SPC v případě autokorelovaných dat

Jiří Michálek, Jan Král

OSSM, 21.6.2012

Pojem korelace

- Statistická vazba mezi veličinami
- Korelace vs. stochastická nezávislost
- Koeficient korelace = míra lineární vazby mezi veličinami
- Autokorelace = korelace mezi sousedícími daty v jedné časové řadě
- Autokorelační funkce, odhad autokorelace pořízený z dat

Předpoklady pro Shewhartovy diagramy

- Normálně rozdělený znak jakosti
- Sbíraná data z procesu jsou nezávislá jak mezi podskupinami, tak i uvnitř podskupin
- Parametr polohy a úroveň variability jsou konstantní
- Regulační meze jsou pak počítány na základě těchto předpokladů

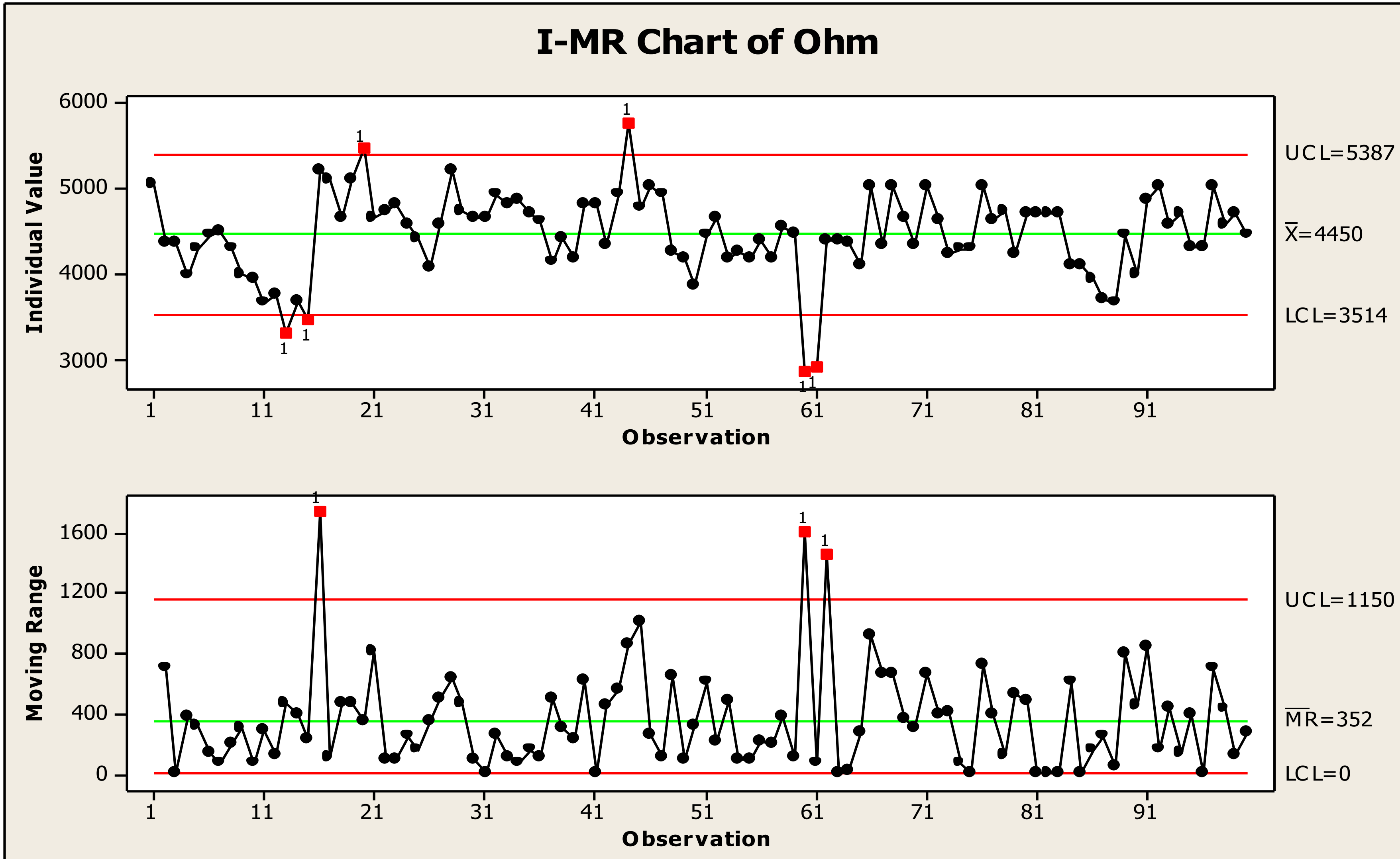
Porušení předpokladů

- Závislost mezi daty v rámci podskupiny – může mít vliv na vyhodnocování průběhu diagramu, odhad směrodatné odchylky pro výpočet regulačních mezí
- Závislost mezi podskupinami – může podstatně ovlivnit výskyt falešných poplachů

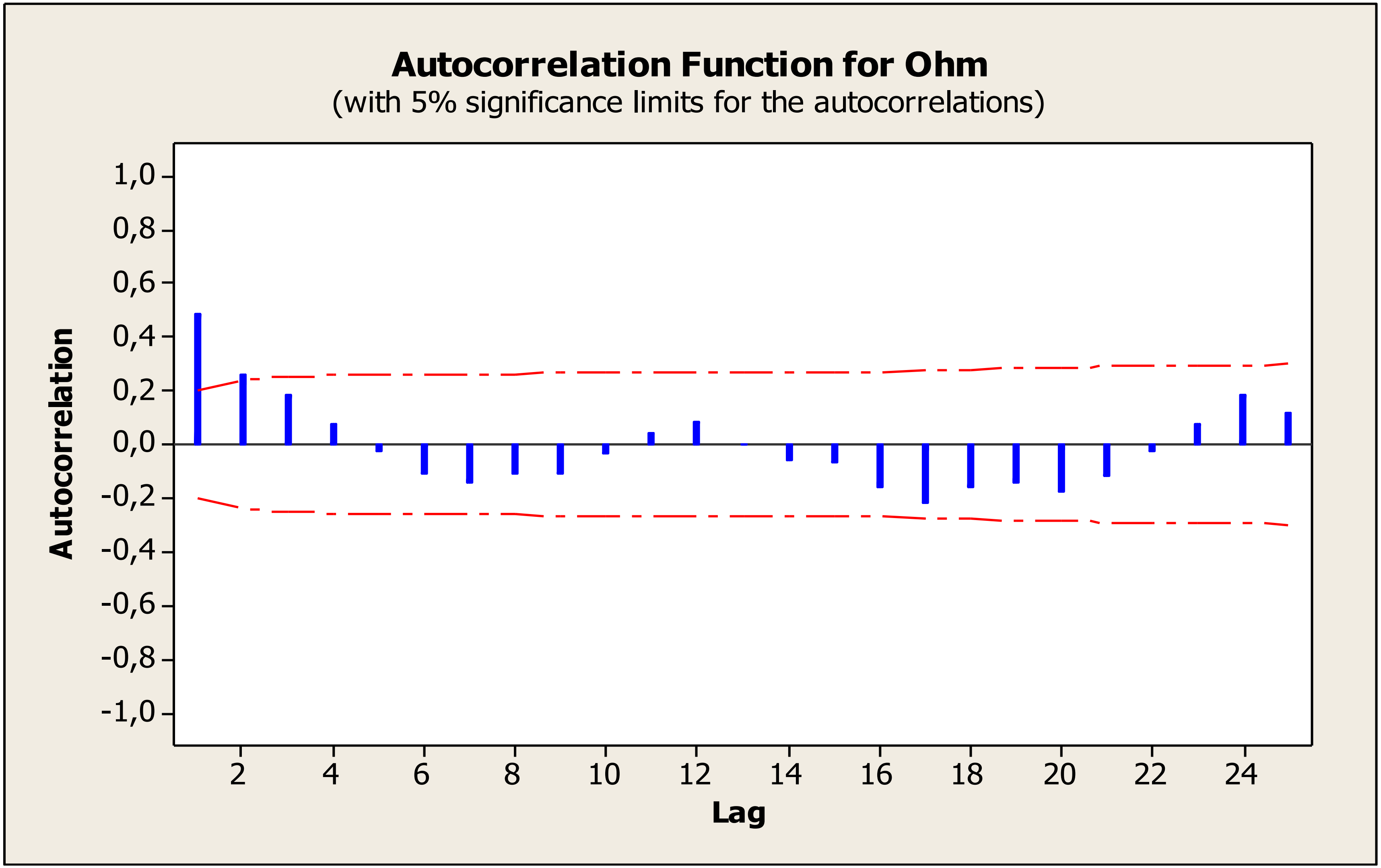
Podezření na autokorelaci

- Vysoká frekvence sběru dat, hlavně při automatickém odběru
- Fyzikální důvody – např. vliv teploty, vliv okolního prostředí
- Setrvačnost v datech vyvolaná např. způsobem měření, systematickými vlivy
- Záměrná manipulace s daty – opisování, falšování

