

REVUE DE STATISTIQUE APPLIQUÉE

C. STELLMANN

P. BORNAREL

Une application de l'analyse de variance à un problème de biologie industrielle

Revue de statistique appliquée, tome 18, n° 1 (1970), p. 99-109

http://www.numdam.org/item?id=RSA_1970__18_1_99_0

© Société française de statistique, 1970, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

UNE APPLICATION DE L'ANALYSE DE VARIANCE A UN PROBLÈME DE BIOLOGIE INDUSTRIELLE

C. STELLMANN et P. BORNAREL (*)

INTRODUCTION

La tendance de la production biologique actuelle est d'acquérir le caractère industriel :

- industriel par les quantités de matières mises en jeu, visant la commercialisation des produits dans des secteurs autrefois non sollicités.

- industriel par la technologie des contrôles et des régulations diverses qui font appel à des techniques physico-chimiques longtemps réservées à l'industrie chimique, tels que la régulation de la température et du H, et les contrôles électroniques qui s'y rattachent.

- industriel par la budgétisation inévitable des moyens pour obtenir une rentabilité optimale d'une production, car l'amélioration d'un rendement jouant sur le prix de revient permet une commercialisation efficiente dans notre société où la concurrence est un facteur d'émulation entre techniciens.

L'analyse statistique pénètre progressivement dans le domaine biologique, au même titre que les techniques perfectionnées de l'électronique et de l'automation. Cette analyse devient indispensable dans la production biologique industrielle par le fait :

- du nombre important de facteurs contrôlés,
- du nombre croissant de données quantitatives qui tendent à supplanter les données qualitatives,
- de la nécessité de "quantitativiser" les conclusions des analyses.

L'analyse de variance, de type factoriel, est un des outils statistiques qui répondent au mieux aux questions des expérimentateurs et des producteurs.

Nous limitons la bibliographie, très fournie, concernant ce problème, à quatre ouvrages, SCHWARTZ [1], PHILIPPE [2], MATHER [3], LISON [8] où nous avons puisé les principes de notre exposé.

(*) Collaboration technique de JP MARTEL (pour la programmation sur ordinateur).

Travail de l'Institut Français de la Fièvre Aphteuse - LYON. Directeur Dr. C. MACKOWIAK.

I - EXPOSE DU PROBLEME

Cet exemple que nous traitons est tiré d'une mise au point de la culture industrielle du virus aphteux selon la technique FRENKEL [4]. De nombreux facteurs influent sur le rendement d'une telle technique et, dans l'expérience présentée, quatre facteurs seront étudiés :

- l'inoculum des cultures qui est la souche virale utilisée pour ensemençer les réacteurs de culture. Cet inoculum, dans notre exemple, est obtenu selon deux procédés différents.

- le milieu de culture dont la composition biochimique est connue. Deux milieux de culture sont testés dans notre essai.

- les échantillons de virus produit, utilisés pour les contrôles, sont traités selon 3 procédés.

- la cinétique de la production en fonction du nombre de passages d'une culture à une autre culture, dans notre exemple 5 passages sont testés.

Le critère de lecture est le titre du virus, estimé par réaction de fixation du complément [5] et exprimé en logarithme décimal pour "normaliser" les résultats.

Les facteurs durée, température de culture et aération sont constants pour l'ensemble de l'expérience.

Le problème pour le producteur est de choisir la meilleure combinaison des 4 paramètres qui entraînera une production optimale dans les conditions industrielles. Un tel problème ne peut être étudié qu'à l'aide d'un plan expérimental rigoureux, qui permet l'étude des interactions éventuelles entre les quatre paramètres.

II - ETABLISSEMENT DE SOUS TABLEAUX POUR LES CALCULS NUMERIQUES.

Lorsque les calculs sont effectués à la machine à calculer de type courant (type ADDOX 4000), le tableau n° 1 doit être décomposé en autant de sous-tableaux que de composants de la variance pour éviter toute erreur et faciliter l'interprétation ultérieure.

Ce mode opératoire est valable dans tous les cas, mais lorsque le nombre de facteurs est supérieur à 4, ou lorsque le nombre de données est de l'ordre de 150, la programmation de tels plans factoriels sur machine à calculer perfectionnée est nécessaire (type OLIVETTI PROGRAMMA 101).

