Uporabna statistika

Gregor Dolinar

Fakulteta za elektrotehniko
Univerza v Ljubljani

15. januar 2013

Definirajmo še en pojem in sicer funkcijo ogroženosti ali funkcijo tveganja h.
Verjetnost, da bo proizvod nehal delovati v časovnem intervalu $[t, t + \delta t]$, je enaka

$$R(t) - R(t + \delta t) = F(t + \delta t) - F(t) \cong f(t)\delta t.$$

Verjetnost, da bo proizvod nehal delovati v intervalu $[t, t + \delta t]$, pri pogoju, da je deloval do časa t, je približno enaka

$$\frac{f(t)\delta t}{R(t)} = h(t)\delta t.$$

Funkcija h nam pove, kolikšna je "nagnjenost" proizvoda, ki se do nekega trenutka še ni pokvaril, da se bo pokvaril.
Torej je

$$h(t) = \frac{f(t)}{R(t)} = \frac{-R'(t)}{R(t)}, \quad \text{oziroma} \quad R(t) = e^{-\int_0^t h(u)du}.$$

Funkcija ogroženosti je pomembna, ker nam pomaga pri interpretaciji in določanju ustreznega modela za življenjsko dobo.

- $h(t) = \lambda$, $R(t) = e^{-\lambda t}$, proizvod se ne obrablja
- h naraščajoča, proizvod se obrablja
- h padajoča, proizvod se "vpelje"
- h ima obliko U, najbolj običajno (proizvod se vpelje, uporabno obdobje, proizvod je izrabljen)

Nekaj najpogostejših modelov za določanje življenjske dobe:

- Poissonova porazdelitev
- Weibullova porazdelitev
- Gumbelova porazdelitev ($\log T$, kjer je T Weibullova)
- normalna logaritemska porazdelitev
- Gama porazdelitev

Poissonova porazdelitev

Proizvod je brez spomina, se ne obrablja.
Okvara je v vsakem trenutku enako verjetna, okvare so neodvisne, funkcija tveganja h je konstantna, torej je $h(t) = \lambda$.
Velja še, da sta gostota verjetnosti in funkcija zanesljivosti

$$f(t) = \lambda e^{-\lambda t}, \quad R(t) = e^{-\lambda t}.$$

Diskretna slučajna spremenljivka X, število okvar v času t, če je $\mu = \lambda t$ pričakovano število okvar v času t,

$$P[X = n] = \frac{e^{-\mu}\mu^n}{n!}$$

Velja

$$R(t) = P[\text{ni okvar do trenutka } t]$$
$$= P[X = 0] = \frac{e^{-\mu}\mu^0}{0!} = e^{-\lambda t}.$$

Weibullova porazdelitev

Najpogosteje uporabljana porazdelitev pri določanju življenjske dobe proizvoda, ima dva parametra:

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} e^{-(\frac{t}{\alpha})^{\beta}}, \quad t \geq 0,$$

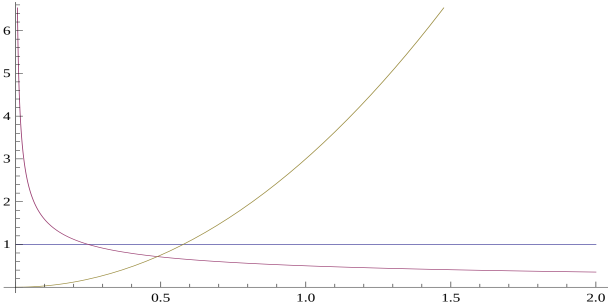
Pogledamo na drugačen način, funkcija tveganja h ima lepo obliko:

$$h(t) = ct^{\beta-1}, \quad c = \frac{\beta}{\alpha^{\beta}}.$$

$$R(t) = e^{-(\frac{t}{\alpha})^{\beta}}, \quad t \geq 0,$$

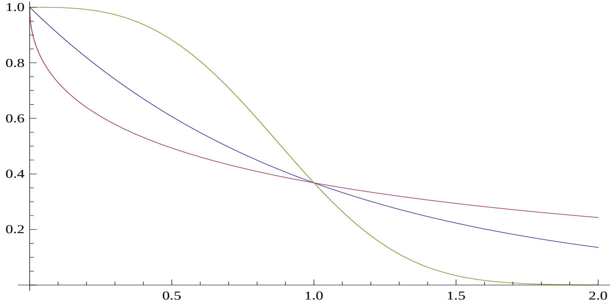
$$f(t) = -R'(t) = h(t)R(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} e^{-(\frac{t}{\alpha})^{\beta}}, \quad t \geq 0.$$

