БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ   ACADÉMIE BULGARE DES SCIENCES
ИЗВЕСТИЯ НА МАТЕМАТИЧЕСКИЯ ИНСТИТУТ
BULLETIN DE L'INSTITUT DE MATHÉMATIQUES
Том (Tome) IX

LE CONTRÔLE STATISTIQUE DE LA QUALITÉ À PLUSIEURS CARACTÈRES SIMULTANÉS[1], I.

LA MÉTHODE DE LA MOYENNE ET LA MÉTHODE DE LA VARIANCE GÉNÉRALISÉE

R. Theodorescu et I. Văduva

Les problèmes du contrôle statistique de la qualité en cours de fabrication, traités dans la littérature, considèrent pour la plupart un seul caractère mesurable (voir, par exemple [3], [6], [7], [8]; dans [2] on fait mention seulement de quelques aspects concernant plusieurs caractères). L'extension à plusieurs caractères mesurables se heurte à des difficultés d'ordre théorique ainsi que d'ordre pratique en ce qui concerne la mise en œuvre du contrôle proprement dit.

Les avantages du contrôle statistique à plusieurs caractères simultanés, sont multiples; il suffit en premier lieu de montrer qu'un tel contrôle tient compte des corrélations existant entre les différents caractères mesurables donnés. Ainsi la surveillance du fonctionnement des mécanismes exécutant les caractères mesurables donnés regarde l'ensemble de ces mécanismes, qui s'influent et se conditionnent les uns les autres pendant la fabrication.

L'introduction du contrôle statistique en cours de fabrication comprend deux stades: (a) l'analyse statistique préliminaire de la fabrication et (b) le choix de la méthode de contrôle et son application. Dans ce qui suit nous nous proposons d'examiner en détail le cas du contrôle à deux caractères simultanés; les résultats peuvent être étendus sans difficultés essentielles à un nombre quelconque de caractères.

1. Le stade de l'analyse statistique préliminaire de la fabrication pose deux problèmes théoriques importants: l'étude des échantillons prélevés et l'estimation de la variabilité (c'est-à-dire l'estimation de la fonction de répartition simultanée des caractères considérés) de la fabrication.

Tout d'abord, nous allons nous occuper du problème des échantillons prélevés. L'étude des prélèvements, dont nous disposons pour l'analyse de la fabrication, montre que très souvent ils ne s'effectuent pas au hasard. Pour vérifier si les prélèvements sont effectués au hasard, nous étendrons quelques résultats classiques concernant les itérations déterminées par la médiane expérimentale. Soit donc Δ l'ensemble des valeurs d'un vecteur aléatoire à deux dimen-

[1] Conférence faite au Colloque sur les applications des mathématiques dans l'industrie tenu en Septembre 1964 à Budapest.

sions $\xi=(\xi_1, \xi_2)$, défini sur un espace de probabilité donné $(\Omega, \mathfrak{K}, P)$, où ξ_1 et ξ_2 représentent dans les applications des caractères mesurables de la même pièce.

D é f i n i t i o n 1. On appelle courbe médiane du vecteur aléatoire ξ une courbe connexe Γ du plan x_1Ox_2 qui sépare l'ensemble Δ en deux parties Δ_1, Δ_2 telles que

$$P(\xi \in \Delta_1) \geq \frac{1}{2} \leq P(\xi \in \Gamma \cup \Delta_2).$$

Evidemment, la courbe médiane ainsi définie n'est pas déterminée de façon unique; elle peut être précisée si elle vérifie des conditions supplémentaires.

Remarquons que si nous prenions comme courbes médianes les frontières des semi-intervalles à deux dimensions ouverts à gauche, c'est-à-dire les frontières des ensembles

$$\{(z_1, z_2);\ z_1 < x_1,\ z_2 < x_2\},$$

le lieu géométrique des points (x_1, x_2) est donné par les relations

$$P(\xi_1 \geq x_1,\ \xi_2 \geq x_2) \geq \frac{1}{2} \leq P(\xi_1 \leq x_1, \xi_2 \leq x_2),$$

ou, sous une forme équivalente,

$$F(x_1, x_2) \leq \frac{1}{2},\ F(x_1+0,\ x_2+0) \geq \frac{1}{2},$$

où F est la fonction de répartition du vecteur aléatoire ξ. En supposant que la fonction de répartition E est continue et croissante par rapport à chaque variable, le lieu cherché dans le plan x_1Ox_2 est la courbe donnée par l'équation

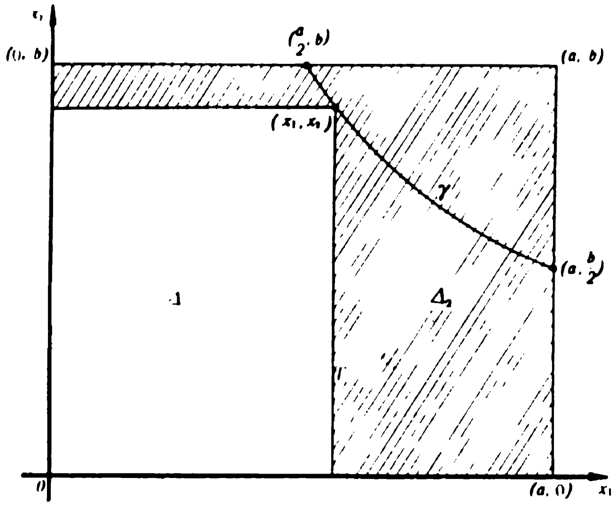

Fig. 1. Droite médiane pour la fonction de répartition uniforme

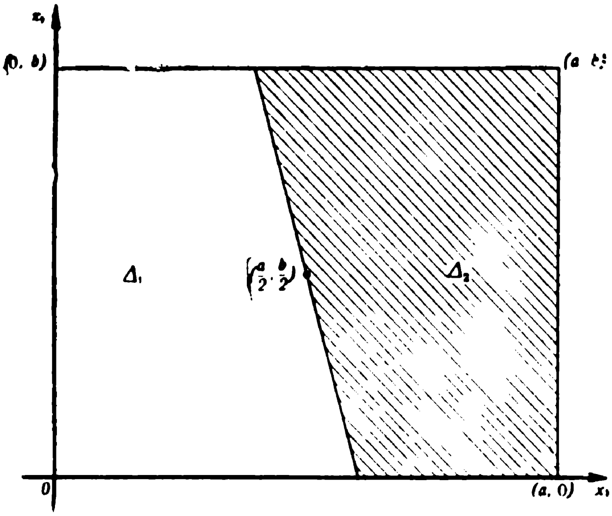

Fig. 2. Le lieu géométrique $F(x_1, x_2)=1/2$ correspondant à la fonction de répartition uniforme

