

Définir, recueillir, décider (2)

Les types de données

- Plusieurs types de données:
 - ▶ Données de type panel: les mêmes personnes interrogées à plusieurs reprises
 - Sur leur situation avant l'entrevue
 - Sur leur situation au moment de l'entrevue
 - Sur leur situation entre les moments d'entrevue
 - ▶ Données de type longitudinal comprenant plusieurs échantillons
 - Qui peuvent avoir été recueillis indépendamment (plusieurs sondages auprès d'échantillons différents mais certaines questions identiques)
 - Qui peuvent être recueillies sous forme de sondage roulant (rolling cross-section): on met sur le terrain un nouveau sous-échantillon à chaque jour (pendant une campagne électorale par exemple).



4

Définir, recueillir, décider (2)

Les types de données (suite)

- ▶ Données de type archive
 - Statistiques institutionnelles,
 - Données économiques, taux de chômage, évolution du PIB, des salaires, etc.
 - Données sur les taux de criminalité, de mortalité, etc.
- ▶ Et même...
 - Données sur les interactions entre deux personnes pendant une entrevue ou une session de thérapie



5

Les problèmes associés

Deux problèmes principaux

- Dépendance des données dans le temps, autocorrélation
 - ▶ Ce qui fait que le postulat de base des analyses de régression ordinaire n'est pas respecté à sa face même. Ceci biaise l'estimation de la variance.
- Données manquantes à certains moments, censure et troncature
 - ▶ Personnes qui "disparaissent", absence d'informations pour certaines unités à certains moments: il faut connaître les raisons de l'absence de données et s'assurer que cette absence est aléatoire, que ça ne biaise pas les analyses



6

Que faire? (1)

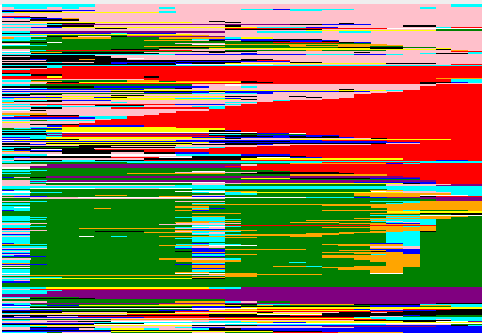
D'abord et avant tout décrire...Visualiser aide à se représenter la situation. A cette étape, on peut repérer des problèmes

- Graphique des transitions (...) (DeGenne, LeBourdais, Renaud, etc. pour exemples)
- Tapis (ex: Degenne (site du CIQSS): description de l'évolution de la situation des finissants)
- Tables de survie (Renaud, Durand, etc.)
- Graphiques de séries chronologiques (Durand, Laroche et Blais, 2005; Durand, 2008, 2011)
- Régressions locales (Durand, à partir de 2014).
- Classifications de trajectoires (Durand et Lacourse; Durand, Pelletier, Wutchiett)



7

Activité post d.u.t (2 ans) De Genne (2003)



Rose = CDD; Rouge = CDI; Violet = alternance; Bleu = intérim;
Orange = stage; Jaune = Service national; Noir = chômage; Blanc = autre situation;
Vert = études ; Cyan = non réponse ou inactivité



8

Que faire (2)

Quelques décisions importantes

- Quel est, quels sont, les événements d'intérêt?
- Qu'est-ce qui donne la mesure du temps? (jours, mois, années, essais)
- Le temps est-il discret ou continu?
- Les événements qui prédisent ou expliquent la variable dépendante se modifient-ils dans le temps?
- Quelle est la forme de l'évolution dans le temps?

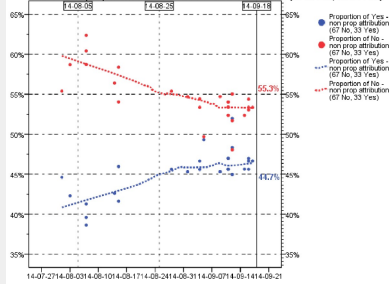


9

Évolution de l'intention de vote Référendum Écosse 2014

après répartition non proportionnelle des discrets

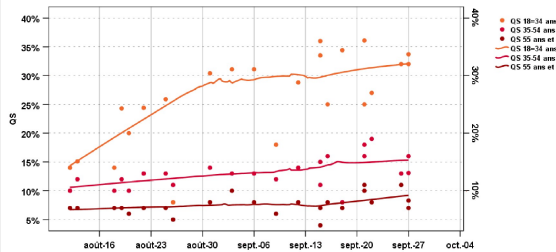
Evolution of support for Scottish Independence since beginning of August 2014 - non proportional attribution of non disclosers (67% NO, 33% Yes)



