

# L'invention des sondages aléatoires

Emmanuel Didier – Laurent Davezies  
ENSAE

# Introduction

- Les sondages aléatoires ne sont pas les “sondages” comme le disent les journalistes en France, c’est-à-dire des enquêtes d’opinion. Ce sont les méthodes de sondage avec plan de sondage et estimation de l’erreur, qui servent le plus souvent pour mesurer des faits, tels qu’on les enseigne à l’ENSAE.
- Ils ont acquis leur caractère d’évidence pendant la grande dépression des années 1930, d’abord aux USA, puis se sont répandus partout.
- Ils ont servi à sortir de la crise, et sont ensuite devenu l’un des outils statistiques les plus utilisés par les administrations. C’est une méthode intimement liée à l’interventionnisme de l’Etat dans l’économie.
- On va raconter leur genèse en 3 parties :
  1. Ils ont d’abord été conceptuellement élaborés en Europe
  2. Les développements techniques nécessaires mais postérieurs ont été menés aux USA
  3. Enfin, nous verrons pourquoi le type de savoirs qu’ils produisent est particulièrement compatible avec l’interventionnisme étatique.

# 1. Comment l'aléa a été qualifié de représentatif.

- La notion centrale à la fin du 19ème est celle de représentativité. Les statisticiens ont dû la définir contre d'autres formes statistiques initialement beaucoup plus légitimes, dont particulièrement le recensement exhaustif.
- C'est de la représentativité qu'est née la notion de tirage aléatoire, qui lui a peu à peu été rendue équivalente.

# 1.1. Apparition de la représentativité à l'IIS

- Qu'est-ce que l'IIS ?
- Notion de “représentativité” présentée par Anders Kiaer, directeur de la statistique norvégienne et travaillée par F. Möhn son subordonné entre 1895 et 1903.
- Représentativité se situe entre monographies et recensements et permet ainsi d'étudier de nouveaux objets : en particulier assurance sociale.
- La représentativité est la réponse à la critique de l'induction exprimée par Von Mayr, directeur de l'institut de Prusse.

# 1.1. Premières méthodes représentatives

- Kiaer explique que la représentativité est obtenue grâce à une sélection intelligente de l'échantillon.
- Les méthodes sont innombrables, il faut les tester. Il parle par exemple de « bonne couverture géographique », de miniaturisation, de l'usage des connaissances annexes (question qui deviendra celle de la stratification).
- La sélection aléatoire est aussi mentionnée – par le Français Lucien March -, mais pas centrale.
- La représentativité de l'échantillon doit ensuite être confirmée grâce à une épreuve d'exactitude.
- Critique de Bortkiewicz : apparemment semblable mais pas de test de similitude.

Professions	Population rurale		Population urbaine	
	Recensement général	Recensement représentatif	Recensement général	Recensement représentatif
	%	%	%	%
Fonctionnaires publics	1,5	1,8	4,8	4,8
Cultivateurs	23,9	23,7	0,5	0,2
Pêcheurs	8,3	7,4	1,9	1,3
Ouvriers agricoles :				
Tenanciers	5,8	5,5	0	0
Fils de cultivateurs	9,6	8,4	0,1	0
Domestiques	4,5	4,9	0,1	0,1
Autres	5,2	4,4	0,6	0,7
Artisans :				
Chefs et patrons	4,7	4,5	7,4	6,8
Ouvriers	3,5	3,7	17,8	18,1
Ouvriers de fabriques	4,3	5,6	12,1	13,1
Marchands, armateurs, etc.	1,3	1,3	6,5	6,6
Employés de commerce, etc.	0,7	0,8	6,2	5,4
Ouvriers de commerce, etc.	0,4	0,3	5,9	4,6
Capitaines, seconds et mécaniciens de navires	0,9	1,8	3,2	4,6
Matelots	2	4,3	5,5	7,6
Autres professions	23,4	21,6	27,4	26,1
Somme totale	100	100	100	100

