

Jak (ne)vyšly předpovědi výsledků parlamentních voleb v ČR 2013

Mirko Navara

1. 11. 2013

1 Vstupní data

http://zpravy.idnes.cz/jak-se-trefily-pruzkumy-0wa-/domaci.aspx?c=A131027_120534_domaci_jw

Rozsah	Volby	Factum	Median	STEM	CVVM	Sanep	Médea	Aisa
ČSSD	20.5	22.8	25.5	25.9	26	23.8	22.2	23
ANO	18.7	12.1	13	16.1	16.5	11.6	16.9	16
KSČM	14.9	17.1	16	13.3	18	16.9	11.8	14
TOP 09	12	13.2	13	11.5	9	11.9	9.6	10.5
ODS	7.7	7.2	8	8.6	6.5	7.5	5.5	7
Úsvit	6.9	3.7	4	5.9	5	5.3	8.2	6
KDU	6.8	5.9	6	4.5	5	5.7	6.2	6
SZ	3.2	4.3	3	2.6	2	3.1	2.9	3
SPOZ	1.5	4.7	5	2.6	3.5	5.2	3.7	4
ostatní	7.8	9	6.5	9	8.5	9	13	10.5
Účast	59.5	54	60	67	63	59.3	71	78
χ^2 -test		≈ 0	≈ 0	$5 \cdot 10^{-6}$	$2 \cdot 10^{-12}$	≈ 0	≈ 0	$3 \cdot 10^{-10}$

(Rozsah výběru je pouze přibližný, agentury neuvedly přesná čísla. Nicméně to nemá velký vliv na výsledek.)

Tučně jsou vyznačeny předpovědi, které se vejdou do 95% symetrického intervalového odhadu.

To, že některé hodnoty vycházejí s chybou do 0.2, je náhoda, protože tolerance odhadů je řádově v procentech (v desetínách procenta u Sanepu, který měl větší rozsah výběru).

Na první pohled je vidět, že např. výsledky první a poslední strany v přehledu jsou všechny nadhodnocené.

Rozdělení jsou porovnána pomocí χ^2 -testu *dobré shody*, který ukazuje, že u nejúspěšnější agentury lze hypotézu o shodě rozdělení zamítnout na hladině významnosti $5 \cdot 10^{-6}$, tj. takový výsledek bychom při shodě rozdělení dostali s pravděpodobností 1/200 000, u ostatních agentur ještě menší.

(Detailní zpracování zůstává jako možné téma pro zápočtové práce.)

2 Metaanalýza

Při *metaanalýze* zpracujeme předpovědi jednotlivých agentur jako nezávislá měření stejné náhodné veličiny.

Odhadneme střední hodnotu (realizací výběrového průměru) a směrodatnou odchylku (realizací výběrové směrodatné odchylky).

Z nich stanovíme symetrický 95% *intervalový odhad*. (Vzhledem k malému rozsahu výběru je nutno použít kvantily Studentova rozdělení, nikoli normálního.)

Jak je vidět z tabulky, tyto odhady vyšly, s výjimkou SPOZ.

Důvodem však je, že odhady založené na malém počtu měření (agentur) dávají velmi široký interval, a tím poměrně malou informaci. Jsou velmi opatrné a nepřekvapivé.

	Volby	průměr	směr. odch.	dolní mez	horní mez
ČSSD	20.5	24.2	1.6	20.2	28.1
ANO	18.7	14.6	2.3	9.0	20.2
KSČM	14.9	15.3	2.3	9.7	20.9
TOP 09	12	11.2	1.6	7.3	15.2
ODS	7.7	7.2	1.0	4.7	9.7
Úsvit	6.9	5.4	1.5	1.8	9.1
KDU	6.8	5.6	0.6	4.1	7.1
SZ	3.2	3.0	0.7	1.3	4.7
SPOZ	1.5	4.1	0.9	1.8	6.4
ostatní	7.8	9.4	2.0	4.5	14.2
Účast	59.5	64.6	8.1	48.8	80.4

3 Proč volební předpovědi nevycházejí?

1. Lidé nejsou losy, své názory mění a nesdělují je vždy správně.
2. Výsledky jsou ovlivněny volební neúčastí, kterou je obtížné předpovídat (závisí např. na počasí).

3. Průzkumy jsou zdlouhavé, tudíž provedené v delším časovém období před volbami.
4. Agentury se snaží o *kvótní výběr*, v němž jsou jednotlivé skupiny obyvatelstva zastoupeny stejně jako v populaci. Otázka je, jaké skupiny mají být rozlišovány (co je podstatné).
5. Kvótní výběr se zcela nepodaří (výběrový soubor nemá odpovídající složení), což se kompenzuje váhováním údajů; tím však vzroste rozptyl.
6. Některé agentury nezveřejňují přímo výsledky průzkumu, ale odhady založené i na výsledcích předchozích voleb a politologických úvahách.
7. **Celý model je špatně, protože měření ovlivňuje zkoumaný systém (stejně jako v kvantové fyzice); seriózní zpracování by tedy mělo používat jiný aparát, který je však složitější. Více se lze dovědět v předmětu Fuzzy logika (doktorandský, resp. volitelný), viz <http://cmp.felk.cvut.cz/~navara/fl/>.**