

Univerzita Pardubice
Fakulta elektrotechniky a informatiky

Zpracování dat pro předmět NMAST

Bc. Lukáš Milar, Bc. Tomáš Prudký

Semestrální práce

2021

OBSAH

Seznam obrázků	3
Seznam tabulek	4
Úvod	5
1 Popis dat	6
2 Popisná statistika	7
3 Základní grafy	9
3.1 Histogram	9
3.2 Bodový graf	13
3.3 Boxplot	18
3.4 Hexbin	19
3.5 3D graf	20
3.6 Chernoff faces	21
4 Testování statistických hypotéz	22
4.1 Jednovýběrový Studentův test vůči střední hodnotě	22
4.2 Dvouvýběrový Studentův test	22
4.3 Fisherův test	22
5 ANOVA	23
6 Korelace	24
6.1 Korelační matice	24
6.2 Bodové grafy hodnot z korelační matice	25
7 Testování v kontingenčních tabulkách	26
7.1 Chí kvadrát test	26
8 Lineární regrese	27
Závěr	29
Použitá literatura	30
Seznam příloh	31
Příloha A	32

SEZNAM OBRÁZKŮ

1	Histogram nových případů smoothed	9
2	Histogram nových případů na milión	9
3	Histogram nových případů na milión smoothed	10
4	Histogram: Hospitalizovaní pacienti v ČR	10
5	Nove testovani v Cesku	11
6	Nove pripady v Cesku	11
7	Nové testy pro Česko a Německo	12
8	Bodový graf zlogaritmovaných nových případů	13
9	Bodový graf nových testů	13
10	Bodový graf reprodukčního čísla	14
11	Bodový graf pacientů na icu	14
12	Bodový graf hospitalizovaných pacientů	15
13	Bodový graf týdenních přírůstků na icu	15
14	Bodový graf týdenních hospitalizací	16
15	Bodový graf pozitivitu testů	16
16	Bodový graf nových očkování	17
17	Bodový graf smrtnosti	17
18	Boxplot graf pro první dvě měny	18
19	Hexbin graf měny s maximální hodnotou šikmosti	19
20	Hexbin graf měny s minimální hodnotou šikmosti	19
21	3D graf graf měny s maximální hodnotou šikmosti	20
22	3D graf graf měny s minimální hodnotou šikmosti	20
23	Chernoff faces graf tabulky popisné statistiky	21
24	Bodový graf maximální hodnoty v korelační matici	25
25	Bodový graf minimální hodnoty v korelační matici	25
26	Graf lineární regrese	28

SEZNAM TABULEK

1	Části popisné statistiky aplikované na data	7
2	Korelační matice	24

ÚVOD

1 POPIS DAT

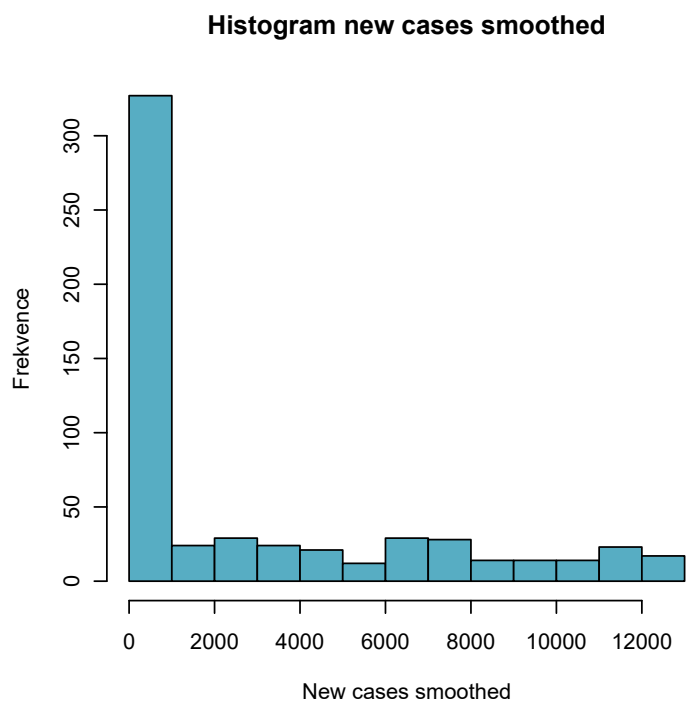
2 POPISNÁ STATISTIKA

	n_p	n_p_s	n_p_n_m	n_p_n_m_s	h_p	h_p_n_m
prumer	2940.58	2936.89	274.19	273.85	2370.49	221.03
modus	75.00	57.57	6.99	5.37	69.00	6.43
median	416.00	422.29	38.79	39.38	339.00	31.61
max	17773.00	12954.86	1657.22	1207.96	9509.00	886.66
min	-2214.00	2.71	-206.44	0.25	0.00	0.00
skewness	1.55	1.16	1.55	1.16	0.86	0.86
kurtosis	1.38	-0.07	1.38	-0.07	-0.86	-0.86
deviation	4277.10	3876.20	398.81	361.43	2973.01	277.22
var	18293577.15	15024928.55	159052.40	130633.34	8838793.84	76848.36

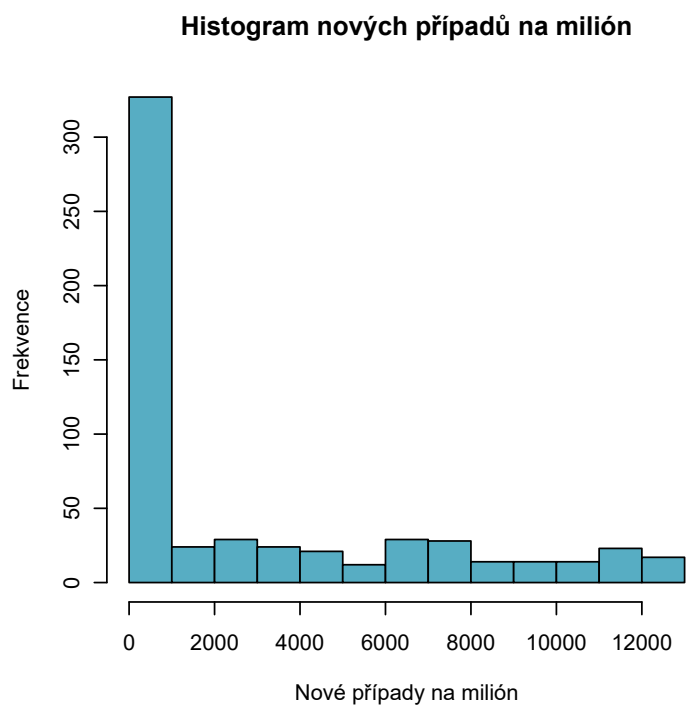
Tabulka 1: Části popisné statistiky aplikované na data

