

Analyse d'ordres de priorités

Thomas Perneger

Quelles devraient être les priorités pour la recherche clinique à Genève?

- A. Maladies cardiovasculaires
- B. Bases génétiques des maladies
- C. Maladies de l'âge avancé
- D. Tests de nouveaux médicaments
- E. Déterminants sociaux des problèmes de santé

- Veuillez inscrire 1 à côté de votre option préférée, 2 à côté de la suivante, etc, et 5 à côté de la moins souhaitable

Pourquoi s'intéresser aux priorités

- Implication du public dans les décisions:
 - Patients
 - Population générale
 - Collaborateurs
- Domaines:
 - Nouveaux programmes de santé/prévention
 - Couverture d'assurance obligatoire
 - Recherche
 - ...

Méthodes diverses

- Evaluations indépendantes de plusieurs options sur la même échelle
 - Utilités
 - Importance, bénéfice attendu
- Choix entre plusieurs options :
 - Election
 - **Ordre de priorité**
 - Discrete choice experiment, conjoint analysis
 - Pie sharing (budget à répartir)

Elections

- But habituel: sélectionner 1 gagnant dans le respect de certains principes
- Énorme littérature sur ce sujet (théorie du choix social, K Arrow)
- Ordre de priorité: permet de simuler des tours multiples avec élimination = « instant run-off »
 - a) un candidat a >50% de 1ers rangs? → stop
 - b) Sinon, éliminer le candidat avec le moins de 1^{ers} rangs et décaler les 2èmes rangs en 1^{ère} position
 - c) Retour à a)

Instant run-off: exemple

- Préférences réelles pour Bush, Clinton, Nader
 - BCN 40%
 - BNC 5%
 - CNB 40%
 - CBN 5%
 - NCB 8%
 - NBC 2%

Bush 45%

Clinton 45%

Nader 10%
- Avec un seul candidat sélectionné, les électeurs de Nader risquent de voter stratégiquement pour leur 2^{ème} choix
- Avec IRO, pas besoin de vote « utile », le résultat final est Bush 47%, Clinton 53%

Evaluations

- Applicables à de nombreux objets
- Tâche simple pour le répondant
- Variables faciles à analyser
- Sensibles au « response set »
- Souvent peu de contrastes

Ordres de priorité

- Tâche plus difficile cognitivement
- Nombre d'objets à classer doit être modeste
- Dépendance (négative) des rangs
- Ordre connu mais pas l'ampleur des différences
- Échelle relative (pas absolue)
- Oblige le répondant à se positionner
- Contrastes plus manifestes entre les objets

Exemple

- Enquête auprès de femmes genevoises éligibles pour une mammographie de dépistage
- Enquête 2012 (mais idem en 1998)
- On leur demande de classer par ordre de priorité 6 programmes de santé publique
- 924 ont participé (ordre complet, sans égalité)

10. Les ressources étant limitées, il est utile de se demander ce que l'on aimerait faire en **priorité**. Dans cette perspective, veuillez classer les six mesures de prévention suivantes selon l'ordre d'importance que vous leur accordez. Notez le chiffre 1 pour la mesure qui vous paraît la plus importante, le chiffre 2 pour la deuxième mesure la plus importante et ainsi de suite jusqu'au chiffre 6 pour la mesure la moins importante.

	Votre classement
a) Les campagnes de prévention du SIDA
b) Les programmes de dépistage du cancer du sein
c) Les programmes de dépistage de la dépression
d) Les campagnes antitabac auprès des jeunes
e) Les programmes de dépistage du cancer de la prostate
f) Les campagnes contre l'alcool au volant

