

Etude de la structure de la consommation
résidentielle en eau potable (sélection
des points de mesure) - Projet Sainte-Foy

INRS-Eau
UNIVERSITE DU QUEBEC
C.P. 7500, Sainte-Foy
Québec G1V 4C7

RAPPORT SCIENTIFIQUE No 14
1973

Rapport rédigé pour
la ville de Sainte-Foy

par
Groupe système urbain

Membres du groupe SYSTEME URBAIN

Bernard BOBEE	Equilibrage, demande
Denis COUILLARD	Demande industrielle et commerciale
Hubert DEMARD	Demande, mesure
Thierry FAURE	Equilibrage, demande résidentielle, mesure
Michel LECLERC	Demande résidentielle
Yves MOREAU	Mesure
Gaëtan THIBAUT	Demande résidentielle
Jean-Pierre VILLENEUVE	Equilibrage, mesure, acquisition des données (Directeur du programme)

TABLE DES MATIERES

	<u>Page</u>
Introduction	1
1 Analyse des données de Montréal	3
1.1 Description et sélection des données	3
1.2 Caractéristiques socio-économiques	4
1.3 Analyse des consommations annuelles	5
1.4 Influence du mode de tarification sur la consommation	10
1.5 Tendances inter-annuelles	12
1.6 Variations saisonnières	14
1.6.1 Été: du 15 mai au 15 octobre	14
1.6.2 Été: du 1er juin au 1er septembre	15
Conclusion	16
2 Analyse des données de Ste-Foy	18
2.1 Description et compilation des données	18
2.2 Caractéristiques des résidences de Ste-Foy	19
2.2.1 Répartition des blocs suivant leur nombre d'appartements	20
2.2.2 Répartition des blocs selon leur rôle d'évaluation	22
2.2.3 Autres variables socio-économiques	24

TABLE DES MATIERES

	<u>Page</u>
2.3 Consommations annuelles	33
2.3.1 Répartition de la consommation annuelle	34
2.3.2 Tests de comparaison des moyennes de consommation annuelle	36
2.4 Relations entre consommations annuelles et évaluations	49
3 Comparaison des résultats obtenus à Montréal et à Ste-Foy	53
3.1 Caractéristiques socio-économiques	53
3.2 Consommations annuelles	53
4 Choix des points de mesure	55
4.1 Regroupement en classe	55
4.2 Répartition des points de mesure par classe	57
4.3 Localisation des points de mesure	

Annexe A1 - Tests statistiques

A1.1	Test de Student (t): comparaison de deux moyennes	A1.1
	A1.1.1 Variances des populations égales	A1.1
	A1.1.2 Variances des populations inégales	A1.3
A1.2	Test d'égalité de deux variances (Fisher)	A1.5
A1.3	Test d'égalité de p variances (Bartlett)	

Annexe A2 - Programmes

A2.1	Caractéristiques socio-économiques de la ville de Ste-Foy	A2.1
	A2.1.1 Création d'une bande contenant les consommations et les évaluations pour chaque numéro de rôle	
	Programme TØTAL	A2.1
	A2.1.2 Etudes des caractéristiques socio-économiques	
	Programme FREQUEN	A2.2
A2.2	Analyse des relevés trimestriels de la ville de Montréal	A2.4
	A2.2.1 Programme MØYAN	A2.4
	A2.2.2 Programme MØYT	A2.5
	A2.2.3 Programme MØIT1	A2.5
	A2.2.4 Programme MØYSØN	A2.6
	A2.2.5 Programme TT (APL/360)	A2.7

		<u>Page</u>
	<u>Annexe A3 - Régression linéaire</u>	A3.1
A3.1	Intervalle de confiance sur B	A3.2
A3.2	Intervalle de confiance sur A	A3.3
A3.3	Intervalle de confiance sur R	A3.3
	<u>Annexe A4 - Méthodes d'échantillonnage</u>	A4-1
A4.1	Position du problème	A4-1
A4.2	Méthode de répartition optimum	A4-2
A4.3	Méthode d'échantillonnage proportionnel	A4-4
A4.4	Comparaison des 2 méthodes	A4-5

LISTE DES TABLEAUX

	<u>Page</u>
1.1 Enquête téléphonique	5
1-2 A Moyenne annuelle - un compteur par logis	7
1-2 B, C Moyenne annuelle - un compteur par bloc	8-9
1.3 Egalité des moyennes des consommations annuelles	11
1.4, 1.5 Variations des moyennes saisonnières	14-15
2.1 Répartition des blocs et des appartements par nombre d'appartements dans le bloc. Source: rôle d'évaluation	21
2.2 Répartition des blocs d'après le nombre de logis des- servis. Source: relevés des compteurs	35
2.3 Matrice logique de comparaison des moyennes	48
2.4 Résultats des régressions	51
4.1 Regroupement en classes des différents types de résidences	56
4.2 Echantillonnage	58

LISTE DES FIGURES

	<u>Page</u>
1.1 Variations inter-annuelles de consommation d'eau basées sur les relevés trimestriels de la ville de Montréal	13
2.1 a Répartition de la consommation annuelle des blocs d'après le nombre d'appartements	23
2.1 b Répartition des blocs d'après leur rôle d'évaluation	23
2.2 à 2.11 Fréquence des blocs de 1, 2*, 2, 3, 4, 6, 8, 12, 18, ≥ 20 appartements en fonction de l'évaluation d'un appartement	25 à 32
2.12 à 2.22 Histogrammes des fréquences de consommation (1, 2*, 2, 3, 4, 6, 8, 12, 16, 18, ≥ 20 logis par compteur)	37 à 47
4.1 à 4.4 Régressions consommation - évaluation (classes 1, 2*, 2, ≥ 3)	56 - 58 à 60

INTRODUCTION

Le projet de Ste-Foy doit nous permettre de déterminer la structure de consommation des résidences multifamiliales; la détermination de cette structure est basée sur la mesure in situ des consommations; il va de soi qu'il est impossible d'effectuer des mesures sur toutes les résidences. Une des premières étapes du projet consiste donc à établir un plan d'échantillonnage qui permette à partir de quelques unités de déterminer cette structure.

L'établissement d'un plan d'échantillonnage est habituellement basé sur une connaissance préalable de la variable que l'on veut mesurer; dans le cas qui nous préoccupe c'est la consommation instantanée; cependant la seule connaissance que nous avons de cette variable consiste en des moyennes annuelles ou trimestrielles. C'est donc à partir de ces variates que nous allons procéder pour établir le plan d'échantillonnage, en n'oubliant pas toutefois que la variable à échantillonner n'est pas une moyenne mais une valeur instantanée.

Le but visé par l'échantillonnage est donc d'obtenir un échantillon qui traduise le mieux possible les variations de la structure de consommation, afin que ces résultats soient significatifs de l'ensemble et permettent une extension à un groupe de consommateurs (quartier, ville, . . .)

La première étape consiste à comparer les moyennes et les variances des consommations pour différentes classes de blocs dans le but d'établir les classes qui ont des moyennes et des variances non significativement différentes.

La deuxième étape consiste à étudier la répartition du nombre d'appartements de la ville de Ste-Foy, élément qui sert dans la répartition des appareils de mesures.

La troisième étape permet en utilisant les variables socio-économiques d'établir à l'intérieur d'une même classe la répartition des points de mesure qui lui ont été attribués.

Après étude des données dont nous disposons pour la CUM et celles dont nous disposons pour la ville de Ste-Foy, nous allons établir la distribution des points de mesure en suivant les différentes étapes énumérées ci-haut.

1 ANALYSE DES DONNEES DE MONTREAL:

L'étude des consommations est basée sur les relevés trimestriels de 367 compteurs pour les blocs appartements de la ville de Pointe aux Trembles sur l'Ile de Montréal. Compte tenu des données disponibles nous avons analysé

- les consommations annuelles pour chaque classe de blocs et pour l'ensemble des blocs.
- les variations saisonnières de la consommation.
- l'influence des compteurs.

1.1 Description et sélection des données

Les relevés effectués trimestriellement sont portés sur des feuilles qui sont identifiées par

- numéro de compte
- numéro et marque du compteur
- adresse
- nom et adresse du propriétaire
- nombre de logis desservis par le compteur

Les relevés eux-mêmes sont notés avec les dates correspondantes et les consommations sont obtenues par soustraction avec les relevés précédents . Les pannes et les changements de compteurs sont également indiqués nous évitant aussi de faire intervenir les mesures provenant de compteurs défectueux.

D'autre part, les relevés des compteurs correspondant à un seul logis ont été écartés lorsque le nombre total de logis dans le bloc n'était pas défini. De plus, les relevés comportant des risques d'erreurs sur la date ou sur les consommations ont été éliminés.

Enfin, la première année d'enregistrement a été supprimée pour éviter les sous-estimations de consommation dues à un faible taux d'occupation et les sur-estimations dues aux consommations relevées à la période de construction du bloc.

1.2 Caractéristiques socio-économiques

Comme nous l'avons vu précédemment, les variations des caractéristiques socio-économiques jouent un rôle important dans le niveau de significativité d'un échantillon. N'ayant aucune indication de ces caractéristiques, nous avons fait une courte enquête téléphonique dont les résultats sont donnés au tableau 1-1.

Il ressort de cette enquête que les caractéristiques des blocs d'une part et des résidents d'autre part sont sensiblement homogènes. Cette homogénéité est particulièrement regrettable car elle limite fondamentalement la généralisation des résultats à d'autres quartiers ou municipalités.

TABLEAU 1-1: ENQUETE TELEPHONIQUE

Questions		Différents types de réponse														
Coût de location \$			90.	80.	85	70	72	80.		95	80		85	85	95.	70.
Absent lors de l'appel (x)	x								x			x				
Gardiennage le jour		x														
Nb. total de résidents		4	4	4	4	5	4	5		3	4		2	4	2	2
Nb. d'adultes résidents		2	3	2	2	4	2	2		3	2		2	2	2	2
Nb. de chambres à coucher		3	3	2	2	2	2	2		2	2		2	2	2	1
Lessiveuse Aut. (automatique) Or. (ordinaire)		aut	or	aut	aut	or	or			aut	or		aut	aut	or	or
Laveuse à vaisselle																
Nb. de toilettes		1	1	1	1	1	1	1		1	1		1	1	1	1
Nb. d'enfants		2		2	2	1	2	3		0	2		0	2	0	0

1.3 Analyse des consommations annuelles

Après sélection des informations, nous avons regroupé les consommations trimestrielles sur une base annuelle par pondération d'après le nombre de jours entre deux relevés dans le but d'éliminer dans un premier temps l'influence des variations saisonnières.

Par ailleurs, un second regroupement a été exécuté d'après les classes de blocs en prenant soin d'établir des classes spéciales pour les appartements munis d'un compteur chacun.

Les résultats de ces regroupements successifs sont présentés dans le tableau 1-2 où figurent la moyenne et l'écart type des consommations annuelles pour chaque classe de blocs et ceci pour les différentes années d'enregistrement (1962 à 1971).

Par ailleurs, des moyennes globales par classe ont également été compilées (Tableau 1-2).

Il ressort de ce dernier tableau que les moyennes globales variant de 99 à 132 GPD (par appartement) dans le cas où un seul compteur enregistre la consommation d'un bloc et de 111 à 146 GPD pour les autres cas. Mentionnons également que la classe la plus nombreuse (2 logis) a une consommation annuelle moyenne de 128 GPD et un écart-type de 56 GPD. Vu la dispersion des moyennes, nous avons fait différents tests statistiques pour vérifier si ces moyennes sont significativement différentes.

Le premier test consiste à vérifier l'égalité des variances entre les différentes classes. Par la suite, les résultats de ce test sont utilisés pour vérifier l'égalité des moyennes (Annexe 1).

CLASSE	ANNEE	MOYENNE (GPD/apt.)	ECART TYPE (GPD/apt.)	MAX. (GPD/apt.)	MIN. (GPD/apt.)	NOMBRE
2 *	66	144	41.7	174	115	2
	67	122	63.0	209	75	4
	68	108	75.9	214	43	4
	69	105	77.3	231	26	5
	70	107	52.8	200	61	6
	71	106	55.5	211	63	6
	66 à 71	111	57.9	231	26	27
4 *	65	147	98.2	217	78	2
	66	154	96.1	222	86	2
	67	146	90.5	210	82	2
	68	135	65.3	204	74	3
	69	172	95.2	272	82	3
	70	147	84.1	241	78	3
	71	132	70.5	232	75	4
	65 à 71	146	69.2	272	74	19
6 *	63	84	6.9	92	80	3
	64	123	46.2	170	67	5
	65	119	38.8	198	77	7
	66	128	55.7	248	47	11
	67	122	42.7	198	53	13
	68	120	34.8	184	66	16
	69	124	30.9	180	78	19
	70	124	38.8	189	58	24
	71	119	46.1	221	21	34
	63 à 71	121	40.6	248	21	132
8 *	62	136	27.5	156	117	2
	63	134	48.0	168	100	2
	64	118	22.0	137	94	3
	65	134	13.0	142	119	3
	66	145	19.4	176	123	5
	67	152	94.1	396	46	11
	68	132	51.7	255	51	22
	69	115	50.5	233	50	29
	70	108	45.1	236	53	39
	71	111	47.9	333	44	70
	62 à 71	118	51.3	396	44	186

TABLEAU 1-2 A: MOYENNE ANNUELLE; UN COMPTEUR PAR LOGIS

CLASSE	ANNEE	MOYENNE (GPD/apt.)	ECART TYPE (GPD/apt.)	MAX. (GPD/apt.)	MIN. (GPD/apt.)	NOMBRE
1	62	122	31.3	156	94	3
	63	101	33.6	168	80	6
	64	130	40.7	199	67	10
	65	126	41.5	217	77	13
	66	136	48.8	248	47	20
	67	134	68.7	396	46	30
	68	126	48.4	255	43	45
	69	120	50.6	272	26	56
	70	115	45.6	241	53	72
	71	114	48.2	333	21	114
	62 à 71	120	49.5	396	21	369
2	61	112	42.8	173	61	7
	62	102	41.2	212	52	18
	63	100	41.4	206	30	23
	64	103	38.3	188	36	29
	65	112	49.0	235	24	32
	66	113	45.5	234	50	36
	67	127	53.6	259	54	45
	68	129	49.0	274	57	57
	69	133	63.3	485	40	77
	70	140	55.5	319	34	85
	71	147	63.8	455	28	93
	61 à 71	128	56.1	485	24	502
3	65	137	88.7	229	52	3
	66	153	94.4	224	46	3
	67	163	111.8	260	43	4
	68	196	145.2	336	40	4
	69	120	62.7	215	27	10
	70	115	61.0	230	26	14
	71	126	53.5	207	30	17
	65 à 71	132	73.4	336	26	55
4	62	82	0.	82	82	1
	63	70	0.	70	70	1
	64	80	0.	80	80	1
	65	115	0.	115	115	1
	66	132	74.2	218	83	3
	67	140	70.5	242	80	4
	68	127	42.7	208	93	6
	69	105	40.4	172	28	8
	70	113	57.8	224	29	8
	71	106	35.9	157	21	10
	62 à 71	113	46.7	242	21	43

TABLEAU 1-2 B: MOYENNE ANNUELLE; UN COMPTEUR PAR BLOC.

CLASSE	ANNEE	MOYENNE (GPD/apt.)	ECART TYPE (GPD/apt.)	MAX. (GPD/apt.)	MIN. (GPD/apt.)	NOMBRE
6	68	122	0.	122	122	1
	69	117	7.0	122	112	2
	70	121	32.9	179	82	9
	71	106	26.5	155	71	11
	68 à 71	113	27.7	179	71	23
8	68	88	18.5	113	69	4
	69	87	21.9	141	39	20
	70	102	29.6	180	52	46
	71	101	23.4	180	65	44
	68 à 71	99	26.1	180	39	114

TABLEAU 1-2 C: MOYENNE ANNUELLE; UN COMPTEUR PAR BLOC

DEFINITION DU CODE DES CLASSES

Classe 1: un logis par compteur: bloc de toute dimension

Classe 2*: un logis par compteur: bloc de 2 logis

Classe 4*: un logis par compteur: bloc de 4 logis

Classe 6*: un logis par compteur: bloc de 6 logis

Classe 8*: un logis par compteur: bloc de 8 logis

Classe 2: deux logis par compteur

Classe 3: trois logis par compteur

Classe 4: quatre logis par compteur

Classe 6: six logis par compteur

Classe 8: huit logis par compteur

Le tableau 1.3 permet d'arriver aux conclusions suivantes:

- Certaines classes peuvent être regroupées: 2 avec 3, 4 avec 6. La classe des blocs de 8 appartements est isolée, elle se rapprocherait peut-être d'une classe de 10 ou 12 appartements.
- Si l'on compare les classes 2* et 2, 3* et 3, 4* et 4, etc., on constate qu'à l'exception de 2* et 2, les moyennes sont significativement différentes. Remarquons également que pour ces mêmes classes, la moyenne des consommations est plus grande lorsqu'il y a un compteur par appartement au lieu d'un compteur pour l'ensemble du bloc. Ce résultat soulève le problème d'influence du mode de tarification sur la consommation.

1.4 Influence du mode de tarification sur la consommation

Les problèmes reliés aux modes de tarification (compteur ou taux fixe, un compteur par bloc ou un compteur par appartement) sont parmi les plus controversés actuellement. Nous avons vu dans le paragraphe précédent qu'à l'exception des blocs de deux appartements, la consommation par appartement dans notre échantillon est plus élevée lorsqu'il y a un compteur par appartement.

Un résumé des résultats est présenté sous forme de matrice symétrique (Tableau 1.3)

Classe	1	2*	4*	6*	8*	2	3	4	6	8
1	1	1	1	1	1	0	1	1	1	0
2*	1	1	0	1	1	1	0	1	1	1
4*	1	0	1	1	1	1	1	0	0	0
6*	1	1	1	1	0	0	1	1	0	0
8*	1	1	1	0	1	0	1	1	1	0
2	0	1	1	0	0	1	1	0	0	0
3	1	0	1	1	1	1	1	0	0	0
4	1	1	0	1	1	0	0	1	1	0
6	1	1	0	0	1	0	0	1	1	0
8	0	1	0	0	0	0	0	0	0	1

TABLEAU 1.3: EGALITE DES MOYENNES DES CONSOMMATIONS ANNUELLES

1 moyennes significativement égales

0 moyennes significativement différentes

Résultat surprenant que l'on devrait étudier plus à fond en tenant compte des éléments suivants:

- a) taux d'occupation
- b) structure de la tarification
- c) précision des mesures par compteur.

1.5 Tendance inter-annuelle

La pratique veut que l'augmentation de consommation d'année en année se situe dans les 2%. Nous avons tenté de voir dans nos relevés si cette tendance à l'augmentation est discernable au niveau de la consommation d'une zone résidentielle.

Le choix de nos classes s'est porté sur celles où il y avait un logis par compteur de façon à éliminer l'effet du taux d'occupation. A partir du tableau 1-2 qui donne la moyenne annuelle de consommation pour chaque année et chaque classe, nous avons construit la figure 1-1.

Il ressort que les variations inter-annuelles mesurées d'après notre échantillon ne semblent pas traduire une tendance et peuvent être interprétées comme un "bruit".

Il semblerait que les augmentations constatées par certains auteurs soient plutôt dues aux nouveaux développements et non à une augmentation systématique de la consommation dans les blocs déjà construits.

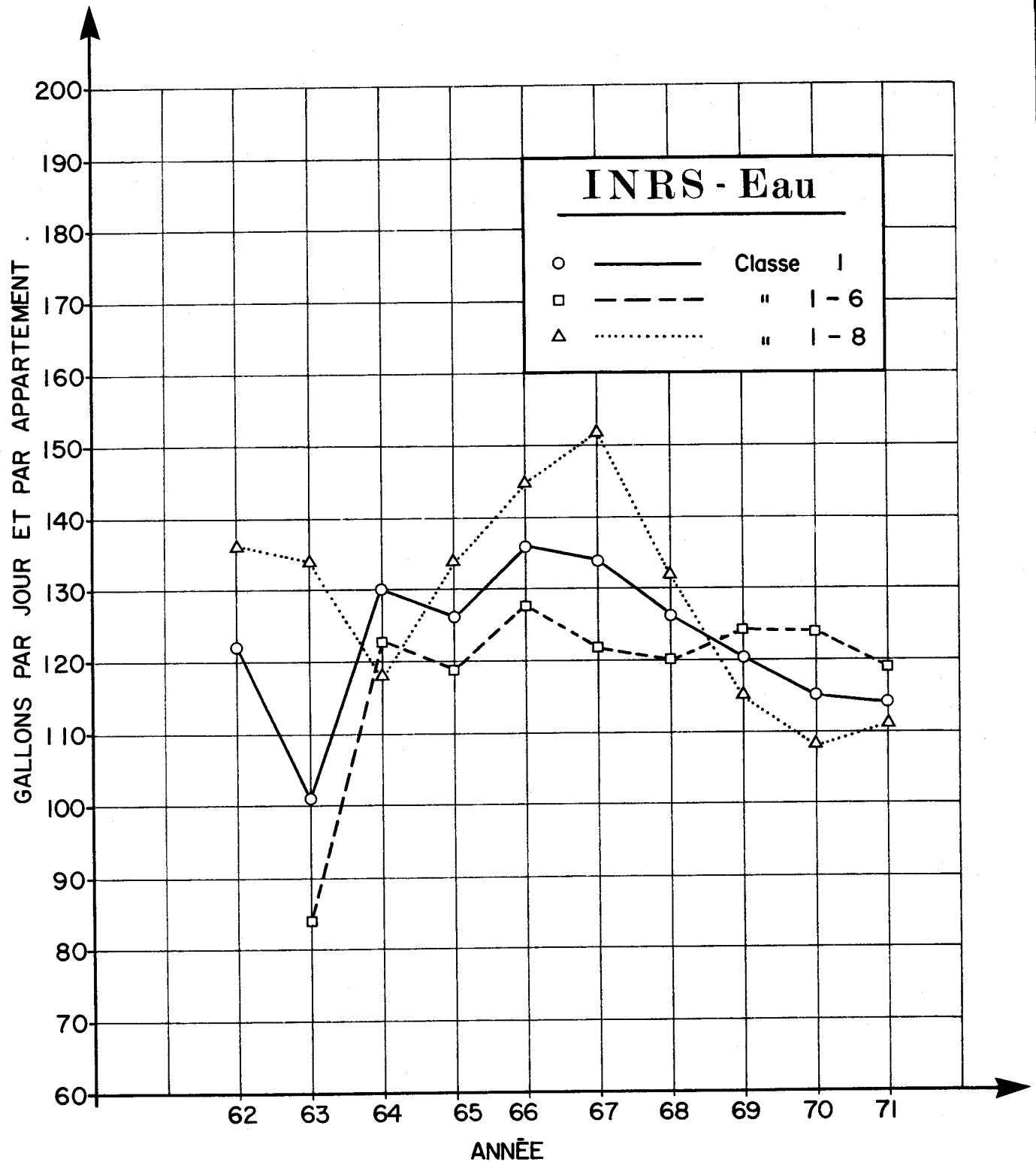


FIGURE:1.1. Variations inter-annuelles de consommation d'eau basées sur les relevés trimestriels de la ville de Montréal.

1.6 Variations saisonnières

Certains usages sont typiquement saisonniers (arrosage et usages extérieurs en général) et déterminent ainsi des variations cycliques. Nous voulons vérifier ici l'importance de ces variations.

Le premier problème est de définir des limites significatives entre l'été et l'hiver. Pour ceci, il faut éliminer les relevés qui couvrent une trop grande période de temps et enfin les relevés qui chevauchent les deux saisons de façon à avoir une idée plus exacte du mode de variation de la consommation entre les saisons.

1.6.1 ETE: du 15 mai au 15 octobre

Classe	Moyenne de l'été	Moyenne de l'hiver	Différence $\frac{M \text{ été} - M \text{ hiver}}{M \text{ été}}$
1	125	117	6 %
2	140	123	12%
3	137	128	6.6%
4	109	115	-5.5%
5	152	145	4.6%
6	116	128	-10.3%
8	93	101	-8.6%
POURCENTAGE DE VARIATION: 8% (en valeur absolue)			

TABEAU 1-4 VARIATIONS DES MOYENNES SAISONNIERES

Ces résultats ne permettant pas de tirer de conclusion, nous avons donc été amené à limiter un peu plus la période "Eté".

1.6.2 ETE: _ _1er_juin_au_1er_Septembre_

Classe	Moyenne de l'été	Hiver	Différence
1	129	116	10%
2	143	124	13.3%
3	120	130	-8.3%
4	128	114	10.9%
5	187	146	21.9%
6	129	123	4.7%
8	98	98	0 %
POURCENTAGE DE LA VARIATION: 16% (en valeur absolue)			

TABLEAU 1.5 VARIATIONS DES MOYENNES SAISONNIERES

On constate cette fois que la moyenne de consommation de l'été est supérieure à celle de l'hiver, sauf les blocs de trois appartements et que cette différence est importante sauf pour la classe des huit appartements.

Ce que l'on retient de ces deux tableaux, c'est d'abord que le simple fait de changer les limites entre les périodes considérées a eu une influence marquée. On peut probablement imputer ce changement au fait que les mois de mai et d'octobre sont généralement des mois où le taux d'occupation des appartements est bas en raison des déménagements. Cependant, le fait d'intégrer des périodes à la saison d'hiver ne nous empêche pas de dire que l'été est une saison où la consommation d'eau dans les unités multifamiliales est plus grande qu'en hiver puisque le pourcentage de variation a quand même doublé par rapport à la moyenne.

