

Jak jsme se učili predikovat výsledky voleb...

Marek Omelka, Ondřej Týbl

Miku(k)lášský den ČStS,
7. 12. 2021

Zadání

Český statistický úřad zveřejňuje průběžné výsledky sčítání.

Je známo, že průběžné výsledky jsou vychýlené.

iROZHLAS



DOMOV SVĚT EKONOMIKA SPORT KULTURA VĚDA KOMENTÁŘE ŽIVOTNÍ STYL VOLBY POČASÍ VINOHRADSKÁ 12

PARLAMENTNÍ VOLBY 2021

VÝSLEDKY

PREDIKCE

KALKULÁTOR KOALIC

SLOŽENÍ SNĚMOVNY

OBCE A KRAJE

POROVNÁVAČ

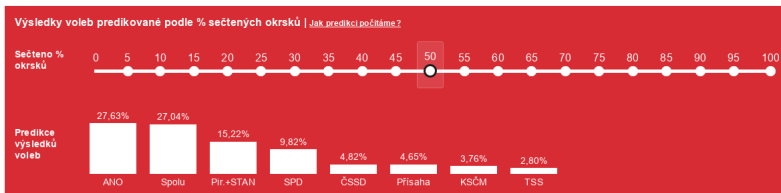
VOLBY NA WEB

STRANY V KRAJÍCH

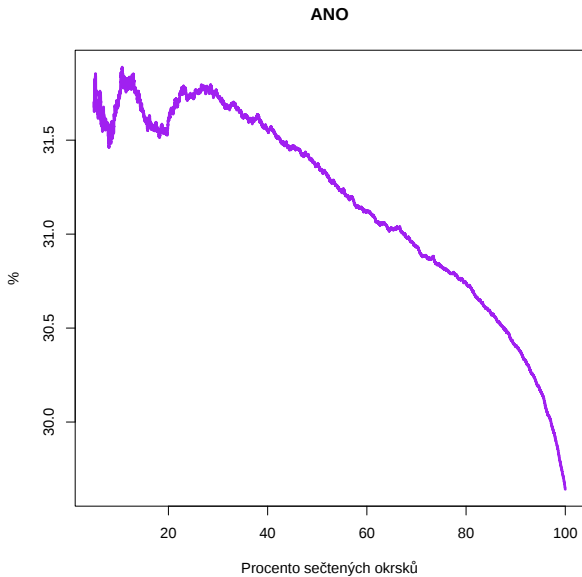
ZAHRANIČÍ

Predikce výsledku voleb

Odhad výsledku voleb na základě průběžných dat, jak ji počítám a její tvůrce najdete níže.

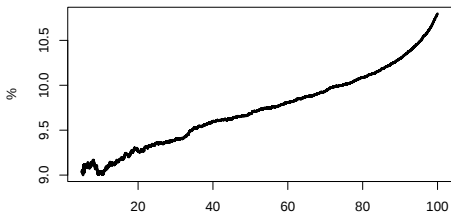


Volby 2017



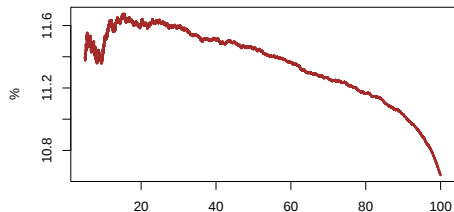
Volby 2017 - Piráti, SPD, ODS

Piráti



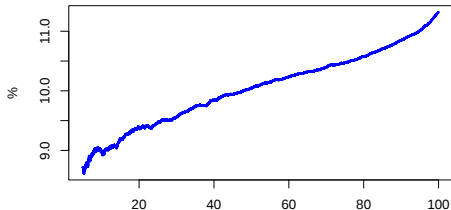
Procento sečtených okrsků

SPD



Procento sečtených okrsků

ODS



Procento sečtených okrsků

Naivní predikce (první nápad)

$$\hat{K}_t^{2021} = A_t^{2021} \frac{K^{2017}}{A_t^{2017}},$$

$\hat{K}_t^{2021}$... odhad konečného procentuálního zisku v roce 2021 při sečtení t % okrsků

