

Bystřina 2021

Jan Hučín

prosinec 2022

1 Obsah

2	Úvod	2
3	Co je to Bystřina	2
3.1	O soutěži	2
3.2	S jakými daty pracujeme	2
4	Účastníci	3
4.1	Celkově	3
4.2	Podobní účastníci	4
4.3	Po kolech	5
5	Hádanky	6
5.1	Celkově	6
5.2	Podle druhů	7
5.3	Podle autorů	9
5.4	Podle délky	10
5.5	Autor, druh, nebo délka?	10
6	Stálost výsledků účastníků v čase	13
7	Existuje specifický talent?	15
8	Předpověď výsledku 13. kola	16
9	Závěr	18
10	Kontakt	18

2 Úvod

Statistika nuda je, ale tato analýza se pokusí být výjimkou. Dozvíte se tu několik zajímavostí z výsledků 72. ročníku soutěže Bystřina (2021), například:

- Byly čtyřřádkové hádanky těžší než dvouřádkové?
- Které druhové operace dělaly řešitelům největší problémy (byly to obměnky a záhadné nápisy, jak by se dalo čekat)? A které byly naopak nejlehčí (tipovali bychom sloučenky a čárkovky)?
- Který autor byl specialista na lavinovky a vlnovky? Kdo byl hlavní tvůrce záhadných nápisů a kdo byl spíš na vsuvky a podobnou klasiku? A dá se říct, že některý autor byl na řešitele nadprůměrně tvrdý, nebo naopak hodný?
- Které hádanky nejvýrazněji určovaly pořadí na předních místech?
- Dá se z výsledku účastníka ve 12 kolech předpovědět počet bodů ve 13. kole? A jak přesná ta předpověď je?

3 Co je to Bystřina

3.1 O soutěži

Ačkoliv je to zvláštní, naprostá většina lidí v naší zemi Bystřinu, časopis KaH ani hádankářsko-křížovkářský svaz nezná. Tento odstavec je hlavně pro ně.

Bystřina je celoroční korespondenční soutěž v řešení **českých slovních hádanek** (laicky můžeme říct “jazykových hlavolamů”). Má 13 kol, zveřejňuje se v časopise **Křížovka a hádanka** jednou za 4 týdny. Je možné, i když celkově nevýhodné se účastnit jen některých kol. V roce 2021 proběhl už 72. ročník, má tedy tradici přibližně stejně dlouhou jako například festival Pražské jaro. Účastní se ho česká špička řešitelů hádanek, ale i příležitostní luštitelé, obsahuje proto jak velmi těžké, tak lehké hádanky.

Za správné řešení hádanky dostane účastník bod, jinak nedostává nic – za neúplné, nesprávné nebo neautorské řešení je nula bodů, jak je uvedeno v propozicích soutěže, které jsou vyhlášeny vždy před zahájením ročníku. Celkové pořadí je dáno součtem bodů, při rovnosti rozhodují pomocná kritéria, která zde pomineme.

3.2 S jakými daty pracujeme

Vedoucí soutěže laskavě poskytl několik datových souborů:

- výsledky účastníků v jednotlivých hádankách po jednotlivých kolech, případně informaci, že se účastník kola nezúčastnil;
- jména autorů a čísla hádanek každého z nich;
- texty hádanek za všechna kola včetně jejich *druhové operace* (pro zjednodušení budeme používat i pojem *druh*) a řešení.

4 Účastníci

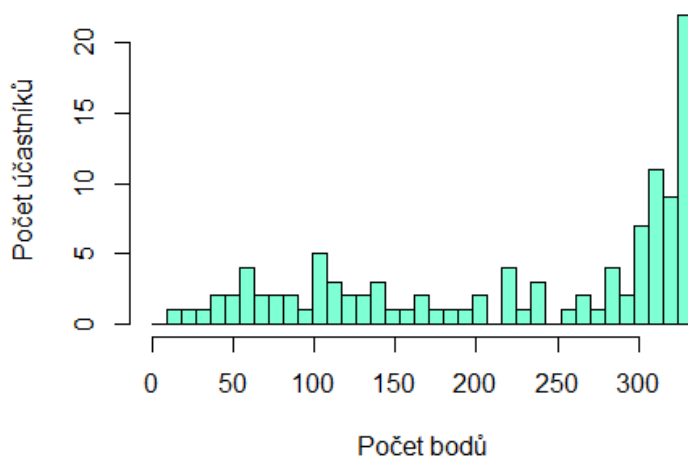
4.1 Celkově

Do soutěže se zapojilo **109 účastníků**, skoro všichni (107) už od 1. kola, zbylí se přidali v průběhu soutěže. Žádné kolo nevynechalo 98 účastníků, tedy naprostá většina startovního pole – budeme je nazývat *kompletní účastníci*.

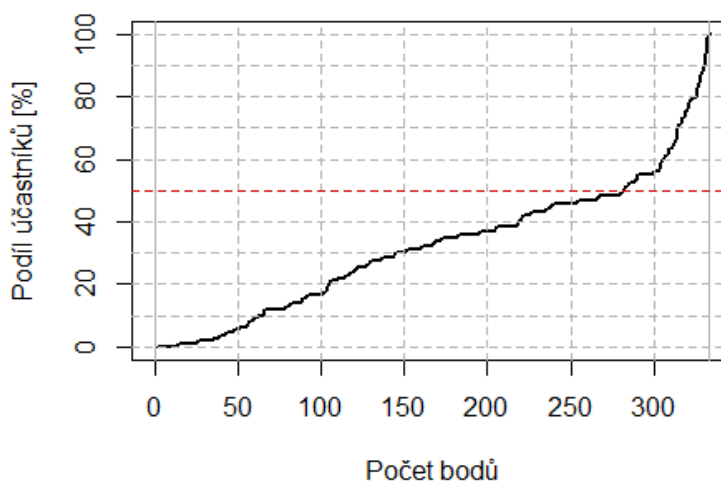
Nejvyšší účast nemělo 1. kolo, ale překvapivě 3. kolo (108 zaslaných řešení), nejmenší účast pak 12. kolo (103 zaslaných řešení).

Celkem bylo v soutěži 333 hádanek a tolik tedy byl i maximální možný počet bodů. Na ten sice nakonec nedosáhl nikdo, ale značná část účastníků se tomu přiblížila. Následující grafy znázorňují rozdělení účastníků podle celkového počtu bodů.

