

Análise de dados COVID-19 em Portugal

Analysis of Portuguese COVID-19 data

by

João F. Pereira

Índice

1. Introdução

Modelos Epidemiológicos

2. Análise Exploratória

Obtenção dos dados

Processamento

Transformação

Exploração

Resultados

3. Demonstração Prática

Introdução



jn.pt, 23 Março 2021. Fonte: Lusa

COVID-19 inCTRL

Financiado pelo programa:
RESEARCH4COVID - FCT

utad UNIVERSIDADE
DE TRÁS-OS-MONTES
E ALTO DOURO

 Instituto Nacional de Saúde
Doutor Ricardo Jorge

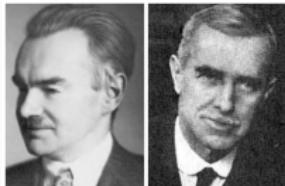
Modelos Epidemiológicos

Modelo SIR



Modelo SEIR

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{SI}{N} \\ \frac{dI}{dt} = \frac{SI}{N} - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases}$$



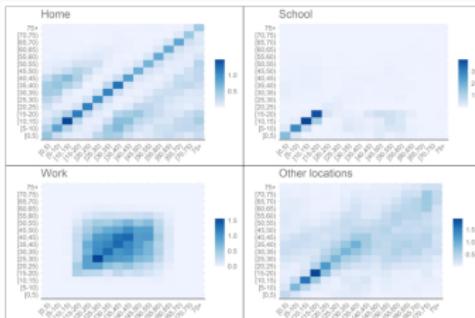
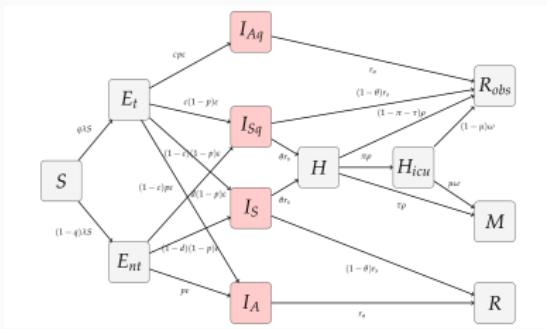
W.O. Kermack e A.G. McKendrick



$$\begin{cases} \frac{dS}{dt} = -\beta \frac{SI}{N} \\ \frac{dE}{dt} = \beta \frac{SI}{N} - \epsilon E \\ \frac{dI}{dt} = \epsilon E - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases}$$

Modelos Epidemiológicos

Modelo COVID-19 inCTRL



$$S' = -\lambda S,$$

$$E'_t = q\lambda S - \epsilon E_t$$

$$I'_A = p\epsilon E_{nt} + (1-c)p\epsilon E_t - r_a I_A$$

$$I'_S = (1-d)(1-p)\epsilon E_{nt} + (1-c)(1-p)\epsilon E_t - r_s I_S$$

$$E'_{nt} = (1-q)\lambda S - \epsilon E_{nt}$$

$$I'_{Aq} = cpe E_t - r_a I_{Aq}$$

$$I'_{Sq} = c(1-p)\epsilon E_t + d(1-p)\epsilon E_{nt} - r_s I_{Sq}$$

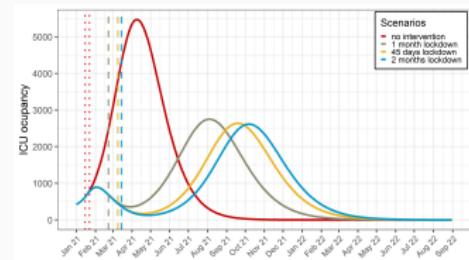
$$H' = \theta r_s (I_S + I_{Sq}) - \rho H$$

$$H'_{icu} = \pi \rho H - \omega H_{icu}$$

$$M' = \mu \omega H_{icu} + \tau \rho H$$

$$R'_{obs} = (1-\theta)r_s I_{Sq} + r_a I_{Aq} + (1-\pi-\tau)\rho H \\ + (1-\mu)\omega H_{icu}$$

$$R' = (1-\theta)r_s I_S + r_a I_A$$



Análise exploratória

Como realizar uma Análise exploratória?

