

Quantifier en ethnographe.

Sur les enjeux d'une émancipation de la représentativité statistique

Julien GROS
IRISSO (DAUPHINE) / LSQ-CREST

Séminaire MetSem - Janvier 2018

Introduction

- ▶ Une réflexion issue d'une immersion tardive dans la sociologie quantitative

Introduction

- ▶ Une réflexion issue d'une immersion tardive dans la sociologie quantitative
- ▶ Une tension sur la généralisation : “Contrairement aux enquêtes statistiques sur échantillon à visée représentative, cette recherche prend le parti du localisme, sans pour autant renoncer à donner à ses résultats une portée générale”, mais “le registre de la montée en généralité ne peut être de même nature que celui défendu par les utilisateurs d'enquêtes par échantillon. Au contraire, la question de la généralisation des conclusions se pose ici dans des termes proches de ceux qu'implique l'ethnographie”.
(Cayouette-Remblière 2015 : 379)

I. Soumettre un cas ethnographique à l'analyse quantitative

II. Le cas contre l'inférence : le problème de la significativité statistique

III. La guerre des étoiles n'aura pas lieu : quantifier sans inférer

I. Soumettre un cas ethnographique à l'analyse quantitative

II. Le cas contre l'inférence : le problème de la significativité statistique

III. La guerre des étoiles n'aura pas lieu : quantifier sans inférer

I. Soumettre un cas ethnographique à l'analyse quantitative

II. Le cas contre l'inférence : le problème de la significativité statistique

III. La guerre des étoiles n'aura pas lieu : quantifier sans inférer

I. Soumettre un cas ethnographique à l'analyse quantitative

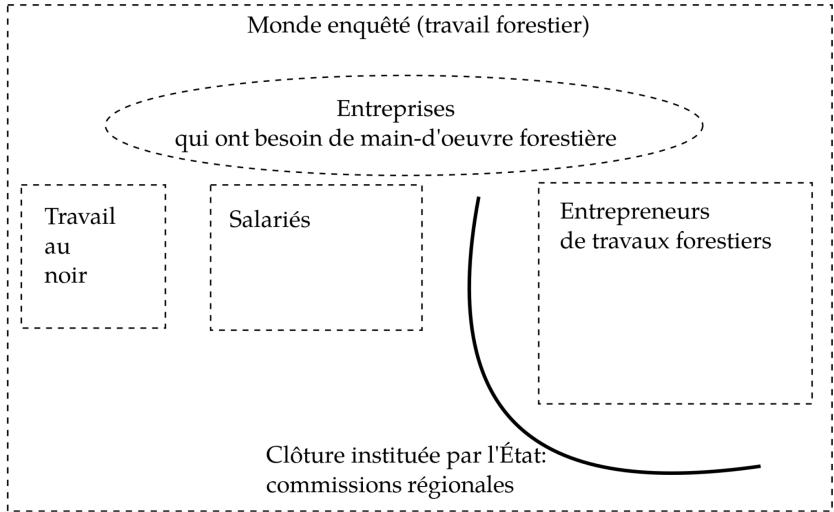
II. Le cas contre l'inférence : le problème de la significativité statistique

III. La guerre des étoiles n'aura pas lieu : quantifier sans inférer

Ethnographie et raisonnement par cas

- ▶ Compréhension de mécanismes singuliers et localisés
- ▶ Le cas n'a pas vocation à être représentatif (Passeron et Revel 1995 ; Beaud 1996 ; Dumez 2015)
- ▶ Le symptématique "n=50" (Small 2005)
- ▶ Définition/ construction du cas

Une enquête de terrain localisée



Une base de données fidèle au cas

- ▶ Construire une base de données fidèle au cas
(Cayouette-Remblière 2011 ; Zalc et Lemerancier 2007)
- ▶ Définir les contours de la population = pouvoir la dénombrer

