

Vícerozměrné regulační diagramy

Josef Křepela, Jiří Michálek

OSSM

22.11.2012

Monitorování a řízení procesu s více proměnnými

Obvykle se uvažuje pouze jeden znak jakosti (proměnná, náhodná veličina) na výstupu procesu, která se monitoruje (sleduje) a řídí pomocí statistických regulačních diagramů. Často se však setkáváme s procesy, kde na výstupu je několik znaků jakosti (proměnných). V tomto případě je možné aplikovat na každou z nich samostatný regulační diagram, což je však zbytečné a někdy dokonce může vést k chybným závěrům. Proto jsou aplikovány metody, které berou v úvahu všechny proměnné společně. Budeme se zabývat regulačními diagramy, které mohou být považovány za rozšíření regulačních diagramů pro jednu proměnnou. Takovými regulačními diagramy jsou Hotellingovy T² regulační diagramy, které jsou analogií Shewhartových diagramů.

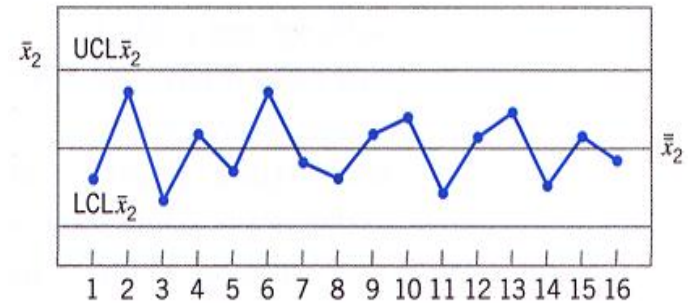
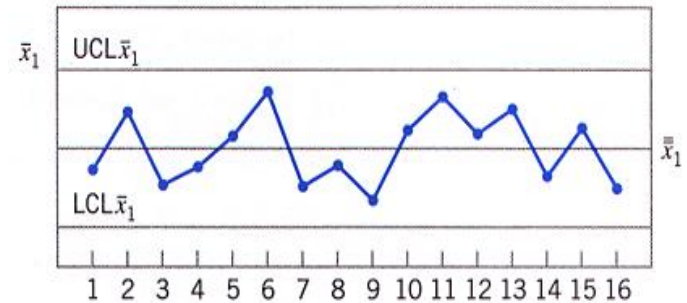
Případ dvou znaků

Uvažujme například ložisko, kde předmětem našeho zájmu je jak vnější průměr (x_1), tak i vnitřní průměr (x_2). Předpokládejme, že jak x_1 tak x_2 jsou nezávislé proměnné a mají normální rozdělení. Obě proměnné můžeme monitorovat pomocí obvyklého regulačního - diagramu. Proces je statisticky zvládnut (pod kontrolou) pouze, pokud výběrové průměry obou rozměrů, padnou do příslušných regulačních mezí; viz. Obr. 1. To je ale analogické tomu, pokud dvojice výběrových průměrů (,) padne do vyznačené oblasti na Obr. 2.

Případ dvou znaků

Regulační diagramy pro průměry obr.

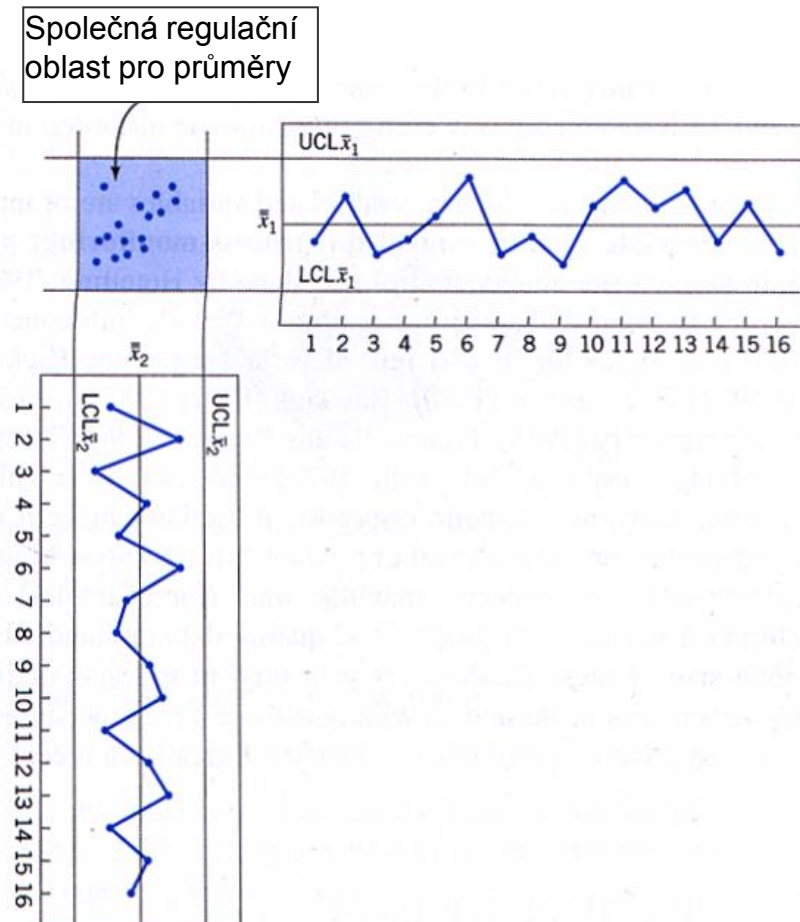
Oba znaky jakosti jsou zvládnuty, pod kontrolou



