

Popis sezónní složky – řešený příklad

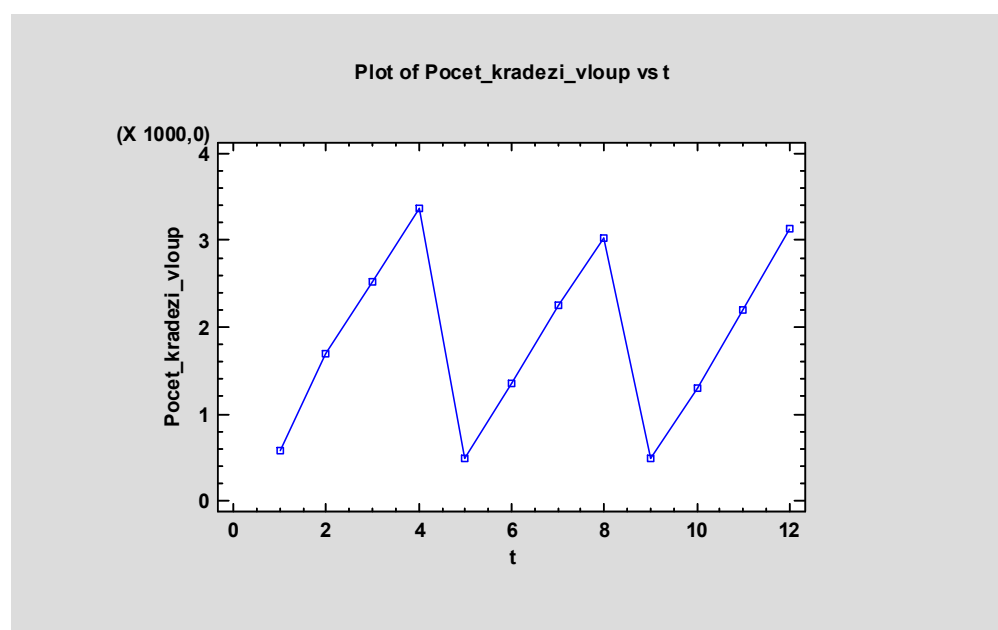
Máme k dispozici údaje o počtu krádeží vloupáním v Libereckém kraji v jednotlivých čtvrtletích let 2016-2018. V této časové řadě byla prokázána existence sezónnosti. Kvantifikujte velikost sezónních výkyvů a očistěte od nich tuto časovou řadu.

Čtvrtletí/Rok	Počet krádeží
I/2016	586
II/2016	1688
III/2016	2526
IV/2016	3365
I/2017	491
II/2017	1361
III/2017	2252
IV/2017	3032
I/2018	488
II/2018	1304
III/2018	2190
IV/2018	3131

Zdroj: <https://www.policie.cz/clanek/statisticke-prehledy-kriminality-za-rok-2018.aspx>

Řešení:

Pokud je v časové řadě prokázána existence sezónnosti, je potřeba nejprve zjistit, jaký má charakter, tedy zda je velikost sezónních výkyvů proporcionální složce trendové nebo zda je sezónní složka od trendu oddělena a velikost sezónních výkyvů je nezávislá na vývoji složky trendové. V tom nám může napomoci bodový diagram (Obr. 1).



Obrázek 1 – Vývoj ukazatele počet krádeží vloupáním v čase

Z Obr. 1 je patrné, že velikost sezónních výkyvů je zhruba konstantní, proto bychom k popisu sezónní složky použili model konstantní sezónnosti, který je často označován také jako aditivní. Aditivní model má tento tvar:

$$y_{ij} = T_{ij} + S_{ij} + \varepsilon_{ij}.$$

V teoretické stati o popisu sezónní složky je uvedeno, že kvantifikace sezónních výkyvů probíhá ve třech krocích:

1. Výpočet empirických sezónních rozdílů, tj. $b_j = y_{ij} - \hat{T}_{ij}$.
2. Výpočet průměrných sezónních rozdílů.
3. Standardizace průměrných sezónních rozdílů, tj. výpočet rozdílových sezónních faktorů.

