Regression Analysis

Eszter Diamant

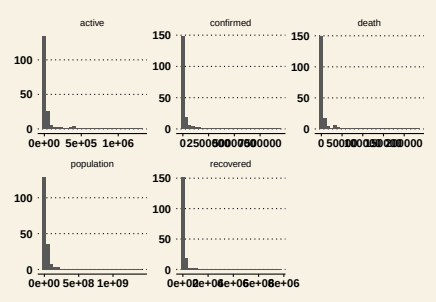
2020 11 29

R Markdown

To write: my research question To write: my variables To write: potential droppings and scaling

Summary statistics

To write: 2-3 sentence about summary statistics



	min	max	mean	median	sd	skew	iq_range
Confirmed	2	9320266	256647.89	21523	1023053.09	7.27	107784.00
Death	0	231713	6588.15	322	24471.42	6.56	1809.00

Table 1: Summary statistics for the confimred COVID-19 cases and death resulted due to the infection

Investigation of variables:

to write: substative reasoning (2-3), statistical reasoning

Investigate different models:

graphs at the appendix to write: what I can see in the charts to write: substative reasoning (2-3), statistical reasoning -> which one to use (in appendix the 2-3 sentence)

Make the three models:

1) level-level:

```
(chart 1): reg1: death = alpha + beta * confirmed
(chart 5): reg2: death = alpha + beta_1 * confirmed + beta_2 * confirmed^2
(chart 6): reg3: death = alpha + beta_1 * confirmed + beta_2 * confirmed^2 + beta_3 * confirmed
```

2) log -log:

```
(chart 4): reg4: log_death = alpha + beta * log_conf
(chart 7): reg5: log_death = alpha + beta_1 * log_conf + beta_2 * log_conf^2
(chart 8): reg6: log_death = alpha + beta_1 * log_conf + beta_2 * log_conf^2 + beta_3 * log_conf
(chart 9): reg7: log_death = alpha + beta_1 * log_conf * 1(log_conf < 7) + beta_2 * log_conf * 1
```

3) extra: weighted-ols:

```
chart 10): reg8: log_death = alpha + beta* log_conf, weights: population
```

	Confirmed cases - linear	Confirmed cases - quadratic	Confirmed cases - cubic	log confirmed cases - linear	log confirmed cases - quadratic	log confirmed cases- cubic	log confirmed cases - PLS	log confirmed cases - weighted linear
(Intercept)	892.20	-420.02		-4.24***	-1.77*	-1.37		-3.02***
	(660.12)	(619.48)		(0.30)	(0.82)	(1.56)		(0.76)
confirmed	0.02***	0.03***	0.04***					
	(0.00)	(0.01)	(0.01)					
conf_sq		-0.00	-0.00					
		(0.00)	(0.00)					
conf_cb			0.00					
			(0.00)					
log_conf				1.02***	0.50**	0.37		0.95***
				(0.03)	(0.16)	(0.53)		(0.06)
log_conf_sq					0.03***	0.04		
					(0.01)	(0.06)		
log_conf_cb						-0.00		
						(0.00)		
lspline(log_conf, cutoff_ln)1							-1.35***	
							(0.14)	
lspline(log_conf, cutoff_ln)2							1.02***	
							(0.03)	
R^2	0.88	0.90	0.90	0.89	0.90	0.90	0.89	0.93
Adj. R^2	0.88	0.89	0.90	0.89	0.89	0.89	0.89	0.93
Num. obs.	170	170	170	170	170	170	170	170
RMSE	8767.93	8217.81	8027.17	0.82	0.80	0.80	0.82	4223.30

***p < 0.001; **p < 0.01; *p < 0.05

Modelling confirmed COVID-19 cases and death resulted among them

Figure 1: Regression Models

Model choice

write the appropiate equation write: 2-3 sent, about interpretation

Hypothesis testing:

H0: log_conf = 0 HA: log_conf != 0

```
## Linear hypothesis test
##
## Hypothesis:
## log_conf = 0
##
## Model 1: restricted model
## Model 2: log_death ~ log_conf
```

```
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1    169
## 2    168  1 1334.2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analysis of residuals:

Countries that relatively saved the most people due to COVID-19:

country	log_death	reg4_y_pred	reg4_res
Burundi	0.000000	2.249810	-2.249810
Iceland	2.484907	4.411990	-1.927083
Qatar	5.446737	7.764013	-2.317276
Singapore	3.332205	6.920522	-3.588317
Sri Lanka	3.044522	5.258958	-2.214436

Countries that relatively lost the most people due to COVID-19:

country	log_death	reg4_y_pred	reg4_res
Chad	4.5849675	3.200339	1.384628
Ecuador	9.4487272	8.011783	1.436945
Fiji	0.6931472	-0.652333	1.345480
Mexico	11.4306302	9.747078	1.683552
Yemen	6.3985949	3.525310	2.873285

Appendix:

Chart 1: level - level linear regression

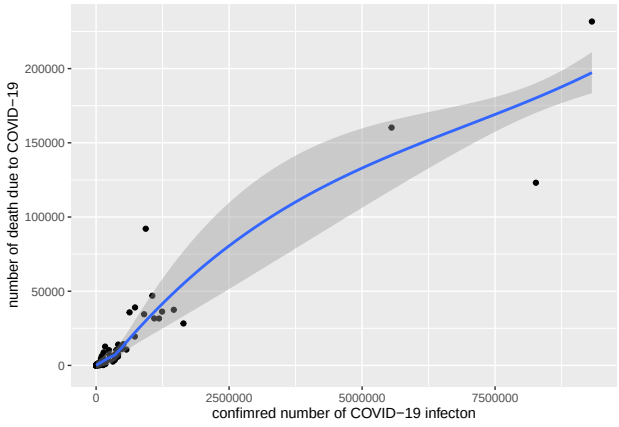


Chart 2: level - log linear regression

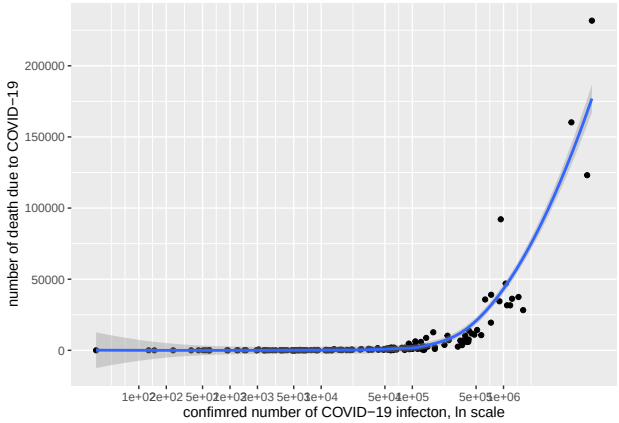


Chart 3: log - level linear regression

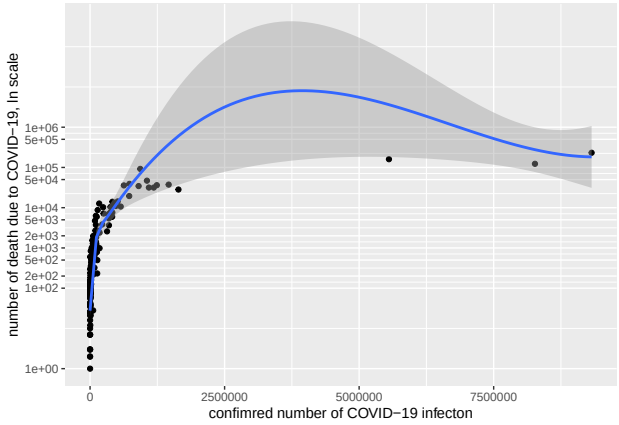


Chart 4: log - log regression linear regression

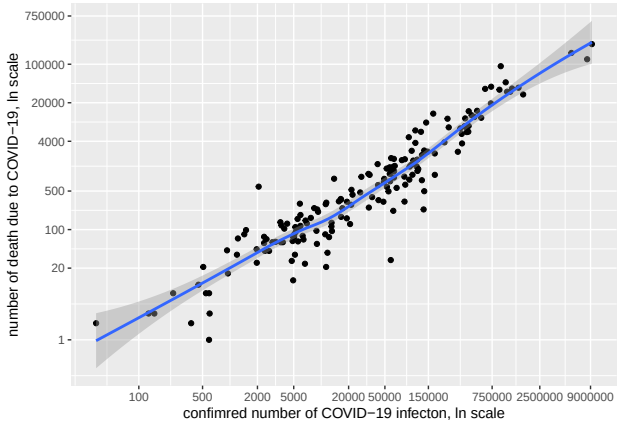


Chart 5: level-level quadratic regression

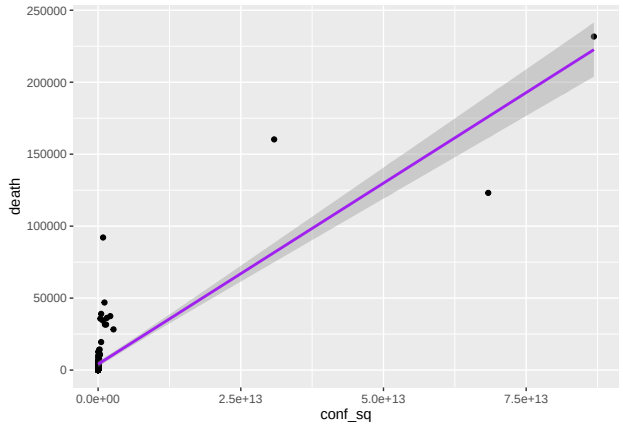


Chart 6: level - level cubic regression

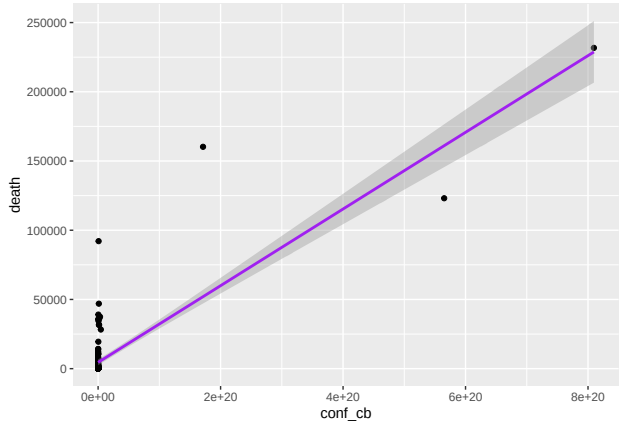


chart 7: log(death) - log(confirmed) quadratic regression

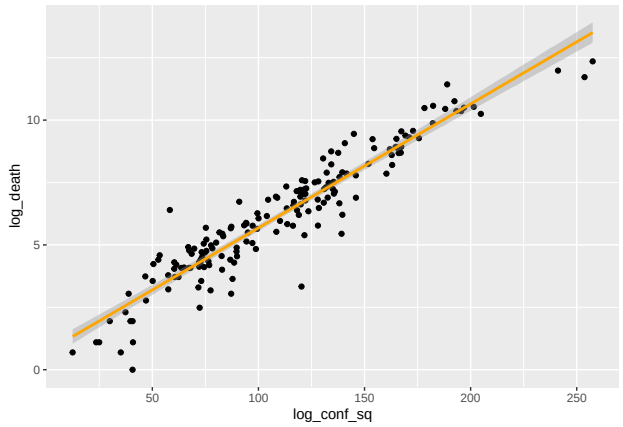


chart 8: log(death) - log(confirmed) cubic regression

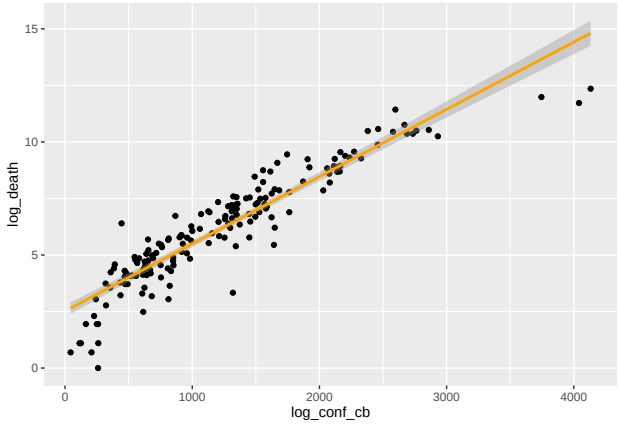


chart 9: Piecewise linear spline regression

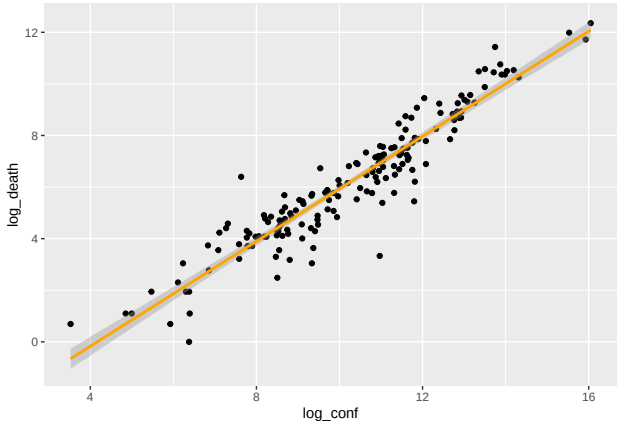


chart 10: Weighted linear regression, using population as weights