Případ dvou znaků

Obr.2

za předpokladu stochastické nezávislosti



Případ dvou znaků

Pravděpodobnost, že každý z nich padne mimo tři-sigma meze je 0,0027. Nicméně společná pravděpodobnost, že oba výběrové průměry překročí své regulační meze současně, když obě proměnné jsou statisticky zvládnuté, je

$$0,0027 \times 0,0027 = 0,00000729,$$

což je výrazně méně než 0,0027. Pravděpodobnost, že oba výběrové průměry budou současně ležet uvnitř regulačních mezí, když proces je statisticky zvládnut, je pouze $0,9973 \times 0,9973 = 0,99460729$. Použití dvou nezávislých diagramů deformuje současné sledování obou znaků jakosti.

Případ více znaků

Obecně, pokud máme p statisticky nezávislých proměnných pro určitý produkt a uvažujeme-li pro každý z nich -diagram s pravděpodobností I druhu α (riziko planého poplachu), potom skutečná pravděpodobnost chyby I druhu pro společný kontrolní postup je

$$\alpha' = 1 - (1 - \alpha)^p$$

a pravděpodobnost, že všech p výběrových průměrů bude současně ležet uvnitř regulačních mezí, když proces je statisticky zvládnut, je

$P_{\text{st}} \{\text{všech } p \text{ výběrových průměrů leží v regulačních mezích}\} = (1 - \alpha)^p$.

Případ více znaků

Problém monitorování procesu, ve kterém je předmětem zájmu několik proměnných (znaků jakosti) se někdy nazývá "řízení jakosti více proměnných" (*multivariate quality control*), nebo "monitorování procesu s více proměnnými" (*multivariate process-monitoring*). Základní práce v tomto směru učinil Hotelling (2) který tento postup aplikoval na data ze zaměřovačů pro shazování bomb za druhé světové války.

Vícerozměrné normální rozdělení

Předpokládejme, že máme p proměnných x_1, x_2, \dots, x_p .
Uspořádejme tyto proměnné do řádkového vektoru

$$\mathbf{x}' = [x_1, x_2, \dots, x_p].$$

Nechť $\boldsymbol{\mu}' = [\mu_1, \mu_2, \dots, \mu_p]$ je vektor středních hodnot uvažovaných proměnných a necht' rozptyly a kovariance proměnných v \mathbf{x} jsou obsaženy v $p \times p$ **kovarianční matici** Σ . Elementy na diagonále Σ jsou rozptyly proměnných v \mathbf{x} a elementy mimo diagonálu jsou kovariance. Potom kvadrát normované vzdálenosti \mathbf{x} od $\boldsymbol{\mu}$ je dán kvadratickou formou