3 ZÁKLADNÍ GRAFY

3.1 Histogram

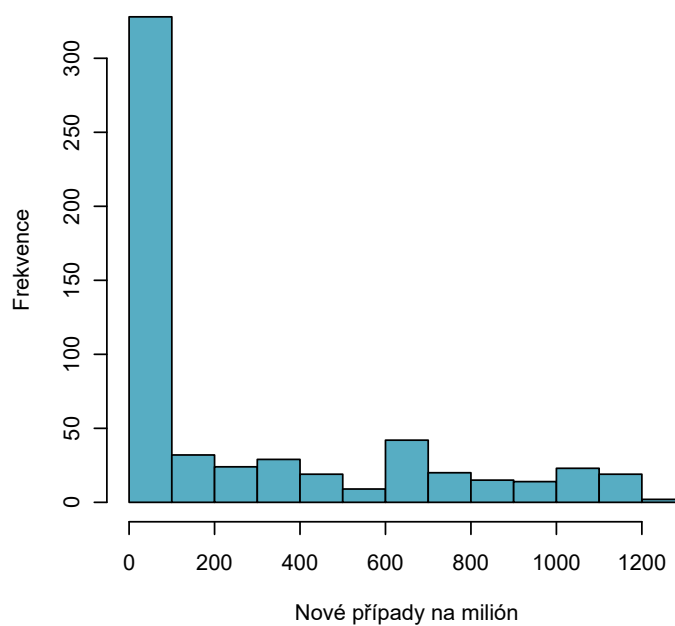


Obrázek 1: Histogram nových případů smoothed



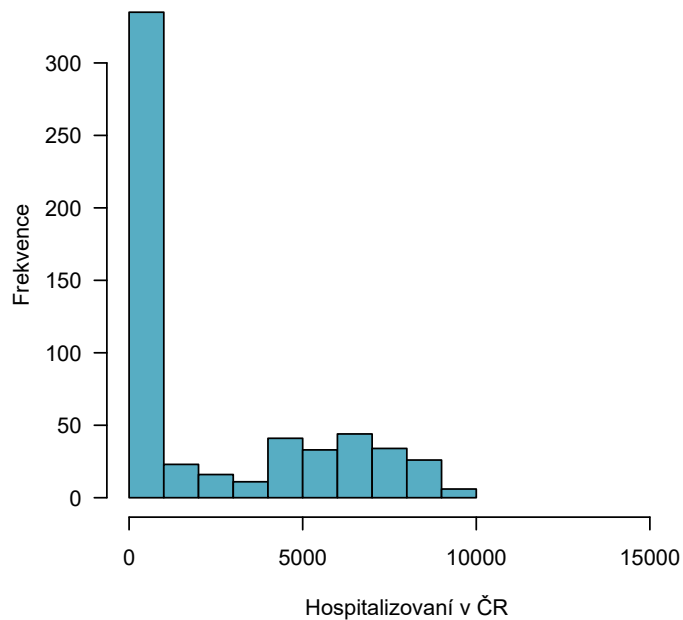
Obrázek 2: Histogram nových případů na milión

Histogram nových případů na milión smoothed

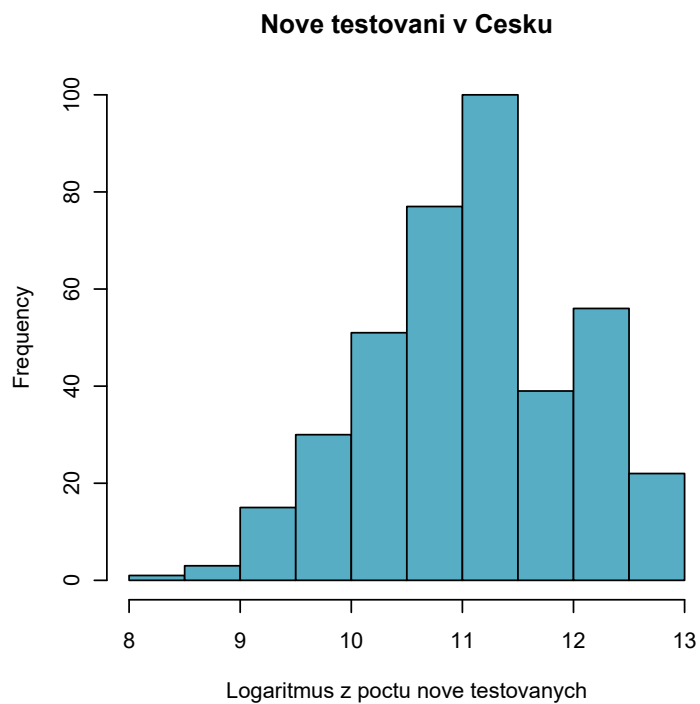


Obrázek 3: Histogram nových případů na milión smoothed

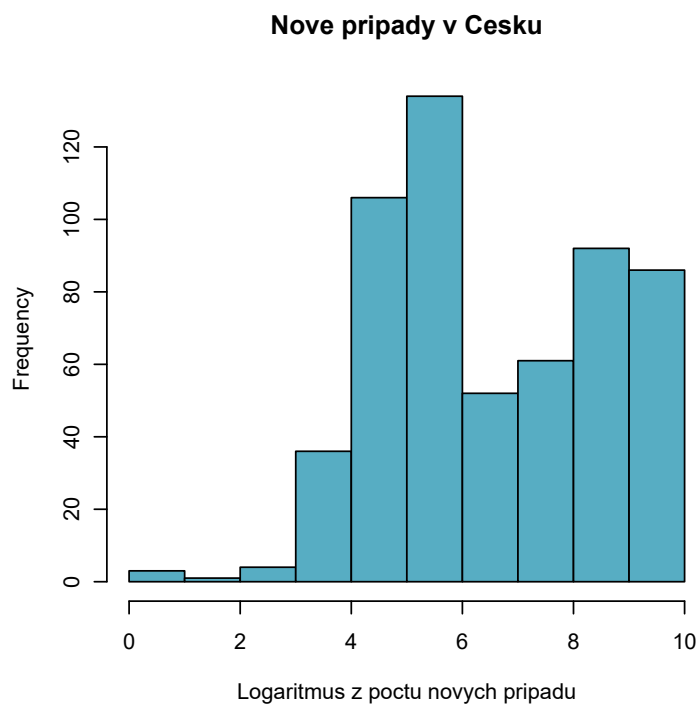
Histogram: Hospitalizovaní pacienti v ČR