**Rozdělení účastníků
podle celkového počtu bodů**



Počet bodů a % předstižených



V prvním grafu (tzv. histogramu) odpovídá každý sloupeček rozpětí 9 bodů, aby se celé rozpětí možných výsledků pokrylo rovnoměrně. Je vidět, že **startovní pole se skládá ze dvou částí**: o přední místa bojoval velký počet účastníků (48 z nich získalo 300 a víc bodů); za nimi je odstup a pak roztroušené skupinky a jednotlivci prakticky po celém rozpětí až k nule (nejnižší počet bodů účastníka hodnoceného na konci soutěže je 27, ale v grafu jsou i ti, kdo odpadli během soutěže s menším počtem bodů na kontě).

Ve druhém grafu si můžeme prohlédnout, kolik bodů bylo potřeba k předstihu jakého podílu ostatních účastníků. Například **abychom předstihli polovinu (50 %) účastníků, potřebovali jsme získat okolo 275 bodů** (místo, kde se protíná červená vodorovná čára s černou křivkou). Pro umístění v nejlepší pětině (předstihu aspoň 80 % ostatních) bylo potřeba okolo 320 bodů. Naopak hranice pro nejhůřší pětinu byla těsně nad 100 body.

Následující tabulka obsahuje průměrný počet bodů a tzv. medián (skóre prostředního účastníka v pořadí) jednak pro všechny účastníky, jednak pro kompletní účastníky (s účastí ve všech 13 kolech). Pod nimi je průměrná úspěšnost v řešení hádanek v obou skupinách. Průměrná úspěšnost se každému počítá pouze za kola, v nichž zaslal řešení.

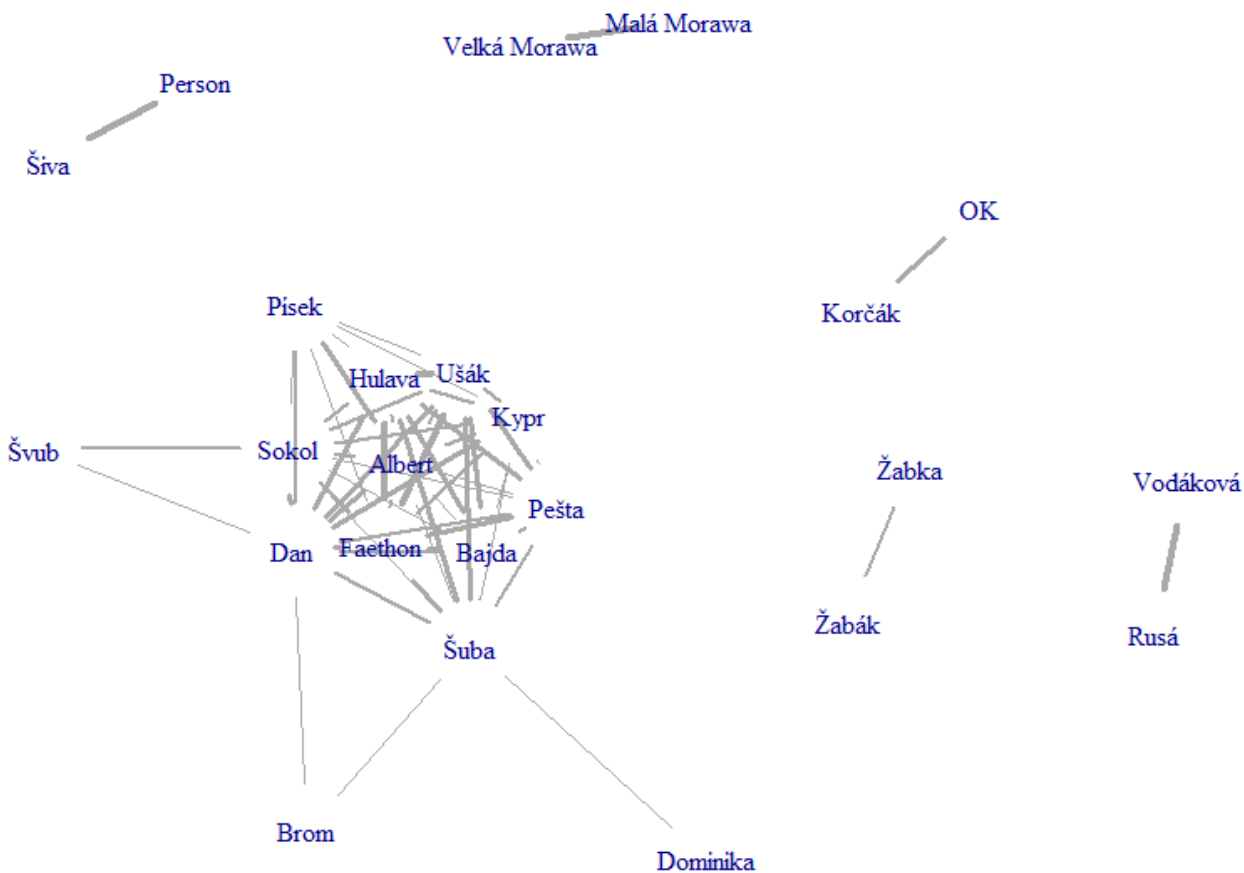
	Všichni	Kompletní
Průměr bodů	226,2	239,1
Medián bodů	281,0	294,0
Prům. úspěšnost (v řešených hádankách) [%]	69,7	71,8

4.2 Podobní účastníci

Ač je Bystřina soutěž jednotlivců, bývá běžné, že se mezi sebou někteří účastníci radí, zejména u obtížnějších hádanek. Z dat lze velmi snadno zjistit, které dvojice účastníků mají vysokou shodu v řešení jednotlivých hádanek. Extrémně úspěšní řešitelé se mohou shodovat přirozeně (např. teoreticky se všichni stoprocentně úspěšní řešitelé navzájem dokonale shodují). U méně úspěšných již vzájemná shoda indikuje, že si poskytují vzájemnou podporu.

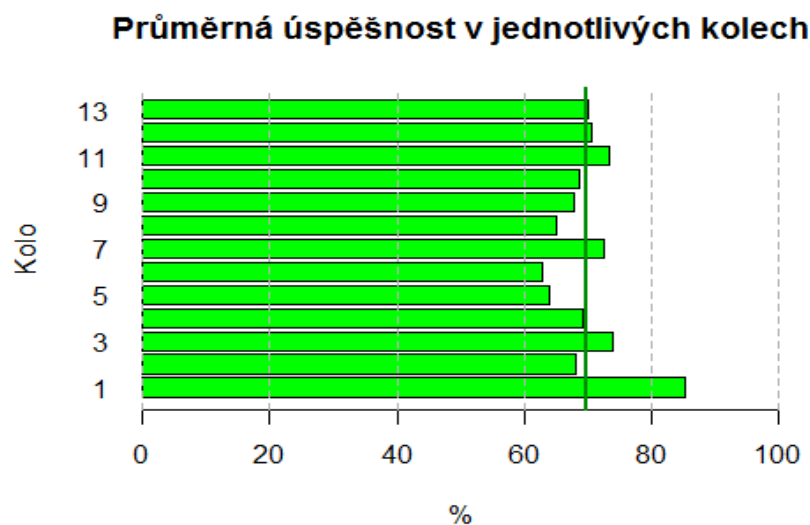
Podíváme se tedy na dvojice kompletních účastníků, kteří mají v úspěšnosti jednotlivých hádanek velmi vysokou podobnost – tedy téměř každou hádanku vyřešili oba zároveň správně, nebo oba zároveň nesprávně. V grafu jsou případy takových velmi vysokých podobností (max. tři rozdíly, tj. nejmeně 330 shod z 333 hádanek).

