

Contrôle statistique de qualité

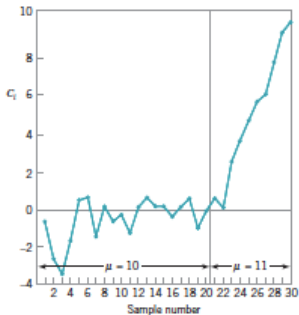
Mitra Fouladirad

Carte de contrôle 2

Exemple

Détection des très petits écarts de la moyenne
CUSUM (CUmulative SUM) somme cumulée utilise la statistique suivante:

$$C_l = \sum_{j=1}^l (\bar{x}_j - \mu_0)$$



CUSUM récursive

Nous utilisons les statistiques

$$C_l^+ = \max(0, x_l - (\mu_0 + K) + C_{l-1}^+)$$

$$C_l^- = \max(0, (\mu_0 + K) - x_l + C_{l-1}^-)$$

où $C_0^+ = C_0^- = 0$ et $K = \frac{\delta}{2}\sigma = \frac{|\mu_1 - \mu_0|}{2}$ est appelée la valeur de référence.

$$\hat{\mu} = \begin{cases} \mu_0 + K + \frac{C_l^+}{N^+} & \text{si } C_l^+ > H \\ \mu_0 - K - \frac{C_l^-}{N^-} & \text{si } C_l^- > H \end{cases}$$

où N^+ et N^- sont des compteurs de nombre de fois C_l^+ et C_l^- dépassent le seuil.

Souvent $H = \sigma h$ et $K = \sigma k$, les seuils sont des multiples de l'écart type σ .

CUSUM récursive

- Comparaison avec le seuil de décision H : $C_l^+ < H$
- ARL de CUSUM unilatéral

$$ARL = \frac{\exp(-2\Delta b) + 2\Delta b - 1}{2\Delta^2}$$

où si $\Delta \neq 0$, $\Delta = \delta^* - k$ pour la borne sup de C_l^+ et $\Delta = -\delta^* - k$ pour la borne inf de C_l^- , $b = h + 1.166$ et $\delta^* = \mu_1 - \mu_0\sigma$. Si $\Delta = 0$ alors on utilise $ARL = b$,

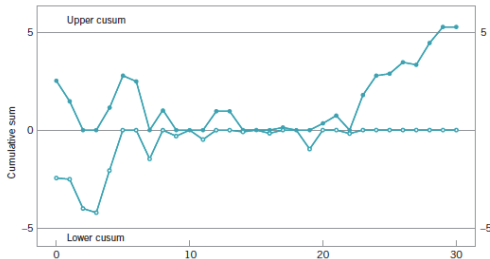
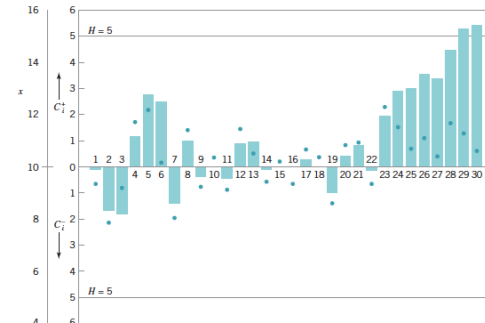
- ARL de CUSUM bilatéral

$$\frac{1}{ARL} = \frac{1}{ARL^+} + \frac{1}{ARL^-}$$

CUSUM: exemple

Period i	x_i	(a)			(b)		
		$x_i - 10.5$	C_i^+	N^+	$9.5 - x_i$	C_i^-	N^-
1	9.45	-1.05	0	0	0.05	0.05	1
2	7.99	-2.51	0	0	1.51	1.56	2
3	9.29	-1.21	0	0	0.21	1.77	3
4	11.66	1.16	1.16	1	-2.16	0	0
5	12.16	1.66	2.82	2	-2.66	0	0
6	10.18	-0.32	2.50	3	-0.68	0	0
7	8.04	-2.46	0.04	4	1.46	1.46	1
8	11.46	0.96	1.00	5	-1.96	0	0
9	9.20	-1.3	0	0	0.30	0.30	1
10	10.34	-0.16	0	0	-0.84	0	0
11	9.03	-1.47	0	0	0.47	0.47	1
12	11.47	0.97	0.97	1	-1.97	0	0
13	10.51	0.01	0.98	2	-1.01	0	0
14	9.40	-1.10	0	0	0.10	0.10	1
15	10.08	-0.42	0	0	-0.58	0	0
16	9.37	-1.13	0	0	0.13	0.13	1
17	10.62	0.12	0.12	1	-1.12	0	0
18	10.31	-0.19	0	0	-0.81	0	0
19	8.52	-1.98	0	0	0.98	0.98	1
20	10.84	0.34	0.34	1	-1.34	0	0
21	10.90	0.40	0.74	2	-1.40	0	0
22	9.33	-1.17	0	0	0.17	0.17	1
23	12.29	1.79	1.79	1	-2.79	0	0
24	11.50	1.00	2.79	2	-2.00	0	0
25	10.60	0.10	2.89	3	-1.10	0	0
26	11.08	0.58	3.47	4	-1.58	0	0
27	10.38	-0.12	3.35	5	-0.88	0	0
28	11.62	1.12	4.47	6	-2.12	0	0
29	11.31	0.81	5.28	7	-1.81	0	0
30	10.52	0.02	5.30	8	-1.02	0	0

CUSUM: exemple



CUSUM: exemple ARL

ARL pour $k = 1/2$ ($K = k\sigma$)

Shift in Mean (multiple of σ)	$h = 4$	$h = 5$
0	168	465
0.25	74.2	139
0.50	26.6	38.0
0.75	13.3	17.0
1.00	8.38	10.4
1.50	4.75	5.75
2.00	3.34	4.01
2.50	2.62	3.11
3.00	2.19	2.57
4.00	1.71	2.01

k et h pour $ARL = 370$ ($K = k\sigma$, $H = h\sigma$)

k	0.25	0.5	0.75	1.0	1.25	1.5
h	8.01	4.77	3.34	2.52	1.99	1.61

CUSUM standardisé

Nous utilisons $y_l = \frac{x_l - \mu_0}{\sigma}$, Dans ce cas nous avons

$$C_l^+ = \max(0, y_l - k + C_{l-1}^+)$$

$$C_l^- = \max(0, k - y_l + C_{l-1}^-)$$

où $C_0^+ = C_0^- = 0$ et k est la valeur de référence et h le seuil.
Il est possible d'avoir des cartes avec un seul indicateur

CUSUM pour détection de la variabilité

Nous utilisons $v_l = \frac{\sqrt{|y_l|} - 0.822}{0.347}$, Hawkins (1981, 1993a) a proposé

$$S_l^+ = \max(0, v_l - k + S_{l-1}^+)$$

$$S_l^- = \max(0, k - v_l + S_{l-1}^-)$$

où $S_0^+ = S_0^- = 0$ et k est la valeur de référence et h le seuil.
Pour la variance, sous l'hypothèse de la normalité avec σ_0 l'écart type sous contrôle et σ_1 l'écart type hors contrôle, Hawkins and Olwell (1998) ont proposé

$$C_l^+ = \max(0, S_i^2 + k + C_{l-1}^+), \quad C_l^- = \max(0, -k + S_i^2 + C_{l-1}^-)$$

où $C_0^+ = C_0^- = 0$, S_i^2 est l'estimateur de la variance et $k = \frac{2 \ln(\sigma_0/\sigma_1) \sigma_0^2 \sigma_1^2}{(\sigma_0^2 - \sigma_1^2)}$ est la valeur de référence et h le seuil.