(1) $$F(x_1, x_2)=\frac{1}{2}.$$

E x e m p l e s. 1°. Supposons que le vecteur aléatoire considéré $\xi=(\xi_1, \xi_2)$ prenne des valeurs dans le rectangle $\Delta=[0, a]\times[0, b]$, $a, b>0$ et ait une fonction de répartition uniforme; dans ce cas, tout arc de courbe qui divise le rectangle Δ en deux parties de même aire, en particulier toute droite qui passe par le centre $\left(\frac{a}{2}, \frac{b}{2}\right)$ du rectangle, est une courbe médiane du vecteur aléatoire $\boldsymbol{\xi}$ au sens de la définition 1 (fig. 1). De même, l'équation (1) devient dans ce cas (fig. 2)

$$\gamma: \quad x_1 x_2 = \frac{1}{2} ab.$$

2°. Supposons que le vecteur aléatoire $\boldsymbol{\xi} = (\xi_1, \xi_2)$ ait une fonction de répartition normale $N(\boldsymbol{m}, \boldsymbol{\Sigma})$, où $\boldsymbol{m} = (m_1, m_2)$ est le vecteur moyenne, et $\boldsymbol{\Sigma}$ la matrice de covariance,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \sigma_i^2 = \mathbf{E}((\xi_i - m_i)^2),\ i = 1, 2,\ \varrho = \frac{\mathbf{E}((\xi_1 - m_1)(\xi_2 - m_2))}{\sigma_1\sigma_2}.$$

Dans ce cas la densité de répartition est[1]

$$(2)\quad f(\boldsymbol{x}) = \frac{1}{2\pi (\det \boldsymbol{\Sigma})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{m})\right) = \frac{1}{2\pi (\det \boldsymbol{\Sigma})^{\frac{1}{2}}} e(\boldsymbol{x}),$$

ou, écrite à l'aide des coordonnées,

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\frac{(x_1 - m_1)^2}{\sigma_1^2} - 2\varrho\frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_1\sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2}\right]\right\}$$

$$(3)\qquad = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} e(x_1, x_2).$$

Soit $\boldsymbol{d} = (d_1, d_2)$ un verseur donné du plan. Nous pouvons prendre comme courbe médiane Γ du vecteur aléatoire $\boldsymbol{\xi}$ la droite à l'équation[2]

$$\Gamma: \quad (\boldsymbol{d}, \boldsymbol{x} - \boldsymbol{m}) = 0.$$

En effet, posons

$$\Delta_1 = \{\boldsymbol{x}; (\boldsymbol{d}, \boldsymbol{x}) < (\boldsymbol{d}, \boldsymbol{m})\}, \quad \Delta_2 = \{\boldsymbol{x}; (\boldsymbol{d}, \boldsymbol{x}) > (\boldsymbol{d}, \boldsymbol{m})\};$$

pour $\boldsymbol{x}' \in \Delta_1$ arbitraire, soit $\boldsymbol{x}''$ le symétrique du point $\boldsymbol{x}'$ par rapport au point $\boldsymbol{m}$. Evidemment, $\boldsymbol{x}'' \in \Delta_2$ et

$$f(\boldsymbol{x}') = f(\boldsymbol{x}'').$$

Il résulte alors

$$P(\boldsymbol{\xi} \in \Delta_1) = \int_{\Delta_1} f(\boldsymbol{x}')\, d\boldsymbol{x}' = \int_{\Delta_2} f(\boldsymbol{x}'')\, d\boldsymbol{x}'' = P(\boldsymbol{\xi} \in \Delta_2).$$

---

[1] Nous avons noté respectivement par $\boldsymbol{A}^{-1}$ et $\boldsymbol{A}'$ l'inverse et la transposée de la matrice $\boldsymbol{A}$.

[2] Si $\boldsymbol{a} = (a_1, a_2)$, $\boldsymbol{b} = (b_1, b_2)$, $(\boldsymbol{a}, \boldsymbol{b}) = a_1 b_1 + a_2 b_2$ représente le produit scalaire des vecteurs $\boldsymbol{a}$ et $\boldsymbol{b}$.

Parce que

$$\int_{\Delta_1 \cup \Gamma \cup \Delta_2} f(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\Delta_1 \cup \Delta_2} f(\boldsymbol{x})\, d\boldsymbol{x} = 1,$$

nous concluons que

$$P(\boldsymbol{\xi} \in \Delta_1) = P(\boldsymbol{\xi} \in \Gamma \cup \Delta_2) = \frac{1}{2},$$

d'où il s'ensuit que la droite Γ est une courbe médiane du vecteur aléatoire $\boldsymbol{\xi}$ au sens de la définition 1. De plus, pour un verseur donné $\boldsymbol{d}$ la droite médiane Γ est, évidemment, déterminée de façon unique. Lorsque le verseur $\boldsymbol{d}$ varie, nous déduisons que les droites médianes forment un faisceau de centre $\boldsymbol{m}$.

3°. Supposons que le vecteur aléatoire $\boldsymbol{\xi} = (\xi_1, \xi_2)$ ait une fonction de répartition arbitraire et fixons un verseur $\boldsymbol{d} = (d_1, d_2)$. L'exemple 2 nous suggère de considérer la variable aléatoire à une seule dimension

$$\eta = (\boldsymbol{d}, \xi);$$

notons par φ la médiane de cette variable aléatoire. Si nous posons

$$\Delta_1 = \{\boldsymbol{x}; (\boldsymbol{d}, \boldsymbol{x}) < \varphi\}, \quad \Delta_2 = \{\boldsymbol{x}; (\boldsymbol{d}, \boldsymbol{x}) > \varphi\}$$

et que Γ soit la droite

$$\Gamma: \quad (\boldsymbol{d}, \boldsymbol{x}) = \varphi$$

on voit aisément que Γ est une courbe médiane du vecteur aléatoire $\boldsymbol{\xi}$ au sens de la définition 1.

Passons maintenant à la médiane expérimentale. Soit donc[1] $\boldsymbol{x}_i^0 = (x_{1,i}^0, x_{2,i}^0)$, $1 \leq i \leq n^0$, un échantillon d'effectif n^0 correspondant au vecteur aléatoire $\boldsymbol{\xi} = (\xi_1, \xi_2)$; au vecteur aléatoire $\boldsymbol{\xi}$ nous associons ensuite le vecteur expérimental $\boldsymbol{\xi}_{(n^0)}$ qui prend les valeurs $\boldsymbol{x}_i^0$, $1 \leq i \leq n^0$, avec des probabilités égales à $\frac{1}{n^0}$. Nous arrivons donc naturellement à la

Définition 2. On appelle courbe médiane expérimentale du vecteur aléatoire $\boldsymbol{\xi}$ une courbe médiane du vecteur expérimental $\boldsymbol{\xi}_{(n^0)}$.

Dans les applications il est utile de faire appel à une droite médiane expérimentale. Une fois donné le verseur $\boldsymbol{d}$, nous considérons la médiane de la variabe aléatoire à une seule dimension

$$(\boldsymbol{d}, \boldsymbol{\xi}_{(n^0)}) = \eta_{(n^0)}.$$

En désignant par z, ou plus précisément, par $z_{(n^0)}$ la médiane des valeurs

$$(\boldsymbol{d}, \boldsymbol{x}_i^0) = z_i, \quad 1 \leq i \leq n^0,$$

la droite médiane expérimentale a l'équation

$$(\boldsymbol{d}, \boldsymbol{x}) = z_{(n^0)}.$$

Si nous supposons que le vecteur $\boldsymbol{\xi}$ admet la densité de répartition f, la variable aléatoire η admet à son tour une densité de répartition g, d'où nous

---

[1] Nous avons mis l'indice supérieur ^{0} pour marquer toutes les grandeurs qui se réfèrent à l'analyse statistique préliminaires de la fabrication.

en déduisons ([4], p. 369) que la médiane expérimentale $z_{(n^0)}$ de la variable aléatoire η est asymptotiquement normale $N\left(\varphi, \frac{1}{2g(\varphi)\sqrt{n_0}}\right)$. Dans ce cas

$$\mathbf{E}(z_{(n^0)}) = \varphi, \lim_{n^0 \to \infty} \mathbf{D}^2 (z_{(n^0)}) = 0,$$

d'où il résulte que la médiane expérimentale $z_{(n^0)}$ converge en probabilité vers la médiane théorique φ lorsque $n^0 \to \infty$. Cela signifie que la droite médiane expérimentale déterminée par le verseur $\boldsymbol{d}$ peut être considérée comme une estimation absolument correcte[1] de la droite médiane théorique déterminée par le même verseur $\boldsymbol{d}$.

La notion de droite médiane expérimentale nous aide à vérifier si les prélèvements sont effectués au hasard. Reprenons donc l'échantillon $\boldsymbol{x}_1^0, \ldots \boldsymbol{x}_{n^0}^0$ considéré plut haut. Etant donné le verseur $\boldsymbol{d}$, le problème revient à l'étude des itérations déterminées par les valeurs $z_1, \ldots, z_{n^0}$ par rapport à la médiane expérimentale $z_{(n^0)}$. Nous supposons que les valeurs $\boldsymbol{x}_1^0, \ldots, \boldsymbol{x}_{n^0}^0$ sont écrites dans l'ordre de leur observation. Tenant compte de l'ordre des valeurs correspondantes z_i, $1 \leq i \leq n^0$, nous pouvons écrire les itérations respectives, ce qui nous permet d'appliquer les tests classiques pour une seule dimension qui utilisent soit le nombre total des itérations soit la longueur maximale des itérations (évidemment, lorsque $n^0 > 30$). Il nous reste à préciser enfin la manière la plus convenable de choisir le verseur $\boldsymbol{d}$ déterminant la droite médiane; on remarque aisément que le verseur $\boldsymbol{d}$ de la petite axe de l'ellipse de concentration s'impose comme le plus indiqué.

2. Un autre problème se reliant à l'analyse statistique préliminaire est l'estimation de la variabilité de la fabrication, c'est-à-dire la détermination de la fonction de répartition simultanée des caractères considérés et l'indication des tests de nullité correspondants. Quoique il y ait de tels tests suffisamment généraux, nous allons nous borner dans ce qui suit, pour fixer les idées, au cas des fonctions de répartition normales, les plus souvent rencontrées en pratique.

Un des tests de nullité connus, d'assez large portée, est le test général χ^2. Supposons donc que les composantes ξ_1 et ξ_2 du vecteur aléatoire $\boldsymbol{\xi}$ représentent des caractères mesurables continus de la même pièce et considérons une partition $(\Delta_i)_{1 \leq i \leq l}$ du domaine Δ. Soit maintenant H l'hypothèse statistique qui affirme que la fonction de répartition du vecteur $\boldsymbol{\xi}$ est $F(\boldsymbol{x}; \boldsymbol{\alpha})$, où $\boldsymbol{x} = (x_1, x_2)$ et $\boldsymbol{\alpha} = (a_1, \ldots, a_k)$ $(k < l)$ sont des paramètres inconnus qui peuvent être estimés comme nous allons le voir.

Introduisons les notations

$$p_i(\boldsymbol{\alpha}) = p_i(a_1, \ldots, a_k) = P(\boldsymbol{\xi} \in \Delta_i), \quad 1 \leq i \leq l;$$

nous supposons que les fonctions p_i, $1 \leq i \leq l$, sont dérivables par rapport aux arguments a_j, $1 \leq j \leq k$.

---

[1] Soient Γ et $(\Gamma_n)_{n \in N^*}$ une courbe et une suite de courbes du plan engendrées par le même groupe de mouvements; nous allons dire que la suite $(\Gamma_n)_{n \in N^*}$ converge en probabilité vers la courbe Γ si les paramètres qui engendrent la suite des courbes considérées convergent en probabilité vers les paramètres de la courbe limite lorsque $n \to \infty$.

Reprenons l'échantillon d'effectif n^0 suffisamment grand $(x_1^0, \ldots, x_{n^0}^0)$, $x_i^0 = (x_{1,i}^0, x_{2,i}^0)$, $1 \leq i < n^0$, correspondant au vecteur aléatoire à deux dimensions ξ, notons par ν_i le nombre des valeurs de l'échantillon qui appartiennent au domaine Δ_i, $1 \leq i \leq l$, et formons l'expression

(4)
$$\chi^2 = \sum_{i=1}^{l} \frac{(\nu_i - n^0 p_i(\alpha))^2}{n^0 p_i(\alpha)},$$

où nous supposons que

(5)
$$p_i(\alpha) > c^2 > 0, \quad 1 \leq i \leq l.$$

Notons par $d_1^*, \ldots, d_k^*$ les estimations des paramètres $a_1, \ldots, a_k$ obtenues par la minimisation de l'expression (4), c'est-à-dire par la méthode du minimum de χ^2 ([4], p. 425). En supposant que nous ayons choisi les probabilités $p_i(\alpha)$, $1 \leq i \leq l$, telles que les dénominateurs qui apparaissent dans l'expression (4) soient approximativement constants (condition satisfaite si, par exemple, l'effectif n^0 de l'échantillon est suffisamment grand, en vertue de la condition (5)), la méthode du minimum de χ^2 se simplifie, conduisant à la méthode du minimum de χ^2 modifiée ([4], p. 426). Dans certaines conditions de régularité ([4], pp. 426—427), la variable aléatoire

(6)
$$\chi^{*2} = \sum_{i=1}^{l} \frac{(\nu_i - n^0 p_i^*)^2}{n^0 p_i^*}$$

où

$$p_i^* = p_i(d^*) = p_i(d_1^*, \ldots, d_k^*), \quad 1 \leq i \leq l,$$

suit asymptotiquent la fonction de répartition χ^2 à $l-k-1$ degrés de liberté

Le test de nullité pour la vérification de l'hypothèse statistique H est basé sur la fonction définie par la formule (6). Pour appliquer ce test, il faut d'abord connaître les estimations des paramètres a_j, $1 \leq j \leq k$, et trouver des procédés simples pour déterminer les probabilités p_i^*, $1 \leq i \leq l$. Dans ce qui suit nous allons analyser la fonction de répartition normale à deux dimensions.

Soit donc la densité de répartition normale à deux dimensions (2) (ou (3)); dans ce cas le nombre des paramètres est $k = 5$ $(m_1, m_2, \sigma_1, \sigma_2, \varrho)$. A l'aide de l'échantillon considéré correspondant au vecteur aléatoire ξ nous estimons les matrices $\boldsymbol{m}$ et $\boldsymbol{\Sigma}$ par la méthode du minimum de χ^2 modifiée qui est basée sur la résolution du système d'équations

$$\sum_{j=1}^{l} \frac{\nu_j}{p_j} \frac{\partial p_j}{\partial a_i} = 0, \quad 1 \leq i \leq k = 5.$$

Ce système nous donne

$$m_i^* = \frac{1}{n^0} \sum_{j=1}^{l} \nu_j \frac{\int\limits_{\Delta_j} x_i e(x_1, x_2)\, dx_1 dx_2}{\int\limits_{\Delta_j} e(x_1, x_2)\, dx_1\, dx_2} \qquad i = 1, 2,$$

$$\sigma_i^{*2} = \frac{1}{n^0} \sum_{j=1}^{l} \nu_j \frac{\int_{\Delta_j} (x_i - m_i^*)^2 e(x_1, x_2)\, dx_1 dx_2}{\int_{\Delta_j} e(x_1, x_2)\, dx_1 dx_2}, \quad i = 1, 2,$$

$$\varrho^* = \frac{1}{n^0 \sigma_1^* \sigma_2^*} \sum_{j=1}^{l} \nu_j \frac{\int_{\Delta_j} (x_1 - m_1^*)(x_2 - m_2^*)\, e(x_1, x_2)\, dx_1 dx_2}{\int_{\Delta_j} e(x_1, x_2)\, dx_1 dx_2}.$$

Si les domaines Δ_i, $1 \leq i \leq l$, sont suffisamment petits, nous pouvons obtenir pour ces valeurs des expressions simples si nous remplaçons les fonctions sous l'intégrale par leurs valeurs en des points fixés des domaines respectifs; nous avons

$$m_i^* = \frac{1}{n^0} \sum_{j=1}^{l} \nu_j y_{i,j}, \quad \sigma_i^{*2} = \frac{1}{n^0} \sum_{j=1}^{l} \nu_j (y_{i,j} - m_i^*)^2, \quad i = 1, 2,$$

$$\varrho^* = \frac{1}{n^0 \sigma_1^* \sigma_2^*} \sum_{j=1}^{l} \nu_j (y_{1,j} - m_1^*)(y_{1,j} - m_2^*),$$

où $y_i = (y_{1,i}, y_{2,i}) \in \Delta_i$, $1 \leq i \leq l$.[1] Il en résulte que pour les matrices $\boldsymbol{m}$ et $\boldsymbol{\Sigma}$ nous pouvons prendre, respectivement, les estimations

$$(7) \qquad \boldsymbol{m}^* = (m_1^*, m_2^*), \quad \boldsymbol{\Sigma}^* = \begin{pmatrix} \sigma_1^{*2} & \varrho^* \sigma_1^* \sigma_2^* \\ \varrho^* \sigma_1^* \sigma_2^* & \sigma_1^{*2} \end{pmatrix},$$

qui sont, comme on le voit, identiques à celles obtenues par le groupement des valeurs de l'échantillon. Mais les estimations obtenues par le groupement des valeurs de l'échantillon donnent des erreurs négligeables par rapport aux estimations de maximum de vraisemblance, données, respectivement, par les formules usuelles suivantes

$$(8) \qquad \bar{\boldsymbol{x}}^0 = (\bar{x}_1^0, \bar{x}_2^0), \quad \boldsymbol{S}^0 = \begin{pmatrix} s_1^{02} & r^0 s_1^0 s_2^0 \\ r^0 s_1^0 s_2^0 & s_2^{02} \end{pmatrix},$$

où

$$\bar{x}_i^0 = \frac{1}{n^0} \sum_{j=1}^{n^0} x_{i,j}^0, \quad s_i^{02} = \frac{1}{n_0} \sum_{j=1}^{n^0} (x_{i,j}^0 - \bar{x}_i^0)^2, \quad i = 1, 2,$$

$$r^0 = \frac{1}{n^0 s_1^0 s_2^0} \sum_{j=1}^{n^0} (x_{1,j}^0 - \bar{x}_1^0)(x_{2,j}^0 - \bar{x}_2^0).$$

---

[1] Comme dans le cas d'un seul caractère ([4], 438), on peut appliquer des corrections aux estimations considérées.

Considérons maintenant l'ellipse de contour

$$\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{m})=0,\quad \theta>0, \tag{9}$$

et notons par E_θ la région formée par l'intérieur de cette ellipse (désormais nous comprendrons par ellipse la région E_θ). Parce que la variable aléatoire $\chi^2=(\boldsymbol{\xi}-\boldsymbol{m})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}-\boldsymbol{m})$ a une fonction de répartition χ^2 à deux degrés de liberté, un calcul simple nous donne

$$P(\boldsymbol{\xi}\in E_\theta)=P\left(\frac{1}{2}\chi^2<\theta\right)=1-e^{-\theta}. \tag{10}$$

D'autre part, soient donnés les nombres $0<\theta_1<\cdots<\theta_2=\infty$; définissons la partitition $(\Delta_i)_{1\leq i\leq l}$ qui correspond au test de nullité χ^2 par les relations

$$\Delta_1=E_{\theta_1},\quad \Delta_j=E_{\theta_j}-E_{\theta_{j-1}},\quad 1<j\leq l,$$

à condition que $l>k+1$, c'est-à-dire $l>6$. En remplaçant les estimations (7) ou (8) dans la formule (9) et en utilisant la relation (10), nous obtenons aisément les probabilités

$$p_1^*=P(\boldsymbol{\xi}\in E_{\theta_1})=1-e^{-\vartheta_1},$$

$$p_j^*=P(\boldsymbol{\xi}\in E_{\theta_j}-E_{\theta_{j-1}})=e^{-\vartheta_{j-1}}-e^{-\vartheta_j},\quad 1<j\leq l.$$

3. On sait que pour la surveillance optimale de la fabrication, il faut effectuer la contrôle eu cours de fabrication d'un paramètre donnant la tendance centrale, et d'un paramètre donnant la dispersion des machines qui interviennent dans la fabrication considérée. Evidemment, dans ce qui suit nous admettons qui la variabilité de la machine est normale.

Le plus important et le plus efficace des paramètres qui donnent la tendance centrale de la machine pour le cas du contrôle statistique à un seul caractère, est la moyenne expérimentale. Il s'impose donc naturellement de considérer pour le contrôle de la tendance centrale correspondant à deux caractères simultanés le vecteur moyen expérimental, calculé à l'aide d'une formule analogue à la première formule (8) en vertu d'un échantillon d'effectif n, où n est beaucoup plus faible que n^0.

Pour établir la méthode de contrôre de la tendance centrale, il faut tenir compte premièrement de la fonction de répartition du vecteur $\bar{\boldsymbol{x}}=(\bar{x}_1,\bar{x}_2)$. Si les valeurs $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ de l'échantillon sont supposées indépendantes, ce qui est tout à fait justifié du point de vue pratique, le vecteur $\bar{\boldsymbol{x}}$ a la fonction de répartition normale à deux dimensions avec le vecteur moyen $\boldsymbol{m}=(m_1, m_2)$ et la matrice de covariance

$$\boldsymbol{\Sigma}_n=\begin{pmatrix}\dfrac{\sigma_1^2}{n} & \dfrac{\varrho\sigma_1\sigma_2}{n}\\[2ex] \dfrac{\varrho\sigma_1\sigma_2}{n} & \dfrac{\sigma_2^2}{n}\end{pmatrix},\quad \boldsymbol{\Sigma}_n^{-1}=n\,\boldsymbol{\Sigma}^{-1}; \tag{11}$$

donc l'expression de la densité de répartition du vecteur $\bar{\boldsymbol{x}}$ sera

$$(12) \qquad f(x) = \frac{1}{2\pi(\det \Sigma_n)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-m)'\Sigma_n^{-1}(x-m)\right).$$

Notons par $t=(t_1, t_2)$ le milieu de l'intervalle de tolérance à deux dimensions $T=[T_{11}, T_{12}]\times[T_{21}, T_{22}]$, où $[T_{11}, T_{12}]$ et $[T_{21}, T_{22}]$ sont respectivement les intervalles de tolérance donnés pour les caractères ξ_1 et ξ_2. Nous admettons qu'on va régler la machine sur le milieu de l'intervalle de tolérance à deux dimensions, c'est-à-dire que la condition $m=t$ est réalisée. Soit alors α_1 une probabilité suffisamment faible, considérons l'ellipse

$$\frac{1}{2}(\bar{x}-t)'\Sigma_n^{-1}(x-t)=\theta(\alpha_1),$$

dont l'intérieur sera designé par $E_{n,\theta(\alpha_1)}$, et choisissons le nombre $\theta(\alpha_1)$ tel que

$$(13) \qquad P(x \in E_{n,\theta(\alpha_1)}) = 1-\alpha_1 = 1-e^{-\theta(\alpha_1)}.$$

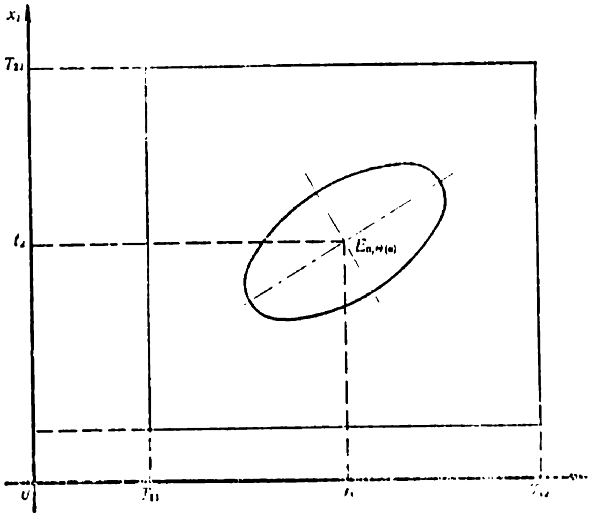

Fig. 3. L'ellipse pour le contrôle de la tendance centrale

Parce que la variable aléatoire $\chi^2=(\bar{x}-t)'\Sigma_n^{-1}(\bar{x}-t)$ a une fonction de répartition χ^2 à deux degrés de liberté, l en résulte, conformément à la formule (10), la relation suivante

$$(14) \qquad P(\bar{x} \in E_{n,\theta(\alpha_1)}) = 1-\alpha_1 = P\left(\frac{1}{2}\chi^2<\theta(\alpha_1)\right) = 1-e^{-\theta(\alpha_1)}.$$

Supposons de plus que

$$(15) \qquad E_{n,\theta(\alpha_1)} \subset T;$$

cette condition revient du point de vue pratique au fait que la machine considérée est adaptée à son travail.

L'ellipse $E_{n,\theta(\alpha_1)}$ peut être utilisée pour le contrôle de la tendance centrale comme il suit. On fixe n et on considère les matrices m et Σ_n estimées à l'aide des relations (7) ou (8) et à l'aide de la formule (11). De la relation (13), pour un α_1 fixé, on obtient $\theta(\alpha_1)$ et on représente graphiquement l'ellipse $E_{n,\theta(\alpha_1)}$ comme dans la fig. 3.

On prélève ensuite n pièces à des intervalles de temps à peu près égaux et on détermine le vecteur $\bar{x}$. Si $\bar{x} \in E_{n,\theta(\alpha_1)}$, le réglage de la machine est correct et la fabrication peut continuer. Par contre, si $\bar{x} \overline{\in} E_{n,\theta(\alpha_1)}$, le réglage de la machine ne correspond plus aux conditions technologiques d'où il s'ensuit que la fabrication ne peut continuer; dans ce cas on effectue un nouveau réglage de la machine et on contrôle à 100% la fabrication depuis le dernier contrôle.

Usine	**Fiche de contrôle statistique pour la tendance centrale**				Pièce	
Section					Effectif de l'échnantillon $n=$	
Machine					Prélèvements toutes les... minutes	
Appareils de mesure					Caractères	Tolérances
Précision des appareils de mesure	α_1	$\theta(\alpha)_1$	β_1	λ_0^2	ξ_1	$[T_{11}, T_{12}]$
Contrôleur					ξ_2	$[T_{21}, T_{22}]$

No de l'échantillon		1		2	
Valeurs de θ_{exp}	θ_{TC}				
	2				
	1				
	0				
Valeurs de l'échantillon	1	$x_{1,1}$	$x_{2,1}$		
	.	.			
	n	$x_{1,n}$	$x_{2,n}$		
Somme		$\sum_{i=1}^{n} x_{1,i}$	$\sum_{i=1}^{n} x_{2,i}$		
Moyenne		$\bar{x}_1$	$\bar{x}_2$		
θ expérimental		θ_{exp}			
Date	Jour, mois, an				
	heure				
Conclusion du contrôleur					

Fig. 4. La fiche θ_{TC}

La fig. 3 nous suggère l'utilisation d'une fiche de contrôle que nous allons appeler „fiche θ_{TC}" (la fiche θ-tendance centrale) (voir fig. 4). Sur cette fiche nous traçons la limite inférieure de contrôle 0 et la limite supérieure de contrôle $\theta_{TC}=\theta(\alpha_1)$, où θ_{TC} est la solution de l'équation en θ

$$1-\alpha_1=1-e^{-\theta}$$

Connaissant les matrices $\boldsymbol{t}$ et $\boldsymbol{\Sigma}_n$, nous calculons la grandeur $\theta_{\exp}$[1] à l'aide de la formule

$$\theta_{\exp}=(\bar{\boldsymbol{x}}-\boldsymbol{t})'\boldsymbol{\Sigma}_n^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{t}) \tag{16}$$

et nous la portons sur la fiche. Si $\theta_{\exp}<\theta_{TC}$, c'est-à-dire si $\theta_{\exp}$ est situé entre les limites de contrôle, le réglage de la machine est correct; par contre si $\theta_{\exp}\geq\theta_{TC}$ la machine doit être arrêtée et réglée.

Il est bien connu que l'application du contrôle statistique est susceptible d'erreurs caractérisées par les risques de 1re et de 2e espèce. La grandeur α_1 [voir la formule (14)] représente le risque de 1re espèce, c'est-à-dire la probabilité que le vecteur moyenne $\bar{\boldsymbol{x}}$ prenne des valeurs extérieures à l'ellipse de contrôle, quand le réglage de la machine est correct. On a donc le risque d'arrêter la machine bien que son réglage soit correct.

De même on peut laisser fonctionner la machine quand son réglage n'est pas correct. Une telle situation se produit quand le vecteur $\bar{\boldsymbol{x}}$ prend des valeurs intérieures à l'ellipse de contrôle après une modification du réglage de la machine; dans ce cas nous sommes en présence du risque de 2e espèce, noté par β_1. En pratique il faut s'assurer que le risque β_1 est suffisamment faible; il faut donc savoir comment calculer ce risque.

Nous remarquons tout d'abord que si le réglage n'est par correct, la fonction de répartition du vecteur aléatoire sera $N(\boldsymbol{m}, \boldsymbol{\Sigma}_n)$ où $\boldsymbol{m}\neq\boldsymbol{t}$. Dans ce cas la variable aléatoire

$$(\bar{\boldsymbol{x}}-\boldsymbol{t})'\boldsymbol{\Sigma}_n^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{t})=n(\bar{\boldsymbol{x}}-\boldsymbol{t})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{t})$$

a la fonction de répartition χ^2 non-centrée à deux degrés de liberté et à paramètre de non-centricité $n(\boldsymbol{m}-\boldsymbol{t})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}-\boldsymbol{t})=\lambda^2$. La grandeur λ^2 représente, du point qe vue statistique, la „distance" entre les fonctions de répartition $N(\boldsymbol{m}, \boldsymbol{\Sigma}_n)$ et $N(\boldsymbol{t}, \boldsymbol{\Sigma}_n)$. Évidemment, le réglage de la machine sera correct tant que $\lambda^2\leq\lambda_0^2$, où λ_0 est une valeur imposée par des considérations d'ordre pratique. La valeur λ_0^2 détermine en fait la zone de réglage à deux dimensions ou l'erreur limite de réglage admissible. Notons par $\chi^2_{2,\lambda}$ la variable χ^2 non-centrée, introduite plus haut; alors le risque de 2e espèce β_1 sera

$$\beta_1=P_{\lambda_0}(\bar{\boldsymbol{x}}\in E_{n,\theta(\alpha_1)})=P\left(\frac{1}{2}\chi^2_{2,\lambda_0}<\theta(\alpha_1)\right). \tag{17}$$

En conclusion pour la construction de la fiche θ_{TC}, il faut choisir dès le commencement les nombres n, α_1 et λ_0^2 tels que les risques de 1re et de 2e espèce, respectivement α_1 et β_1, soient convenablement faibles. Si les nécessités pratiques imposent un contrôle rigoureux, nous déterminons pour

---

[1] En pratique, cette grandeur peut être calculée suffisamment vite, par exemple, pour $n=7$.

un β_1 donné la grandeur $\theta(\alpha_1)$ de la formule (17) et ensuite nous construisons l'ellipse de contrôle $E_{n,\theta(\alpha_1)}$; conformément à la première égalité (14), cette ellipse nous permet de déterminer le risque de 1re espèce α_1.

4. Pour le contrôle statistique en cours de fabrication de la dispersion de la machine nous allons indiquer une méthode que utilise la variance généralisée.

Etant donné un vecteur aléatoire $\boldsymbol{\xi}$ à deux dimensions qui a une fonction de répartition normale $N(\boldsymbol{m}, \boldsymbol{\Sigma})$, la variance généralisée est définie par la relation

$$\sigma^2 = \det \boldsymbol{\Sigma}.$$

La variance généralisée expérimentale est donnée par la relation

$$s^2 = \det \boldsymbol{S},$$

où $\boldsymbol{S}$ est la matrice de covariance expérimentale et

$$\boldsymbol{S} = \frac{1}{n-1} \boldsymbol{A}, \quad \boldsymbol{A} = \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'.$$

Il s'ensuit que

$$\det \boldsymbol{S} = \frac{1}{(n-1)^2} \det \boldsymbol{A}$$

et tenant compte que

$$\det \boldsymbol{A} = \det \boldsymbol{\Sigma}\, \chi^2_{n-1}\, \chi^2_{n-2}$$

nous obtenons

$$(n-1)\left(\frac{\det \boldsymbol{S}}{\det \boldsymbol{\Sigma}}\right)^{1/2} = \chi_{n-1}\, \chi_{n-2}.$$

Ici χ^2_{n-1} et χ^2_{n-2} sont des variables aléatoires ayant des fonctions de répartition χ^2 à $n-1$ et $n-2$ degrés de liberté, respectivement.

Si nous notons

$$V = \frac{\det \boldsymbol{S}}{\det \boldsymbol{\Sigma}},$$

on sait [1] que la variable aléatoire $\widetilde{V} = 2\sqrt{V}$ a une fonction de répartition χ^2 à $2n-4$ degrés de liberté.

Dans ce qui suit nous prendrons comme paramètre donnant la dispersion l'expression

$$\widetilde{V} = 2\sqrt{V}.$$

Soit maintenant α_2, $0 < \alpha_2 < 1$, une probabilité suffisamment faible. À l'aide des tables pour la fonction de répartition χ^2, nous pouvons alors déterminer la valeur $V^2_{\alpha_2} = \theta_D$ telle que

$$P(\widetilde{V} < \widetilde{V}_{\alpha_2}) = 1 - \alpha_2. \tag{18}$$

Pour le contrôle de la dispersion on procède ainsi: pour chaque échantillon $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ d'effectif $n > 2$ fixé, prélevé à des intervalles de temps à peu

Usine	**Fiche de contrôle statistique pour la dispersion**	Pièce
Section		Effectif de l'échantillon $n=$
Machine		Prélèvements toutes les . . . minutes
Appareils de mesure		Caractères | Tolérances
Précision des appareils de mesure	α_2 | det $\boldsymbol{\Sigma}$ | β_2 | δ_0	ξ_1 | $[T_{11}, T_{12}]$
Contrôleur		ξ_2 | $[T_{21}, T_{22}]$

No de l'échantillon		1		2	
Valeurs de $\tilde{V}$	θ_D				
	2				
	1				
	0				
Valeurs de l'échantillon	1	$x_{1,1}$	$x_{2,1}$		
	.	.	.		
	n	$x_{1,n}$	$x_{2,n}$		
$V=\frac{\det S}{\det \boldsymbol{\Sigma}}$					
$\tilde{V}=2\sqrt{V}$					
Date	jour, mois, an				
	heure				
Conclusion du contrôleur					

Fig. 5. La fiche θ_D

près égaux on calcule la valeur expérimentale $\widetilde{V}$. Si $\widetilde{V}<\widetilde{V}_{\frac{\alpha}{2}}$, la dispersion de la machine est acceptable et la fabrication peut continuer; par contre, si $\widetilde{V}>\widetilde{V}_{\frac{\alpha}{2}}$, la dispersion de la machine ne correspond plus aux conditions technologiques, elle doit donc être arrêtée et revsiée; ce qui a été fabriqué depuis le dernier contrôle doit être contrôlé à 100%.

De même que dans le cas du paramètre donnant la tendance centrale nous pouvons construire une fiche simple, que nous allons appeler „fiche θ_D" (la fiche θ-dispersion); sa limite inférieure de contrôle sera aussi 0 et la limite supérieure de contrôle $\widetilde{V}_{\frac{\alpha}{2}}$. Comme il en résulte de la fig. 5, pour effectuer le contrôle de la dispersion nous devons calculer les déterminants de $\boldsymbol{S}$ et de $\boldsymbol{\Sigma}$ et ensuite la valeur $\widetilde{V}$.

Il faut faire aussi quelques remarques sur les risques qui proviennent du contrôle de la dispersion. Le risque de 1re espèce sera ici α_2. Pour définir le risque de 2e espèce nous remarquons préalablement que si la réelle matrice de covariance est $\boldsymbol{\Sigma}_1$, alors

$$\widetilde{V}=2\sqrt{\frac{\det \boldsymbol{S}}{\det \boldsymbol{\Sigma}_1}\cdot\frac{\det \boldsymbol{\Sigma}_1}{\det \boldsymbol{\Sigma}}}=\delta\widetilde{V}_1,$$

où

$$\widetilde{V}_1=2\sqrt{\frac{\det \boldsymbol{S}}{\det \boldsymbol{\Sigma}_1}},\qquad \delta^2=\frac{\det \boldsymbol{\Sigma}_1}{\det \boldsymbol{\Sigma}}.$$

Soit maintenant δ_0, $\delta_0>1$, la limite admise pour la croissance de δ, c'est-à-dire $1<\delta\leq\delta_0$, telle que la dispersion de la machine reste encore acceptable. Le risque de 1e espèce sera donc

$$P(\delta_0\widetilde{V}_1<\widetilde{V}_{\frac{\alpha}{2}})=\beta_2. \tag{19}$$

En fixant les grandeurs n, α_2 et δ_0 nous pouvons déterminer le risque de 2e espèce β_2. Mais si la rigueur du contrôle exige un β_2 donné, la connaissance des grandeurs n et δ_0 nous permet d'obtenir $\widetilde{V}_{\frac{\alpha}{2}}$ de l'équation (19) et de calculer le risque de 1re espèce α_2 à partir de l'équation (18).

En réunissant ces deux procédés de contrôle exposés plus haut nous obtenons une méthode combinée pour le contrôle statistique en cours de fabrication à deux caractères simultanés pour la tendance centrale et la dispersion. La fiche de contrôle combinée correspondante sera la fiche θ_{TCD} (la fiche θ-tendance centrale-dispersion) de la fig. 6 qui a l'avantage d'être très simple.

5. Si on a $h>2$ caractères, les procédés utilisée plus haut s'étendent aisément. Dans ce cas le vecteur aléatoire $\boldsymbol{\xi}$ a une fonction de répartition normale $N(\boldsymbol{m}, \boldsymbol{\Sigma})$ à h dimensions.

La vérification, si les prélèvements sont effectués au hasard reste inchangée. Quant à l'application du test χ^2 et la méthode de contrôle de la tendance centrale, il faut tenir compte que les variables aléatoires $(\boldsymbol{\xi}-\boldsymbol{m})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}-\boldsymbol{m})$ et $(\bar{\boldsymbol{x}}-\boldsymbol{t})'\boldsymbol{\Sigma}_n^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{t})$ ont [1] des fonctions de répartition χ_h^2 (ou $\chi_{h,\lambda}^2$) à h degrés de liberté. Il s'ensuit que nous allons choisir les ellipsoïdes E_{θ_j}, $1\leq j\leq l$, tels que

Usine	**Fiche de contrôle statistique pour la tendance centrale et la dispersion**								Pièce		
Section									Effectif de l'échantillon $n=5$		
Machine									Prélèvements toutes les... minutes		
Appareils de mesure									Caractères	Tolérances	
Précision des appareils de mesure	α_1	$\theta(\alpha_1)$	β_1	λ_0^2	α_2	det $\mathbf{\Sigma}$	β_2	δ_0	ξ_1	$[T_{11}, T_{12}]$	
Contrôleur									ξ_2	$[T_{21}, T_{22}]$	

No de l'échantillon		1		2	
Valeurs de $\widetilde{V}$	θ_D				
	2				
	1				
	0				
Valeurs de θ_{exp}	θ_{TC}				
	2				
	1				
	0				
Valeurs de l'échantillon et les valeurs θ_i	1	$x_{1,1}$	$x_{2,1}$		
	2	$x_{1,2}$	$x_{2,2}$		
	3	$x_{1,3}$	$x_{2,3}$		
	4	$x_{1,4}$	$x_{2,4}$		
	5	$x_{1,5}$	$x_{2,5}$		
Somme		$\sum_{i=1}^{5} x_{1,i}$	$\sum_{i=1}^{5} x_{2,i}$		
Moyenne		$\bar{x}_1$	$\bar{x}_2$		
		$\theta_{exp}=$ | $V=$	$\widetilde{V}=$		
Date	jour, mois, an.				
	heure				
Conclusion du contrôleur					

Fig. 6. La fiche θ_{TCD}

$$p_1^* = P(\xi \in E_{\theta_1}) = P\left(\frac{1}{2}\chi_h^2 < \theta_1\right),$$

$$p_j^* = P(\xi \in E_{\theta_j} \setminus E_{\theta_{j-1}}) = P\left(\theta_{j-1} \leq \frac{1}{2}\chi_h^2 < \theta_j\right), \quad 1 < j \leq l,$$

et l'ellipsoïde $E_{n,\, \theta(a_1)}$ tel que

$$P\left(\frac{1}{2}\chi_h^2 < \theta(a_1)\right) = 1 - a_1.$$

La méthode de contrôle de la dispersion se modifie ayant en vue le fait que la variable aléatoire $V^{1/h}$ a comme densité de probabilité approximative l'expression [1]

$$\frac{c^{\frac{1}{2}h(n-h)}\, y^{\frac{1}{2}h(n-h)}\, e^{-cy}}{\Gamma\left(\frac{1}{2}h(n-h)\right)},$$

où

$$c = \frac{h}{2}\left(1 - \frac{(h-1)(h-2)}{2n}\right)^{1/h};$$

évidemment, dans ce cas pour déterminer le risque de 2e espèce nous prenons

$$V^{\frac{1}{h}} = V_1^{\frac{1}{h}} \delta^{\frac{2}{h}}, \quad V = \frac{\det S}{\det \Sigma}, \quad V_1 = \frac{\det S}{\det \Sigma_1} \tag{20}$$

Ces précisions nous permettent de construire les fiches de contrôle θ_{TC} et θ_D ou θ_{TCD}, que l'on n'utilise plus à l'aide de méthodes graphiques, mais seulement à l'aide de méthodes analytiques conduisant à la détermination de θ_{exp} (d'après (16)), et de $V^{1/h}$ (d'après (20)). Ces fiches de contrôle sont donc pratiques seulement si les calculs nécessaires à leur application peuvent être effectués rapidement, ce qui ne comporte aucune difficulté si l'on dispose d'une machine à calculer simple.

Centre de statistique mathématique
l'Académie de la République Popilaire Roumaine
Bucarest

BIBLIOGRAPHIE

1. Anderson T. W., An introduction to Multivariate Statistical Analysis, New York, 1958.
2. Cave R., Le contrôle statistique das fabrications, Paris, 1961.
3. Cowden D. J., Statistical methods in quality control, Englewood Cliffs, N. J., 1957.
4. Cramer H., Mathematical methods of Statistics, Princeton, 1964.
5. Дунин-Барковский И. В., Н. В. Смирнов, Теория вероятностей и математическая статистика (общая часть), Москва, 1955.
6. Hald A., Statistical theory with engineering applications, New-York, 1952.
7. Rancu N., L. Tövissi, Statistica matematică cu aplicaţii în producţie, Bucureşti, 1963.
8. Schindowski E., O. Schürz, Statististische Qualitätskontrolle, Berlin, 1959.

Постъпила на 23. IX. 1964 г.

СТАТИСТИЧЕСКИЙ КОНТРОЛЬ КАЧЕСТВА ПО НЕСКОЛЬКИМ ПРИЗНАКАМ

I. МЕТОД СРЕДНЕГО И МЕТОД ОБОБЩЕННОЙ ДИСПЕРСИИ

Р. Теодореску и Й. Вадуда

(Резюме)

В работе рассматривается статистический текущий контроль качества продукции в случае, когда наблюдается одновременно за несколькими (больше одного) признаками одного и того же изделия. Приводятся методы для проверки случайности двумерной выборки, с помощью т. н. медианной кривой. Дальше изучаются критерии для проверки правильности наладки машины, как в отношении отклонения среднего, так и по отношению двумерного рассеяния. Методы снабжены контрольными картами.