

Análise de Dados
CCP 1069
60 horas

DS4PS: Ciência de Dados para Ciência Política
Pós-Graduação

Professor: Davi Moreira

Monitores: Bhreno Vieira

Ementa

O curso aborda técnicas exploratórias e inferenciais para análise de dados a partir de linguagens de programação com ênfase sobre modelos de regressão. Enquanto a análise exploratória é necessária para o desenvolvimento de hipóteses que serão investigadas através de dados coletados, a inferência estatística é o processo pelo qual são tiradas conclusões e realizadas previsões a partir da modelagem de sua incerteza. Com bases de dados reais, serão abordadas técnicas estatísticas multivariadas e modelos de regressão linear usuais em ciência política. Serão conhecidas as melhores práticas para apresentação de resultados com foco sobre o desenvolvimento de relatórios automatizados.

Objetivos

Apresentar o fluxo de trabalho do cientista de dados a partir de exemplos práticos aplicados à ciência política. Ao final, o aluno será capaz de criar projetos e apresentar resultados por meio de relatórios automatizados.

Pré-requisito: Não há.

Dinâmica das aulas

As aulas virtuais ou presenciais irão alternar exposição e aplicação. Elas terão intervalos curtos a depender da dinâmica de cada uma. O tempo inicial da primeira aula será utilizado para apresentação do conteúdo desse programa. A partir da segunda aula, o tempo inicial poderá ser usado para dúvidas. Será exigida postura ativa dos alunos no desenvolvimento das tarefas.

Link da aula: <https://meet.google.com/wtw-dfhd-yyx>

Link da turma no classroom: <https://classroom.google.com/c/NDU1NzlyMjMyNDY0?cjc=k7gspid>

Software e linguagem

A linguagem R e o RStudio serão utilizados no curso¹.

Sobre o professor:

Davi Moreira: Cientista Político. Vencedor do prêmio CAPES de melhor tese de doutorado em Ciência Política e Relações Internacionais em 2017. Suas pesquisas focam nas seguintes áreas: Metodologia em Ciência Política, Ciência de Dados, *Text as Data* e Política Comparada.

Webpage: www.davimoreira.com

E-mail: davi.moreira@ufpe.br

¹ Para detalhes, ver: <https://cran.r-project.org/> e <https://www.rstudio.com/>. Acessado em 15 de setembro de 2021.

Sobre os monitores:

Bhreno Vieira: Doutorando em Ciência Política na Universidade Federal de Pernambuco (UFPE). Sua pesquisa se concentra amplamente em estudos legislativos, com o objetivo particular de entender (i) como são formuladas as políticas públicas e (ii) as relações entre o Executivo e o Legislativo em regimes presidenciais multipartidários. Também é bastante interessado em *Text as Data*.

E-mail: bhreno.vieira@ufpe.br

Avaliação

Tarefas: No decorrer do curso serão solicitadas tarefas específicas, de mesmo peso. Cada tarefa deverá ser entregue em até uma semana. Todas as tarefas serão avaliadas de acordo com o atendimento das instruções e sua replicabilidade em três escalas: insatisfatória, satisfatória e plenamente satisfatória, valendo respectivamente 0, 7 e 10 pontos.

As entregas devem ser feitas pelo *Classroom*. Os arquivos enviados devem ter a seguinte nomenclatura: “NN_ds4ps_grad_assignment_NOME_SOBRENOME.rmd”. Exemplo: “01_ds4ps_grad_assignment_davi_moreira.rmd”.

Seminários: Na primeira aula os alunos serão divididos em grupos. Por sorteio, cada grupo deverá se dedicar a um ou mais capítulos do livro [DSS]. As apresentações deverão ocorrer nas datas previstas nesse programa.

Presença: Os alunos devem comparecer a todas as aulas do semestre. O aluno que não comparecer a pelo menos 75% das aulas será reprovado por falta. [Consulte o manual acadêmico da UFPE para mais detalhes.](#)

Aprovação: Os conceitos finais obedecerão aos critérios abaixo:

Conceito	Crítérios Mínimos	Resultado
A	- Todas as tarefas entregues e - 90% dos pontos possíveis e - Seminário apresentado e - 75% de presença.	Aprovado
B	- 80% dos pontos possíveis e - Seminário apresentado e - 75% de presença.	Aprovado
C	- 60% dos pontos possíveis e - Seminário apresentado e - 75% de presença.	Aprovado
D	- Menos de 60% dos pontos possíveis ou - Seminário não apresentado ou - Menos de 75% de presença.	Reprovado

Código de honra e ética

- Os alunos são incentivados a discutir estratégias de resolução dos exercícios, mas espera-se que os resultados representem seus esforços individuais.