CUSUM avec d'autres statistiques

Lorsque nous disposons des données de comptage avec un taux faible, nous pouvons construire un CUSUM pour l'intervalle entre l'arrivée des événements.

Par exemple, pour un processus de Poisson (l'intervalle inter-arrivée des événements suit une loi exponentielle) nous pouvons considérer

$$C_i^- = \max(0, K - T_i + C_{i-1}^-)$$

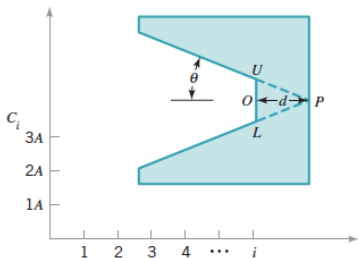
où K est la valeur de référence et T_i le temps écoulé depuis la dernière observation .

Masque V

Cette méthode est une alternative au CUSUM récursif où on utilise la statistique

$$C_i = \sum_{j=1}^i y_j = y_i + C_{i-1}$$

où $y_i = \frac{x_i - \mu_0}{\sigma}$ Le masque est défini dans la Figure



Masque V

Pour $k = A \tan \theta$ et $h = d A \tan \theta = dk$ où A est la distance horizontal entre deux points successifs.

Johnson (1961) a proposé, $\theta = \tan^{-1}(\frac{\delta}{2A})$ et $d = (\frac{2}{\delta^2} \ln(\frac{1-\beta}{\alpha}))$ où 2α la valeur maximale de fausse alarme acceptable et β la probabilité de ne pas détecter un écart de taille δ .

Les inconvénients:

- Une statistique qui ne démarre pas à zéro n'est pas applicable
- Parfois difficile à déterminer les branches de la masque
- Ambiguïté sur α et β .

Actual Values of ARL_0 for a V-Mask Scheme Designed Using Johnson's Method [Adapted from Table 2 in Woodall and Adams (1993)]

Shift to Be Detected, δ	Values of α [Desired Value of $ARL_0 = 1/(2\alpha)$]	
	0.00135 (370)	0.001 (500)
1.0	2350.6	3184.5
2.0	1804.5	2435.8
3.0	2194.8	2975.4

CUSUM "autonome" (self-starting)

CUSUM autonome proposé par Hawkins (1987) lorsque la moyenne et la variance sont inconnues.

Facile à implanter pour la moyenne d'un échantillon gaussien.

Opérationnelle immédiatement après la mise en place pas besoin de période initiale d'estimation

Soit $w_n = \sum_{i=1}^n (x_i - \bar{x}_n)^2$, alors $s_n^2 = w_n / (n - 1)$ nous considérons

$$\bar{x}_n = \bar{x}_{n-1} + \frac{x_n - \bar{x}_{n-1}}{n}, \quad w_n = w_{n-1} + \frac{(n-1)(x_n - \bar{x}_{n-1})^2}{n}$$

Nous utilisons

$$\sqrt{\frac{n-1}{n}} T_n, \quad T_n = \frac{x_n - \bar{x}_{n-1}}{s_{n-1}}$$

qui suit une loi de Student avec $n - 1$ degré de liberté.

CUSUM "autonome" (self-starting)

Pour utiliser les tables de loi normale nous allons considérer la statistique suivante:

$$U_n = \Phi^{-1}(\mathbb{P}(T_n < t)) = \Phi^{-1}(F_{n-2}(\sqrt{\frac{n-1}{n}}T_n))$$

où Φ^{-1} est l'inverse de la fonction de répartition d'une variable normale centrée réduite et F_n est le CFD d'une loi de Student de $n - 1$ de degré de liberté.

EWMA: Moyenne mobile pondérée exponentiellement

EWMA (exponentially weighted moving average) est introduit par Roberts (1959), étudié notamment par (1987a, 1989), Lucas and Saccucci (1990), pour détecter des très petits changements et ses propriétés sont proches de celles de CUSUM.

La statistique utilisée:

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1}$$

où $0 \leq \lambda \leq 1$, et le point de départ $z_0 = \mu_0$ la moyenne sous contrôle.

En développant,

$$z_i = \lambda \sum_{j=0}^{i-1} (1-\lambda)^j x_{j+1} + (1-\lambda)^i z_0 = (1 - (1-\lambda)^i) \bar{x} + (1-\lambda)^i z_0$$

EWMA: Moyenne mobile pondérée exponentiellement

Parfois appelé moyenne mobile géométrique (geometric moving average GMA).

Moyenne mobile donc adaptée à l'hypothèse de normalité

Si les x_i sont indépendantes, alors les z_i aussi et nous avons la variance

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2i}) \right)$$

Les bornes de la carte

$$UCL = \mu_0 + L\sigma \sqrt{\left(\frac{\lambda}{2 - \lambda} \right) (1 - (1 - \lambda)^{2i})}$$

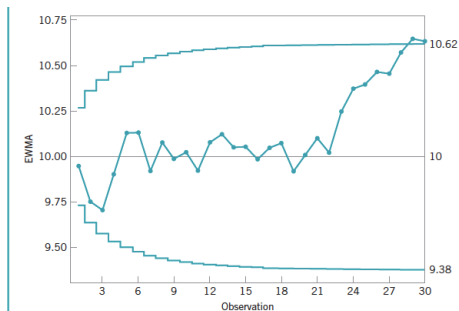
$$LCL = \mu_0 - L\sigma \sqrt{\left(\frac{\lambda}{2 - \lambda} \right) (1 - (1 - \lambda)^{2i})}$$

et μ_0 comme ligne centrale.

EWMA: Moyenne mobile pondérée

Si la carte est lancée depuis longtemps (i grand) nous avons les approximations suivantes:

$$UCL = \mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}}, \quad LCL = \mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}}$$



Carte EWMA sur les données de l'exemple

EWMA: Moyenne mobile pondérée exponentiellement

Average Run Lengths for Several EWMA Control Schemes
[Adapted from Lucas and Saccucci (1990)]

Shift in Mean (multiple of σ)	$L = 3.054$ $\lambda = 0.40$	2.998 0.25	2.962 0.20	2.814 0.10	2.615 0.05
0	500	500	500	500	500
0.25	224	170	150	106	84.1
0.50	71.2	48.2	41.8	31.3	28.8
0.75	28.4	20.1	18.2	15.9	16.4
1.00	14.3	11.1	10.5	10.3	11.4
1.50	5.9	5.5	5.5	6.1	7.1
2.00	3.5	3.6	3.7	4.4	5.2
2.50	2.5	2.7	2.9	3.4	4.2
3.00	2.0	2.3	2.4	2.9	3.5
4.00	1.4	1.7	1.9	2.2	2.7

ARL pour différentes L , λ

EWMA: Moyenne mobile pondérée

Woodall and Mahmoud (2005) ont défini la **résistance** qui est le plus grand écart normalisé de la moyenne à la valeur cible ne menant pas au signal d'alarme.