Lorsque le nombre de facteurs est supérieur à 5, ou le nombre de données supérieur à 150, la programmation sur ordinateur est indispensable (type IBM 1130) ou autre (360/40) et dans ce cas la présentation des données peut être différente.

Les sous-tableaux sont établis en calculant la somme des données dans un ordre précis pour chaque composant de la variance et le calcul de chaque variance est lié à une formule générale :

TABLEAU n° 1

TABLEAU DES DONNEES

		PASSAGES												
SOUCHES	MILIEUX	ECHANTILLONS	2		3		6		7		9		ε _a = 10.874	
			a*	b										
S ₁	M ₁	E ₁	1.918		2.279		1.806		2.176		2.495			
		E ₂		1.756	2.025		1.568		2.050		2.312		ε _b = 9.711	
		E ₃				2.398		2.090		2.290		2.845	ε _c = 11.703	
	M ₂	E ₁	a'	1.93	2.193		1.969		2.158		2.536		ε _{a'} = 10.786	
		E ₂	b'	1.748	2.09		1.819		2.097		2.243		ε _{b'} = 9.997	
		E ₃	c'		2.131	2.447		2.300		2.431		2.869	ε _{c'} = 12.034	
	S ₂	M ₁	E ₁	a''	2.190	2.14		2.057		2.210		2.434		ε _{a''} = 11.031
			E ₂	b''	2.012	1.964		1.903		2.064		2.267		ε _{b''} = 10.211
			E ₃	c''		2.424		2.239		2.407		2.584		ε _{c''} = 12.034
M ₂		E ₁	a'''	2.021	2.241		2.164		2.288		2.516		ε _{a'''} = 11.232	
		E ₂	b'''	1.832	2.079		2.017		2.343		2.390		ε _{b'''} = 10.661	
		E ₃	c'''		2.243	2.455		2.362		2.681		2.615	ε _{c'''} = 12.356	
				ε _A	8.059	8.855		7.996		8.832		9.981		
				ε _B	7.348	8.158		7.308		8.554		9.212		ε _X = 132.574
				ε _C	8.834	9.724		8.991		9.809		10.913		ε _{X²} = 297.0325

(*) Titre fixant le complément des virus obtenus.

$$s^2 = \frac{1}{\text{ddl}} \left[\frac{1}{n} \sum X^2 - \text{terme correctif} \right]$$

Le terme entre parenthèse est la somme des carrés de chaque composant.

III - TEST DE F ET SIGNIFICATIVITE

Pour tester la significativité des calculs précédents, nous effectuons, pour chaque source de variation le rapport :

$$F_{N_2}^{N_1} = \frac{\text{Variance de la source de variation}}{\text{Variance résiduelle}}$$

En fonction de :

N1 = ddl du numérateur

N2 = ddl du dénominateur

des tables (SCHWARTZ, PHILIPPE, MATHER) donnent la probabilité limite, accordée par le hasard, pour obtenir les valeurs de F.

L'interprétation des probabilités obtenues est, en général, la suivante :

- lorsque la probabilité est inférieure à 0.01, l'effet de la source de variation est considéré comme une certitude,
- lorsque la probabilité est comprise entre 0.01 et 0.05, l'effet est vraisemblable et significatif.
- lorsque la probabilité est comprise entre 0.05 et 0.10 l'effet nécessite une confirmation expérimentale en multipliant les essais.
- lorsque la probabilité est supérieure à 0.10, l'effet est considéré comme peu vraisemblable et non significatif.

Ces règles sont édictées par les besoins de l'expérimentateur.