- Desonestidade acadêmica não será tolerada. Todos os alunos devem ler a cartilha sobre plágio acadêmico (<https://www3.ufpe.br/dqf/images/documentos/cartilha-plagio-academico.pdf>) e evitar tal prática. Qualquer tipo de desonestidade acadêmica será reportado às instâncias superiores responsáveis da UFPE.
- Relatório da Comissão de Integridade de Pesquisa do CNPq: <http://www.cnpq.br/documents/10157/a8927840-2b8f-43b9-8962-5a2ccfa74dda>
- Código de Boas Práticas Científicas: <http://www.fapesp.br/8577>
- Matérias referentes ao plágio e a como evitá-lo:
<http://revistapesquisa.fapesp.br/2017/07/18/universidades-brasileiras-contra-o-plagio/>
<http://revistapesquisa.fapesp.br/2018/09/14/quatro-tons-de-plagio-academico/>

Observação geral

De acordo com a demanda e dinâmica do curso, ajustes podem ser feitos no programa. Ademais, outras referências serão passadas no material do curso e durante as aulas.

Programa

Tópicos	Tema	Bibliografia
1	Introdução	[R4DS] Caps. 1, 2 e 27 [DSS] Cap. 1 [QSS] Cap. 1
2	Processamento de dados	[R4DS] Caps. 5, 11 e 12
3	Visualização de dados I: Gráficos	[R4DS] Caps. 3 e 28 [DV] Caps. 1 e 3
4	Análise exploratória	[R4DS] Cap. 7 [DBRFCor1] [DBRFCor2]
5	Visualização de dados II: Mapas	[DV] Caps. 7 e 8 [QSS] Cap. 5
6	Inferência, Amostragem, Estimação	[QSS] Cap. 7 [MDR] Cap. 7 e 8 [FPSR] Cap. 5 a 7 [DSS] Caps. 3, 4 e 7
7.1	Estimação e Modelos de Regressão Múltipla	[FPSR] Cap. 8 [R4DS] Caps. 22, 23 e 24 [MDR] Cap. 5, 6, 9 e 10 [DSS] Caps. 3, 4 e 7
7.2	Modelos de Regressão Múltipla (pressupostos e viés de variável omitida)	[FPSR] Cap. 8 e 9 [DBRFReg]
7.3	Modelos de Regressão Múltipla (transformações logarítmicas e variável independente categórica)	[FPSR] Cap. 9 e 10
7.4	Modelos de Regressão Múltipla (interação entre variáveis independentes)	[FPSR] Cap. 10
7.5	Modelos lineares generalizados (variável dependente categórica binária e nominal)	[FPSR] Cap. 10 e 11
11	Desenvolvendo uma pesquisa científica	[FPSR] Cap. 12 [DBRF7R] [RPREP] [DBRFSig1]

		[DBRFSig2] [DBRFSig3]
--	--	--------------------------

Calendário de aulas

#	data	tema
1	09/03/2022	Apresentação do curso.
2	16/03/2022	Tópico 1: introdução
3	23/03/2022	Tópico 2: processamento de dados
4	30/03/2022	Tópico 3: visualização de dados I - gráficos
5	06/04/2022	Tópico 4: análise exploratória
6	13/04/2022	Tópico 5: visualização de dados II - mapas
7	20/04/2022	Tópico 6: inferência, amostragem, estimação
8	27/04/2022	Tópico 7.1: estimação e modelos de regressão múltipla
9	04/05/2022	Tópico 7.2: modelos de regressão múltipla (pressupostos e viés de variável omitida)
10	11/05/2022	Tópico 7.3: modelos de regressão múltipla (transformações logarítmicas e variável independente categórica)
11	18/05/2022	Tópico 7.4: modelos de regressão múltipla (interação entre variáveis independentes)
12	25/05/2022	Tópico 7.5: modelos lineares generalizados (variável dependente categórica binária e nominal)
13	01/06/2022	Tópico 11: desenvolvendo uma pesquisa científica
14	08/06/2022	Seminário: capítulos 2, 3 e 4
15	15/06/2022	Seminário: capítulos 5, 6 e 7

Bibliografia

[DSS] LLAUDET, E. IMAI, K. Data Analysis for Social Science, A Friendly Introduction with Kosuke Imai. Princeton University Press.

[QSS] IMAI, K. Quantitative Social Science: An Introduction. Princeton University Press, 2018.

[FPSR] Kellstedt, P. M., & Whitten, G. D. (2013). The Fundamentals of Political Science Research (2 edition). Cambridge: Cambridge University Press.

[MDR] ISMAY, C.; KIM, A. Y. Statistical Inference via Data Science: A ModernDive into R and the Tidyverse. 1st edition ed. Boca Raton: Chapman and Hall/CRC, 2019. <https://moderndive.com/index.html>

[R4DS] WICKHAM, H.; GROLEMUND, G. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 1 edition ed. Sebastopol, CA: O'Reilly Media, 2017. Disponível em: <https://r4ds.had.co.nz/>

[TXTBRA] IZUMI, M. Y.; MOREIRA, D. C. O Texto como Dado: Desafios e Oportunidades para as Ciências Sociais. REVISTA BRASILEIRA DE INFORMAÇÃO BIBLIOGRÁFICA EM CIÊNCIAS SOCIAIS - BIB, v. 2, n. 86, p. 138–174, 2018.