I. Soumettre un cas ethnographique à l'analyse quantitative

II. Le cas contre l'inférence : le problème de la significativité statistique

III. La guerre des étoiles n'aura pas lieu : quantifier sans inférer

Une base de données fidèle au cas



NICOLAS MARIOT
CLAIRE ZALC

FACE À LA PERSÉCUTION

991 JUIFS DANS LA GUERRE



Fondation pour la Mémoire de la Shoah

Une base de données fidèle au cas

- ▶ 101 candidatures
- ▶ Une fois la population définie, échantillonner (population de grande taille : Perdoncin, collectif Onze, Bruno) ou travailler sur la population entière (Zalc et Mariot, Cayouette Remblière)

Un grain de sable dans ma "statistique ethnographique"

	Avis défavorable	Avis favorable	total
Aucun diplôme (ou inconnu)	23	35	58
CAP ou diplôme supérieur	3	40	43
Ensemble	26	75	101

χ^2 significatif au seuil de 1% ($p=0.0045$).

I. Soumettre un cas ethnographique à l'analyse quantitative

II. Le cas contre l'inférence : le problème de la significativité statistique

III. La guerre des étoiles n'aura pas lieu : quantifier sans inférer

I. Soumettre un cas ethnographique à l'analyse quantitative

II. Le cas contre l'inférence : le problème de la significativité statistique

III. La guerre des étoiles n'aura pas lieu : quantifier sans inférer

Le rôle de l'inférence statistique dans les enquêtes sur échantillon

- ▶ Parler du tout à partir d'une partie
- ▶ Mesurer les aléas d'échantillonnage : la formalisation du hasard par les probabilités
- ▶ La condition du raisonnement inférentiel : disposer d'un échantillon probabiliste, ou aléatoire (Ardilly 2006)
- ▶ Déterminer précisément la marge d'erreur des résultats issus de l'échantillon

La logique du test s'oppose à la logique du cas

- ▶ Tests (khi-deux par exemple) et p -value
- ▶ Seuils (0.1 , 0.05, 0.01 etc.) et étoiles (*, **, ***)

La logique du test s'oppose à la logique du cas

- ▶ Une hypothèse (que l'on voudrait réfuter) : "Dans la population mère, il n'existe aucune relation entre les variables étudiées".
- ▶ Une condition : l'échantillonnage aléatoire
- ▶ Une question : Cette indépendance étant supposée dans la population, quelle est la probabilité de tirer au hasard un échantillon dans lequel on observe une corrélation au moins aussi marquée ?
- ▶ Une réponse : la p -value

La logique du test s'oppose à la logique du cas

- ▶ Un échantillon permet d'approcher la "vraie valeur", celle que l'on connaîtrait si l'on disposait de toute l'information (la population entière)
- ▶ Une population construite comme un cas donne la valeur *dans ce cas*.

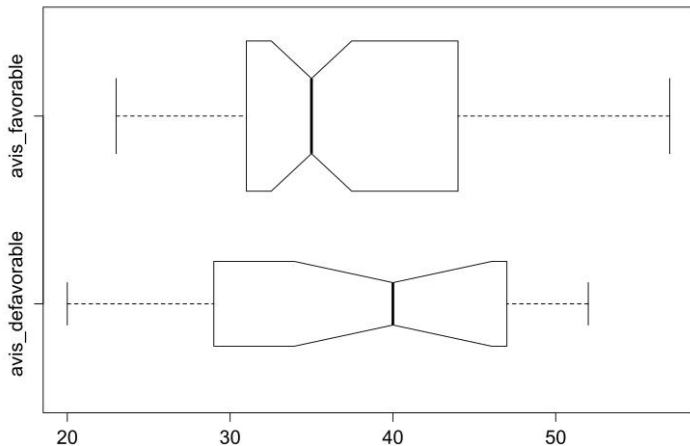
La logique du test s'oppose à la logique du cas

- ▶ Mais cette population singulière ne peut-elle pas être considérée comme l'échantillon d'une population plus large ?
- ▶ Premier problème : cet échantillon n'est pas aléatoire
- ▶ Deuxième problème : on annule les traits singuliers qui font le cas
- ▶ Troisième problème : on s'expose aux critiques (justifiées) sur la qualité de l'échantillon, qui ici souffre par nature de multiples biais incontrôlés.

