

L'apparition des sondages
aléatoires et des statisticiens
agricoles aux
États-Unis (1920-1940)

Emmanuel Didier

1ères JHS

ENSAE – 15 et 16 février 2006

Argument

- Nous allons montrer comment les statisticiens ont adopté les méthodes de sondage probabilistes aux États-Unis pendant l'entre-deux-guerres.
- Nous allons montrer comment cela a constitué la profession de statisticien agricole et en même temps renouvelé profondément les sondages aléatoires.
- Nous allons en même temps réfléchir à ce que signifient des phrases comme « apparaître » quand il s'agit d'une méthode ou d'une profession.

La Division

- Archives d'un bureau du ministère de l'Agriculture US pendant l'entre-deux-guerres chargé des statistiques : la Division.
- But : estimer les totaux annuels de différentes productions (céréales, bétail, etc.) pour stabiliser les cours.
- Avant les années 1920, travaillent avec un très grand nombre de Correspondants locaux
 - Volontaires
 - Fidèles
 - Information exhaustive
- **Méthode incompatible avec la sélection aléatoire.**
- **Aucun fonctionnaire n'a été formé aux statistiques.**

Les données de la Division et l'erreur probable



$$p.e = 0,6745 \frac{\sigma}{\sqrt{n-1}}$$

TABLE 11.—*Corn: Yields per acre. Selected illustrations of size of sample, measures of dispersion, and probable error—Continued*

State, year, and district	Reports	Average yield (arithmetic mean)	Standard deviation of reported yields	Coefficient of variation	Probable error of the average yield, or mean	Relative probable error
Mississippi: 1927.....	Number 430	Bushels 17.60	Bushels 5.51	Per cent 31.3	Bushels 0.18	Per cent 1.0
1.....	17	22.90	6.75	29.5	1.11	4.8
2.....	52	16.80	5.30	31.5	.50	3.0
3.....	59	18.00	5.60	31.1	.49	2.7
4.....	27	17.80	8.85	49.7	1.15	6.5
5.....	68	17.10	4.48	26.2	.37	2.2
6.....	58	16.20	3.81	23.5	.34	2.1
7.....	38	16.60	4.46	26.9	.49	3.0
8.....	47	17.60	5.35	30.4	.53	3.0
9.....	64	18.90	5.10	27.0	.43	2.3

Charles Sarle – Statisticien de l'Iowa en 1922 de **formation universitaire**

Sarle veut articuler la **représentativité** des données et le **tirage aléatoire** de la formule de Bowley. Articles de 1927 et 1932.

Ses travaux transforment les travaux de Bowley et montrent que :

Données de la Division apparaissent comme **non exhaustives**.

Formule donne lieu à une **interprétation réduite** : erreur par rapport à un échantillon de taille infinie tiré dans les mêmes conditions.

Rapprochement de la Division et de l'Université

- La Division
 - Sarle et ses deux rapports
 - Projet de loi Buchanan
 - Impulsion probable du ministre H. Wallace
 - Recherche de compétences techniques
- L'université d'Etat d'Iowa
 - Laboratoire récent et de taille modeste
 - Influence de R. Fisher
 - Agronomie
 - Liens déjà établis avec les Statisticiens Agricoles
 - Recherche de données

Organisation d'une **conférence** en 1936 :

On Statistical Methods of Sampling Agricultural Data

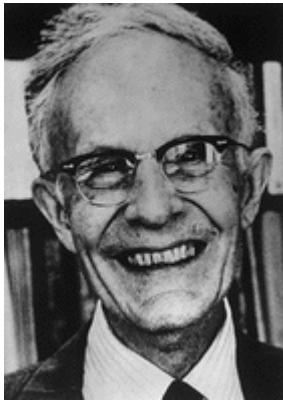
Organisée conjointement par la Division et

l'Université : comment administrer une éventuelle enquête nationale si le projet Buchanan est voté ?

Avantage mutuel à l'organisation commune.