Distributions des rangs 2012

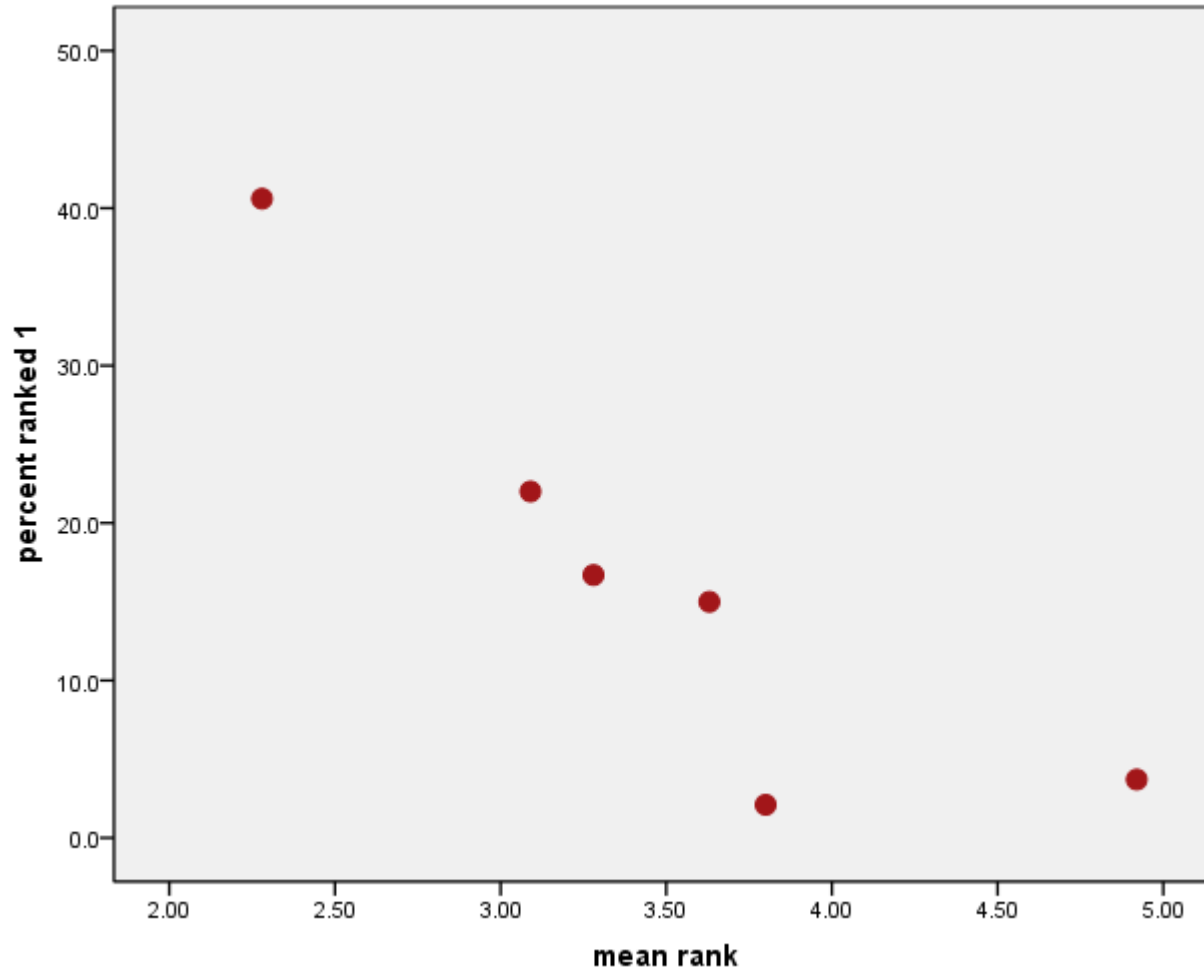
Rang de priorité

2012	1	2	3	4	5	6
Breast cancer screening	40.6	21.2	16.6	14.6	5.5	1.5
AIDS prevention	22.0	15.7	24.6	14.1	16.5	7.3
Campaigns against drunk driving	16.7	17.9	18.2	22.7	16.8	7.8
Prostate cancer screening	2.1	25.0	17.0	18.9	21.0	16.0
Prevention of smoking in youth	15.0	15.3	14.3	17.5	23.2	14.7
Screening for depression	3.7	5.0	9.4	12.1	17.1	52.7

Rang moyen et % de premiers rangs

2012	Mean rank (SD)	Proportion as 1 st choice
Breast cancer screening programs	2.28 (1.35)	375 (40.6)
AIDS prevention programs	3.09 (1.58)	203 (22.0)
Campaigns against drunk driving	3.28 (1.54)	154 (16.7)
Campaigns against teenage smoking	3.63 (1.67)	139 (15.0)
Prostate cancer screening programs	3.80 (1.47)	19 (2.1)
Depression screening programs	4.92 (1.43)	34 (3.7)