Pour la carte Shewhart chart, la résistance est égale à L le multiplicateur utilisé pour les bornes. Alors une résistance de signal constante

Pour la carte de contrôle EWMA, nous avons,

$$RE = \frac{L\sqrt{\frac{\lambda}{2-\lambda}} - (1-\lambda)w}{\lambda}$$

où w est la valeur de la statistique EWMA.

EWMA: Moyenne mobile pondérée

Robustness à la non normalité:

Tableaux des ARL sous différentes lois et paramètres

λ L	EWMA			Shewhart
	0.05 2.492	0.1 2.703	0.2 2.86	1 3.00
Normal	370.4	370.8	370.5	370.4
Gam(4, 1)	372	341	259	97
Gam(3, 1)	372	332	238	85
Gam(2, 1)	372	315	208	71
Gam(1, 1)	369	274	163	55
Gam(0.5, 1)	357	229	131	45

λ L	EWMA			Shewhart
	0.05 2.492	0.1 2.703	0.2 2.86	1 3.00
Normal	370.4	370.8	370.5	370.4
t_{50}	369	365	353	283
t_{40}	369	363	348	266
t_{30}	368	361	341	242
t_{20}	367	355	325	204
t_{15}	365	349	310	176
t_{10}	361	335	280	137
t_8	358	324	259	117
t_6	351	305	229	96
t_4	343	274	188	76

EWMA: Moyenne mobile pondérée

Les extensions de EWMA:

Si pas données individuelles mais échantillons de taille $n > 1$ on fait la carte pour \bar{x}_i

Détection rapide si un décalage dès le début:

- Deux cartes unilatérales avec des bornes dépendantes du temps Rhoads, Montgomery, and Mastrangelo (1996)
- Une carte avec borne Steiner (1999),

$$\pm L\sigma(1 - (1 - f)^{1+a(t-1)})\sqrt{\frac{\lambda}{2 - \lambda}(1 - (1 - \lambda)^{2t})}$$

les constantes f et a sont définies telles que le délai de détection moyen soit égale à une valeur donnée.

EWMA: Moyenne mobile pondérée, extensions

Carte du EWMA pour détecter un changement dans l'écart type: EWMS

$$S_i^2 = \lambda(x_i - \mu)^2 + (1 - \lambda)S_{i-1}^2$$

$E(S_i^2) \rightarrow \sigma^2$ lorsque $i \rightarrow \infty$ et si les observations sont i.i.d. et normalement distribuées, $\frac{S_i^2}{\sigma^2} \sim \chi^2\left(\frac{2-\lambda}{\lambda}\right)$.

Si σ_0 est l'écart type sous contrôle, nous pouvons alors considérer une carte de contrôle avec les seuils

$$UCL = \sigma \sqrt{\frac{\chi^2(\nu, \alpha/2)}{\nu}}, \quad UCL = \sigma \sqrt{\frac{\chi^2(\nu, 1 - \alpha/2)}{\nu}}$$

Afin de rendre la statistique insensible aux variations de μ , on peut utiliser un estimateur ponctuel par exemple:

$$S_i^2 = \lambda(x_i - z_i)^2 + (1 - \lambda)S_{i-1}^2$$

où $z_i = \lambda x_i + (1 - \lambda)z_{i-1}$, $z_0 = 0$

Carte de moyenne mobile (Moving Average Chart)

Moyenne mobile non pondérée sur une fenêtre w :

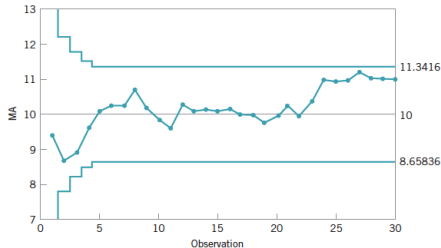
$$M_i = \frac{x_i + x_{i-1} + \cdots + x_{i-w+1}}{w}$$

avec les bornes

$$UCL = \mu_0 + \frac{3\sigma}{\sqrt{w}}, \quad LCL = \mu_0 - \frac{3\sigma}{\sqrt{w}}.$$

Carte de moyenne mobile (Moving Average Chart)

Observation, i	x_i	M_i	Observation, i	x_i	M_i
1	9.45	9.45	16	9.37	10.166
2	7.99	8.72	17	10.62	9.996
3	9.29	8.91	18	10.31	9.956
4	11.66	9.5975	19	8.52	9.78
5	12.16	10.11	20	10.84	9.932
6	10.18	10.256	21	10.9	10.238
7	8.04	10.266	22	9.33	9.98
8	11.46	10.7	23	12.29	10.376
9	9.2	10.208	24	11.5	10.972
10	10.34	9.844	25	10.6	10.924
11	9.03	9.614	26	11.08	10.96
12	11.47	10.3	27	10.38	11.17
13	10.51	10.11	28	11.62	11.036
14	9.4	10.15	29	11.31	10.998
15	10.08	10.098	30	10.52	10.982



Carte de contrôle pour les durées courtes

Une carte \bar{x} et R .

La statistique sera basée sur

$$x_i = M_i - T_A$$

où M_i est la mesure et T_A la valeur de référence

Nous pouvons aussi construire une carte \bar{x} et R standardisée avec les statistiques:

$$R_i^s = \frac{R_i}{\bar{R}_j}, \quad \bar{x}_i^s = \frac{\bar{M}_i - T_j}{\bar{R}_j}$$

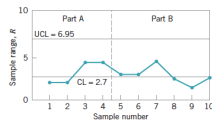
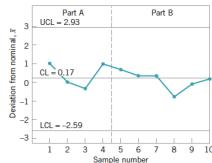
où \bar{R}_j et T_j sont respectivement l'étendue moyenne et la valeur nominal de x de la j ème calsse de données ,

$LCL = D_3$ et $UCL = D_4$ pour R_i^s et $LCL = -A_2$ et $UCL = A_2$ pour \bar{x}_i^s .

Pour les autres cartes p , np u et c les statistiques habituelles sont efficaces.