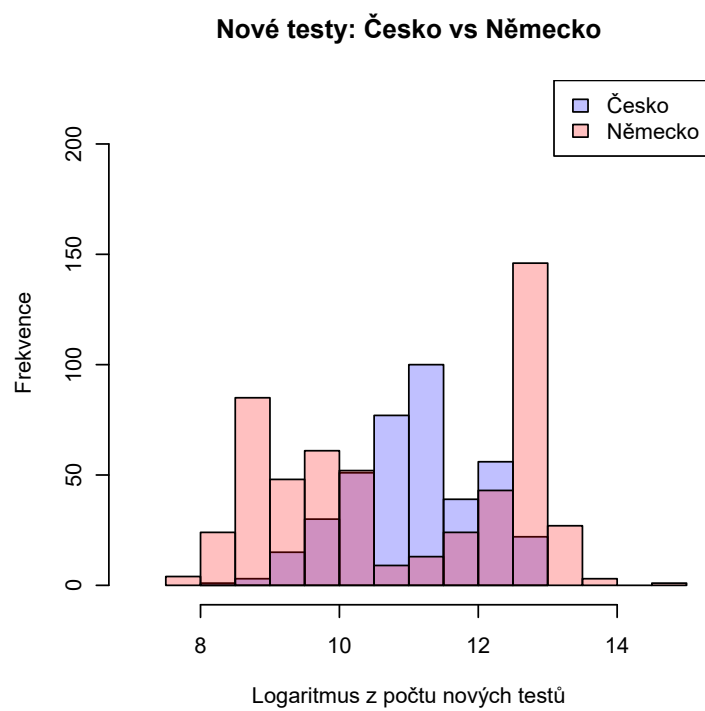
Obrázek 4: Histogram: Hospitalizovaní pacienti v ČR