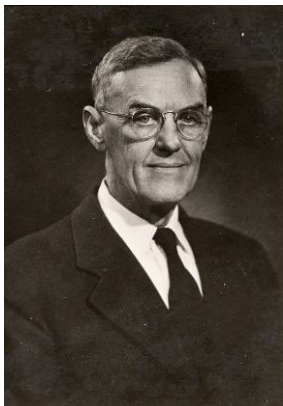
Affrontement sur la notion de représentativité

Il ne suffit pas d'organiser une conférence pour qu'elle réussisse ; il faut aussi que les débats soient riches.



Theodore W. Schultz (1902-1998), qui sera prix Nobel en 1979 pour ses travaux sur l'économie du développement, professeur d'économie rurale à Iowa State : il faut représenter des **facteurs**. Il méprise l'aléa.

Charles S. Sarle : il faut une méthode facile à **administrer**.



George Snedecor (1882-1974), fondateur du département de statistiques d'Iowa State : il faut trouver les **unités d'échantillonnage** qui donnent les résultats les plus **précis**. Il présuppose l'aléa.

Conclusion de Sarle dans un compte rendu : il faut **expérimenter** pour **comparer** les deux méthodes et trouver la meilleure.

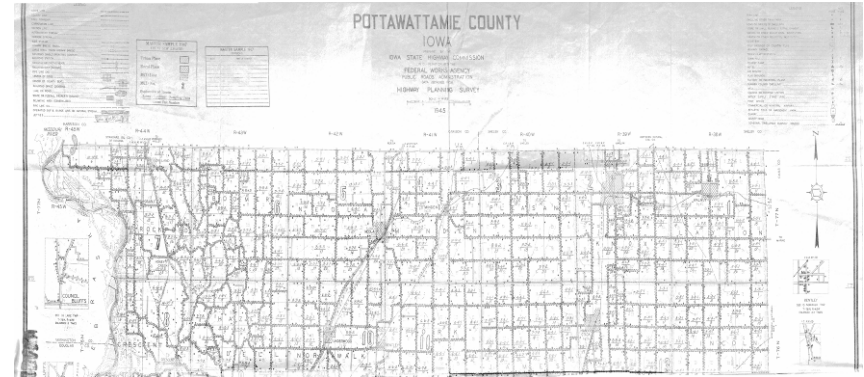
Qu'est-ce que comparer deux méthodes représentatives? (1)



Raymond Jessen,
(1910-2003)
doctorant au
laboratoire en 1938 ;
sa thèse a pour
ambition de trouver
les meilleurs unités à
échantillonner

Travail grâce aux fonds obtenus par
l'administration à la suite de la conférence :
loi Bankhead-Jones.

Avec des **cartes** là où il y en a
(Administration des Travaux Civils) et des
photos aériennes lorsque c'est nécessaire,
il construit un univers **exhaustif** des US
dans lequel des **parcelles de territoire**
peuvent être dessinées ce qui rend possible
la sélection. Comment ?



Qu'est-ce que comparer deux méthodes représentatives? (2)