CONCLUSION

D'après les relevés trimestriels de 367 compteurs situés dans des blocs appartements à Pointe aux Trembles et après avoir procédé à une sélection de l'information, nous arrivons aux conclusions suivantes:

- les consommations moyennes annuelles par appartement varient de 99 à 132 GPD dans le cas où un seul compteur enregistre la consommation de l'ensemble du bloc et de 111 à 146 GPD dans le cas où il y a un compteur par appartement.

- il est possible de regrouper comme suit les appartements d'après leur consommation annuelle moyenne:
2 et 3 appartements, 4 et 6 appartements, enfin 8 appartements.
- les résultats de l'analyse des consommations annuelles ne permettent pas de distinguer nettement l'influence des compteurs sur la consommation.
- l'analyse des consommations annuelles permet d'affirmer que sur les 12 ans d'enregistrements à Pointe aux Trembles il n'est pas possible d'identifier une tendance inter-annuelle.
- l'analyse des consommations trimestrielles a permis de montrer une augmentation moyenne de l'ordre de 15% de la consommation pendant la période du 1er juin au 1er septembre par rapport à la moyenne du reste de l'année.
- les caractéristiques socio-économiques des blocs ainsi que des habitants des blocs sont trop homogènes pour que l'ensemble des conclusions puisse être généralisé à d'autres quartiers ou d'autres villes.

2 ANALYSE DES DONNEES DE STE-FOY

Compte tenu des données disponibles à Ste-Foy, nous avons étudié les distributions des caractéristiques socio-économiques d'une part et d'autre part celles des consommations moyennes annuelles. Par la suite, nous avons cherché à établir des relations entre ces deux séries de variables.

2.1 Description et compilation des données

Nous avons essayé de réunir le plus d'information possible sur les caractéristiques socio-économiques ainsi que sur la consommation d'eau des résidences de Ste-Foy en nous limitant cependant aux données utilisables directement sur ordinateur.

Grâce au service informatique de la ville de Ste-Foy, nous avons pu compiler les fichiers suivants:

- numéros de rôle;
- superficies et évaluations des terrains;
- évaluations des résidences;
- nombres d'appartements dans les résidences.

De plus, des bandes magnétiques préparées pour la CUQ ont été mises à notre disposition et nous y trouvons les informations suivantes:

- somme des relevés des compteurs pour 70-71
- nombre d'appartements desservis par compteur
- date de relevés.

D'autres informations sont accessibles telles que loyers, nombre d'habitants mais n'ont pu être compilées car elles ne sont pas disponibles sous une forme directement utilisable sur ordinateur.

Une première compilation des fichiers a mis en évidence un certain nombre de cas où le nombre d'appartements porté sur le rôle d'évaluation est différent de celui indiqué sur les relevés de compteurs. Dans la majorité des cas, une vérification a permis de constater qu'il s'agissait d'une erreur sur la bande des consommations. Un cas a cependant attiré particulièrement notre attention, il s'agit des résidences bifamiliales pour lesquelles la différence entre les nombres d'appartements portés sur les deux fichiers nous a permis d'identifier des résidences bifamiliales à 2 compteurs et des résidences bifamiliales à un compteur.

2.2 Caractéristiques des résidences de Ste-Foy

L'ensemble des résidences de Ste-Foy constitue la population dont sera tirée notre échantillon aussi nous sommes nous efforcés, compte tenu des données disponibles pour cette étude préalable, de caractériser au mieux la répartition de ces résidences en fonction de quelques variables jugées significatives.

2.2.1 Répartition des blocs suivant leur nombre d'appartements

Le nombre d'appartements par bloc nous est apparu comme une variable indicative de la grosseur des blocs. La répartition des différents blocs en fonction du nombre d'appartements d'après le rôle d'évaluation conduit aux résultats représentés sur le tableau 2-1.

On retrouvera sur ce tableau deux classes de résidences bifamiliales correspondant à un ou deux compteurs par résidence. Une brève enquête nous a permis de vérifier que la plupart des résidences bifamiliales à 2 compteurs sont du type juxtaposé alors que celles à un seul compteur sont plutôt du type superposé plus ancien ainsi que des résidences unifamiliales où le sous-sol a été aménagé en appartement. Rentrent également dans cette dernière classe un certain nombre de résidences unifamiliales juxtaposées à un seul compteur, de construction identique aux résidences unifamiliales juxtaposées à deux compteurs. Il ressort finalement que:

- a) 83% des résidences de Ste-Foy sont unifamiliales. Si on s'intéresse cette fois à la répartition des appartements dans les blocs, on constate que 40% des appartements tombent dans des résidences unifamiliales. Or le nombre de personnes est habituellement plus élevé dans les résidences unifamiliales que dans les autres. On peut donc penser qu'environ la moitié de la population de Ste-Foy habite des résidences unifamiliales.

Type de bloc	Nombre de blocs	% cumulés	Nombre d'appartements	% cumulés
1	6,599	83.3	6,599	40.0
2*	202	85.8	404	42.5
2	442	91.4	884	47.9
3	139	93.2	417	50.4
4	92	94.3	368	52.6
5	40	94.8	200	53.8
6	76	95.8	456	56.6
7	9	95.9	63	57.0
8	74	96.9	592	60.6
9	8	97.0	72	61.0
10	9	97.1	90	61.6
11	3	97.1	33	61.8
12	27	97.4	324	63.7
13	4	97.5	52	64.1
14	3	97.5	42	64.3
15	8	97.8	120	65.0
16	26	98.0	416	67.6
17	0	98.0	0	67.6
18	31	98.4	558	71.0
19	17	98.6	323	72.9
20 et plus	109	100	4,450	100
TOTAL	7,918		16,463	

TABLEAU 2-1

Répartition des blocs et des appartements par nombre d'appartements dans le bloc. - Source: rôle d'évaluation.

- b) les classes de blocs les plus nombreuses sont dans l'ordre: 2, 3, 20 et plus, 4, 6, 8, puis 18, 12 et 16. En ce qui concerne les classes d'appartements, celle des 20 et plus qui contient 28% du total arrive immédiatement après celle des résidences unifamiliales (40%). Il est cependant difficile d'établir des classes pour les blocs de 20 appartements et plus en raison de leur grande dispersion dans le nombre d'appartements. Leur importance ne devra cependant pas être minimisée.

2.2.2 Répartition des blocs d'après leur rôle d'évaluation

La connaissance de relations entre la valeur estimée des bâtisses et la consommation dans le cas de résidences unifamiliales nous a poussés à étudier la répartition des blocs d'après leur rôle d'évaluation.

Les moyennes et les écarts-types des distributions de l'évaluation pour chacune des classes de blocs sont représentés en Figure 2-1B. Il ressort que les résidences unifamiliales ont une évaluation moyenne de \$16,700.00 environ, les bifamiliales à deux compteurs \$13,300.00, celles à un seul compteur \$10,300.00 et l'ensemble des autres semble assez homogène avec une évaluation moyenne de \$7,000 à \$8,000.00.

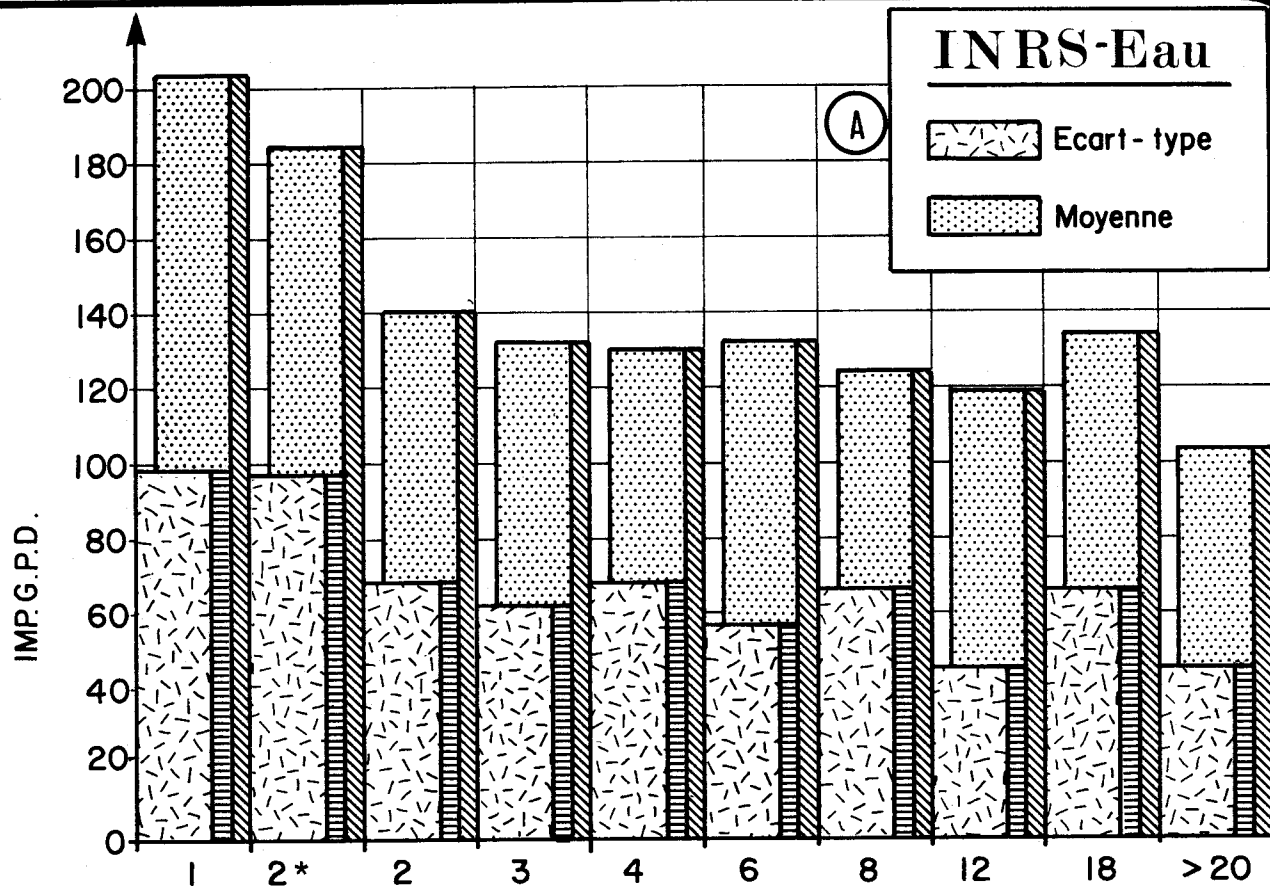
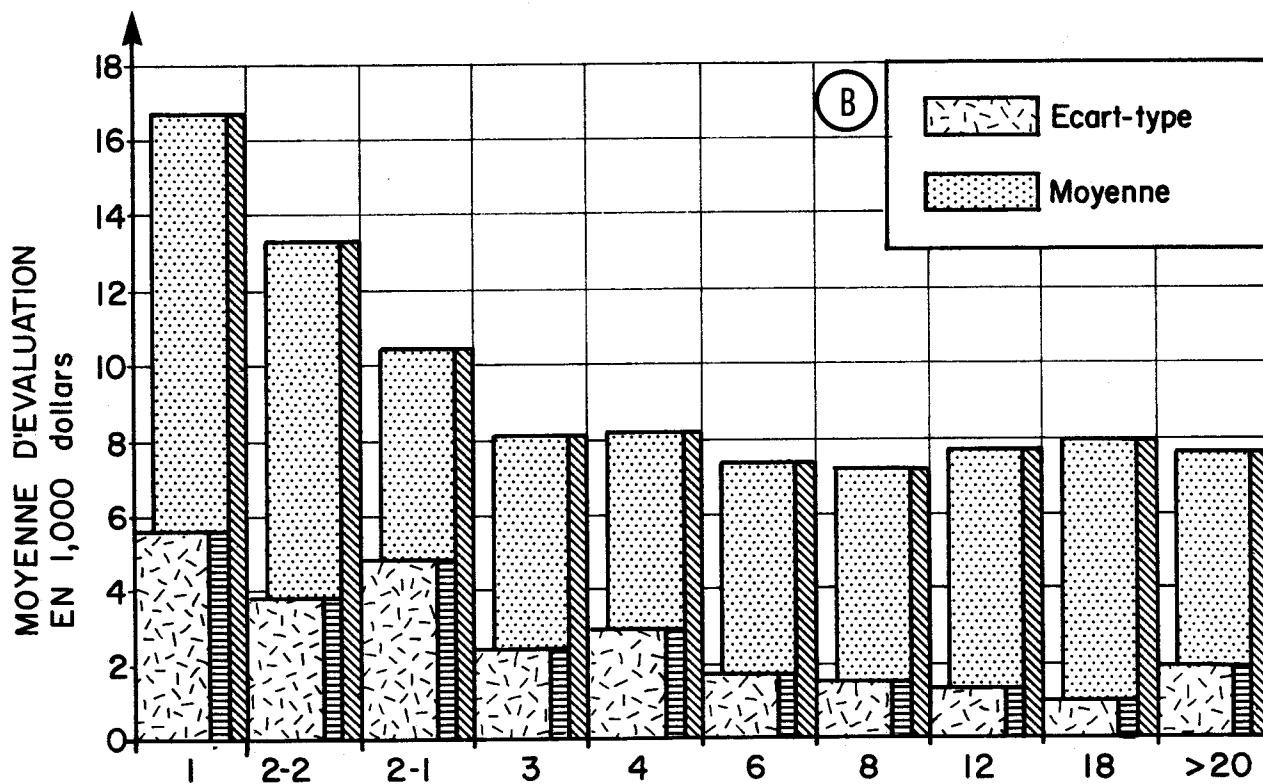


FIGURE 2.1a Répartition de la consommation annuelle des blocs d'après le nombre appartements.



Répartition des blocs d'après leur rôle d'évaluation.

FIGURE 2.1b

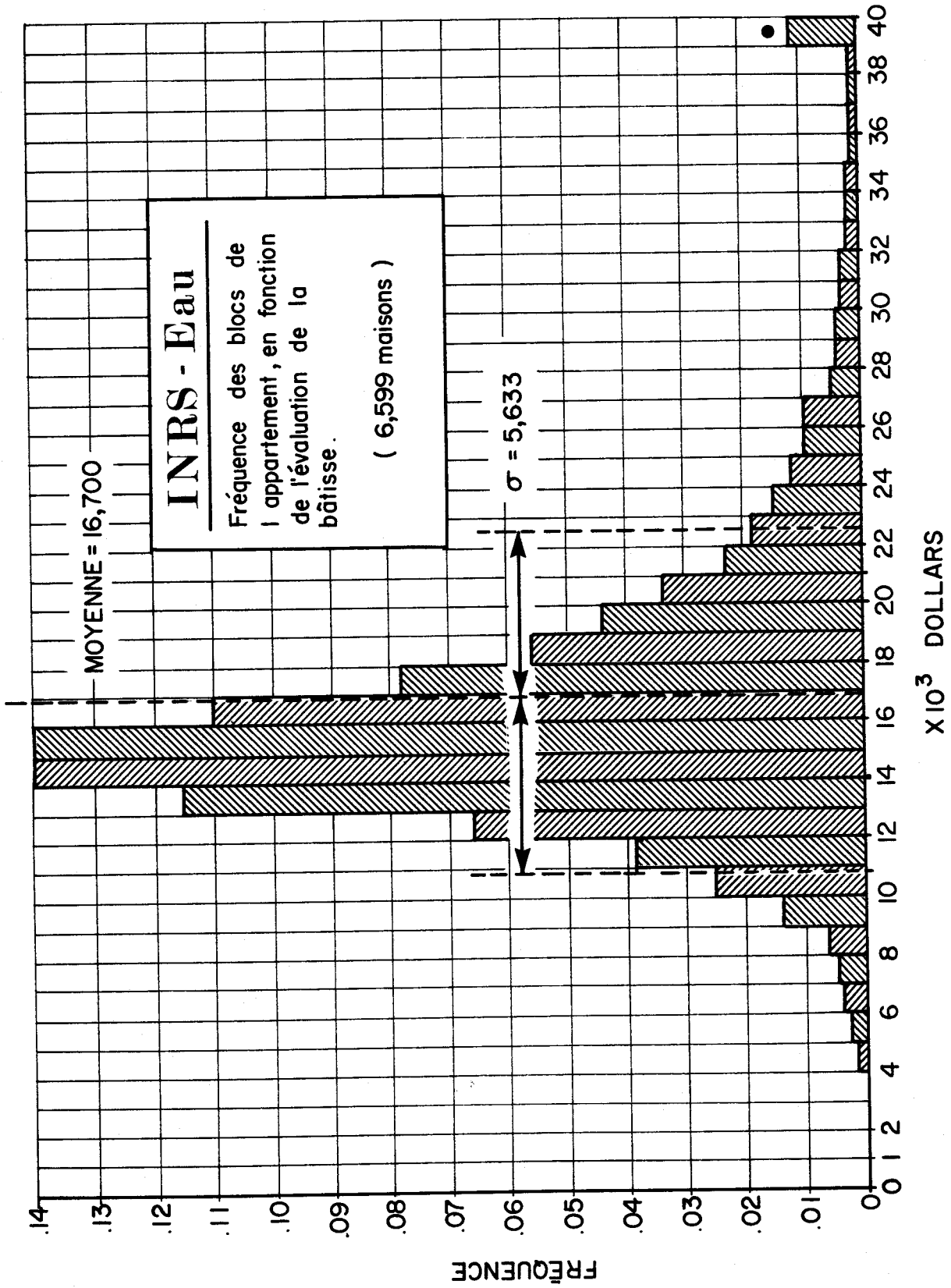
L'étude des écarts-types permet d'arriver aux conclusions suivantes:

- la dispersion semble sensiblement constante et assez importante pour les 1, 2*, 3, 4.
- la dispersion est très importante pour les 2 traduisant l'hétérogénéité de cette classe, ce que l'on a constaté précédemment.
- la dispersion diminue progressivement avec le nombre d'appartements par classe (construction standardisée); cette tendance n'est plus vraie pour les 20 et plus qui sont assez hétérogènes.

Les histogrammes des distributions pour chacune des classes (Figures 2-2 à 2-11) sont d'autant plus réguliers que la taille de la classe augmente. La figure 2-4 se rapportant aux évaluations des résidences bifamiliales à un seul compteur met en évidence une première distribution assez régulière sur laquelle semble être superposé un nouveau pic vers \$13,500. (le même pic qui existe pour les résidences bifamiliales à 2 compteurs). Cette anomalie nous a semblée assez frappante pour que nous la mentionnions ici.

2.2.3 Autres variables socio-économiques

Nous disposons également des superficies et des évaluations des terrains mais ces variables ne nous ont pas semblé suffisamment



● = Cette classe comprend les appartements de 39,000 à 106,000 dollars.

FIGURE 2-2

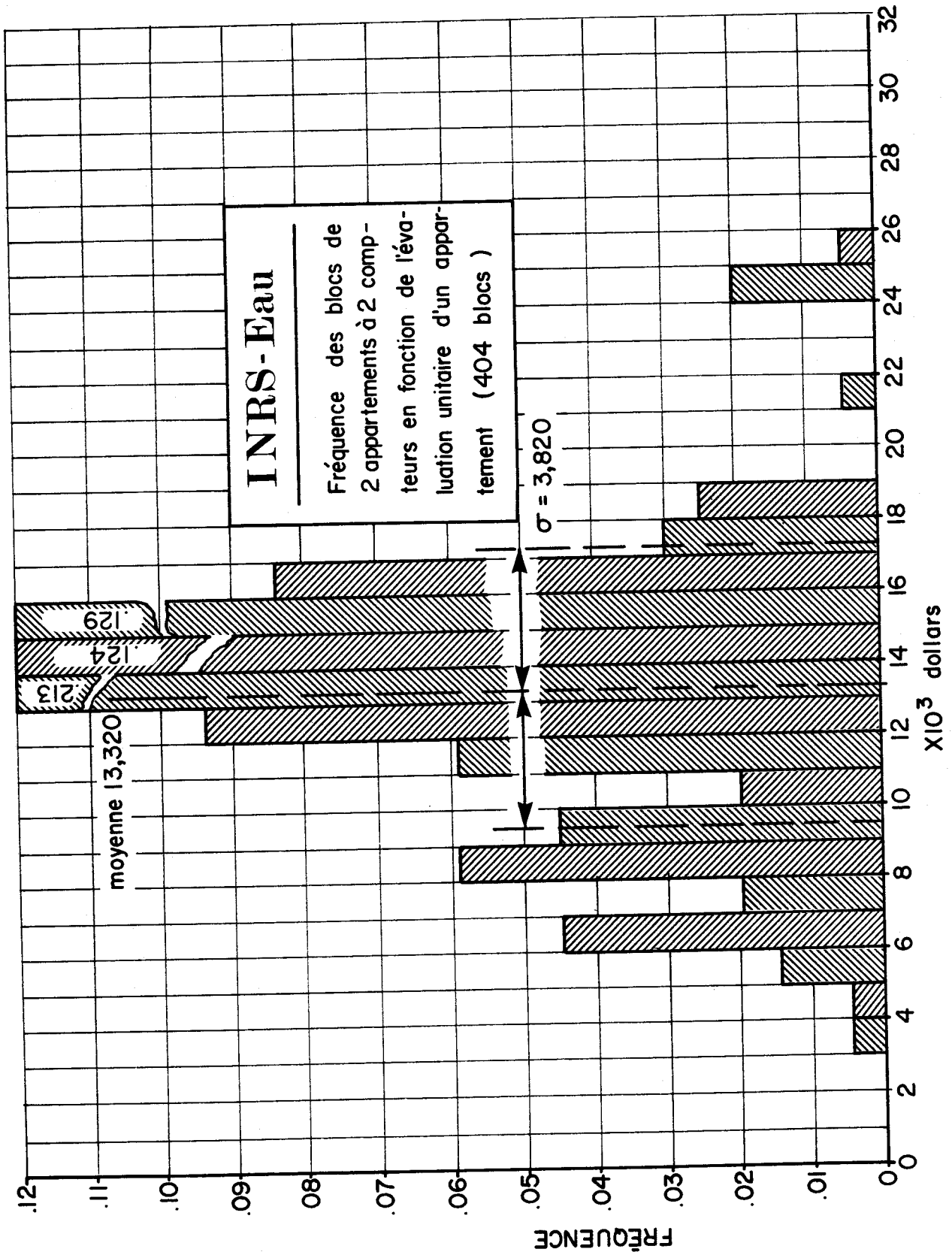
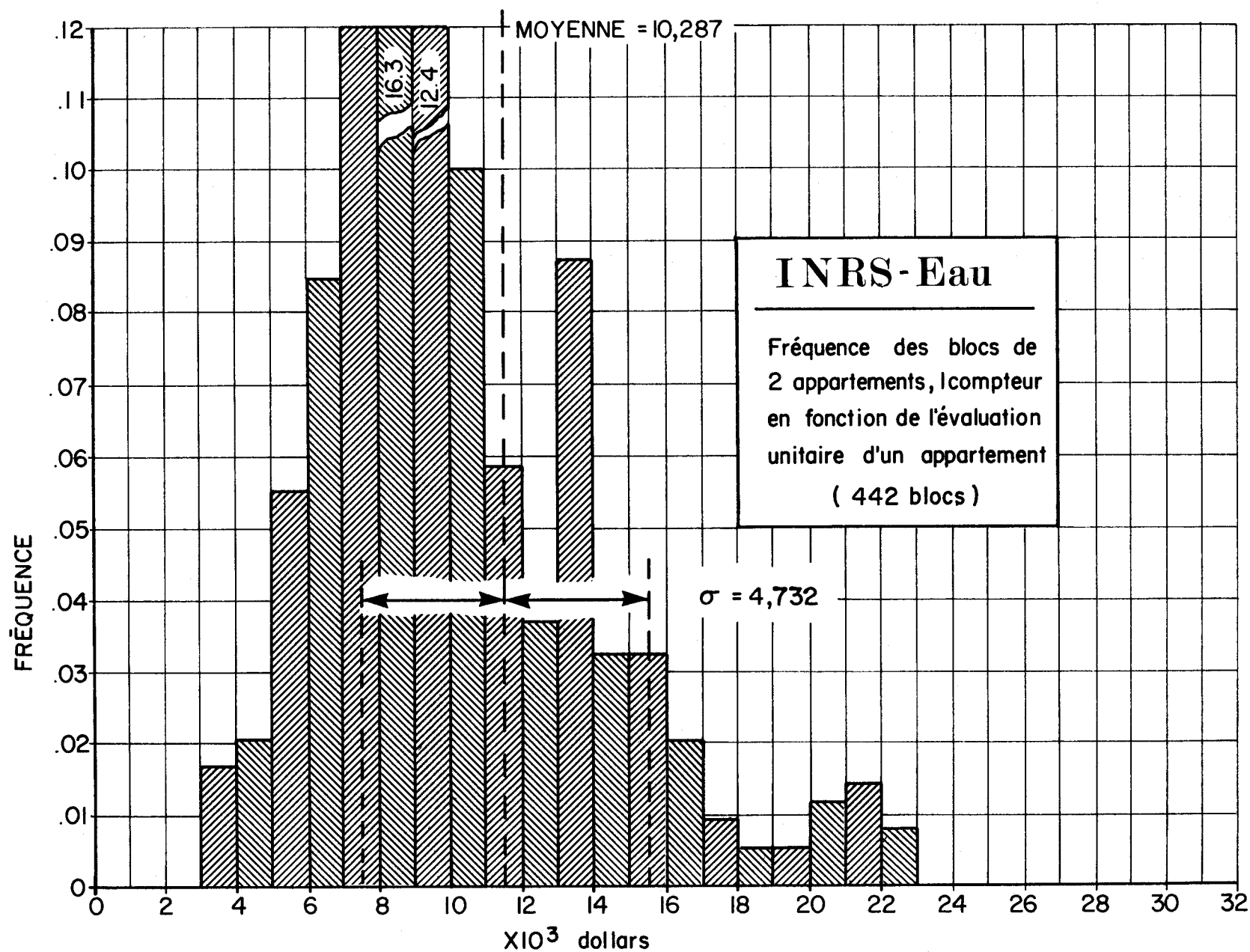