Corrélation du rang moyen avec proportion de rang=1



Sein versus dépression

p_sein * p_depression Crosstabulation

Count

		p_depression						Total
		1	2	3	4	5	6	
p_sein	1	0	24	35	52	60	204	375
	2	13	0	21	26	31	105	196
	3	9	5	0	11	41	87	153
	4	4	11	19	0	25	76	135
	5	6	5	8	17	0	15	51
	6	2	1	4	6	1	0	14
Total		34	46	87	112	158	487	924

Sein versus prostate

p_sein * p_prostate Crosstabulation

Count

		p_prostate						Total
		1	2	3	4	5	6	
p_sein	1	0	214	49	48	39	25	375
	2	14	0	92	37	24	29	196
	3	2	13	0	78	37	23	153
	4	0	3	12	0	85	35	135
	5	1	0	2	12	0	36	51
	6	2	1	2	0	9	0	14
Total		19	231	157	175	194	148	924

Corrélations (Spearman)

2012	Breast cancer screening	AIDS prevention	Against drunk driving	Prostate cancer screening	Prevention of smoking in youth	Screening for depression
Breast cancer screening	1	-0.25	-0.50	0.55	-0.47	-0.09
AIDS prevention		1	-0.14	-0.25	-0.24	-0.15
Campaigns against drunk driving			1	-0.44	0.16	-0.16
Prostate cancer screening				1	-0.45	-0.25
Prevention of smoking in youth					1	-0.19
Screening for depression						1

Sous H_0 $r = -1/(k-1)$

Analyse factorielle (varimax)

Rotated Component Matrix^a

	Component		
	1	2	3
p_sida	-.049	-.984	-.097
p_sein	.840	.187	-.121
p_depression	-.009	.093	.987
p_tabac	-.689	.385	-.119
p_prostate	.794	.233	-.324
p_alcool	-.690	.115	-.276

Extraction Method: Principal Component Analysis.

Tests statistiques: comparaison de paires de priorités successives

2012	Mean rank (SD)	P versus next Wilcoxon	Proportion as 1 st choice	P versus next McNemar
Breast cancer screening programs	2.28 (1.35)	<0.001	375 (40.6)	<0.001
AIDS prevention programs	3.09 (1.58)	0.012	203 (22.0)	0.011
Campaigns against drunk driving	3.28 (1.54)	<0.001	154 (16.7)	0.41
Campaigns against teenage smoking	3.63 (1.67)	0.04	139 (15.0)	<0.001
Prostate cancer screening programs	3.80 (1.47)	<0.001	19 (2.1)	0.053
Depression screening programs	4.92 (1.43)	-	34 (3.7)	-

Comparer toutes les paires

- Ex: k objets A, B, C, D,... à classer
- $k(k-1)/2$ comparaisons de paires
- $k!$ ordres possibles (ex BADC, ACDB, ...)

- Note
 - La préférence est transitive pour un individu
 - La majorité de préférences n'est pas transitive dans une population

Proportion (%) qui classe programme de la rangée plus haut que programme colonne



2012	Breast cancer screening	AIDS prevention	Against drunk driving	Prostate cancer screening	Prevention of smoking in youth	Screening for depression
Breast cancer screening	-	61.5	63.9	92.1	66.8	88.0
AIDS prevention	38.5	-	55.0	57.9	59.0	80.5
Campaigns against drunk driving	36.1	45.0	-	54.2	58.0	78.1
Prostate cancer screening	7.9	42.1	45.8	-	50.4	73.9
Prevention of smoking in youth	33.2	41.0	42.0	49.6	-	71.5
Screening for depression	12.0	19.5	21.9	26.1	28.5	-

Odds ratios de McNemar

Preference for row over column, 2012	AIDS prevention	Against drunk driving	Prostate cancer screening	Prevention of smoking in youth	Screening for depression
Breast cancer screening	1.59	1.76	11.68	2.02	7.35
AIDS prevention		1.22	1.88	1.44	4.15
Campaigns against drunk driving			1.18	1.38	3.59
Prostate cancer screening				(1.02)	2.84
Prevention of smoking in youth					2.52

(): non significatif

Analogie avec systèmes d'élection du 18^{ème} siècle français

- Condorcet
 - On considère toutes les paires de candidats
 - Celui qui défait chacun des opposants est un « vainqueur selon Condorcet »
 - Principe n'est pas transitif (pierre-caillou-ciseau)
- Borda
 - Si k candidats, le premier reçoit k points, le second $k-1$, etc, donc $\text{points} = k - \text{rang}$
 - (alternative à Nauru: $\text{points} = 1/\text{rang}$)
 - Celui qui a le plus de points gagne (=nb de victoires en confrontation de chaque paire d'options chez chaque votant)
 - Équivalent au rang de priorité moyen le plus bas

Quel système choisir?