$$(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) .$$

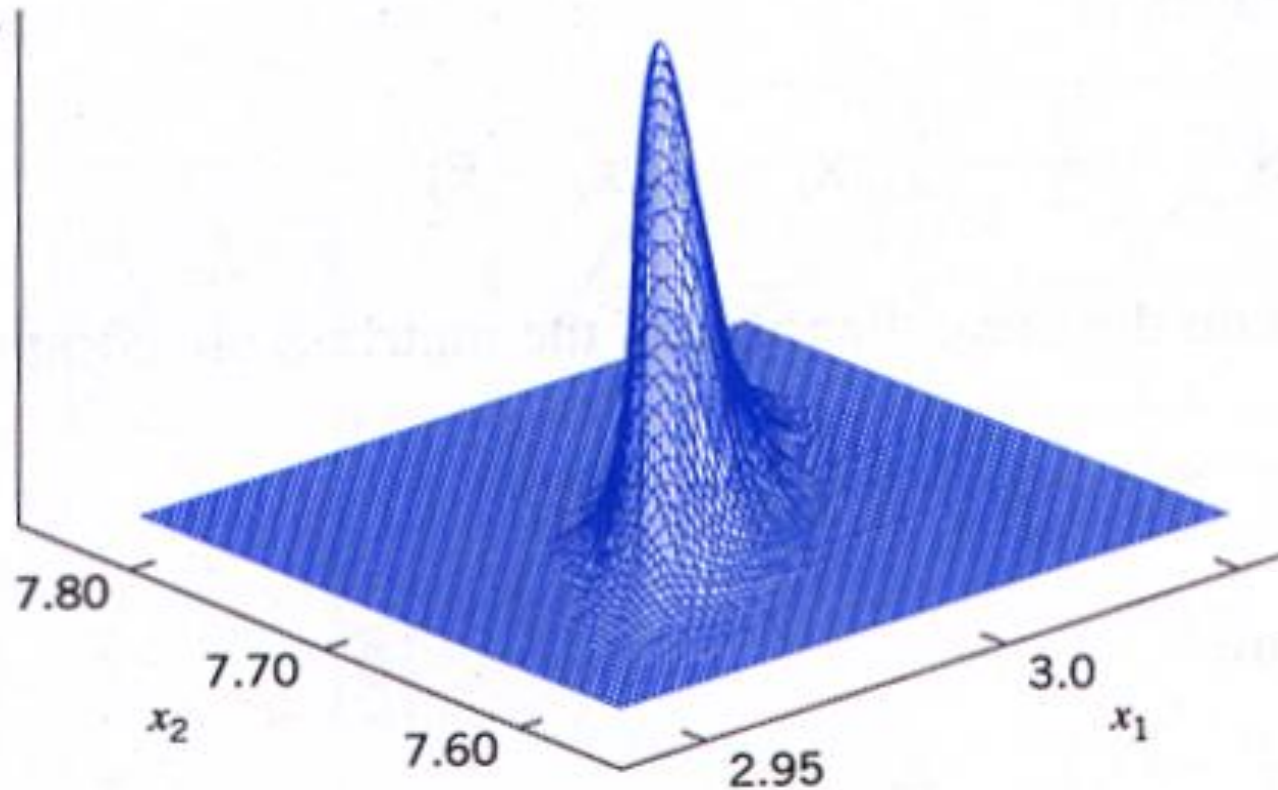
Vícerozměrné normální rozdělení

Vzorec pro hustotu rozdělení (kovarianční matice musí být regulární)

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

DVOUROZMĚRNÉ NORMÁLNÍ ROZDĚLENÍ

Příklad dvourozměrné hustoty



Odhady parametrů rozdělení

Uvažujme náhodný výběr z vícerozměrného normálního rozdělení $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$.

Vektor \mathbf{x}_i (i-tý výběrový vektor) obsahuje pozorování $x_{i1}, x_{i2}, \dots, x_{ip}$. Potom vektor výběrových průměrů je $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$

a výběrová kovarianční matice je

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

Hotellingovy T^2 -diagramy

Nejrozšířenější postup monitorování a řízení procesu s více proměnnými (vícerozměrný proces) jsou Hotellingovy T^2 regulační diagramy pro sledování vektoru výběrových průměrů (mean vector) procesu. Je to analogie Shewhartových regulačních diagramů pro jednu proměnnou. Budeme uvažovat dvě verze Hotellingova T^2 diagramu, jednu pro data seskupená do podskupin, druhou pro individuální pozorování.

Hotellingovy T^2 -diagramy

Data seskupená do podskupin

Předpokládejme, že dva znaky jakosti, proměnné x_1 a x_2 , mají společně dvojrozměrné normální rozdělení. Necht' jsou μ_1 a μ_2 střední hodnoty, σ_1 a σ_2 směrodatné odchylky proměnných x_1 a x_2 . Kovariance mezi x_1 a x_2 je σ_{12} . Předpokládejme, že σ_1 , σ_2 a σ_{12} jsou známy. Potom statistika

$$\chi_0^2 = \frac{n}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \left[\sigma_2^2 (\bar{x}_1 - \mu_1)^2 + \sigma_1^2 (\bar{x}_2 - \mu_2)^2 - 2\sigma_{12} (\bar{x}_1 - \mu_1)(\bar{x}_2 - \mu_2) \right]$$

má χ^2 -rozdělení o dvou stupních volnosti.

Hotellingovy T^2 -diagramy

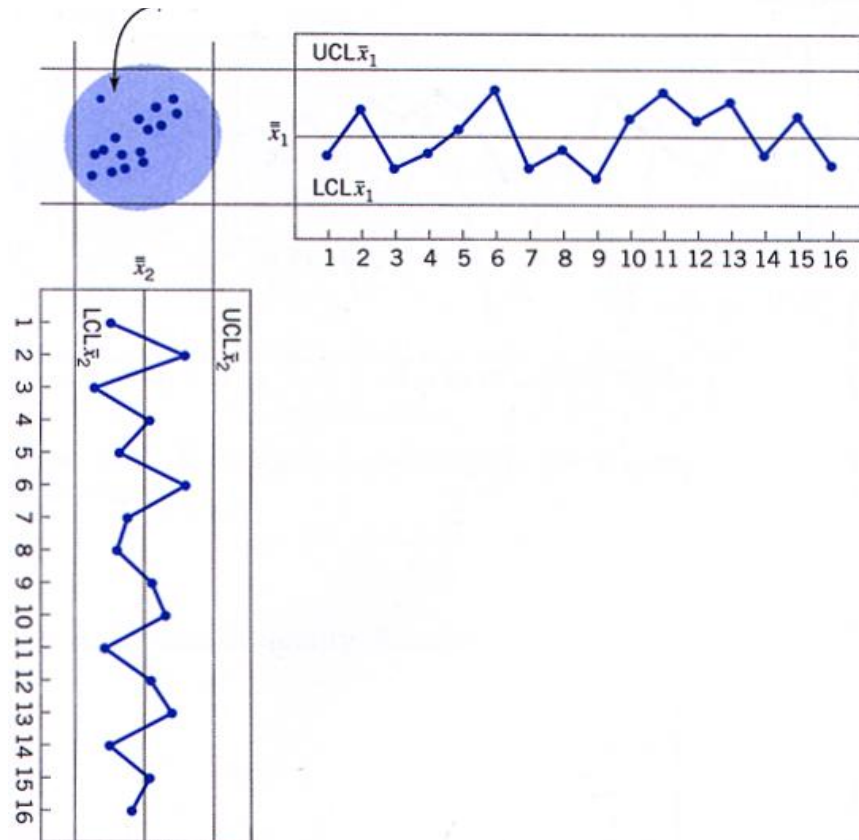
Předchozí rovnice může být použita jako základ regulačního diagramu pro střední hodnoty procesu. Pokud střední hodnoty procesu budou rovny hodnotám μ_1 a μ_2 , potom hodnota bude menší než horní regulační mez $UCL = \chi_{\alpha,2}$, to je α -kritická hodnota chi-kvadrát rozdělení se 2 stupni volnosti. Pokud nejméně jedna ze středních hodnot se změní, nabude jiné hodnoty (dostane se mimo kontrolu), potom pravděpodobnost, že statistika překročí horní regulační mez, vzroste.