FIGURE 2-3

FIGURE 2-4



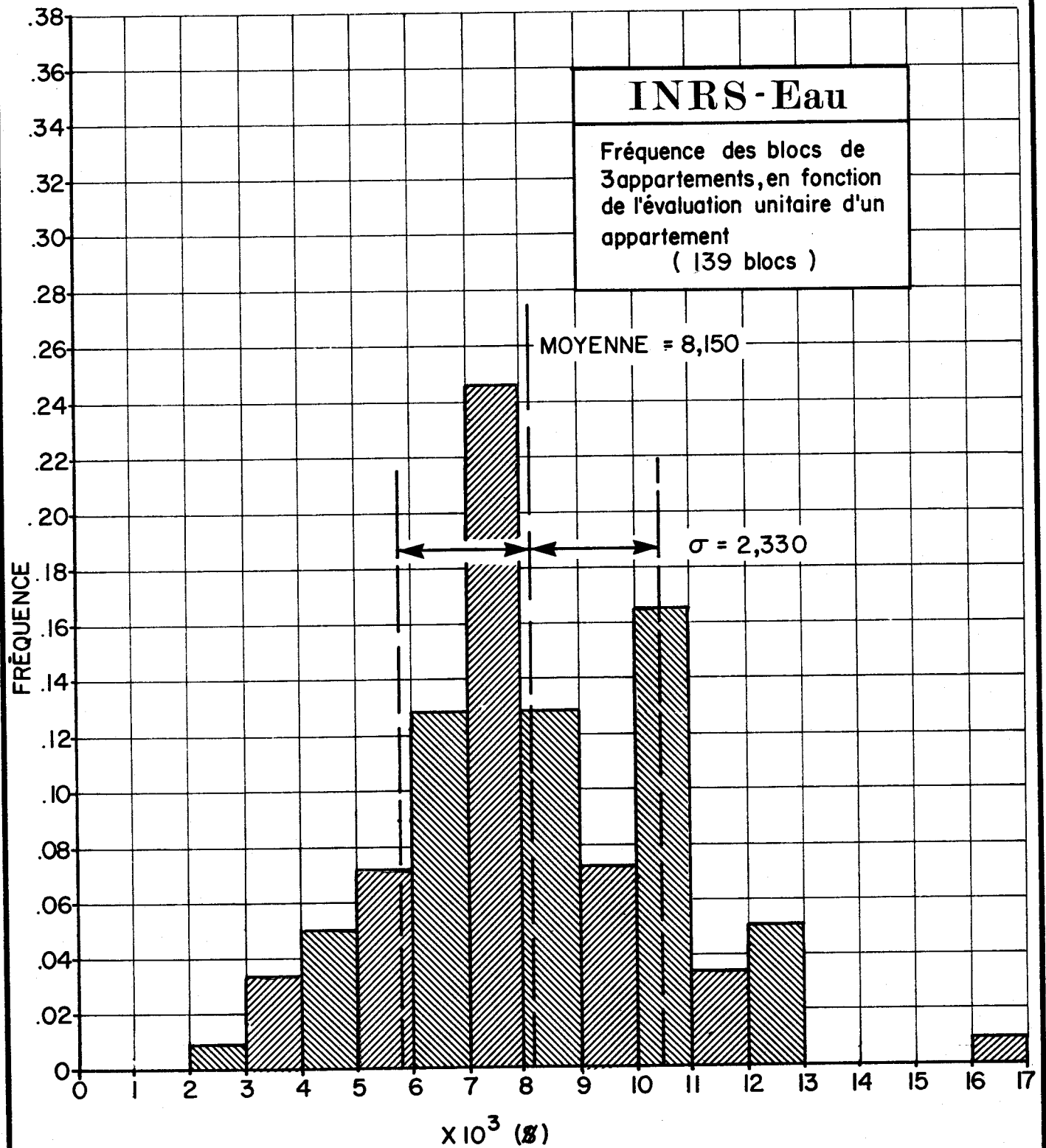


FIGURE 2-5

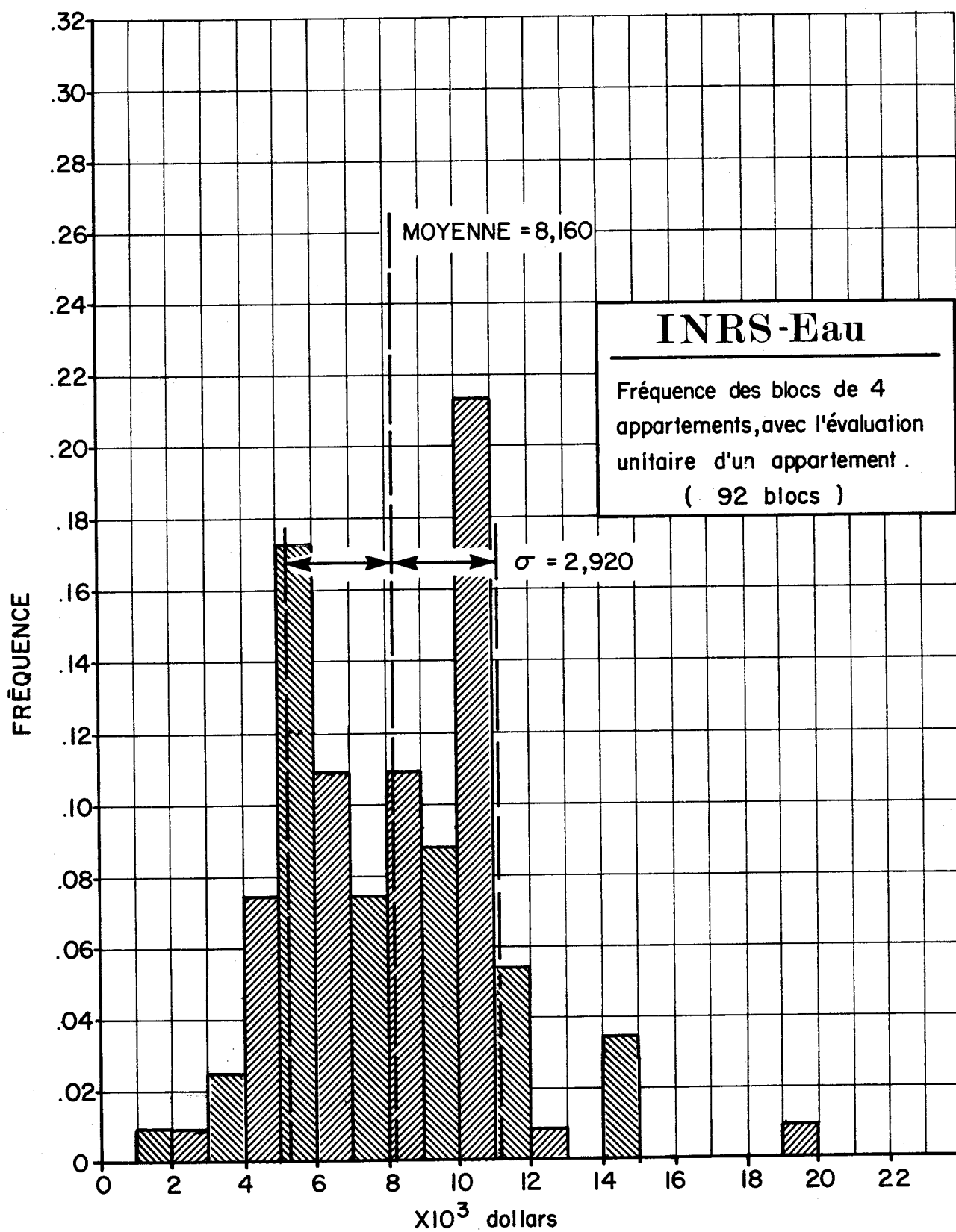


FIGURE 2-6

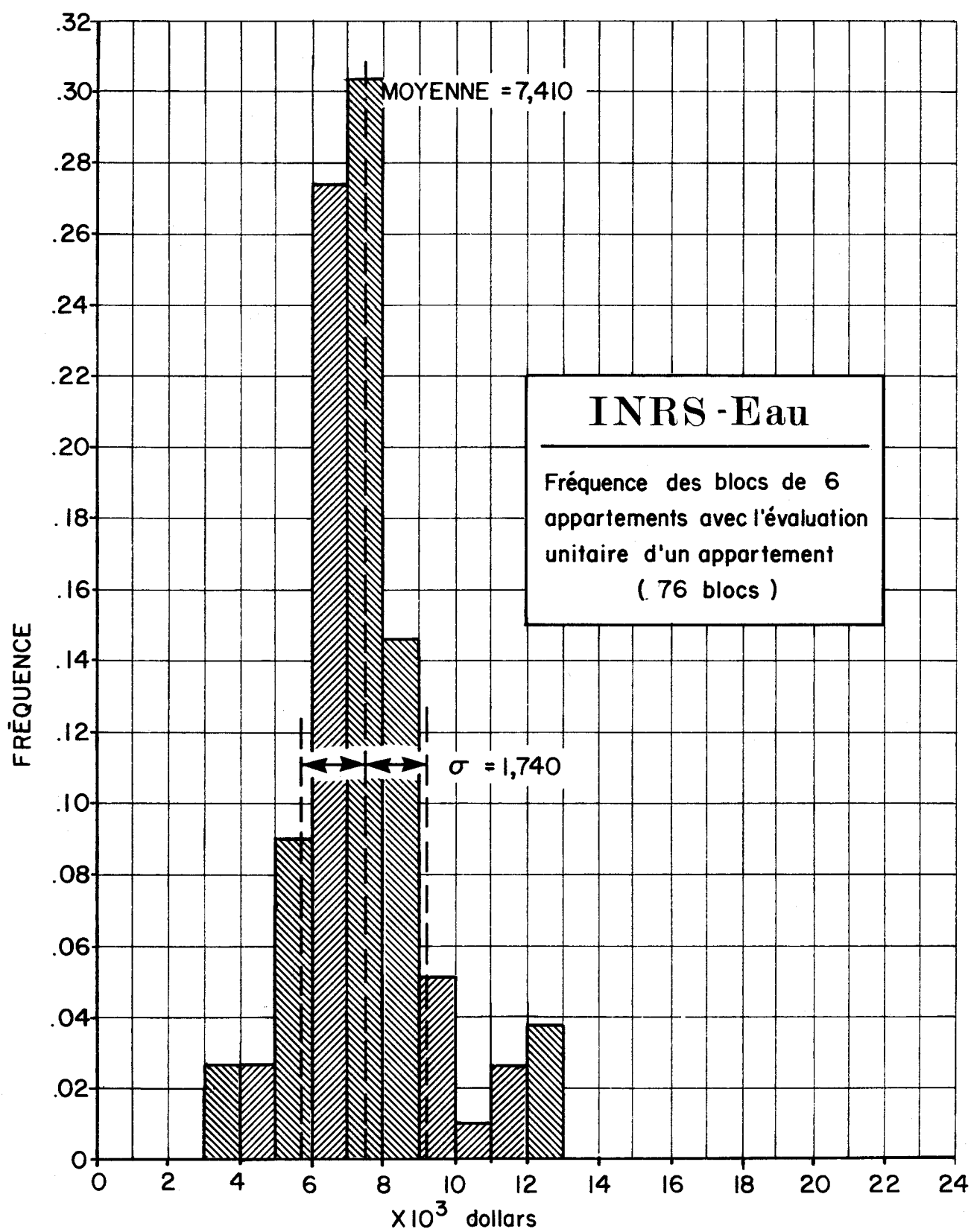


FIGURE 2-7

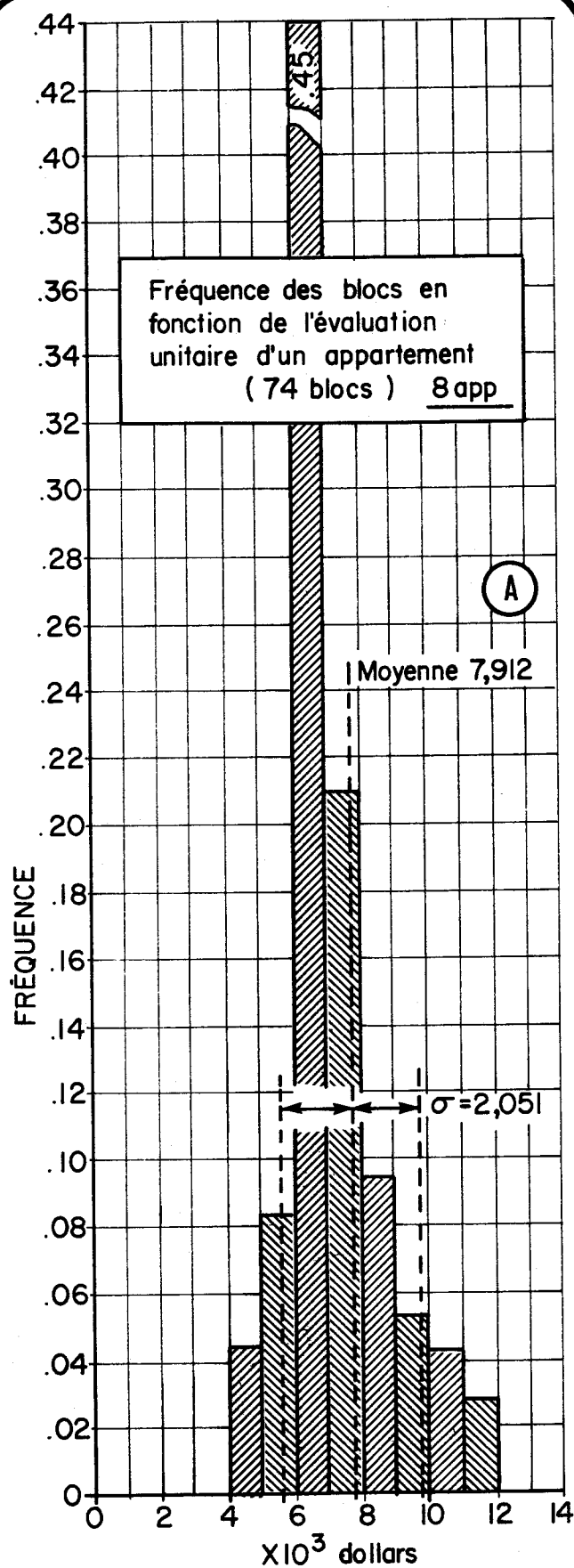


FIGURE 2-8

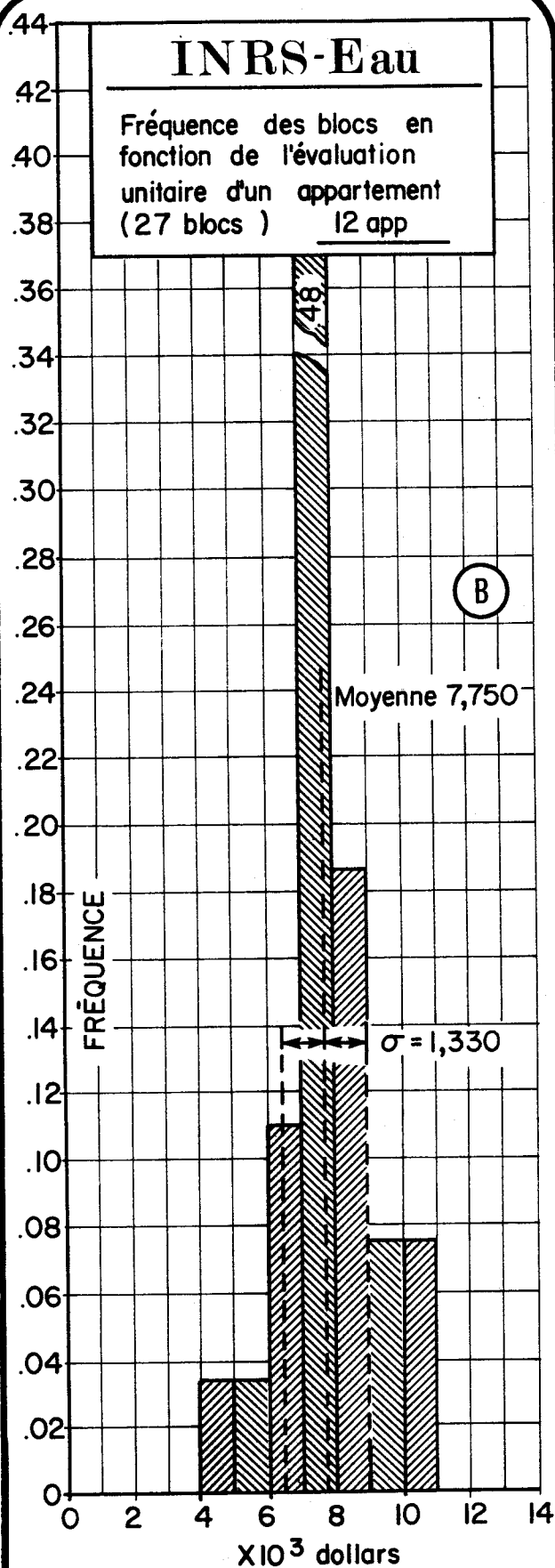


FIGURE 2-9

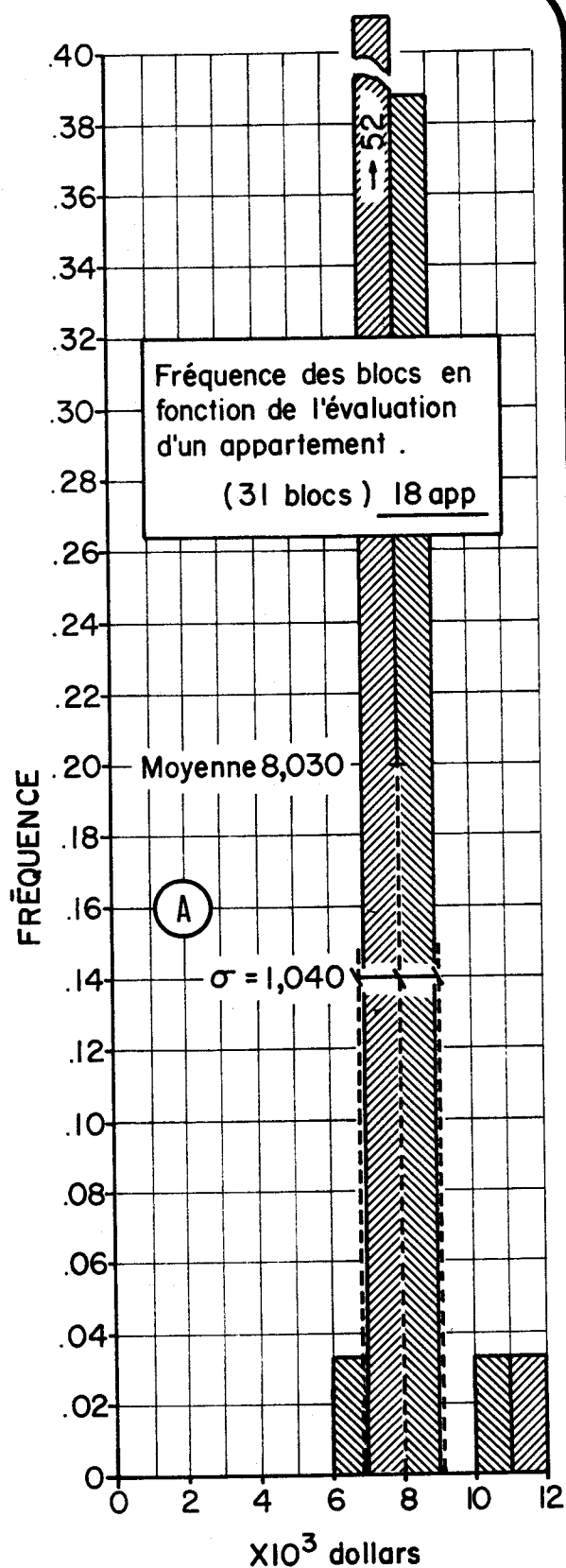


FIGURE 2-10

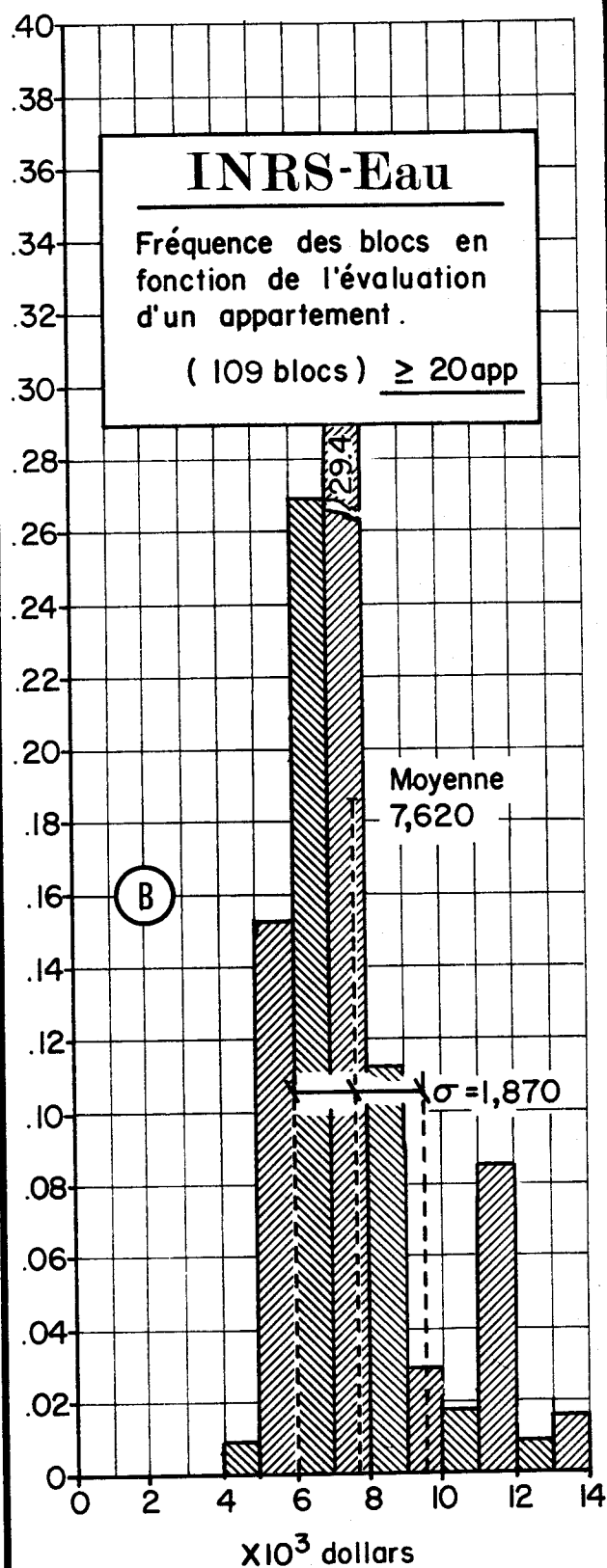


FIGURE 2-11

représentatives pour que nous procédions de la même façon qu'avec l'évaluation des bâtisses. Par ailleurs, des corrélations intéressantes doivent certainement exister entre des variables socio-économiques telles que loyer, nombre de pièces, nombre de personnes mais ce genre d'étude dépasse le cadre du problème posé par l'échantillonnage et sera abordé lors d'une prochaine étape.

2.3 Consommations annuelles

Nous avons utilisé dans cette étude les relevés de compteurs fournis par le service d'informatique de la Ville de Ste-Foy. Ces relevés comportent outre les enregistrements des compteurs, les dates de relevés et nous nous sommes ramenés à des consommations annuelles. Pour faciliter les comparaisons, nous avons également ramené les consommations en gallons par jour et par appartement (GPD). Faute de données, nous n'avons pas pu faire l'analyse des variations saisonnières. Une sélection de l'information a été réalisée pour éliminer les cas d'erreurs suivants:

- erreur dans le nombre d'appartements desservis par un compteur: lorsque la différence entre ce nombre et le nombre indiqué par le rôle d'évaluation dépasse 2, nous avons supprimé l'enregistrement correspondant. Si cette différence est inférieure à 2, nous avons après enquête, considéré comme exact le nombre porté sur le rôle d'évaluation.

- compteurs en panne, relevés inexacts, consommations anormales: les relevés conduisant à des consommations inférieures à 10 GPD ou supérieures à 1,000 GPD sur une base annuelle ont été éliminés.

