

Covid-19: Vorhersage der Hospitalisierungsrate

Autoren

Alexander Marquard, Qian Feng, Phu Nguyen

Informationen

Die folgenden Datensätze sind im Zusammenhang mit dem Statistischen Praktikum an der LMU München entstanden. Sie wurden erstellt um eine Antwort auf die Frage zu geben: Kann die Hospitalisierungsrate vorhergesagt werden. Bei den Datensätzen handelt es sich um Modifikationen der originalen Datensätze, wie diese auf der Github-Seite des RKI gefunden werden können. Zur Beantwortung der oben genannten Frage wurden die Datensätze VOC und impfung lediglich zur explorativen Analyse genutzt. Der Datensatz final_df hingegen wurde genutzt um mit Hilfe eines ARIMA-Modells eine Vorhersage der Hospitalisierungsrate in den einzelnen deutschen Bundesländern zu treffen.

Im Folgenden sind die Datensätze detailliert erklärt.

COVID-19-Fälle und hospitalisierte COVID-19-Fälle nach Bundesland und Altersgruppe (final_df.RDS)

Grundlegende Informationen zu den Daten

Dieser Datensatz zeigt die Summe der COVID-19-Fälle oder hospitalisierten COVID-19-Fälle innerhalb einer Kalenderwoche (Montag bis Freitag) direkt lesbar, stratifiziert nach Altersgruppen und nach Bundesländern. Um diesen Datensatz zu erstellen, werden Hospitalisierungsdaten (aus Robert-Koch-Institut), Infektionsdaten (aus Robert-Koch-Institut) und Bevölkerungsdaten (aus KitMetricsLab) verwendet.

Meldedatum und Hospitalisierungsdatum

Typischerweise findet der Meldeprozess zwischen dem Tag, an dem ein positives COVID-Testergebnis bei einer Person aufgewiesen ist, und dem Tag, an dem die Person mit starken Symptomen im Krankenhaus eingewiesen wird, statt. Eine Person mit positivem Testergebnis gilt als ein Covid-19-Fall, der zunächst bei dem lokalen Gesundheitsamt, und danach erst bei dem RKI gemeldet wird. Das Meldedatum ist das Datum, an dem ein COVID-Fall dem lokalen Gesundheitsamt bekannt wird. Abhängig von der Schwere des Verlaufs müssen einige Personen im Krankenhaus eingewiesen werden. Das Datum, an dem ein Patient ins Krankenhaus eingeliefert wird, wird als Hospitalisierungsdatum genannt. Aufgrund fehlender Daten bezüglich der Hospitalisierung werden die hospitalisierten COVID-19-Fälle zeitlich nach Meldedatum ausgewiesen. Somit entsprechend alle Daten im Datensatz dem Meldedatum der Infektion und eine Unterscheidung zwischen dem Meldedatum der Hospitalisierung und dem tatsächlichen Hospitalisierungsdatum ist nicht mehr möglich.

Geografischer Bezug der Hospitalisierungs- und Infektionsdaten

Bezüglich der Erfassung der Infektions- und Hospitalisierungsdaten werden seitens des RKI's Unterschiede bei der räumlichen Lage gemacht. Daten der Hospitalisierung werden in der Regel von dem Gesundheitsamt an das RKI übermittelt, das für den Wohnort der betroffenen Person zuständig ist. Wenn die betroffene Person z.B. in Berlin wohnhaft

ist, aber in Brandenburg hospitalisiert wird, dann erfolgt die Zuordnung zum Bundesland Berlin. Infektionsdaten hingegen werden entsprechend dem Ort des positiven Tests erfasst, unabhängig vom Wohnort der infizierten Person. Wurde demnach eine betroffene Person, wohnhaft z.B. in Berlin positiv in Brandenburg getestet, so erfolgt die Zuweisung allerdings in Brandenburg. Eine Zuordnung der Bettenbelegung in Krankenhäusern wird dadurch unmöglich, ebenso wenig, wie eine Betrachtung der Hospitalisierung in einzelnen Bundesländern, mit der Infektionsrate als abhängige Variable.