Abychom mohli provést první krok, je potřeba vypočítat trendové hodnoty. U časových řad, které jsou charakteristické velkými výkyvy v trendu, je vhodné použít k jejich vyrovnání klouzavé průměry. Vzhledem k tomu, že jde o časovou řadu periodickou, je potřeba stanovit rozsah klouzavé části období interpolace podle délky dané periody, zde tedy $m = 4$. Tuto časovou řadu tedy vyrovnáme pomocí čtyřčlenných centrovaných klouzavých průměrů. Výpočet těchto průměrů je podrobně uveden v příkladu 8_3 u Tématu 8 v e-learningovém kurzu KSY/ST2. Výsledné hodnoty klouzavých průměrů uvedeme do Tab. 1.

Tabulka 1 – Hodnoty čtyřčlenných centrovaných klouzavých průměrů

Čtvrtletí/Rok	Počet krádeží	\hat{T}_{ij}
I/2016	586	•
II/2016	1688	•
III/2016	2526	2029,375
IV/2016	3365	1976,625
I/2017	491	1901,5
II/2017	1361	1825,625
III/2017	2252	1783,625
IV/2017	3032	1776,125
I/2018	488	1761,25
II/2018	1304	1765,875
III/2018	2190	•
IV/2018	3131	•

Nyní již můžeme vypočítat empirické sezónní rozdíly a uvést je opět do tabulky – viz Tab. 2.

Tabulka 2 – Hodnoty empirických sezónních rozdílů

Čtvrtletí/Rok	Počet krádeží	\hat{T}_{ij}	Sezónní rozdíly
I/2016	586	•	•
II/2016	1688	•	•
III/2016	2526	2029,375	496,625
IV/2016	3365	1976,625	1388,375
I/2017	491	1901,5	-1410,500
II/2017	1361	1825,625	-464,625
III/2017	2252	1783,625	468,375
IV/2017	3032	1776,125	1255,875
I/2018	488	1761,25	-1273,25
II/2018	1304	1765,875	-461,875
III/2018	2190	•	•
IV/2018	3131	•	•

První dvě a poslední dvě hodnoty zůstanou neznámé, protože nemáme hodnoty klouzavých průměrů, se kterými bychom porovnávali napozorovaná data.

Dalším krokem má být výpočet průměrných sezónních rozdílů. Počítají se proto, že každé dílčí období v rámci roku, zde čtvrtletí, by mělo být zastoupeno jednou hodnotou, charakterizující velikost sezónního výkyvu. Zde však máme pro jednotlivá čtvrtletí hodnoty dvě. Je proto potřeba určit jejich průměrnou hodnotu. Ta se vypočítá podle:

$$\bar{b}_j = \frac{1}{m} \sum_{i=1}^m (y_{ij} - \hat{T}_{ij}) .$$

První hodnotu tak vypočítáme jako $\bar{b}_1 = \frac{1}{2}(-1410,5 - 1273,25) = -1341,875$. Další průměrný sezónní index bude $\bar{b}_2 = \frac{1}{2}(-464,625 - 461,875) = -463,25$. Třetí hodnota se vypočítá jako $\bar{b}_3 = \frac{1}{2}(496,625 + 468,375) = 482,5$ a obdobně pak hodnota pro IV. čtvrtletí, jak vidíme v tabulce 3.

Tabulka 3 – Hodnoty průměrných sezónních rozdílů

Čtvrtletí	\bar{b}_j
I	-1341,875
II	-463,250
III	482,500
IV	1322,125
Součet	-0,500

Na sezónní rozdíly je obvykle kladena normalizační podmínka, která říká, že vliv sezónních faktorů by se měl v rámci periody vykompenzovat, proto by součet sezónních rozdílů, a tedy i průměrných sezónních rozdílů, měl být nulový. Zde je však roven hodnotě -0,5. Proto přistoupíme ke kroku třetímu, kdy uplatněním výše uvedené podmínky dostaneme hodnoty rozdílových sezónních faktorů. Jak tedy docílit toho, aby součet průměrných sezónních rozdílů byl nulový? Hodnotu, která do nuly chybí, dělíme počtem faktorů a ke každému faktoru pak výslednou hodnotu přičteme, abychom ji zvětšili a dostali se tak v součtu na nulu. Zde tedy provedeme výpočet $0,5 \div 4 = 0,125$. O číslo 0,125 nyní zvětšíme všechny hodnoty průměrných sezónních rozdílů, a tím dostaneme hodnoty rozdílových sezónních faktorů – viz Tab. 4.