Análise exploratória

Como realizar uma Análise exploratória?

1. Obtenção dos dados

Análise exploratória

Como realizar uma Análise exploratória?

1. Obtenção dos dados
2. Processamento

Análise exploratória

Como realizar uma Análise exploratória?

1. Obtenção dos dados
2. Processamento
3. Transformação

Análise exploratória

Como realizar uma Análise exploratória?

1. Obtenção dos dados
2. Processamento
3. Transformação
4. Exploração

Análise exploratória

Como realizar uma Análise exploratória?

1. Obtenção dos dados
2. Processamento
3. Transformação
4. Exploração
5. Resultados

Obtenção dos dados

- Os dados são relativos a APR-DRG (all-patient refined diagnosis-related groups) de Portugal, relativamente a pacientes com diagnóstico de COVID-19
- Disponibilizados pelos SPMS /ACSS de Portugal, na plataforma de acesso restrito BI-MH
- Entradas de 4 de janeiro 2020 até saídas a 22 de junho de 2021
- 56 666 entradas/saídas hospitalares registadas (linhas), cada uma contendo 13 variáveis (colunas)

Apesar da informação contida ser pseudo-anonimizada, por se tratarem de dados hospitalares, foi necessário a aprovação por um comité ético a utilização dos dados.

Processamento

Objetivo: Transformar os dados originais para outros mais utilizáveis e verificar possíveis incongruências.

1. Identificação das variáveis
2. Alteração de designações
3. Transformação de variáveis
(ex: idade, grupos etários, LoS, etc.)
4. Problema de representação
(consideradas apenas entradas entre 1 de Março de 2020 e 31 de Março de 2021)
5. Remoção de variáveis sem interesse

Dados Originais

Variable	Description
Health institution	Hospital where the entry took place
Gender	Gender of the patient
Date of birth	Date of birth of the patient
Hospital process	Process identification number
Hospital episode	Episode identification number
Destination after discharge	Includes 21 types of discharges, including fatality
Date of entry	Date of entry of the patient in hospital
Admission to ICU	If patient received Intensive Care Unit (ICU) treatment during hospitalization
Type of episode	Designation of the length of stay in hospital
Diagnostic code	Code of the diagnosis, according with APR-DIG regulations
Discharge date	The date of discharge for the diagnostic code
Order of diagnosis	The order the patient arrived in that day
Date of discharge	Date when the patient was discharged



Dados Processados

Variable	Designation	Levels	Description
Region	Region	Noite, Centro, LVT, Alentejo and Algarve	General Health Administration (ARS) of the health institution.
Type of hospitalization	Type	nonICU and ICU	Non-ICU patients did not receive intensive care unit (ICU) treatment. ICU patients did.
Age	Age	Whole numbers from 0 to 105	Age of the patient, at the time of entry
Age Group	Group	[0,5), [5,10), (...), [75,80), [80,150]	Designated age group of the patient, based on age at entry.
Gender	Gender	Feminine and Masculine	Gender of the patient.
Entry date	Entry.date	Date in format "%d/%m/%Y"	Date of patient's entry into the institution.
Discharge date	Discharge.date	Date in format "%d/%m/%Y"	Patient discharge date from the health institution.
Outcome	Outcome	Deceased and Discharged	If the patient died in hospital care or was discharged.
Length of stay	Time.stay	Whole numbers	Number of days between the discharge and entry date.
Month of entry	Entry.month	Date in format "%m/%Y"	Month of entry in hospital care.
Month of discharge	Discharge.month	Date in format "%m/%Y"	Month of discharge from hospital care.

Transformação

Objetivo: Transformar os dados noutros com nova informação.