16

Évolution de l'intention de vote pour Québec Solidaire, selon l'âge (2018)

Appui à Québec Solidaire, depuis août 2018, par groupe d'âge

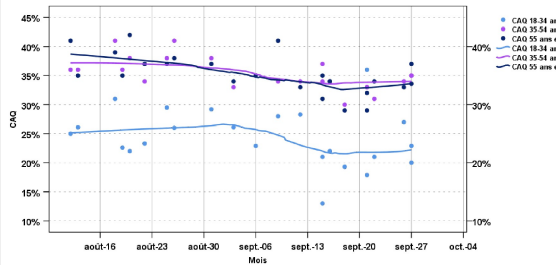


Chaque point représente un sondage positionné dans le mois où il a été effectué. Les lignes représentent les estimations faites par régression locale (Loess) utilisant Epanchukov SS, B.C. Turkel, 2018.

17

Évolution de l'intention de vote pour la Coalition Avenir Québec, selon l'âge (2018)

Appui à la Coalition Avenir Québec, depuis août 2018, par groupe d'âge

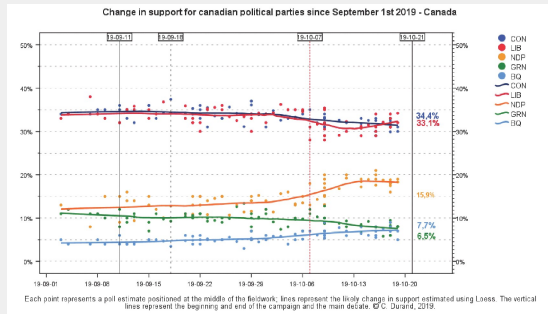


Chaque point représente un sondage positionné dans le mois où il a été effectué. Les lignes représentent les estimations faites par régression locale (Loess) utilisant Epanchukov SS, B.C. Turkel, 2018.

18

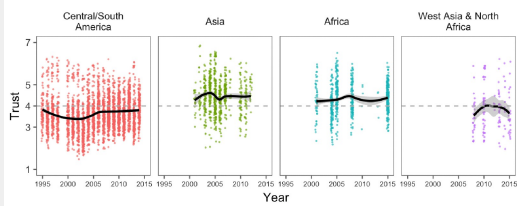
Évolution de l'intention de vote - Canada 2019 - selon les sondages publiés

Régressions locales



19

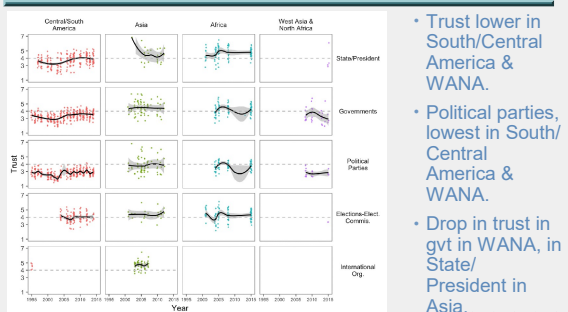
Une vue synthétique de l'évolution de la confiance institutionnelle par région du monde.



- En moyenne, la confiance est stable
- Plus élevée en Asie et en Afrique qu'en Amérique latine.
- A diminué beaucoup depuis 2011 en Afrique du Nord & Asie de l'Ouest.

20

Confiance dans les institutions politiques.



- Trust lower in South/Central America & WANA.
- Political parties, lowest in South/Central America & WANA.
- Drop in trust in gvt in WANA, in State/President in Asia.

21

Que faire? (5)

Multiples analyses disponibles selon la situation

■ Analyses multi-niveaux longitudinales

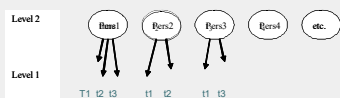
- ▶ Dans ce cas, le temps est considéré comme un niveau: les diverses mesures prises sont "nichées" dans les individus qui peuvent eux-mêmes être nichés dans des unités (classes, équipes de travail, familles,...).
- ▶ L'intérêt est la flexibilité de la méthode, entre autres parce qu'il n'est pas obligatoire d'avoir des mesures à chaque moment et au même moment pour tous les sujets.
- ▶ Il faut que le niveau supérieur (2 ou 3) soit un échantillon (n> 40).
- ▶ Exemple: évolution de l'emploi durant un certain temps, évolution de la confiance institutionnelle dans le temps pour un certain nombre de pays (échantillon de pays ou de périodes).