Tabulka 4 – Hodnoty rozdílových sezónních faktorů

Čtvrtletí	\bar{b}_j	Rozdílové sezónní faktory
I	-1341,875	-1341,75
II	-463,250	-463,125
III	482,500	482,625
IV	1322,125	1322,250
Součet	-0,500	0,00000

Rozdílové sezónní faktory představují finální vyčíslení velikosti sezónních výkyvů. Lze je také jednoduše interpretovat. První z nich říká, že v I. čtvrtletí jsou hodnoty krádeží vloupáním v průměru o 1341,75 krádeží pod dlouhodobým normálem. Naopak ve III. čtvrtletí jsou hodnoty krádeží vloupáním v průměru o 482,625 krádeží nad dlouhodobým normálem.

Posledním úkolem je očistit časovou řadu od sezónních vlivů. To v případě aditivního modelu znamená, že od původních hodnot časové řady krádeží vloupáním odečteme hodnoty příslušných sezónních faktorů. První očistěnou hodnotu získáme jako rozdíl 586 a sezónního rozdílového faktoru pro I. čtvrtletí, tedy $586 - (-1341,75) = 1927,75$. Druhá hodnota pak bude $1688 - (-463,125) = 2151,13$. Další očistěné hodnoty jsou uvedeny v tabulce 5.

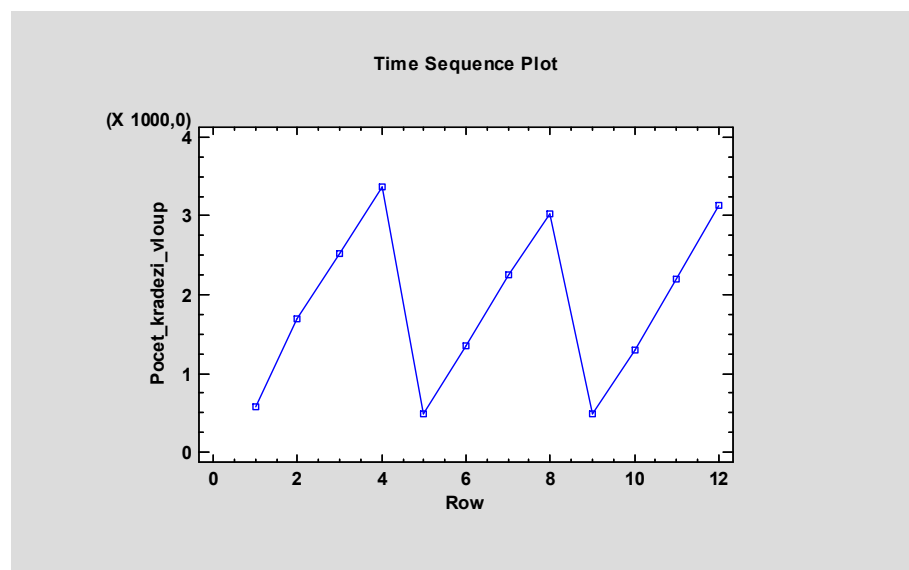
Tabulka 5 – Hodnoty časové řady očistěné od sezónních vlivů

Čtvrtletí/Rok	Počet krádeží	\hat{T}_{ij}	Sezónní rozdíly	Očistěná data
I/2016	586	•	•	1927,750
II/2016	1688	•	•	2151,125
III/2016	2526	2029,375	496,625	2043,375
IV/2016	3365	1976,625	1388,375	2042,750
I/2017	491	1901,5	-1410,500	1832,750
II/2017	1361	1825,625	-464,625	1824,125
III/2017	2252	1783,625	468,375	1769,375
IV/2017	3032	1776,125	1255,875	1709,750
I/2018	488	1761,25	-1273,25	1829,750
II/2018	1304	1765,875	-461,875	1767,125
III/2018	2190	•	•	1707,375
IV/2018	3131	•	•	1808,750

Pro takto sezónně očistěnou časovou řadu je pak možné najít vhodnou trendovou funkci, kterou se očistěné údaje vyrovnají a vytvářet předpovědi budoucího vývoje prodloužením trendu do budoucnosti. Nesmíme však zapomenout, že když vytvoříme tímto způsobem předpověď na dalších několik čtvrtletí, je vždy potřeba do časové řady sezónnost vrátit, tj. k vytvořené předpovědi pro I. čtvrtletí přidáme rozdílový sezónní faktor pro I. čtvrtletí atd.