Il est à remarquer que le taux d'occupation joue également un rôle dans le calcul de la consommation. Nous n'avons cependant pas pu en tenir compte faute d'information. Nous avons ainsi défini l'ensemble des résidences dont la consommation est considérée comme valable. Cet ensemble est légèrement plus petit que celui correspondant au rôle d'évaluation (Tableau 2-2). Mentionnons également qu'il s'agit de compteurs volumétriques, conventionnels (disque en rotation) dont nous ne pouvons en aucun cas garantir la précision pour des débits inférieurs à .6 à .8 GPM (ϕ 5/8").

2.3.1 Répartition de la consommation annuelle

Pour les différentes classes de blocs mentionnées précédemment, nous avons trouvé des consommations annuelles moyennes décroissant de 205 GPD (résidences unifamiliales - figure 2-12) à 103 GPD (vingt appartements et plus) (figure 2-22). Ceci est à rapprocher de la figure 2-1a où l'évaluation décroît également en fonction du nombre d'appartements. Une analyse statistique permettra de regrouper ces différentes consommations (paragraphe 3.2).

<u>Type de bloc</u>	<u>Nombre de blocs d'après les relevés de compteurs</u>	<u>Différence avec rôle d'évaluation</u>	<u>Nombre d'appartements</u>
1	6,444	155	6,444
2*	201	1	402
2	364	78	728
3	139	0	417
4	84	8	336
5	31	9	155
6	72	4	432
7	9	0	63
8	72	2	576
9	6	2	54
10	9	0	90
11	3	0	33
12	21	6	252
13	3	1	39
14	3	0	42
15	5	3	75
16	19	7	304
17	0	0	0
18	28	3	504
19	18	-1	342
20 et plus	74	35	2,749
TOTAL:	7,605	313	14,037

TABLEAU 2-2: Répartition des blocs d'après le nombre de logis desservis. Source: relevés des compteurs.

Note: Il y a un bloc considéré comme ayant 19 appartements au point de vue des consommations, mais étant en fait un 18 appartements, au point de vue des évaluations.

La dispersion des consommations autour des valeurs moyennes semble plus aléatoire que pour l'évaluation mais indique cependant une concentration plus forte pour les 12 et les 16 alors que les 18 semblent correspondre à un point singulier. Le groupe des 20 et plus se distingue comme pour l'évaluation par une dispersion assez grande semblable en fait à celle des groupes de 1 à 10.

Les histogrammes des distributions de consommation pour chacune des classes sont représentés dans les figures 2.12 à 2.22. Mentionnons que dans le cas de la classe 2, on ne retrouve qu'un seul pic, ce qui est difficilement explicable.

2.3.2 Tests de comparaison des moyennes de consommation annuelle

Les tests de comparaison de moyennes d'échantillons utilisés précédemment ont été appliqués ici dans le but d'établir des regroupements de classes. Les résultats sont également donnés sous forme de matrice (Tableau 2.3).

Il ressort que:

- les résidences unifamiliales constituent une classe à part (204 GPD) ainsi que les bifamiliales à 2 compteurs (183 GPD). Ceci s'interprète par le fait que cette seconde classe correspond à des résidences dont les caractéristiques s'apparentent plus à celles des résidences unifamiliales qu'à celles des appartements.