# 1.2. Restriction à deux méthodes

- EN 1905, la question de la représentativité disparaît complètement des débats de l'IIS.
- Rôle important d'Arthur Bowley, professeur à la LSE, qui mena des enquêtes sur la pauvreté dès 1906.
- Utilise exclusivement des sélection aléatoires et propose la notion d'erreur probable.
- *“Tout dépend d’une compréhension complète de la nature de la mesure quand nous disons, par exemple, à partir de certaines données que l’estimation la plus probable de la moyenne des salaires est de 24 shillings. Il y a, cependant, une chance sur deux qu’elle s’écarte de 4 penny de cette valeur : l’écart-type est de 6 pennies ; il y a une chance sur 10 que la moyenne soit au-dessus de 24 shillings et 8 pennies, et 1 sur 100 000 qu’elle soit supérieure à 26 shillings. C’est à ce type d’affirmation que mènent les calculs. Le résultat peut être noté brièvement 24s. ±6 pennies si la ‘déviation standard’ est adoptée comme mesure de la précision.”*
- En Italie, Corrado Gini (celui de l'indice) veut faire de la place parmi les questionnaire du recensement et met en place une méthode de sélection reposant sur l'ajustement aux marges qu'il appelle « choix judicieux ».

# 1.2. Deux méthodes légitimées par l'IIS

- Session de 1925 de l'IIS avec un comité sur les sondages.
- Rapporteurs A. Bowley et Jensen du Danemark.
- Les enquêtes représentatives sont soit aléatoires, soit par choix judicieux.

*“L’Institut international de statistiques*

*I - Attire l’attention sur les avantages très considérables qui peuvent être obtenus en appliquant la méthode représentative dans les conditions suivantes:*

*- Les résultats d’une investigation partielle ne peuvent être généralisés que si l’échantillon utilisé est par nature suffisamment représentatif de la totalité. Pour ceci, l’échantillon peut être sélectionné de différentes façons ; les deux principaux cas suivants peuvent cependant être distingués :*

*A) Sélection aléatoire : un certain nombre d’unités sont sélectionnées de façon à ce que l’exacte égalité de chance d’inclusion soit la règle dominante.*

*B) Sélection par choix judicieux : un certain nombre de groupes d’unités sont sélectionnés qui, ensemble, fournissent presque les mêmes caractéristiques que la totalité. Pour avoir connaissance de la précision des estimateurs, il est nécessaire que suffisamment de groupes soient inclus de façon à permettre la mesure des variations entre les caractères des groupes.*

*II - Recommande que les investigations soient menées de façon à permettre le traitement mathématique de la précision des résultats, et avec ces résultats une indication de la taille de l’erreur qu’ils sont susceptibles de faire commettre doit être indiquée.*

*III - Répète le souhait que dans les rapports donnant les résultats de toute investigation représentative, une description explicite et détaillée de la méthode adoptée pour sélectionner l’échantillon soit donnée.”*

# 1.3 Neyman et les intervalles de confiance

- Jerzy Neyman publie : « on the two different aspects of the representative method : the method of stratified sampling and the method of purposive selection », JRSS, 1934.
- Un article jugé fondateur :
  - Il définit les intervalles de confiance
  - Il montre que la sélection optimale par strate et proportionnelle à la dispersion et non à la moyenne
  - Il rejette la sélection par choix judicieux comme non optimale.
- En d'autres termes, il ne laisse subsister que l'aléa comme méthode représentative.
- Attention : représentativité reste absolument centrale



## 2. Développement des méthodes d'enquêtes

- L'équivalence représentativité – aléatoire restait très largement théorique et schématique à ce stade.
- Il fallait encore développer tout l'appareil méthodologique permettant de procéder concrètement aux enquêtes.
- C'est ce qu'on fait les Américains principalement à partir des années 1930.
- On va présenter quelques figures de ces travaux puis se concentrer sur la question de la constitution des bases de sondage, qui est au fondement de tout enquête.