- Arguments politiques/philosophiques
- Condorcet:
 - « Winner takes all »
 - Système majoritaire
 - Vainqueur Condorcet n'existe pas toujours!
- Borda:
 - Chaque voix compte
 - Estime des « distances » entre les options
 - Système proportionnel
 - Est « sensible aux alternatives non-relevantes »

- 5 votants, 4 candidats:

Ordre votant 1: A C D B

Ordre votant 2: A C D B

Ordre votant 3: A C D B

Ordre votant 4: B C D A

Ordre votant 5: B C D A

- Vainqueur Condorcet: **A**
bat chaque autre candidat par 3 à 2
- Vainqueur Borda: **C** (10 points)
devant A (9), B (6) et D (5)
- Excluant D, Vainqueur Borda: **A** (6 points)
devant C (5) et B (4)

Ramon Llull, 13^{ème} siècle

5. De arte electionis

(Sankt Nikolaus-Hospital / Cusanusstift, Cod. Cus. 83, f. 47v-48r)

<i>bc</i>	<i>cd</i>	<i>de</i>	<i>ef</i>	<i>fg</i>	<i>gh</i>	<i>hi</i>	<i>ik</i>
<i>bd</i>	<i>ce</i>	<i>df</i>	<i>eg</i>	<i>fh</i>	<i>gi</i>	<i>hk</i>	
<i>be</i>	<i>cf</i>	<i>dg</i>	<i>eh</i>	<i>fi</i>	<i>gk</i>		
<i>bf</i>	<i>cg</i>	<i>dh</i>	<i>el</i>	<i>fk</i>			
<i>bg</i>	<i>ch</i>	<i>di</i>	<i>ek</i>				
<i>bh</i>	<i>ci</i>	<i>dk</i>					
<i>bi</i>	<i>ck</i>						
<i>bk</i>							

Désignation de l'évêque, mère supérieure, etc

Précurseur de Condorcet: gagne celui qui a le plus de victoires en confrontation directe

Nicolaus Cusanus, 15^{ème} siècle



Le premier à compter les pulsations
(pouls radial)

Mais aussi:
A exigé que les juifs portent un insigne
qui les identifie

Système de désignation de l'empereur
Précurseur de Borda: gagne celui qui a le plus de votes

Rank-ordered logit

- Rappel: régression logistique conditionnelle = modèle de Cox avec évènement = statut de cas
- Extension possible à un ordre de priorité?
- « Observation » = choix de priorité pour un programme par une personne (k par personne)
- « Cluster » = répondant
- « Evènement » = sélection comme xème priorité
- « Temps » = rang de priorité
- « Hazard » = probabilité d'être choisi en xème position
- « Relative hazard » = priorité relative

ASSESSING THE POTENTIAL DEMAND FOR ELECTRIC CARS*

S. BEGGS and S. CARDELL

Charles River Associates

J. HAUSMAN

Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received March 1980, final version received April 1981

An ordered logit specification for use on ranked individual data is used to analyze survey data on potential consumer demand for electric cars. In many situations in economics and marketing we would like to be able to forecast consumer demands for goods which have not yet appeared in actual markets. By defining goods as a bundle of underlying attributes, we can use discrete choice models to estimate consumer evaluations. Then new good demand is forecast by use of the estimated coefficients to compare consumer evaluation of the new good to existing choices. When ranked individual data are available, we can estimate separate coefficients for each individual rather than assuming identical coefficients as is usual with logit models. Our results indicate considerable dispersion in individual coefficients. This finding can have important implications for new product analysis.