Obrázek 5: Nove testovani v Cesku

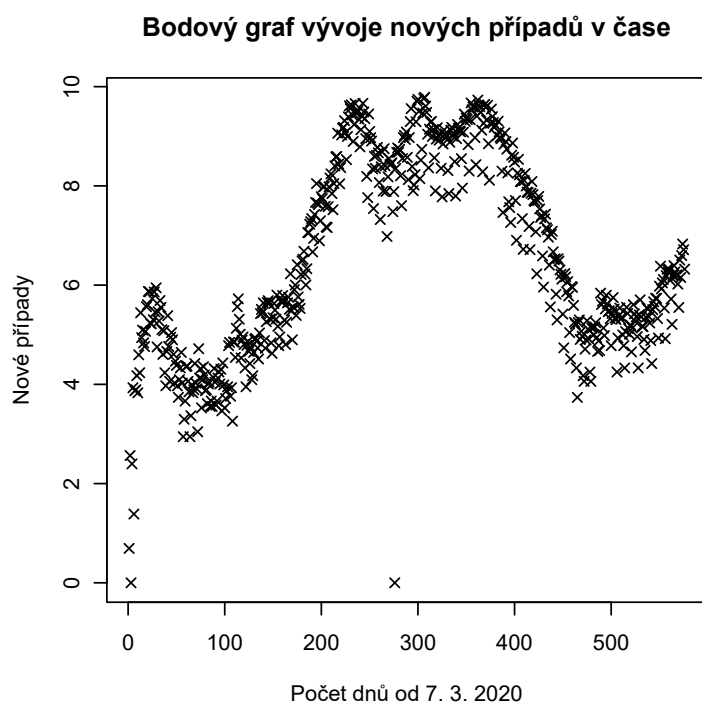


Obrázek 6: Nove pripady v Cesku

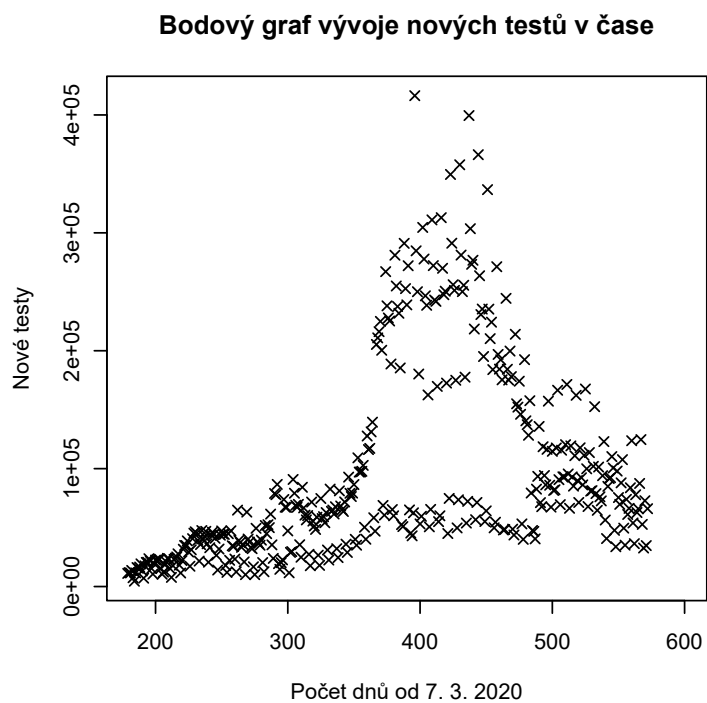


Obrázek 7: Nové testy pro Česko a Německo

3.2 Bodový graf

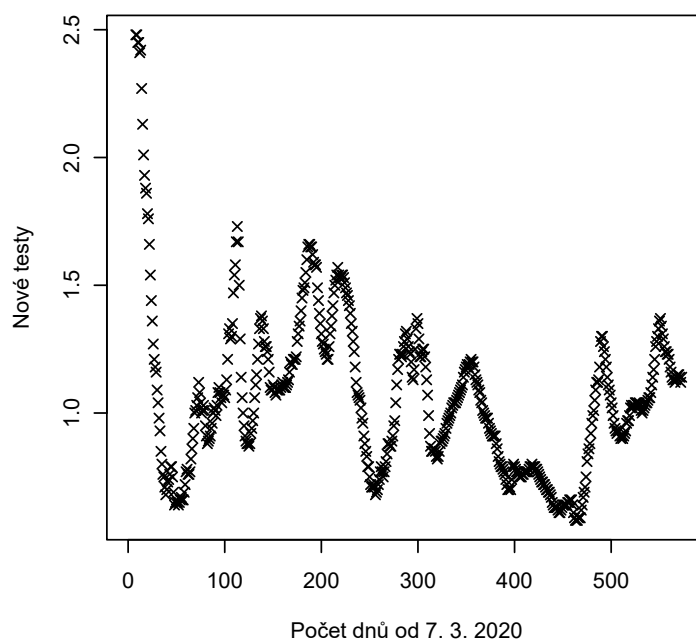


Obrázek 8: Bodový graf zlogaritmovaných nových případů



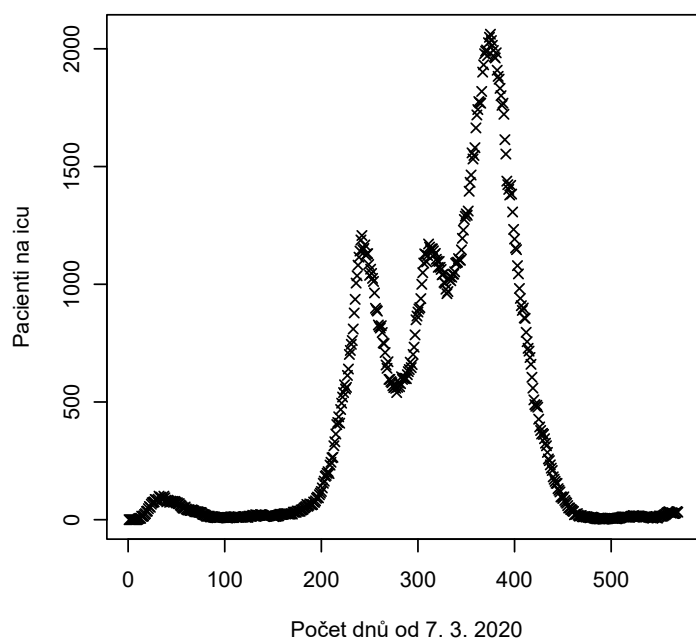
Obrázek 9: Bodový graf nových testů

Bodový graf vývoje reprodukčního čísla v čase



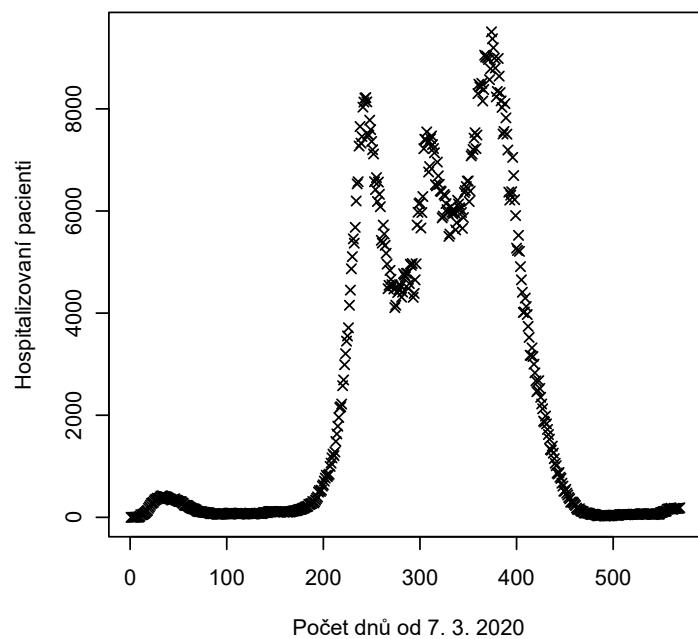
Obrázek 10: Bodový graf reprodukčního čísla