IV - RESULTATS ET INTERPRETATION DE L'ANALYSE

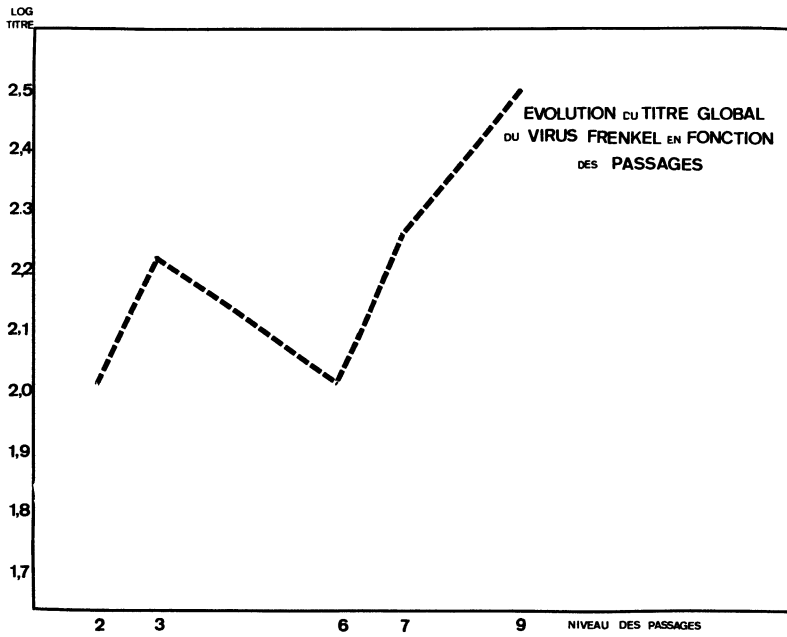
Les résultats des calculs sont transcrits dans le tableau n° 2. L'interprétation d'un tel tableau doit répondre à un double problème :

- le problème purement statistique ou les conclusions seront la synthèse d'un certain nombre d'affirmations, reflets de l'analyse,
- le problème purement technique où les conclusions doivent être pratiques et réalisables.

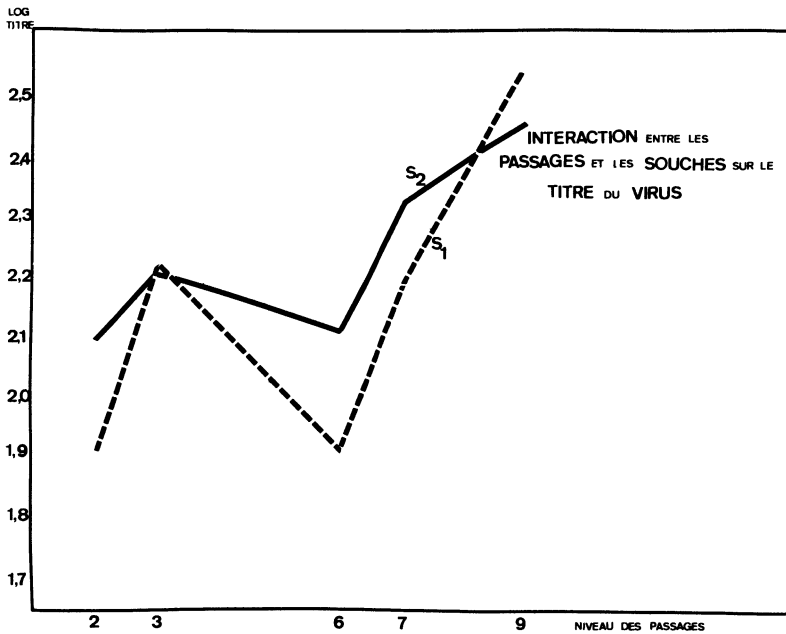
Nous nous attacherons particulièrement au premier volet statistique.

Le tableau n° 2 indique que tous les effets principaux P.E.S. et M sont hautement significatifs, mais du fait de la significativité des interactions de premier ordre, PS et PM, et de deuxième ordre, PES et PSM, l'interprétation des effets principaux est fonction de l'interaction de l'ordre le plus élevé.

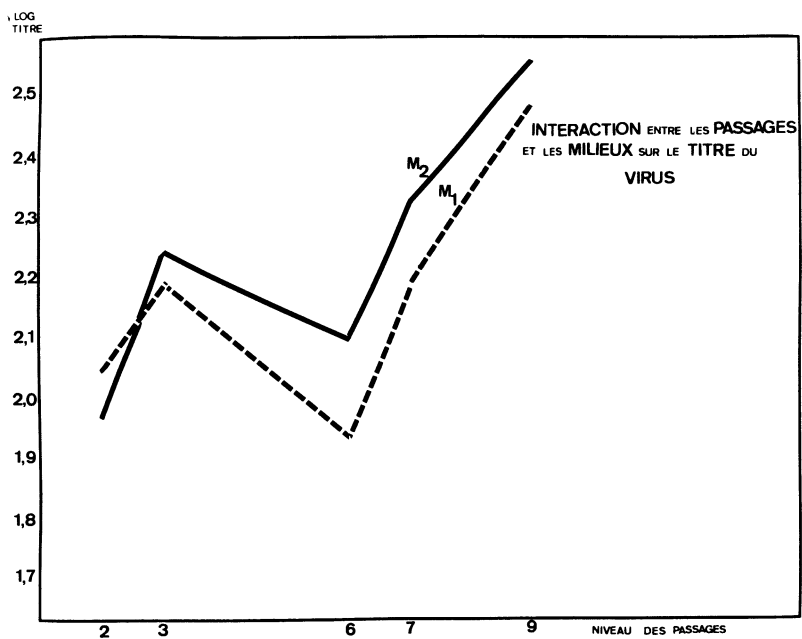
Les graphiques n° 1, 2, 3, 4 et 5 concrétisent les fluctuations dues aux divers facteurs, et nous remarquons qu'une interaction significative se traduit par un non parallélisme des courbes représentatives des réponses.



