

kek

UFPB - Regressão I

Paulo Ricardo Seganfredo Campana

Gabriel de Jesus Pereira

25 de outubro de 2023

Resumo

ESCREVER MERDA

Introdução

O concreto é um dos materiais de construção mais utilizados na engenharia civil devido à sua durabilidade, versatilidade e resistência. Ele é composto por agregados, água e principalmente cimento. Analisando-se todo o cenário e necessidades da engenharia civil, foi-se proposto a modelagem da força compressiva do concreto. Para isso, utilizamos um banco de dados que contém o tempo que o concreto foi deixado para secar e as substâncias que o compõem.

Fazendo uso desses dados, da modelagem e análise estatística, foi possível chegar a dois modelos de regressão linear, um mais simples e outro mais complexo, com o intuito de poder saber o que leva um concreto a ser mais resistente do que outro e para previsão da força compressiva de certo concreto baseado nas variáveis de estudo.

Metodologia

As análises a seguir foram realizadas usando a linguagem de programação R (R Core Team 2023) com o *framework* de modelagem estatística *tidymodels* (Kuhn e Wickham 2020). Os códigos utilizados estão disponíveis no github (Campana e J. Pereira 2023) e os documentos do relatório e apresentação foram feitos com Quarto (Allaire et al. 2022), um sistema de escrita e publicação científica.

Utilizamos o conjunto de dados sobre concreto de alta performance (Yeh 2006) onde foi estudado um modelo de regressão linear múltipla em que a `força_compressiva` do concreto é

explicada a princípio pelas variáveis que achamos importantes no estudo: o tempo de secagem da mistura final (em dias) e os matérias que compõem a mistura: cimento, escória_de_aço, cinzas_pulverizadas, água, superplastificante, agregado_graúdo e agregado_miúdo (em quilogramas por metro cúbico).

Sendo assim, ajustamos dois modelos, um primeiro mais simples utilizando apenas 4 dos regressores e algumas transformações com o objetivo de estabelecer uma relação compreensível das substâncias que mais interferem na força compressiva do concreto. O segundo modelo é mais complexo, trazendo a interação entre as variáveis e composição das mesmas em novas medidas, este foca no poder preditivo da regressão.

Como se tratam de dados experimentais sobre o concreto, esse conjunto inclui várias combinações de valores diferentes entre os regressores, desse modo, a correlação entre as variáveis é baixa, atingindo no máximo 50%.

Fizemos uso da transformação Yeo-Johnson (Yeo e Johnson 2000), que de maneira similar a Box-Cox, é uma transformação feita para tornar a distribuição dos regressores mais normais e estabilizar a variância, porém também funciona para dados que contêm valores 0 e números negativos. O parâmetro λ é estimado por máxima verossimilhança.

$$\psi(\lambda, y) = \begin{cases} ((1+y)^\lambda - 1)/\lambda & \lambda \neq 0, y \geq 0 \\ \ln(1+y) & \lambda = 0, y \geq 0 \\ ((1-y)^{2-\lambda} - 1)/(\lambda - 2) & \lambda \neq 2, y < 0 \\ -\ln(1-y) & \lambda = 2, y < 0 \end{cases}$$

Resultados

```
oper 1 step rm [training]
oper 2 step YeoJohnson [training]
The retained training set is ~ 0.03 Mb in memory.
```

```
# A tibble: 4 x 3
  terms      value id
  <chr>      <dbl> <chr>
1 cimento    0.197 YeoJohnson_sbvK0
2 escória_de_aço 0.0660 YeoJohnson_sbvK0
3 água       0.960 YeoJohnson_sbvK0
4 tempo     -0.00633 YeoJohnson_sbvK0
```