$h(t) = ct^{\beta-1}, \quad c = \beta/\alpha^{\beta}$



$\beta > 1$, se obrablja; $\beta = 1$, se ne obrablja; $\beta < 1$, vedno bolje.
$(\alpha = 1,\ \beta_1 = 1, \beta_2 = 0.5, \beta_3 = 3)$

$$R(t) = e^{-(\frac{t}{\alpha})^{\beta}}, \quad t \geq 0$$



Izbira modela

Ustrezno porazdelitev izberemo na podlagi:

- Poznavanja lastnosti procesa
- Pretekle izkušnje
- Podatkov iz preteklosti

Natančneje

- parametrični način
  Izberemo družino porazdelitev, nato moramo določiti še parametre.
- neparametrični način
  Empirično določimo funkcijo zanesljivosti in funkcijo tveganja.

Pri preverjanju ustreznosti modela moramo med drugim upoštevati:

- omejeno dostopnost do podatkov - krnjenje:
  - prekinemo testiranje n proizvodov po določenem času
  - prekinemo testiranje n proizvodov, ko pride do napake pri k proizvodih - Kaplan-Meier
- kakšne vrste model imamo
- način preverjanja (cena, izvedljivost)

Pri proizvodih - sistemih je potrebno upoštevati, da so sistemi večinoma popravljivi. Če pride do okvare, lahko zamenjamo ustrezni sestavni del sistema.
Zanima nas:

- kako zamenjava vpliva na življenjsko dobo sistema
- medsebojna odvisnost posameznih sestavnih delov sistema
- čas zastoja (razpoložljivost elementov sistema in čas aktivnega popravila)
  in stroški zamenjave (nedelovanje sistema in nov sestavni del)

Pri analizi zanesljivosti si pomagamo z drevesno analizo odpovedi:

- vzporedni (sistem deluje, če deluje vsaj en sestavni del), zaporedni sistemi (sistem deluje, če delujejo vsi sestavni deli)
- redundantnost sistema: pasivni (dodaten sestavni del v rezervi, se aktivira, če potrebno), aktivni (več aktivnih sestavnih delov, kot je potrebno - letalski motor)
- odkrivanje najšibkejšega člena

Statistično obvladovanje procesa (SPC)

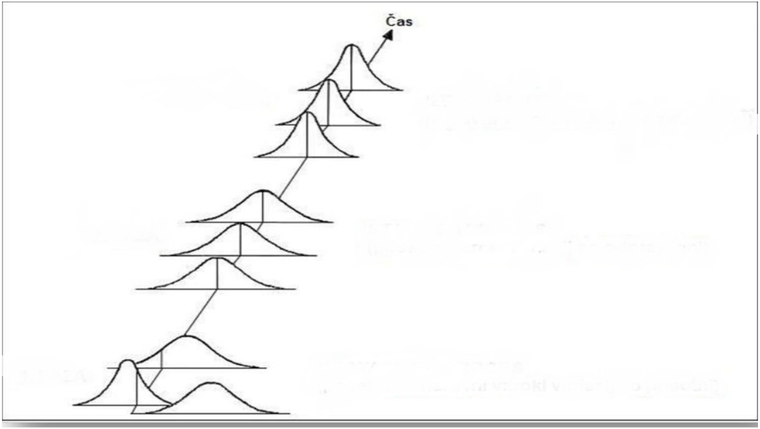
V 20. letih prejšnjega stoletja so se inženirji začeli zavedati:

- pomembnosti zmanjšanja variabilnosti,
- neprestano prilagajanje procesa, kot reakcija na posamezna odstopanja, poveča variabilnost.