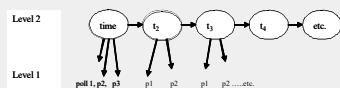
22

Modèle multiniveaux longitudinal

- Le temps peut être au niveau 1, par exemple, personnes (niveau 2) avec mesures prises à divers moments (niveau 1)



Le temps peut être au niveau 2, par exemple, mois (niveau 2) avec résultats de sondages faits à chaque mois (niveau 1)



23

Évolution de la confiance institutionnelle

Variable	Trust in institutions - basic models			
	Model 0	Model 1	Model 2	Model 3
Intercept	1.1233***	4.43***	2.06***	0.52***
Level Measure				
Work (SD)				
CEPR	0.28***	0.23***	0.22***	0.23***
Public Admin.	0.28***	0.23***	0.22***	0.23***
CEC - Civil Society	0.28***	0.23***	0.22***	0.23***
Env. quality	0.28***	0.23***	0.22***	0.23***
Public Admin.	0.28***	0.23***	0.22***	0.23***
Industry	0.28***	0.23***	0.22***	0.23***
Finance	0.28***	0.23***	0.22***	0.23***
Education	0.28***	0.23***	0.22***	0.23***
Level Respondent				
Age	0.0016	0.0016	0.0016	0.0016
Young (less than 35)	0.0017	0.0017	0.0017	0.0017
Sex	0.0017	0.0017	0.0017	0.0017
Civil ID (poll)	0.0017	0.0017	0.0017	0.0017
Wgt	0.0017	0.0017	0.0017	0.0017
Prog. Res. inst.	0.0017	0.0017	0.0017	0.0017
Level Country Year				
Time	0.0017	0.0017	0.0017	0.0017
Cont.	0.0017	0.0017	0.0017	0.0017
Level Country cluster				
Cont (North America)	0.0017	0.0017	0.0017	0.0017
Lat.	0.0017	0.0017	0.0017	0.0017
W. Asia & Africa	0.0017	0.0017	0.0017	0.0017
Europe (incl. UK)	0.0017	0.0017	0.0017	0.0017
Variances				
Intercept	2.735	42.91	1.945	0.738
Work (SD)	0.007	0.007	0.007	0.007
CEPR	0.007	0.007	0.007	0.007
Public Admin.	0.007	0.007	0.007	0.007
CEC (Civil Society)	0.007	0.007	0.007	0.007
Env. quality	0.007	0.007	0.007	0.007
Public Admin.	0.007	0.007	0.007	0.007
Industry	0.007	0.007	0.007	0.007
Finance	0.007	0.007	0.007	0.007
Education	0.007	0.007	0.007	0.007
Age	0.007	0.007	0.007	0.007
Young (less than 35)	0.007	0.007	0.007	0.007
Sex	0.007	0.007	0.007	0.007
Civil ID (poll)	0.007	0.007	0.007	0.007
Wgt	0.007	0.007	0.007	0.007
Prog. Res. inst.	0.007	0.007	0.007	0.007
Time	0.007	0.007	0.007	0.007
Cont.	0.007	0.007	0.007	0.007
Cont (North America)	0.007	0.007	0.007	0.007
Lat.	0.007	0.007	0.007	0.007
W. Asia & Africa	0.007	0.007	0.007	0.007
Europe (incl. UK)	0.007	0.007	0.007	0.007



24

Focus sur les niveaux 2 & 3

Trust in institutions - basic models			
	Model 2	Model 3	
Intercept	4,290 ***	4,452 ***	
Level Respondent			
woman	-0,004 ns	-0,004 ns	
Young (Less than 30) time	0,009 **	0,009 **	
Old (60 plus) time	0,095 ***	0,094 ***	
Prop_Non-resp.	0,003 ***	0,004 **	
Level Country-Year			
Time			0,001 ns
Time2			0,001 *
Variance			
Measures	2,380	60,7%	2,380 60,7%
Respondents	1,109	28,3%	1,109 28,3%
Country-Year	0,109	2,8%	0,107 2,7%
Country-Source	0,324	8,3%	0,327 8,3%
Total	3,922		3,924
	92%		92%

- Individual level:
 - ♦ Sex is not significant
 - ♦ Compared to middle age:
 - being less than 30: +.009;
 - being 60+: +.094 .
 - ♦ Prop. Non-response: +.003.
- Niveau pays-année:
 - ♦ Le temps au carré est significatif.
- Variance explained: minimal