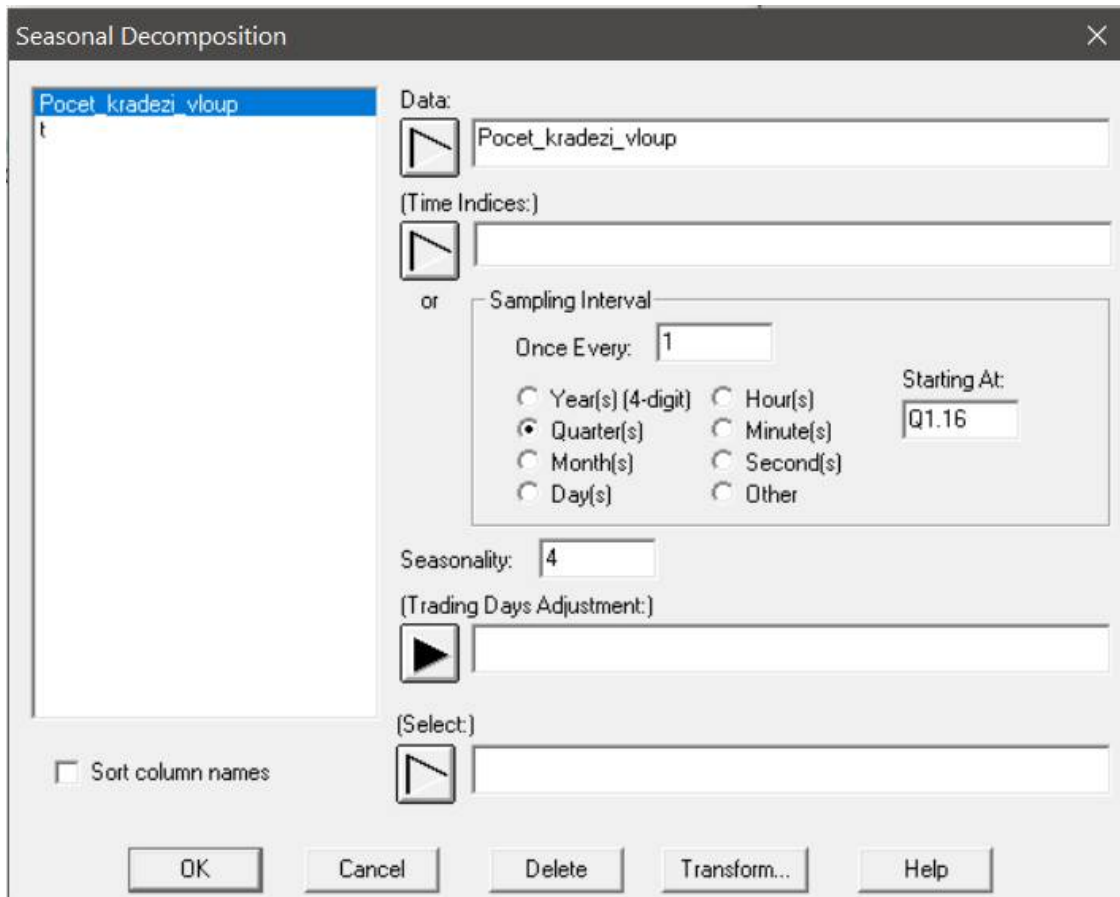
Řešení v SGP:

Do datového listu zadáme hodnoty krádeží – do prvního sloupce. Pokud se chceme podívat na „vzhled“ sezónních výkyvů, zvolíme posloupnost procedur **Plot – Scatterplots – X Plot ...** Do řádku *Data* vložíme hodnoty krádeží a potvrdíme OK. Nabídku *Tables and Graphs* potvrdíme tlačítkem OK, jak je. Zajímá nás zde výstup *Time Sequence Plot* – viz Obr. 2.