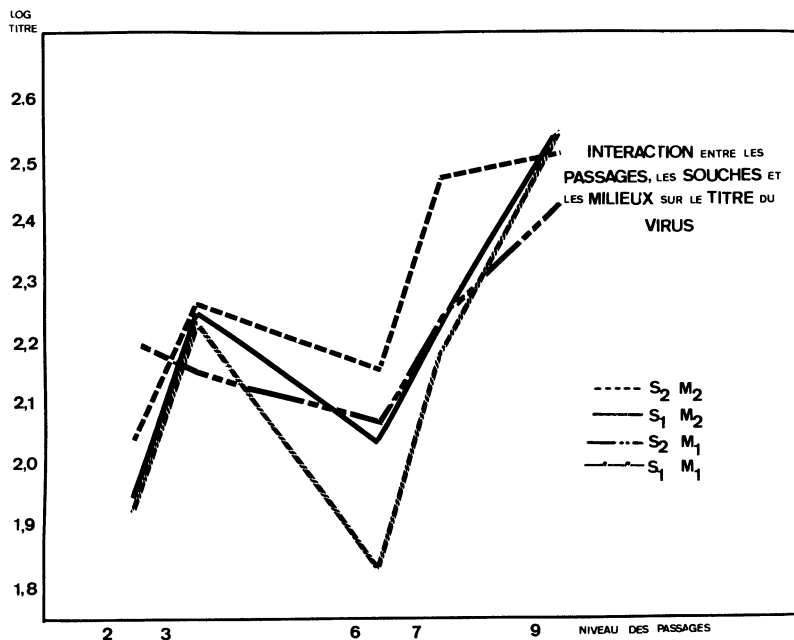
Graphique n° 1



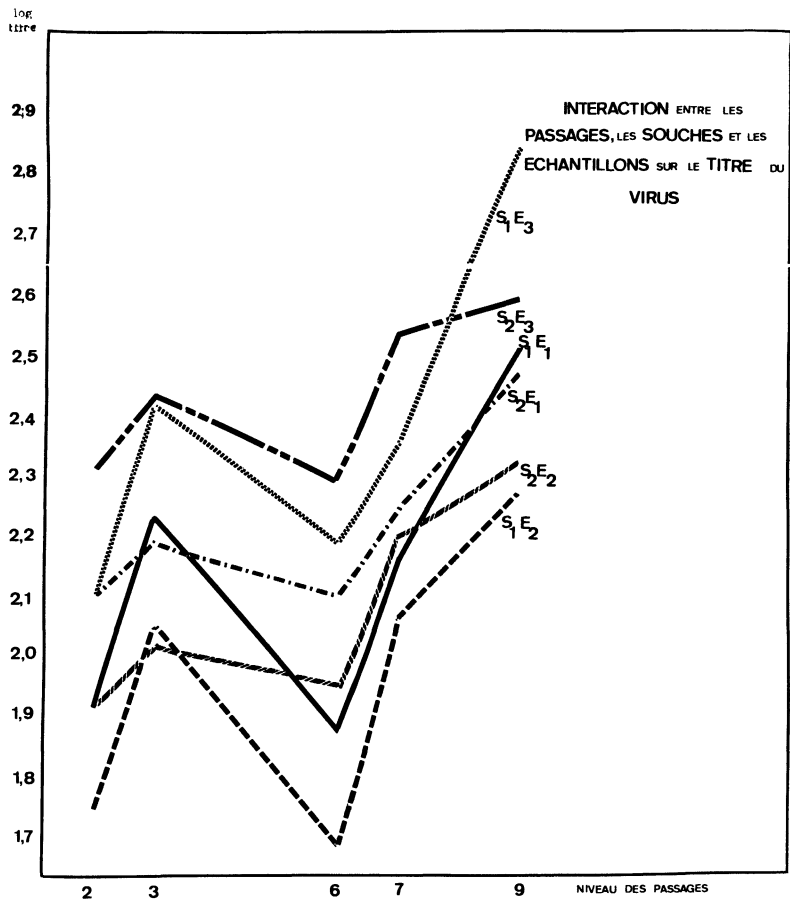
Graphique n° 2



Graphique n° 3



Graphique n° 4



Graphique n° 5

TABLEAU n° 2

RESULTATS DE L'ANALYSE : VARIANCE ET TEST DE SIGNIFICATIVITE

SOURCE DE VARIATION	SOMME DES CARRES	DEGRE DE LIBERTE	VARIANCE	TEST DE F	PROBABILITE
Effets principaux					
P ²	1.95916	4*	0.48976	F ₈ ⁴ = 360.14	Très inférieur à 0.001
E ²	1.49568	2	0.7478	F ₈ ² = 549.88	Très inférieur à 0.001
S ²	0.10192	1	0.10192	F ₈ ¹ = 74.94	Inférieur à 0.001
M ²	0.05673	1	0.0567	F ₈ ¹ = 41.71	Inférieur à 0.001
Interaction 1er ordre					
PE ²	0.02672	8**	0.00334	F ₈ ⁸ = 2.456	Supérieur à 0.10
PS ²	0.19352	4	0.0484	F ₈ ⁴ = 35.57	Inférieur à 0.001
PM ²	0.09108	4	0.0253	F ₈ ⁴ = 18.58	Inférieur à 0.001
ES ²	0.01072	2	0.00536	F ₈ ² = 3.94	Entre 0.05 et 0.10
EM ²	0.00702	2	0.0035	F ₈ ² = 2.58	Supérieur à 0.10
SM ²	0.00016	1	0.00016	F ₈ ¹ = 0.11	Supérieur à 0.50
Interaction 2ème ordre					
PES ²	0.06044	8	0.00768	F ₈ ⁸ = 5.64	Entre 0.01 et 0.025
PEM ²	0.01680	8	0.00210	F ₈ ⁸ = 1.544	Supérieur à 0.10
PSM ²	0.05724	4	0.014309	F ₈ ⁴ = 10.522	Entre 0.01 et 0.001
ESM ²	0.00274	2	0.00137	F ₈ ² = 1.0	Très supérieur à 0.10
Interaction 3ème ordre					
Résiduelle					
PESM ²	0.01088	8	0.00136		
Total	4.08807	59			

* ddl = N-1

N = niveau du facteur