25

Que faire? (6)

Multiples analyses disponibles selon la situation

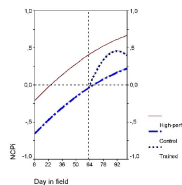
- Analyse des trajectoires
 - Il s'agit de faire une classification des trajectoires "individuelles" -- ça peut être des individus mais aussi des groupes, des pays, etc. -- pour en arriver à des regroupements de parcours.
 - Méthode en développement, relativement récente mais en voie d'être intégrée dans les principaux logiciels.
 - Exemple: Les trajectoires de délinquance entre l'âge de 5 ans et de 18 ans, au moyen de mesures similaires prises à divers moments durant cette période.
 - Problème: prédire le passé avec le futur.



26

Analyses de trajectoires, avant et après formation, avec groupe contrôle (bleu)

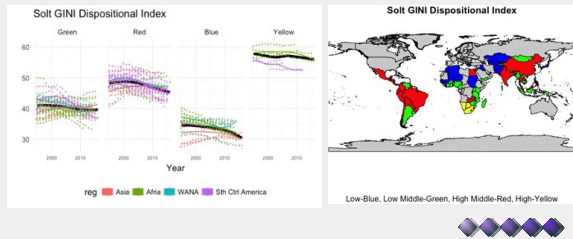
Trajectoires de performance (NCPI) des interviewers.
 Groupe rouge: bonne performance, non formés
 Groupe bleu: faible performance, non formés
 Groupe gris: faible performance, formés



27

Analyse de trajectoires de mesures (K-means long.)

L'évolution de certaines mesures dans le temps:
http://www.mapageweb.umontreal.ca/durandc/Recherche/Publications/confiance/WAPOR2018_CD.pdf



28

Que faire? (7)

Multiplés analyses disponibles selon la situation

- Analyse de variance pour mesures répétées
 - Analyse traditionnelle en psychologie. D'une certaine manière, c'est la base.
 - L'idée est de faire une intervention et de mesurer la variable dépendante à divers moments fixés pour analyser l'évolution entre les moments.
 - Exemple :
 - Mesure de la dépression à divers moments durant un processus thérapeutique, évolution de la performance en mathématique avant et après une intervention de remédiation, etc.
 - Problème: On ne peut garder que les cas pour lesquels on a de l'information à **tous** les temps de mesure.

29

Quelle forme prend l'évolution dans le temps?

- Pour la plupart des analyses, il est très important de se demander quelle forme prend l'évolution dans le temps en soi ou suite à un événement (voir Effet "Bouchard" pendant la campagne référendaire de 1995 au Québec, Durand, 2008)
- L'évolution peut être simplement linéaire mais elle peut aussi être quadratique en "U", cubique (en forme de dos de poisson), voir Applied Longitudinal Data Analysis, Singer & Willett, 2003.
- Un événement peut provoquer un saut. Il peut aussi provoquer une modification de l'évolution: accélération, plafonnement, etc.

30

Avantages et inconvénients

Interventions sur les données

- Dans les analyses de ce type, il est souvent nécessaire de faire des interventions sur les fichiers
 - Pour les analyses de survie (tables, régressions de survie), il faut "rectangulariser" le fichier (voir procédure restructurer de SPSS) (voir travail sur les fichiers http://www.mapageweb.umontreal.ca/durandc/menuMethodesQuantitatives.htm#travail_fichier).
 - Il faut parfois créer des variables qui indiquent le moment où un événement survient.
 - Pour les analyses multi-niveaux: avec HLM, il faut faire un fichier par niveau (procédures Agréger ou Restructurer dans SPSS)



34

Choix des logiciels

- Le logiciel STATA est probablement le plus approprié pour les régressions de survie mais SPSS réussit généralement à faire la même chose. Stata a des modèles de risques simultanés (competing risks): équivalent à logistique multinomiale de survie.
- Pour les analyses multi-niveaux, les logiciels spécifiques sont HLM, MLWin, M+ et R et dans une moindre mesure Lisrel, Stata, SPSS, SAS.
- Le transfert de bases de données d'un logiciel à un autre est habituellement facile. Les logiciels spécifiques lisent les fichiers de SPSS, STATA, R ou SAS.



35

Conclusion

- Au départ, il peut être plus difficile de travailler sur les fichiers pour pouvoir faire les analyses appropriées, MAIS
- Le jeu en vaut la chandelle. Une fois la base de données créée, tout devient nettement plus simple.
- Ne pas oublier que la première étape est de décrire, de visualiser.



36