Variablen

Die Daten differenzieren nach verschiedenen Merkmalen einer sozialen Gruppe. Pro Eintrag bzw. Zeile ist eine eindeutige soziale Gruppe abgebildet. Eine soziale Gruppe wird grundlegend durch folgende Merkmale charakterisiert (in den Klammern finden sich die Variablen dieser Merkmale):

- Ort der Infektionen/Wohnort der hospitalisierten COVID-19-Fälle (Bundesland)
- Altersgruppe der COVID-19-Fälle/hospitalisierten COVID-19-Fälle (Altersgruppe)

Zusätzlich werden folgende Variablen angegeben:

- das Datum (Sonntag), an dem COVID-19-Fälle innerhalb einer respektiven Kalenderwoche (Montag bis Freitag) dem lokalen Gesundheitsamt bekannt wird (Meldedatum)
- Kalenderjahr respektives Meldedatums (Kalenderjahr)
- Kalenderwoche respektives Meldedatums (Kalenderwoche)
- Summe der COVID-19-Fälle innerhalb einer Kalenderwoche (Montag bis Freitag) in der Gruppe (Neuerkrankung)
- Summe der hospitalisierten COVID-19-Fälle innerhalb einer Kalenderwoche (Montag bis Freitag) in der Gruppe (Hospitalisierung)
- Anzahl der Bevölkerung in der Gruppe (population)
- Reihenfolgennummer der Kalenderwoche (index)

Eine soziale Gruppe nimmt eine eindeutige Ausprägung hinsichtlich der Anzahl der COVID-19-Fälle/hospitalisierten COVID-19-Fälle eines Bundeslandes und einer Altersgruppe an. Für jede Impfgruppe wird die wöchentliche Anzahl neuer COVID-19-Fälle/hospitalisierten COVID-19-Fälle ausgewiesen.

Variablenausprägungen

Die Daten enthalten die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen:

Variable	Typ	Ausprägung	Beschreibung
Meldedatum	Datum	JJJJ-MM-TT	das Datum (Sonntag), an dem COVID-19-Fälle innerhalb einer respektiven Kalenderwoche (Montag bis Freitag) dem lokalen Gesundheitsamt bekannt wird.
Kalenderjahr	Text	2020 2021 2022	Kalenderjahr respektives Meldedatums
Kalenderwoche	Text	1	Kalenderwoche respektives

		...	Melddatums
		53	
Bundesland	Text	Schleswig-Holstein ...	Name des Bundeslandes
		Thueringen	
Altersgruppe	Text	00-04, 05-14, 15-34, 35-59, 60-79, 80+	Altersgruppe der COVID-19-Fälle/hospitalisierten COVID-19-Fälle
Neuerkrankung	Natürliche Zahl	≥0	Summe der COVID-19-Fälle innerhalb einer Kalenderwoche (Montag bis Freitag) in der sozialen Gruppe
Hospitalisierung	Natürliche Zahl	≥0	Summe der hospitalisierten COVID-19-Fälle innerhalb einer Kalenderwoche (Montag bis Freitag) in der sozialen Gruppe
population	Natürliche Zahl	≥0	Anzahl der Bevölkerung in der sozialen Gruppe
index	Natürliche Zahl	≥1	Reihenfolgennummer der Kalenderwoche. Zum Beispiel ist die elfte Kalenderwoche im Jahr 2020 die elfte Woche in dem ganzen Datensatz, wird der „index“ dieser Beobachtung mit „11“ zugewiesen.

COVID-19 Impfdaten auf Ebene der Bundesländer (impfung.RDS)

Grundlegende Informationen zu den Daten

Dieser Datensatz zeigt die Anzahl der täglichen Impfungen, stratifiziert nach Bundesländern und nach Ständen vom Impfschutz (nur Vollständiger Impfschutz und Aufgefrischter Impfschutz). Um diesen Datensatz zu erstellen, werden Impfungsdaten (aus Robert-Koch-Institut) und Bevölkerungsdaten (aus KitMetricsLab) verwendet.