Příklad regulačního diagramu



Odhad autokorelační funkce



Co říká koeficient korelace?

- Koeficient korelace je mezi -1 a 1
- Čím je koeficient korelace v absolutní hodnotě blíže k 1, tím je větší stochastická vazba mezi veličinami
- Čím je koeficient korelace v absolutní hodnotě blíže 0, tím je menší stochastická vazba mezi veličinami
- Pro hodnoty -1 a 1 stochastická vazba přechází dokonce ve funkční lineární vztah

Test korelace mezi 2 veličinami

- Data jsou spárována do dvojic
- Vypočte se odhad *r* koeficientu korelace
- Vypočte se statistika

$$t = r\sqrt{(n-2)} \; / \; \sqrt{(1-r^2)} \quad ,$$

která za nulové hypotézy o nekorelovanosti má přibližně t-rozdělení o n-2 stupňů volnosti

(n je počet dvojic)

Pozn. U normálně rozdělených veličin nekorelovanost = nezávislost

Jak se s korelací vyrovnat?

- Pro praktické účely má cenu se zabývat korelací, pokud odhad koeficientu korelace má hodnotu nad 0,75 – 0,80
- Znaky kvality na jednom výrobku mohou být vzájemně korelovány – vylepšení jednoho znaku může vyvolat zhoršení jiného znaku
- Použití vícerozměrných regulačních diagramů – odhad korelační matice

Jak se s korelací vyrovnat?

- Při zjištěné silné korelaci uvnitř podskupin

je nutno toto zohlednit při výpočtu regulačních mezí – zásadní roli hraje špatný odhad směrodatné odchylky na základě výběrového rozpětí či výběrové směrodatné odchylky počítaných uvnitř podskupin

Korelace může způsobit podhodnocení úrovně variability a toto může vést ke zvýšenému počtu falešných poplachů, protože regulační meze jsou pak blíže u sebe

Jak se s korelací vyrovnat?

- Zvláště je nutné si dát pozor u regulačního diagramu pro individuální data při sériové korelaci – odhad založený na klouzavém rozpětí opět může podhodnotit úroveň variability
- V tomto případě v praxi stačí se zaměřit pouze na autokorelaci 1.řádu a použít faktor $\sqrt{(1 - r^2)}$ pro úpravu regulačních mezí

Jak se s korelací vyrovnat?

- Pro výpočet klasických regulačních mezí nepoužijeme *Rbar* či *sbar*, ale jejich hodnoty získané z dat podělíme právě faktorem $\sqrt{(1 - r^2)}$, čímž dosáhneme rozšíření regulačních mezí
- Pokud jsou mezi daty autokorelace vyšších řádů, je obvykle možno v praxi je ignorovat

Jak se vyrovnat s korelací?

- Komplikovanější přístup je založen na nalezení vhodného modelu pro sledovanou časovou řadu (AR, ARMA, ARIMA či další modely) a pracovat s rezidui získanými při použití vhodného modelu
- Obvykle se vyžaduje, aby rezidua byla vzájemně nekorelovaná se střední hodnotou nula a konstantní úrovní variability (či dokonce normálně rozdělená) a na jejich sledování pak lze použít klasické Shewhartovy regulační diagramy

Příklad na úpravu mezí

- Na následujícím diagramu je vidět, že první 3-4 zleva, které jsou mimo regulační meze, by mohly znamenat pouze falěšný poplach
- Je nutné prošetřit autokorelaci a zjistíme, že korelační koeficient mezi sousedícími daty bude významný, jeho odhad je sice pouze 0,48, ale je mimo konfidenční meze

Příklad na úpravu mezí

- Původní meze pro individuální hodnoty:

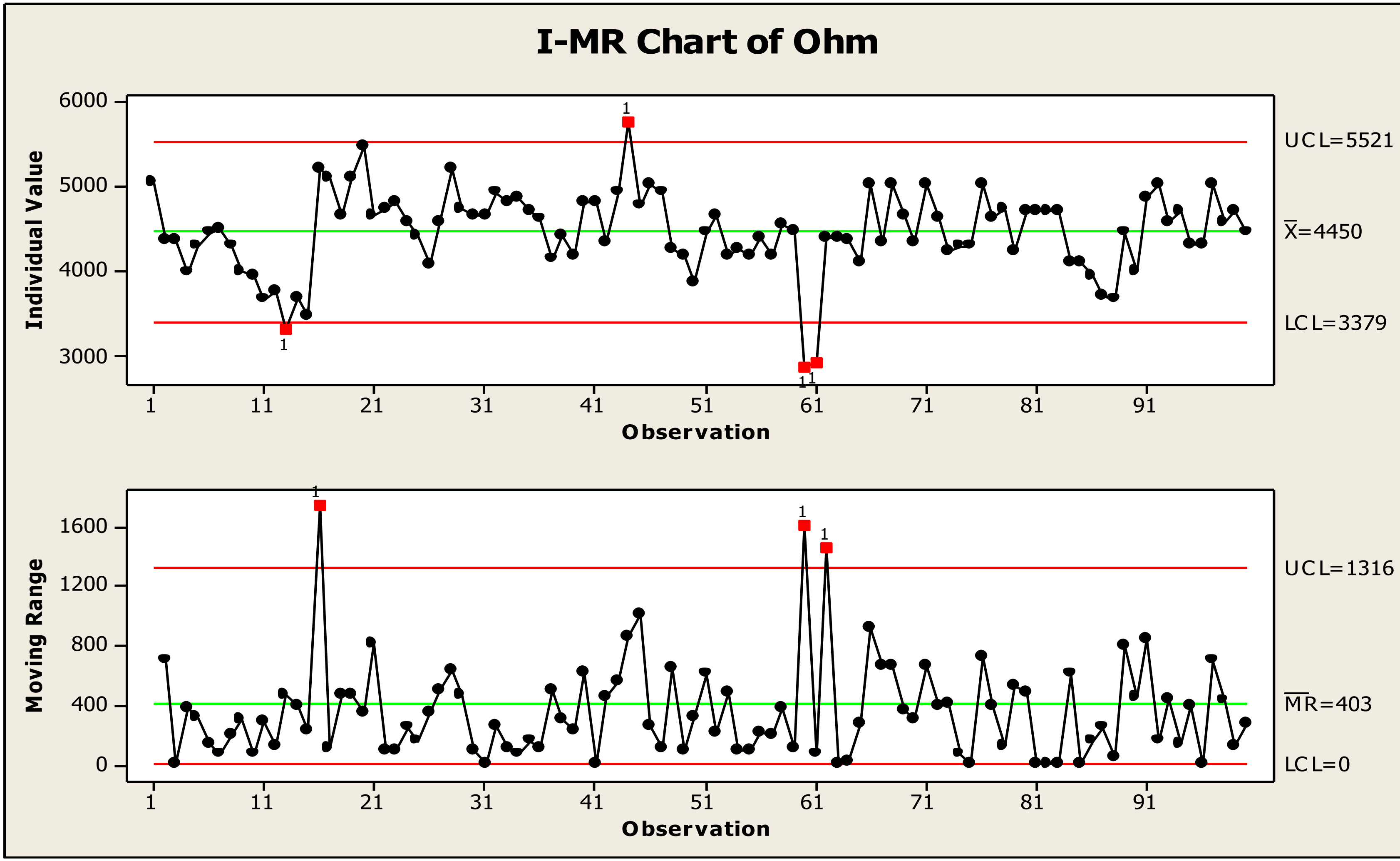
  UCL = 5387, LCL = 3514

  CL = 4450

  Faktor opravy je při r = 0,4845 roven 0,8748 a přepočtené meze jsou:

  UCL = 5521, LCL = 3379

Graf s upravenými mezemi



Použití modelu časové řady

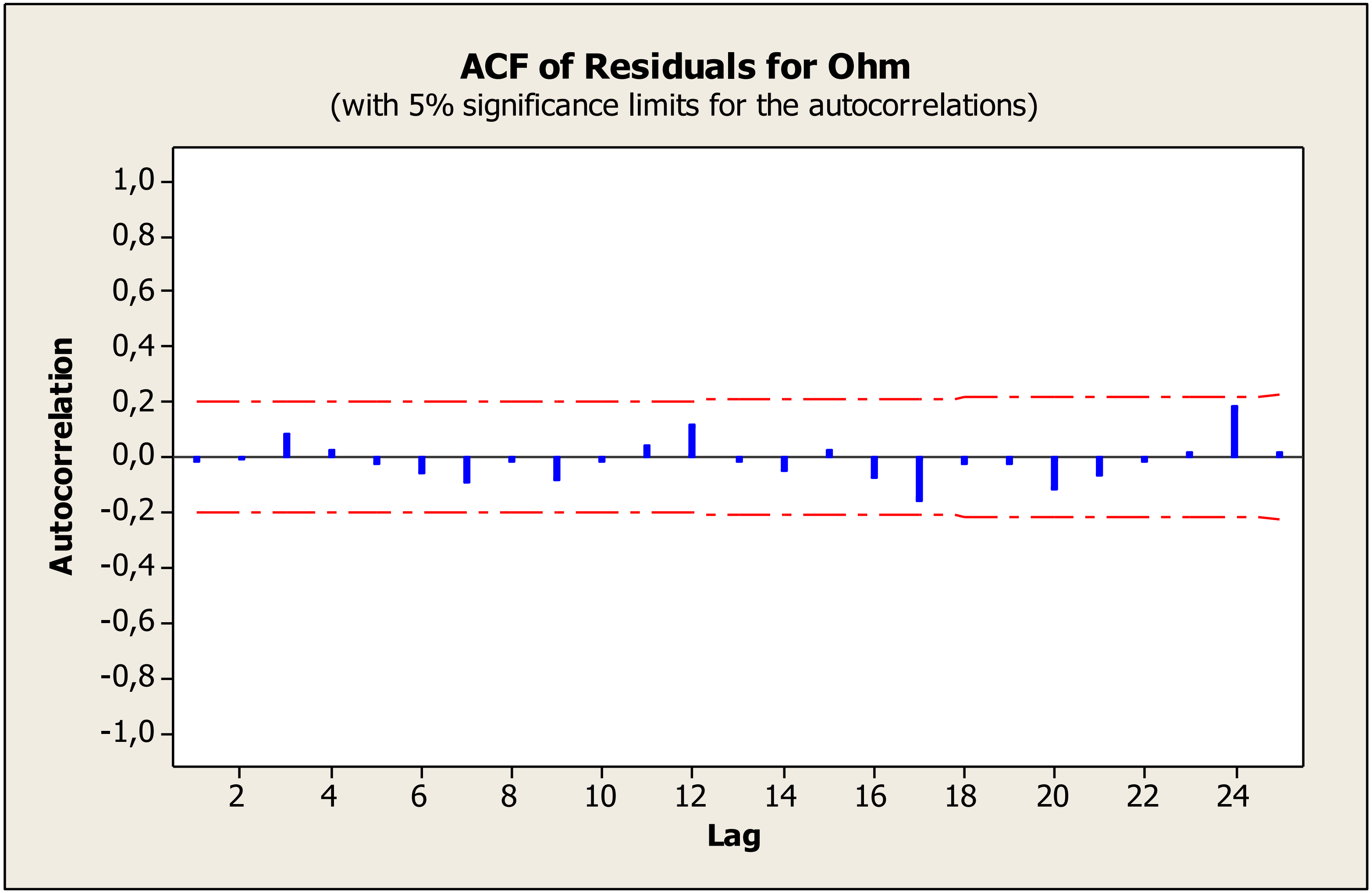
- Na základě odhadu autokorelační funkce budeme uvažovat model typu AR(1):