LOGIT MODELS FOR SETS OF RANKED ITEMS

*Paul D. Allison**

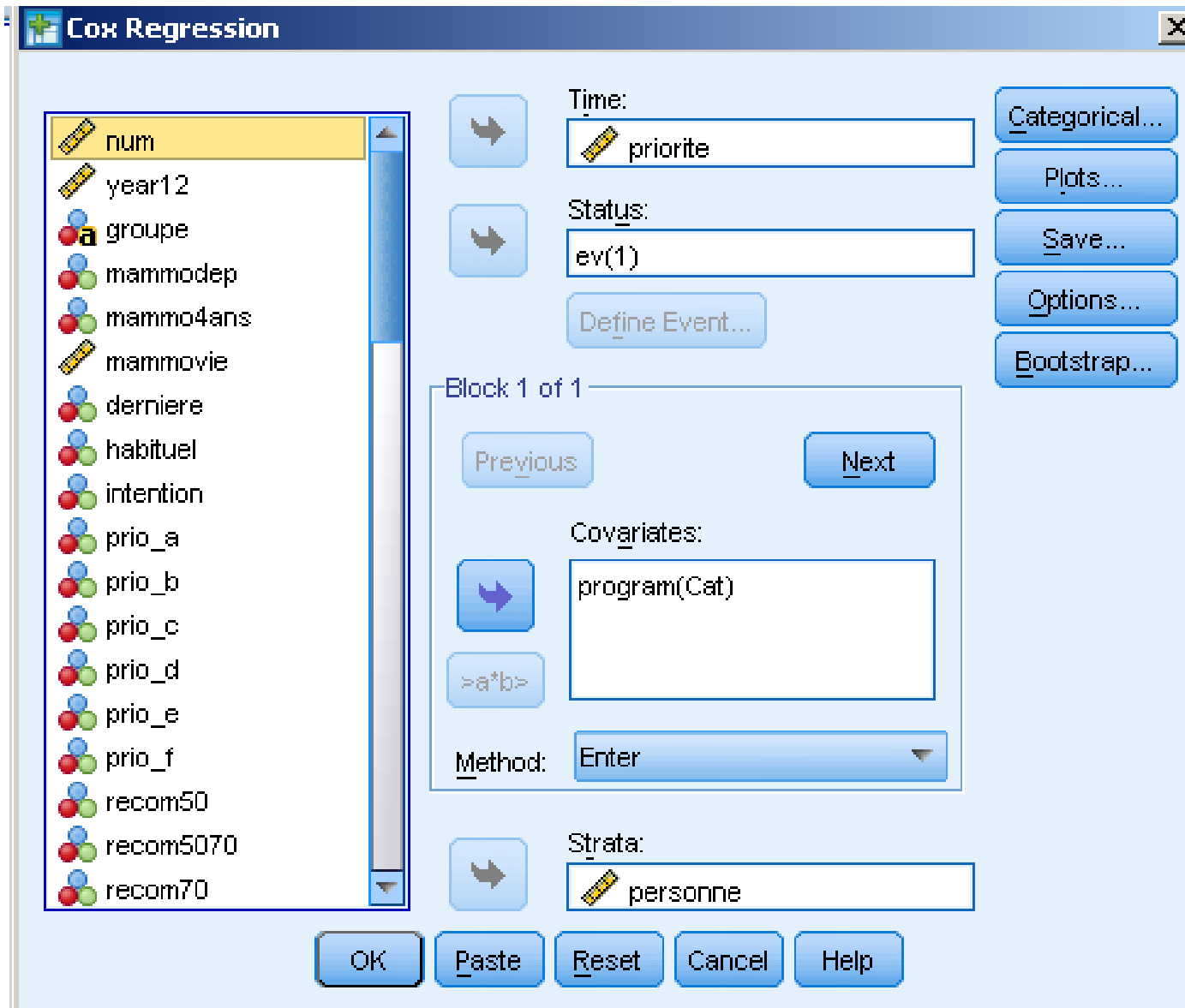
*Nicholas A. Christakis**

Methods are presented for analyzing data generated by asking respondents to rank a set of items. Based on a conditional logit model, these methods allow us to estimate and test for differences among items in respondents' preferences for them; to test for differences in item preferences across subpopulations; and to incorporate predictor variables describing respondents, items, or both. The models can be easily estimated with programs for proportional hazards models, and they can be generalized to allow for ties in the rankings. Detailed examples are given.

Sociological Methodology, Vol. 24. (1994), pp. 199-228.