Carte de contrôle pour les durées courtes: exemple

Sample Number	Part Number	(a) Measurements			(b) Deviation from Nominal				\bar{x}	R
		M_1	M_2	M_3	x_1	x_2	x_3			
1	A	50	51	52	0	1	2	1.00	2	
2	A	49	50	51	-1	0	1	0.00	2	
3	A	48	49	52	-2	-1	2	-0.33	4	
4	A	49	53	51	-1	3	1	1.00	4	
5	B	24	27	26	-1	2	1	0.67	3	
6	B	25	27	24	0	2	-1	0.33	3	
7	B	27	26	23	2	1	-2	0.33	4	
8	B	25	24	23	0	-1	-2	-1.00	2	
9	B	24	25	25	-1	0	0	-0.33	1	
10	B	26	24	25	1	-1	0	0.00	2	
									$\bar{\bar{x}} = 0.17$	$\bar{R} = 2.7$



Carte de contrôle modifiée

Changer les bornes, pour être moins stricte

Si $PCR = \frac{USL - LSL}{6\sigma}$ atteint la borne souhaitée

$\mu_L = LSL + Z_\delta \sigma$ et $\mu_U = USL - Z_\delta \sigma$ avec Z_δ la quantile alors

$$UCL = \mu_U + \frac{Z_\delta}{\sqrt{n}}, \quad LCL = \mu_L - \frac{Z_\delta}{\sqrt{n}}$$

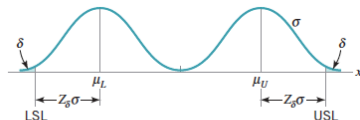
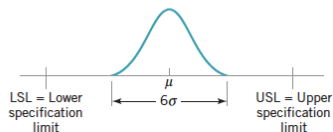
ou autrement

$$UCL = USL - (Z_\delta \frac{3}{\sqrt{n}})\sigma, \quad LCL = LSL + (Z_\delta \frac{3}{\sqrt{n}})\sigma$$

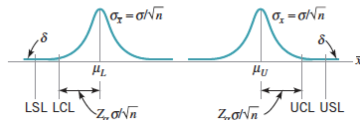
La carte modifiée est équivalente à un test sur la moyenne d'un échantillon gaussien $\mu_L \leq \mu \leq \mu_U$

Attention carte valable si nous disposons d'un estimateur robuste de σ

Carte de contrôle modifiée



(a)

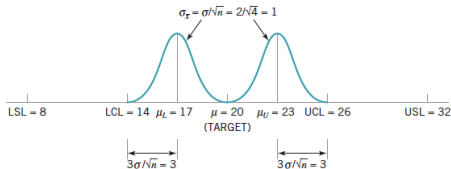
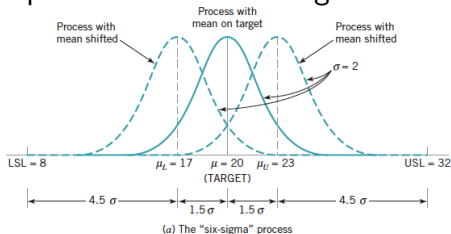


(b)

(gauche) Tolérance limite naturelle inférieure à la dispersion des limites de spécifications, $6\sigma < USL - LSL$ (droite) (a) Densité des sorties. (b) Densité des moyennes.

Carte de contrôle modifiée: exemple

Soit un process gaussien, avec la valeur moyenne cible $\mu = 20$ et $\sigma = 2$, $LSL = 8$, $USL = 32$. Alors si le processus est autour de 2, nous sommes dans le cadre de 6sigma. Adaptons la carte pour détecter un changement de moyenne à 1.5 ave



$$n = 4$$

Carte de contrôle pour acceptation

Changer les bornes, pour être moins stricte. Soient n la taille de chaque échantillon, la fraction de non conformité δ , et α l'erreur de première espèce. On peut interpréter δ comme la fraction de nonconformité d'un process qu'on accepte avec probabilité $1 - \alpha$. Freund (1957) a développé une carte pour prendre en compte les erreurs de première et seconde espèce α et β (on rejette avec probabilité $1 - \beta$).

Carte de contrôle pour acceptation

Freund a proposé d'utiliser la carte \bar{x} pour surveiller le pourcentage de non-conformité

$$UCL = \mu_U - \frac{Z_\beta \sigma}{\sqrt{n}} = USL - (Z_\gamma + \frac{Z_\beta}{\sqrt{n}})\sigma$$

$$LCL = \mu_L + \frac{Z_\beta \sigma}{\sqrt{n}} = LSL + (Z_\gamma + \frac{Z_\beta}{\sqrt{n}})\sigma$$

où Z_p est le seuil supérieur à $100(1 - p)$ pour cent d'une loi de normale centrée réduite.

Carte de contrôle pour acceptation

Lorsque n , γ et β sont spécifiés les limites de contrôle sont entre μ_L et μ_U et mènent à une fraction de non-conformité γ .
Au contraire lorsque n , δ et α sont spécifiés la limite de contrôle inférieure est entre μ_L et LSL et la limite de contrôle supérieure est entre μ_U et USL

En fixant γ , α et β on peut définir $n = \left(\frac{Z_\alpha + Z_\beta}{Z_\delta - Z_\gamma} \right)^2$

Carte de contrôle pour des flux multiples

Deux problématiques,

- un changement dans les mesures d'une voie
- un changement dans les mesures de toutes les voies

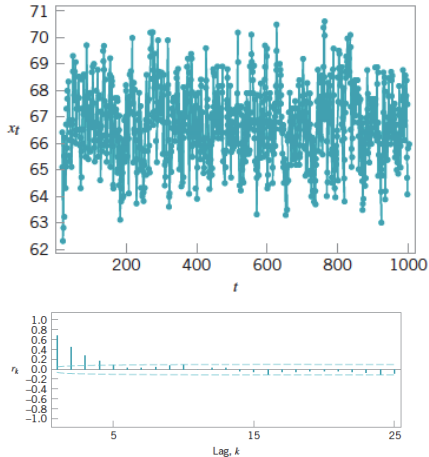
Carte de contrôle pour les données auto-corrélées: avec modèle

$$x_k = \mu + \epsilon_k, \quad k = 1, 2, \dots$$

Problème étudié par: Berthouex, Hunter, and Pallesen (1978), Alwan and Roberts (1988), Montgomery and Friedman (1989), Alwan (1992), Harris and Ross (1991), Montgomery and Mastrangelo (1991), and Maragah and Woodall (1992). On ne peut pas ignorer la dépendance. Nous pouvons surveiller le coefficient de corrélation empirique:

$$r_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i-k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad k = 0, 1, 2, \dots, K$$

Carte de contrôle pour les données auto-corrélées: avec modèle



Carte de contrôle pour les données auto-corrélées: avec modèle

Une autre méthode est d'utiliser les séries temporelles pour décrire la corrélation, AR, ARMA, MA. Par exemple, un $AR(1)$

$$x_t = \xi + \phi x_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon)$$

où ξ et $-1 < \phi < 1$ sont des constantes inconnues.

Utiliser EWMA pour détecter un changement dans les séries temporelles.