Call:

```
stats::lm(formula = força_compressiva ~ . - 1, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-22.9938	-5.3658	-0.1437	5.2579	24.8154

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
cimento	7.98320	0.16112	49.547	< 2e-16 ***
escória_de_aço	2.07304	0.09500	21.821	< 2e-16 ***
água	-0.33674	0.01096	-30.713	< 2e-16 ***
tempo_poly_1	283.48975	7.62659	37.171	< 2e-16 ***
tempo_poly_2	-55.04679	7.60523	-7.238	1.11e-12 ***

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

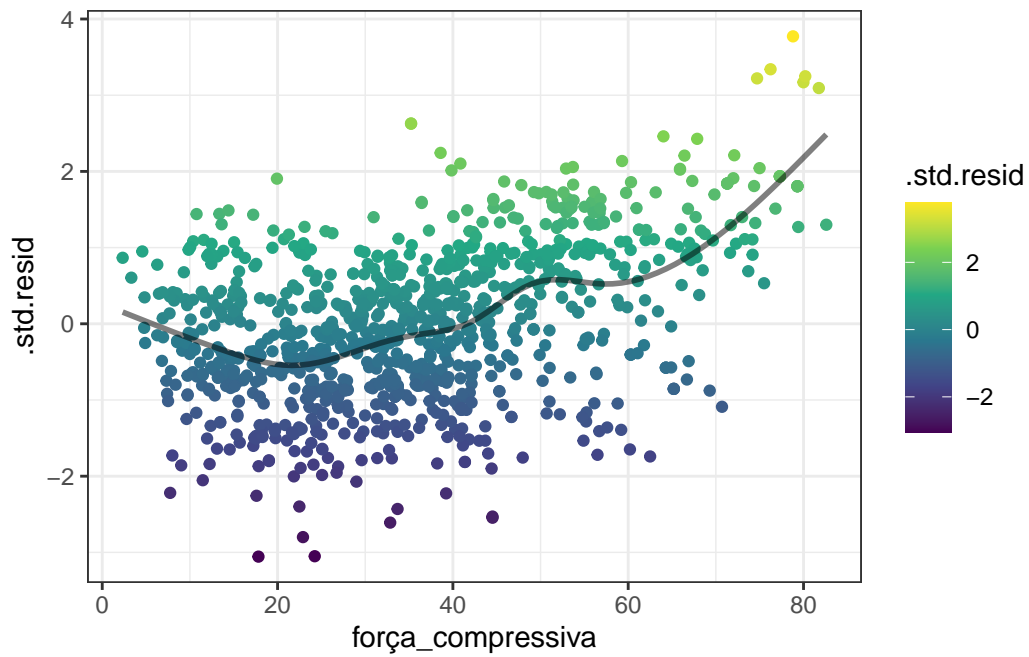
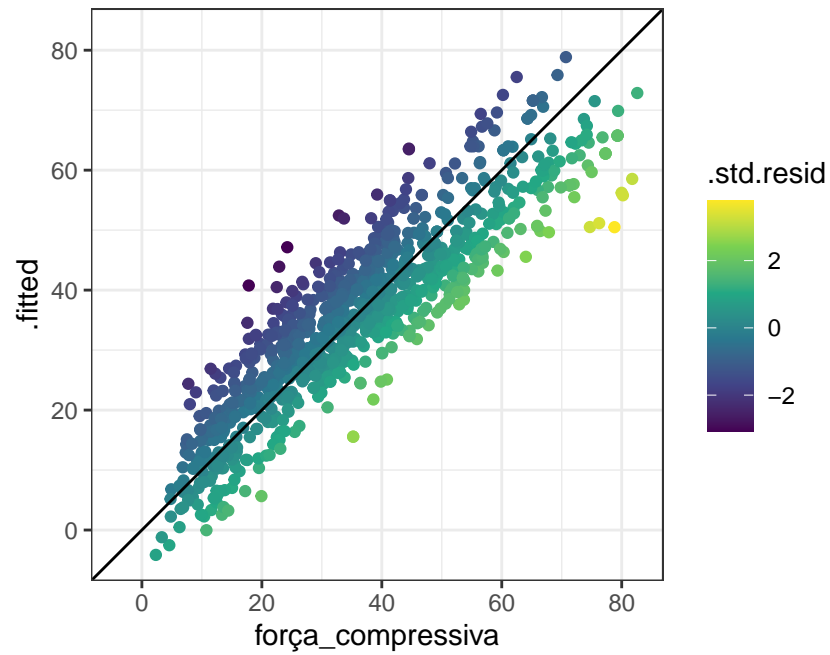
Residual standard error: 7.578 on 767 degrees of freedom

Multiple R-squared: 0.964, Adjusted R-squared: 0.9638

F-statistic: 4112 on 5 and 767 DF, p-value: < 2.2e-16

A tibble: 2 x 4

	.metric	.estimator	.estimate	.config
	<chr>	<chr>	<dbl>	<chr>
1	rmse	standard	7.33	Preprocessor1_Model1
2	rsq	standard	0.790	Preprocessor1_Model1



Call:
stats::lm(formula = força_compressiva ~ . - 1, data = data)

Residuals:

Min	1Q	Median	3Q	Max
-21.9914	-3.6080	-0.0712	3.3141	21.3136

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
cinzas_pulverizadas	3.234e+00	2.021e-01	16.005	< 2e-16	***
agregado_graúdo	6.486e-02	7.359e-04	88.134	< 2e-16	***
água_poly_1	-6.407e+01	8.040e+00	-7.969	5.91e-15	***