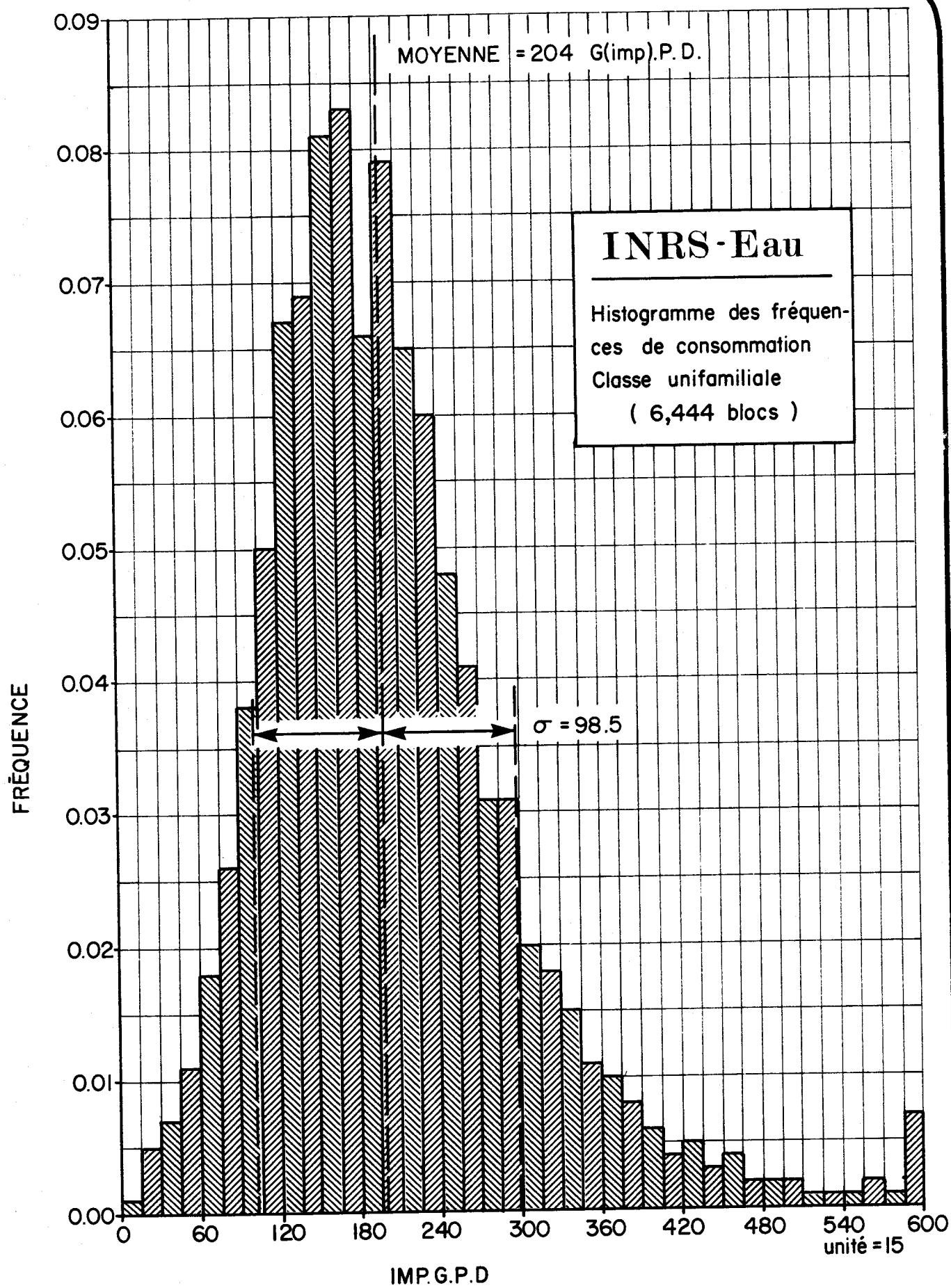


FIGURE 2-12

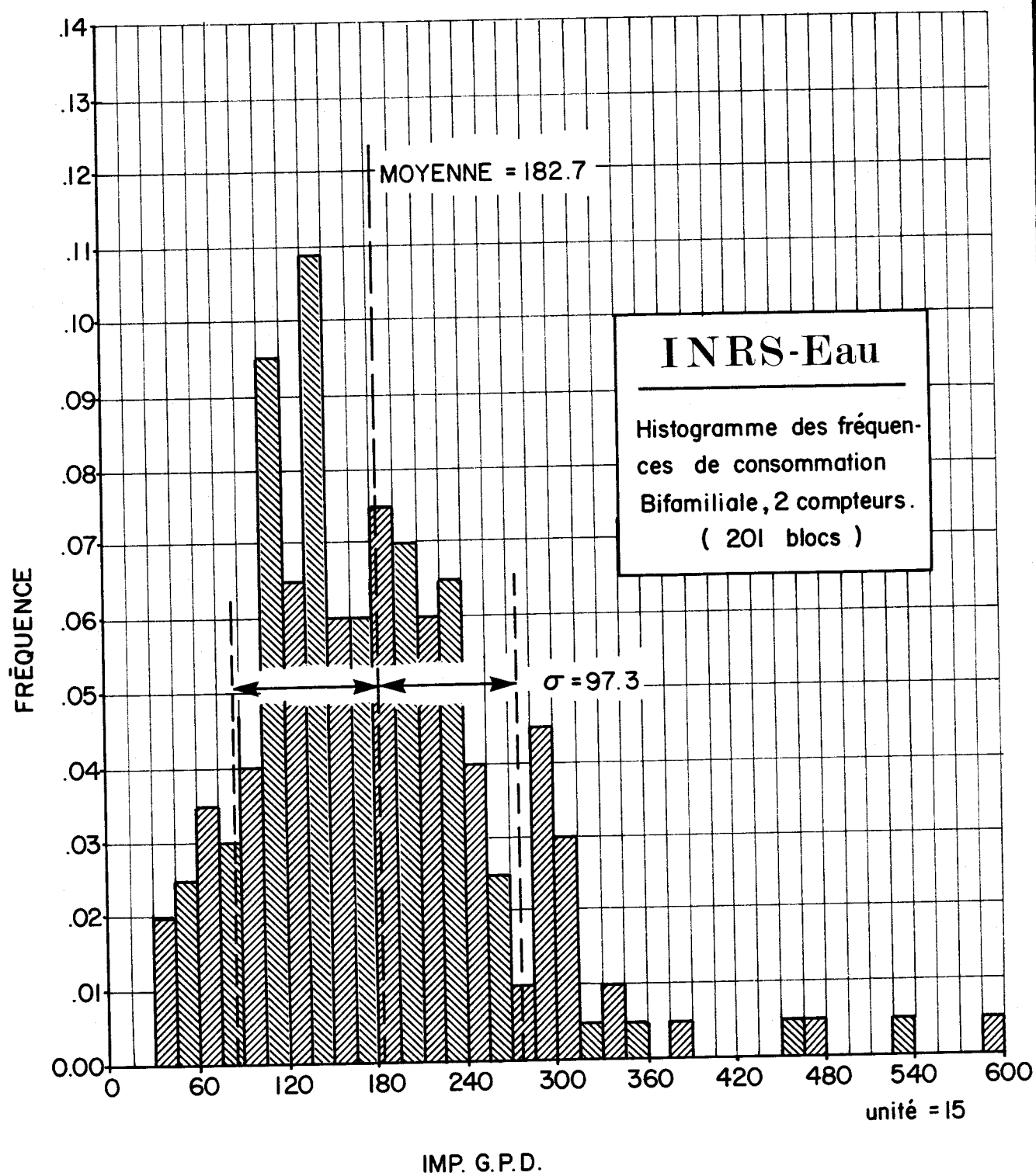


FIGURE 2-13

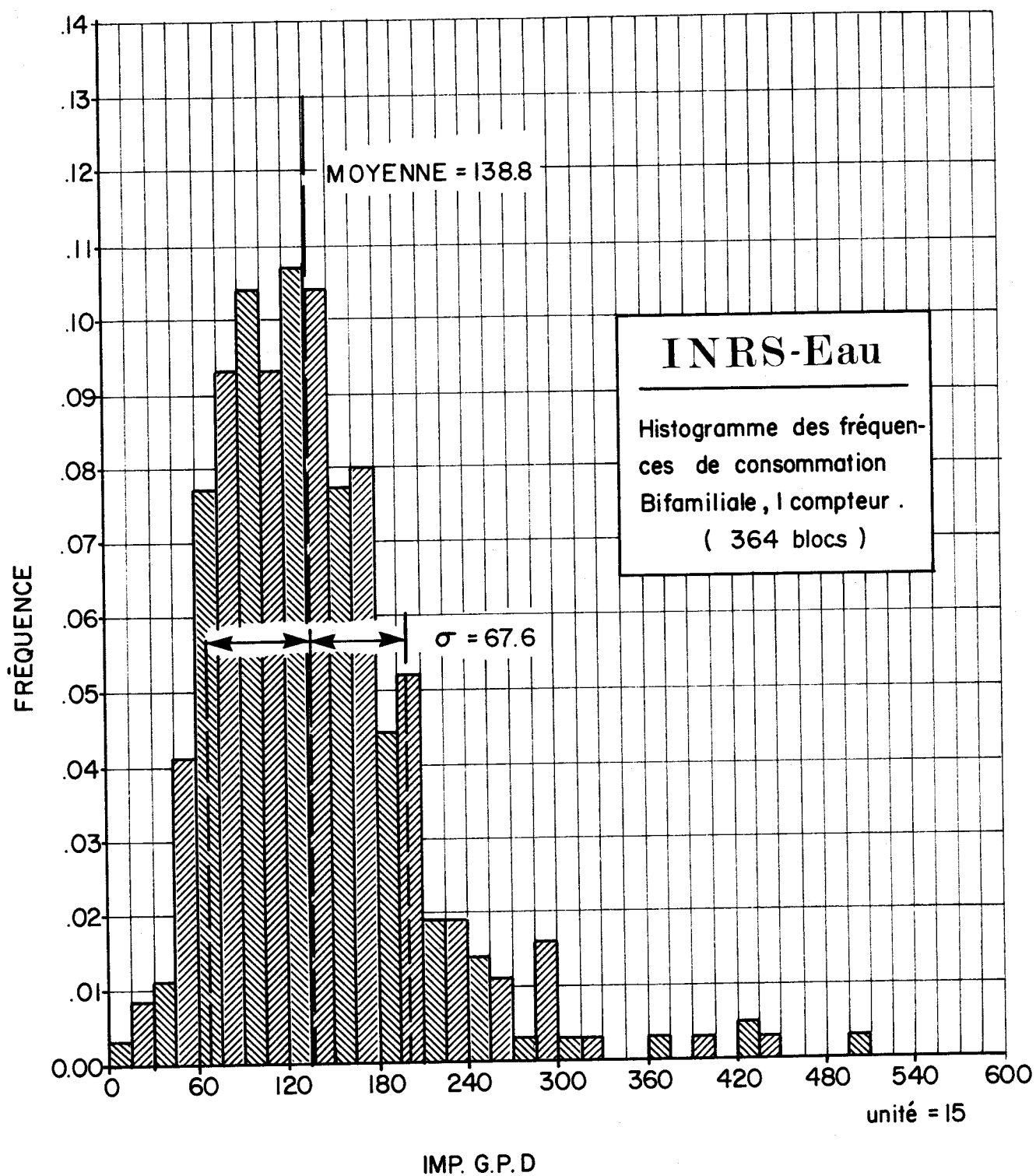


FIGURE 2-14

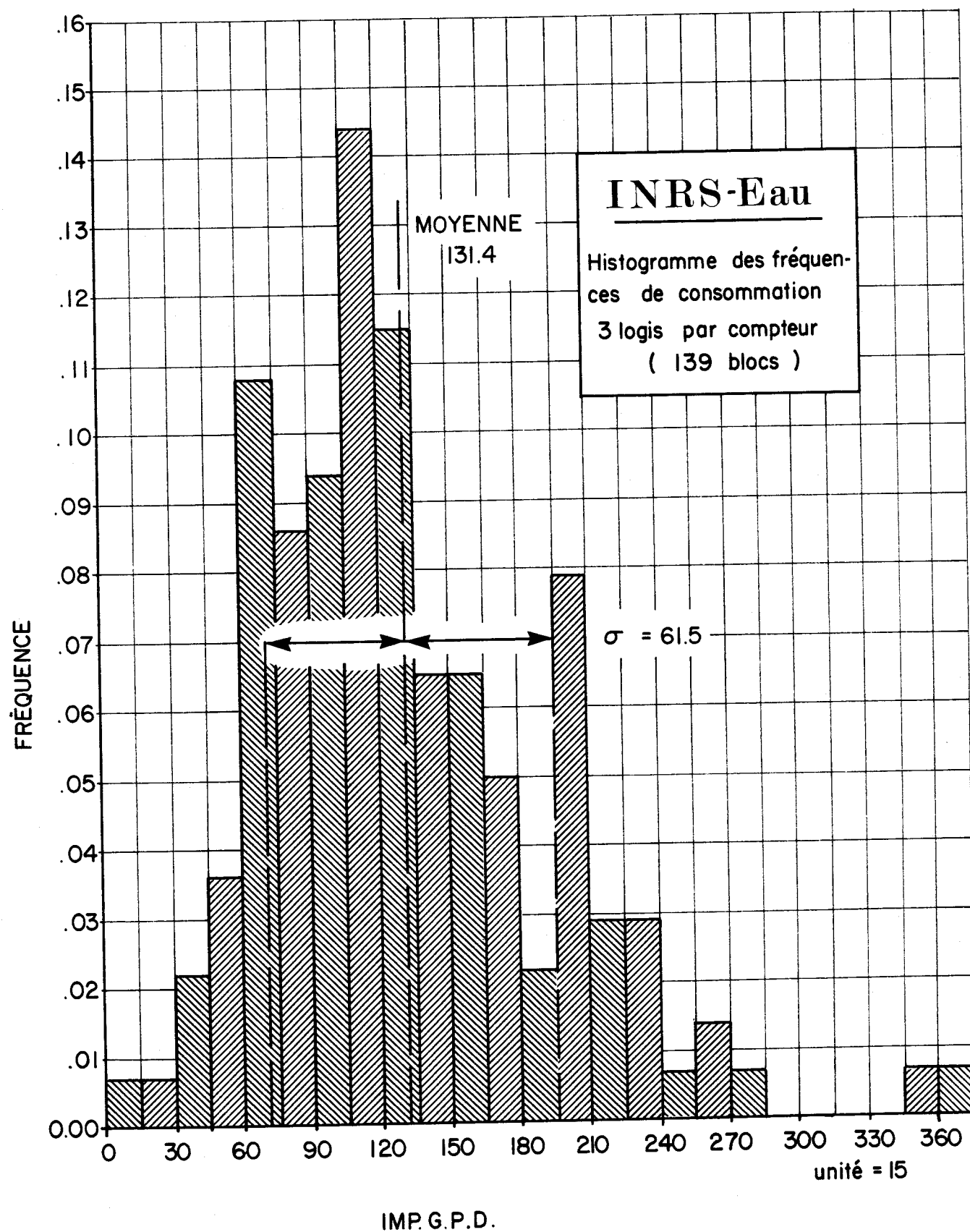


FIGURE 2-15

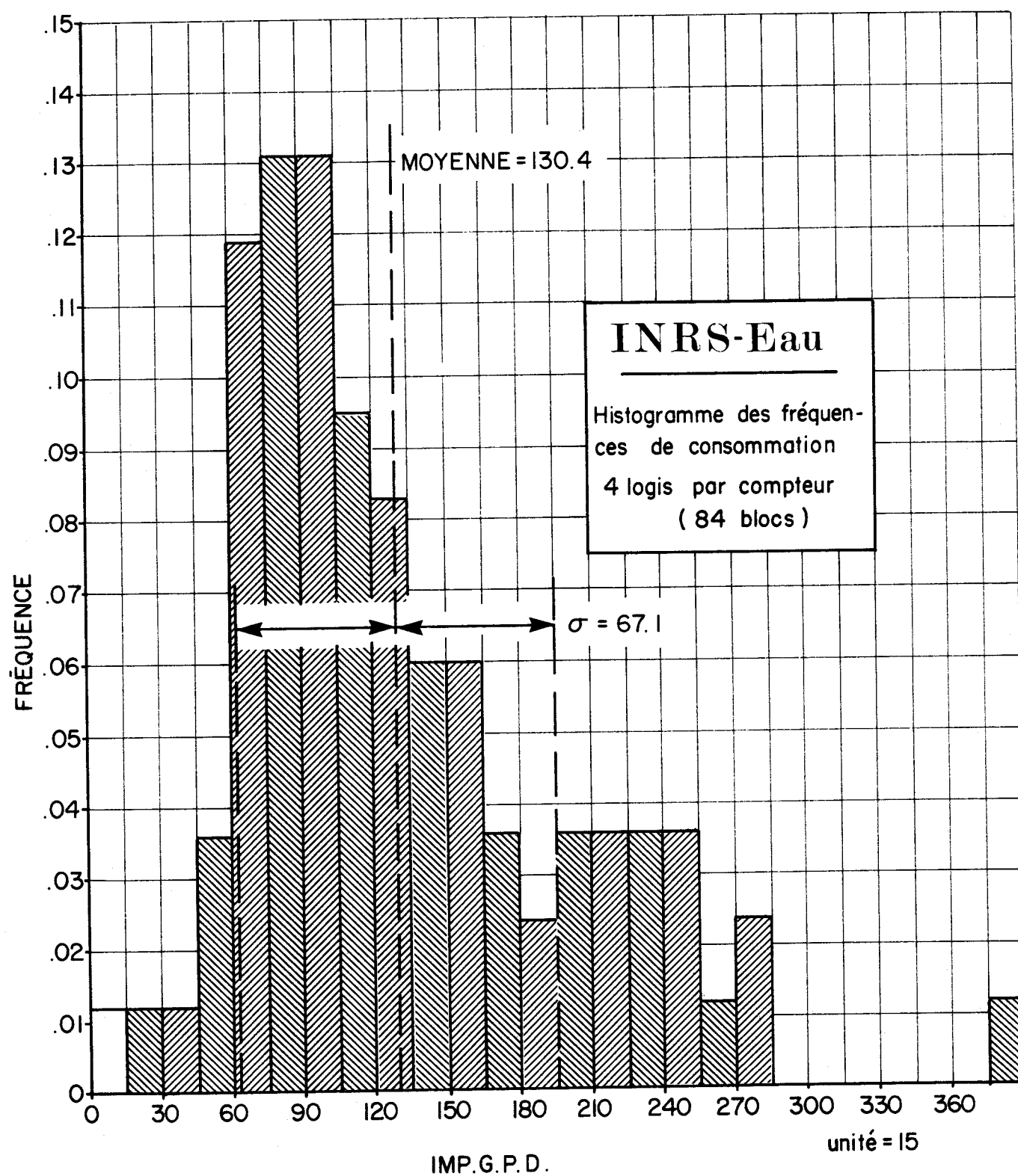


FIGURE 2-16

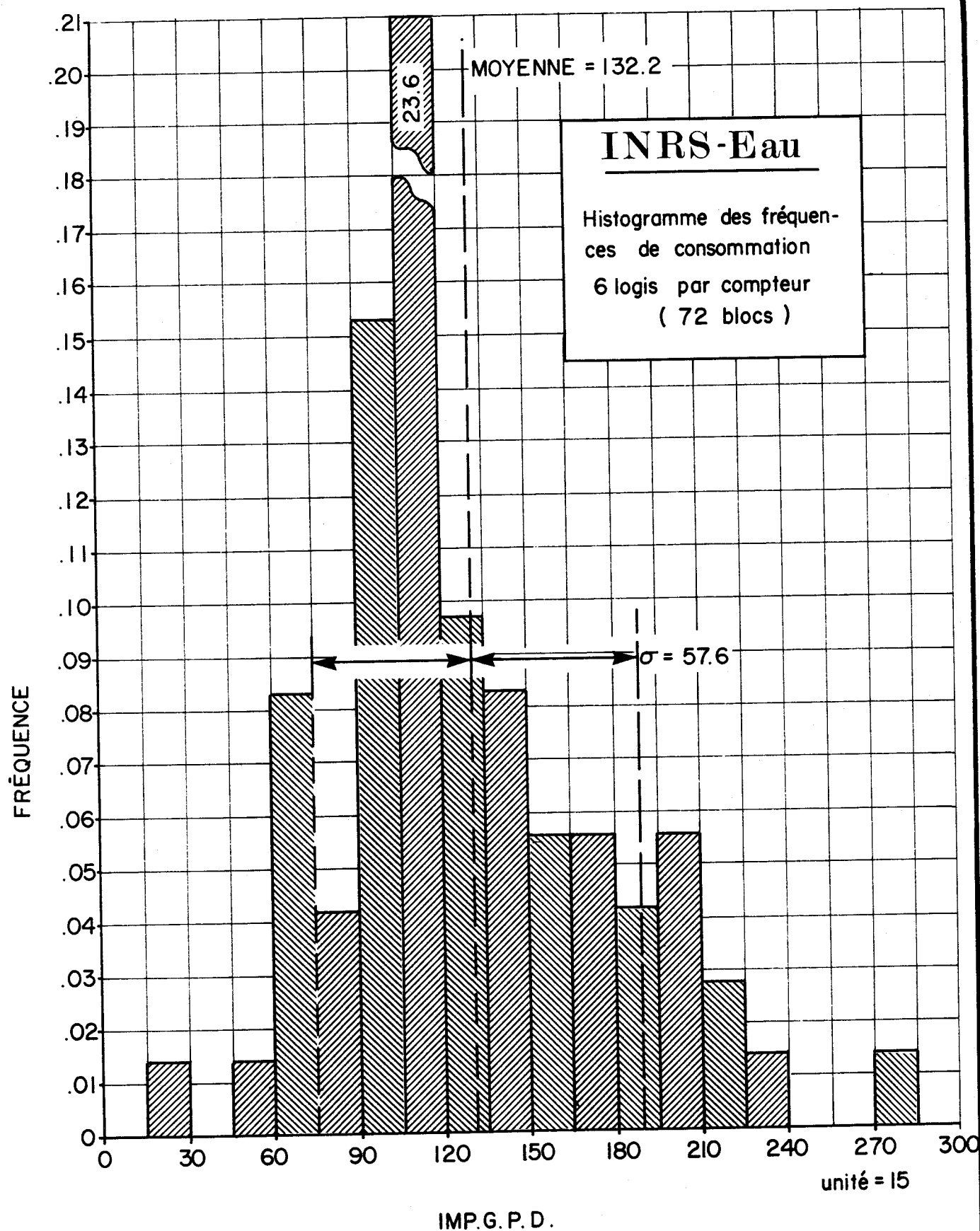


FIGURE 2-17

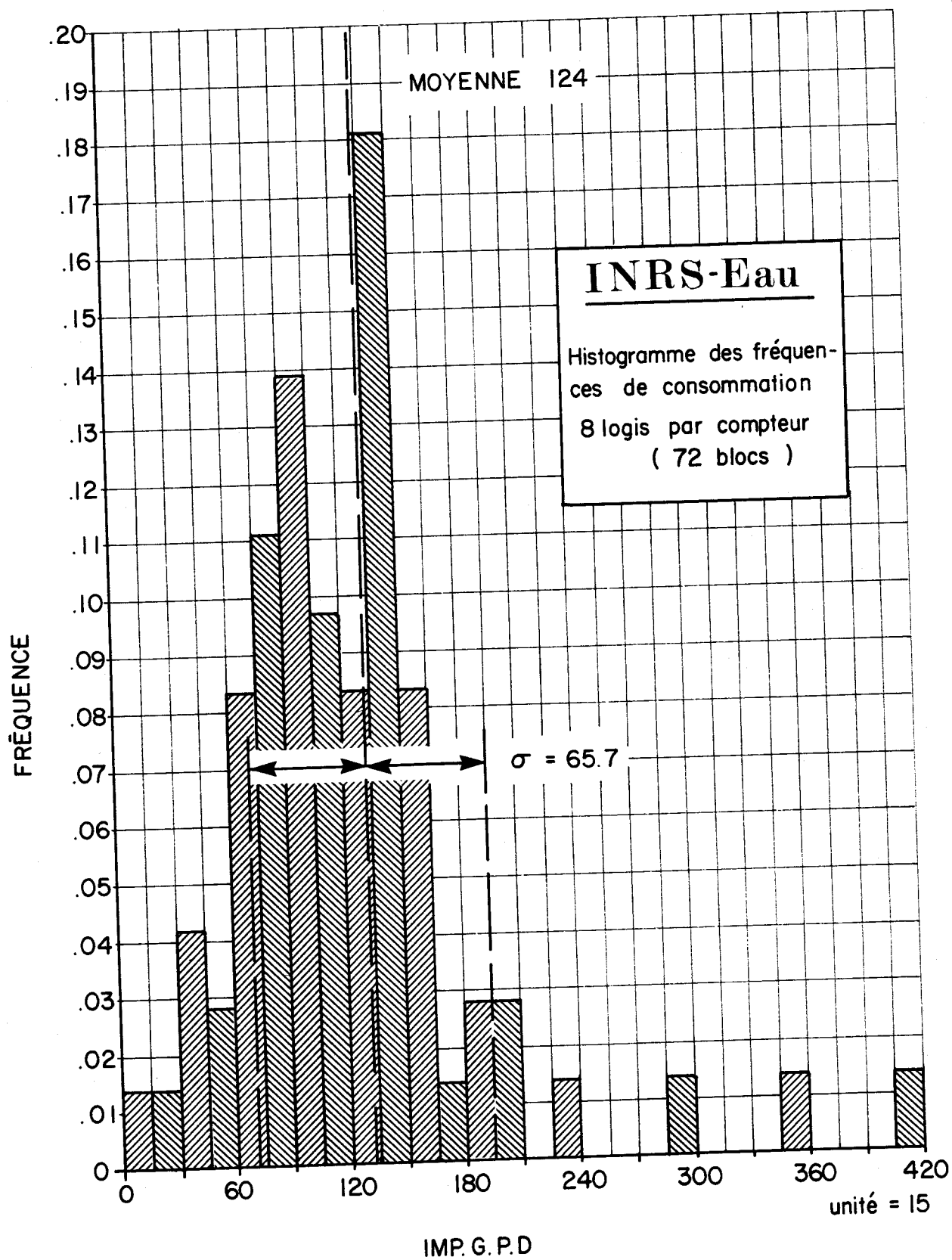


FIGURE 2-18

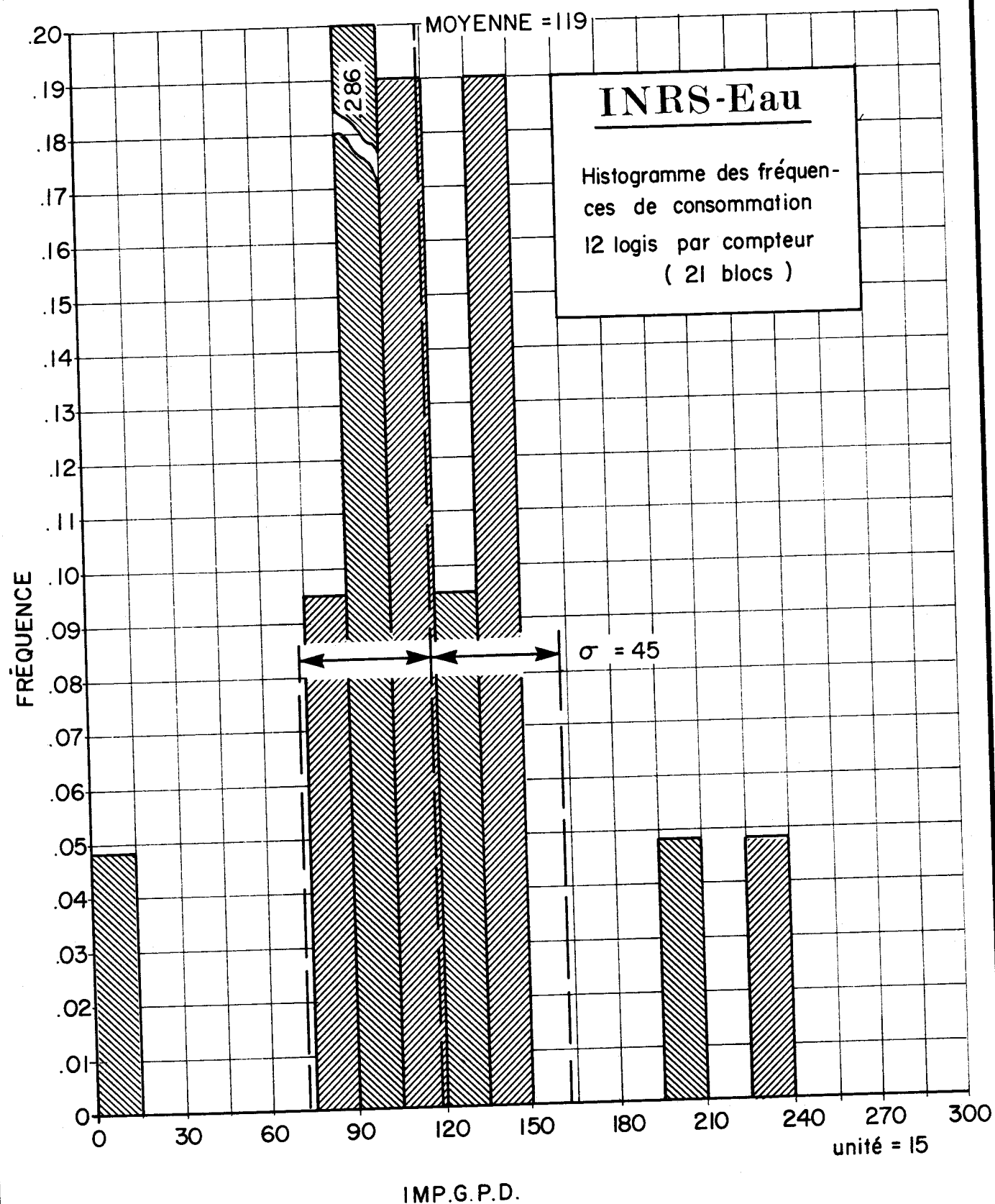


FIGURE 2-19

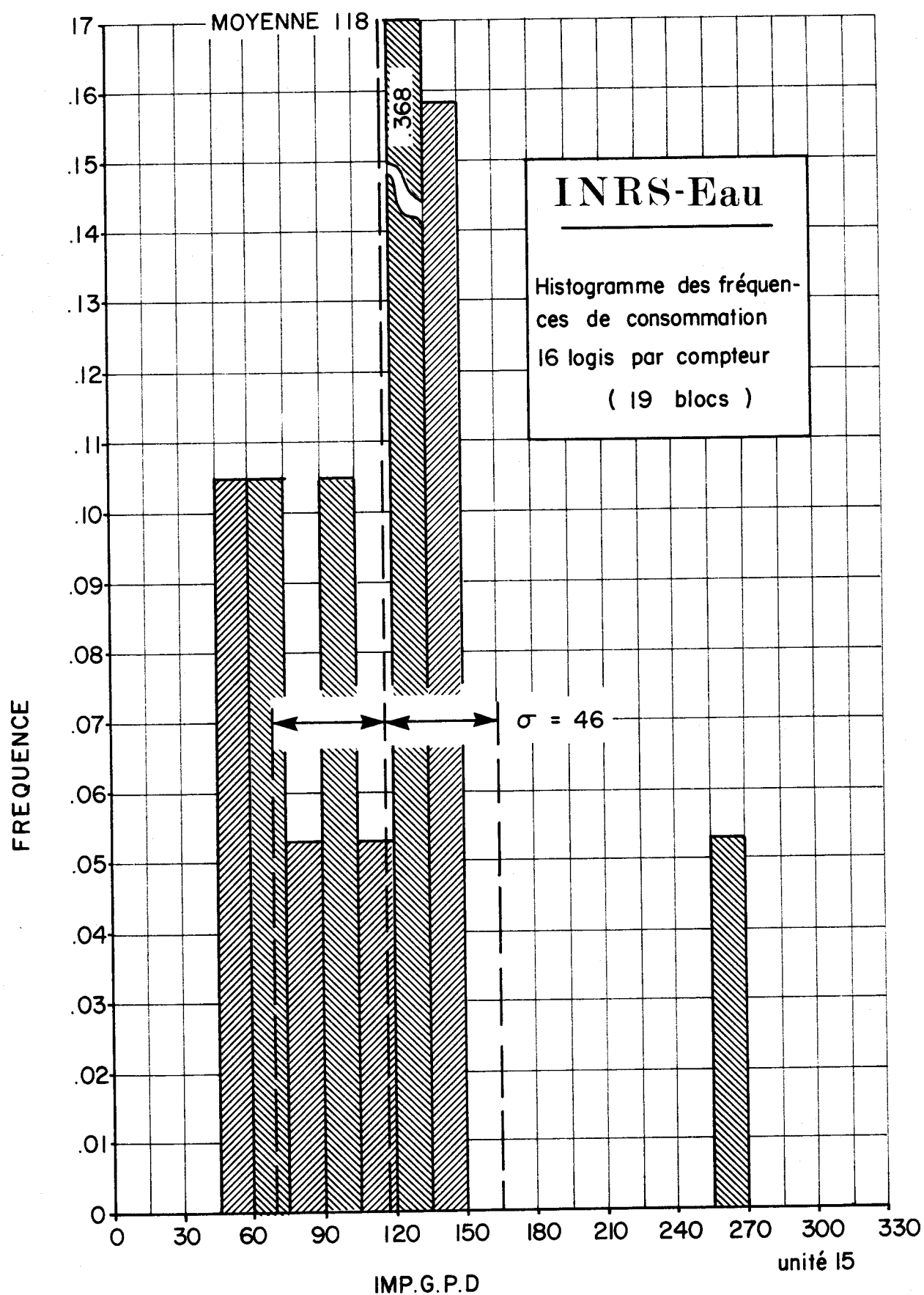


FIGURE 2-20

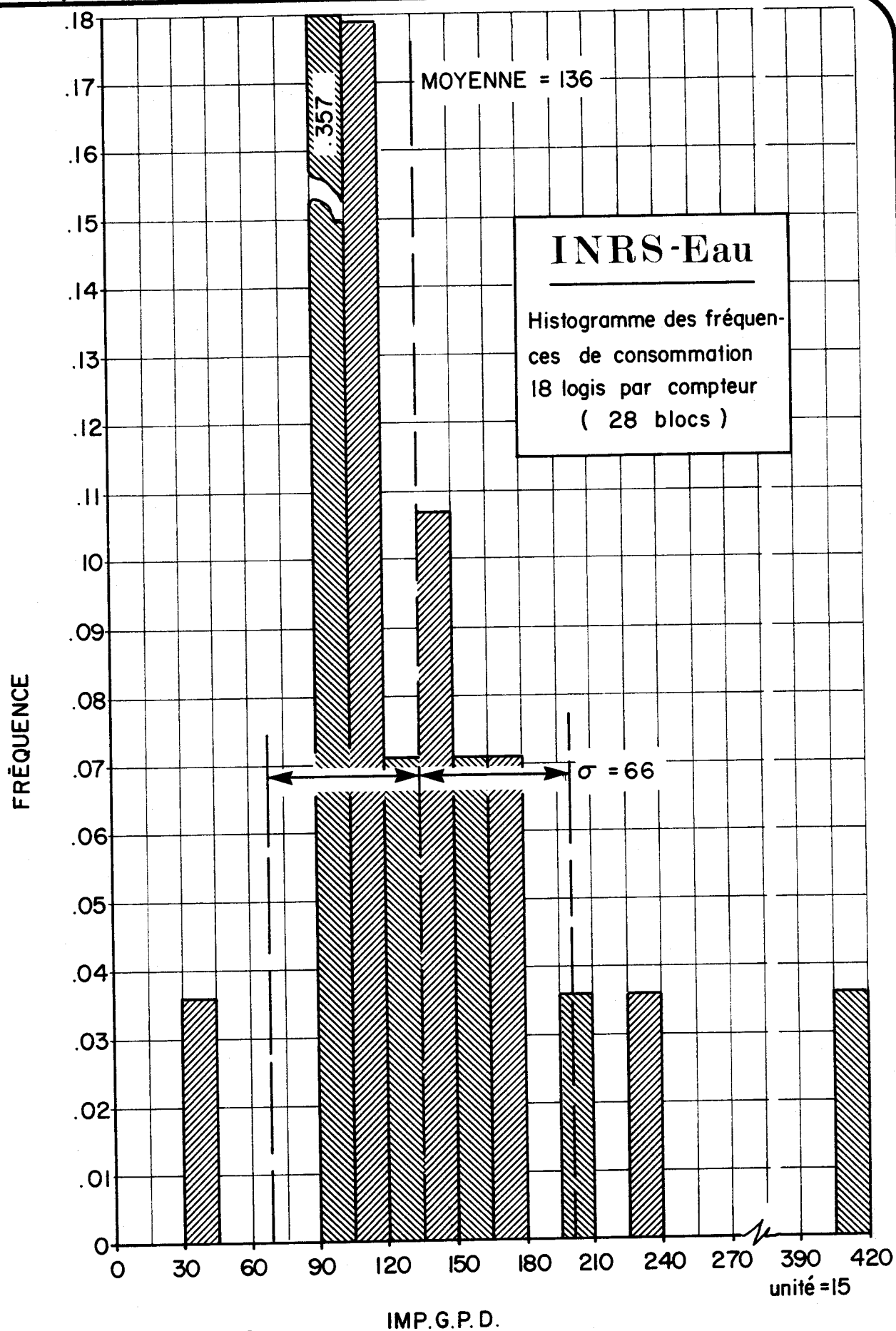


FIGURE 2-21

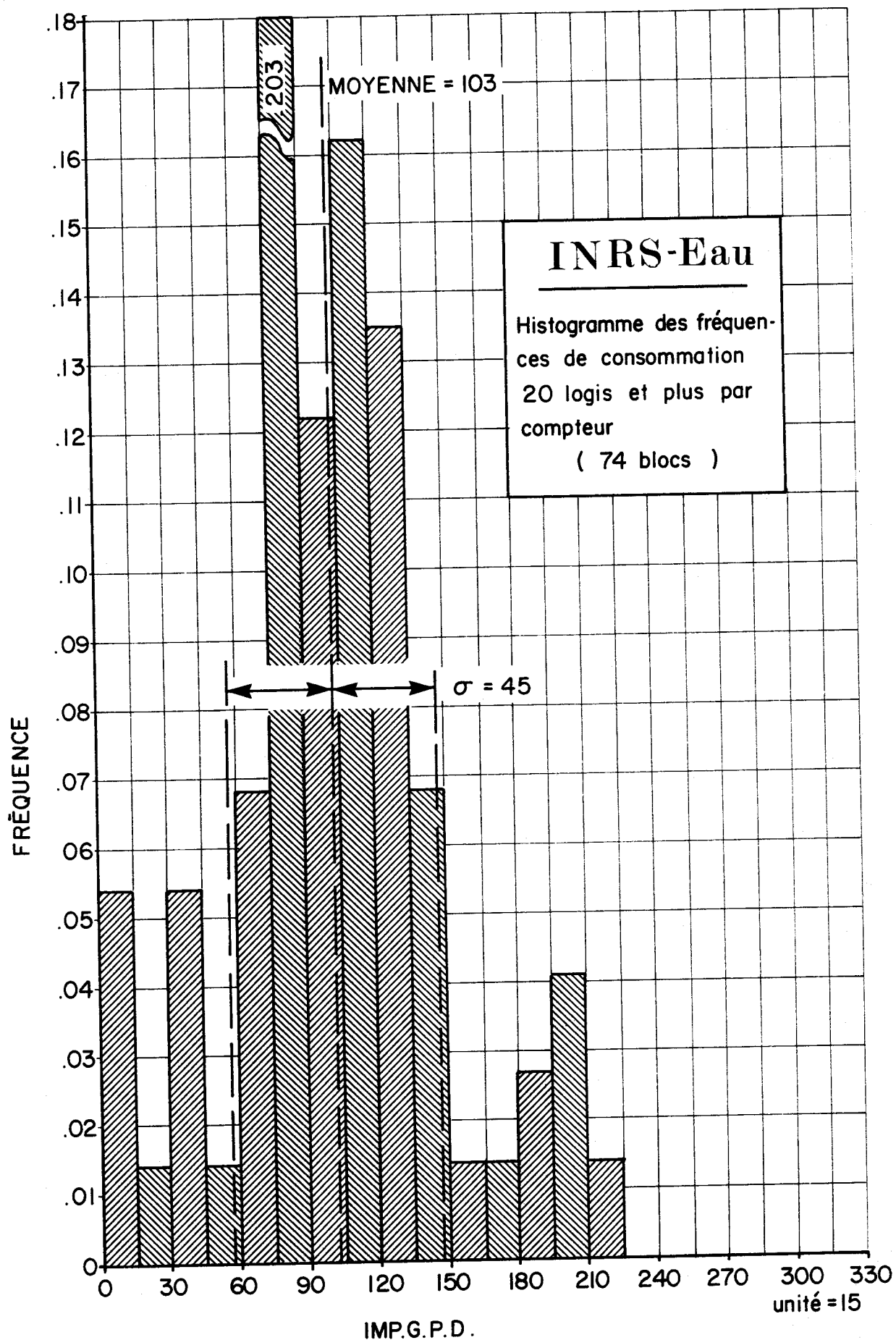


FIGURE 2-22

Classes	1	2	2*	3	4	6	8	12	16	18	> 20
1	1	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	1	1	0	0	0	1	0
2*	0	0	1	0	0	0	0	0	0	0	0
3	0	0	0	1	1	1	0	1	1	1	0
4	0	1	0	1	1	0	0	0	0	1	0
6	0	1	0	1	0	1	0	0	0	1	0
8	0	0	0	0	0	0	1	1	1	1	0
12	0	0	0	1	0	0	1	1	0	0	0
16	0	0	0	1	0	0	1	0	1	0	0
18	0	0	0	1	1	1	1	0	0	1	0
> 20	0	0	0	0	0	0	0	0	0	0	1

TABEAU 2-3: Matrice logique de comparaison des moyennes

1 moyennes significativement égales

0 moyennes significativement différentes

- les résidences bifamiliales à un compteur diffèrent sensiblement de celles à 2 compteurs et s'apparentent au groupe des classes 3, 4, 6 (130 GPD).
- les classes de 8 et plus semblent plutôt homogènes à l'exception des 18 où des vérifications devraient s'imposer.

2.4 Relations entre consommations annuelles et évaluations

La troisième étape du plan d'échantillonnage consiste en un choix des points de mesure pour une classe de blocs donnée. Ce choix doit tenir compte des variations des caractéristiques socio-économiques d'une part, pour que l'échantillon soit le plus représentatif possible et, d'autre part pour tenir compte des relations entre consommations annuelles et évaluations. Nous avons développé ce genre de relation pour les classes de blocs suivantes: résidences unifamiliales, bifamiliales à 2 compteurs, bifamiliales à un compteur, blocs de 3 appartements et plus. Cette séparation s'appuie sur la répartition des blocs par nombre d'appartements ainsi que sur la répartition d'après leurs caractéristiques socio-économiques et leurs consommations. Il aurait été possible de regrouper les résidences bifamiliales à un compteur avec les blocs de 3 appartements et plus. Cependant, le nombre relativement élevé (343) des premières aurait perturbé la relation consommation - évaluation des seconds (577 blocs); aussi avons-nous séparé ces deux classes.

Nous avons recherché des relations du type:

$$Q = A (EV - EV_m) + B$$

où Q est la consommation annuelle en GPD par appartement

EV l'évaluation en \$

EV_m l'évaluation moyenne pour la classe de blocs en \$

A et B les coefficients de la régression linéaire.

Quelques développements statistiques sur les régressions linéaires sont donnés en annexe 3. Les résultats des régressions apparaissent dans le Tableau 2-4 où figurent le coefficient de corrélation et les coefficients des régressions avec leurs limites de confiance.

Il ressort que:

- les corrélations entre consommations et évaluations sont faites pour les résidences unifamiliales, les bifamiliales à 2 compteurs et pour l'ensemble des blocs de 3 appartements et plus.
Pour les résidences bifamiliales à un seul compteur, le coefficient de corrélation est plus faible. Ceci est à rapprocher de la grande dispersion trouvée pour cette classe dans les répartitions des évaluations.
- la comparaison entre les intervalles de confiance sur les paramètres de la régression permet d'assurer que les droites de régression sont globalement différentes confirmant ainsi la nécessité de former des classes pour cette étude.

	CLASSES DE BLOCS			
	Unifamiliale	Bi-familiale 1 compteur	Bi-familiale 2 compteurs	Blocs de 3 apt. et plus
Nombre	6,437	343	185	577
Evaluation moyenne \$	22,000	9,500	12,000	8,400
Nombre de classes	18	11	8	11
A	.0056	.0081	.0109	.0080
Intervalle de con- fiance à 95%	.0047 à .0065	.0029 à .0133	.0063 à .0155	.0055 à .0105
B	236	144	168	126
Intervalle de con- fiance à 95%	227 à 245	188 à 160	152 à 183	120 à 133
Coefficient de cor- rélation R	.96	.76	.92	.92
Intervalle de con- fiance de R à 95%	.89 à .985	.297 à .934	.62 à .985	.708 à .979

TABLEAU 2-4: RESULTATS DES REGRESSIONS

Régression de la forme

$$Q = A (EV - EV_m) + B$$

Q = Consommation annuelle GPD/appartement

EV = Evaluation en \$

 EV_m = Evaluation moyenne

A et B = Coefficients de la régression linéaire

Les droites de régression pour chacune des classes sont représentées sur les figures 4.1 à 4.4 ou sont également indiqués les écarts-types pour chacune des classes d'évaluation.

Il est à remarquer que deux des hypothèses inhérentes aux calculs de régression linéaire ne sont pas strictement vérifiées:

- les écarts-types des consommations devraient être identiques pour chaque classe d'évaluation;
- les consommations dans chaque classe d'évaluation devraient suivre une loi normale.

Nous nous contentons dans cette étape des résultats mentionnés plus haut; une analyse plus complète sera faite lors de l'étape finale.

3 COMPARAISON DES RESULTATS OBTENUS A MONTREAL ET A STE-FOY

3.1 Caractéristiques socio-économiques

L'étude pour la ville de Ste-Foy a été effectuée sur l'ensemble des résidences unifamiliales et multifamiliales; nous disposons ainsi de l'information complète relative à l'évaluation et au nombre d'appartements d'un bloc donné.

A Montréal par contre, nous ne disposons que de données de consommations trimestrielles et une rapide enquête téléphonique nous a montré que ces données provenaient d'un quartier bien précis (Pointe aux Trembles) et aux caractéristiques socio-économiques homogènes. De plus, ce quartier n'était en aucune manière représentatif de l'ensemble des résidences de Montréal, et difficilement comparable aux résidences de Ste-Foy.

3.2 Consommations annuelles

En raison de l'hétérogénéité des caractéristiques socio-économiques, il est inutile de comparer les consommations annuelles pour des résidences de taille identique. En effet,

- d'une part, on a montré (ch. 2) qu'il y a une forte corrélation entre l'évaluation et la consommation annuelle moyenne pour une taille de blocs donnée.

- d'autre part, certains autres facteurs qui conditionnent la consommation (taux d'occupation, mode de vie ...) n'ont pu être pris en compte et leur influence n'a pu être appréciée.

Par contre, il est possible de comparer les regroupements par classe de moyenne de consommations voisines:

- A Ste-Foy, nous obtenons les classes suivantes:
(2, 3, 4, 6); (8, 12, 16); 18; 20 et plus.
- A Montréal, nous avons:
(2, 3); (4, 6); 8.