Walter A. Shewhart (1891-1967), družba Bell Telephone Laboratories, kontrolne karte.
W. Edwards Deming (1900-1993), poskus s frnikolami. "Če ni pokvarjeno, ne popravljaj."

V vsakem procesu prihaja do odstopanj:

- običajni vzroki (chance cause - naključni vzroki): vsota veliko manjših neizogibnih vplivov
- posebni vzroki (assignable cause - določljivi vzroki): neprimerne nastavitve proizvodnih sredstev, napake upravljalcev, napake pri surovinah, ...

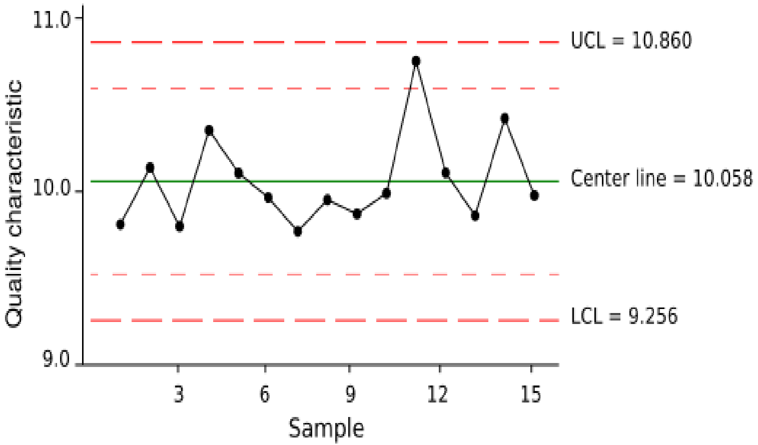


Vir: Kokol 2004, 7.

- preverjanje kakovosti končnih izdelkov, izločanje neustreznih (plačujemo za proizvodnjo določenega deleža proizvodov, ki jih mečemo stran).
- zagotavljanje kakovosti, proces mora biti v vseh fazah stabilen
- izboljševanje kakovosti, proces stabilen in mora delovati s čim manj variabilnosti okrog zahtevane vrednosti

Če prihaja do odstopanj samo zaradi običajnih razlogov, je proces pod kontrolo (stanje procesa je sprejemljivo).
Če prihaja do odstopanj tudi zaradi posebnih razlogov, potem proces ni pod kontrolo (stanje procesa ni sprejemljivo), delež neustreznih proizvodov je večji.
Glavna naloga statistične kontrole kakovosti je hitro odkrivanje posebnih razlogov variabilnosti.
Eno glavnih orodij pri tem so kontrolne karte.

Kontrolne karte



Ločimo dve veliki skupini kontrolnih kart:

- Kontrolne karte vrednosti spremenljivk
- Kontrolne karte lastnosti spremenljivk

Bistvene lastnosti kontrolne karte:

- srednja kontrolna črta,
  CL - center line
- zgornja in spodnja kontrolna meja,
  UCL - upper control limit, LCL - lower control limit

Oglejmo si

- Kontrolna karta za povprečje
- Kontrolna karta za variacijski razmik

Kontrolna karta za povprečje

Naj bo X spremenljivka, za katero velja $E(X) = \mu$, $\sigma(X) = \sigma$. Naj bo velikost vzorca n. Potem je

$$E(\overline{X}) = \mu, \;\; \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}.$$

Ločimo dve možnosti:

- Poznamo μ in σ
- Ne poznamo μ in σ

Če poznamo μ in σ, potem:

- $CL = \mu$
- $UCL = CL + 3\sigma$, $LCL = CL - 3\sigma$

Da je pri normalni porazdelitvi vrednost izven območja 6σ, je verjetnost le 0.27 %.
To pri milijonu proizvodov pomeni 2700 neustreznih proizvodov.