Hotellingovy T^2 -diagramy

Procedura monitorování procesu může být znázorněna graficky. Uvažujme případ, kdy dvě náhodné proměnné jsou nezávislé, tj. $\sigma_{12} = 0$, potom rovnice (11) definuje elipsu se středem v (μ_1, μ_2) a s hlavními osami rovnoběžnými s osami, viz Obr. 4. Pokud hodnota padne dovnitř této elipsy, znamená to, že proces je statisticky zvládnut, zatímco padne-li mimo tuto elipsu, proces je mimo kontrolu. Obrázek 4. je často nazýván "regulační elipsa" (*control ellipse*).

Hotellingovy T^2 -diagramy

Společná regulační
oblast pro průměry



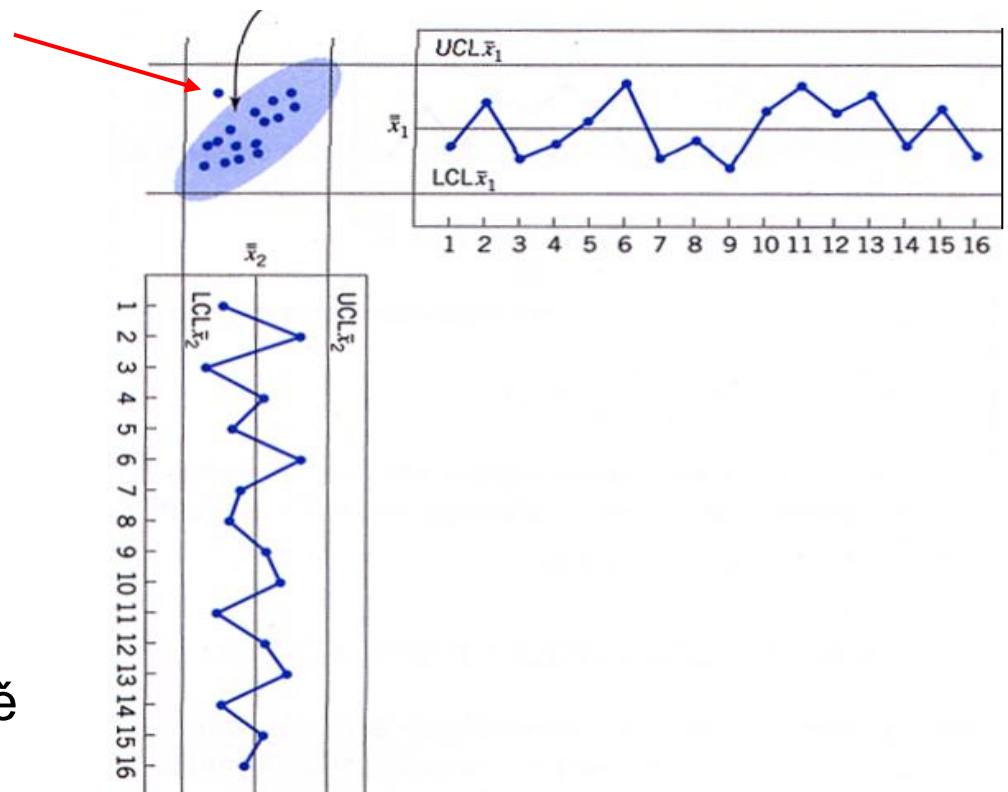
Obr. 4
Kontrolní elipsa pro dvě
nezávislé proměnné

Hotellingovy T^2 -diagramy

Pokud dvě náhodné proměnné jsou závislé, potom $\sigma_{12} \neq 0$ a odpovídající "regulační elipsa" je znázorněna na Obr. 5. V tomto případě hlavní osy elipsy nejsou rovnoběžné s osami.. Výběrový bod 11, který padl mimo "regulační elipsu" signalizuje přítomnost zvláštní příčiny variability. Nicméně bod 11 je uvnitř regulačních mezí na obou regulačních diagramech. Podle všeho není nic mimořádného okolo bodu 11 (podskupiny 11), pokud ho posuzujeme individuálně v každé souřadnici.