Bodový graf vývoje pacientů na icu v čase



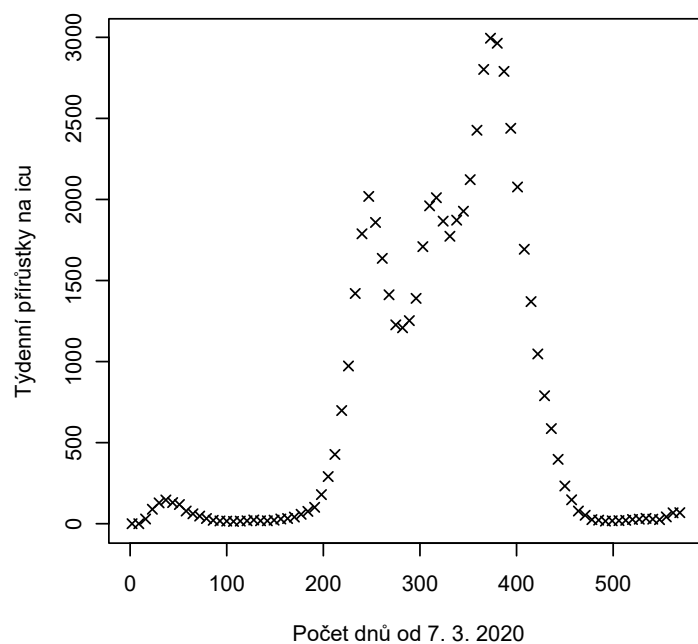
Obrázek 11: Bodový graf pacientů na icu

Bodový graf vývoje hospitalizovaných pacientů v čase

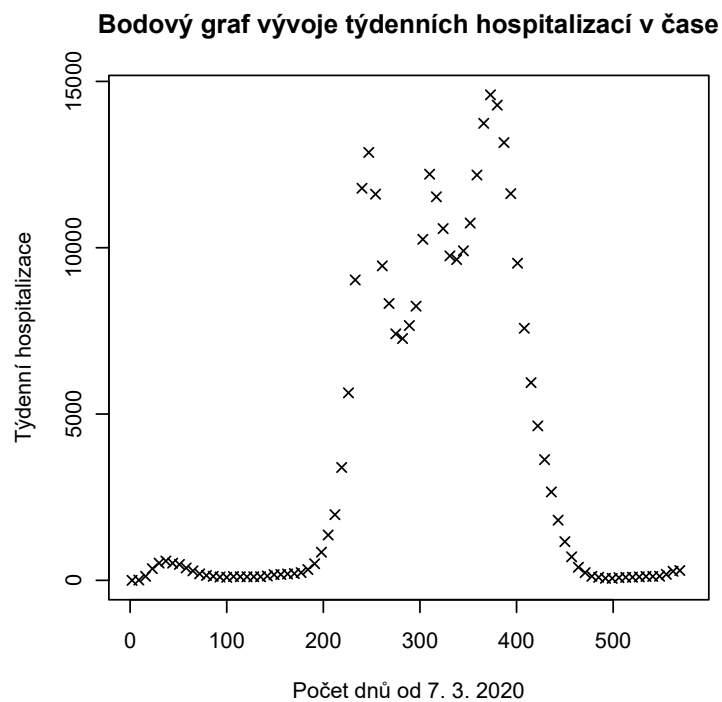


Obrázek 12: Bodový graf hospitalizovaných pacientů

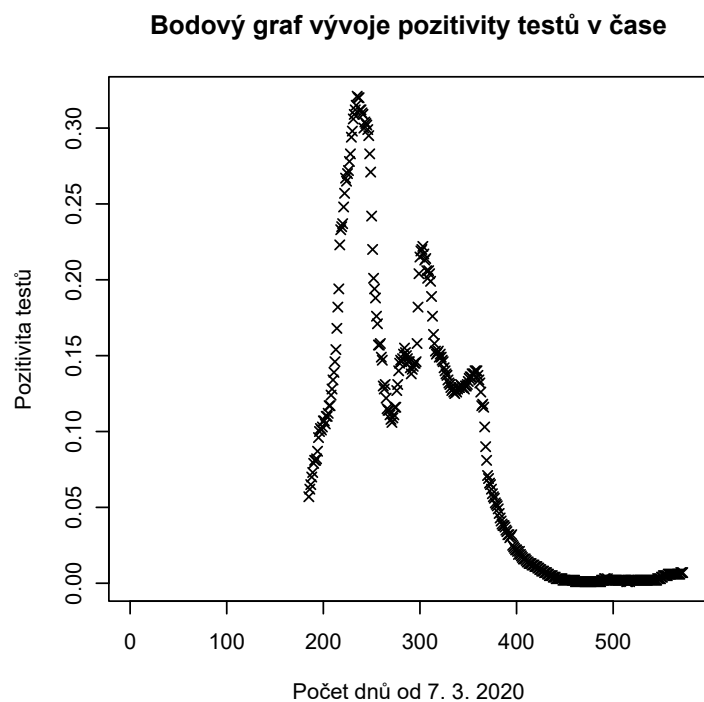
Bodový graf vývoje týdenních přírůstků na icu v čase



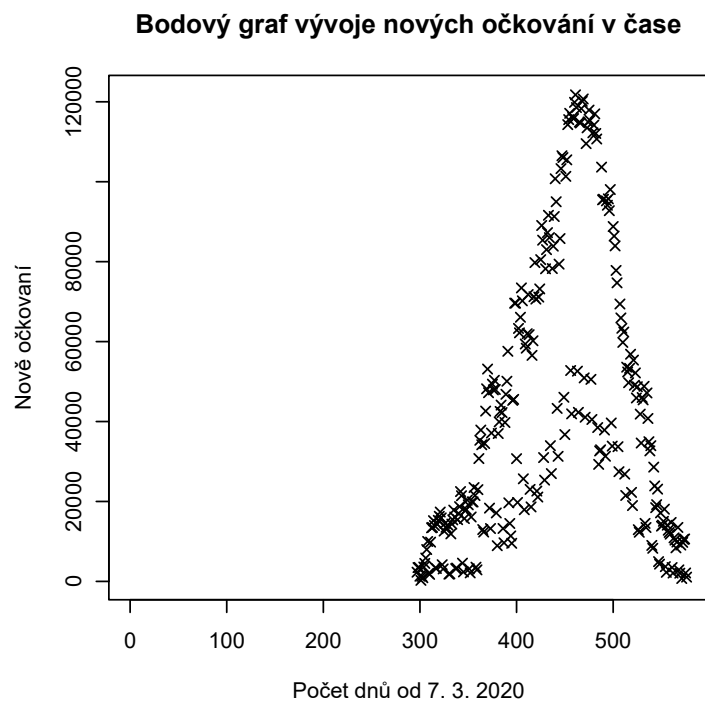
Obrázek 13: Bodový graf týdenních přírůstků na icu



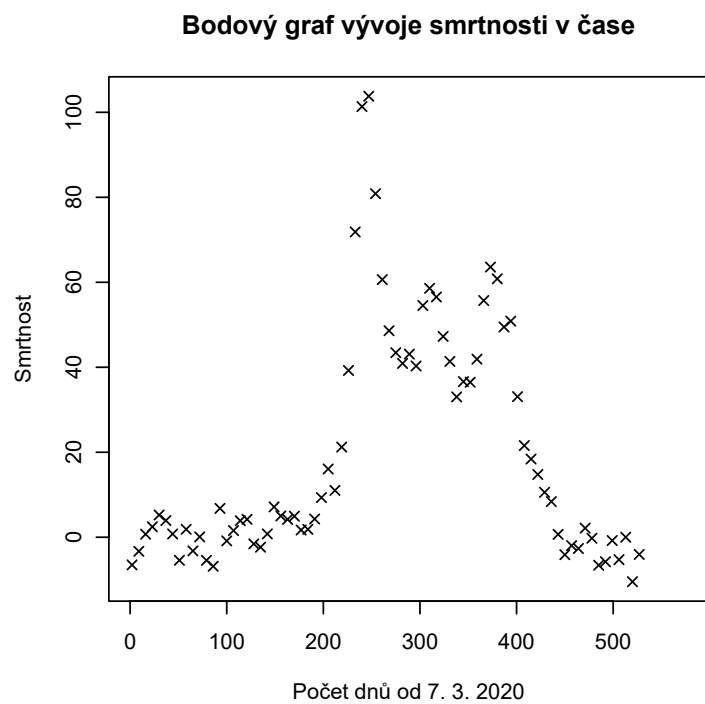
Obrázek 14: Bodový graf týdenních hospitalizací