[TXT4RI] MOREIRA, D., PIRES, A. MEDEIROS (*working paper*). Do 'texto como texto' ao 'texto como dado': o potencial das pesquisas em Relações Internacionais.

[TXT4CS] MOREIRA, D. C. (2019). Text as Data for Social Sciences. Independent Publication. Disponível em: https://bookdown.org/davi_moreira/txt4cs/

[MQCP] FILHO, D. B. F. Métodos quantitativos em ciência política. 1ª edição ed. [s.l.] InterSaberes, 2019.

- [RMD] XIE, Y. bookdown: Authoring Books and Technical Documents with R Markdown. Disponível em: <https://bookdown.org/yihui/bookdown/>
- [DV] HEALY, K. Data Visualization: A Practical Introduction. Princeton University Press, 2018.
- [DBRFCor1] Filho, D. B. F., & Junior, J. A. S. (2010). Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r). Revista Política Hoje - ISSN: 0104-7094, 18(1). Recuperado de <https://periodicos.ufpe.br/revistas/politica hoje/article/view/3852>
- [DBRFCor2] Paranhos, R., Filho, D. B. F., Rocha, E. C. da, Júnior, J. A. da S., Neves, J. A. B., & Santos, M. L. W. D. (2014). Desvendando os Mistérios do Coeficiente de Correlação de Pearson: o Retorno. Leviathan (São Paulo), (8), 66–95. <https://doi.org/10.11606/issn.2237-4485.lev.2014.132346>
- [RPRRep] Paranhos, R., Filho, D. B. F., Rocha, E. C. da, & Carmo, E. F. do. (2014). A Importância da Replicabilidade na Ciência Política: O Caso do SIGOBR. Revista Política Hoje - ISSN: 0104-7094, 22(2), 213–229.
- [DBRFSig1] Filho, F., Britto, D., Paranhos, R., Rocha, E. C. da, Batista, M., Jr, S., Marino, J. G. (2013). When is statistical significance not significant? Brazilian Political Science Review, 7(1), 31–55. <https://doi.org/10.1590/S1981-38212013000100002>
- [DBRFSig2] Silva, G. P. da, Guarnieri, F. H., Silva, G. P. da, & Guarnieri, F. H. (2014). Comments on When is Statistical Significance not Significant? Brazilian Political Science Review, 8(2), 133–136. <https://doi.org/10.1590/1981-38212014000100018>
- [DBRFSig3] Filho, F., B, D., Rocha, E. C. da, Batista, M., Paranhos, R., Jr, S., ... Da, J. A. (2014). Reply on the Comments on When is Statistical Significance not Significant? Brazilian Political Science Review, 8(3), 141–150. <https://doi.org/10.1590/1981-38212014000100024>
- [DBRFRReg] Filho, D. F., Nunes, F., Rocha, E. C. da, Santos, M. L., Batista, M., & Júnior, J. A. S. (2011). O Que Fazer e o Que Não Fazer Com a Regressão: pressupostos e aplicações do modelo linear de Mínimos Quadrados Ordinários (MQO). Revista Política Hoje - ISSN: 0104-7094, 20(1). Recuperado de <https://periodicos.ufpe.br/revistas/politica hoje/article/view/3808>
- [DBRF7R] FILHO, D. F. et al. Seven Reasons Why: A User's Guide to Transparency and Reproducibility. Bras. Political Sci. Rev., v. 13, 29 jul. 2019.

Bancos de dados

- AidData: <http://aiddata.org/>
- Banco Mundial (World Bank). <http://data.worldbank.org/>
- CIA World Factbook. <https://www.cia.gov/library/publications/the-world-factbook/rankorder/rankorderguide.html>
- CEPESP data: <https://cepespdata.io/>
- Comparative Study of Electoral Systems (CSES). <http://www.cses.org/>
- Correlates of War. http://www.correlatesofwar.org/data-sets/folder_listing
- FMI. <http://data.imf.org/>
- Global Health Observatory (OMS). <http://apps.who.int/gho/data/?theme=home>
- Human Development Reports (ONU) (IDH, Inequality, Gender Development, Gender Inequality, Poverty). <http://hdr.undp.org/en/data-explorer>
- Latinobarómetro. <http://www.latinobarometro.org/latContents.jsp>
- OCDE. <https://data.oecd.org/>
- OMC. https://www.wto.org/english/res_e/statis_e/statis_e.htm
- ONU. <http://data.un.org/Explorer.aspx>
- Paul Hensel's International Relations Data Site. <http://paulhensel.org/data.html>
- Penn World Tables. <https://pwt.sas.upenn.edu/>
- Quality of Government Institute. <http://qog.pol.gu.se/data/datadownloads>

Transparency International (Corruption, Bribe Payer). <https://www.transparency.org/research/>
World Values Survey. <http://www.worldvaluessurvey.org/WVSContents.jsp>