Hotellingovy T^2 -diagramy

Společná regulační oblast
pro

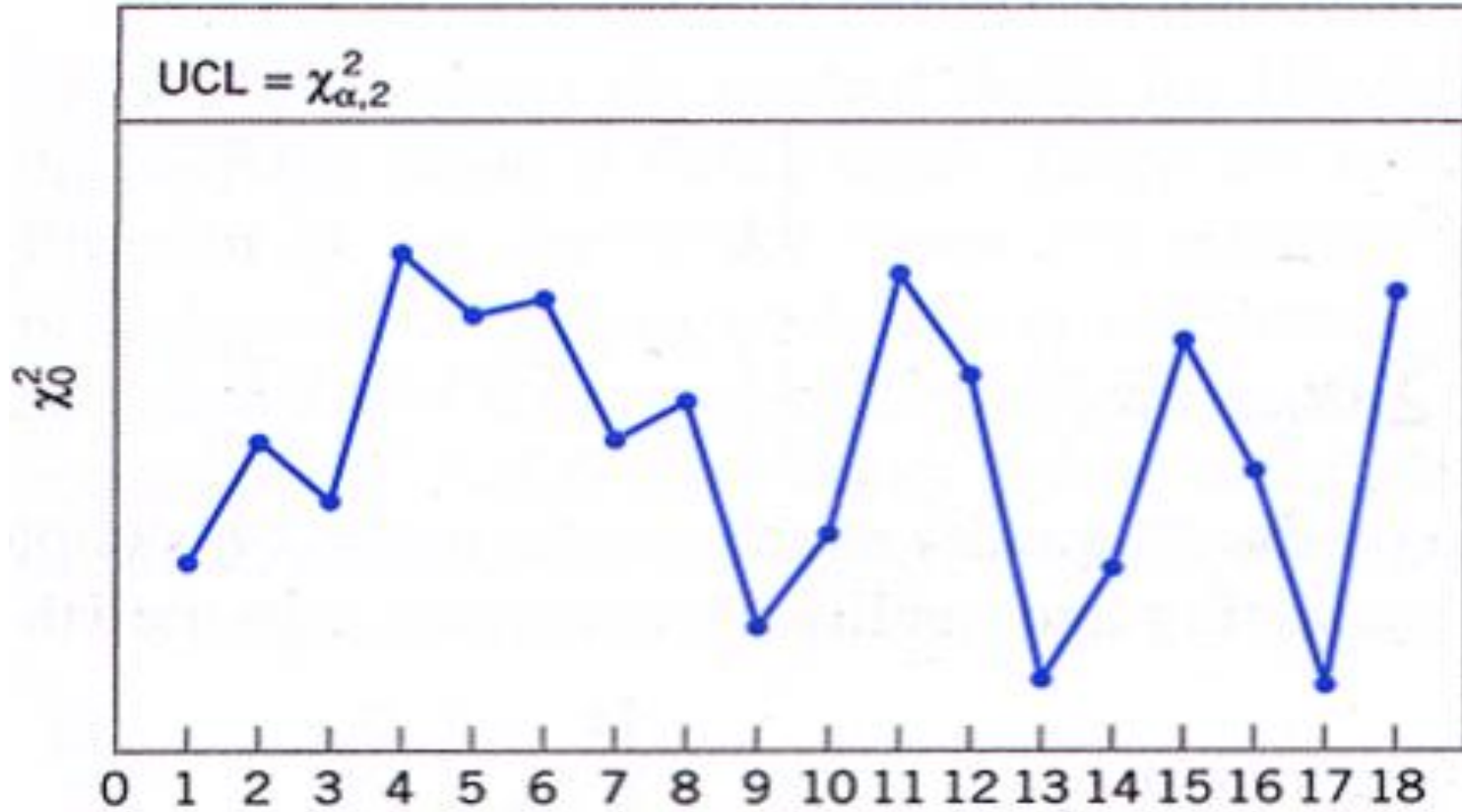


Obr. 5
Kontrolní elipsa pro dvě
závislé proměnné

Hotellingovy T^2 -diagramy

S kontrolními elipsami jsou spojeny dva podstatné nedostatky. **První** spočívá v tom, že se ztrácí časová následnost zakreslovaných bodů. **Druhá**, daleko závažnější, spočívá v tom, že je nesnadné konstruovat elipsu pro více jak dva znaky jakosti. Abychom se vyhnuli těmto nedostatkům vynášíme obvykle do regulačního diagramu z každého výběru vypočítané hodnoty (podle vzorce), které porovnáваме s horní regulační mezí UCL, tak jak je uvedeno v Obr. 6. Tento regulační diagram se obvykle nazývá "chi-kvadrát diagram" (*chi-square control chart*). V tomto případě je zachována časová souslednost, takže mohou být analyzovány všechny posloupnosti a nenáhodná seskupení. To má i tu výhodu, že stav procesu je charakterizován jedním číslem - hodnotou statistiky .