Obrázek 15: Bodový graf pozitivity testů



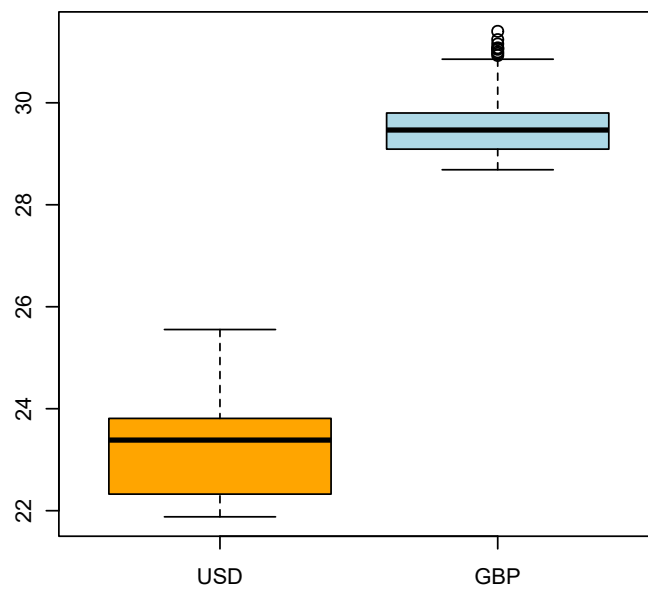
Obrázek 16: Bodový graf nových očkování



Obrázek 17: Bodový graf smrtnosti

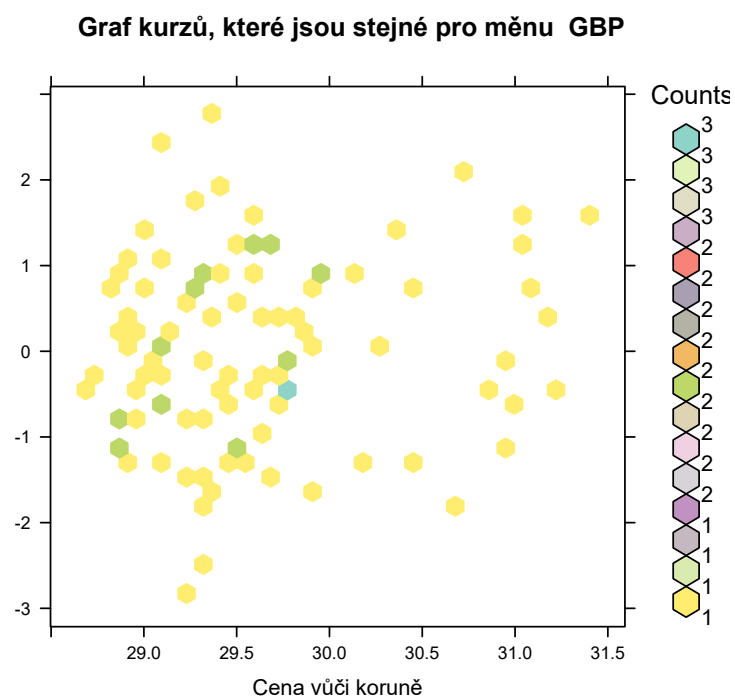
3.3 Boxplot

Boxplot pro srovnání kvartilů měn USD a GBP



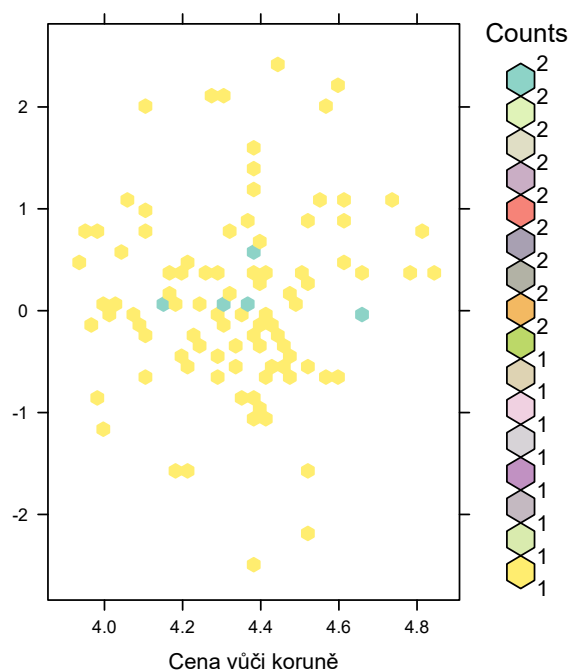
Obrázek 18: Boxplot graf pro první dvě měny

3.4 Hexbin



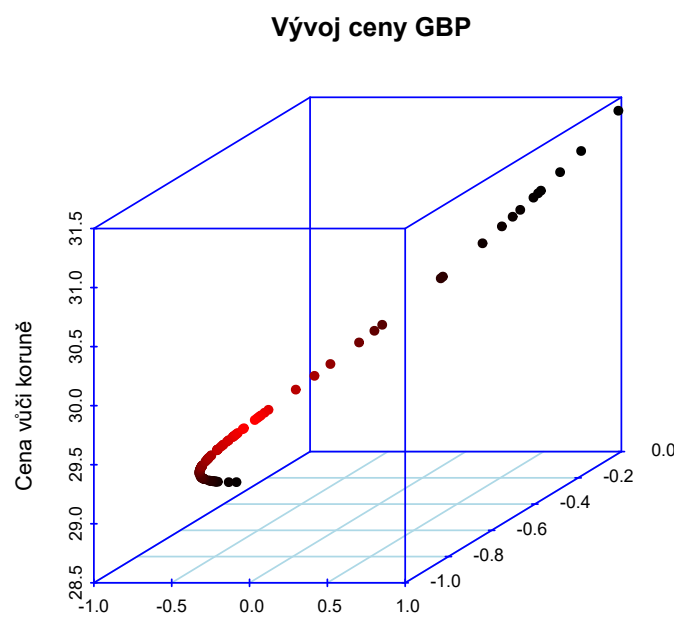
Obrázek 19: Hexbin graf měny s maximální hodnotou šikmosti