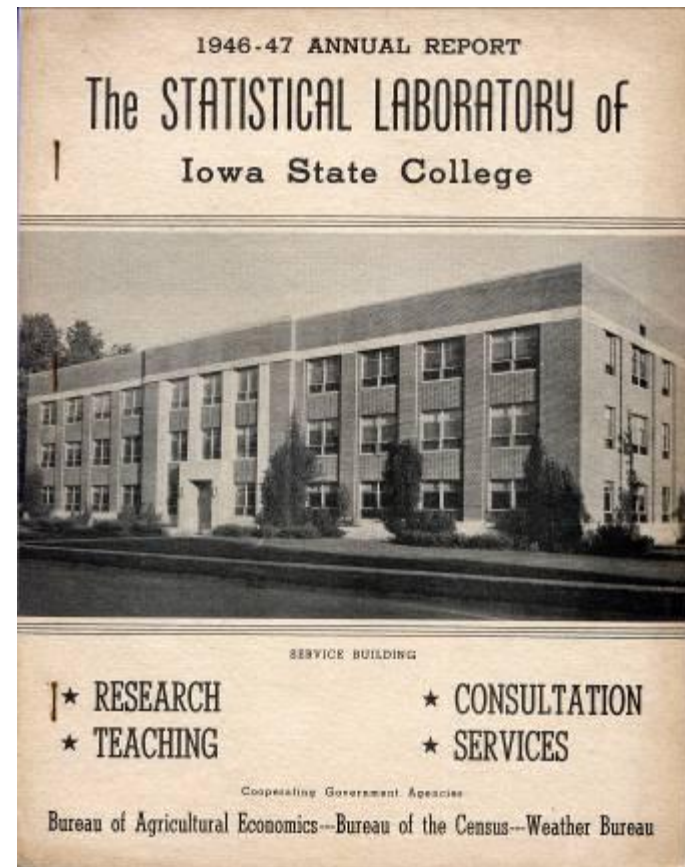
## 2.1.1 Des administrateurs

- Charles Sarle est le premier à avoir parlé de « l'erreur probable » (à la fin des années 1920) et a essayé de l'utiliser au Ministère de l'agriculture
- Corrington Gill qui a travaillé à l'administration du secours
- Calvert Dedrick deviendra le directeur de la recherche du Bureau du Recensement
- Sort tragique de Margaret Hogg

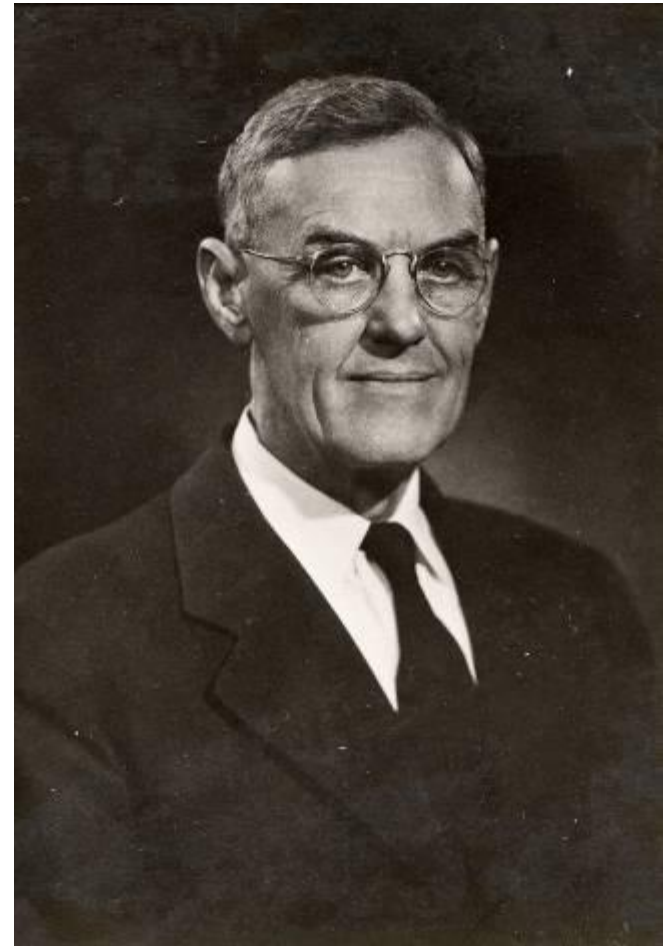


## 2.1.2 L'université

- Pour développer ses recherches, Sarle s'est trouvé des alliés à l'Université d'Etat D'Iowa, où un des premiers laboratoires de statistiques était en train de se monter.



# George Snedecor



- Le laboratoire était dirigé par G. Snedecor, qui avait écrit un article avec le Ministre de l'agriculture Wallace au début des années 1920. Il a dirigé la thèse de Ms. Cox, l'une des toutes premières en statistiques aux Etats-Unis, sur la décomposition de la variance. (elle est au centre de la tablée).

# Raymond Jessen

- Jessen a été formé aux US, Il a joué un rôle crucial dans l'invention de la sélection aléatoire. C'est l'inventeur des sondages aréolaires. Chaque chemise, sur son bureau, contient un échantillon aléatoire d'un comté de l'Iowa.



# Frederick Stephan

- Formé à l'université de Chicago puis y deviendra professeur.
- Consultant pour le gouvernement – surtout l'administration des secours et le bureau des statistiques du travail.

# Sam Stouffer

- Formé à l'Université de Chicago.
- Voyage d'étude en Grande Bretagne avec Pearson
- Connu pour avoir travaillé pendant la guerre à *The American Soldier* – une somme sur la motivation des troupes

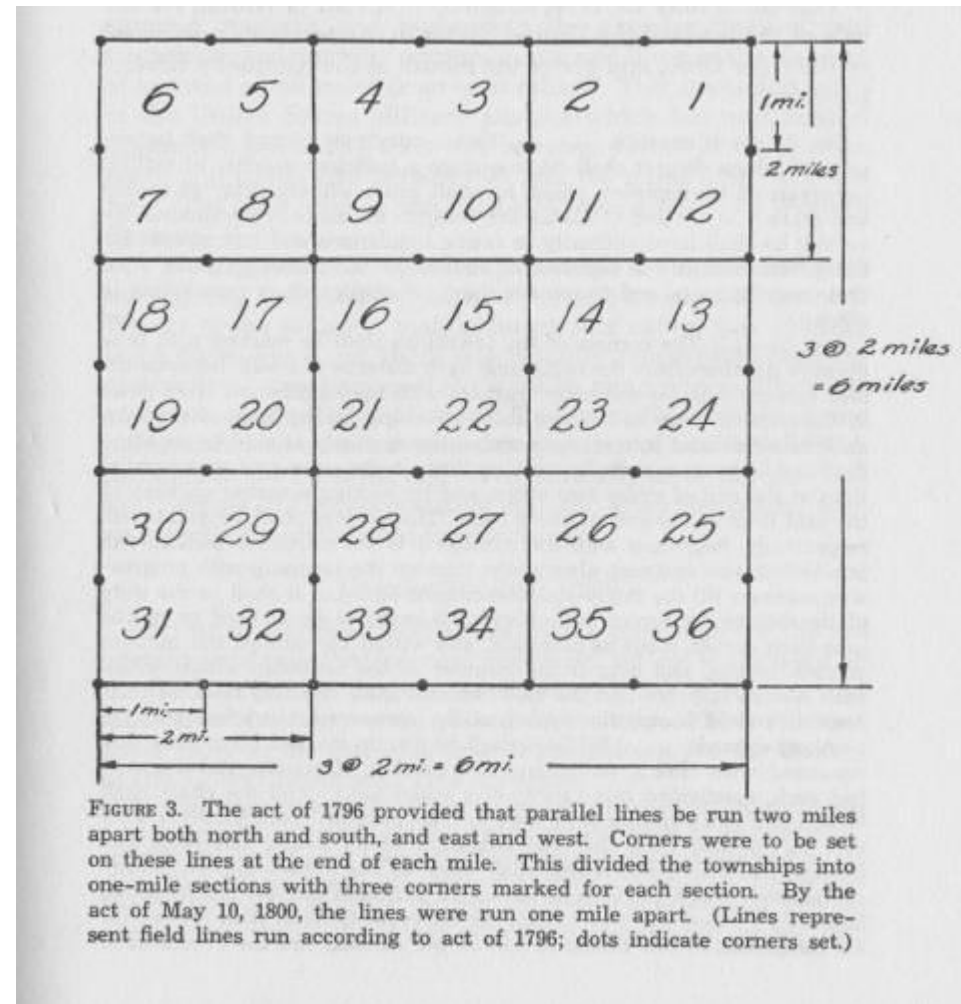
## 2.2 La base d'échantillonnage

- L'hypothèse pour tirage aléatoire est que chaque unité a une probabilité mesurable d'être échantillonnée.
- Il faut donc une base d'éléments exhaustive et sans répétition.
- Mais aux US il n'y a pas d'Etat civil. Le téléphone, aujourd'hui utilisé, ne convenait évidemment pas alors.



## 2.2.1. Un antécédent

- La vente des terres publiques pendant le 19ème siècle avait nécessité la réalisation d'un « survey »

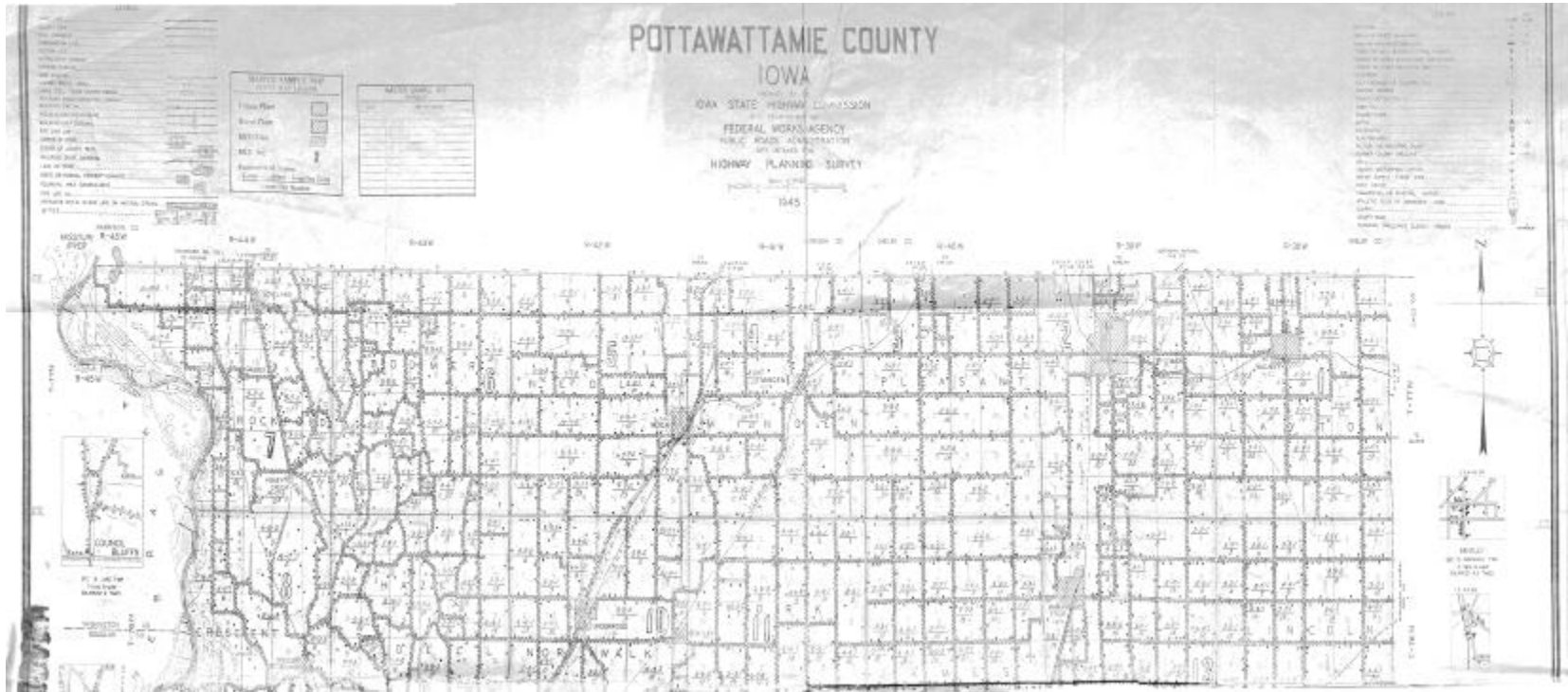


## 2.2.1. Finalement inutilisable

- Aurait été parfait car précis et régulier
- Mais n'était pas exhaustif.
- Et les archives étaient trop disséminées



## 2.2.2. Les cartes : stratégie d'exhaustivité



- Une solution (utilisée surtout par les laboratoires ruraux) fut de viser l'exhaustivité nationale. Ils utilisèrent les cartes routières des Etats pour construire les strates puis les unités d'échantillonnage. Ici l'exemple d'un comté d'Iowa.

## 2.2.2. Les photos aériennes

- Mais par endroit les cartes manquaient. Il fallut alors utiliser des photos aériennes.





## 2.2.4. Les listes : stratégie locale

- Pour les villes les cartes étaient trop différentes d'une ville à l'autre.
- Utilisation de toutes les listes disponibles :
  - Abonnés au gaz
  - Listes électorales
  - Listes des rues de la ville.

### 3. Les sondages comme nouveau type d'action

- La connaissance produite par les sondages aléatoire n'est pas utilisée de la même façon que celle qui était produite par le passé.
  - Situation du statisticien
  - Relation à l'enquêté
  - Usage des mathématiques



# 3.1 Une nouvelle place pour le statisticien

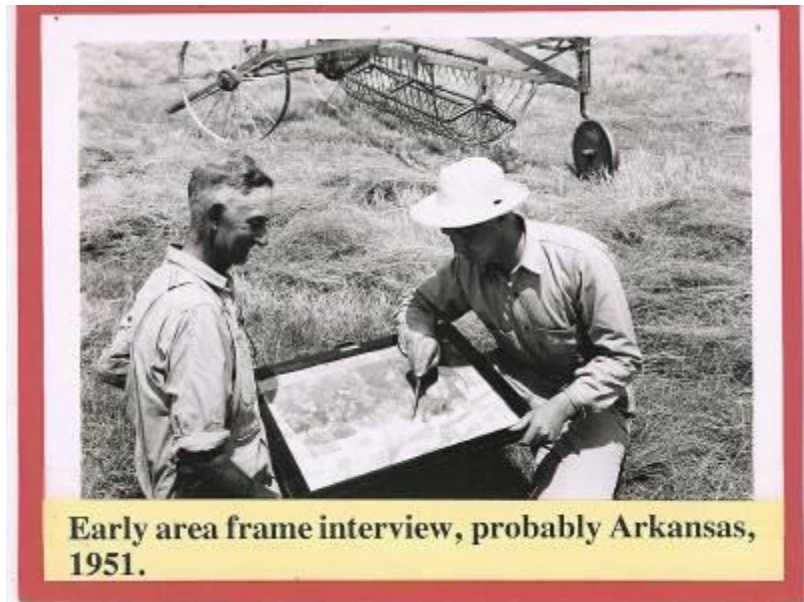
- Au contraire du Statisticien agricole qui était intégré dans son objet, le statisticien probabiliste domine son objet. Il est au dessus de la carte.
- C'est lui qui maintenant établit les rapport de corrélation entre variables : il détermine celles qui sont explicative et celles qui sont expliquées





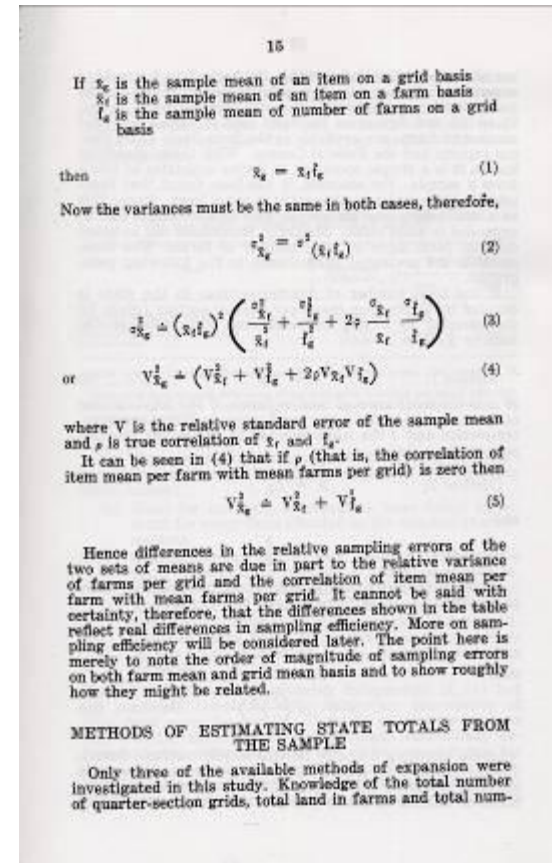
## 3.2 Nouvelles interviews

- 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
- L'enquêté ne s'autodésigne plus.



# 3.3 Nouveaux outils

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



## 3.4. Les sondages d'opinion et les quotas

- Les sondages d'opinion sont complètement marginalisés dans l'histoire technique.
- Pourtant ils ont eu une visibilité publique énorme en 1936 avec George Gallup, Elmo Roper et Archibald Crossley : ont prévu la réélection de Roosevelt contre les « straw polls »
- Mais sont considérés par les praticiens de l'aléa comme du tâtonnement arriérés (car non aléatoires).
- Reste que les gouvernants, y compris Roosevelt lui-même avec une figure comme Emil Hurja les ont énormément utilisés.

# Conclusion

- Sondages sont apparus à l'intersection entre les universitaires et les administrateurs. Depuis l'origine c'est un type de savoir lié à l'action gouvernementale.
- Ils ont permis les interventions de l'Etat dans l'économie. Fournissant des connaissances sur la main d'oeuvre, le chômage, la population rurale, etc. ils ont guidé les actions interventionnistes.
- Cela a été critiqué (C. Wright Mills) comme une massification car la population était prise en main par l'Etat plutôt qu'elle ne se dressait elle-même pour lutter contre l'adversité.
- Mais c'est aussi la raison d'être de l'Etat-providence.