

Základy statistického zpracování údajů – praktické cvičení

zadáno 29. dubna 2020

Datový soubor „tennis.men.2012-14“ – analýza dat

V souboru máte 40 nejlepších hráčů světového žebříčku a jejich celoroční výsledky separátně za roky 2012, 2013 a 2014. Nezapomínejte, že někteří hráči v datech opakovaně figurují za různé roky. Pro jednoznačné pochopení zadaných úloh jsem připojil ukázková řešení. Vaše výsledky mají být stejné.

Úloha A) Sestavte kontingenční tabulku, která ukáže pro každou zemi, kolik hráčů v datech figuruje pro jednotlivé roky.

- *Nápověda:* Veškerá data pro všechny roky lze zkopírovat na jeden list.

Úloha B) Kolik různých hráčů je pro každou zemi v datech celkově?

- *Nápověda:* Z originálních dat je nutno nejprve odstranit duplicity pomocí filtru (Data—Filtr—Upřesnit—Bez duplicitních záznamů), a pak vytvořit kontingenční tabulku.

Úloha C) Pomocí kontingenční tabulky nalezněte všechny hráče, kteří v letech 2012–2014 sehráli alespoň 150 utkání. Berte přitom do úvahy pouze hráče ze zemí Španělsko, Itálie, Česká republika, Spojené státy a Francie.

- *Nápověda:* Nejprve sestavte kontingenční tabulku a nastavte filtr na zemi. Potom můžete data v ní seřadit podle počtu sehraných utkání.

Úloha D) Pro každou zemi zjistěte, kolik má hráčů, kteří v každém roce vyhráli alespoň 3/5 svých zápasů.

- *Nápověda:* V tabulce se všemi hráči za všechny roky nejprve vyfiltrujte pouze řádky, kde je ve sloupci PCT hodnota alespoň 60%. Z těchto údajů sestavte kontingenční tabulku tak, že v každém jejím políčku bude údaj (0–3) o tom, v kolika letech se danému hráči podařilo vyhrát alespoň 60% svých zápasů. Nakonec pro každou zemi spočítejte, kolik je hráčů, kteří podmínku filtru splňují pro všechny tři roky, a to jednoduše tak, že zjistíte, kolik trojek se nachází v řádku pro danou zemi – viz ukázkové řešení níže.
- *Jiné možné řešení:* Původní tabulku s daty můžete také ponechat kompletní, a pak až ve vytvořené kontingenční tabulce nastavíte filtr na hodnotu PCT. Výsledek by měl být totožný.

Důležité – odevzdejte emailem nejpozději v pondělí 4. 5. 2020.

- V souboru, který odevzdáte, zachovejte původní data, jak jste je obdrželi.
- Řešení každé úlohy provádějte na zvláštním listu, kam si podle potřeby data nakopírujete.
- Listy s řešením pojmenujte A, B, C, D.
- Pokud budete používat pomocné listy, kde si sestavíte pomocné tabulky, v odevzdaném souboru je rovněž zachovejte.

Ukázkové řešení A)

Count - Player	Year			Total Result	
Country	2012	2013	2014		
ARG		3	2	2	7
AUS		1			1
AUT			1		1
BRA		1			1
BUL			1	1	2
CAN		1	1	1	3
COL				1	1
CRO		1	1	2	4
CYP		1			1
CZE		1	2	2	5
ESP		7	7	6	20
FIN		1	1	1	3
FRA		5	6	7	18
GBR		1	1	1	3
GER		2	3	2	7
ITA		1	2	2	5
JPN		1	1	1	3
LAT			1	1	2
NED			1		1
POL		1			1
POR				1	1
RSA		1	1	1	3
RUS		2	2		4
SRB		3	1	1	5
SUI		2	2	2	6
UKR		1		1	2
USA		2	2	3	7
UZB		1	1	1	3
Total Result		40	40	40	120

Ukázkové řešení B)

POZOR: Konkrétní hráč nesmí být za jednu zemi započítán vícekrát – proto je třeba z datové tabulky odstranit duplicity.

Počet hráčů za každou zemi -- bez duplicit

Popisky řádků	Počet z Country
ARG	5
AUS	1
AUT	1
BRA	1
BUL	1
CAN	1
COL	1
CRO	3
CYP	1
CZE	2
ESP	11
FIN	1
FRA	8
GBR	1
GER	3
ITA	2
JPN	1
LAT	1
NED	1
POL	1
POR	1
RSA	1
RUS	3
SRB	3
SUI	2
UKR	1
USA	3
UZB	1
Celkový součet	62

Ukázkové řešení C)

Player	Country	Součet z Matches
Ferrer	ESP	253
Berdych	CZE	236
Tsonga	FRA	190
Nadal	ESP	189
Isner	USA	188
Gasquet	FRA	187
Simon	FRA	180
Seppi	ITA	178
Lopez	ESP	167
Querrey	USA	160
Benneteau	FRA	152
Almagro	ESP	147
Fognini	ITA	135
Robredo	ESP	125
Paire	FRA	114
Rosol	CZE	108
Monfils	FRA	106
Verdasco	ESP	106
Agut	ESP	68
Chardy	FRA	60
Ramos-Vinolas	ESP	55
Granollers	ESP	52
Garcia-Lopez	ESP	51
Roger-Vasselin	FRA	49
Andujar	ESP	48
Sock	USA	47