Format long

	priorite	order_pres	ev	person	prog
	1	5	1	4276	3
	2	2	1	4276	2
	3	1	1	4276	4
	4	3	1	4276	1
	5	4	1	4276	6
	6	6	1	4276	5
	1	2	1	4277	4
	2	4	1	4277	3
	3	3	1	4277	5
	4	6	1	4277	6
	5	5	1	4277	2
	6	1	1	4277	1

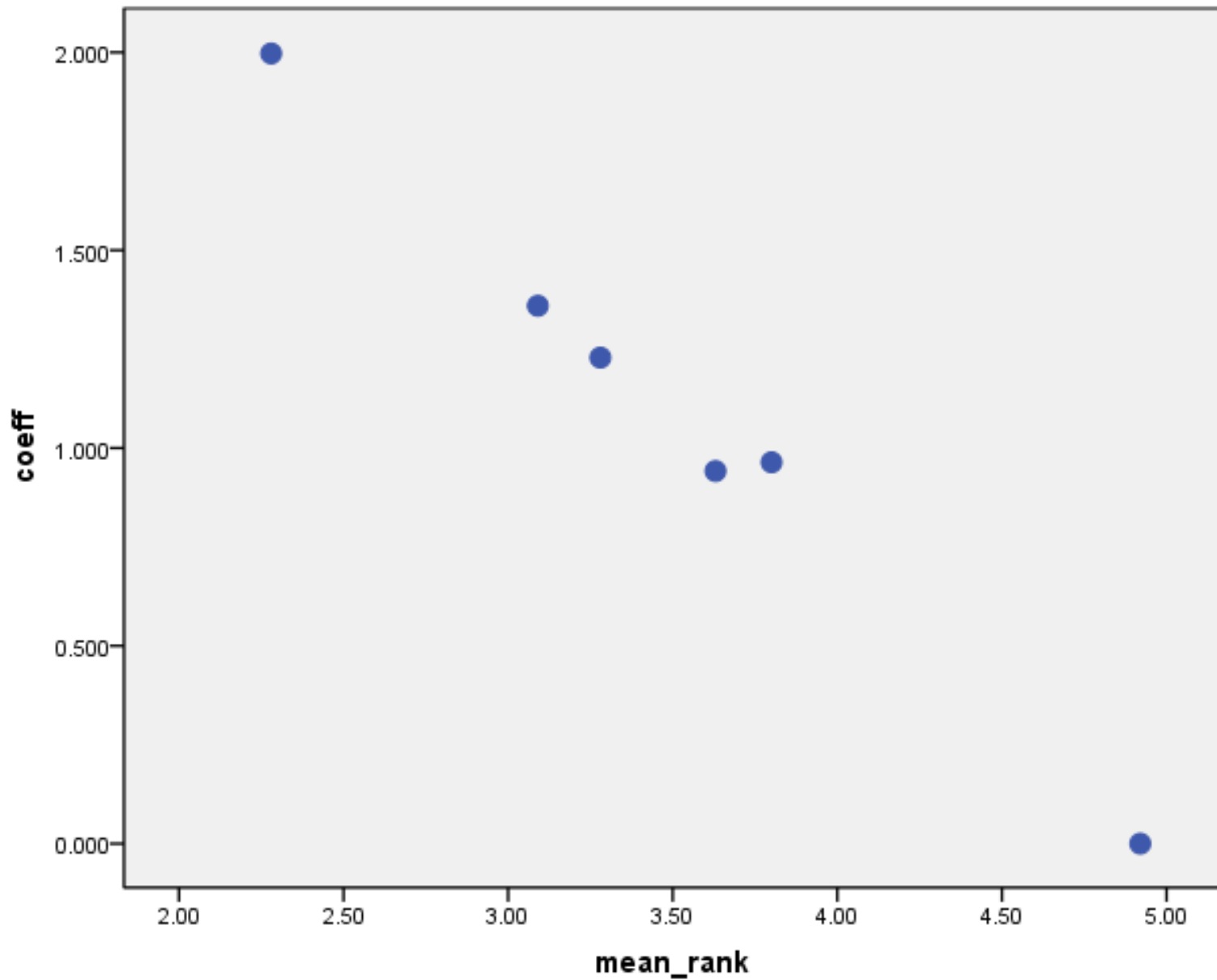


Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
prog			1001.600	5	.000			
prog(1)	1.997	.065	931.998	1	.000	7.369	6.482	8.377
prog(2)	1.360	.063	462.812	1	.000	3.895	3.441	4.409
prog(3)	1.229	.063	385.368	1	.000	3.416	3.022	3.862
prog(4)	.964	.063	236.314	1	.000	2.622	2.318	2.964
prog(5)	.942	.062	229.869	1	.000	2.565	2.271	2.897