K^{2017} ... konečný procentuální zisk v roce 2017

A_t^{2021} ... aktuální procentuální zisk v roce 2021

A_t^{2017} ... aktuální procentuální zisk v roce 2017

Příklad

Sečteno okrsků:	5 %	25 %	50 %	75 %
$\frac{K^{2017}}{A_t^{2017}}$ pro ANO	0,899	0,902	0,920	0,944

Stratifikace pomocí krajů

$$\hat{K}_t^{2021} = \sum_{j=1}^{14} w_{j,t} A_{j,t}^{2021}$$

$A_{j,t}^{2021}$... aktuální procentuální zisk v **j -tém kraji** v roce 2021

$w_{j,t}$... váha j -tého kraje, která odhaduje $\frac{\text{podíl platných hlasů v } j\text{-tém kraji}}{\text{podíl platných hlasů v ČR}}$

Model pro data z voleb 2017

$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iK})^\top \sim \text{Mult}_K(n_i, \mathbf{p}_i)$, kde $\mathbf{p} = (p_{i1}, \dots, p_{iK})$

n_i počet platných hlasů v i -tém okrsku,

p_{ik} pravděpodobnost, že volič v i -tém okrsku bude volit k -tou stranu.

Model směsí (mixture model):

$\mathbf{Y}_i \sim v_1 \text{Mult}_K(n_i, \mathbf{p}^{(1)}) + v_2 \text{Mult}_K(n_i, \mathbf{p}^{(2)}) \dots + v_G \text{Mult}_K(n_i, \mathbf{p}^{(G)}),$

$v_j \geq 0$, $\sum_{j=1}^G v_j = 1$ a G je předem zvoleno.

Obecná stratifikace

$$\widehat{K}_t^{2021} = \sum_{g=1}^G w_{g,t} A_t^{g,2021}.$$

$A_t^{g,2021}$... aktuální procentuální zisk **v g -tém stratu** v roce 2021.

$w_{g,t}$... váha g -tého strata, která odhaduje $\frac{\text{podíl platných hlasů v } g\text{-tém stratu}}{\text{podíl platných hlasů v ČR}}$.

Přehled metod

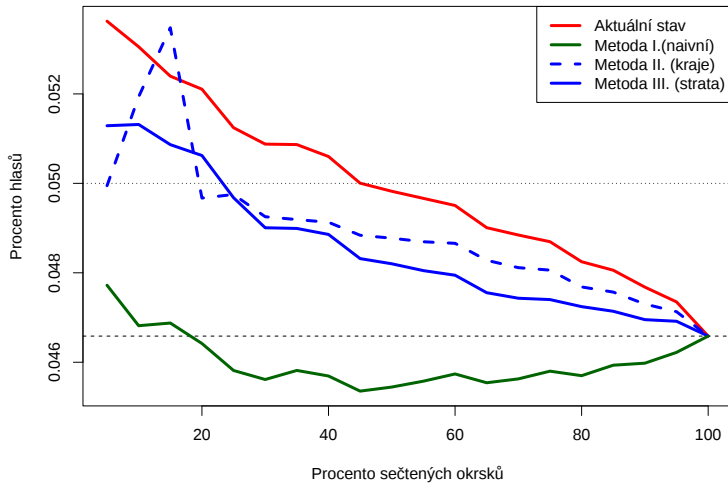
Metoda I. - naivní metoda, tj. $\hat{K}_t^{2021} = A_t^{2021} \frac{K^{2017}}{A_t^{2017}}$.

Metoda II. - stratifikace pomocí krajů.

Metoda III. - stratifikace pomocí strat vytvořených shlukovací analýzou (na základě dat z roku 2017).