Poznámka: Pro jakoukoliv dvojici kompletních účastníků je možné statisticky určit, zda můžeme shodu jejich úspěšností v jednotlivých hádankách považovat ještě za náhodnou, nebo je již významně nenáhodná. Už by to ovšem přesáhlo rozumný rozsah studie, proto si to necháme na některou příští příležitost.



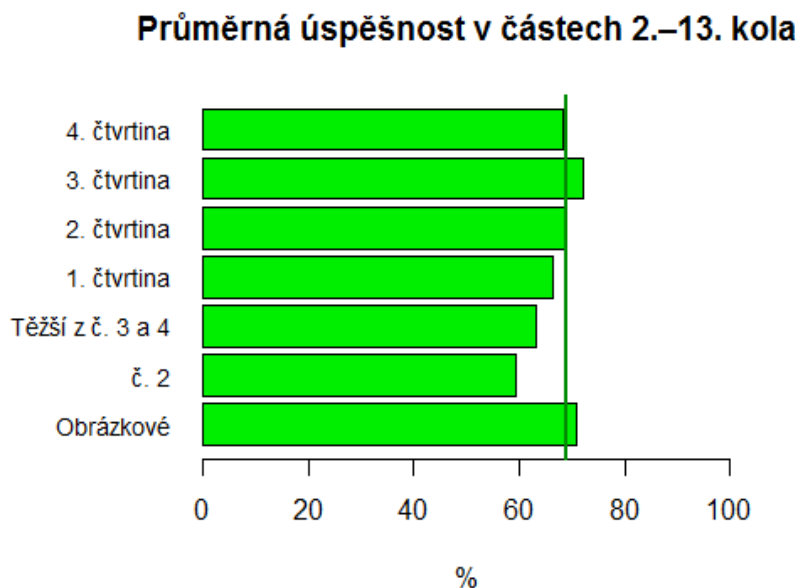
4.3 Po kolech

Byla jednotlivá kola stejně obtížná? Pravidelní účastníci podobných soutěží vědí, že 1. kolo bývá zpravidla nejlehčí, u ostatních kol obtížnost kolísá, někdy bývá poslední kolo nejtěžší. I v Bystřině 2021 byly mezi koly patrné rozdíly v obtížnosti, jak ukazuje následující graf.



Princip “náborového” lehkého 1. kola platil, ale **nejtěžší kola byla uprostřed soutěže**, úspěšnost posledních dvou kol odpovídala celkovému průměru (ten je vyznačený svislou zelenou čarou).

Každé kolo Bystřiny má určité pevné body: v první a poslední hádance je použit kreslený námět a druhá hádanka je na aktuální téma. Podívejme se, jestli jsou některé části kol (od 2. do 13. kola) pravidelně lehčí nebo těžší než celkový průměr.



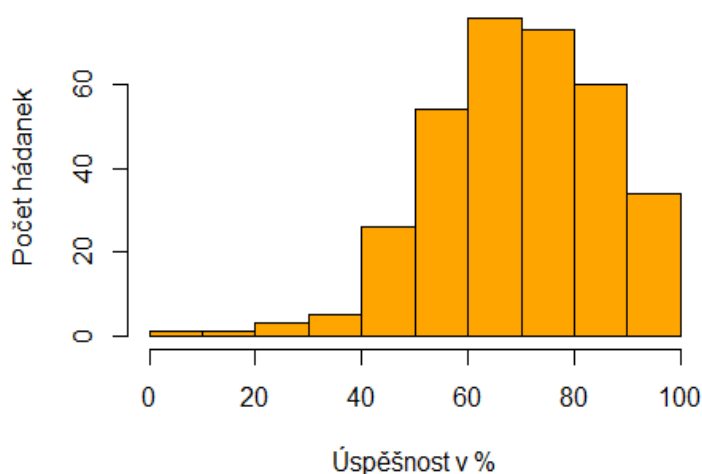
Rozdíly v obtížnosti mezi jednotlivými částmi kol existují. V průměru **nejtěžší byly pro řešitele hádanky první čtvrtiny kola**, na čemž měly velký podíl hádanky na pozicích 2 až 4. Naopak obrázkové hádanky byly celkově o něco lehčí než průměr. Nejlehčí hádanky čekaly ve třetí čtvrtině (což řada luštitelů intuitivně vycítila a řešila kola Bystřiny zpravidla odzadu).

5 Hádanky

5.1 Celkově

Bystřina, ač určená spíše zkušenějším a lepším luštitelům, obsahuje v každém kole i několik velmi snadných hádanek pro širší řešitelskou obec. Následující graf ukazuje, že v ročníku 2021 se vyskytovaly hádanky všech stupňů obtížnosti, jak téměř stoprocentně řešené, tak ty, které vyřešila méně než pětina startovního pole.

Rozdělení hádanek podle prům. úspěšnosti



5.2 Podle druhů

Druhů hádanek existuje téměř sto a v Bystřině se tato pestrost projevila. I když nedošlo na všechny druhy, celkem se v Bystřině 2021 objevilo 51 různých druhů hádanek. V následující tabulce jsou uvedeny všechny druhy, které byly zastoupeny v ročníku aspoň sedmi hádankami.

Druh	Počet	Druh	Počet
ZÁMĚNKA	32	CHYBOVKA	10
ZÁHADNÝ NÁPIS	26	PŘÍDAVKA	9
VSUVKA	19	OBNAŽKA	8
VÝMĚNKA	18	SLOUČENKA	8
VÝPUSTKA	16	RÁČEK	8
ODSUVKA	14	VÝLUČKA	8
OBMĚNKA	14	HÁČKOVKA	8
PŘESUNKA	13	LAVINOVKA	7
OBLOŽKA	12	CITOSLOVKA	7
PŘÍSUVKA	11	VOKÁLOVKA	7
PÁDOVKA	10		

Druhy hádanek, které se v Bystřině vyskytly málo, přihodíme k podobným druhům – můžeme například spojit přídavky a výlučky, lavinovky s pyramidkami a vlnovkami apod. Vytvoříme tak 23 kategorií podle “řešitelské” podobnosti a podíváme se, které druhy či “rodiny” druhů (kategorie) byly v Bystřině pro účastníky celkově nejnáročnější a které nejjednodušší.