Compte tenu des restrictions précédentes, les résultats de Montréal n'infirmement pas ceux de Ste-Foy.

4 CHOIX DES POINTS DE MESURE

4.1 Regroupement en classe

Ce regroupement est essentiellement basé sur la comparaison des moyennes de consommations annuelles (ch. 2) pour les résidences de taille donnée. Les classes que l'on a pu identifier sont: (voir Tableau 4.1)

- Résidences unifamiliales	Classe A
- Résidences bifamiliales à un seul compteur	Classe B
- Blocs de taille 3, 4, 6	Classe C
- Blocs de taille 8, 12, 16	Classe D
- Blocs de taille 18	Classe E
- Blocs de 20 appartements et plus	Classe F

En ce qui concerne les résidences bifamiliales à 2 compteurs, leur comportement est semblable à celui des résidences unifamiliales et ce groupe est écarté de l'échantillonnage. Les résidences unifamiliales seront traitées séparément dans l'échantillonnage car la structure de la demande pour ce groupe fait l'objet d'une vérification.

TYPE	BLOCS		EVALUATION			CONSOMMATION			CLASSE	CONSOMMATION	
	NOMBRE	POURCENT	MOYENNE	ECART-TYPE	CV	MOYENNE	ECART-TYPE	CV		TAILLE	ECART-TYPE*
Unifamilial	6,600	83	16,700	5,600	.33	204	99	.48	A	6,600	99
Bifamilial 2 compteurs	202	2.5	13,300	3,800	.28	183	97	.53	**		
Bifamilial 1 compteur	442	5.6	10,300	4,700	.45	139	68	.49	B	442	68
3	139	1.8	8,150	2,300	.28	131	61	.46	C	307	62
4	92	1.2	8,160	2,900	.35	130	67	.51			
6	76	1.0	7,410	1,750	.24	132	58	.44			
8	74	.9	7,240	1,440	.20	124	66	.53	D	127	58
12	27	.3	7,750	1,330	.17	119	45	.38			
16	26	.3	6,750	1,280	.19	118	46	.39			
18	31	.4	8,028	1,040	.13	136	66	.48	E	31	66
20 et plus	109	1.3	7,610	1,870	.24	103	45	.43	F	109	45

TABLEAU 4-1

REGROUPEMENT EN CLASSES DES DIFFERENTS TYPES DE RESIDENCES

** Eliminé

* Ecart-type des classes

4.2 Répartition des points de mesure par classe

En tenant compte du coût approximatif (achat du compteur, tête de transmission, installation et location des lignes téléphoniques, traitement du signal) de chaque point de mesure dans chaque classe et en nous fixant un coût global maximum, nous avons pu, en utilisant les résultats de l'Annexe A 4, déterminer pour chaque classe le nombre de points de mesure. (La classe A est traitée séparément). Ces résultats sont récapitulés dans le tableau 4.2.

Les deux méthodes utilisées donnent sensiblement les mêmes résultats; en effet, le rapport des variances des estimations obtenues par les deux méthodes donne une efficacité de 99% à la méthode d'échantillonnage proportionnel, ce qui montre dans le cas considéré l'équivalence des deux méthodes dans le cas considéré.

La taille des échantillons pour chaque classe est finalement:

Classe A	n = 7	(Résidences unifamiliales)
Classe B	n = 14	(Résidences bifamiliales, 1 compteur)
Classe C	n = 10	(Blocs à 3, 4, 6 appartements)
Classe D	n = 4	(Blocs à 8, 12, 16 appartements)
Classe E	n = 2	(Blocs à 18 appartements)
Classe F	n = 3	(Blocs de 20 appartements et plus)

TABLEAU 4-2 ECHANTILLONNAGE

CLASSE	COUT UNITAIRE* (C_i) \$	TAILLE (N_i)	ECART TYPE (σ_i)	ALLOC. OPT. (n_i) ₁	ECH. PROP. (n_i) ₂	VALEUR RETENUE (n_i)
A**	500	6,600	99	7	7	7
B	500	442	68	14.86	14.17	14
C	500	307	62	9.4	9.84	10
D	700	127	58	4.3	4.07	4
E	700	31	66	1.2	1.	2
F	700	109	45	2.9	3.49	3

* Le coût unitaire comprend l'achat du compteur et de la tête de transmission, l'installation et la location des lignes téléphoniques, le traitement du signal.

** La classe A est traitée séparément.

4.3 Localisation des points de mesure

Après avoir déterminé le nombre de points de mesure à installer pour chaque classe, il faut maintenant répartir les points à l'intérieur de chaque classe.

Pour effectuer cette répartition, nous nous sommes basés sur les principes suivants:

- localisation des points de mesure dans les mêmes quartiers quand c'est faisable;
- Tenir compte autant que possible des singularités dans les distributions.

La méthodologie a été la suivante:

- a) Pour une classe comportant une seule taille donnée de blocs (A,B,) et dont le nombre d'éléments à choisir est supérieur à deux, nous utilisons la distribution de l'évaluation que nous divisons en intervalles équiprobables. A l'intérieur de chaque classe, nous prenons l'évaluation correspondant au point milieu de l'intervalle. Nous déterminons la consommation moyenne relative à cette évaluation par la courbe de régression évaluation - consommation et dans le fichier de l'ensemble des blocs, nous cherchons celui ou ceux dont l'évaluation et la consommation sont voisines de celles calculées.

Exemple - Classe A: Résidences unifamiliales

Nous devons choisir sept éléments dans cette classe. A l'aide de l'histogramme de la figure 2-12, nous déterminons les points milieux de chaque classe, et en utilisant la régression évaluation - consommation de la figure 4-1, nous déterminons la consommation correspondante.

Pour le premier intervalle, on ne prend pas exactement le point milieu qui est non représentatif (on se rapproche de la borne supérieure du premier intervalle).

Intervalle	1	2	3	4	5	6	7
Evaluation du point milieu \$	10,500	13,000	14,200	15,200	16,500	18,200	23,000
Consommation GPD	170	185	192	198	205	215	242

- b) Classes comportant une seule taille de blocs, dont le nombre de points de mesure est inférieur ou égal à deux (Exemple Classe E).

Dans ce cas, étant donné le petit nombre de points de mesure à déterminer, nous prenons un ou deux blocs dont l'évaluation est voisine de la moyenne.

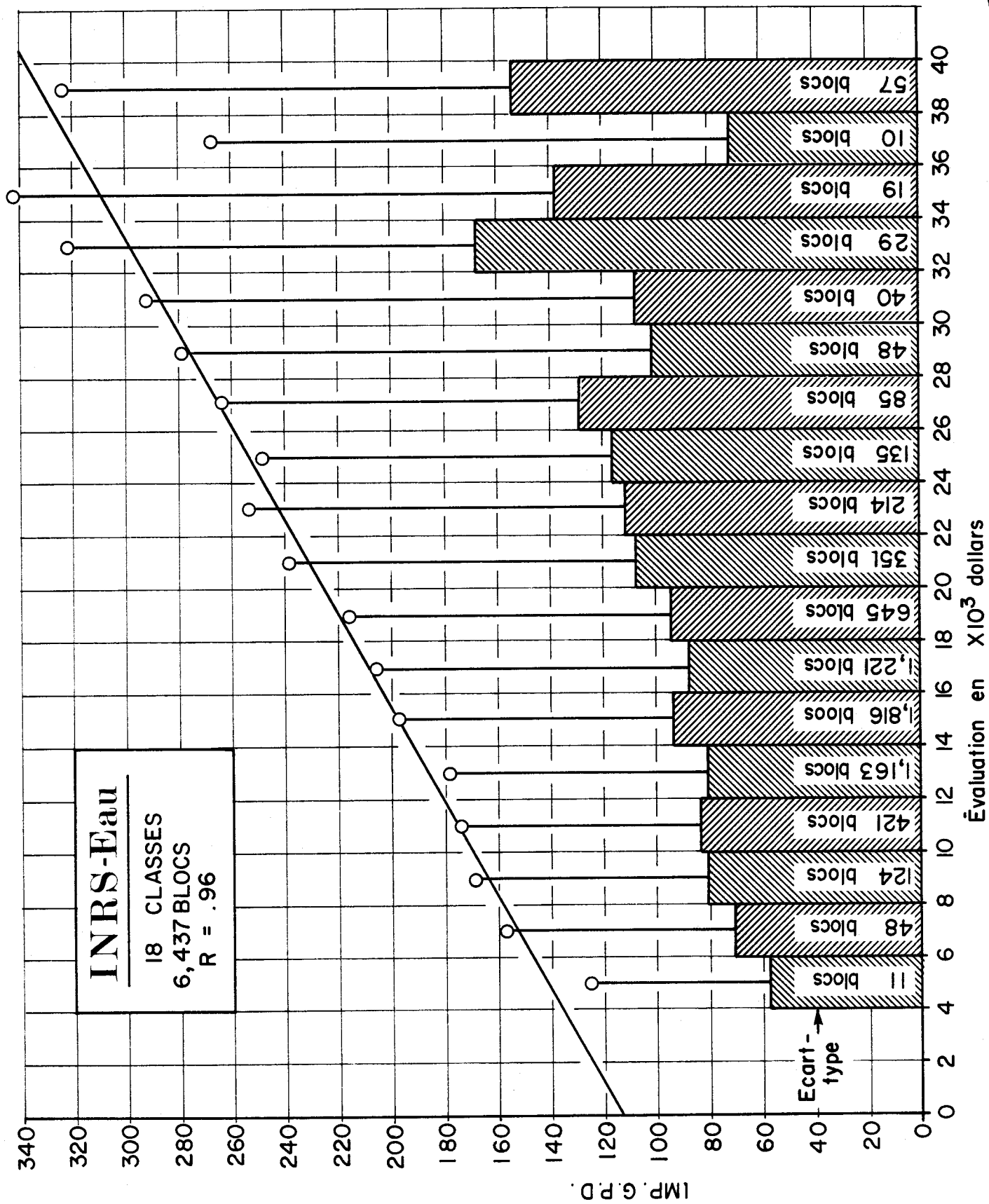


FIGURE 4-1

Regression consommation-évaluation, résidences unifamiliales

Exemple - Classe E: (18 appartements)

Nous devons choisir deux points de mesure.

L'évaluation moyenne est de \$8,030.00 (d'après l'histogramme de la figure 2-10), la consommation annuelle moyenne correspondante est 122 GPD d'après la régression de la figure 4-4.

Nous choisissons des blocs dont l'évaluation est proche de \$8,030.00 et la consommation aux alentours de 122 GPD.

c) Classe comportant plusieurs tailles de blocs (classes C, D, F).

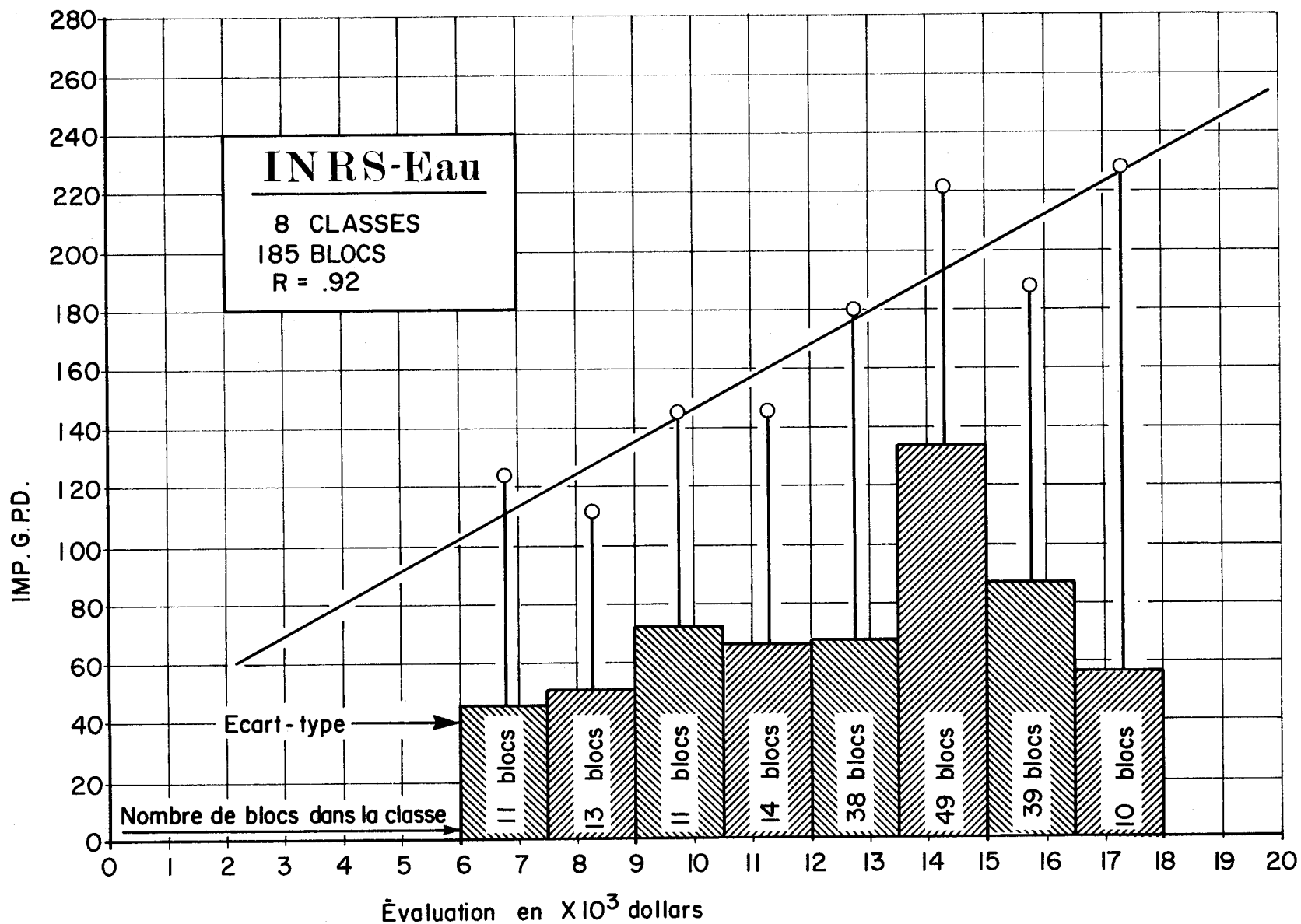
On choisit une taille donnée que l'on étudie plus précisément et on répartit les autres points de mesure sur les tailles restantes.

Exemple - Classe C: (3, 4, 6 appartements)

Nous devons choisir dix points de mesure que l'on répartit comme suit:

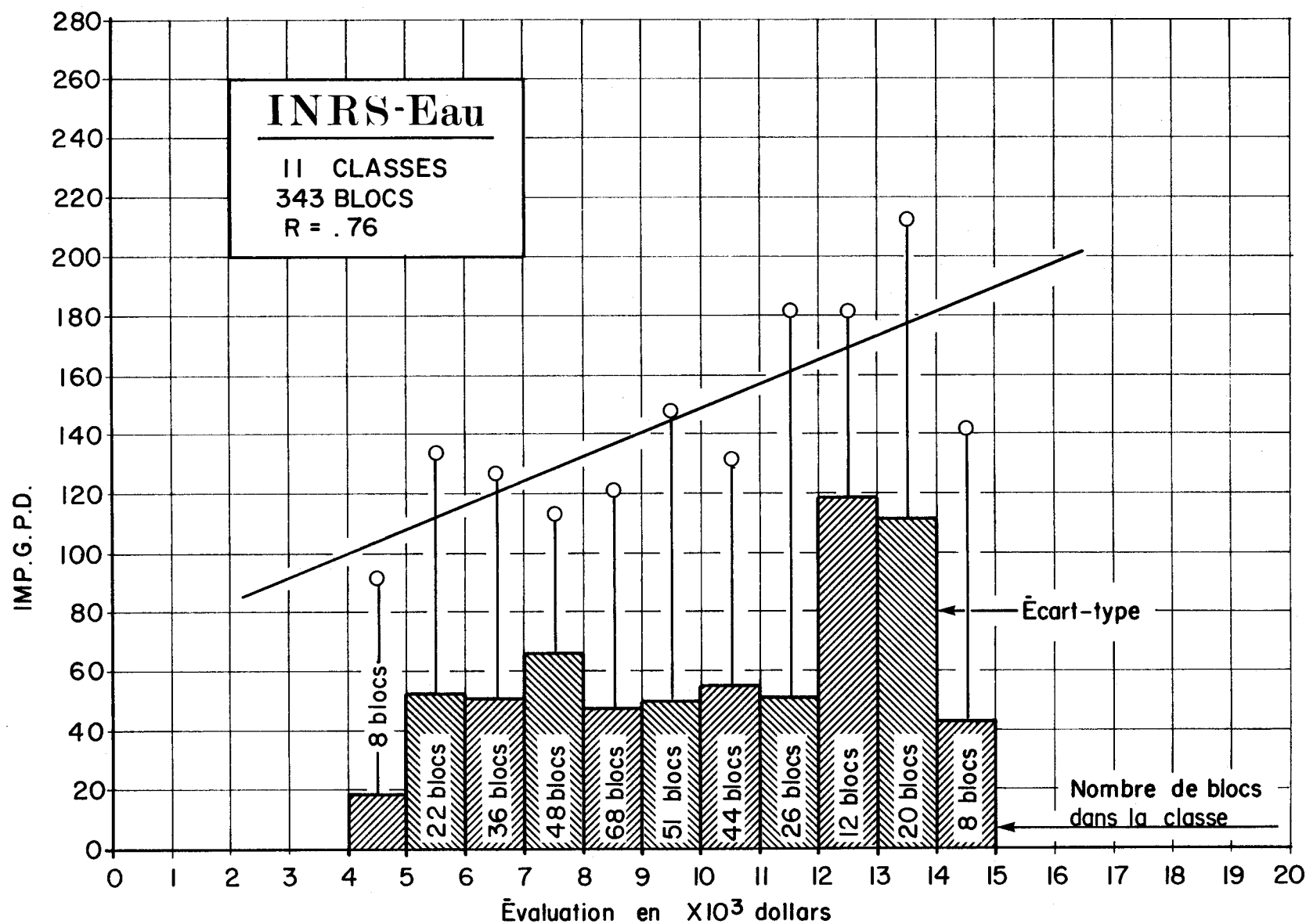
On met six points de mesure sur les 6 appartements, car pour cette taille, les histogrammes de consommation et d'évaluation montrent une moins grande dispersion. Ceci nous permettra une connaissance approfondie des variations de la structure autour de la moyenne. La répartition pour cette taille est effectuée suivant les mêmes principes qu'en a).

FIGURE 4-2



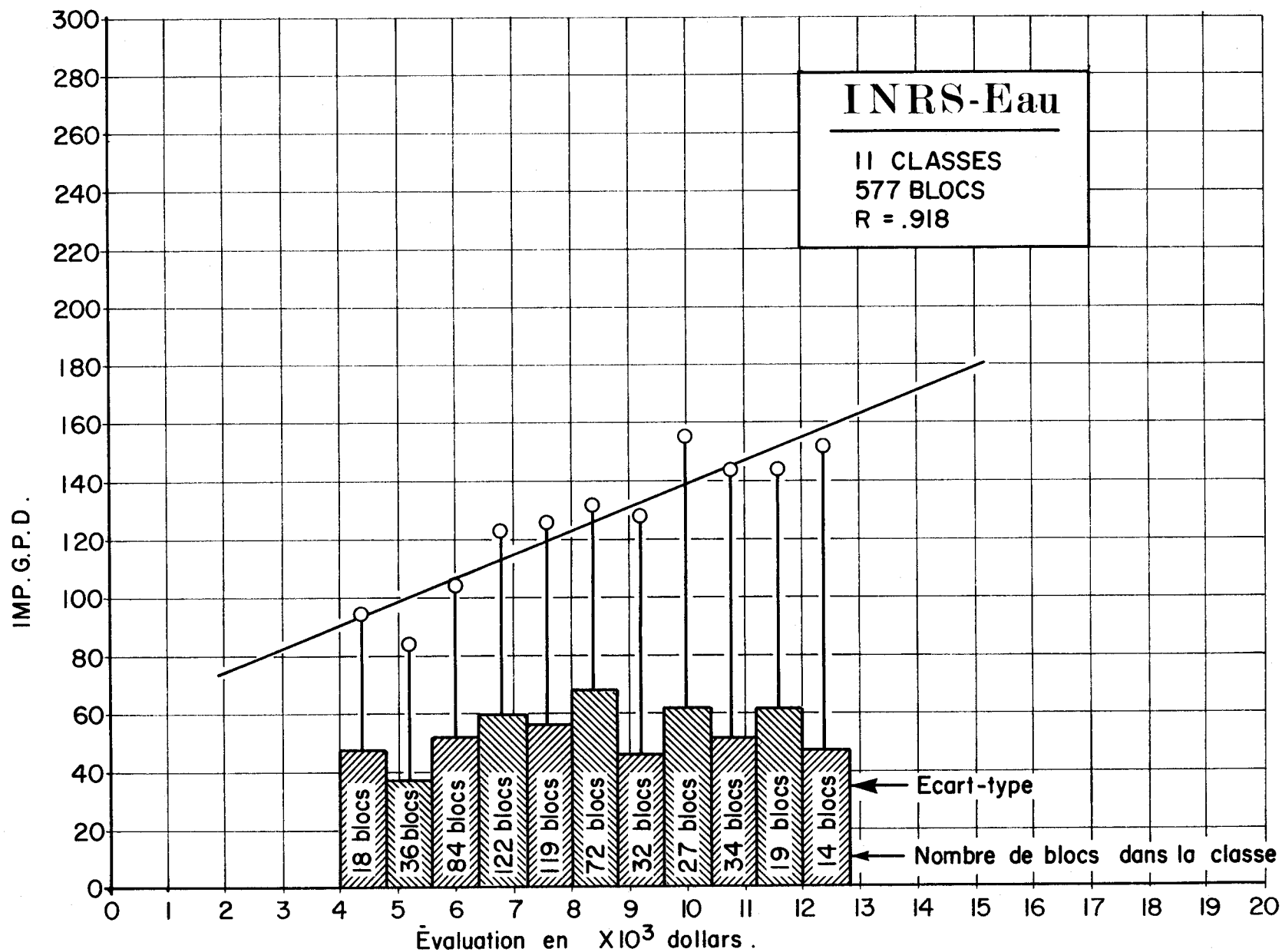
Regression consommation-évaluation bifamiliale à deux compteurs.

FIGURE 4-3



Regression consommation-évaluation bifamiliale à un compteur .

FIGURE 4-4



Regression consommation-évaluation blocs de 3 appartements et plus.

Pour vérifier le comportement des 3 et 4 appartements, nous installons deux points de mesure pour les 3 et les 4 appartements suivant les mêmes principes qu'en b) en utilisant les figures 2-5 et 2-6. L'histogramme de l'évaluation des blocs de 4 appartements (figure 2-5) présentant deux modes à \$5,000 et \$10,000, nous avons choisi les deux points de mesure autour de ces valeurs.

On obtient le tableau récapitulatif suivant:

Taille des blocs	6 app.						3 app.		4 app.	
Points de mesure	1	2	3	4	5	6	1	2	1	2
Evaluation \$	5,250	6,400	7,000	7,600	8,250	10,000	8,150	8,150	5,500	10,500
Consommation GPD	102	110	115	120	125	139	125	125	103	143

Les autres classes sont échantillonnées avec les mêmes méthodes.

Exemple - Classe B: (Bifamilial à un compteur)

Dans cette classe, on peut distinguer deux catégories:

- les résidences unifamiliales avec sous-sol aménagé en appartement; leur nombre total est de 60 (catégorie 2-SS)

- les résidences bifamiliales à deux appartements superposés; leur nombre est de 304 (catégorie 2).

Il y a 80 résidences pour lesquelles la classification est impossible, faute d'informations. Nous n'avons pas tenu compte de ces résidences dans notre échantillonnage.

Il y a 14 points de mesure à implanter dans cette classe que nous avons répartie proportionnellement à la taille comme suit:

- 3 pour la catégorie des 2-SS
- 11 pour la catégorie des 2.

Catégorie 2-SS

Point de mesure	1	2	3
Evaluation \$	8,000	9,000	10,000
Consommation GPD	100	110	120

Catégorie 2

Point de mesure	1	2	3	4	5	6	7	8	9	10	11
Evaluation \$	4,700	6,000	7,000	7,650	8,200	8,800	9,500	10,500	12,400	13,500	17,000
Consommation GPD	105	110	125	130	135	140	145	152	165	175	200

Exemple - Classe D: 8, 12, 16 appartements

Nous avons choisi:

- 2 points de mesure pour les 8 appartements
- 1 point de mesure pour les 12 appartements
- 1 point de mesure pour les 16 appartements.

En effet, les 12 et les 16 appartements ont le même comportement en ce qui concerne les variations de consommation annuelle et d'évaluation.

Nombre d'appartements	8	12	16
Evaluation \$	6,500	8,000	7,750
Consommation GPD	110	122	121

Exemple - Classe F:

Les blocs de 20 appartements et plus (figure 2-11) constituent une classe dont le nombre d'appartements est très dispersé:

- 20 à 30 appartements: 50% des blocs
- 30 à 40 appartements: 20% des blocs
- plus de 40 appartements: 30% des blocs.

Nous nous attachons à :

- a) tenir compte de cette répartition;
- b) choisir les blocs dans un secteur isolable dans
le but d'étudier la consommation du secteur.