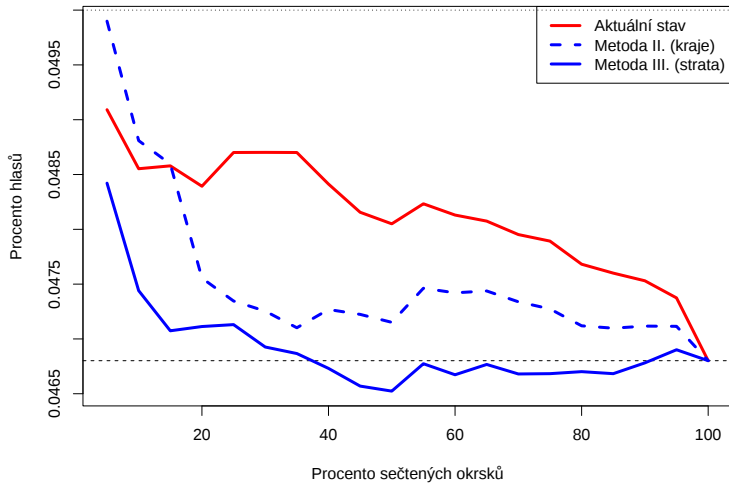
Výsledky 2021

ČSSD



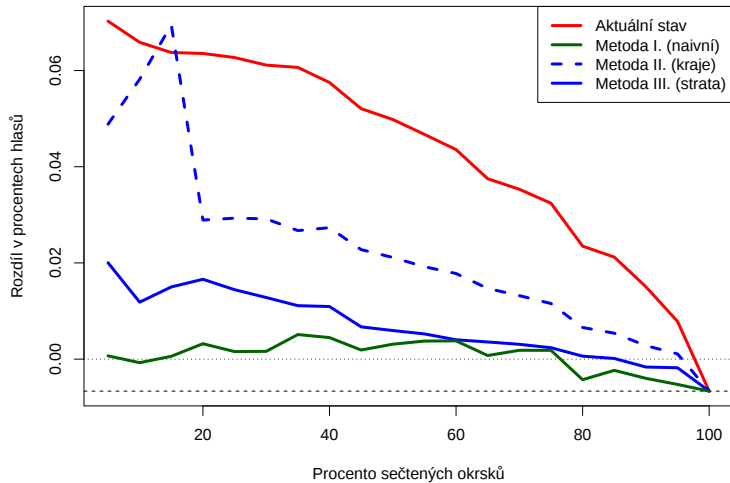
Výsledky 2021

Přísaha



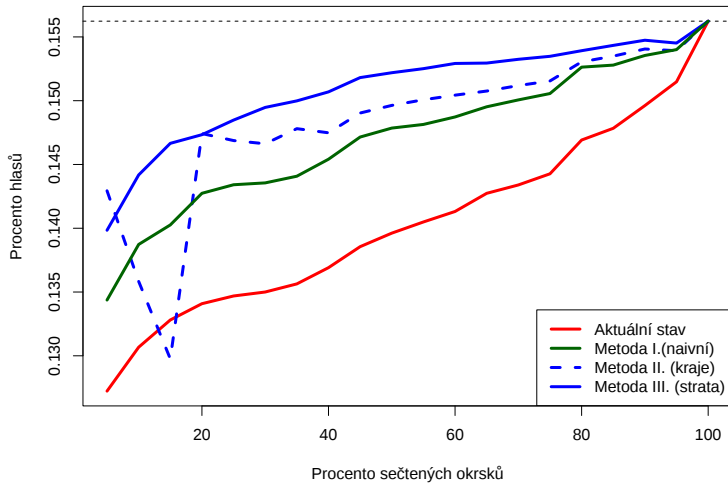
Výsledky 2021

ANO minus SPOLU



Výsledky 2021

PirStan



Děkuji za pozornost.