Tempo de permanência

LoS	mean	median	25%	75%	90%	95%	99%	max	IQR
Non-ICU	12.54	8	4	15	27	38	67	348	11
ICU	24.31	18	11.0	30.3	51.0	66.0	112.0	216	19.3

Dados por dia

Designations	Description
date	Row day, from 1 March 2020 to 31 March 2021.
entries	Entries by day.
discharges	Discharges by day.
entries.nonICU	Entries of non-ICU patients, by day.
discharges.nonICU	Discharges of non-ICU patients, by day.
in.nonICU	Non-ICU patients in hospital care, on the respective day.
entries.ICU	Entries of ICU patients, by day.
discharges.ICU	Discharges of ICU patients, by day.
in.ICU	CU patients in hospital care, on the respective day.
deaths.nonICU	Deceased non-ICU patients, by day.
deaths.ICU	Deceased ICU patients, by day.
sum.deaths.nonICU	Cumulative sum of deceased non-ICU patients.
sum.deaths.ICU	Cumulative sum of deceased ICU patients.

Dados Processados

Variable	Designation	Levels	Description
Region	Region	Norte, Centro, IAT, Alentejo and Algarve	General Health Administration (ARS) of the health institution.
Type of hospitalization	Type	nonICU and ICU	Non-ICU patients did not receive intensive care unit (ICU) treatment, ICU patients did.
Age	Age	Whole numbers from 0 to 105	Age of the patient, at the time of entry.
Age Group	Group	[0,5], [5,10], (...), [75,80], [80,150]	Designated age group of the patient, based on age at entry.
Gender	Gender	Promotional and Masculine	Gender of the patient.
Entry date	Entry.date	Date in format "%d/%m/%Y"	Date of patient's entry into the institution.
Discharge date	Discharge.date	Date in format "%d/%m/%Y"	Patient discharge date from the health institution.
Outcome	Outcome	Decased and Discharged	If the patient died in hospital care or was discharged.
Length of stay	Time.stay	Whole numbers	Number of days between the discharge and entry date.
Month of entry	Entry.month	date in format "%m/%Y"	Month of entry in hospital care.
Month of discharge	Discharge.month	date in format "%m/%Y"	Month of discharge from hospital care.



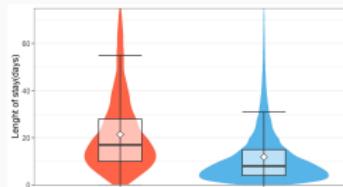
Parâmetros em relação a variáveis

Variable	Designations	Description
Group	Group	One of the following: Total, Region, Gender, Group and Entry.month.
Level	Level	Row level, according to group.
Entries	Entries.(Type)	Number of entries.
Percentage to ICU	Prop.ICU = $\frac{\text{Entries.ICU}}{\text{Entries.Total}}$	ICU patients percentage in Total population.
Fatalities	Fatalities.(Type)	Number of fatalities.
Fatality rate	Fatal_rate.(Type) = $\frac{\text{Fatalities.(Type)}}{\text{Entries.(Type)}}$	Proportion of entries with the Deceased outcome.
Proportion of all entries	Properties.(Type) = $\frac{\text{Entries.(Type)}}{\sum(\text{Entries.(Type)})}$	Proportion of the total Type entries.
Population	Pop	Population numbers.
Percentage of the population	PropProperties.Pop.(Type) = $\frac{\text{Properties.(Type)}}{\text{Pop}}$	Proportion of the population of each Level.

Exploração

Objetivo: Relacionar as diferentes variáveis de forma a encontrar relações de interesse, utilizando variados tipos de visualização.

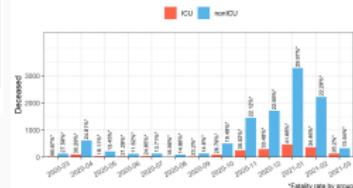
Diagramas de caixa /
Gráficos de violino



Gráficos de linha



Gráficos de barras



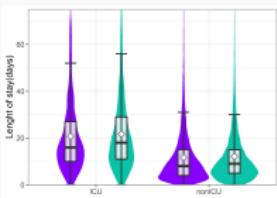
Exploração

Variáveis a comparar

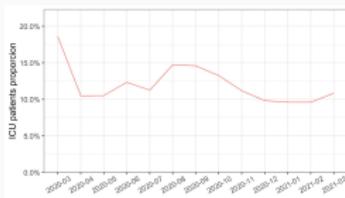
Tipo de hospitalização (com e sem UCI)