$$X(i+1) = \mu + \alpha X(i) + e(i+1),$$

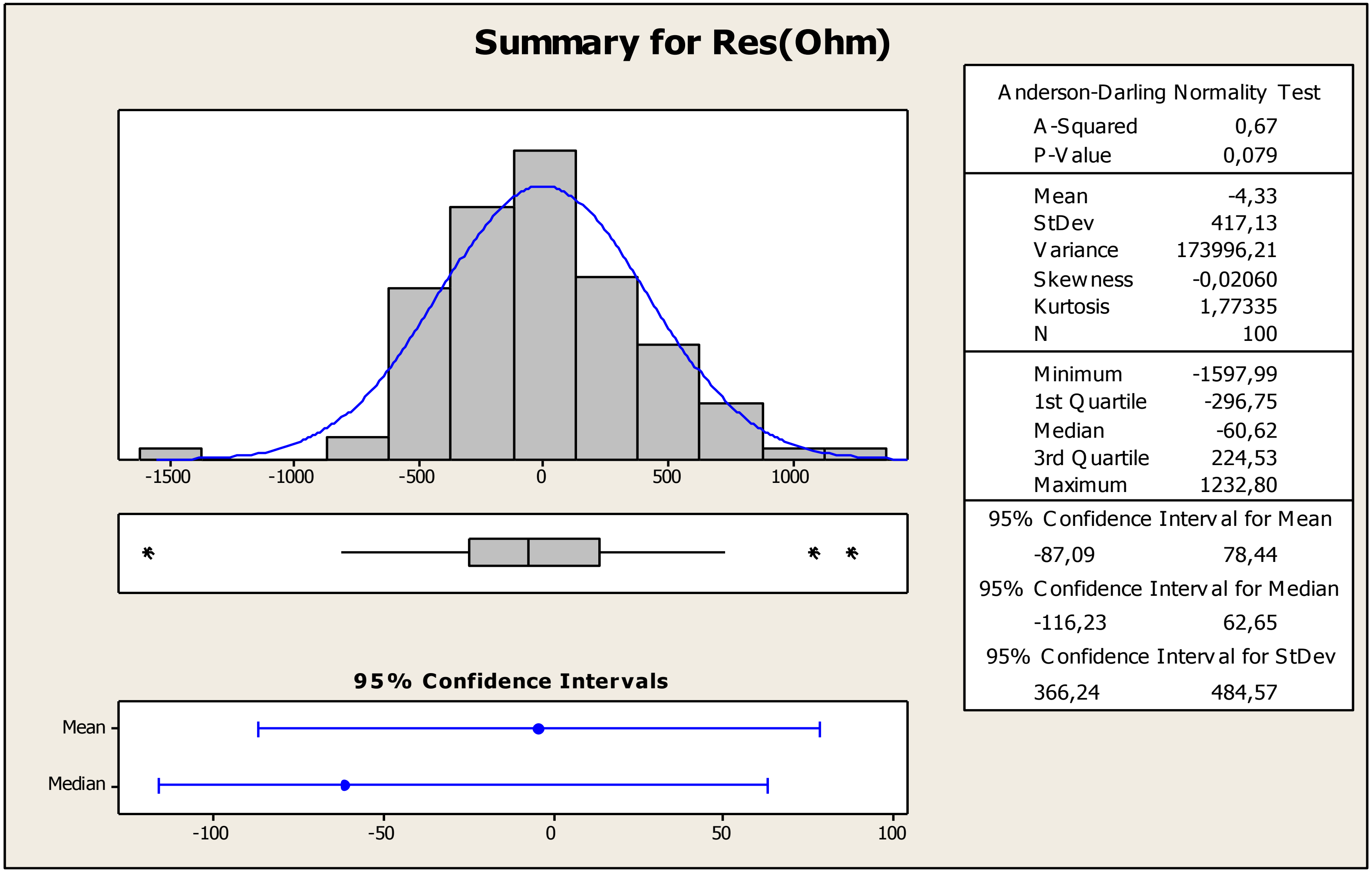
kde chyby jsou normálně rozdělené se střední hodnotou 0 a nezávislé, μ je aditivní konstanta

Vhodnost modelu posoudíme podle chování reziduí a průběh původního procesu pomocí regulačního diagramu pro rezidua