Soit $z_t = \hat{x}_{t+1}(t)$ la prédiction de x_{t+1} à l'instant t . Nous avons $z_t = \lambda x_t + (1 - \lambda)z_t$ qui est un EWMA.

Considérons l'erreur $e_t = x_t - \hat{x}_t(t - 1)$. Nous pouvons faire une carte pour surveiller e_t . Nous pouvons considérer les bornes constantes de Shewhart ou des bornes dynamiques.

Carte de contrôle pour les données auto-corrélées: avec modèle

Sachant que e_t est normalement distribuée

$$\mathbb{P}(-3\sigma \leq e_t \leq 3\sigma) = 0.9973$$

$$\Leftrightarrow \mathbb{P}(\hat{x}_t(t-1) - 3\sigma \leq x_t \leq \hat{x}_t(t-1) + 3\sigma) = 0.9973$$

On pourra donc utiliser les bornes $\pm 3\sigma$ pour e_t ou les bornes suivantes pour x_t :

$$\pm \hat{x}_t(t-1) + 3\sigma = \pm z_{t-1} + 3\sigma$$

Carte de contrôle pour les données auto-corrélées : avec modèle

On pourra aussi surveiller la variance, avec ses prédictions, de deux façon en utilisant soit

- la valeur absolue de l'erreur de prédiction

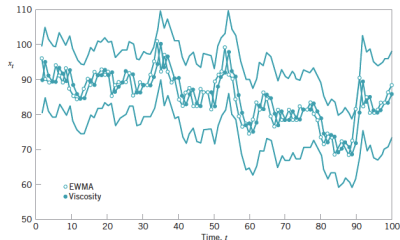
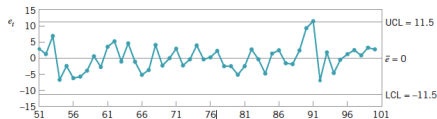
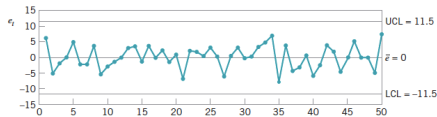
$$\Delta_t = \alpha |e_t| + (1 - \alpha) \Delta_{t-1}, \quad 0 < \alpha \leq 1$$

Pour le cas normal, un lien directe avec la variance,

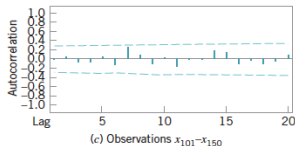
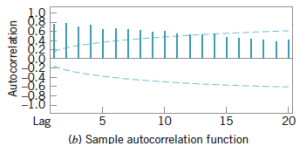
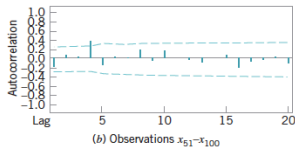
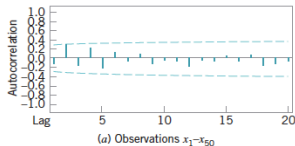
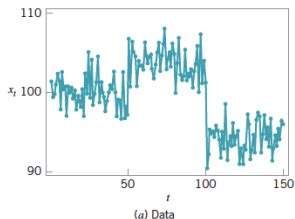
- la variance lissée

$$\hat{\sigma}_t^2 = \alpha e_t^2 + (1 - \alpha) \hat{\sigma}_{t-1}^2, \quad 0 < \alpha \leq 1$$

Carte de contrôle pour les données auto-corrélées : avec modèle



Carte de contrôle pour les données auto-corrélées : avec modèle



Absence d'autocorrelation lorsque les données hors limites sont éliminées

Carte de contrôle pour les données auto-corrélées: sans modèle

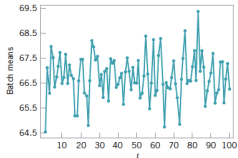
Nous pouvons aussi envisager la carte de contrôle sur la moyenne des sous groupes (voir Runger and Willemain (1996)).

Soit \bar{x}_j la moyenne du sous-groupe j de taille b définie par:

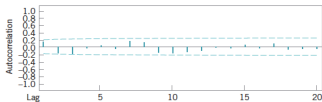
$$\bar{x}_j = \frac{1}{b} \sum_{i=1}^b x_{(j-1)b+1}$$

Le choix de b est très important

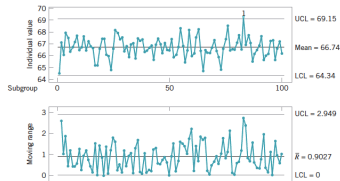
Carte de contrôle pour les données auto-corrélées : sans modèle



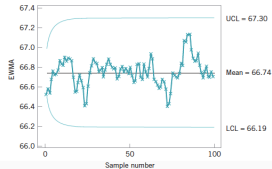
(a) Plot of batch means using batch size $b = 10$



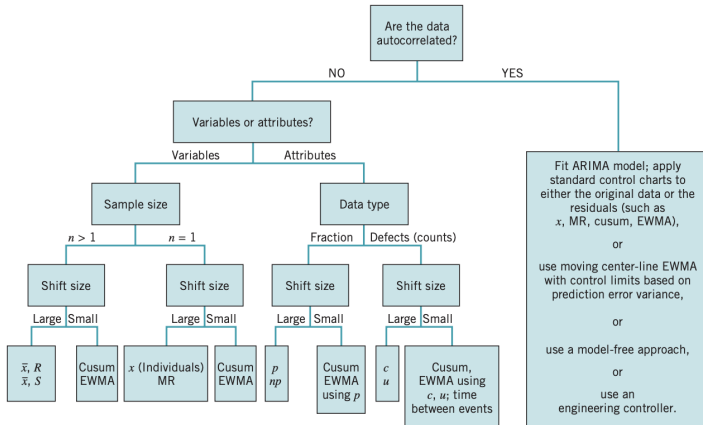
(b) Sample autocorrelation function



(a) Individuals and moving range control charts



Carte de contrôle pour les données auto-corrélées : sans modèle



Carte de contrôle Cuscore

Une carte proposée par Box et Ramirez (1992), Box et Luceo (1997), Ramirez (1998), Luceo (1999), Runger, Testik (2003).
A partir de score de Fisher,

$$Q = \sum e_{t0} r_t$$

où les e_{t0} sont les résidus d'un procédé sous contrôle et r_t est un détecteur qui mesure le taux de changement des résidus d'un procédé sous contrôle lorsque le procédé est hors contrôle.

Carte de contrôle Cuscore

Exemple dans le cas linéaire,

$$x_t = \mu + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma)$$

pour un changement de taille δ ,

$$e_{t0} = x_t - \mu, \quad e_t = x_t - \mu - \delta, \quad r_t = \frac{e_{t0} - e_t}{\delta}$$

et $Q = \sum (x_t - \mu)$

Application possible à d'autres types de données corrélées.