Les tests affaiblissent les résultats plus qu'ils ne les renforcent

- ▶ Généraliser abusivement
- ▶ Passer à côté d'effets parce qu'ils ne sont pas significatifs.

Age et sélection : quelle relation ?



Les tests affaiblissent les résultats plus qu'ils ne les renforcent

- ▶ Le résultat du test est directement lié à la taille de l'échantillon (et non de la population)
 - ▶ L'ethnographe a plus de chance de rencontrer des tests non significatifs parce qu'il travaille sur des petites populations
 - ▶ La posture prudente : renoncement et déception
 - ▶ La posture erronée : conclure qu'il n'y a pas d'effet
-
- ▶ Risque énorme de reprendre à son compte les erreurs classiques et extrêmement fréquentes dans l'usage des tests, y compris chez les quantitativistes (Chakhaia et Leopold 2017).

I. Soumettre un cas ethnographique à l'analyse quantitative

II. Le cas contre l'inférence : le problème de la significativité statistique

III. La guerre des étoiles n'aura pas lieu : quantifier sans inférer

Une posture descriptive face aux statistiques

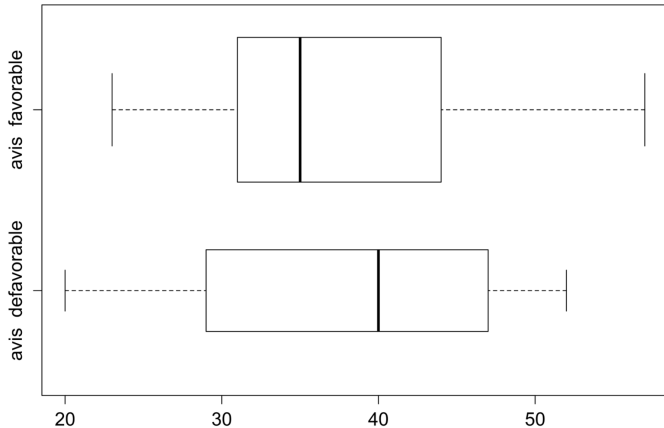
Distinguer techniques d'enquête et techniques d'analyse

- ▶ Les techniques d'enquête (enquête de terrain localisée, enquête sur échantillon) dictent leurs principes aux techniques d'analyse (quantitatives, qualitatives). Cf. (Bugeja-Bloch et Couto, 2015).
- ▶ Ici, l'étude de cas comme technique d'enquête et le quantitatif comme technique d'analyse.

Privilégier une approche descriptive des statistiques

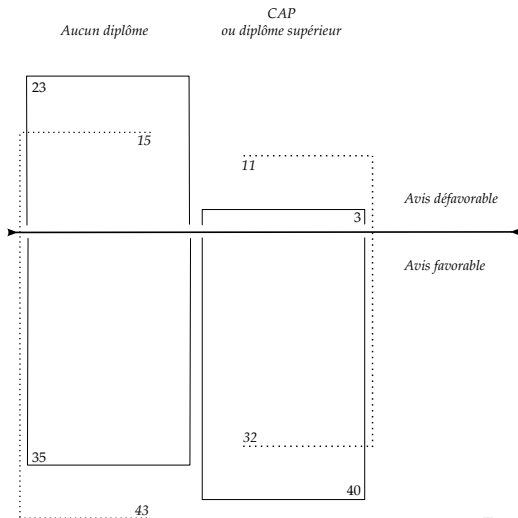
Objectif : Raisonnement, à l'aide d'outils statistiques, sur une population dans laquelle se déroulent des phénomènes que l'on souhaite expliquer.	Statistique inférentielle : L'ensemble des individus de la base de données constitue un échantillon aléatoire de la population que l'on cherche à décrire. On ne connaît pas la vraie valeur, on cherche à l'estimer. L'échantillon n'existe pas en dehors du protocole expérimental – il ne constitue pas un groupe social.
	Statistique descriptive : L'ensemble des individus de la base de données constitue la population que l'on cherche à décrire. On connaît la vraie valeur. La population a un sens en elle-même en dehors de l'expérience, elle est l'objet de l'analyse.