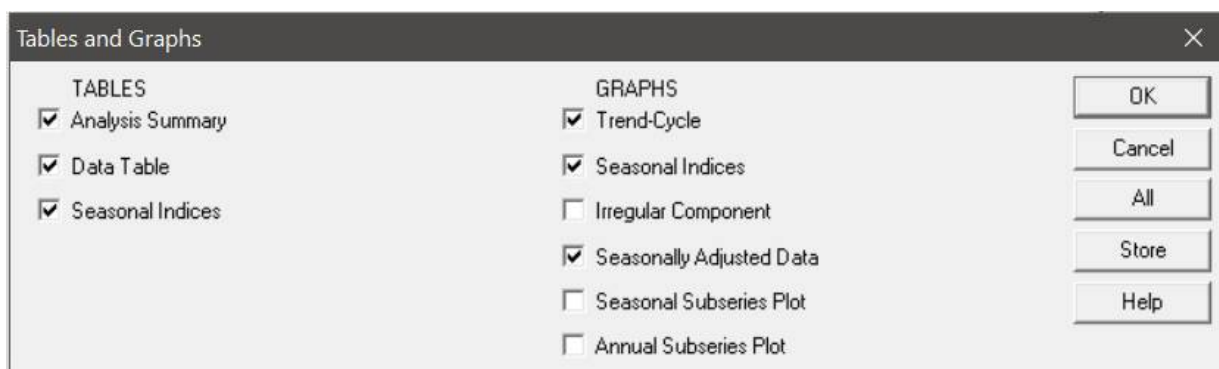
Obrázek 2 – Grafické znázornění počtu krádeží vloupáním v Libereckém kraji

Zde vidíme podobu sezónních výkyvů. Pokud však chceme sezónní výkyvy kvantifikovat a očistit od nich časovou řadu, je potřeba zvolit **Describe – Time Series – Seasonal Decomposition ...** Do vstupního panelu do řádku *Data* zadáme proměnnou krádeže. Dále zaškrtneme v oddílu *Sampling Interval* položku *Quarters*. V oddílu *Starting At* můžeme (ale nemusíme) vyplnit Q1.16, abychom měli ve výsledné tabulce přehledně uspořádaná data podle období, ke kterým přísluší. V oddílu *Seasonality* vyplníme hodnotu periody dané řady, tedy 4, a následně stiskneme tlačítko OK (Obr. 3).



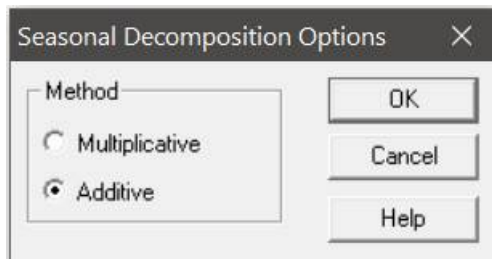
Obrázek 3 – Vstupní panel Seasonal Decomposition

Objeví se před námi panel *Adjustment Options*, kde nic neměníme a potvrdíme jen OK. Potom se objeví nabídka *Tables and Graphs*, kde v tabulkových výstupech zaškrtneme ještě *Data Table* a v grafických *Seasonally Adjusted Data* – viz Obr. 4.



Obrázek 4 – Nabídka Tables and Graphs v proceduře Seasonal Decomposition

Nejprve se podívejme na výstup *Data Table*. Hned v úvodní části výstupu je uvedena informace „Seasonal Decomposition Method: Multiplicative“. Pokud chceme použít metodu aditivní, nikoli multiplikativní, je potřeba na výstupu *Data Table* vyvolat např. pravým tlačítkem myši políčkovou volbu *Pane Options* a zde změnit použitou metodu na aditivní, jak ukazuje Obr. 5.



Obrázek 5 – Volba modelu popisu sezónní složky v *Seasonal Decomposition Options*

Ve výstupu na Obr. 6 je tabulka, kde v prvním sloupci je označení časové periody, ke které se vztahují údaje o počtu krádeží vloupáním, které jsou ve druhém sloupci. Sloupec nazvaný *Trend-Cycle* obsahuje hodnoty centrovaných čtyřčlenných klouzavých průměrů, kterými je vyrovnána časová řada krádeží. Sloupec *Seasonality* obsahuje hodnoty sezónních rozdílů, tak jak jsme je při ručním výpočtu vypočítali v Tab. 2, v posledním sloupci. Ve sloupci *Irregular* jsou hodnoty náhodné složky a v posledním sloupci jsou sezónně očištěná data. Stejných hodnot jsme se dopočítali při ručním výpočtu v posledním sloupci Tab. 5.