Poznámka: kategorie jsme mohli samozřejmě utvořit i jinak, mimo jiné i podle zařazení ve Směrnících. My jsme dali přednost hledisku podobnosti uvažování při řešení.

Kategorie	Druhy	Počet	Úspěšnost [%]
LAV/PYR/VLN	LAVINOVKA, VLNOVKA, PYRAMIDKA	13	47,4
PŘÍD/VÝL	PŘÍDAVKA, VÝLUČKA	17	60,5
ZN	ZÁHADNÝ NÁPIS	26	61,3
OBMĚNKA	OBMĚNKA	14	63,6
PRON/POLOV/VYB/VVSL	VÝPUSTKA SLOVNÍ, PRONIKAČKA, POLOVSUVKA, VSUVKA SLOVNÍ, VYBÍRAČKA	12	63,9
ZÁM	ZÁMĚNKA	32	64,5
VOKÁLOVKA	VOKÁLOVKA	7	65,7
PŘESUNKA+SLAB	PŘESUNKA, PŘESUNKA SLABIKOVÁ	14	66,4
VÝMĚNKA+SLAB	VÝMĚNKA, VÝMĚNKA SLABIKOVÁ	19	69,5
PŘESM/PERM	PŘESMYČKA, PERMUTOVKA	9	70,7
RÁČEK/OBRAT	RÁČEK, OBRATKA	12	71,6
OBNAŽKA	OBNAŽKA	8	71,6
POSUNKA	POSUNKA	5	72,7
PÁD/ČAS	ČASOVAČKA, PÁDOVKA	12	72,7
PŘÍSUVKA	PŘÍSUVKA	11	72,8
OSTATNÍ	HLÁSKOVKA, ČTVERCOVKA SLABIKOVÁ, ČÍSELKA, KLÍČOVKA, MNOŽENKA, ZDVOJENKA	9	73,6
HÁČ/MĚK/YPS	YPSILONOVKA, MĚKČENKA, HÁČKOVKA	14	73,8
ODSUVKA	ODSUVKA	14	74,6
VS/VÝP	VÝPUSTKA, VSUVKA	35	75,7
POMVÝR	PRVKOVKA, PRAVĚKOVKA, OPĚTOVAČKA, CITOSLOVKA, ZÁPORKA	12	78,2
CHYBOVKA	CHYBOVKA	10	79,0
OBLOŽKA	OBLOŽKA	12	80,2
SSSSČ	SLOUČENKA, STŘÍDAVKA, SKRÝVAČKA, SPOJENKA, ČÁRKOVKA	16	85,1

Možná někoho překvapí, že *záhadný nápis* a *obměnka*, řešiteli poněkud neoblíbené druhy, nebyly nejtěžší. Těmi se staly s velkým náskokem *lavinovky*, *pyramidky* a *vlnovky*, na druhém místě pak *výlučky* a *přídavky*. Naopak se dalo čekat, že nejlépe půjdou řešitelům hádanky s minimálními změnami mezi náznaky a jádrem, tedy rodina *sloučenek*, *střídavek*, *skrývaček*, *spojenek* a *čárkovek*.

5.3 Podle autorů

Někoho může napadnout, že obtížnost hádanky je daná nejen jejím druhem, ale i osobou autora. Kdo chce řešitele potrápit, dokáže to i sloučenkou, kdo chce být hodný, vymyslí snadnou lavinovku. Dá se říct, že někteří autoři byli nadmíru drsní a někteří naopak vlídní?

Autorsky se na ročníku Bystřiny podílelo 31 autorů, někteří přispěli třeba jen jedinou hádankou, někteří měli v každém kole dvě či více. Následující tabulka uvádí autory s aspoň 10 hádankami, u každého pak nejčastější druhy a průměrnou úspěšnost řešitelů na všech hádankách příslušného autora.

Autor	Počet hádanek	Nejčastější druhy	Úspěšnost [%]
Rias	22	LAV/PYR/VLN 8, PRON/POLOV/VYB/VVSL 4, PŘESUNKA+SLAB 3	52,7
Vokatý	23	VÝMĚNKA+SLAB 5, PŘÍD/VÝL 4, PŘESUNKA+SLAB 3	57,5
Bohuš	18	HÁČ/MĚK/YPS 3, VS/VÝP 3, ZN 3	59,7
Šimek	13	ZÁM 8, OBMĚNKA 2, OBNAŽKA 1	61,3
Kožušník	10	POMVÝR 2, VS/VÝP 2, ZN 2	68,3
Roemer	13	VS/VÝP 3, VÝMĚNKA+SLAB 3, ZÁM 2	69,2
Zahajský	13	LAV/PYR/VLN 2, OBNAŽKA 1, ODSUVKA 1	69,4
Konečný	14	POMVÝR 3, HÁČ/MĚK/YPS 2, SSSSČ 2	72,9
Lančarič	27	OBLOŽKA 6, VS/VÝP 4, CHYBOVKA 2	73,6
Říha	30	ZN 7, PÁD/ČAS 3, PŘESM/PERM 3	74,0
Kokeš	34	CHYBOVKA 6, PŘESM/PERM 4, VS/VÝP 4	74,8
Kousal	27	VS/VÝP 6, SSSSČ 5, ZÁM 4	75,5
Ambrož	11	PÁD/ČAS 3, ZÁM 2, OBLOŽKA 1	76,4
Salavec	17	ODSUVKA 3, OBMĚNKA 2, POMVÝR 2	79,8

Přítel **Rias** přispěl osmi hádankami do rodiny nejnáročnějších druhů (lavinovky s pyramidkami a vlnovkami), což přispělo k tomu, že se stal nejdrsnějším autorem – jeho taje zvládli účastníci v průměru jen s poloviční úspěšností. Hlavním dodavatelem výluček a přidavek byl přítel **Vokatý** s druhým místem mezi nejnáročnějšími autory. Že lze tvořit jako těžké i háčkovky či vsuvky (které patří mezi řešitelsky oblíbené), prokázal přítel **Bohuš** s úspěšností řešitelů těsně pod 60 %. Na druhém konci žebříčku vidíme autory spíš tradičnějších druhů, nicméně ani jejich vysoké zastoupení nepomohlo příteli **Kousalovi**, aby se stal největším miláčkem luštitelů – post plodného autora s nejvyšší průměrnou úspěšností získal pro rok 2021 přítel **Salavec**, u něhož uspěli v průměru vždy čtyři z pěti účastníků.