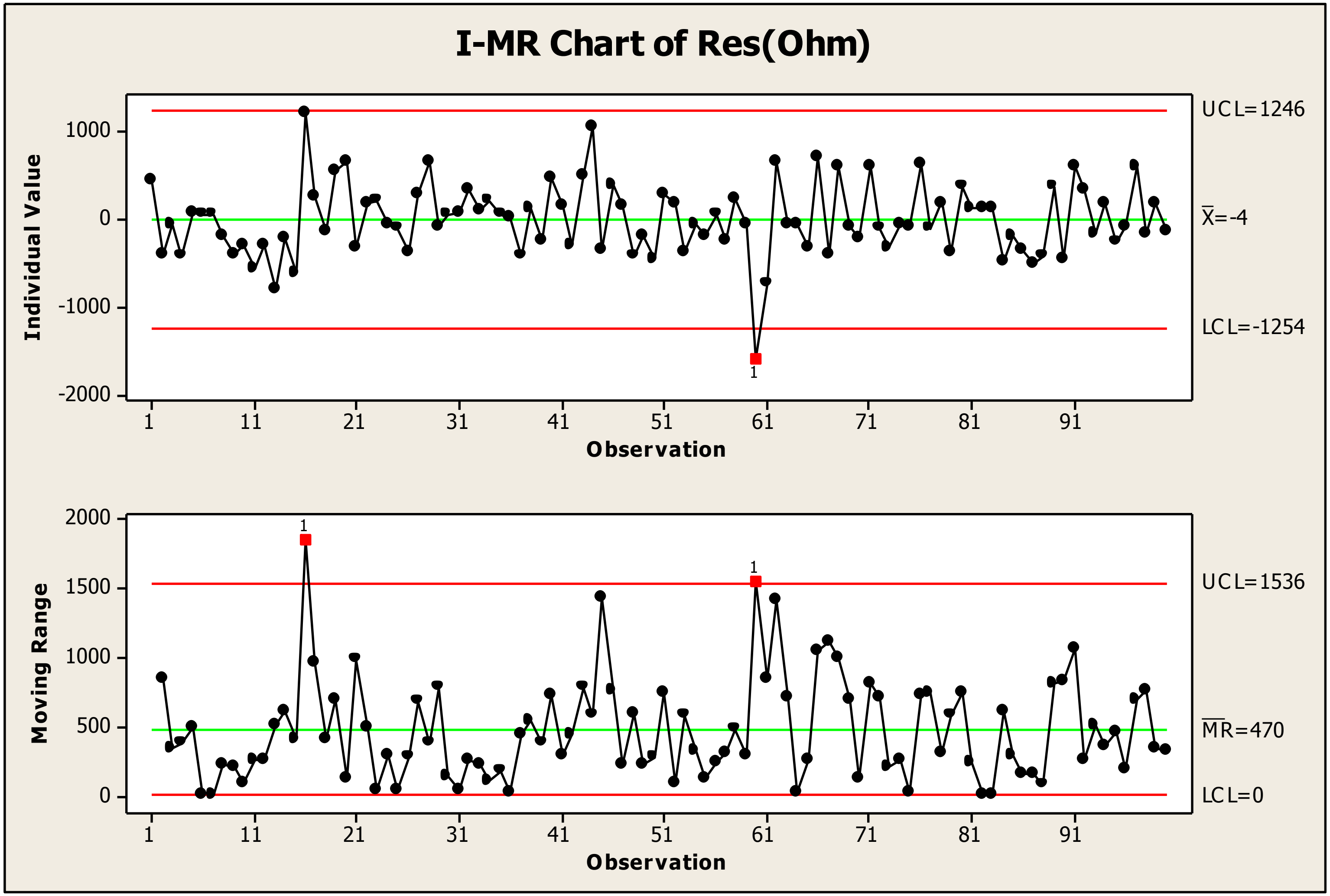
Použití modelu časové řady



Použití modelu časové řady



Použití modelu časové řady



Použití modelu časové řady

Nalezený model má tvar:

$$X(i+1) = 2262{,}73 + 0{,}4922X(i) + e(i+1),$$

kde rezidua jakožto odhady chyb mají rozdělení $N(\mu,\sigma^2)$.

$$\mu \approx -4{,}33 \qquad \sigma \approx 417{,}13$$

Srovnání obou přístupů

Lze říci, že přístup založený na úpravě regulačních mezí se zdá přísnější, ale v žádném případě si oba přístupy neodporují a možné ovlivněné podskupiny označily podobně.

Je zřejmé, že přístup založený na úpravě mezí je daleko jednodušší, přístup založený na modelování časové řady vyžaduje již vhodný software.