C'est ainsi que nous avons choisi un bloc de 24 appartements
(les 24 appartements représentent 20% des blocs de 20 à 30),
un bloc de 32 appartements, un bloc de 41 appartements.

ANNEXE A1TESTS STATISTIQUESA1.1 Test de Student (t): Comparaison de 2 moyennes

On dispose de 2 échantillons de taille N_1 et N_2 dont les observations sont:

$$y_{1i} \quad i = 1, \dots, N_1$$

$$y_{2i} \quad i = 1, \dots, N_2$$

On veut comparer les moyennes des populations μ_1 et μ_2 dont les échantillons sont tirés.

Si N_1 et N_2 sont élevés ($N_1, N_2 > 30$) les populations sont supposées quelconques, si N_1 et N_2 sont petits les populations doivent être supposées normales.

Dans ce test les variances des populations sont inconnues, deux cas pratiques sont à considérer.

A1.1.1 Variances des populations égales

Soient: $\bar{y}_1$ et $\bar{y}_2$ les moyennes calculées sur les échantillons

s^2 la variance combinée telle que

$$s^2 = \frac{\sum_{i=1}^{N_1} (y_{1i} - \bar{y}_1)^2 + \sum_{i=1}^{N_2} (y_{2i} - \bar{y}_2)^2}{N_1 + N_2 - 2}$$

La variable
$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{s \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

suit une loi de Student à $V = N_1 + N_2 - 2$ degrés de liberté.

Il est alors possible de tester l'hypothèse nulle H_0 telle que:

$$H_0: \mu_1 - \mu_2 = 0 \quad (\text{ou } \mu_1 = \mu_2)$$

On détermine pour le niveau de signification α désiré, la valeur $t_{\alpha, v}$

si $|t| < t_{\alpha/2, v}$ H_0 est acceptée et $\mu_1 = \mu_2$

si $|t| > t_{\alpha/2, v}$ H_0 est refusée et $\mu_1 \neq \mu_2$

A1.1.2 Variances des populations inégales

Les variances non biaisées des échantillons sont:

$$s_1^2 = \frac{\sum_{i=1}^{N_1} (y_{1i} - \bar{y}_1)^2}{N_1 - 1} \quad s_2^2 = \frac{\sum_{i=1}^{N_2} (y_{2i} - \bar{y}_2)^2}{N_2 - 1}$$

Dans ce cas

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

soit une variable de Student à ν degrés de liberté avec

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)^2 (N_1 - 1) (N_2 - 1)}{\left(\frac{s_1^2}{N_1} \right)^2 (N_2 - 1) + \left(\frac{s_2^2}{N_2} \right)^2 (N_1 - 1)}$$

A1.2 Test d'égalité de 2 variances (Fisher)

Soient deux échantillons de taille N_1 et N_2 que l'on suppose tirés de 2 populations normales, on veut tester l'égalité des variances σ_1^2 et σ_2^2 des 2 populations.

Si s_1^2 et s_2^2 sont les variances non biaisées des 2 échantillons
 la variable $F = \frac{s_1^2}{s_2^2}$ suit une loi de Fisher à $v_1 = (N_1 - 1)$
 et $v_2 = (N_2 - 1)$ degrés de liberté

L'hypothèse H_0 à considérer est:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad \left(\text{ou} \quad \frac{\sigma_1^2}{\sigma_2^2} = 1 \right)$$

Pour un niveau de signification donné α , on détermine la zone
 d'acceptation de H_0 :

$$F_1 < F < F_2 \quad \text{avec} \quad F_2 = F_{v_1, v_2}(\alpha/2)$$

$$F_1 = F_{v_1, v_2}(1 - \alpha/2) = \frac{1}{F_{v_2, v_1}(\alpha/2)}$$

- . Si $F_1 < F < F_2$ on accepte H_0 et les variances des 2 populations sont égales.
- . Si $F < F_1$ ou $F > F_2$ on rejette H_0 et les variances des 2 populations sont différentes.

Ce test peut être utilisé comme préliminaire du test de Student
 pour savoir si l'on se trouve dans le cas A1.1.1 ou A1.1.2.

A1.3 Test d'égalité de p variances (Bartlett)

Ce test est une généralisation du test précédent, qui permet de comparer p variances ($p \geq 2$).

On considère p échantillons de taille $N_1, N_2 \dots N_p$, tirés de populations normales.

Soient: $\sigma_1^2, \dots, \sigma_p^2$ les variances des populations,
 $s_1^2, \dots, s_p^2$ les variances non biaisées des échantillons.

On veut tester l'hypothèse H_0 :

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_p^2 = \sigma^2$$

Si H_0 est vraie les échantillons sont tirés de populations normales dont les variances sont égales.

La variance combinée de ces échantillons est:

$$s^2 = \frac{\sum_{i=1}^p (N_i - 1) s_i^2}{\left(\sum_{i=1}^p N_i \right) - p}$$

La quantité K telle que

$$K = -\frac{1}{C} \sum_{i=1}^p N_i \operatorname{Log} \left(\frac{s_i^2}{s^2} \right)$$

$$\text{avec } C = 1 + \frac{1}{3(p-1)} \left[\sum_{i=1}^p \left(\frac{1}{N_i} \right) - \frac{1}{\sum_{i=1}^p N_i} \right]$$

Suit une distribution χ^2 (chi-deux) avec $\nu = (p-1)$ degrés de liberté.

Pour un niveau de signification α donné, il est donc possible de déterminer la zone d'acceptation.

Si $0 < K < \chi_{\alpha, \nu}^2$ l'hypothèse H_0 est acceptée et les variances des populations sont égales $[\chi_{\alpha, \nu}^2 \text{ est tel que } \operatorname{Prob} (\chi^2 > \chi_{\alpha, \nu}^2) = \alpha]$

ANNEXE 2
PROGRAMMES

- A2-1 Caractéristiques socio-économiques de la ville de Ste-Foy
- A2-1-1 Création d'une bande contenant les consommations et les évalua-
tions pour chaque numéro de rôle
Programme TØTAL

Le service d'informatique de la ville de Ste-Foy nous a donné deux bandes:

- bande des évaluations des bâtisses et terrains pour chaque numéro de rôle
- bande des consommations et date de relevé pour chaque numéro de rôle.

La première bande contient aussi tous les sous-rôles (01, 02, etc.) dans le cas où plusieurs personnes sont propriétaires de la même parcelle.

A partir de ces deux fichiers, une troisième bande a été créée, regroupant sous le même numéro de rôle la consommation, les nombres d'appartements tels qu'indiqués sur les deux bandes, les évaluations et superficie des terrains, ainsi que l'adresse du rôle.

Au cours de ce processus, nous avons:

- éliminé les sous-rôles, qui ne contenaient aucune information
- ramené la consommation à une valeur moyenne annuelle (en se basant sur les mois des dates des relevés)
- identifié les rôles qui avaient des nombres d'appartements différents sur les deux bandes.

Ainsi, à partir des 10,367 rôles et sous-rôles inscrits sur la bande des évaluations et des 7,990 rôles inscrits sur celle des consommations, nous avons créé une bande récapitulative de 7,980 rôles qui a servi à identifier les diverses caractéristiques socio-économiques des résidences de Ste-Foy.

A2-1-2 Etude des caractéristiques socio-économiques
Programme FREQUEN

A partir du nouveau fichier, nous avons pu obtenir les fréquences des consommations et des évaluations et faire les régressions entre ces deux paramètres.

Le programme ci-joint, donné comme exemple, calcule les moyennes et écarts-types des consommations pour chaque type de bloc. De nombreux programmes semblables ont été utilisés. Dans tous ces programmes se trouvent des tests qui éliminent:

- les rôles possédant des nombres d'appartements nuls
- les rôles ayant une différence dans les nombres d'appartements de plus de deux
- les rôles ayant une consommation totale de moins de 10 GPD
- les appartements ayant une consommation supérieure à 1,000 GPD
- une consommation anormalement grande (600 GPD) pour un bloc de 5 appartements
- une évaluation anormalement grande (622,000\$) pour un bloc de 16 appartements
- les rôles où la date de relevé des compteurs était absente
- les rôles ayant une évaluation de la bâtisse inexistante.

A2-2 Analyse des relevés trimestriels de la ville de Montréal

A2-2-1 Programme MØYAN

Ce programme a pour but de générer un ruban magnétique contenant:

- les moyennes annuelles (gallons/jour) de consommations pour chaque compteur ainsi que le nombre de jours pour lesquels chaque moyenne est calculée.

Les calculs sont effectués à partir de rubans composés des relevés trimestriels de la ville de Montréal. Dans ces calculs, on tient compte des cas particuliers de données engendrés par:

- i) des pannes de compteurs
- ii) l'installation de nouveaux compteurs

Le ruban engendré par ce programme est utilisé dans les calculs d'analyse des consommations.

A2-2-2 Programme MØYT

Ce programme évalue, à partir des données sur le ruban généré par le programme MØYAN:

- i) les moyennes annuelles (gallons/jour) pour chaque classe de compteurs (établie selon le nombre de logis desservis par un compteur)
- ii) l'écart-type, le maximum, le minimum, la taille de l'échantillon (i.e. le nombre de compteurs dans cette classe) correspondant à chaque année
- iii) la moyenne interannuelle (gallons/jour) pour chaque classe de compteurs, avec l'écart-type, le maximum et le minimum correspondant.

On a systématiquement éliminé des calculs la première année de relevés.

A2-2-3 Programme MØIT1

Ce programme exécute les mêmes calculs que le programme MØYT en restreignant les données aux compteurs de la classe 1 (i.e. ceux desservant un seul logis). On subdivise ces données en nouvelles classes qui tiennent compte du type de bloc (i.e. du nombre de logis dans ce bloc) auquel appartient le logis.

A2-2-4 Programme MØYSØN

A partir des données des rubans composés des relevés trimestriels de consommation d'eau des unités multifamiliales de Pointe aux Trembles et Montréal-Est, ce programme étudie:

- i) les moyennes saisonnières (gallons/jour) pour chaque classe de compteurs (établie selon le nombre de logis desservis par un compteur). L'été est défini comme la période s'étalant du 1er juin au 1er septembre. L'hiver correspond à ce qui reste dans l'année. Les relevés qui chevauchent sur deux saisons sont éliminés du calcul.
- ii) les moyennes totales ainsi que l'écart-type correspondant (compilés sur l'ensemble des moyennes saisonnières de chaque classe de compteurs
- iii) la moyenne et l'écart-type de l'ensemble total des moyennes saisonnières (toutes classes de compteurs).

A2-2-5 Programme TT (APL/360)

Ce programme a pour but de tester statistiquement l'égalité des moyennes entre elles et les écarts-types correspondants. Les valeurs sont entrées sous forme de vecteurs et le test s'effectue sur toutes les combinaisons possibles de chacun des vecteurs.

L'égalité des moyennes est testée à l'aide du test de Student (Himmelblau D.M., 1970) qui pose comme hypothèse que:

- i) les échantillons sont tirés de populations distribuées normalement
- ii) les écarts-types des échantillons sont inconnus et différents.

L'égalité des variances est testée à l'aide du test de Fisher (Himmelblau D.M., 1970) qui pose comme hypothèse que les échantillons sont tirés de populations distribuées normalement.

```

1 PROGRAM TOTAL (INPUT, OUTPUT, TAPE1, TAPE2, TAPE3)
C CE PROGRAM AFICHE UN SEUL FICHIER AVEC CONSONNATION ET EVALUATION.
1 FORMAT(1X,15)
11 FOR AT( * ERREUR DE SEQUENCE SUR EVALUATION* 418)
12 FOR AT(8X,211,46,33X,13)
13 FORMAT(17,17,8X,1F,8X,15,46X,17,1F)
14 FORMAT(1X,15,4,12,2X,212,8X,212,6X,18)
15 FOR AT( * ERREUR DE SEQUENCE SUR CONSONNATION* 418)
16 FOR AT( * X DE LOGEMENTS DIFFERENT*215* ,POUR LE ROLE* 17)
17 FORMAT(// * X DE LOGEMENTS DIFFERENT SUR LA BANDE* 17//
1 * X DE LOGEMENTS SUR EVALUATION* 17//
2 * X DE LOGEMENTS SUR CONSONNATION* 17)
18 FORMAT(1X,11,2X,17,15)
19 TITRE = '1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,182,183,184,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,208,209,210,211,212,213,214,215,216,217,218,219,220,221,222,223,224,225,226,227,228,229,230,231,232,233,234,235,236,237,238,239,240,241,242,243,244,245,246,247,248,249,250,251,252,253,254,255,256,257,258,259,260,261,262,263,264,265,266,267,268,269,270,271,272,273,274,275,276,277,278,279,280,281,282,283,284,285,286,287,288,289,290,291,292,293,294,295,296,297,298,299,300,301,302,303,304,305,306,307,308,309,310,311,312,313,314,315,316,317,318,319,320,321,322,323,324,325,326,327,328,329,330,331,332,333,334,335,336,337,338,339,340,341,342,343,344,345,346,347,348,349,350,351,352,353,354,355,356,357,358,359,360,361,362,363,364,365,366,367,368,369,370,371,372,373,374,375,376,377,378,379,380,381,382,383,384,385,386,387,388,389,390,391,392,393,394,395,396,397,398,399,400,401,402,403,404,405,406,407,408,409,410,411,412,413,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428,429,430,431,432,433,434,435,436,437,438,439,440,441,442,443,444,445,446,447,448,449,450,451,452,453,454,455,456,457,458,459,460,461,462,463,464,465,466,467,468,469,470,471,472,473,474,475,476,477,478,479,480,481,482,483,484,485,486,487,488,489,490,491,492,493,494,495,496,497,498,499,500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,515,516,517,518,519,520,521,522,523,524,525,526,527,528,529,530,531,532,533,534,535,536,537,538,539,540,541,542,543,544,545,546,547,548,549,550,551,552,553,554,555,556,557,558,559,560,561,562,563,564,565,566,567,568,569,570,571,572,573,574,575,576,577,578,579,580,581,582,583,584,585,586,587,588,589,590,591,592,593,594,595,596,597,598,599,600,601,602,603,604,605,606,607,608,609,610,611,612,613,614,615,616,617,618,619,620,621,622,623,624,625,626,627,628,629,630,631,632,633,634,635,636,637,638,639,640,641,642,643,644,645,646,647,648,649,650,651,652,653,654,655,656,657,658,659,660,661,662,663,664,665,666,667,668,669,670,671,672,673,674,675,676,677,678,679,680,681,682,683,684,685,686,687,688,689,690,691,692,693,694,695,696,697,698,699,700,701,702,703,704,705,706,707,708,709,710,711,712,713,714,715,716,717,718,719,720,721,722,723,724,725,726,727,728,729,730,731,732,733,734,735,736,737,738,739,740,741,742,743,744,745,746,747,748,749,750,751,752,753,754,755,756,757,758,759,760,761,762,763,764,765,766,767,768,769,770,771,772,773,774,775,776,777,778,779,780,781,782,783,784,785,786,787,788,789,790,791,792,793,794,795,796,797,798,799,800,801,802,803,804,805,806,807,808,809,810,811,812,813,814,815,816,817,818,819,820,821,822,823,824,825,826,827,828,829,830,831,832,833,834,835,836,837,838,839,840,841,842,843,844,845,846,847,848,849,850,851,852,853,854,855,856,857,858,859,860,861,862,863,864,865,866,867,868,869,870,871,872,873,874,875,876,877,878,879,880,881,882,883,884,885,886,887,888,889,890,891,892,893,894,895,896,897,898,899,900,901,902,903,904,905,906,907,908,909,910,911,912,913,914,915,916,917,918,919,920,921,922,923,924,925,926,927,928,929,930,931,932,933,934,935,936,937,938,939,940,941,942,943,944,945,946,947,948,949,950,951,952,953,954,955,956,957,958,959,960,961,962,963,964,965,966,967,968,969,970,971,972,973,
```

```

30 DECODE(65,12,A(16)) ADRES,NBAPT
   DECODE(103,13,A(32)) S1,S2,S3,IMP1,IMP2
   COUNT=COUNT+1
   IF(APT.NE.NBAPT)PRINT 16,APT,NBAPT,ROLE
   S1=S1+S2+S3
   DEN=(M1-M2+12.*(A1-A2)) /32.8767
   IF(DEN.LT..00001) DEN=.00001
   COUS=ICONS/DEN
   C(1,COUNT)=ROLE
   C(2,COUNT)=COUS
   C(3,COUNT)=IMP1
   C(4,COUNT)=IMP2
   C(5,COUNT)=APT
   C(6,COUNT)=NBAPT
   C(7,COUNT)=S1
   C(8,COUNT)=ADRES(1)
   C(9,COUNT)=ADRES(2)
   C(10,COUNT)=ADRES(3)
   TOT1=TOT1+1
   TDEFS=14400
   IF(COUNT-1) 33,34,33
34 WRITE(5)((C(L,-),L=1,10),4=1,10)
   CONT=0
   COUNTOT=COUNTOT+1
33 BUFFER IN(1,9) (C(1),A(S1))
   IF(COUNT(1)) 35,2000,33
35 BUFFER IN(5,14,A(1)) ROLE
   IF(ROLE.EQ.ROLE1) GO TO 35
   PRINT 11,ROLE,ROLE1,11,12
36 I1=I1+1
   GO TO 33
2000 PRINT 17,COUNTOT,I1,12
   STOP
   END

```

```

PROGRAM FREQUEN (INPUT,OUTPUT,TAPF1)
C  COMPTAGE DES BLOCS ET DES APPARTEMENTS
C  MOYENNE, ECART TYPE DES CONSO-MATIONS
    DIMENSION XLOG(21),YLOG(21),LOG(21,30)
    REAL MO(21),S(21),AMA(21),AMI(21)
    INTEGER    ROLE,APT,M1,A1,M2,A2,A(10,10),APT1
    REWIND 1
10  FORMAT(1X,I5,9X,I3,2X,2I2,8X,2I2,6X,I8)
15  FORMAT(1H1,* NOMBRE DE BLOCS      *F10.0,
    1* NOMBRE D APPARTEMENTS* F10.0,
    2///* NOMBRE      DES BLOCS,NOMBRE      DES APPARTEMENTS*)
16  FORMAT(F12.0,F25.0,I4)
17  FORMAT(1H1,
    1///* NOTE L INDICE 20 CORRESPOND AUX DUPLEX A 2 COMPTEURS*//
    2* NB DE LOG SUR LE COMPTEUR,MOYENNE,ECART TYPE,MINI,MAXI*//)
18  FORMAT(I13,F21.2,3F9.2)
200  FORMAT(7I10,3A10)
    COUNT =0.
    COUNT1 = 0.
    CONTO =0.
    DO 4 I =1,21
    MO(I)=0.
    S(I)=0.
    AMA(I)=0.
    AMI(I)=1000000.
    XLOG(I)=0.
    4  YLOG(I)=0.
    2  READ (1) A
    IF (EOF(1)) 2000,1
    1  DO 3 K =1,10
    APT=A(5,K)
    APT1=A(6,K)
    IF(A(2,K).LT.10) GO TO 3
    IF(APT.LT..5) GO TO 3
    IF(APT1.EQ.APT) GO TO 60
    IF(APT1.LT..5) GO TO 3
    IF(IABS(APT-APT1).GT.2) GO TO 3
    IF((APT+APT1).EQ.3) GO TO 61
    APT=MAX0(APT,APT1)
    GO TO 60
    61  IF(APT.NE.1) GO TO 60
    APT=20
    CONS=A(2,K)/2.
    IF(CONS.GT.1000.) GO TO 3
    CONTO=CONTO+2.
    APT2=2
    GO TO 62
    60  CONS=A(2,K)/APT
    IF(APT.NE.5) GO TO 63
    IF(CONS.GT.600) GO TO 3
    IF(CONS.GT.1000) GO TO 3
    CONTO=CONTO+APT
    APT2=APT

```

```

      IF (APT.GE.20) APT=21
62  COUNT=COUNT+1
      MC(APT)=MC(APT)+CONS
      S(APT) = S(APT) + CONS * CONS
      XLOG(APT) = XLOG(APT) +1.
      YLOG(APT) = YLOG(APT) +APT2
      AMA(APT) =AMAX1(AMA(APT),CONS)
      AMI(APT) =AMIN1(AMI(APT),CONS)
      3  CONTINUE
      GO TO 2
2000 PRINT 15,COUNT,CONTO
      DO 6 K=1,21
      IF (XLOG(K).LT..5) XLOG(K)=.2
      MC(K)=MC(K)/XLOG(K)
      6  S(K) =SQRT(S(K)/XLOG(K) - MC(K)*MC(K))
      PRINT 16,( XLOG(K),YLOG(K),K ,K=1,21 )
      PRINT 17
      PRINT 18,( K,MC(K),S(K),AMI(K),AMA(K) ,K=1,21)
      STOP
      END

```

MOYAN

```

PROGRAM MOYAN(INPUT,OUTPUT,TAPE1,TAPE2)
C  ETUDE PRIMAIRE DES RELEVES TRIMESTRIELS DES COMPTEURS D EAU DE MTL-EST E
C  POINTE AUX TREMBLES
C  CE PROGRAMME COMPILE DES MOYENNES ANNUELLES DE CONSOMMATION D EAU RESIDEN
C  LE. NUMERO DU TAPE D ENTREE EST SS3009 ET CELUI DE SORTIE EST SS3261
C  VARIABLES IMPORTANTES, JDAT%JDAT%DATE EN JOURS COMPTES
C                                  JAN %ANNEE
C                                  JCONS%CONSOMMATION EN GALLONS IMP.
C                                  JOTOT%NOMBRE TOTAL DE JOURS ECHANTILLONNES DANS L A
C                                  IAPT %LE NOMBRE DE LOGIS PAR COMPTEUR
C                                  IBLOC%NOMBRE DE LOGIS PAR BLOC
C                                  MOY(K)%MOYENNE JOURNALIERE ANNUELLE PAR COMPTEUR
C
  DIMENSION JCOM(50),JDAT(50),JCONS(50),JAN(50),JA(12),MOY(12)
  DIMENSION IA(24)
  INTEGER REST,RESI,SSI
  PRINT 99
  99 FORMAT(1H1,7X,*CODE*,10X,*MOYENNE*,5X,*NB TOTAL JOURS*,5X,*NB MESU
  IRES*,5X,*ANNEE*//)
  100 M=0
      K =1
      NDATE=0
      MDATE=0
      MC=0
      IPANS=0
      SSI=0
      DO 1000 II=1,12
        JA(II)=0
  1000 MOY(II)=0
      READ(1)ICOM,(JCOM(I),I=1,50)
      IF(EOF(1))1,2
C  OPERATION DE DECODAGE -DEBUT
      2 IAPT=ICOM/100000000
      IBLOC=(ICOM/1000000)-(IAPT*100)
      N=ICOM-(ICOM/100)*100
      DO 3 I=1,50
        JDAT(I)=JCOM(I)/100000000
        JAN(I)= (JCOM(I)-100000000*JDAT(I))/1000000
      3 JCONS(I)=JCOM(I)-(JCOM(I)/1000000)*1000000
C  OPERATION DE DECODAGE -FIN
      PRINT 4,ICOM
      4 FORMAT(5X,I10)
      ... L=N-1
      DO 5 I=1,L
        IDATS =JDAT(I+1)-JDAT(I)
        IF(JCONS(I+1).NE.999999)GO TO 6
C  JCONS(I+1) %999999 SI LA CONSOMMATION A PU ETRE AFFECTEE PAR UNE PANNE
        IF(JAN(I).NE.JAN(I+1))GO TO 11
        IPANS=IDATS
        GO TO 5
      11 NDATE=365-JDAT(I)  $ MDATE=JDAT(I+1)
        IF(MC.EQ.0)GO TO 13
        JOTOT=365-IPANS-NDATE
      14 MC=1
        RESI=0
        GO TO 9

```

```

13 JOTOT=365-JDAT(1)-NDATE-IPANS
   GO TO 14
6  IF(JAN(1).EQ.JAN(I+1))GO TO 7
   REST=(JCONS(I+1)*(365-JDAT(1)))/(IDATS+365)
   RES1=JCONS(I+1)-REST
   SS1=SS1+REST
   MC=MC+1
   IF(MC.EQ.1)GO TO 8
12 JOTOT=365-IPANS-MDATE
   MDATE=0
   GO TO 9
8  JOTOT=365-JDAT(1)-IPANS
9  MOY(K)=SS1/(JOTOT*IAPT)
   JA(K)=JAN(1)
   M=M+1
   PRINT 10,MOY(K),JOTOT,M,JAN(1)
10 FORMAT(13X,I4,10X,I4,15X,I3,9X,I2)
   K=K+1
   SS1=RES1
   M=1
   IPANS=0
   GO TO 5
7  SS1=SS1+JCONS(I+1)
   M=M+1
5  CONTINUE
   WRITE(2) IAPT,IBLOC,(JA(K),MOY(K),K=1,12)
   GO TO 100
1  CONTINUE
   REWIND 2
   DO 2004 K=1,1000
   READ(2)IAPT,IBLOC,(IA(1),I=1,24)
   IF(EOF(2))2003,2004
2004 PRINT2101,IAPT,IBLOC,(IA(1),I=1,24)
2101 FORMAT(5X,I4,5X,I4/,2X,6(I6,1X,I3,1X),/6(I6,1X,I3,1X))
2003 STOP
      END