5.4 Podle délky

Řešiteli se na první pohled obvykle zdá dvouřádková hádanka lehčí než čtyř- nebo víceřádková. Je tomu tak ale doopravdy?

Počet řádků	Počet hádanek	Úspěšnost [%]	Počet řádků	Počet hádanek	Úspěšnost [%]
2	121	71,3	4	193	68,9
3	2	65,7	6 a více	17	67,9

Tabulka ukazuje, že dvouřádkové hádanky byly sice v průměru lehčí než čtyřřádkové, ale rozdíl byl malý. Ani šestiřádkové hádanky nijak extrémně nevybočovaly z průměru. Jiné počty řádků než 2, 4 nebo 6 již byly zřídka.

5.5 Autor, druh, nebo délka?

Co tedy mělo na obtížnost hádanek hlavní vliv? Byla to osobnost autora? Nebo druh hádanky? Nebo něco jiného (délka, obrázkové podání)?

Díky pokročilým matematickým metodám dokážeme oddělit vliv autora, druhu i dalších faktorů. Zde jsme použili tzv. [logistickou regresi](#). Pokoušíme se vysvětlit rozdíly v průměrné úspěšnosti řešení hádanky pomocí jejího autora, druhu (kategorie), typu podání a počtu řádků. Vzali jsme přitom jen hádanky těch autorů, kteří měli v ročníku aspoň 10 hádanek (vyřadili jsme minoritní autory), tj. 272 hádanek.

Výsledný model sice nefunguje nijak zázračně (vysvětluje jen 29,9 % rozdílů mezi hádankami), ale aspoň nějakou informaci nám dává. Když vezmeme v úvahu všechny faktory najednou (autor, druh, počet řádků, typ podání):

- počet řádků sám o sobě nemá na úspěšnost vliv;
- obrázkové podání hádanky zvyšuje snadnost hádanky přibližně 1,7krát;
- významně těžší hádanky (po odečtení vlivu konkrétního druhu hádanky) dodávali Bohuš, Rias a Vokatý; významně lehčí hádanky pak Kokeš, Říha, Salavec a Ambrož;
- významně těžší kategorie druhů hádanek (po odečtení vlivu konkrétního autora) jsou lavinovka+pyramidka+vlnovka, obměnka, výlučka+přídavka, záměnka; významně lehčí kategorie jsou SSSSČ, obložka, “pomocné výrazy”.

Ze závěrů matematického modelování tak vyplývá například, že vysoká průměrná úspěšnost Kousalových hádanek (viz výše kapitola *Podle autorů*) není dána pouhou “snadností” autora, ale hlavně tím, že tvořil snazší druhy hádanek. Konkrétní srovnání “snadnosti” či “obtížnosti” jednotlivých autorů i jednotlivých druhů hádanek mezi sebou obsahují následující dvě tabulky. Uvedená čísla ukazují, **kolikrát snazší hádanky produkoval autor v levém sloupci ve srovnání s autorem v horním řádku**, případně **kolikrát snazší byl druh v levém sloupci ve srovnání s druhem v horním řádku**. Jde jen o výstupy modelu založeného na Bystřině 2021, nelze je zobecnit na kompletní tvorbu autorů ani na všechny druhy hádanek, nicméně určitý pohled nabízejí.

Co znamená, že určitý druh (autor) je např. dvakrát (obecně K-krát) lehčí než jiný?

Snadnost či obtížnost hádanky zde vyjadřujeme tzv. *šanci*, tedy počtem úspěšných pokusů k počtu neúspěšných pokusů. Pokud by například účastník vyřešil polovinu hádanek určitého druhu, má šanci 1:1. Dvakrát snazší druh znamená, že počet úspěšných pokusů se zdvojnásobí a počet neúspěšných zůstane stejný – úspěšné jsou tak dva pokusy ze tří a šance se změnila na 2:1. Pokud byl výchozí poměr 1:4 (účastník vyřešil správně jen jednu z pěti hádanek určitého druhu), u dvakrát lehčího druhu by už byla jeho šance $2:4=1:2$, tedy měl by vyřešit jednu ze tří.

Vliv autora a druhu lze mezi sebou násobit. Přísvuka od D. Lančariče by tak podle modelu měla být čtyřikrát snazší než výlučka od F. Vokatého: přísvuky jsou dvakrát lehčí než výlučky a hádanky D. Lančariče dvakrát lehčí než hádanky F. Vokatého.

	Bohuš	Vokatý	Rias	Kožušník	Roemer	Konečný	Šimek	Lančarič	Zahajský	Kousal	Ambrož	Říha	Kokeš
Salavec	3,6	3,4	3,0	2,2	2,1	1,9	1,9	1,7	1,6	1,5	1,3	1,3	1,3
Kokeš	2,8	2,6	2,4	1,7	1,7	1,5	1,5	1,3	1,2	1,2	1,0	1,0	
Říha	2,7	2,6	2,3	1,7	1,6	1,5	1,5	1,3	1,2	1,1	1,0		
Ambrož	2,7	2,6	2,3	1,7	1,6	1,5	1,5	1,3	1,2	1,1			
Kousal	2,4	2,3	2,1	1,5	1,4	1,3	1,3	1,1	1,1				
Zahajský	2,3	2,2	2,0	1,4	1,4	1,3	1,2	1,1					
Lančarič	2,1	2,0	1,8	1,3	1,3	1,2	1,2						
Šimek	1,9	1,8	1,6	1,1	1,1	1,0							
Konečný	1,9	1,8	1,6	1,1	1,1								
Roemer	1,7	1,6	1,4	1,0									
Kožušník	1,7	1,5	1,4										
Rias	1,2	1,1											
Vokatý	1,1												