Age et sélection : quelle relation ?



Privilégier une approche descriptive des statistiques

Niveau de diplôme dans le secteur forestier



Une parenthèse pour finir : Indépendance et aléa

- ▶ On compare à une situation d'indépendance pour apprécier la force de l'attraction des variables, de la même manière que l'on calcule la distance à l'indépendance lors du test du khi-deux
- ▶ Mais c'est une comparaison qui ne fait pas intervenir les probabilités
- ▶ Sauf si l'on cherche à prouver que les agents n'ont pas agité aléatoirement...
- ▶ Mais n'est-ce pas une hypothèse invérifiable et métaphysique que celle d'un hasard fondamental qu'il faudrait mesurer et annuler ?
- ▶ Une posture descriptive semble épistémologiquement plus légère et prudente

Conclusion

Quelques intérêts :

- ▶ Ne pas entretenir l'ambiguïté qui entoure le terme "significatif".
- ▶ Éviter les généralisations abusives et éviter de passer à côté d'effets sociologiquement importants
- ▶ Encourager la recherche d'outils d'analyse des relations statistiques qui ne soient pas inférentiels
- ▶ Ne pas contribuer à l'usage automatique et abusif des tests, contre lequel se battent de plus en plus de quantitativistes, dans toutes les disciplines
- ▶ Une question qui concerne des quantitativistes purs qui travaillent sur des populations : recensements, bid data (et leur fameux "n=all"), données administratives.
- ▶ Raisonner en ethnographe et non, inconsciemment, en sondeur.

Beaud S. (1996), « L'usage de l'entretien en sciences sociales. Plaidoyer pour l'«entretien ethnographique» », Politix, Vol. 9, n.35, pp. 226-257.

Bruno A.S. (2010), Les chemins de la mobilité : migrants de Tunisie et marché du travail parisien depuis 1956, Paris, Éd. de l'EHESS.

Bruno A.S., Rygiel P., Spire A. et Zalc C. (2007), « Jugés sur pièces. Le traitement des dossiers de séjour et de travail des étrangers en France (1917-1984) », Population, vol.61, n.5, pp. 737-762.

Bugeja-Bloch F et Couto M.P. (2015), Les méthodes quantitatives, Paris, PUF « Que sais-je ? ».

Bernardi, F., Chakhaia L. et Leopold L. (2017), «“Sing Me a Song with Social Significance” : Te (Mis)Use of Statistical Significance Testing in European Sociological Research», European Sociological Review, vol. 33, no 1

Cayouette-Remblière J., (2011), « Reconstituer une cohorte d'élèves à partir de dossiers scolaires. La construction d'une statistique ethnographique. », Genèses vol.4, n.85 , p. 115-133.

Dumez H. (2015), « What Is a Case, and What Is a Case Study ? », BMS, vol. 127, n.1.

Lemerrier C. et Zalc C. (2007), Méthodes quantitatives pour l'historien, Paris, la Découverte.

Mariot N. et Zalc C. (2010), Face à la persécution. 991 Juifs dans la guerre, Paris, Odile Jacob.

Passeron J.C. et Revel J. (2005) Penser par cas, Paris, École des Hautes Études en Sciences Sociales.

Ziliak S. T. et McCloskey D.N. (2007), The Cult of Statistical Significance : How the Standard Error Costs Us Jobs, Justice, and Lives, University of Michigan.