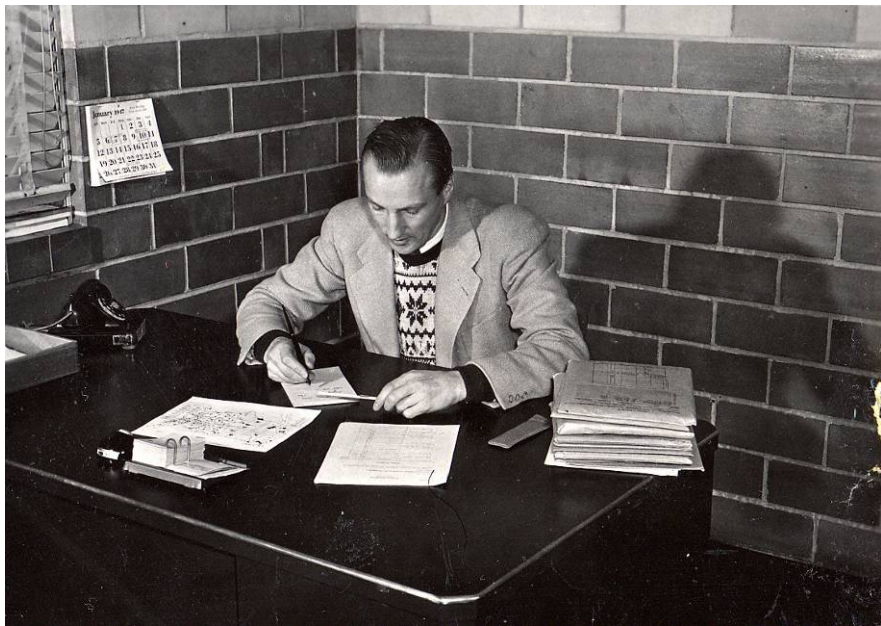
Division of Census
 Wash. D.C., 1949
 Form 502
 District

Statistical Laboratory of the Census Bureau
 Office of Representative Statistics

How State Counts

Block, Town, Precinct, County, State

Block No.	Town	Precinct	County	State	Block No.	Town	Precinct	County	State
101	1	1	1	1	101	1	1	1	1
102	1	1	1	1	102	1	1	1	1
103	1	1	1	1	103	1	1	1	1
104	1	1	1	1	104	1	1	1	1
105	1	1	1	1	105	1	1	1	1
106	1	1	1	1	106	1	1	1	1
107	1	1	1	1	107	1	1	1	1
108	1	1	1	1	108	1	1	1	1
109	1	1	1	1	109	1	1	1	1
110	1	1	1	1	110	1	1	1	1
111	1	1	1	1	111	1	1	1	1
112	1	1	1	1	112	1	1	1	1
113	1	1	1	1	113	1	1	1	1
114	1	1	1	1	114	1	1	1	1
115	1	1	1	1	115	1	1	1	1
116	1	1	1	1	116	1	1	1	1
117	1	1	1	1	117	1	1	1	1
118	1	1	1	1	118	1	1	1	1
119	1	1	1	1	119	1	1	1	1
120	1	1	1	1	120	1	1	1	1
121	1	1	1	1	121	1	1	1	1
122	1	1	1	1	122	1	1	1	1
123	1	1	1	1	123	1	1	1	1
124	1	1	1	1	124	1	1	1	1
125	1	1	1	1	125	1	1	1	1
126	1	1	1	1	126	1	1	1	1
127	1	1	1	1	127	1	1	1	1
128	1	1	1	1	128	1	1	1	1
129	1	1	1	1	129	1	1	1	1
130	1	1	1	1	130	1	1	1	1
131	1	1	1	1	131	1	1	1	1
132	1	1	1	1	132	1	1	1	1
133	1	1	1	1	133	1	1	1	1
134	1	1	1	1	134	1	1	1	1
135	1	1	1	1	135	1	1	1	1
136	1	1	1	1	136	1	1	1	1
137	1	1	1	1	137	1	1	1	1
138	1	1	1	1	138	1	1	1	1
139	1	1	1	1	139	1	1	1	1
140	1	1	1	1	140	1	1	1	1
141	1	1	1	1	141	1	1	1	1
142	1	1	1	1	142	1	1	1	1
143	1	1	1	1	143	1	1	1	1
144	1	1	1	1	144	1	1	1	1
145	1	1	1	1	145	1	1	1	1
146	1	1	1	1	146	1	1	1	1
147	1	1	1	1	147	1	1	1	1
148	1	1	1	1	148	1	1	1	1
149	1	1	1	1	149	1	1	1	1
150	1	1	1	1	150	1	1	1	1
151	1	1	1	1	151	1	1	1	1
152	1	1	1	1	152	1	1	1	1
153	1	1	1	1	153	1	1	1	1
154	1	1	1	1	154	1	1	1	1
155	1	1	1	1	155	1	1	1	1
156	1	1	1	1	156	1	1	1	1
157	1	1	1	1	157	1	1	1	1
158	1	1	1	1	158	1	1	1	1
159	1	1	1	1	159	1	1	1	1
160	1	1	1	1	160	1	1	1	1
161	1	1	1	1	161	1	1	1	1
162	1	1	1	1	162	1	1	1	1
163	1	1	1	1	163	1	1	1	1
164	1	1	1	1	164	1	1	1	1
165	1	1	1	1	165	1	1	1	1
166	1	1	1	1	166	1	1	1	1
167	1	1	1	1	167	1	1	1	1
168	1	1	1	1	168	1	1	1	1
169	1	1	1	1	169	1	1	1	1
170	1	1	1	1	170	1	1	1	1
171	1	1	1	1	171	1	1	1	1
172	1	1	1	1	172	1	1	1	1
173	1	1	1	1	173	1	1	1	1
174	1	1	1	1	174	1	1	1	1
175	1	1	1	1	175	1	1	1	1
176	1	1	1	1	176	1	1	1	1
177	1	1	1	1	177	1	1	1	1
178	1	1	1	1	178	1	1	1	1
179	1	1	1	1	179	1	1	1	1
180	1	1	1	1	180	1	1	1	1
181	1	1	1	1	181	1	1	1	1
182	1	1	1	1	182	1	1	1	1
183	1	1	1	1	183	1	1	1	1
184	1	1	1	1	184	1	1	1	1
185	1	1	1	1	185	1	1	1	1
186	1	1	1	1	186	1	1	1	1
187	1	1	1	1	187	1	1	1	1
188	1	1	1	1	188	1	1	1	1
189	1	1	1	1	189	1	1	1	1
190	1	1	1	1	190	1	1	1	1
191	1	1	1	1	191	1	1	1	1
192	1	1	1	1	192	1	1	1	1
193	1	1	1	1	193	1	1	1	1
194	1	1	1	1	194	1	1	1	1
195	1	1	1	1	195	1	1	1	1
196	1	1	1	1	196	1	1	1	1
197	1	1	1	1	197	1	1	1	1
198	1	1	1	1	198	1	1	1	1
199	1	1	1	1	199	1	1	1	1
200	1	1	1	1	200	1	1	1	1



Dans un tableau, chaque Comté (carte) était découpé en unités de comptage (nombre de fermes et nombre d'habitations) puis en unité d'échantillonnage. Pour le premier sondage, on tira un échantillon systématique de 1/18. La carte du Comté et les tableaux de conversion furent ensuite rangés dans des enveloppes (celles que nous voyons à côté de Jessen). Facilité d'administration.

Cette technique va « notablement améliorer la précision des estimateurs »

Qu'est-ce que comparer deux méthodes représentatives ? (3)

- Parallèlement aux travaux de Jessen, le Journal of Farm Economics publie en 1939 un numéro spécial sur le centenaire des statistiques agricoles (premier recensement agricole US : 1840).
- Snedecor y inscrit l'échantillonnage aléatoire dans la tradition européenne des débats sur la représentativité : (première mention de l'IIS et de Neyman 1934 dans mon corpus)
- Sarle défend l'aléa au motif que la stratification assure la représentativité.
- Jessen présente ses travaux.

JOURNAL of FARM ECONOMICS

Contents

A Century of Agricultural Statistics	697
..... Henry C. Taylor	
A Close-up View of the Development of Agricultural Statistics from 1900 to 1920	707
..... Nat C. Murray	
Why the Government Entered the Field of Crop Re- porting and Forecasting	718
..... Walter H. Ebling	
Development of Agricultural Statistics in the Bureau of the Census	735
..... Murray R. Benedict	
Progress of Agricultural Statistics in the World	761
..... E. S. and W. S. Woynitsky	
Agricultural Price Statistics in the United States and Abroad	788
..... F. A. Pearson and G. E. Brandow	
Developments in Crop and Livestock Reporting Since 1920	799
..... Joseph A. Becker and C. L. Harlan	
Estimating Local Market Prices and Farm Labor Since 1920	828
..... Roger F. Hale	
Future Improvement in Agricultural Statistics	838
..... Charles F. Sarle	
Design of Sampling Experiments in the Social Sciences	846
..... George W. Snedecor	
An Experiment in the Design of Agricultural Surveys	856
..... Raymond J. Jessen	

Published by

THE AMERICAN FARM ECONOMIC ASSOCIATION

Volume XXI

NOVEMBER, 1939

Number 4

Preuve est faite que **la sélection aléatoire est efficace**.
Cette preuve remplace la *comparaison* projetée initialement.
C'est le Master Sample qui s'imposera en 1940.

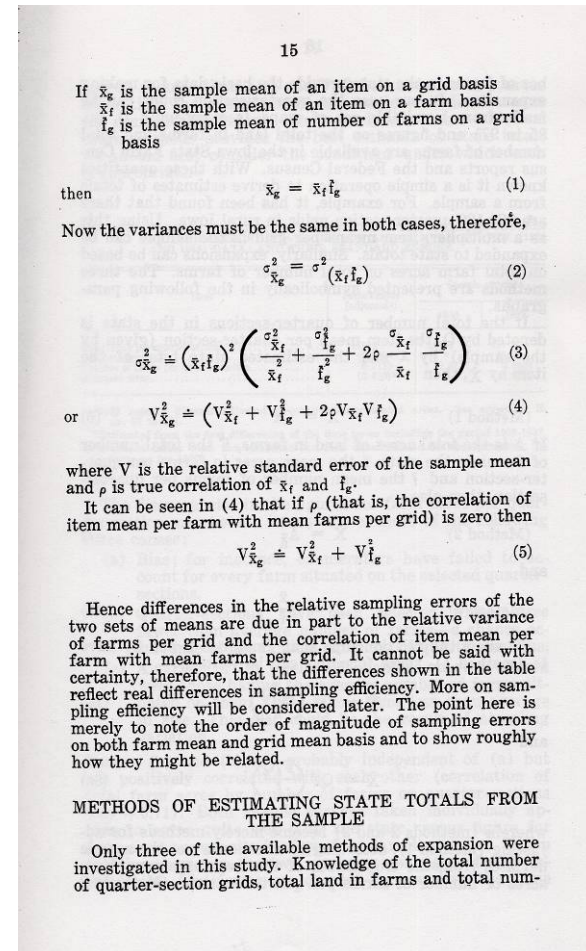
Une nouvelle place pour le statisticien

- Contrairement au Statisticien agricole des années 1920, qui était plongé dans son objet, le statisticien probabiliste domine son objet. Il est au-dessus de la carte.
- Dans le même temps apparaît donc l'enquêteur, qui se distingue du statisticien en tant que tel.



Nouveaux outils

- La statistique demande dorénavant de nouveaux outils, en particulier des formules mathématiques. Ici une page de Jessen publiée en 1942, où il fait l'analyse d'une formule de variance.



Nouvelles interviews

- L'enquête se fait maintenant en face à face, on cherche à reconnaître sur le terrain, avec le fermier, les limites de parcelles indiquées sur la carte
- Cette part importante du travail distingue aussi le statisticien agricole de son ancêtre uniquement universitaire.



Early area frame interview, probably Arkansas, 1951.

L'adoption des méthodes aléatoires est-elle une affaire de profession ?

- Les traits pertinents de la profession et de la méthode sont apparues ensemble par leur rapprochement mutuel. **C'est une production simultanée.**
- Qu'est-ce qu'apparaître ?
 - C'est local (on apparaît quelque part de bien précis).
 - C'est conjoint (deux choses se font mutuellement apparaître)
 - C'est une série de petits événements, peu notables en soi, mais rapprochés dans le temps.
- Donc réponse : oui, si « l'affaire » est une *expression mutuelle*.