	LAV/PYR/VLN	OBMĚNKA	PŘÍD/VÝL	ZÁM	ZN	PÁD/ČAS	PŘESUNKA+SLAB	ODSUVKA	RÁČEK/OBRAT	PRON/POLOV/VYB/VVSL	VÝMĚNKA+SLAB	PŘÍSUVKA	CHYBOVKA	VS/VÝP	HÁČ/MĚK/YPS	POMVÝR	OBLOŽKA
SSSSČ	5,2	4,5	3,7	3,6	3,1	2,5	2,3	2,2	2,2	2,0	1,9	1,8	1,8	1,7	1,7	1,4	1,1
OBLOŽKA	4,6	4,0	3,3	3,2	2,8	2,2	2,1	2,0	2,0	1,8	1,7	1,6	1,6	1,5	1,5	1,2	
POMVÝR	3,7	3,3	2,7	2,6	2,2	1,8	1,7	1,6	1,6	1,4	1,4	1,3	1,3	1,2	1,2		
HÁČ/MĚK/YPS	3,1	2,7	2,2	2,1	1,9	1,5	1,4	1,3	1,3	1,2	1,1	1,1	1,1	1,0			
VS/VÝP	3,0	2,7	2,2	2,1	1,8	1,5	1,4	1,3	1,3	1,2	1,1	1,1	1,1				
CHYBOVKA	2,9	2,5	2,1	2,0	1,8	1,4	1,3	1,3	1,2	1,1	1,1	1,0					
PŘÍSUVKA	2,8	2,5	2,0	2,0	1,7	1,4	1,3	1,2	1,2	1,1	1,0						
VÝMĚNKA+SLAB	2,7	2,4	2,0	1,9	1,6	1,3	1,2	1,2	1,2	1,0							
PRON/POLOV/VYB/VVSL	2,6	2,3	1,9	1,8	1,6	1,2	1,2	1,1	1,1								
RÁČEK/OBRAT	2,3	2,0	1,7	1,6	1,4	1,1	1,1	1,0									
ODSUVKA	2,3	2,0	1,7	1,6	1,4	1,1	1,0										
PŘESUNKA+SLAB	2,2	2,0	1,6	1,6	1,3	1,1											
PÁD/ČAS	2,1	1,8	1,5	1,5	1,3												
ZN	1,7	1,5	1,2	1,2													
ZÁM	1,4	1,3	1,0														
PŘÍD/VÝL	1,4	1,2															
OBMĚNKA	1,1																

Příklad: po odečtení vlivu autorů vyšly odsuvky jako druh dvakrát snazší než obměnky. Šimkova obměnka by však vyšla stejně těžká jako Bohušova odsuvka, neboť Šimek je dvakrát „snazší“.

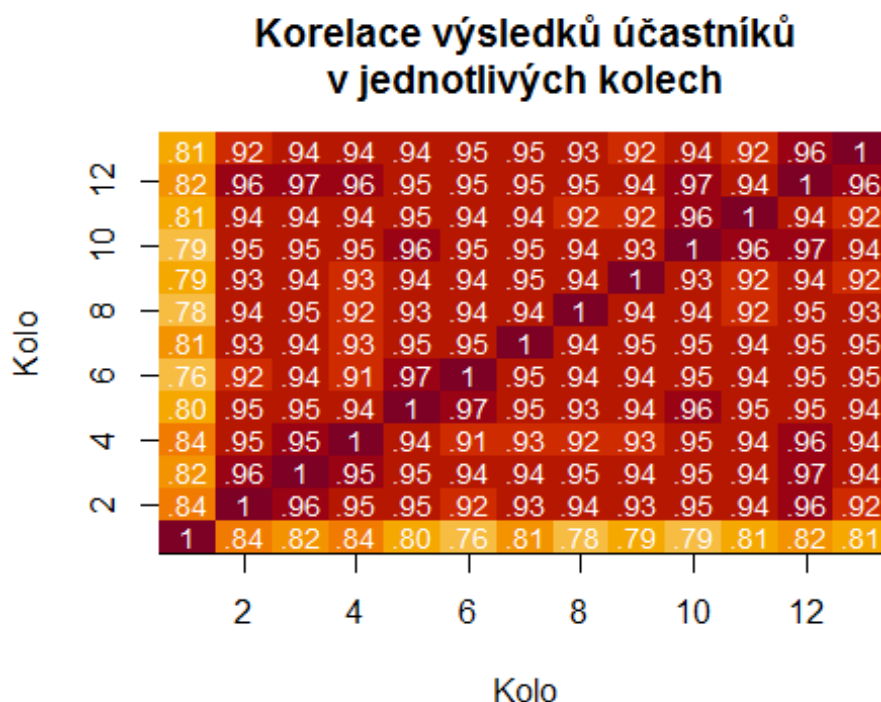
6 Stálost výsledků účastníků v čase

Při 13 kolech rozprostřených do celého roku a velkém množství druhů hádanek se nabízí několik otázek:

- Daří se účastníkům držet stabilní výkonnost během roku?
- Do jaké míry je úspěšné vyřešení jednotlivých hádanek dáno náhodou?
- Znamená nadprůměrná úspěšnost v jednom druhu hádanky, že se účastníkovi budou nadprůměrně dařit i ostatní druhy?

Základním statistickým nástrojem pro analýzu podobných otázek je **korelace**. Ta číselně vyjadřuje, jak spolu u stejného účastníka souvisejí hodnoty více veličin (např. počty bodů v jednotlivých kolech). Nejlepší účastníci ve všech kolech získali plný nebo téměř plný počet bodů, takže jejich výkonnost byla v čase stálá. Platí to ale i pro ostatní? Víme (viz výše kapitola *Účastníci*), že asi polovina startovního pole měla celkový počet bodů rozprostřený v širokém rozmezí od 0 do 280.

Následující obrázek ukazuje hodnoty korelačního koeficientu mezi výsledky účastníků v jednotlivých kolech. (Kvůli úspoře místa jsou desetinná čísla vypsána bez úvodní nuly.)



Je poměrně překvapivé, že výsledky účastníků ve všech kolech od 2. do 13. mají velmi vysokou korelaci (teoreticky maximální hodnota je 1). Dokonce i nejvzdálenější kola (2. a 13.) mají korelaci nad 0,9. Korelace s 1. kolem je o něco nižší (okolo 0,8), to je proto, že ve snadném 1. kole

dosáhli plného počtu bodů i ti, kdo v dalších obtížnějších kolech již pravidelně ztráceli. **Výsledky účastníků jsou v čase velmi stabilní.**

Celkové skóre řešitele je určeno součtem bodů za jednotlivé hádanky. Každá jednotlivá hádanka přitom rozděluje účastníky do dvou skupin podle toho, zda ji vyřešili. Pokud některé hádanky "krájí" startovní pole ve shodě s celkovým skóre, tedy vyřeší je pouze ti celkově lepší, dalo by se z výsledku účastníka v těchto několika hádankách odhadnout, jak účastník dopadne celkově. Úspěšné řešení takových hádanek pak určitě není náhodné, ale odpovídá celkovým schopnostem účastníka.

Pomůže nám opět korelace. U každé hádanky spočítáme korelaci mezi výsledkem účastníků v této hádance a součtem bodů za všechny zbývající hádanky. (V tomto případě bereme v úvahu jen kompletní účastníky.) Následující tabulka uvádí čísla hádanek s nejvyšší hodnotou korelace, tedy hádanky, jejichž výsledek nejtěsněji souvisí s celkovým skóre.