Seasonal Decomposition - Pocet_kradezi_vloup					
Data Table for Pocet_kradezi_vloup					
Seasonal decomposition method: Additive					
					Seasonally
Period	Data	Trend-Cycle	Seasonality	Irregular	Adjusted
Q1.16	586,0				1927,75
Q2.16	1688,0				2151,13
Q3.16	2526,0	2029,38	496,625	14,0	2043,38
Q4.16	3365,0	1976,63	1388,38	66,125	2042,75
Q1.17	491,0	1901,5	-1410,5	-68,75	1832,75
Q2.17	1361,0	1825,63	-464,625	-1,5	1824,13
Q3.17	2252,0	1783,63	468,375	-14,25	1769,38
Q4.17	3032,0	1776,13	1255,88	-66,375	1709,75
Q1.18	488,0	1761,25	-1273,25	68,5	1829,75
Q2.18	1304,0	1765,88	-461,875	1,25	1767,13
Q3.18	2190,0				1707,38
Q4.18	3131,0				1808,75

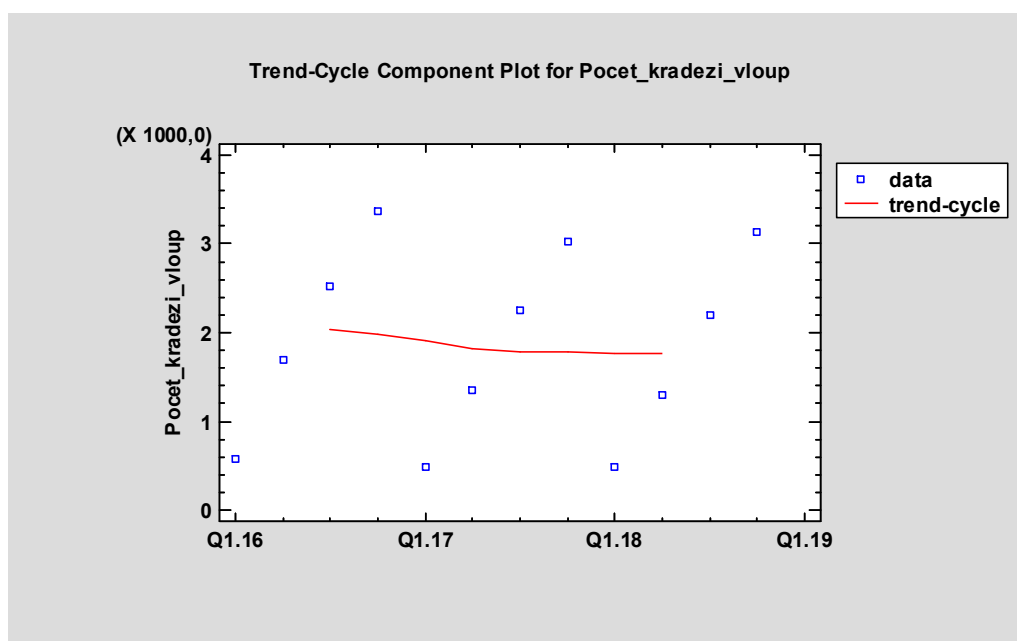
Obrázek 6 – Výstup *Data Table* v *Seasonal Decomposition*

Hodnoty rozdílových sezónních faktorů pak najdeme v tabulkovém výstupu nazvaném *Seasonal Indices* – viz Obr. 7. Už jsou zde finální hodnoty, které vyčíslují velikost sezónních výkyvů, tj. rozdílové sezónní faktory, tak, jak jsme je ručně vypočítali v posledním sloupci tabulky 4.