Variablen

Die Impfungsdaten differenzieren nach verschiedenen Merkmalen einer Impfgruppe. Pro Eintrag bzw. Zeile ist eine eindeutige Impfgruppe abgebildet. Eine Impfgruppe wird grundlegend durch folgende Merkmale charakterisiert (in den Klammern finden sich die Variablen dieser Merkmale):

- Ort der Impfung (Bundesland)
- Angabe zur Vollständigkeit/Auffrischung des Impfschutzes (Impfschutz)

Zusätzlich werden folgende Variablen angegeben:

- Datum der Impfung (Impfdatum)

- Anzahl der Impfungen in der Gruppe (Anzahl)
- Anzahl der Bevölkerung respektives Bundeslandes (population)

Eine Impfgruppe nimmt eine eindeutige Ausprägung hinsichtlich der Anzahl der Impfungen eines Bundeslandes der Geimpften und der Ausprägung des Impfschutzes an. Für jede Impfgruppe wird die tägliche Anzahl neuer Impfungen ausgewiesen, sofern diese größer vier ist.

Variablenausprägungen

Die Impfdaten enthalten die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen:

Variable	Typ	Ausprägung	Beschreibung
Impfdatum	Datum	JJJJ-MM-TT	Datum der Impfungen
Bundesland	Text	Schleswig-Holstein ... Thüringen	Name des Bundeslandes
Impfschutz	Natürliche Zahl	2: Vollständiger Impfschutz 3: Aufgefrischter Impfschutz	Angabe zum Impfschutz Vollständiger Impfschutz besteht bei zweifacher Impfung, Impfung mit Janssen und einfach Geimpften mit überstandener SARS-CoV-2 Infektion
population	Natürliche Zahl	≥0	Anzahl der Bevölkerung respektives Bundeslandes
Anzahl	Natürliche Zahl	≥5	Anzahl der Impfungen in der Impfgruppe

Ist die Anzahl an Impfungen einer Impfgruppe an einem Tag kleiner als fünf, werden, aus Gründen des Datenschutzes, an diesem Tag keine Impfungen für die Impfgruppe ausgewiesen. Um dennoch einen genauen Überblick über die Gesamtzahl der Impfungen zu ermöglichen, werden Impfgruppen mit weniger als fünf Impfungen zu Impfungen der Folgetage derselben Impfgruppe hinzuaddiert, bis die kumulierte Anzahl der Impfungen an einem Tag den Wert von fünf übersteigt.

Daten über besorgniserregende Varianten (VOC.RDS)

Grundlegende Informationen zu den Daten

Dieser Datensatz zeigt den Anteil in der Stichprobe (Sequenzdaten) detektierten und der über das Meldewesen (IfSG) übermittelten besorgniserregenden und unter Beobachtung stehenden Varianten (VOC). Um diesen Datensatz zu erstellen, werden VOC-Daten (aus Robert-Koch-Institut) verwendet.

Variablen

Die VOC-Daten differenzieren verschiedene Merkmale. Grundlegend nach folgende Merkmale differenziert (in den Klammern finden sich die Variablen dieser Merkmale):

- Kalenderjahr (KJ)
- Kalenderwoche (KW)
- Kalenderjahr-Kalenderwoche (KWKJ)
- Name von der Variante des Virus (Variante)
- Anteil der Variante des Virus in der Stichprobe (Anteil)

Variablenausprägungen

Die VOC-Daten enthalten die in der folgenden Tabelle abgebildeten Variablen und deren Ausprägungen:

Variable	Typ	Ausprägung	Beschreibung
KJ	Text	2021 2022	Kalenderjahr
KW	Text	1 ... 52	Kalenderwoche
KWKJ	Text	JJJJ-WW	Kalenderjahr-Kalenderwoche: JJJJ bedeutet Kalenderjahr, WW bedeutet Kalenderwoche
Variante	Text	Alpha Beta Delta Gamma Omikron	Name von der Variante des Virus
Anteil	Rationale Zahl	≥ 0	Anteil der Variante des Virus in der Stichprobe