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
sein	1.997	.065	931.998	1	.000	7.369	6.482	8.377
sida	1.360	.063	462.812	1	.000	3.895	3.441	4.409
alcool	1.229	.063	385.368	1	.000	3.416	3.022	3.862
prostate	.964	.063	236.314	1	.000	2.622	2.318	2.964
tabac	.942	.062	229.869	1	.000	2.565	2.271	2.897



Variables in the Equation

year1 2 year		B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
								Lower	Upper
1998	sein	1.894	.103	336.563	1	.000	6.649	5.430	8.140
	sida	1.874	.104	323.237	1	.000	6.513	5.310	7.989
	alcool	1.052	.099	112.955	1	.000	2.864	2.359	3.477
	prostate	1.033	.099	108.612	1	.000	2.810	2.314	3.413
	tabac	.845	.099	72.454	1	.000	2.328	1.916	2.828
2012	sein	1.997	.065	931.998	1	.000	7.369	6.482	8.377
	sida	1.360	.063	462.812	1	.000	3.895	3.441	4.409
	alcool	1.229	.063	385.368	1	.000	3.416	3.022	3.862
	prostate	.964	.063	236.314	1	.000	2.622	2.318	2.964
	tabac	.942	.062	229.869	1	.000	2.565	2.271	2.897

Sans conditionnement

Variables in the Equation^a

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
sein	1.371	.049	784.729	1	.000	3.940	3.580	4.337
sida	.867	.048	332.035	1	.000	2.381	2.169	2.613
alcool	.793	.048	277.558	1	.000	2.209	2.013	2.425
prostate	.537	.047	130.083	1	.000	1.710	1.559	1.875
tabac	.579	.047	151.233	1	.000	1.784	1.627	1.956

a. year12 year = 2012

Effet d'ordre?

a) Liste des priorités (selon version du questionnaire, de A à F)

A	B	C	D	E	F
HIV	Alcool	Dépression	Sein	Prostate	Tabac
Sein	Tabac	Alcool	Prostate	HIV	Dépression
Dépression	Prostate	HIV	Tabac	Sein	Alcool
Tabac	Sein	Prostate	Alcool	Dépression	HIV
Prostate	Dépression	Tabac	HIV	Alcool	Sein
Alcool	HIV	Sein	Dépression	Tabac	Prostate

Modèle multivarié

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
order_pres			59.416	5	.000			
order_pres(1)	.344	.051	46.022	1	.000	1.411	1.277	1.558
order_pres(2)	.152	.050	9.226	1	.002	1.165	1.056	1.285
order_pres(3)	.127	.050	6.510	1	.011	1.135	1.030	1.252
order_pres(4)	.145	.049	8.758	1	.003	1.156	1.050	1.272
order_pres(5)	.016	.050	.104	1	.747	1.016	.922	1.120
sein	2.015	.066	941.307	1	.000	7.501	6.595	8.532
sida	1.375	.063	470.121	1	.000	3.955	3.492	4.478
alcool	1.239	.063	390.591	1	.000	3.453	3.054	3.904
prostate	.978	.063	242.007	1	.000	2.659	2.350	3.007
tabac	.945	.062	230.441	1	.000	2.574	2.278	2.908
sein*year98	-.079	.123	.419	1	.518	.924	.726	1.175
sida*year98	.523	.122	18.351	1	.000	1.688	1.328	2.144
alcool*year98	-.160	.117	1.852	1	.174	.852	.677	1.073
prostate*year98	.073	.118	.385	1	.535	1.076	.854	1.354
tabac*year98	-.084	.117	.518	1	.472	.919	.730	1.157

2012

Avantages du modèle de Cox ou « rank-ordered logit »

- Résultat simplifié par rapport à Condorcet: $k-1$ odds ratios plutôt que $k(k-1)/2$ odds ratios de McNemar
- Permet ajustement pour
 - Facteurs liés à l'étude (ordre de présentation,...)
 - Caractéristiques des programmes à classer (p. ex. dépistage de cancers versus autre programme)
- Permet de tester des interactions entre:
 - Programme et conditions de l'étude (année d'enquête)
 - Programme et sous-groupe de participants

Conclusions

- Méthodes classiques
 - Rang moyen de chaque option (Borda)
 - Comparer toutes les paires d'options (Condorcet)
- Les 2 peuvent être justifiés!
- Regardez les tableaux croisés 2x2 pour détecter des patterns particuliers
- Rank-ordered conditional logistic regression pour analyse multivariée