Graf kurzů, které jsou stejné pro měnu BRL

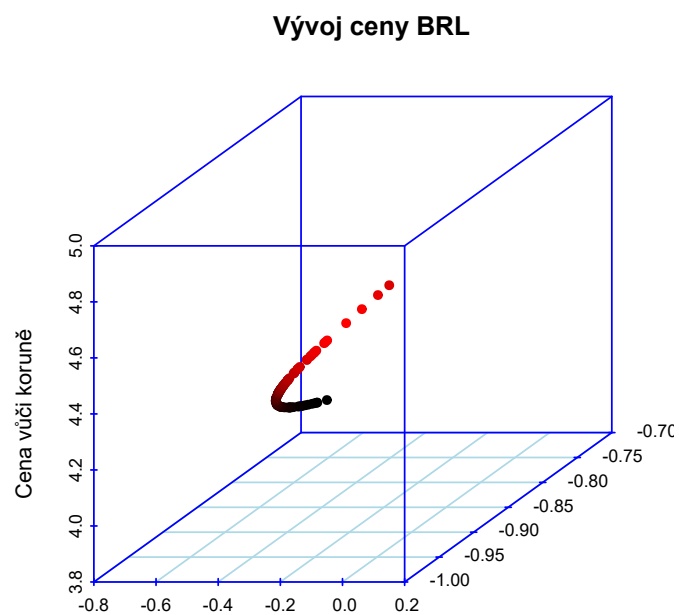


Obrázek 20: Hexbin graf měny s minimální hodnotou šikmosti

3.5 3D graf



Obrázek 21: 3D graf graf měny s maximální hodnotou šikmosti

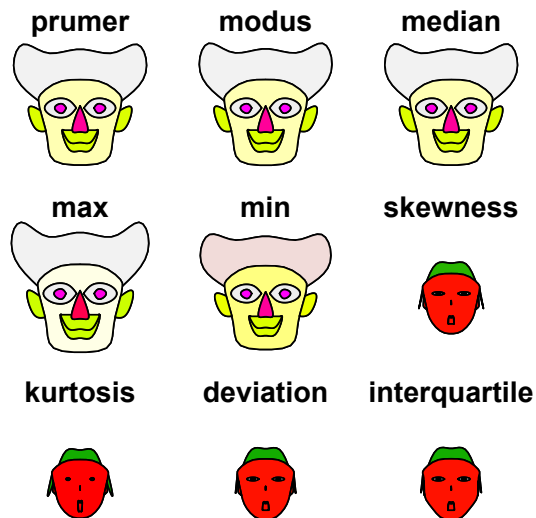


Obrázek 22: 3D graf graf měny s minimální hodnotou šikmosti

3.6 Chernoff faces

effect of variables:

modified item	Var
"height of face	" "USD"
"width of face	" "GBP"
"structure of face"	"CHF"
"height of mouth	" "EUR"
"width of mouth	" "BRL"
"smiling	" "ILS"
"height of eyes	" "CAD"
"width of eyes	" "MXN"
"height of hair	" "BGN"
"width of hair	" "RON"
"style of hair	" "USD"
"height of nose	" "GBP"
"width of nose	" "CHF"
"width of ear	" "EUR"
"height of ear	" "BRL"



Obrázek 23: Chernoff faces graf tabulky popisné statistiky

4 TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

4.1 Jednovýběrový Studentův test vůči střední hodnotě

One Sample t-test

```
data: TestHypCur1
t = 31.433, df = 100, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 20
95 percent confidence interval:
 23.08508 23.50076
sample estimates:
mean of x
 23.29292
```

4.2 Dvouvýběrový Studentův test

Welch Two Sample t-test

```
data: TestHypCur1 and TestHypCur2
t = -51.224, df = 166.54, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.549041 -6.062939
sample estimates:
mean of x mean of y
 23.29292 29.59891
```

4.3 Fisherův test

F test to compare two variances

```
data: TestHypCur1 and TestHypCur2
F = 2.6248, num df = 100, denom df = 100, p-value = 2.308e-06
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.769629 3.893242
sample estimates:
ratio of variances
 2.624803
```

5 ANOVA

6 KORELACE

6.1 Korelační matice

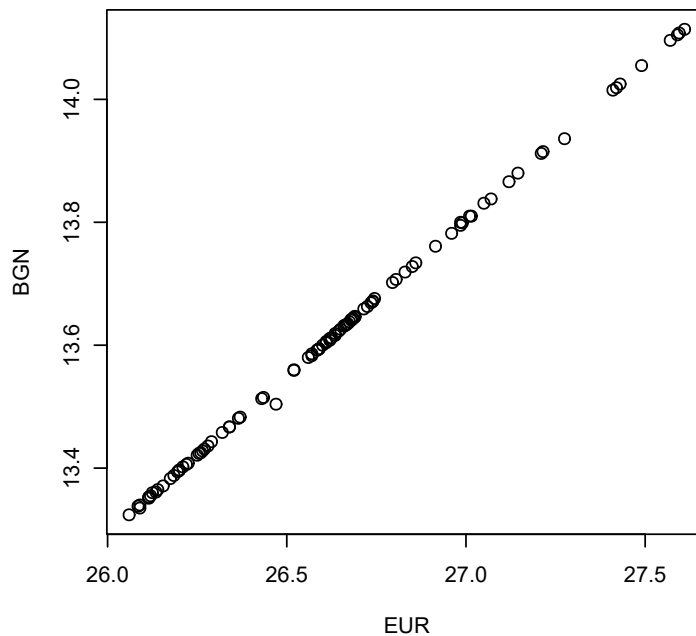
	USD	GBP	CHF	EUR	BRL	ILS	CAD
USD	1.0000000	0.9107230	0.9460180	0.9179658	0.6392782	0.9883873	0.9626825
GBP	0.9107230	1.0000000	0.8948676	0.8658716	0.4673841	0.9088316	0.8908392
CHF	0.9460180	0.8948676	1.0000000	0.9692083	0.4617732	0.9327882	0.9217226
EUR	0.9179658	0.8658716	0.9692083	1.0000000	0.5020093	0.9016465	0.9312468
BRL	0.6392782	0.4673841	0.4617732	0.5020093	1.0000000	0.6334782	0.6587888
ILS	0.9883873	0.9088316	0.9327882	0.9016465	0.6334782	1.0000000	0.9636070
CAD	0.9626825	0.8908392	0.9217226	0.9312468	0.6587888	0.9636070	1.0000000
MXN	0.6587680	0.5420454	0.5987799	0.7068223	0.7230531	0.6424965	0.7797695
BGN	0.9186576	0.8658093	0.9696949	0.9998681	0.5025878	0.9021541	0.9312349
RON	0.9480076	0.9012998	0.9807330	0.9928694	0.5247833	0.9318752	0.9408933
	MXN	BGN	RON				
USD	0.6587680	0.9186576	0.9480076				
GBP	0.5420454	0.8658093	0.9012998				
CHF	0.5987799	0.9696949	0.9807330				
EUR	0.7068223	0.9998681	0.9928694				
BRL	0.7230531	0.5025878	0.5247833				
ILS	0.6424965	0.9021541	0.9318752				
CAD	0.7797695	0.9312349	0.9408933				
MXN	1.0000000	0.7060044	0.6736579				
BGN	0.7060044	1.0000000	0.9931094				
RON	0.6736579	0.9931094	1.0000000				

