2.2e-16

# Inclusão de Nova variável  
  
  
dados3$perc\_pop\_rural <- dados3$populacao\_rural / dados3$populacao\_total  
  
  
mod1.1 <- mod1 <- lm(dados3$mort\_infantil ~ dados3$renda\_per\_capita   
 + dados3$indice\_gini  
 + dados3$salario\_medio\_mensal  
 + dados3$perc\_criancas\_extrem\_pobres  
 + dados3$perc\_criancas\_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)  
  
  
  
summary(mod1.1)

##   
## Call:  
## lm(formula = dados3$mort\_infantil ~ dados3$renda\_per\_capita +   
## dados3$indice\_gini + dados3$salario\_medio\_mensal + dados3$perc\_criancas\_extrem\_pobres +   
## dados3$perc\_criancas\_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)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.7386 -2.4369 -0.3806 1.8390 19.1262   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) 20.9745918 0.8084047  
## dados3$renda\_per\_capita -0.0010319 0.0005661  
## dados3$indice\_gini -17.2296204 1.2337441  
## dados3$salario\_medio\_mensal -0.2391021 0.0931745  
## dados3$perc\_criancas\_extrem\_pobres 0.0665409 0.0120308  
## dados3$perc\_criancas\_pobres 0.2278958 0.0112549  
## dados3$perc\_pessoas\_dom\_agua\_estogo\_inadequados 0.0551911 0.0058982  
## dados3$perc\_pessoas\_dom\_paredes\_inadequadas 0.0459096 0.0077553  
## dados3$perc\_pop\_dom\_com\_coleta\_lixo -0.0006609 0.0063926  
## dados3$perc\_pop\_rural -4.6229326 0.2931518  
## t value Pr(>|t|)   
## (Intercept) 25.946 < 2e-16 \*\*\*  
## dados3$renda\_per\_capita -1.823 0.0684 .   
## dados3$indice\_gini -13.965 < 2e-16 \*\*\*  
## dados3$salario\_medio\_mensal -2.566 0.0103 \*   
## dados3$perc\_criancas\_extrem\_pobres 5.531 0.00000003332 \*\*\*  
## dados3$perc\_criancas\_pobres 20.249 < 2e-16 \*\*\*  
## dados3$perc\_pessoas\_dom\_agua\_estogo\_inadequados 9.357 < 2e-16 \*\*\*  
## dados3$perc\_pessoas\_dom\_paredes\_inadequadas 5.920 0.00000000342 \*\*\*  
## dados3$perc\_pop\_dom\_com\_coleta\_lixo -0.103 0.9177   
## dados3$perc\_pop\_rural -15.770 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.951 on 5554 degrees of freedom  
## Multiple R-squared: 0.6941, Adjusted R-squared: 0.6936   
## F-statistic: 1400 on 9 and 5554 DF, p-value: < 2.2e-16