Č.	Druh	Pořadí		Autor	Úspěšnost [%]	Korelace
		Kolo	v kole			
116	ZÁHADNÝ NÁPIS	5	20	Říha	64,3	0,895
210	VÝPUSTKA	9	10	Bohuš	65,3	0,892
144	CHYBOVKA	6	22	Kokeš	69,4	0,890
207	VSUVKA	9	7	Kousal	62,2	0,888
111	POLOVSUVKA	5	15	Kousal	65,3	0,887
158	OBNAŽKA	7	10	Vokatý	67,3	0,887
287	ZÁMĚNKA	12	6	Říha	65,3	0,884
246	PERMUTOVKA	10	19	Říha	70,4	0,881
126	ZÁHADNÝ NÁPIS	6	4	Kožušník	61,2	0,880
283	ZÁMĚNKA	12	2	Šimek	60,2	0,879
28	ZÁMĚNKA	2	10	Zahajský	66,3	0,877
249	VÝPUSTKA	10	22	Bohuš	59,2	0,875
167	VÝPUSTKA	7	19	Ambrož	70,4	0,875
180	OBMĚNKA	8	6	Kokeš	65,3	0,870
173	OBMĚNKA	7	25	Kokeš	61,2	0,866
41	ZÁHADNÝ NÁPIS	2	23	Lančarič	71,4	0,865
139	PERMUTOVKA	6	17	Zahajský	67,3	0,865
242	OBMĚNKA	10	15	Salavec	68,4	0,865
241	VSUVKA	10	14	Lančarič	66,3	0,863
306	VSUVKA	12	25	Adamčík	71,4	0,862

Jak vidíme, pro schopnost rozlišit mezi celkově lepšími a horšími účastníky nemusí být hádanka nijak extrémně těžká ani neobvyklého druhu. V tomto směru dobře fungují záhadné nápisy, obměnky, ale i tradičnější záměnky, vsuvky a výpustky. Hodnoty korelace jsou jednotlivě velmi vysoké. Pokud bychom každému účastníkovi spočítali skóre jen na výše uvedených 20 hádankách, pak korelace tohoto redukovaného skóre a celkového skóre (na všech

333 hádankách) je 0,977, tedy velmi vysoká. Dá se říct, že **výsledek za pouhých 20 uvedených hádanek by stačil k velmi přesné předpovědi celkového výsledku.**

Poznámka: samozřejmě vedle celkového rozlišení účastníků je důležité i rozlišení mezi účastníky v naprosté špičce, proto hádanky s úspěšností okolo 65 % musí být doplněny sadou velmi těžkých hádanek. Stejně je důležité neodradit slabší řešitele a zařadit i lehké hádanky.

Pro zajímavost uvádíme i deset hádanek, které jsou na opačné straně – výsledek v nich s celkovým skóre koreluje slabě. V naprosté většině se jedná o hádanky velmi lehké, je ale zajímavé, že pouze jediná z nich pochází z náborového 1. kola. Celkově lze říct, že **naprostá většina hádanek silně koreluje s celkovým skóre, a proto pouhá náhoda či štěstí k jejich vyřešení zpravidla nestačí.**

Č.	Druh	Kolo	Pořadí v kole	Autor	Úspěšnost [%]	Korelace
91	LAVINOVKA	4	21	Vokatý	2,0	0,120
190	OBLOŽKA	8	16	Kokeš	99,0	0,121
323	VSUVKA	13	16	Tršická	88,8	0,202
260	VÝPUSTKA	11	6	Lančarič	96,9	0,243
218	ČÁRKOVKA	9	18	Zahajský	96,9	0,295
44	VÝPUSTKA	2	26	Roemer	82,7	0,304
50	SLOUČENKA	3	6	Kousal	95,9	0,313
181	MNOŽENKA	8	7	Adamčík	96,9	0,317
6	ODSUVKA	1	6	Adamčík	94,9	0,318
105	OBLOŽKA	5	9	Lančarič	95,9	0,330

7 Existuje specifický talent?

Jak již bylo uvedeno, druhů hádanek existuje téměř sto, v Bystřině 2021 jich bylo zastoupeno 51. I typy prováděných operací v hádankách se liší, u některých to jsou tzv. mutace (přidávání, ubírání či změna písmen), u jiných zase tvaroslovné změny atd. Nabízí se otázka, zda různí řešitelé mají specifické nadání na řešení určitých druhů hádanek – např. někomu jdou lépe mutační než tvaroslovné a jinému zase naopak.

Otázku lze položit i jinak. Do jaké míry byl celkový výsledek účastníka dán jediným globálním faktorem (“všeobecným nadáním na hádanky”) a jak moc se na něm podílely specifické faktory (např. “nadání na permutační hádanky”)?

Matematicky se taková věc dá zjistit tak, že výsledek každého účastníka spočítáme po skupinách hádanek stejného druhu (přesněji řečeno za kategorii podobných druhů) a zjistíme, zda výsledek ve všech skupinách spolu těsně souvisí, nebo zda některé takové skupiny tvoří specifické faktory relativně málo závislé na jiných. Použijeme k tomu metodu nazvanou [analýza hlavních komponent](#) (označovanou též anglickou zkratkou PCA). Jiná možná metoda by byla [faktorová analýza](#) používaná např. v sociologii, nicméně zjistili jsme, že obě metody vedou k velmi podobným závěrům.

Z analýzy metodou hlavních komponent vyplývá, že **výsledek účastníka je z 91,1 % dán jedním hlavním faktorem**, který můžeme označit jako “faktor všeobecných schopností”. Do faktoru přispívají všechny druhy hádanek v podstatě stejnou měrou.

Druhý nejvýznamnější faktor má už jen 2,2 % vlivu na celkový výsledek. Tabulka níže ukazuje, jakou vahou která kategorie druhů hádanek do druhého faktoru přispívá (váhy jsou na škále 0 až 10).

Kategorie druhu	Váha ve faktoru	Kategorie druhu	Váha ve faktoru
LAV/PYR/VLN	9,4	PŘESUNKA+SLAB	4,8
PŘÍD/VÝL	7,9	ODSUVKA	4,7
VOKÁLOVKA	7,3	RÁČEK/OBRAT	4,5
PRON/POLOV/VYB/VVSL	7,3	HÁČ/MĚK/YPS	4,4
ZÁM	6,7	OBLOŽKA	3,6
ZN	6,4	VS/VÝP	3,6
PŘESM/PERM	6,4	POMVÝR	3,0
OBMĚNKA	5,9	SSSSČ	1,9
OBNAŽKA	5,7	POSUNKA	1,8
VÝMĚNKA+SLAB	5,5	CHYBOVKA	1,6
OSTATNÍ	5,3	PÁD/ČAS	1,5
PŘÍSUVKA	5,2		

Jak vidíme, největší váhu ve druhém faktoru mají hádanky nejtěžších druhů (viz kapitulu o obtížnosti druhů hádanek výše). Závěr tedy je, že **nic jako specifický talent na řešení hádanek určitého druhu v datech nevidíme**; pro celkový výsledek v soutěži je důležitá, vedle obecné řešitelské schopnosti, schopnost se vypořádat s těžkými hádankami, které mohou být různého druhu.