χ^2 -diagram



χ^2 -diagram

Tento výsledek je možno rozšířit na případ, kdy uvažujeme p znaků jakosti, které jsou ve vztahu a jsou kontrolovány společně. Předpokládá se, že společné rozdělení p proměnných je p -rozměrné normální rozdělení. Postup vyžaduje vypočítat výběrové průměry pro každou proměnnou z výběru rozsahu n .

χ^2 -diagram

Tento soubor výběrových průměrů je representován $(p \times 1)$ -vektorem

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}$$

Testová statistika zakreslovaná do chi-kvadrát diagramu pro každý výběr je

$$\chi_0^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$$

χ^2 -diagram

kde $\mu' = [\mu_1, \mu_2, \dots, \mu_p]$ je vektor středních hodnot statisticky zvládnutého procesu (*in-control means*) a Σ je kovarianční matice.

Horní regulační mez je

$$UCL = \chi_{\alpha, n}.$$

Pro tento diagram je nutno znát parametry rozdělení.

Odhady parametrů

V praxi je obvykle nezbytné odhadnout μ a Σ z předběžné analýzy náhodných výběrů rozsahu n vzatých z procesu, o kterém je možno předpokládat, že je statisticky zvládnut (pod kontrolou). Předpokládejme, že je takových výběrů m . Z každého výběru se vypočítá výběrový průměr a rozptyl.

Odhady parametrů

X_{ijk} je i -té pozorování j -tého znaku v k -tém výběru

$$\bar{x}_{jk} = \frac{1}{n} \sum_{i=1}^n x_{ijk}$$

$$S_{jk}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{x}_{jk})^2$$

kovariance mezi znaky jakosti j a h v k -tém výběru je

$$S_{jhk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{x}_{jk})(x_{ihk} - \bar{x}_{hk})$$

Odhady parametrů

Statistiky \bar{x}_{jk} , S_{jk}^2 a S_{jhk} jsou potom zprůměrovány přes všech m provedených výběrů

$$\bar{x}_j = \frac{1}{m} \sum_{k=1}^m \bar{x}_{jk} \quad j = 1, 2, \dots, p$$

$$\bar{S}_j^2 = \frac{1}{m} \sum_{k=1}^m S_{jk}^2 \quad j = 1, 2, \dots, p$$

$$\bar{S}_{jh} = \frac{1}{m} \sum_{k=1}^m S_{jhk} \quad j \neq h$$

\bar{x}_j jsou složky vektoru $\bar{\mathbf{x}}$

Odhady parametrů

Symetrická matice \mathbf{S} je tvořena z průměrných kovariancí \bar{S}_{jh} a rozptylů \bar{S}_{jj}

$$\mathbf{S} = \begin{bmatrix} \bar{S}_1^2 & \bar{S}_{12} & \bar{S}_{13} & \cdots & \bar{S}_{1p} \\ & \bar{S}_2^2 & \bar{S}_{23} & \cdots & \bar{S}_{2p} \\ & & \bar{S}_3^2 & \cdots & \bar{S}_{3p} \\ & & & \ddots & \vdots \\ & & & & \bar{S}_p^2 \end{bmatrix}$$

Hotellingova T^2 -statistika

Nyní předpokládejme, že matice \mathbf{S} definovaná rovnicí (18) se použije k odhadu Σ a že vektor $\bar{\mathbf{x}}$ je použit jako odhad vektoru μ středních hodnot statisticky zvládnutého procesu. Jestliže nahradíme μ hodnotou $\bar{\mathbf{x}}$ a Σ hodnotou \mathbf{S} v rovnici pro χ^2 dostaneme testovou statistiku T^2 .

Hotellingova T^2 -statistika

Vzorec pro T^2 -statistiku

$$T^2 = n(\bar{\mathbf{x}} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{x}})$$

V tomto tvaru je procedura obvykle nazývána Hotellingovy T^2 regulační diagramy.

Je třeba věnovat pozornost stanovení regulačních mezí pro tuto Hotellingovu T^2 statistiku. Ty závisí na tom, jak se diagramy použijí. Jsou dvě odlišné fáze použití těchto regulačních diagramů.

Výpočet regulačních mezí

Fáze 1 předpokládá použití regulačních diagramů pro zavedení kontroly; tj. testovat, zda proces byl statisticky zvládnut (byl pod kontrolou) když bylo odebráno m podskupin rozsahu n a byly vypočítány výběrové statistiky, které byly použity pro odhady parametrů.