Če ne poznamo μ in σ, potem:

- $CL = \frac{1}{k}\sum_{i=1}^{k}\overline{x}_i$,
  $\overline{x}_i$ aritmetična sredina i-tega vzorca
- $UCL = CL + A_2\overline{VR}$, $LCL = CL - A_2\overline{VR}$,
  $\overline{VR} = \frac{1}{k}\sum_{i=1}^{k} VR_i$,
  VR_i maksimalni variacijski razmik i-tega vzorca,
  A_2 konstanta, odvisna od velikosti vzorca, npr. za $n = 2$ je $A_2 = 1.88$, za $n = 10$ je $A_2 = 0.308$.

Kontrolna karta za variacijski razmik

Če ne poznamo μ in σ, potem so kontrolne črte kontrolne karte za variacijski razmik:

- $CL = \overline{VR}$
- $UCL = D_4\overline{VR}$, $LCL = D_3\overline{VR}$,
  D_3, D_4 konstanti, odvisni od velikosti vzorca, npr. za $n = 2$ je $D_3 = 0$, $D_4 = 3.267$, za $n = 10$ je $D_3 = 0.223$, $D_4 = 1.777$.

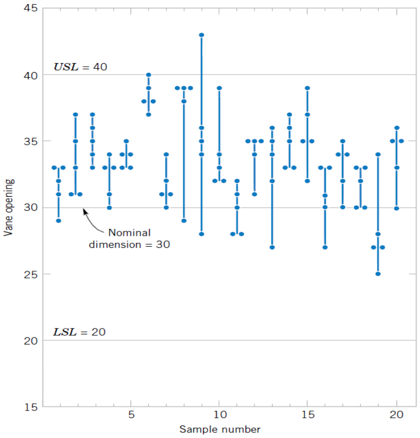
Kdaj proces ni pod kontrolo?

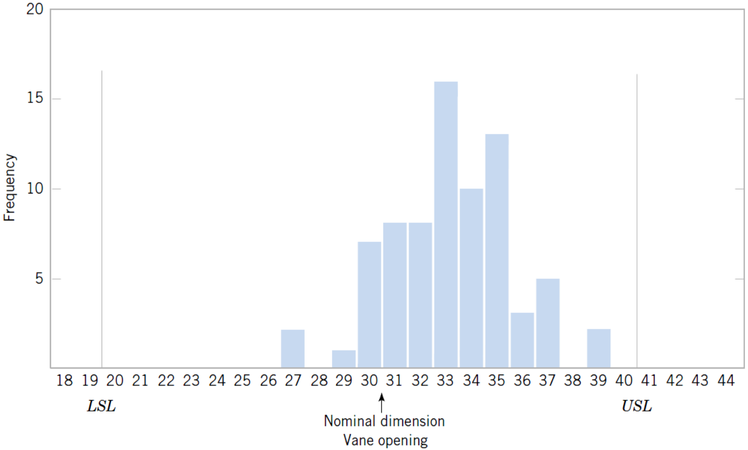
- 1 vrednost izven kontrolnih mej
- 7 ali več zaporednih vrednosti nad (ali pod) srednjo kontrolno črto
- 6 zaporednih vrednosti naraščajočih - naraščajoči trend, ali 6 zaporednih vrednosti padajočih (padajoči trend)
- 2 od 3 točk v območju med $\mu + 2\sigma$ in $\mu + 3\sigma$ ali 2 od 3 točk v območju med $\mu - 2\sigma$ in $\mu - 3\sigma$

Spremljanje delovanja tehnološkega procesa

Ločiti moramo med predpisanimi mejami (tolerančne meje, specificirane) in kontrolnimi mejami (proces pod kontrolo).
Ločimo:
USL (upper specification limit) in UCL (upper control limit)
LSL (lower specification limit) in LCL (lower control limit).
Ni primerno risati USL in UCL na kontrolno karto.