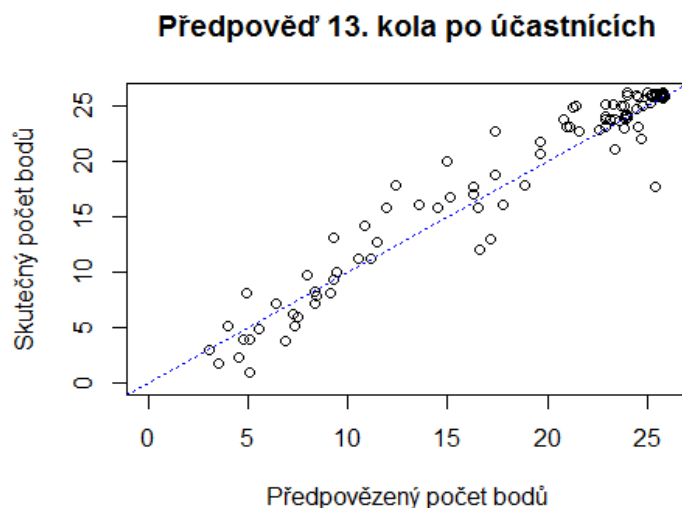
8 Předpověď výsledku 13. kola

Na závěr ilustrujeme dosavadní poznatky experimentem. Představme si, že bylo právě zveřejněno 13. kolo, v němž známe autory i druhy jednotlivých hádanek. Jak přesně dokážeme předpovědět pro každého kompletního účastníka, kolik bodů ve 13. kole získá?

Je celkem jasné, že o **úspěšnosti účastníka při řešení konkrétní hádanky** rozhoduje **jeho úroveň** (řešitelská kvalita) a **obtížnost hádanky**. Úroveň účastníka je v čase stabilní a lze předpokládat, že ve 13. kole bude podobná jako v souhrnu předchozích kol. Obtížnost je pak možné rozdělit na vliv autora (“obtížní” vs. “snadní” autoři), vliv druhu hádanky a vliv podání (obrázkové vs. textové), viz výše kapitulu *Autor, druh, nebo délka?*.

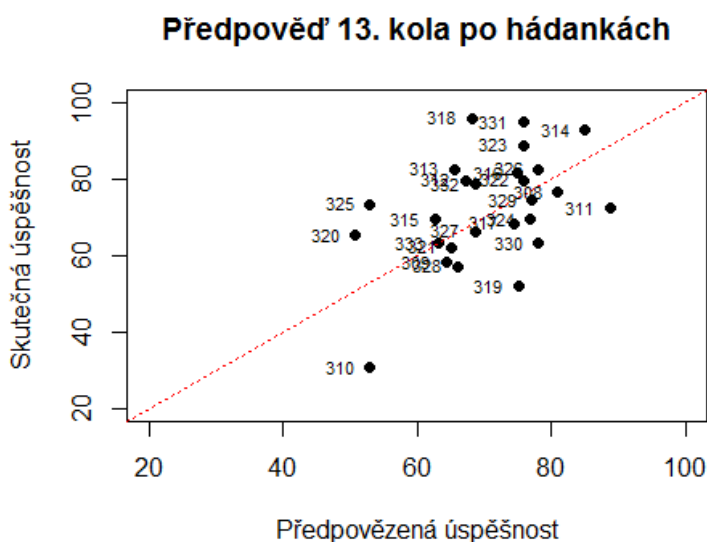
Pomocí matematického modelu odhadneme z dat 1.–12. kola, jakou průměrnou úspěšnost očekáváme v jednotlivých hádankách 13. kola podle jejich autorů, druhu a podání. Dále u každého účastníka spočítáme dosavadní průměrnou úspěšnost za 12 kol. Tyto dvě informace

zkombinujeme a odhadneme pro každého účastníka, kolik bodů celkem získá. Následující dva grafy ukazují, jak se naše odhady trefily do reality.



Každý bod je jeden účastník. Body na tečkované čáře jsou případy, kdy se předpověď přesně trefila do skutečného bodového zisku. Body nad touto čarou znamenají vyšší bodový zisk, než tvrdila předpověď, body pod čarou pak nižší bodový zisk oproti předpovědi. Jak je vidět, **předpovědi získaných bodů za účastníky se docela dobře trefily**. Korelace předpovězených a skutečných počtů bodů je 0,97, tedy velmi vysoká.

V průměru účastníci dopadli nepatrně lépe, než očekávala předpověď. Průměrná předpovězená úspěšnost byla 70,5 %, kdežto průměrná skutečná úspěšnost 72,3 %.



Jak vidíme, předpovědi “snadnosti” či “obtížnosti” se u jednotlivých hádanek už tak dobře netrefily, i když přibližně polovina z nich je poměrně blízko tečkované čáry, a tedy se pro předpověď od skutečnosti nelišila až tak zásadně. Nicméně např. hádanka č. 320 měla být podle

předpovědi nejtěžší a č. 311 zase nejsnazší, ale ve skutečnosti skončily obě někde okolo průměru. **Odhadovat obtížnost hádanky pouze podle autora, druhu a typu podání je tedy vcelku nepřesné.**

9 Závěr

Tato studie se pokusila najít v datech ze soutěže Bystřina 2021 zajímavé informace a pomocí statistických postupů vydolovat skryté souvislosti. Podařilo se mimo jiné ukázat, že:

- výkonnost účastníků je po celou dobu dost stabilní a výsledek každého dalšího kola lze celkem přesně předpovědět;
- naprostá většina hádanek je řešena úměrně schopnostem řešitele a vliv náhody je minimální;
- do jisté míry lze určit, které druhy hádanek a kteří autoři se účastníkům řeší snáze;
- schopnost účastníků řešit hádanky je všestranná, nedá se říct, že by pro určité druhy hádanek byl potřeba specifický talent.

Analytické postupy lze dále rozvíjet, například lze detekovat, kdy je shoda odpovědí (úspěšností) ve dvojici či ve větší skupině účastníků již nad rámec náhody.

10 Kontakt

Jan Hučín, e-mail honza@hucin.cz, tel. 737 904 870