Výpočet regulačních mezí

Fáze 1

Cílem fáze 1 je získat soubor statisticky zvládnutých pozorování (in-control set of observations), tak aby mohly být stanoveny regulační meze pro Fázi 2, která spočívá v monitorování budoucí produkce. Fáze 1 se někdy nazývá "retrospektivní analýzou". Regulační meze pro tuto fázi jsou dány vztahy:

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha, p, mn-m-p+1} \quad LCL = 0$$

Výpočet regulačních mezí

Fáze 2

Tato fáze předpokládá použití regulačních diagramů k monitorování budoucí produkce, regulační meze jsou pro tuto fázi následující:

$$UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha, p, mn-m-p+1}$$

$$LCL = 0$$

Výpočet regulačních mezí

Pokud jsou odhady parametrů spolehlivé (např. na základě dlouhodobé stability procesu), lze horní regulační mez nahradit

$$UCL = \chi^2_{\alpha,p}$$

($\alpha 100\%$ - kritická hodnota)

Lze použít jako horní regulační mez jak pro Fázi 1, tak pro Fázi 2. Stejně jako v případě jedné proměnné (jednoho znaku jakosti) se doporučuje pracovat s nejméně $m \geq 25$ předběžnými výběry. Rozlišení mezi mezemi pro Fázi 1 a Fázi 2 je obvykle nepodstatné.

Výpočet regulačních mezí

Individuální pozorování

V řadě případů je velikost podskupiny $n = 1$. Předpokládejme m podskupin a p znaků jakosti. Necht' \mathbf{x} a \mathbf{S} jsou vektor výběrových průměrů a kovarianční matice. Hotellingova T^2 statistika je potom ve tvaru

$$T^2 = (\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}})$$

Regulační meze pro Fázi 2 jsou

$$UCL = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha, p, m-p}$$

$$LCL = 0$$

Výpočet regulačních mezí

Velkým problémem v případě individuálních hodnot je spolehlivý odhad kovarianční matice. V literatuře lze najít řadu různých doporučení, jak odhady vylepšit (např. pomocí rozdílů mezi po sobě jdoucími pozorováními). Kvalita použitého odhadu silně ovlivňuje hodnotu horní regulační meze.

Použití regulačních diagramů

Výhoda: je možno sledovat několik znaků jakosti současně na základě jejich možné korelovanosti, což individuální diagramy nemohou.

Nevýhoda: při detekci zvláštní příčiny, nemusí být jasné, kterých znaků jakosti se příčina týká. Tomuto problému se literatura velice usilovně věnuje, jak vlivy jednotlivých znaků rozpoznat.

Monitoring variability

Stejně tak, jako je třeba monitorovat vektor středních hodnot μ , je třeba monitorovat i variabilitu procesu. Variabilita procesu je charakterizována $p \times p$ maticí kovariancí Σ . Prvky na hlavní diagonále této matice jsou rozptyly jednotlivých proměnných procesu a prvky mimo diagonálu jsou kovariance.

Monitoring variability

Často používaný přístup je založen na "zobecněném" výběrovém rozptylu $|\mathbf{S}|$. Tato statistika, která je determinantem výběrové kovarianční matice, je mírou vícerozměrné variability. Montgomery a Wadsworth použili asymptoticky normální aproximace ke stanovení regulačního diagramu pro $|\mathbf{S}|$. Metoda využívá průměr a rozptyl $|\mathbf{S}|$ - tj. $E(|\mathbf{S}|)$ a $V(|\mathbf{S}|)$ - a vlastnosti že převážná část rozdělení pravděpodobnosti $|\mathbf{S}|$ je obsažena v intervalu $E(|\mathbf{S}|) \pm 3\sqrt{V(|\mathbf{S}|)}$.

Monitoring variability

Potom lze ukázat, že

$$E(|\mathbf{S}|) = b_1 |\Sigma|$$

a

$$V(|\mathbf{S}|) = b_2 |\Sigma|^2$$

kde

$$b_1 = \frac{1}{(n-1)^p} \prod_{i=1}^p (n-i)$$

$$b_2 = \frac{1}{(n-1)^{2p}} \prod_{i=1}^p (n-i) \left[\prod_{j=1}^p (n-j+2) - \prod_{j=1}^p (n-j) \right]$$