Carte de contrôle Cuscore

Pour un modèle linéaire suivant:

$$x_t = \mu + \beta t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma)$$

$$e_{t0} = x_t - \mu - \beta t, \quad e_t = x_t - \mu - \delta - \beta t, \quad r_t = \frac{e_{t0} - e_t}{\delta}$$

et

$$Q = \sum (x_t - \mu) = \sum (x_t - \mu - \beta t)t.$$

Les modèles de détection de changement

Supposons avoir des données issues d'une des deux lois telles que

$$x_t \sim \mathcal{N}(\mu_0, \sigma^2), t = 1, 2 \dots \tau$$

$$x_t \sim \mathcal{N}(\mu_1, \sigma^2), t = \tau + 1, \tau + 2 \dots n$$

Nous cherchons à détecter l'instant de changement τ . Nous pouvons par exemple utiliser la statistique:

$$t_{jn} = \sqrt{\frac{j(n-j)}{n}} \frac{\bar{x}_{jn} - \bar{x}_{jn}^*}{\hat{\sigma}_{jn}}$$

où \bar{x}_{jn} est la moyenne des j premières observations, \bar{x}_{jn}^* la moyenne des $n - j$ dernières observations, $\hat{\sigma}_{jn}$ l'estimation de la variance et $1 \leq j \leq n - 1$. Pour détecter τ , on cherche j ,

$$\max_{1 \leq j \leq n-1} |t_{jn}| > h_n$$

Surveillance de profile (covariables)

$$y_{i,j} = \beta_0 + \beta_1(x_1 - \bar{x}) + \epsilon_{i,j} = \beta_0 + \beta_1 x_i$$

où les $\epsilon_{i,j}$ sont indépendentes et $\epsilon_{i,j} \sim \mathcal{N}(\mu_0, \sigma^2)$

Les bornes pour les cartes de contrôle sur les estimations des paramètres de regression $\hat{\beta}_0$ et $\hat{\beta}_1$ sont comme suit

$$UCL = \beta_0 + Z_{\alpha/2} \sqrt{\sigma^2} n, \quad LCL = \beta_0 - Z_{\alpha/2} \sqrt{\sigma^2} n$$

$$UCL = \beta_1 + Z_{\alpha/2} \sqrt{\sigma^2} S_{xx}, \quad LCL = \beta_1 - Z_{\alpha/2} \sqrt{\sigma^2} S_{xx},$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2.$$

Les limites pour l'erreur sur la variance

$$UCL = \frac{\sigma^2}{n-2} \chi_{\alpha/2, n-2}^2, \quad LCL = \frac{\sigma^2}{n-2} \chi_{1-\alpha/2, n-2}^2$$

Carte de contrôle pour les covariables

Day	Position	x	y
1	L	0.76	1.12
1	M	3.29	3.49
1	U	8.89	9.11
2	L	0.76	0.99
2	M	3.29	3.53
2	U	8.89	8.89
3	L	0.76	1.05
3	M	3.29	3.46
3	U	8.89	9.02
4	L	0.76	0.76
4	M	3.29	3.75
4	U	8.89	9.3
5	L	0.76	0.96
5	M	3.29	3.53
5	U	8.89	9.05
6	L	0.76	1.03
6	M	3.29	3.52
6	U	8.89	9.02

Carte de contrôle pour les données pour les covariables

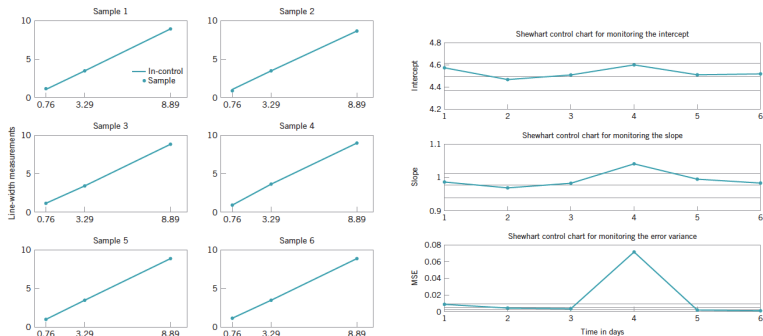


Figure: Dépendance (gauche), carte pour $\hat{\beta}_0$, $\hat{\beta}_1$ et S_{xx} (droite)

Données censurées

Lorsque le taux de censure est inférieure à 50% on la néglige. Si il est supérieur à 95% une carte de np est utilisée pour contrôler le nombre d'observation censuré.

Sinon comme proposé par Steiner et MacKay (2000) des observations gaussiennes de moyenne μ et de variance σ et le niveau de censure C , la probabilité de censure est

$$p_c = 1 - \Phi\left(\frac{C - \mu}{\sigma}\right) = Q(C)$$

où Φ est le CDF de la loi normale.

L'espérance conditionnelle d'une v.a censurée est:

$$w_c = \mathbb{E}(X|X > C) = \mu + \sigma \left(\frac{\phi(z_c)}{\Phi(z_c)} \right)$$

où ϕ est la densité de la loi normale centrée réduite et $z_c = (C - \mu)/\sigma$

Données censurées

Les données permettant de construire une carte de contrôle sont les suivantes

$$w = \begin{cases} x & \text{si } x \leq C \\ w_c & \text{si } x > C \end{cases}$$

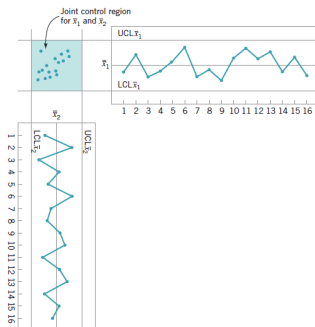
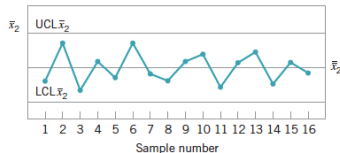
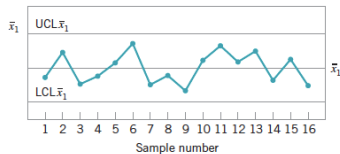
Les limites de contrôle ne sont pas facilement calculables, on a recours aux simulations.

Problème de contrôle qualité multivarié

Si nous avons p indicateurs de qualité indépendants à suivre et si pour une carte \bar{x} , nous avons une erreur de première espèce α alors

$$Pp(\text{toutes les moyennes soient sous contrôle}) = (1 - \alpha)p$$

Problème de contrôle qualité multivarié avec données gaussiennes

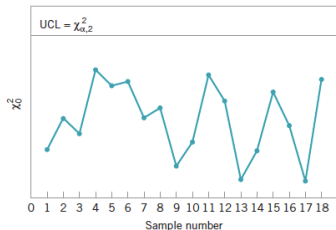


Carte de contrôle multivariée: 2 échantillons

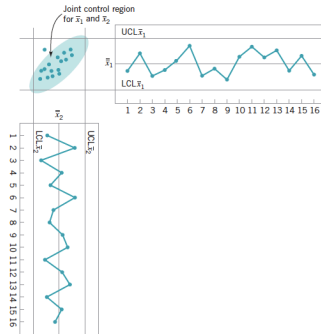
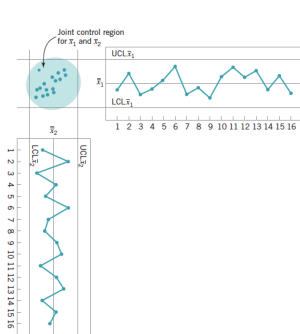
Soient $x_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $x_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ deux caractéristiques de qualité avec σ_{12} leur covariance. Soient \bar{x}_1 et \bar{x}_2 leur moyenne empirique. La statistique suivante

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

Si le procédé reste avec une moyenne entre μ_1 et μ_2 alors $\chi_0^2 < \chi_{\alpha,2}^2$.