```

MOYI

```

PROGRAM MOYT(INPUT,OUTPUT,TAPE2)
C ETUDE SECONDAIRE DES RELEVES TRIMESTRIELS DES COMPTEURS D EAU DE MTL-EST ET PO
C INTE AUX TREMBLES
C CE PROGRAMME EVALUE LA MOYENNE ANNUELLE ET LA VARIANCE POUR CHAQUE ANNEE EN ME
C ME TEMPS QUE POUR CHAQUE TYPE DE COMPTEUR
C VARIABLES IMPORTANTES, IAPT=NOMBRE DE LOGIS PAR COMPTEUR
C                               IBLOC=NOMBRE DE LOGIS PAR BLOC
C                               IAN(I)=L ANNEE CORRESPONDANT A ICON(I)
C                               ICON(I)=MOYENNE ANNUELLE POUR UN COMPTEUR
C
  DIMENSION IAN(12),N(12),SS1(12),SS2(12),ICON(12),MAX1(12),MIN1(12)
  INTEGER SS1,SS2,S1,S2
  IAP =1
C
  1000 NN =S1=S2=MAX2=0  $ MIN2=9999
  DO 1 J=1,12
    N(J) =SS1(J) =SS2(J) =MAX1(J)=0  $ MIN1(J)=9999
  1 CONTINUE
  2 M=1      IBLOC
  READ(2) IAPT,(IAN(I),ICON(I),I=1,12)
  IF(EOF(2))3,4
  4 IF(IAPT.NE.IAP)GO TO 5
  DO 100 I=1,12
C MAXIMUM DE 12 ANNEES D ECHANTILLONAGE DONC I VARIE DE 1 A 12
  IF(IAN(I).EQ.0)GO TO 2
  J=IAN(I)-59
  IF(IAN(I).GE.60.AND.IAN(I).LE.71)GO TO 200
  PRINT 201,IAPT,IAN(I)
  201 FORMAT(2I4,1X)
C                               SS1=SOMME DES MOYENNES DU TAPE1 SS3009
C                               SS2=SOMME DES CARRES DES MOYENNES
  200 N(J)=N(J)+1
  SS1(J)=SS1(J)+ICON(I)
  SS2(J)=SS2(J)+ICON(I)**2
  IF(ICON(I).GE.MAX1(J)) MAX1(J)=ICON(I)
  IF(ICON(I).LE.MIN1(J)) MIN1(J)=ICON(I)
  100 CONTINUE
  GO TO 2
  5 PRINT 6,IAP
  6 FORMAT(1H1,20X,*MOYENNE ANNUELLE POUR*,I4,*LOGIS PAR COMPTEUR*/,10
  1X,*ANNEE*,10X,*MOYENNE*,12X,*ECART TYPE*,10X,*MAX*,10X,*MIN*,10X,*
  2NB*//)
  IAP =IAPT
  11 DO 101 J=1,12
  IF(N(J).EQ.0)GO TO 101
C                               MOY=MOYENNE DES MOYENNES
  MOY =SS1(J)/N(J)
C                               ECAR=ECART TYPE CORRESPONDANT
  ECAR= ((SS2(J)-(SS1(J)**2)/N(J))/(N(J)-1))**.5
  NN = NN+N(J)
  S1 =S1+SS1(J)
  S2 =S2+SS2(J)
  IN =J+59
C ON IMPRIME ICI LA MOYENNE POUR CHAQUE ANNEE
  PRINT 7,IN,MOY,ECAR,MAX1(J),MIN1(J),N(J)
  7 FORMAT(11X,I2,14X,I3,12X,E15.6,7X,I3,10X,I3,10X,I3)

```

```

      IF (MAX1(J).GE.MAX2) MAX2=MAX1(J)
      IF (MIN1(J).LE.MIN2) MIN2=MIN1(J)
101  CONTINUE
C      MOT =MOYENNE GLOBALE CUMULANT LES ANNEES
      MOT =S1/NN
C      ECT =ECART TYPE CORRESPONDANT
      ECT= ((S2-(S1**2)/NN)/(NN-1))**.5
C ON IMPRIME ICI LA MOYENNE GLOBALE POUR CHAQUE TYPE DE COMPTEUR
      PRINT 9,MOT,ECT,MAX2,MIN2,NN
      9 FORMAT(1X//,10X,*MOYENNE GLOBALE*,I4,10X,*ECART TYPE *,E15.6,10X,*
      1MAX *,I4,5X,*MIN *,I3,5X,*NBRE *,I4)
      IF (M.EQ.2) GO TO 10
C
      BACKSPACE 2
      GO TO 1000
      3 M=2
      GO TO 11
10  STOP
      END

```

MOIT1

```

PROGRAM MOIT1(INPUT,OUTPUT,TAPE2)
C ETUDE SECONDAIRE DES RELEVES TRIMESTRIELS DES COMPTEURS D EAU DE MTL-EST ET PO
C INTE AUX TREMBLES
C CE PROGRAMME EVALUE LA MOYENNE ANNUELLE ET LA VARIANCE POUR CHAQUE ANNEE EN ME
C ME TEMPS QUE POUR CHAQUE TYPE DE BLOC,TOUT EN SE LIMITANT AUX COMPTEURS DESERV
C ANT UN SEUL LOGIS
C VARIABLES IMPORTANTES, IAPT=NUMBER DE LOGIS PAR COMPTEUR
C                               IBLOC=NUMBER DE LOGIS PAR BLOC
C                               IAN(I)=L ANNEE CORRESPONDANT A ICON(I)
C                               ICON(I)=MOYENNE ANNUELLE POUR UN COMPTEUR
C
  DIMENSION IAN(12),ICON(12),N(8,12),SS1(8,12),SS2(8,12),MAX1(8,12),
  MIN(8,12)
  INTEGER SS1,SS2
  ISS1=NN=MAX2=ISS2=0 $ MIN2=9999
  DO 2 K=2,8
  DO 2 J=1,12
  N(K,J)=SS1(K,J)=SS2(K,J)=MAX1(K,J)=0 $ MIN(K,J)=9999
2 CONTINUE
3 READ(2) IAPT,IBLOC,(IAN(I),ICON(I),I=1,12)
  IF(IAPT.NE.1) GO TO 99
  DO 100 I=2,12
C MAXIMUM DE 12 ANNEES D ECHANTILLONAGE DONC I VARIE DE 1 A 12
  IF(IAN(I).EQ.0) GO TO 3
  J=IAN(I)-59
  K=IBLOC
  N(K,J)=N(K,J)+1
C                               SS1=SOMME DES MOYENNES DU TAPE1 SS3009
  SS1(K,J)=SS1(K,J)+ICON(I)
C                               SS2=SOMME DES CARRES DES MOYENNES
  SS2(K,J)=SS2(K,J)+ICON(I)**2
  IF(ICON(I).GE.MAX1(K,J)) MAX1(K,J)=ICON(I)
  IF(ICON(I).LE.MIN(K,J)) MIN(K,J)=ICON(I)
100 CONTINUE
  GO TO 3
99 PRINT 98
98 FORMAT(1H1,*MOYENNE ANNUELLE POUR UN LOGIS PAR COMPTEUR,COMPTE TEN
  1U DES DIMENSIONS DU BLOC*//,10X,*BLOC*,10X,*ANNEE*,10X,*MOYENNE*,1
  22X,*ECART TYPE*,10X,*MAX*,10X,*MIN*,10X,*NB*//)
  DO 101 K=2,8
  PRINT 97,K
97 FORMAT(11X,I2)
  DO 102 J=1,12
  IF(N(K,J).LE.1) GO TO 102
C                               MOY=MOYENNE DES MOYENNES
  MOY=SS1(K,J)/N(K,J)
C                               ECAR=ECART TYPE CORRESPONDANT
  ECAR=((SS2(K,J)-(SS1(K,J)**2)/N(K,J))/(N(K,J)-1))**.5
  IN=J+59
C ON IMPRIME ICI LA MOYENNE POUR CHAQUE ANNEE
  PRINT 96,IN,MOY,ECAR,MAX1(K,J),MIN(K,J),N(K,J)
96 FORMAT(25X,I2,14X,I3,12X,E15.6,7X,I3,10X,I3,10X,I3)
  ISS1=ISS1+SS1(K,J)
  NN=NN+N(K,J)
  ISS2=ISS2+SS2(K,J)
  IF(MAX1(K,J).GE.MAX2) MAX2=MAX1(K,J)

```

```
IF(MIN(K,J).LE.MIN2) MIN2=MIN(K,J)
102 CONTINUE
IF(NN.LE.1) GO TO 101
C      MOYT=MOYENNE GLOBALE CUMULANT LES ANNEES
C      MOYT=ISS1/NN
C      EC4 =ECART TYPE CORRESPONDANT
C      ECA=((ISS2-(ISS1**2)/NN)/(NN-1))**.5
C ON IMPRIME ICI LA MOYENNE GLOBALE POUR CHAQUE TYPE DE BLOC.
PRINT 95,MOYT,ECA,MAX2,MIN2,NN
95 FORMAT(19X,*MOYENNE GLOBALE*/.41X,I3,12X,E15.6,7X,I3,2(10X,I3)/)
ISS1=NN=MAX2=ISS2=0 $ MIN2=9999
101 CONTINUE
STOP
END
```

MOYSON

```

      PROGRAM MOYSON(INPUT,OUTPUT,TAPE1)
C   CE PROGRAMME FAIT L ETUDE DES VARIATIONS SAISONNIERES SUR LES RELEVES
C   TRIMESTRIELS DE LA VILLE DE MTL.
      DIMENSION JDAT(50),JAN(50),JCONS(50),JCOM(50),NDATE(50),MC(4500),N
      IN(4500),IS(2),ISS(2),IN(2),MC2(4500)
      MOE=MOH=MECE=MECH=MME=MMH=0
      IAP =1
C   K PEUT VALOIR AU MAX. 4500 CAR IL INDICE CHAQUE JOUR DE 1960 A 1970
      DO 1 K=1,4500
        MC2(K)=0
        1 MC(K)=NN(K)=0
        DO 2 II=1,2
          2 IS(II)=ISS(II)=IN(II)=0
C
C
        3 M=1
C   LECTURE DU TAPE SS3009 ET DECODAGE
        READ(1)ICOM,(JCOM(I),I=1,50)
        IF (EOF(1))4,5
        5 IAPT=ICOM/100000000
        N=ICOM-(ICOM/100)*100
        DO 6 I=1,50
          JDAT(I)=JCOM(I)/100000000
          JAN(I)=(JCOM(I)-100000000*JDAT(I))/1000000
          NDATE(I)=JDAT(I)+(JAN(I)-60)*365
        6 JCONS(I)=JCOM(I)-(JCOM(I)/1000000)*1000000
C   TEST SUR LE CHANGEMENT DE CLASSE
        IF (IAPT.EQ.IAP)GO TO 7
        8 DO 9 K=1,4500
          JO= MOD(K,365)
          II=1
C   TEST SUR LA POSITION SAISONNIERE (151=PREMIER JUIN,243=PREMIER SEPT.)
          IF (JO.LT.151.OR.JO.GT.243) II=2
          IS(II)=IS(II)+MC(K)
          ISS(II)=ISS(II)+MC2(K)
          IN(II)=IN(II)+NN(K)
        9 CONTINUE
        DO 10 II=1,2
          MOY=IS(II)/IN(II)
          ECT=((ISS(II)-((IS(II)**2)/IN(II)))/(IN(II)-1))**.5
          IF (II.EQ.2)GO TO 11
C   CALCULS POUR LA PERIODE DE L ETE
          MOE= MOE+IS(II)
          MECE=MECE+ISS(II)
          MME=MME+IN(II)
          PRINT 12,IAP,MOY,ECT,IN(II)
        12 FORMAT(10X,*ETE, TYPE;*,I3,5X,*MOYENNE;*,I4,5X,*ECART TYPE;*,E15.6
          1,5X,*NBRE*,I6/)
          GO TO 10
C   CALCULS POUR LA PERIODE DE L HIVER
        11 MOH=MOH+IS(II)
          MECH=MECH+ISS(II)
          MMH=MMH+IN(II)
          PRINT 13,IAP,MOY,ECT,IN(II)
        13 FORMAT(10X,*HIV.,TYPE;*,I3,5X,*MOYENNE;*,I4,5X,*ECART TYPE;*,E15.6

```

```

1,5X,*NBRE*,I6//)
10 CONTINUE
  IF(M.EQ.2)GO TO 14
  IAP=IAPT
  BACKSPACE 1
  GO TO 1000
7 L=N-1
  DO 15 I=1,L
    IF(JCONS(I+1).EQ.999999)GO TO 15
    N=NDATE(I+1)-NDATE(I)
    ICJ=JCONS(I+1)/(N*IAPT)
    K1=NDATE(I)
    K2=NDATE(I+1)
    K3 =K2-K1
    IF(K3.GT.272) GO TO 15
    J01=MOD(K1,365)
    J02=MOD(K2,365)
C TEST DE SELECTION DES DONNEES
    IF(J01.GE.151.AND.J02.LE.243) GO TO 99
    IF(J01.GT.243) GO TO 99
    IF(J02.LT.151) GO TO 99
    GO TO 15
99 DO 16 K=K1,K2
    MC(K)=MC(K)+ICJ
    MC2(K)=MC2(K)+(ICJ**2)
16 NN(K)=NN(K)+1
15 CONTINUE
    GO TO 3
4 M=2
    GO TO 8
C CALCUL FINAL POUR OBTENIR LA VARIATION GLOBALE
14 MOYH=MOH/MMH
    MOYE =MOE/MME
    ECH =((MECH-((MOH**2)/MMH))/(MMH-1))**.5
    ECE =((MECE-((MOE**2)/MME))/(MME-1))**.5
    PRINT 20,MOYH,ECH,MMH,MOYE,ECE,MME
20 FORMAT(1H1,9X,*MOYENNE HIVER*,I4,5X,*ECART TYPE *,E15.6,5X,*NBRE*,
1I6/,20X,*ETE*,I4,16X,E15.6,9X,I6)
    STOP
    END

```

∇ TT

- [1] $XA + (XA - QXA + (N, N) \rho XX) + XA - QXA + (N, N + \rho XX) \rho XA + (P + (SA * 2) \div NX) * 0.5$
- [2] $B + -2 + ((B + QB + (N, N) \rho B) * 2) \div (X + QX + (N, N) \rho ((B * 2) \div NX + 1))$
- [3] $X + XP \quad P$
- [4] $G1 + (0.25 \times X + X * 3), (((3 \times X) + (16 \times X * 3) + 5 \times X * 5) \div 96), ((-15 \times X) + (17 \times X * 3) + (19 \times X * 5) + 3 \times X * 7) + 384$
- [5] $G1 + G1, ((79 \times X * 9) + (776 \times X * 7) + (1482 \times X * 5) + (-1920 \times X * 3) + (-945 \times X)) + 92160$
- [6] $B + X + (G1[1] \div B) + (G1[2] \div B * 2) + (G1[3] \div B * 3) + (G1[4] \div B * 4)$
- [7] $SAA + ((N, N) \rho SA * 2) + Q(N, N) \rho SA * 2$
- [8] ''
- [9] $B + (|XA) \leq B$
- [10] $SAA + (((|SAA) \geq FF \quad 1 - P1) \wedge (|SAA) \leq FF \quad P1)$
- [11] $\square + SAA \wedge B$

ANNEXE A3REGRESSION LINEAIRE

On dispose de 2 séries

$$\begin{array}{ll} x_i & i = 1, \dots, N \\ y_i & i = 1, \dots, N \end{array} \quad \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

On cherche une régression de la forme:

$$Y = B + A (x - \bar{x})$$

Les coefficients A et B sont déterminés de manière que

$$\sum_{i=1}^N (Y_i - y_i)^2 \text{ soit minimum (méthode des moindres carrés)}$$

et on a:

$$\begin{aligned} B &= \frac{1}{N} \sum_{i=1}^N y_i = \bar{y} \\ A &= \frac{\sum_{i=1}^N y_i (x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} \end{aligned}$$

Le coefficient de corrélation R qui mesure l'association entre les 2 variables x et y est:

$$R = \frac{\sum_{i=1}^N y_i (x_i - \bar{x})}{\sqrt{\left(\sum_{i=1}^N (x_i - \bar{x})^2 \right) \left(\sum_{i=1}^N (y_i - \bar{y})^2 \right)}}$$

A3.1 Intervalle de confiance sur B

Si β est la valeur vraie de B (estimée par B)

$$\frac{B - \beta}{s/\sqrt{N}} = t \text{ suit une loi de Student à } (N-2)^0 \text{ de liberté}$$

$$\text{avec } s^2 = \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N-2}$$

On a donc pour l'intervalle de confiance à $(1-\alpha) \%$

$$B - t_{\alpha/2, N-2} \frac{s}{\sqrt{N}} \leq \beta \leq B + t_{\alpha/2, N-2} \frac{s}{\sqrt{N}}$$

$$t_{\alpha/2, N-2} \text{ est tel que: } \text{Prob} [t \geq t_{\alpha/2, N-2}] = \frac{\alpha}{2}$$

A3.2 Intervalle de confiance sur A

De la même manière on montre que l'intervalle de confiance de A au niveau $(1-\alpha) \%$ est:

$$A - t_{\alpha/2, N-2} \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} \leq \alpha_0 \leq A + t_{\alpha/2, N-2} \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}}$$

α_0 est la valeur vraie, inconnue, de A.

A3.3 Intervalle de confiance sur R

$$\text{On pose } z = \frac{1}{2} \text{Log} \left(\frac{1 + R}{1 - R} \right)$$

$$\text{Si } \bar{z} = \frac{1}{2} \text{Log} \left(\frac{1 + \rho}{1 - \rho} \right)$$

ρ étant la vraie valeur du coefficient de corrélation.

On peut déterminer l'intervalle de confiance de ρ au niveau $(1-\alpha) \%$

$$\rho_1 \leq \rho \leq \rho_2$$

avec ρ_1 et ρ_2 tels que

$$\frac{1}{2} \operatorname{Log} \left(\frac{1 + \rho_1}{1 - \rho_1} \right) = \frac{1}{2} \operatorname{Log} \left(\frac{1 + R}{1 - R} \right) - u_{\alpha/2} \cdot \frac{1}{\sqrt{N-3}}$$

$$\frac{1}{2} \operatorname{Log} \left(\frac{1 + \rho_2}{1 - \rho_2} \right) = \frac{1}{2} \operatorname{Log} \left(\frac{1 + R}{1 - R} \right) + u_{\alpha/2} \cdot \frac{1}{\sqrt{N-3}}$$

où $u_{\alpha/2}$ variable normale standardisée telle que:

$$\operatorname{Prob} \left[u \geq u_{\alpha/2} \right] = \frac{\alpha}{2}$$

REFERENCE

D.M. MINNELBLAU: Process Analysis by statistical methods,

John Wiley and sons

ANNEXE A4METHODES D'ECHANTILLONNAGEA4.1 Position du problème

On considère une population que l'on décompose en k classes de taille $N_1, N_2, \dots, N_k$.

On veut estimer la valeur d'une certaine caractéristique ξ de la population (par exemple: consommation annuelle moyenne), la moyenne et l'écart-type de cette caractéristique pour la classe i sont μ_i et σ_i .

Pour estimer ξ , on prend dans chaque classe un échantillon de taille n_i , la moyenne de la caractéristique calculée à partir de l'échantillon pour la classe i est $\bar{x}_i$, qui est distribuée avec une moyenne μ_i et une variance $\frac{\sigma_i^2}{n_i}$. Il est alors possible d'estimer ξ par

$$X = \sum_{i=1}^k N_i \bar{x}_i$$

et l'on a

$$\text{var } X = \text{var} \left(\sum_{i=1}^k N_i \bar{x}_i \right) = \sum_{i=1}^k N_i^2 \frac{\sigma_i^2}{n_i} \quad (\text{Si } n_i \ll N)$$

Le problème est le choix des tailles des échantillons n_i , pour cela nous disposons de 2 méthodes:

- . méthode de répartition optimum;
- . méthode d'échantillonnage proportionnel.

A4.2 Méthode de répartition optimum

On veut choisir les tailles et échantillons n_i de manière à minimiser la variance de l'estimation compte tenu du coût total impliqué par l'échantillonnage.

Si C est le coût total de l'échantillonnage, et si C_i est le coût pour chaque unité de la classe i on a :

$$C = \sum_{i=1}^k n_i C_i$$

On veut donc minimiser $\text{Var}(X)$ avec la contrainte de coût précédente ce qui revient à minimiser F telle que :

$$F = \text{var } X + \frac{1}{k^2} \left(\sum_{i=1}^k n_i C_i - C \right)$$

On a donc:

$$\frac{\partial F}{\partial n_i} = 0 \quad i = 1, \dots, k$$

$$\frac{\partial F}{\partial n_i} = -N_i^2 \frac{\sigma_i^2}{n_i^2} + \frac{1}{k^2} C_i = 0$$

d'où

$$n_i = \frac{N_i \sigma_i}{k} \sqrt{c_i} \quad (1)$$

D'autre part

$$C = \sum_{i=1}^k n_i c_i = \sum_{i=1}^k \frac{N_i \sigma_i}{k} c_i^{3/2}$$

d'où

$$k = \frac{1}{C} \sum_{i=1}^k N_i \sigma_i c_i^{3/2}$$

Si on élimine k dans la relation (1) on a:

$$n_i = C \frac{N_i \sigma_i \sqrt{c_i}}{\sum_{i=1}^k N_i \sigma_i c_i^{3/2}} \quad (2)$$

on a alors

$$\text{var } X = \frac{1}{C} \left[\sum_{i=1}^k N_i \sigma_i c_i^{3/2} \right] \left[\sum_{i=1}^k N_i \sigma_i / \sqrt{c_i} \right] \quad (3)$$

Cas particulier

Si les coûts unitaires sont égaux pour chaque classe

$$C_1 = \dots = C_i = \dots = C_k = C_0$$

$$\text{on a alors } C = (n_1 + n_2 + \dots + n_k) C_0$$

Si on pose

$$p = \sum_{i=1}^k n_i \quad (p \text{ est le nombre d'éléments échantillonnés})$$

on a alors:

$$n_i = p \frac{N_i \sigma_i}{\sum_{i=1}^k N_i \sigma_i} \quad (4)$$

$$\text{var } X = \frac{1}{p} \left[\sum_{i=1}^k N_i \sigma_i \right]^2 \quad (5)$$

A4.3 Méthode d'échantillonnage proportionnel

Dans ce cas, on a choisi à priori d'échantillonner dans chaque classe un nombre d'éléments n_i proportionnel au nombre N_i d'éléments de la classe:

$$\frac{n_1}{N_1} = \dots = \frac{n_i}{N_i} = \dots = \frac{n_k}{N_k} = \lambda$$

Si l'on pose comme précédemment :

$$C = \sum_{i=1}^k c_i n_i \quad \text{on a} \quad C = \lambda \sum_{i=1}^k c_i N_i$$

d'où en éliminant λ :

$$n_i = \frac{C N_i}{\sum_{i=1}^k c_i N_i} \quad (6)$$

et

$$\text{var } X = \frac{1}{C} \left(\sum_{i=1}^k c_i N_i \right) \left(\sum_{i=1}^k N_i \sigma_i^2 \right) \quad (7)$$

Cas particulier

Si les coûts unitaires sont égaux pour chaque classe

$$C_1 = \dots = C_i = \dots = C_k = C_0 \quad \text{et} \quad \frac{C}{C_0} = p = \sum_{i=1}^k n_i$$

on a alors:

$$n_i = \frac{N_i}{\sum_{i=1}^k N_i} \quad (8)$$

$$\text{var } X = \frac{1}{p} \left(\sum_{i=1}^k N_i \right) \left(\sum_{i=1}^k N_i \sigma_i^2 \right) \quad (9)$$

A4.4 Comparaison des 2 méthodes

Les deux méthodes recherchent une répartition adéquate pour un coût total donné.

La méthode de répartition optimum tient compte de la variabilité de la taille des classes et des écarts-types et est de ce fait préférable au moins en théorie à la méthode d'échantillonnage proportionnel. Cependant, il arrive souvent que les deux méthodes donnent des résultats voisins et alors la méthode d'échantillonnage proportionnel, plus simple à mettre en oeuvre, sera préférée.

Pour comparer ces deux méthodes, on peut comparer les variances des estimations X de ξ pour un même coût global:

Si $(\text{Var } X)_1$ pour la méthode de répartition optimum;

Si $(\text{Var } X)_2$ pour la méthode d'échantillonnage proportionnel.

La quantité $I = \frac{(\text{Var } X)_1}{(\text{Var } X)_2}$ mesure l'efficacité de l'échan-

tillonnage proportionnel par rapport à la méthode optimum, on a

$$I = \frac{\left(\sum_{i=1}^k N_i \sigma_i c_i^{3/2} \right) \left(\sum_{i=1}^k \frac{N_i \sigma_i}{\sqrt{c_i}} \right)}{\left(\sum_{i=1}^k N_i c_i \right) \left(\sum_{i=1}^k N_i \sigma_i^2 \right)} \quad (10)$$

Si les coûts unitaires sont identiques pour chaque classe, on trouve

$$I = \frac{\left(\sum_{i=1}^k N_i \sigma_i \right)^2}{\left(\sum_{i=1}^k N_i \right) \left(\sum_{i=1}^k N_i \sigma_i^2 \right)}$$

Si I est voisin de 1, la méthode d'échantillonnage proportionnel est comparable à la méthode de répartition optimum.