água_poly_2	-5.260e+02	1.033e+02	-5.092	4.49e-07	***
água_poly_3	-4.352e+02	8.628e+01	-5.044	5.71e-07	***
cimento_poly_1	3.832e+02	8.848e+00	43.312	< 2e-16	***
cimento_poly_2	7.179e+01	7.353e+00	9.763	< 2e-16	***
escória_de_aço_poly_1	2.565e+02	7.677e+00	33.405	< 2e-16	***
escória_de_aço_poly_2	1.078e+02	7.893e+00	13.660	< 2e-16	***
agregado_miúdo_poly_1	6.257e+01	8.125e+00	7.701	4.26e-14	***
agregado_miúdo_poly_2	-4.185e+01	7.060e+00	-5.928	4.66e-09	***
tempo_poly_1	2.875e+02	6.273e+00	45.832	< 2e-16	***
tempo_poly_2	-2.174e+01	7.068e+00	-3.076	0.002173	**
cinzas_pulverizadas_x_cimento_poly_2	-1.488e+01	4.393e+00	-3.386	0.000745	***
agregado_graúdo_x_água_poly_2	1.189e+00	2.143e-01	5.547	4.02e-08	***
agregado_graúdo_x_água_poly_3	1.039e+00	1.775e-01	5.854	7.17e-09	***
água_poly_1_x_tempo_poly_1	-7.499e+02	1.767e+02	-4.243	2.48e-05	***
agregado_miúdo_poly_2_x_tempo_poly_1	-8.561e+02	1.646e+02	-5.202	2.55e-07	***
tempo_poly_1_x_escória_de_aço_poly_1	1.040e+03	1.697e+02	6.130	1.42e-09	***
tempo_poly_2_x_cimento_poly_1	-7.673e+02	1.896e+02	-4.047	5.73e-05	***

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

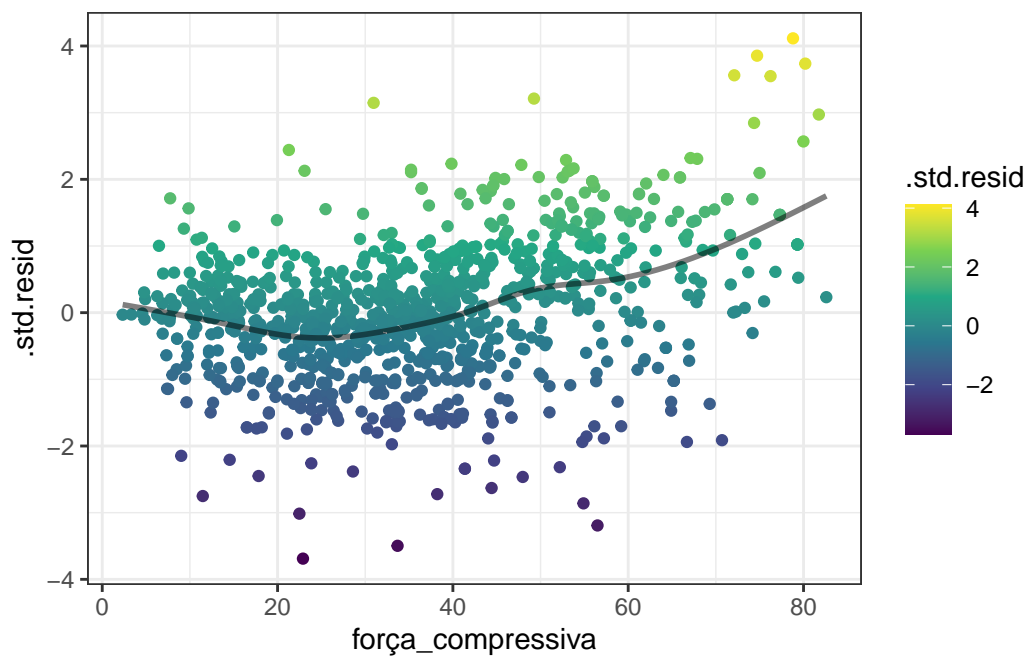
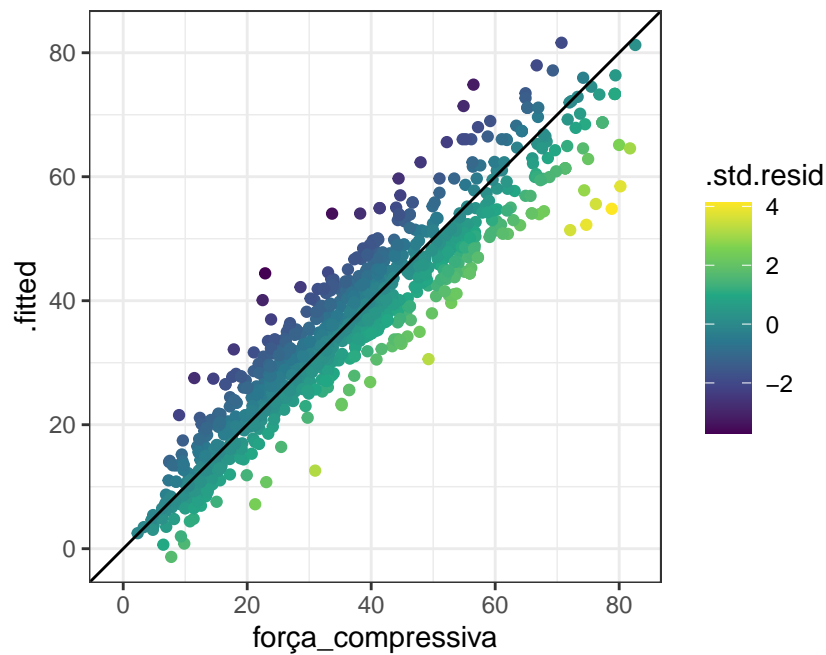
Residual standard error: 5.875 on 752 degrees of freedom

Multiple R-squared: 0.9788, Adjusted R-squared: 0.9782