Season	Index
1	-1341,75
2	-463,125
3	482,625
4	1322,25

Obrázek 7 – Výstup Seasonal Indices v Seasonal Decomposition

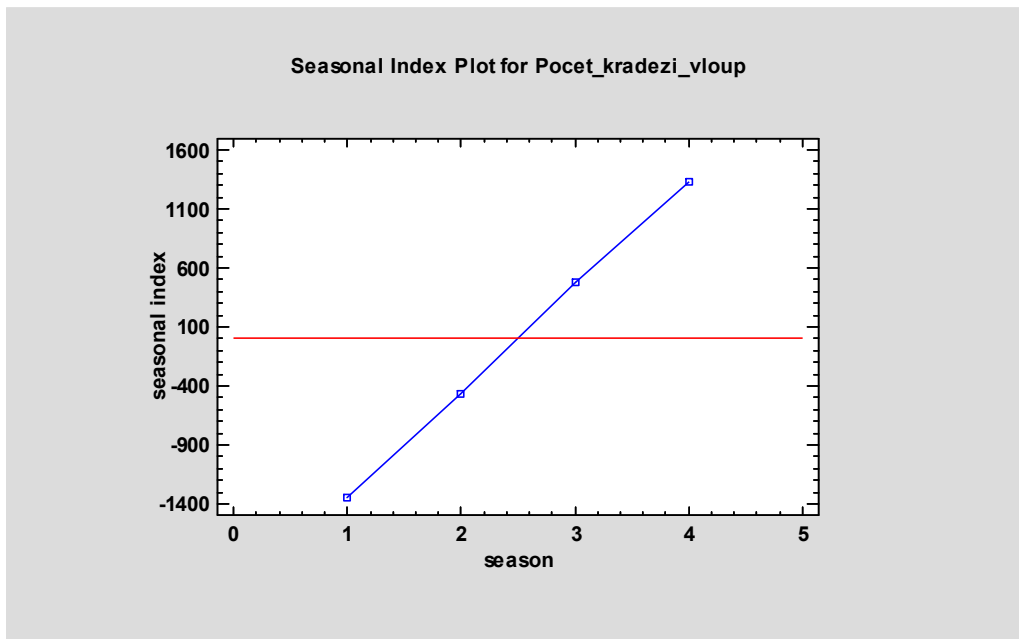
Zajímavé jsou také grafické výstupy, kdy Obr. 8 ukazuje původní hodnoty a jejich vyrovnání pomocí čtyřčlenných centrovaných klouzavých průměrů.



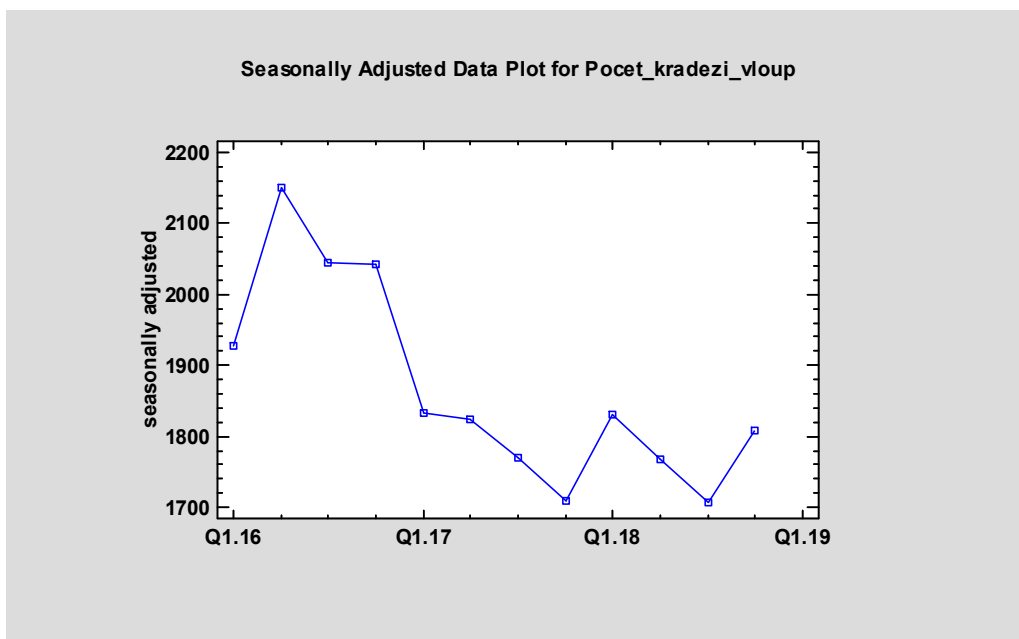
Obrázek 8 – Vyrovnání časové řady krádeží vloupáním pomocí čtyřčlenných centrovaných klouzavých průměrů

Obr. 9 pak zachycuje grafické znázornění rozdílových sezónních faktorů. Je zde velmi přehledně zobrazeno, kdy jsou hodnoty časové řady v průměru pod dlouhodobým normálem (první a druhé čtvrtletí) a kdy nad ním (třetí a čtvrté čtvrtletí).

Poslední Obr. 10 ukazuje sezónně očištěná data, tj. jak by data o krádežích vloupáním vypadala, kdyby zde nepůsobily sezónní vlivy.



Obrázek 9 – Rozdílové sezónní faktory



Obrázek 10 – Sezónně očištěná data počtu krádeží vloupáním v Libereckém kraji