Grafični predstavitvi delovanja tehnološkega procesa




20
15
10
5
Frequency
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44
LSL
USL
Nominal dimension
Vane opening

Opis delovanja tehnološkega procesa s pomočjo kazalnikov (process capability indices)

Definicija

Kazalnik sposobnosti procesa

$$C_p = \frac{USL - LSL}{6\sigma}$$

$$\hat{C}_p = \frac{USL - LSL}{6\hat{\sigma}},$$

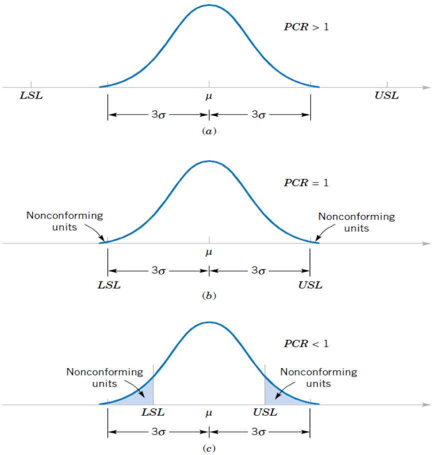
kjer je

$$\hat{\sigma} = \overline{VR} d_2.$$

Kazalnik C_p ima naravno interpretacijo, izraz

$$\frac{1}{C_p}$$

pove, kolikšen del tolerančnega intervala uporablja proces, ki je pod kontrolo.


PCR > 1
LSL
μ
USL
3σ
3σ
(a)
PCR = 1
Nonconforming units
Nonconforming units
μ
3σ
3σ
LSL
USL
(b)
PCR < 1
Nonconforming units
Nonconforming units
LSL
μ
USL
3σ
3σ
(c)

Vrednosti C_p:

- $C_p = 0.5$, neustreznih proizvodov 13%, (133614 na milijon)
- $C_p = 1$, neustreznih proizvodov 0.27%, (2700 na milijon)
- Priporočeni minimum za $C_p = 1.5$, (od 1.33 do 1.67), neustreznih proizvodov 0.0007%, (7 na milijon)
- $C_p = 2$, neustreznih proizvodov skorajda ni (0.0018 na milijon) Tako imenovani 6σ proces (Motorola).

Pri kazalniku C_p je bila bistvena predpostavka, da je proces centriran.
Če proces ni centriran, so ocene o sposobnosti tehnološkega procesa, da ostaja znotraj tolerančnih mej, preveč optimistične.

Definicija

Kazalnik centriranosti procesa

$$C_{pk} = \min\{C_{pu}, C_{pl}\},$$

kjer je

$$C_{pu} = \frac{USL - \mu}{3\sigma}, \quad C_{pl} = \frac{\mu - LSL}{3\sigma}.$$

Definicija

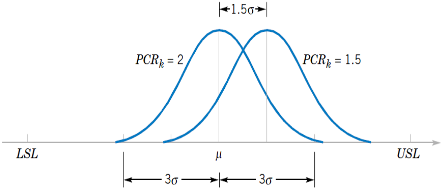
$$\hat{C_{pk}} = \min\{\hat{C_{pu}}, \hat{C_{pl}}\},$$

kjer je

$$\hat{C_{pu}} = \frac{USL - \overline{x}}{3\hat{\sigma}}, \quad \hat{C_{pl}} = \frac{\overline{x} - LSL}{3\hat{\sigma}}.$$

Če je proces centriran, je $C_p = C_{pk}$.

Kazalnik C_p meri možno, kazalnik C_{pk} pa dejansko sposobnost procesa. Proces je težko ohranjati centriran na daljše obdobje.
Če se sredina premakne za 1.5σ, se pri 6σ procesu C_{pk} zmanjša z 2 na 1.5, kar je še vedno sprejemljivo (neustreznih 3.4 na milijon)



Definicija

Taguchijeva kazalnika sposobnosti procesa

$$C_{pm} = \frac{C_p}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}$$

in

$$C_{pmk} = \frac{C_{pk}}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}},$$

kjer je T ciljna vrednost, ki je lahko različna od μ.

Osnovna literatura

- Douglas C. Montgomery, *Applied Statistics and Probability for Engineers*
- Douglas C. Montgomery, *Introduction to Statistical Quality Control*
- Linda. C. Wolstenholme, *Reliability Modeling, A Statistical Approach*