Monitoring variability

Meze regulačního diagramu pro $|\mathbf{S}|$ budou

$$\text{UCL} = |\Sigma| (b_1 + 3b_2^{1/2})$$

$$\text{CL} = b_1 |\Sigma|$$

$$\text{LCL} = |\Sigma| (b_1 - 3b_2^{1/2})$$

V praxi jsou parametry nahrazeny odhady

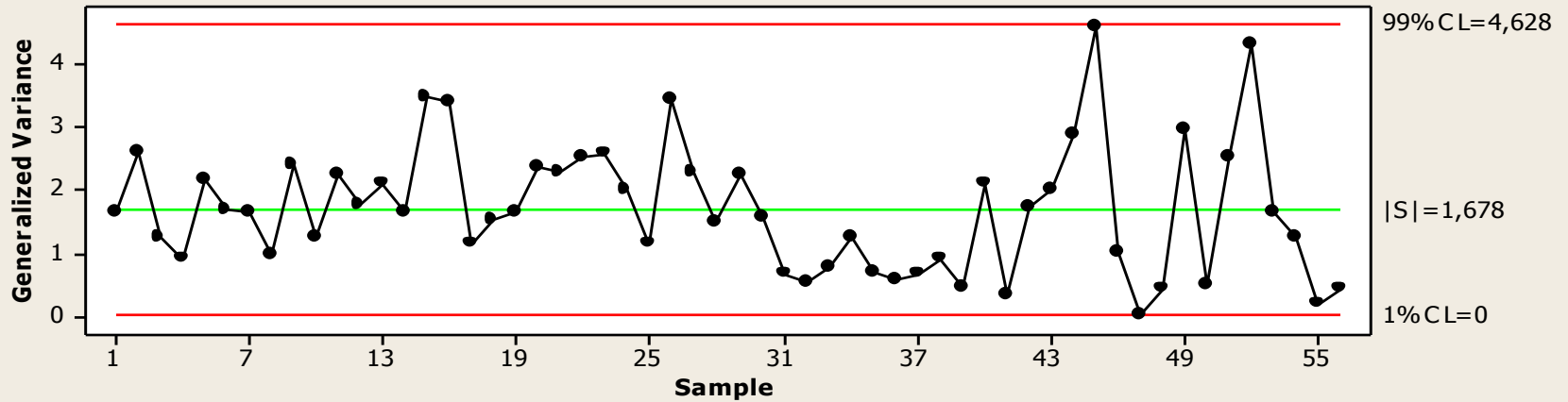
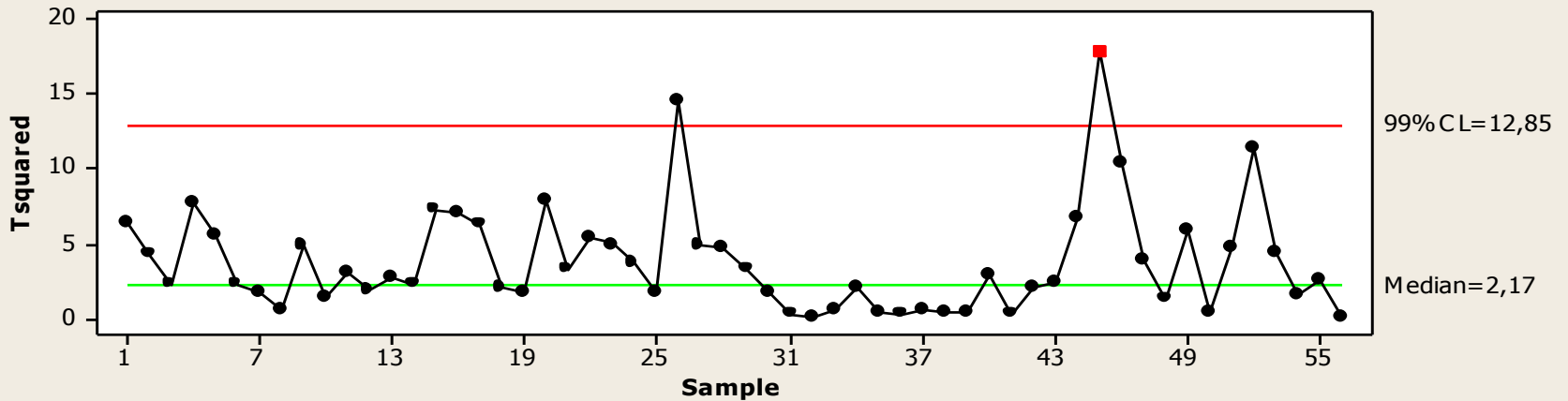
Příklad

Příklad byl řešen pomocí Minitab 16
2 znaky jakosti normálně rozděleny,
jsou závislé, výběr. koeficient korelace
je -0,769, UCL a LCL jsou nastaveny na
99% a 1%

Proces se zdá být zvládnutý v úrovni
variability, ale v parametrech polohy dělá
problémy

Příklad

Tsquared-Generalized Variance Chart of X1; X2



- **Literatura:**

- (1) Montgomery D. C.: "*Introduction to Statistical Quality Control*" ; 4th Edition; John Wiley & Sons, 2001
- (2) Hotelling, H.: "*Multivariate Quality Control*" edited by Eisenhart, Hastay and Wallis; McGraw-Hill, New York, 1947.
- (3) Lowry, C. A., and Montgomery D. C. (1995). "*A Review of Multivariate Control Charts*", IIE Transactions Vol. 26.
- (4) Tracy, N. D., Young, J. C., Mason, R. L. (1992). "*Multivariate Control Charts for Individual Observations*", Journal of Quality Technology, Vol. 24.
- (5) Sullivan, J. H., Woodall, W. H. (1995). "*A Comparison of Multivariate Quality Control Charts for Individual Observations*". Journal of Quality Technology, Vol. 27.
- (6) Holmes, D. S., Mergen, A. E. (1993). "*Improving the Performance of the T2 Control Chart*". Quality Engineering, Vol. 5.
- (7) Montgomery, D. C., Wadsworth, H. M. Jr. (1972). "*Some Techniques for Multivariate Quality Control Applications*", ASQC Technical Conference Transactions, Washington, DC.

A to je konec.....

Díky za pozornost