** Produit des ddl des facteurs, contenu dans l'interaction

ddl (P) = 4 ddl (E) = 2 ddl (PE) = 8

Conclusions statistiques

a) La première conclusion pratique est que le titre du virus est maximum entre le 7ème et le 9ème passage (graphique n° 1).

b) Le choix de la souche inoculum est lié, à la fois au milieu, à l'échantillon et au passage, nous pouvons tout d'abord étudier le rôle du milieu :

- d'après le graphique n° 2 : S_2 est supérieur à S_1 , d'une valeur moyenne de 0.082 (différence logarithmique), soit un facteur de multiplication 1.203.

- d'après le graphique n° 3 : M_2 est supérieur à M_1 , d'une valeur moyenne de 0.062, soit un facteur de multiplication 1.154.

- d'après le graphique n° 4 :

S_2M_2 est supérieur à S_2M_1 , différence moyenne : 0.06 (X 1.149)

S_2M_1 est supérieur à S_1M_2 , différence moyenne : 0.04 (X 1.097)

S_1M_2 est supérieur à S_1M_1 , différence moyenne : 0.035 (X 1.084)

donc $S_2M_2 > S_2M_1 > S_1M_2 > S_1M_1$

C'est dire que la souche S_2 est d'autant plus active qu'elle est cultivée avec le milieu M_2 , non par simple addition des deux composants ($S_2 + M_2$) mais par multiplication ($S_2 \times M_2$), cette interaction fort intéressante pour le producteur ne pourrait être mise en évidence, à bon compte, que pour ce plan expérimental.