Problème de contrôle qualité multivarié



Carte de contrôle avec matrice de variance covariance connue

Lorsque nous disposons de plus de 2 caractéristique à surveiller (x_1, \dots, x_p) , $x_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ et $\mu = (\mu_1, \dots, \mu_p)$ et Σ la matrice de variance covariance. Notons $\bar{x} = (\bar{x}_1, \dots, \bar{x}_p)$ alors le contrôle de la moyenne se fera sur la base de la statistique

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

avec le seuil supérieur

$$UCL = \chi_{\alpha,p}^2$$

Carte de contrôle Hotelling T^2 : matrice de variance covariance inconnue

Considérons m échantillons pour chaque caractéristique. Pour $j = 1, 2, \dots, p$ et $k = 1, 2, \dots, m$, notons

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

l'estimateur de la moyenne et de la variance de la j ème caractéristique dans le k ème échantillon. De plus

$$s_{jhk}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{x}_{jk})(x_{ihk} - \bar{x}_{hk}), \quad j \neq h.$$

$$\bar{\bar{x}}_j = \frac{1}{m} \sum_{k=1}^m \bar{x}_{jk}, \quad \bar{s}_j^2 = \frac{1}{m} \sum_{k=1}^m s_{jk}^2, \quad \bar{s}_{jh}^2 = \frac{1}{m} \sum_{k=1}^m s_{jhk}^2$$

Carte de contrôle Hotelling T^2

Lorsque les paramètres sont inconnus, le contrôle de la moyenne se fera sur la base de la statistique

$$T^2 = n(\bar{x} - \bar{\bar{x}})S^{-1}(\bar{x} - \bar{\bar{x}})^T$$

avec les seuils

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

Carte de contrôle Hotelling T^2

Lorsque $n = 1$, $T^2 = n(x - \bar{x})S^{-1}(x - \bar{x})^T$ avec les seuils

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

Dans ce cas l'estimateur de la matrice de v.c:

$$S_1 = \frac{1}{m-1} \sum_{k=1}^m (x_i - \bar{x})(x_i - \bar{x})^T$$

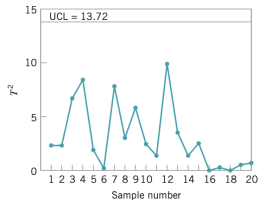
ou

$$S_2 = \frac{1}{2} \frac{VV^T}{(m-1)}$$

où $V = (v_1, \dots, v_{m-1})$, $v_i = x_{i+1} - x_i$.

Problème de contrôle qualité multivarié: T^2

Sample Number k	(a) Means		(b) Variances and Covariances			(c) Control Chart Statistics	
	Tensile Strength (\bar{x}_{1k})	Diameter (\bar{x}_{2k})	s_{1k}^2	s_{2k}^2	s_{12k}	T_k^2	$ S_k $
1	115.25	1.04	1.25	0.87	0.80	2.16	0.45
2	115.91	1.06	1.26	0.85	0.81	2.14	0.41
3	115.05	1.09	1.30	0.90	0.82	6.77	0.50
4	116.21	1.05	1.02	0.85	0.81	8.29	0.21
5	115.90	1.07	1.16	0.73	0.80	1.89	0.21
6	115.55	1.06	1.01	0.80	0.76	0.03	0.23
7	114.98	1.05	1.25	0.78	0.75	7.54	0.41
8	115.25	1.10	1.40	0.83	0.80	3.01	0.52
9	116.15	1.09	1.19	0.87	0.83	5.92	0.35
10	115.92	1.05	1.17	0.86	0.95	2.41	0.10
11	115.75	0.99	1.45	0.79	0.78	1.13	0.54
12	114.90	1.06	1.24	0.82	0.81	9.96	0.36
13	116.01	1.05	1.26	0.55	0.72	3.86	0.17
14	115.83	1.07	1.17	0.76	0.75	1.11	0.33
15	115.29	1.11	1.23	0.89	0.82	2.56	0.42
16	115.63	1.04	1.24	0.91	0.83	0.08	0.44
17	115.47	1.03	1.20	0.95	0.70	0.19	0.65
18	115.58	1.05	1.18	0.83	0.79	0.00	0.36
19	115.72	1.06	1.31	0.89	0.76	0.35	0.59
20	115.40	1.04	1.29	0.85	0.68	0.62	0.63
Averages	$\bar{\bar{x}}_1 = 115.59$	$\bar{\bar{x}}_2 = 1.06$	$\bar{s}_1^2 = 1.23$	$\bar{s}_2^2 = 0.83$	$\bar{s}_{12} = 0.79$		



EWMA multivarié: MEWMA

$$Z_i = \lambda x_i + (1 - \lambda)Z_{i-1}$$

La statistique utilisée pour le contrôle est:

$$T^2 = Z_i^T \Sigma_{Z_i}^{-1} Z_i$$

où $\Sigma_{Z_i} = \frac{\lambda}{2-\lambda}(1 - (1 - \lambda)^{2i})\Sigma$.

Hotelling est plutôt adapté pour tester deux hypothèses simples. En réalité, plus souvent on a une contre hypothèse composite. De plus, il est sensible aux changements de la variance.

Un écart est modélisé par le paramètre de non-centralité de la loi de χ^2 : $\delta = (\mu' \Sigma^{-1} \mu)^{-1}$

EWMA multivarié: MEWMA

Pour un décalage de $\delta = (\mu' \Sigma^{-1} \mu)^{-1}$ et $UCL = H$

δ	ARL ₀ =	$p = 4$		$p = 10$		$p = 20$	
		500	1000	500	1000	500	1000
0.5	λ	0.04	0.03	0.03	0.025	0.03	0.025
	H	13.37	14.68	22.69	24.70	37.09	39.63
	ARL _{min}	42.22	49.86	55.94	66.15	70.20	83.77
1.0	λ	0.105	0.09	0.085	0.075	0.075	0.065
	H	15.26	16.79	25.42	27.38	40.09	42.47
	ARL _{min}	14.60	16.52	19.29	21.74	24.51	27.65
1.5	λ	0.18	0.18	0.16	0.14	0.14	0.12
	H	16.03	17.71	26.58	28.46	41.54	43.80
	ARL _{min}	7.65	8.50	10.01	11.07	12.70	14.01
2.0	λ	0.28	0.26	0.24	0.22	0.20	0.18
	H	16.49	18.06	27.11	29.02	42.15	44.45
	ARL _{min}	4.82	5.30	6.25	6.84	7.88	8.60
3.0	λ	0.52	0.46	0.42	0.40	0.36	0.34
	H	16.84	18.37	27.55	29.45	42.80	45.08
	ARL _{min}	2.55	2.77	3.24	3.50	4.04	4.35

EWMA multivarié: MEWMA

Average Run Lengths (zero state) for the MEWMA Control Chart [from Prabhu and Runger (1997)]

p	δ	λ							
		0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.80
2	$H = 7.35$	8.64	9.65	10.08	10.31	10.44	10.52	10.58	
	0.0	199.93	199.98	199.91	199.82	199.83	200.16	200.04	200.20
	0.5	26.61	28.07	35.17	44.10	53.82	64.07	74.50	95.88
	1.0	11.23	10.15	10.20	11.36	13.26	15.88	19.24	28.65
	1.5	7.14	6.11	5.49	5.48	5.78	6.36	7.25	10.28
	2.0	5.28	4.42	3.78	3.56	3.53	3.62	3.84	4.79
	3.0	3.56	2.93	2.42	2.20	2.05	1.95	1.90	1.91
4	$H = 11.22$	12.73	13.87	14.34	14.58	14.71	14.78	14.85	
	0.0	199.84	200.12	199.94	199.91	199.96	200.05	199.99	200.05
	0.5	32.29	35.11	46.30	59.28	72.43	85.28	97.56	120.27
	1.0	13.48	12.17	12.67	14.81	18.12	22.54	28.06	42.58
	1.5	8.54	7.22	6.53	6.68	7.31	8.40	10.03	15.40
	2.0	6.31	5.19	4.41	4.20	4.24	4.48	4.93	6.75
	3.0	4.23	3.41	2.77	2.50	2.36	2.27	2.24	2.37
6	$H = 14.60$	16.27	17.51	18.01	18.26	18.39	18.47	18.54	
	0.0	200.11	200.03	200.11	200.18	199.81	200.01	199.87	200.17
	0.5	36.39	40.38	54.71	70.30	85.10	99.01	111.65	133.91
	1.0	15.08	13.66	14.63	17.71	22.27	28.22	35.44	53.51
	1.5	9.54	8.01	7.32	7.65	8.60	10.20	12.53	20.05
	2.0	7.05	5.74	4.88	4.68	4.80	5.20	5.89	8.60
	3.0	4.72	3.76	3.03	2.72	2.58	2.51	2.51	2.77
10	$H = 20.72$	22.67	24.07	24.62	24.89	25.03	25.11	25.17	
	0.0	199.91	199.95	200.08	200.01	199.98	199.84	200.12	200.00
	0.5	42.49	48.52	67.25	85.68	102.05	116.25	128.82	148.96
	1.0	17.48	15.98	17.92	22.72	29.47	37.81	47.54	69.71
	1.5	11.04	9.23	8.58	9.28	10.91	13.49	17.17	28.33
	2.0	8.15	6.57	5.60	5.47	5.77	6.48	7.68	12.15
	3.0	5.45	4.28	3.43	3.07	2.93	2.90	2.97	3.54
15	$H = 27.82$	30.03	31.59	32.19	32.48	32.63	32.71	32.79	
	0.0	199.95	199.89	200.08	200.03	199.96	199.91	199.93	200.16
	0.5	48.20	56.19	78.41	98.54	115.36	129.36	141.10	159.55
	1.0	19.77	18.28	21.40	28.06	36.96	47.44	59.03	83.86
	1.5	12.46	10.41	9.89	11.08	13.53	17.26	22.38	37.07
	2.0	9.20	7.36	6.32	6.30	6.84	7.97	9.80	16.36
	3.0	6.16	4.78	3.80	3.43	3.29	3.31	3.49	4.49

Contrôle de la variabilité

Une extension de la carte univarié s^2 avec la matrice vc Σ .
Statistique utilisée

$$W_i = -pn + pn \ln(n) - n \ln(|A_i|/|\Sigma|) + \text{tr}(\Sigma^{-1}A_i)$$

où $A_i = (n-1)S_i$, S_i la matrice v-c de l'échantillon i avec comme borne

$$UCL = \chi^2_{\alpha, p(p+1)/2}$$

Sinon nous pouvons utiliser la statistique $|S|$ où

$$\mathbb{E}(|S|) = b_1|\Sigma|, \quad \text{var}(|S|) = b_2|\Sigma|^2$$

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

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

alors ligne avec une ligne centrale $b_1|\Sigma|$ nous avons les bornes

$$UCL = |\Sigma|(b_1 + 3b_2^{1/2}), \quad LCL = |\Sigma|(b_1 - 3b_2^{1/2})$$

Problème de contrôle qualité multivarié: $|S|$

Sample Number k	(a) Means		(b) Variances and Covariances			(c) Control Chart Statistics	
	Tensile Strength (\bar{x}_{1k})	Diameter (\bar{x}_{2k})	s_{1k}^2	s_{2k}^2	s_{12k}	T_k^2	$ S_k $
1	115.25	1.04	1.25	0.87	0.80	2.16	0.45
2	115.91	1.06	1.26	0.85	0.81	2.14	0.41
3	115.05	1.09	1.30	0.90	0.82	6.77	0.50
4	116.21	1.05	1.02	0.85	0.81	8.29	0.21
5	115.90	1.07	1.16	0.73	0.80	1.89	0.21
6	115.55	1.06	1.01	0.80	0.76	0.03	0.23
7	114.98	1.05	1.25	0.78	0.75	7.54	0.41
8	115.25	1.10	1.40	0.83	0.80	3.01	0.52
9	116.15	1.09	1.19	0.87	0.83	5.92	0.35
10	115.92	1.05	1.17	0.86	0.95	2.41	0.10
11	115.75	0.99	1.45	0.79	0.78	1.13	0.54
12	114.90	1.06	1.24	0.82	0.81	9.96	0.36
13	116.01	1.05	1.26	0.55	0.72	3.86	0.17
14	115.83	1.07	1.17	0.76	0.75	1.11	0.33
15	115.29	1.11	1.23	0.89	0.82	2.56	0.42
16	115.63	1.04	1.24	0.91	0.83	0.08	0.44
17	115.47	1.03	1.20	0.95	0.70	0.19	0.65
18	115.58	1.05	1.18	0.83	0.79	0.00	0.36
19	115.72	1.06	1.31	0.89	0.76	0.35	0.59
20	115.40	1.04	1.29	0.85	0.68	0.62	0.63
Averages	$\bar{\bar{x}}_1 = 115.59$	$\bar{\bar{x}}_2 = 1.06$	$\bar{s}_1^2 = 1.23$	$\bar{s}_2^2 = 0.83$	$\bar{s}_{12} = 0.79$		