Tabulka 2: Korelační matice

6.2 Bodové grafy hodnot z korelační matice

Hodnota korelace pro tento graf je: 0.999868090159611

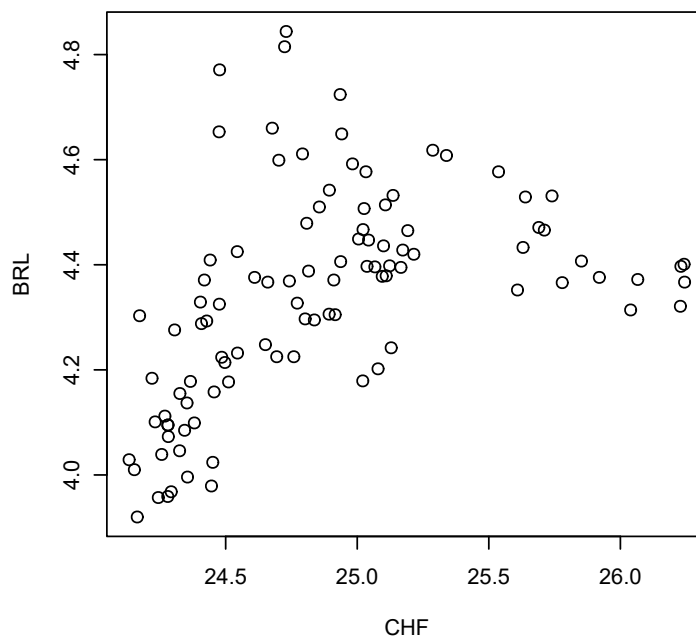
Graf závislosti ceny BGN na ceně EUR



Obrázek 24: Bodový graf maximální hodnoty v korelační matici

Hodnota korelace pro tento graf je: 0.461773203642033

Graf závislosti ceny BRL na ceně CHF



Obrázek 25: Bodový graf minimální hodnoty v korelační matici

7 TESTOVÁNÍ V KONTINGENČNÍCH TABULKÁCH

7.1 Chí kvadrát test

Chi-squared test for given probabilities

```
data: smoke.freq
```

```
X-squared = 24.076, df = 3, p-value = 2.408e-05
```

Pearson's Chi-squared test

```
data: kontTab
```

```
X-squared = 13.062, df = 6, p-value = 0.04206
```

Pearson's Chi-squared test

```
data: kontTabSpojena
```

```
X-squared = 9.7686, df = 3, p-value = 0.02064
```

8 LINEÁRNÍ REGRESE

Call:

```
lm(formula = DATAregrese$lRegreseCur2 ~ DATAregrese$lRegreseCur1)
```

Coefficients:

(Intercept)	DATAregrese\$lRegreseCur1
0.007772	0.510974

Call:

```
lm(formula = DATAregrese$lRegreseCur2 ~ DATAregrese$lRegreseCur1)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.0292594	-0.0001743	0.0003621	0.0011365	0.0035889

Coefficients:

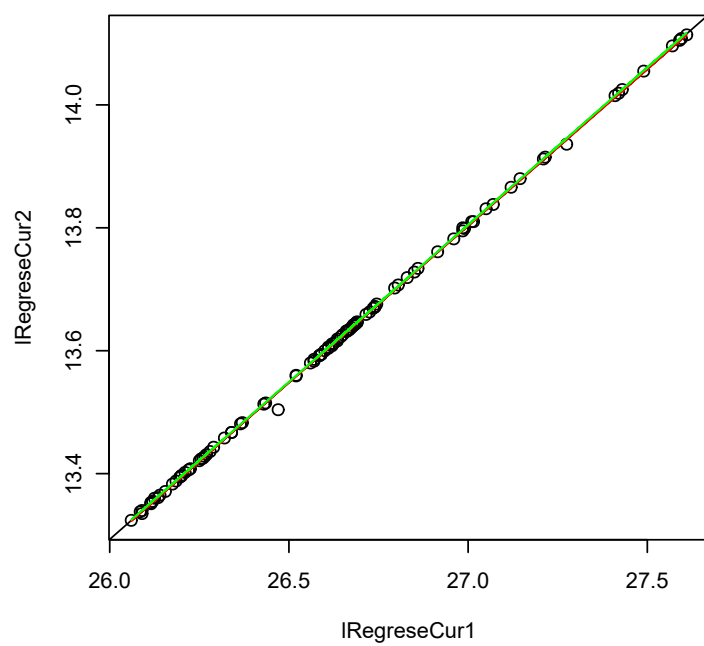
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0077722	0.0222232	0.35	0.727
DATAregrese\$lRegreseCur1	0.5109742	0.0008342	612.52	<2e-16 ***

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

Residual standard error: 0.003296 on 99 degrees of freedom

Multiple R-squared: 0.9997, Adjusted R-squared: 0.9997

F-statistic: 3.752e+05 on 1 and 99 DF, p-value: < 2.2e-16



Obrázek 26: Graf lineární regrese

ZÁVĚR

POUŽITÁ LITERATURA

- [1] Our World in Data *Data on COVID-19 (coronavirus)* [online]. 2021 [cit. 2021-11-18]. Dostupné z: <https://github.com/owid/covid-19-data/tree/master/public/data>

SEZNAM PŘÍLOH

Příloha A	??
-----------------	----

PŘÍLOHA A

Příloha A zahrnuje ZIP soubor, který obsahuje: