REVUE DE STATISTIQUE APPLIQUÉE

J. LAROCHE
J. MOTHES

Des problèmes que pose le contrôle industriel. Remarques relatives au contrôle de l'épaisseur des papiers destinés à l'enrubannage des câbles électriques

Revue de statistique appliquée, tome 4, nº 2 (1956), p. 75-84
<http://www.numdam.org/item?id=RSA_1956__4_2_75_0>

REMARQUES RELATIVES AU CONTRÔLE DE L'ÉPAISSEUR DES PAPIERS DESTINÉS A L'ENRUBANNAGE DES CABLES ÉLECTRIQUES

par

J. LAROCHE

Directeur de la Papeterie de Nanterre

et

J. MOTHES

Ancien Élève de l'École Polytechnique

Compte tenu de l'usage auquel on destine un produit on est généralement amené, en pratique industrielle, à imposer certaines conditions aux propriétés physiques ou chimiques du produit en question. C'est ainsi qu'on est amené dans le cas des fabrications mécaniques à fixer des conditions dimensionnelles aux pièces usinées en vue de leur utilisation ultérieure. C'est ainsi, également, qu'on est amené à imposer aux papiers destinés à l'enrubannage des câbles électriques des conditions dont l'une vise l'épaisseur, mieux encore la régularité d'épaisseur à l'intérieur de chacun des rouleaux livrés par le papetier, etc... etc...

Bien entendu, le choix des propriétés à contrôler pose des problèmes plus ou moins complexes.

Dans certains cas il découle sans ambiguité de la nature même des choses. Lorsqu'on usine des pistons destinés à entrer dans des cylindres le contrôle du diamètre des pistons s'impose nécessairement.

Dans d'autres cas, il résulte de considérations techniques beaucoup moins immédiates. Il implique, en effet, une étude de liaison entre les résultats à atteindre et les propriétés des produits à mettre en œuvre en vue d'une part de diagnostiquer les propriétés à contrôler et, d'autre part, de définir les spécifications à leur imposer. Il arrive même qu'à ce stade préliminaire se pose le problème de la nature des essais à effectuer dans le cadre du contrôle des propriétés sur lesquelles on entend porter son attention.

La statistique mathématique, on l'ignore trop souvent, est à même de jouer un rôle considérable au stade du choix tant des propriétés à contrôler que de la nature des essais à effectuer. Une fois résolus ces problèmes préalables, elle reste, par ailleurs, l'instrument fondamental d'interprétation des résultats de contrôle.

Ayant, en présence d'une certaine quantité d'un certain produit (constituant ce que nous appellerons une fourniture), à contrôler une propriété du produit en question, on sait parfaitement que les "valeurs" ou "qualités" de la propriété surveillée, au sein de la fourniture soumise à examen, ne sont nullement invariables mais se "distribuent". Pratiquement le contrôle de cette propriété consiste donc sous une forme ou sous une autre à saisir sa distribution au sein de la fourniture c'est-à-dire les caractéristiques qui la définissent complètement (ou tout au moins certaines d'entre elles) et cela, pour des raisons évidentes d'économie,

par l'intermédiaire d'échantillons. Les problèmes de contrôle industriel se ramènent ainsi à des problèmes strictement statistiques du type :

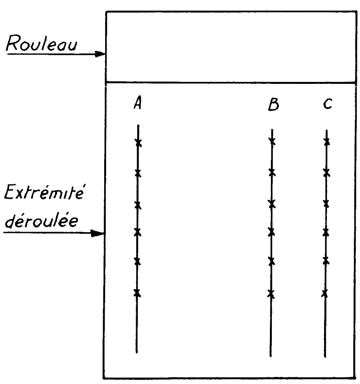
- Estimer sur échantillon la moyenne ou la dispersion d'une distribution.
- Vérifier sur échantillon que la moyenne d'une distribution n'est pas inférieure à une certaine valeur et que la dispersion n'est pas supérieure à une certaine valeur, etc... etc...

Cela étant, il ressort de l'examen de la plupart des normes existantes qu'il n'a pas été tenu compte, lors de leur élaboration, du caractère nécessairement statistique de l'interprétation des résultats auxquels conduit leur application. Il en résulte de regrettables divergences entre les enseignements **réels** apportés par les opérations de contrôle et ceux qu'on en tire effectivement.

Nous nous proposons au cours de cette étude d'illustrer cette remarque en procédant à l'analyse critique des règles de contrôle de l'épaisseur des papiers destinés à l'enrubannage des câbles électriques. Nous préciserons ensuite les modifications qu'il conviendrait d'apporter à ces règles pour rétablir leur cohérence.

I. MODALITÉS ACTUELLES DU CONTRÔLE DE L'ÉPAISSEUR DES PAPIERS DESTINÉS A L'ENRUBANNAGE DES CABLES ÉLECTRIQUES

Nous examinerons dans ce qui suit le cas des papiers V/I B - 50 grs/m².



Le contrôle de l'épaisseur, à la réception, porte sur **chacun** des rouleaux réceptionnés.

On déroule l'extrémité libre du rouleau sur une certaine longueur, puis on trace trois lignes A B C parallèles au sens du déroulement. Les lignes A et C sont tracées à 5 cms au plus des bords extrêmes de la feuille, la ligne intermédiaire B pouvant occuper une position quelconque.

Cela fait, on procède à 6 mesures d'épaisseur sur chacune des lignes A, B, C, en des points distants de 20 cms, puis on établit la moyenne des épaisseurs relatives à chacune des lignes.

Pour être jugé convenable le rouleau examiné doit fournir des moyennes de ligne comprises entre 69 et 79 μ, la différence entre la plus forte et la plus faible moyenne (qu'on désigne en statistique sous le terme d'**étendue**) n'excédant pas 4 % soit environ 3 μ.

Exemples :

a) rouleau accepté			b) rouleau refusé		
A	B	C	A	B	C
74	75	77	73	75	78
76	77	75	74	77	79
79	78	77	78	75	77
78	76	74	73	72	79
76	73	78	75	69	77
78	75	75	76	75	75
76.8	75.6	76.0	74.8	73.8	77.5

Etendue : 76.8-75.6 = 1,2   77.5-73.8 = 3.7

Les objectifs réels d'un tel contrôle apparaissent clairement à travers ses modalités.

Le réceptionnaire sait parfaitement qu'aucun rouleau de papier ne peut présenter une épaisseur invariable en chacun de ses points. Il distingue donc, dans le sens de la longueur, l'épaisseur moyenne susceptible d'être observée sur une ligne donnée et, dans le sens de la largeur, les fluctuations d'épaisseur moyenne susceptibles d'être enregistrées d'une ligne à l'autre. En d'autres termes, en désignant de façon générale par m_i l'épaisseur moyenne d'une ligne (i), et par σ_t l'écart type de la distribution des m_i dans le sens de la largeur, le réceptionnaire entend contrôler m_i d'une part, σ_t d'autre part.

Pratiquement, il essaie de contrôler les moyennes m_i par l'intermédiaire de 3 d'entre elles, saisies sur 3 lignes différentes A, B, C, puis de contrôler le σ_t à partir de l'étendue de ces 3 moyennes.

Cette optique est parfaitement logique, mais ses conditions d'application soulèvent des difficultés que les normalisateurs semblent n'avoir pas perçues. Le réceptionnaire ne saisit pas les moyennes m_i, mais seulement 3 estimations m'_A m'_B m'_C des m_i sur 3 lignes A, B, C, estimations résultant chacune d'une série de 6 observations. **Il ne peut donc raisonner sur les quantités m'_A m'_B m'_C comme il le ferait sur les moyennes m_i elles-mêmes.**

En vue de concrétiser le type des difficultés rencontrées il nous suffira de nous référer à des résultats pratiques de contrôle.

Nous avons prélevé au hasard dans un ensemble de fiches de contrôle, 10 fiches **concernant des rouleaux refusés**. En principe ces rouleaux avaient été refusés parce que, d'après le contrôle opéré, les moyennes des lignes m_i différaient, à l'intérieur de chaque rouleau, de plus de 3 µ, mais nous nous sommes demandés si les résultats des **échantillonnages effectués** permettaient de conclure **ne serait-ce qu'à l'inégalité des m_i**.

Pour résoudre ce problème (classique en statistique) nous avons appliqué aux données portées sur les 10 fiches de contrôle la technique dite d'**analyse de la variance** (1). Les résultats se sont alors établis comme indiqué ci-dessous.

Données ramenées à la moyenne 75				Tableau d'analyse de la variance Limite significative Ve/Vr = 6,4 au seuil de probabilité 1 %
Rouleau n° 1 = Etendue 3,2				
	- 2	0	3	Ve = 17,4
	- 1	2	4	$\frac{Ve}{Vr} = 5.0$
	3	0	2	Vr = 3,5
	- 2	- 2	2	Différences des moyennes de
	0	- 4	4	ligne NON SIGNIFICATIVES
	1	0	0	
total	- 1	- 4	+15	

(1) L'application de l'analyse de la variance implique (cf. annexe I) :
- la normalité de la distribution Di des épaisseurs dans chaque ligne (i);
- l'égalité, d'une ligne à l'autre, de la dispersion des distributions Di.

Dans le cas présent, comme d'ailleurs dans la majorité des cas rencontrés en pratique, ces deux hypothèses sont justifiées.

Données ramenées à la moyenne 75				Tableau d'analyse de la variance Limite significative Ve/Vr = 6,4 au seuil de probabilité 1 %
		Rouleau n° 5 = Etendue 3.0		
	- 3	- 3	- 4	Ve = 14.0
	- 2	- 6	- 2	$\frac{Ve}{Vr} = 2.8$
	1	- 2	- 7	Vr = 5.0
	0	- 3	- 2	Différences des moyennes de
	0	- 2	0	ligne NON SIGNIFICATIVES
	2	- 4	1	
total	- 2	-20	-14	
		Rouleau n° 7 = Etendue 4.8		
	- 1	- 2	2	Ve = 36.7
	- 2	0	0	$\frac{Ve}{Vr} = 9.0$
	- 3	- 4	4	Vr = 4.1
	- 3	- 4	- 1	Différences des moyennes de
	0	- 6	- 1	ligne SIGNIFICATIVES
	- 4	- 6	3	
total	-13	-22	7	
		Rouleau n° 10 = Etendue 3.7		
	2	0	- 1	Ve = 20.8
	- 2	- 3	0	$\frac{Ve}{Vr} = 4.7$
	- 5	- 2	3	Vr = 4.4
	0	- 6	- 1	Différences des moyennes de
	- 3	- 4	2	lignes NON SIGNIFICATIVES
	- 1	- 2	2	
total	- 9	-17	5	
		Rouleau n° 12 = Etendue 3.2		
	- 1	- 2	- 2	Ve = 16.7
	2	- 3	- 1	$\frac{Ve}{Vr} = 3.5$
	3	- 6	0	Vr = 4.7
	2	+ 1	- 4	Différences des moyennes de
	1	- 3	0	lignes NON SIGNIFICATIVES
	- 1	0	- 2	
total	6	-13	- 9	
		Rouleau n° 20 = Etendue 4.3		
	2	- 2	2	Ve = 28.4
	0	- 1	4	$\frac{Ve}{Vr} = 6.5$
	1	- 3	4	Vr = 4.4
	- 1	- 6	- 3	Différences des moyennes de
	0	0	3	lignes SIGNIFICATIVES
	- 2	- 3	1	
total	0	-15	11	

Données ramenées à la moyenne 75				Tableau d'analyse de la variance Limite significative Ve/Vr = 6,4 au seuil de probabilité 1 %
		Rouleau n° 22 = Etendue 6.1		
	3	- 2	0	Ve = 58.2
	0	- 4	- 3	$\frac{Ve}{Vr} = 16.6$
	3	- 6	- 1	Vr = 3.5
	3	- 5	4	Différences des moyennes de
	0	- 3	- 1	lignes SIGNIFICATIVES
	4	- 4	0	
total	13	-24	- 1	
		Rouleau n° 30 = Etendue 3.7		
	2	- 6	- 7	Ve = 20.7
	3	0	- 2	$\frac{Ve}{Vr} = 2.7$
	2	- 2	2	Vr = 7.7
	0	- 2	- 4	Différences des moyennes de
	- 4	- 4	- 2	lignes NON SIGNIFICATIVES
	- 1	- 6	- 1	
total	2	-20	- 14	
		Rouleau n° 31 = Etendue 4.9		
	- 4	- 3	2	Ve = 39.5
	- 1	- 2	3	$\frac{Ve}{Vr} = 18.8$
	0	- 4	4	Vr = 2.1
	- 2	- 5	1	Différences des moyennes de
	- 2	- 2	2	lignes SIGNIFICATIVES
	1	0	1	
total	- 8	- 16	13	
		Rouleau n° 52 = Etendue 3.9		
	- 2	- 6	2	Ve = 22.4
	- 3	- 6	- 3	$\frac{Ve}{Vr} = 5.4$
	- 2	- 2	1	Vr = 2.1
	- 1	- 3	3	Différences des moyennes de
	- 4	- 2	- 4	lignes NON SIGNIFICATIVES
	- 2	- 4	1	
total	-14	-23	0	

Ces résultats montrent donc que dans 6 cas, sur les 10 examinés, les règles actuelles de contrôle ont conduit à refuser des rouleaux pour différences entre m_i supérieures à 3 μ **alors que rien ne permettait, en fait, de dire qu'il y avait des différences significatives entre les m_i.**

II. PROPOSITIONS DE RÉVISION DES RÈGLES ACTUELLES DE CONTRÔLE

En vue de rétablir la cohérence des règles actuelles de contrôle de l'épaisseur il suffit de tenir compte du fait que les moyennes de lignes observées m'_A m'_B m'_C ne sont que le résultat d'un échantillonnage de 6 mesures d'épaisseur sur chacune des lignes A, B, C, ces lignes ne constituant elles-mêmes qu'un échantillon de l'ensemble des lignes (i) susceptibles d'être distinguées dans la largeur d'un rouleau.

Déterminer, de façon générale, la loi de distribution des moyennes m' observées est aisé. Si nous désignons par σ_t l'écart type de la distribution des m_i et par σ l'écart-type de la distribution des épaisseurs dans une ligne (écart-type que nous admettons, comme indiqué plus haut, constant d'une ligne à l'autre), l'écart-type de la distribution des m' s'écrit (1) :

$$\sqrt{\sigma_t^2 + \frac{\sigma^2}{6}}$$

Dans ces conditions les 3 moyennes m'_A m'_B m'_C peuvent être considérées comme le résultat d'un échantillonnage effectué dans une population d'écart-type $\sqrt{\sigma_t^2 + \frac{\sigma^2}{6}}$ et la théorie nous indique (cf. annexe II) qu'il y a 998 chances sur 1000 pour que leur étendue se situe au-dessous de la limite

$$4.8 \sqrt{\sigma_t^2 + \frac{\sigma^2}{6}}$$

Une normalisation correcte implique donc d'abord qu'on se fixe σ_t et σ puis ensuite qu'on adopte comme limite acceptable de la valeur de l'étendue de l'échantillon (m'_A m'_B m'_C) la limite

$$4.8 \sqrt{\sigma_t^2 + \frac{\sigma^2}{6}}$$

Il est à noter que la normalisation actuelle fixe implicitement σ_t et σ.

Imposer aux moyennes de lignes de ne ~~pas différer~~ différer entre elles de plus de 3 µ revient, en effet, à fixer à σ_t la limite $3/6.18$ (au seuil de probabilité 2 ‰). D'autre part, imposer à chaque moyenne ~~de ligne~~ ligne de tomber dans l'intervalle 69 µ - 79 µ revient à fixer à σ la limite $10\Big/\frac{6.18}{\sqrt{6}}$ (toujours au seuil de probabilité 2 ‰.

Ainsi dans le cadre de la norme existante, la limite à fixer à l'étendue des 3 moyennes m'_A m'_B m'_C n'est pas 3 µ , mais

$$4.8 \sqrt{\frac{3^2}{6.18^2} + \frac{10^2}{6.18^2}} = 8.1\,\mu$$

En adoptant une telle limite on observera qu'aucun des 10 rouleaux précédemment considérés n'aurait du être refusé. Et il n'y a là rien qui puisse étonner. Nous avons déjà constaté, par application de l'analyse de la variance, que parmi les 10 rouleaux il y en avait 6 ne présentant pas de différences significatives entre m_i . En ce qui concerne les 4 autres il y a bien différences significatives entre m_i mais ces différences n'atteignent pas en réalité les 3 µ retenus comme limite d'acceptation.

En bref, les remarques qui précèdent permettent de mettre en évidence sur un exemple particulier les problèmes que soulève la rédaction d'un cahier des charges.

A l'heure actuelle, on peut affirmer que les conditions de réception visant l'épaisseur des papiers utilisés à l'enrubannage des câbles électriques ne sont

(1) Bien entendu, il parait inutile d'insister sur le fait que nous nous référons, dans tout cet exposé, à des distributions du type normal.

Nous ne tenons d'autre part pas compte des erreurs de mesures, celles-ci s'étant avérées négligeables.

pas cohérentes. Il en résulte une perte économique certaine par refus de fourniture convenable.

Pour rétablir la cohérence de la norme, dans le cadre du processus expérimental en usage et des désiderata implicitement exprimés dans la norme actuelle, par les normalisateurs, il convient de fixer comme limite de l'étendue de m'_A m'_B m'_C la limite de $8.1\,\mu$ et non la limite de $3\,\mu$.

ANNEXE I

ANALYSE DE LA VARIANCE

Un des problèmes les plus fréquemment rencontrés en pratique industrielle s'énonce de la façon suivante : on se trouve en présence d'une population dans laquelle peuvent être distinguées des sous-populations susceptibles d'avoir été influencées par une ou plusieurs causes d'hétérogénéité agissant non pas sur la dispersion des résultats observables dans chaque sous-population mais sur leur moyenne et l'on se demande si l'intervention des causes en question a ou non effectivement provoqué, d'une sous-population à l'autre, une différence significative des moyennes. En d'autres termes, considérant (par exemple) une cause d'hétérogénéité A susceptible de prendre les formes A_1 A_2 .... A_p et désignant par x_{ij} un résultat observé dans la sous-population A_i, on écrit ce résultat :

$$x_{ij} = a_i + \xi_{ij}$$

en distinguant a_i moyenne correspondant à l'intervention de A_i et ξ_{ij} fluctuation aléatoire de moyenne nulle et de variance constante σ^2 (1), puis on se demande si la moyenne a_i varie effectivement avec les modalités d'intervention de la cause A.

Exemple simple :

Des ampoules électriques d'un certain modèle peuvent être fabriquées avec des filaments de trois types différents F_1 F_2 F_3. La nature du filament a-t-elle une influence sur la durée moyenne de vie des ampoules ?

Désignant de façon générale par :

$$x_{1j} = f_1 + \xi_{1j}$$
$$x_{2k} = f_2 + \xi_{2k}$$
$$x_{3l} = f_3 + \xi_{3l}$$

la durée d'une ampoule fabriquée avec un filament de type F_1 ou F_2 ou F_3 cela revient à se demander s'il y a effectivement différence significative de f_1 f_2 et f_3.

Ce problème, lorsque l'hypothèse de la normalité de la distribution des fluctuations aléatoires est fondée (2), est aisément résolu par l'analyse de la variance.

Considérant la population dans son ensemble et les résultats x_{ij} observés dans cette population, le numérateur de la variance totale de ces résultats s'écrit:

$$\Sigma_i\Sigma_j\,(x_{ij} - x_{..})^2 = \Sigma_i\Sigma_j\,(x_{ij} - x_{i.} + x_{i.} - x_{..})^2$$
$$= \Sigma_i\Sigma_j\,(x_{ij} - x_{i.})^2 + \Sigma_i\, n_i\,(x_{i.} - x_{..})^2$$

en désignant par $x_{..}$ la moyenne générale des observations, par $x_{i.}$ la moyenne

(1) L'hypothèse de la constance de la variance des fluctuations aléatoires est moins restrictive qu'elle n'en a l'air. On constate en pratique que les causes d'hétérogénéité des fabrications jouent beaucoup plus sur les moyennes que sur les dispersions.

(2) Elle l'est dans la majorité des cas.

des observations appartenant à la sous-population A_i , par n_i le nombre d'observations effectuées dans la sous-population A_i .

Si l'on admet d'autre part que les modalités d'intervention de la cause d'hétérogénéïté (A) sont sans influence sur les a_i tout se passe comme si on avait tiré l'échantillon observé d'une population normale de variance σ^2 et de moyenne :

$$a_1 = a_2 = .. = a_i = .. = a_p = a.$$

Dans ces conditions on démontre alors :

1° que la variable

$$\sum_i \sum_j \frac{(x_{ij} - x_{i.})^2}{\sigma^2}$$

suit une loi dite de χ^2 à n - 1 degrés de Liberté (n étant le nombre total d'observations)

2° que la variable

$$\sum_i \frac{n_i (x_{i.} - x_{..})^2}{\sigma^2}$$

suit une loi de χ^2 à p - 1 degrés de Liberté (en désignant par p le nombre des formes d'intervention possibles de A);

3° que la variable

$$\frac{\sum_i \sum_j (x_{ij} - x_{i.})^2}{\sigma^2}$$

suit une loi de χ^2 à n - p degrés de Liberté;

autrement dit que les 2 variables

$$\frac{\sum_i n_i (x_{i.} - x_{..})^2}{\sigma^2} \quad \text{et} \quad \frac{\sum_i \sum_j (x_{ij} - x_{i.})^2}{\sigma^2}$$

sont distribuées suivant des lois de χ^2 à p - 1 et n - p degrés de Liberté et **que leur somme est également distribuée suivant une loi de χ^2 avec un nombre de degrés de liberté égal à la somme des degrés de liberté (p-1) + (n-1) = n - 1.**

C'est là, d'après un théorème dû à COCHRAN, la preuve de **l'indépendance** des deux variables en question.

A ce stade la théorie de l'estimation statistique indique que

$$\frac{\sum_i n_i (x_{i.} - x_{..})^2}{p - 1}$$

et

$$\frac{\sum_i \sum_j (x_{ij} - x_{i.})^2}{n - p}$$

sont deux estimations de σ^2 et, bien entendu, compte tenu de la remarque précédente, sont deux estimations **indépendantes** basées l'une sur p - 1, l'autre sur n - p degrés de Liberté.

Ainsi, **dans l'hypothèse de non intervention de A, on doit constater l'identité statistique de ces deux estimations.**

Le test consiste à former le rapport :

$$F = \frac{\sum_i n_i \, (x_{i.} - x_{..})^2 \,/\, (p - 1)}{\sum_i \sum_j (x_{ij} - x_{i.})^2 \,/\, n - p}$$

puis à vérifier que ce rapport ne dépasse pas une valeur limite (fonction du nombre des degrés de liberté p - 1 et n - p ainsi que du seuil de probabilité considéré comme négligeable) indiquée dans une table dite de SNEDECOR.

Le tableau classique d'analyse de la variance se présente, dans le cas de l'étude d'une seule cause d'hétérogénéité (1), sous la forme suivante :

Sommes quadratiques	Degrés de Liberté	Estimations de variance
$\sum_i n_i (x_{i.} - x_{..})^2$	$p - 1$	$V_A = \frac{\sum_i n_i (x_{i.} - x_{..})^2}{p - 1}$
$\sum_i \sum_j (x_{ij} - x_{i.})^2$	$n - p$	$V_R = \frac{\sum_i \sum_j (x_{ij} - x_{i.})^2}{n - p}$
$\sum_i \sum_j (x_{ij} - x_{..})^2$	$n - 1$	

Les calculs proprement dits ne présentent aucune difficulté. On calcule tout d'abord :

$$\sum_i \sum_j (x_{ij} - x_{..})^2 = \sum_i \sum_j x_{ij}^2 - n\, x_{..}^2$$

puis

$$\sum_i n_i (x_{i.} - x_{..})^2 = \sum_i n_i \, x_{i.}^2 - n\, x_{..}^2$$

et de là, par différence, $\sum_i \sum_j (x_{ij} - x_{i.})^2$.

<u>Remarque importante</u>

La variance désignée par V_R (variance résiduelle) reste toujours une estimation de σ^2 même dans l'hypothèse d'intervention de la cause d'hétérogénéité A.

On a en effet :

$$x_{ij} = a_i + \xi_{ij}$$
$$x_{i.} = a_i + \xi_{i.}$$

et par conséquent

$$\frac{\sum_i \sum_j (x_{ij} - x_{i.})^2}{n - p} = \frac{\sum_i \sum_j (\xi_{ij} - \xi_{i.})^2}{n - p}$$

Or, $\frac{\sum_i \sum_j (\xi_{ij} - \xi_{i.})^2}{n - p}$ fournit bien, toujours, une estimation de σ^2 puisque, par définition, les ξ_{ij} ont la même moyenne (nulle) quelles que soient les modalités d'intervention de A.

(1) L'analyse de la variance est généralisable à n causes d'hétérogénéité.

ANNEXE II

DISTRIBUTION DES ETENDUES D'ECHANTILLONS EXTRAITS D'UNE POPULATION NORMALE

Etant donné un échantillon

$$x_1 \; x_2 \; \ldots \; x_n$$

on désigne sous le terme d'**étendue** la différence w entre le résultat observé le plus fort et le résultat observé le plus faible et la détermination de la distribution des étendues d'échantillons de n observations extraits d'une population, elle-même distribuée suivant une loi connue, est théoriquement aisée.

Dans l'hypothèse d'une population distribuée normalement (moyenne m, écart-type σ) on peut, en particulier, déterminer sans difficulté la loi $f_n(\frac{w}{\sigma})$ mais l'expression de cette loi n'est pas simple.

En vue des utilisations pratiques le statisticien anglais E.S. PEARSON a donc calculé un certain nombre de valeurs spécifiques de la fonction de distribution de $f_n\,(\frac{w}{\sigma})$ c'est-à-dire de

$$\int_0^{w/\sigma} f_n \left(\frac{w}{\sigma}\right) d \; \frac{w}{\sigma}$$

pour des valeurs de n variant de 2 à 12.

En se reportant à la table qu'il a établie, on constate, par exemple, qu'il y a une probabilité égale à 998/1.000 d'observer $\frac{w}{\sigma} \leqslant 4.8$ lorsque les échantillons sont de taille n = 3, ce qui explique les raisons de l'adoption de ce coefficient au cours de la présente étude.