Les différences moyennes ne sont valables que sur l'ensemble des passages, mais le graphique n° 4 permet de calculer les différences pour chaque passage.

c) L'interaction PES, nous guide davantage quant au choix de la souche ; en effet, des études antérieures [6] ont montré que la valeur immunogène du virus FRENKEL est optimale lorsque :

$$\left\{ \begin{array}{l} E_2 \text{ est maximum} \\ \text{La différence } E_2 - E_1 \text{ est minimale} \\ \text{La différence } E_3 - E_1 \text{ est minimale} \end{array} \right.$$

Le graphique n° 5 résume les variations de ces valeurs et nous constatons que E_2S_2 est toujours supérieur à E_2S_1 et que la différence $S_2E_2 - S_2E_1$ est toujours supérieure à $S_1E_2 - S_1E_1$.

d) En définitive la meilleure combinaison des 4 paramètres est la suivante :

"Souches S_2 , Milieu M_2 et Passage 7 à 9", et nous pouvons quantifier l'augmentation du rendement : la combinaison S_2M_2 multiplie le rendement par 1.365 en moyenne par rapport à S_1M_1 .

Le technicien, en dernier ressort, choisira la solution la plus propice à son problème par la comparaison des différents rendements quantitatifs

qu'il pourrait espérer obtenir, et le calcul des erreurs lui donnera les limites expérimentales théoriques de son choix.

V - CALCUL DE L'ERREUR SUR LES VALEURS NUMERIQUES

La variance résiduelle est $s^2 = 0.00136$.

- pour une réponse, l'écart-type sera $s = 0.037$ et $2s = 0.074$ pour un risque d'erreur de $\alpha = 0.05$.

- pour une somme de N réponses, la variance sera s^2/N et l'écart type s/\sqrt{N}

- pour une différence entre deux moyennes de N réponses chacune, la variance sera $2 s^2/N$.

CONCLUSION

Le progrès de la technologie biologique et les nécessités de la rentabilité optimale de la production biologique ont rapidement entraîné l'emploi d'analyses statistiques dans la recherche et la mise au point des procédés de fabrication.

L'analyse de variance de type factoriel permet d'étudier l'action simultanée de plusieurs facteurs (qui peuvent être sélectionnés auparavant par une analyse classique), ce qui apporte une économie des moyens et une précision sur les interactions éventuelles entre ces facteurs.

Ce type de plan expérimental entraîne certaines obligations, peu restrictives, et les calculs numériques complexes sont facilités par l'utilisation de machines à calculer perfectionnées et/ou d'ordinateurs.

BIBLIOGRAPHIE

- [1] SCHWARTZ D. - Méthodes statistiques à l'usage des médecins et des biologistes. Editions médicales FLAMMARION PARIS 1969.
- [2] PHILIPPE J. - Les méthodes statistiques en pharmacie et en chimie. Monographie de pharmacie MASSON & Cie, PARIS 1967.
- [3] MATHER K. - Analyse statistique en biologie. ACTA Editions GAUTHIER-VILLARS, PARIS 1965.
- [4] DUBOUCARD C., GILBERT H., FONTAINE J. et ROUMIANTZEFF M. - Nouveaux aspects de la culture industrielle du virus de la fièvre aphteuse. IXe congrès international de Microbiologie de MOSCOU 24-30 Juillet 1966.
- [5] ROUMIANTZEFF M., STELLMANN C. et DUBOUCARD C. - Technique de fixation quantitative du complément appliquée à l'étude du virus de la fièvre aphteuse. Bull. Soc. Sci. Vét. LYON, 67, 1965, 243-69.

- [6] ROUMIANTZEFF M., FONTAINE J. et DUBOUCARD C. - Evaluation du pouvoir immunogène du virus aphteux par mesure du pouvoir fixant le complément après traitement par un fluorocarbone. C.R. Acad. Sci. PARIS 261, 1965, 598-601.
- [7] ROUMIANTZEFF M., et STELLMANN C. - Emploi de méthodes automatiques de fixation du complément pour le titrage des antigènes viraux et des sérums correspondants. Symposium Technicon, Européen - Automatisation en chimie analytique - PARIS 2-4 Nov. 1966.
- [8] LISON L. - Statistique appliquée à la biologie expérimentale. La planification de l'Expérience et l'Analyse des Résultats. Gauthier - Villars PARIS 1968.