F-statistic: 1737 on 20 and 752 DF, p-value: < 2.2e-16

A tibble: 2 x 4

	.metric	.estimator	.estimate	.config
	<chr>	<chr>	<dbl>	<chr>
1	rmse	standard	5.71	Preprocessor1_Model1
2	rsq	standard	0.873	Preprocessor1_Model1



Allaire, J. J., Charles Teague, Carlos Scheidegger, Yihui Xie, e Christophe Dervieux. 2022. «Quarto». 2022. <https://quarto.org>.

Campana, Paulo R. S., e Gabriel de J. Pereira. 2023. «Códigos dos modelos de regressão e análise». 2023. <https://github.com/cowvin0/conkrekt>.

Kuhn, Max, e Hadley Wickham. 2020. *Tidymodels: a collection of packages for modeling and*

- machine learning using tidyverse principles*. <https://www.tidymodels.org>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing* (versão 4.3.1). Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Yeh, I-Cheng. 2006. «Analysis of Strength of Concrete Using Design of Experiments and Neural Networks». *Journal of Materials in Civil Engineering* 18 (4): 597–604. [https://doi.org/10.1061/\(ASCE\)0899-1561\(2006\)18:4\(597\)](https://doi.org/10.1061/(ASCE)0899-1561(2006)18:4(597)).
- Yeo, In-Kwon, e Richard A. Johnson. 2000. «A New Family of Power Transformations to Improve Normality or Symmetry». *Biometrika* 87 (4): 954–59. <http://www.jstor.org/stable/2673623>.