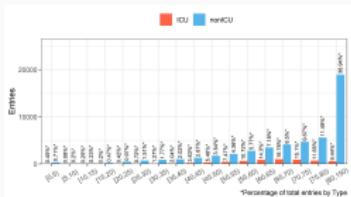
Género



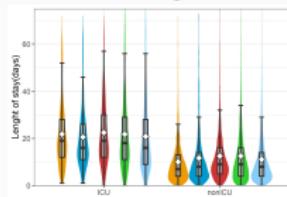
Tempo



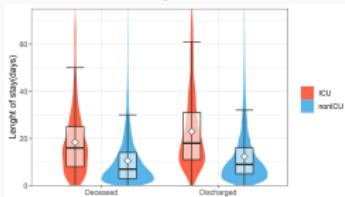
Grupo etário



Região



Resultado da hospitalização



Resultados

Objetivo: Identificar variáveis e relações de interesse para a questão em estudo.

$$\text{Proporção de pacientes UCI} = \frac{\text{entradas UCI}}{\text{Total entradas}} \times 100\% = 10.92\%$$

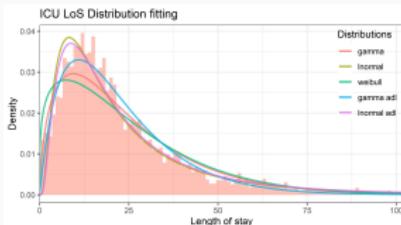
$$\text{Mortalidade de pacientes ICU} = \frac{\text{fatalidades não-UCI}}{\text{entradas não-UCI}} \times 100\% = 22.60\%$$

$$\text{Mortalidade de pacientes não-UCI} = \frac{\text{fatalidades UCI}}{\text{entradas UCI}} \times 100\% = 32.84\%$$

Parâmetros de interesse, em relação a várias variáveis

Variable	Level	mean non-ICU LoS	mean ICU LoS	ICU percentage	non-ICU fatality rate	ICU fatality rate
Total	Total	12.54	24.31	0.11	0.23	0.33
Gender	Female	12.32	23.82	0.07	0.22	0.31
Gender	Male	12.74	24.53	0.14	0.23	0.34
Group	(0,5]	6.28	33.15	0.07	0.60	0.04
Group	[5,10)	5.91	16.40	0.05	0.60	0.09
Group	[10,15)	7.13	13.87	0.12	0.60	0.07
Group	[15,20)	7.86	20.42	0.05	0.60	0.09
Group	[20,25)	7.08	18.64	0.05	0.60	0.09
Group	[25,30)	7.49	20.95	0.06	0.01	0.02
Group	[30,35)	7.12	14.14	0.08	0.01	0.09
Group	[35,40)	8.88	17.76	0.11	0.02	0.10
Group	[40,45)	9.17	20.10	0.14	0.02	0.11
Group	[45,50)	9.92	21.16	0.16	0.04	0.11
Group	[50,55)	10.86	23.31	0.17	0.04	0.10
Group	[55,60)	11.11	24.70	0.19	0.06	0.23
Group	[60,65)	12.26	25.09	0.20	0.08	0.26
Group	[65,70)	13.05	25.50	0.19	0.12	0.37
Group	[70,75)	14.06	26.46	0.16	0.17	0.43
Group	[75,80)	14.32	25.79	0.11	0.26	0.52
Group	[80,150]	13.44	22.67	0.03	0.40	0.57

Ajuste a distribuições estatísticas



Resultados

Permitindo tirar as seguintes conclusões de maior interesse:

- ▶ Pacientes UCI passam mais tempo em hospital, independentemente do resultado
- ▶ Idade condiciona taxa de hospitalização, tempo de permanência, entre outros
- ▶ Homens são mais hospitalizados em UCI, mas tem igual tempo de permanecia e taxa de mortalidade
- ▶ Percentagem pacientes UCI e taxa de mortalidade variaram ao longo do tempo

Demonstração Prática

Convido todos os presentes a acompanharem a demonstração nos seus computadores.

A demonstração encontra-se disponível em :

<https://github.com/Pereirajpf/XXX.git>

, bem como os ficheiros desta apresentação em:
https://github.com/Pereirajpf/Thesis_Presentation_JPereira.git

