

Statistiques et informatique

appliquées aux sciences sociales

Analyse d'un tableau croisé : indicateurs avancés

Mardi 10/10/2023 15h30-17h Salle D2 Censier

Licence de sciences sociales 3^e année
Université Paris 1 Panthéon Sorbonne

Plan de la séance

- Quelle information est significative ?
- Évaluer la force des écarts (suite) :
le V de Cramer
- Récapitulatif : indicateurs, formules
- Pour aller plus loin : analyse factorielle et
analyse des correspondances multiples

Des écarts pondérés au Khi²

On cherche à présent à construire un indicateur mesurant l'information apportée par une case du tableau toujours par rapport à la situation théorique d'indépendance

- On s'attend à ce que les écarts à l'indépendance ne puissent pas être beaucoup plus grands que les effectifs attendus (théoriques) en valeur absolue.
- On construit donc un poids égal à l'écart à l'indépendance divisé par l'effectif théorique.
- On multiplie ce poids par l'écart à l'indépendance pour obtenir un écart pondéré, qui est de signe positif
- La somme de ces écarts pondérés est appelé le Khi² du tableau
- On connaît la loi que suivrait ce Khi² si les variables étaient indépendantes, compte tenu du nombre de lignes et de colonnes
- (à condition qu'aucun effectif théorique ne soit inférieur à 5 individus)

Tableau permettant le calcul du Khi²

écarts pondérés	Priorités entre carrière et vie familiale			
	1:Privilégie sa carrière professionnelle	2:Accorde une égale importance...	3:Privilégie sa famille	Ensemble
genre				
0:homme	0,41	0,03	0,11	0,56
1:femme	0,40	0,03	0,11	0,54
Ensemble	0,81	0,07	0,23	1,10

Khi²=1,1

Degré de liberté = 2

prob(Khi²>1,1) ≈ 0,576

Il y avait donc 0,576 chances soit environ 6 chances sur 10 d'observer ces écarts à l'indépendance par le hasard de l'échantillonnage sous l'hypothèse qu'ils soient nuls dans la population-cible.

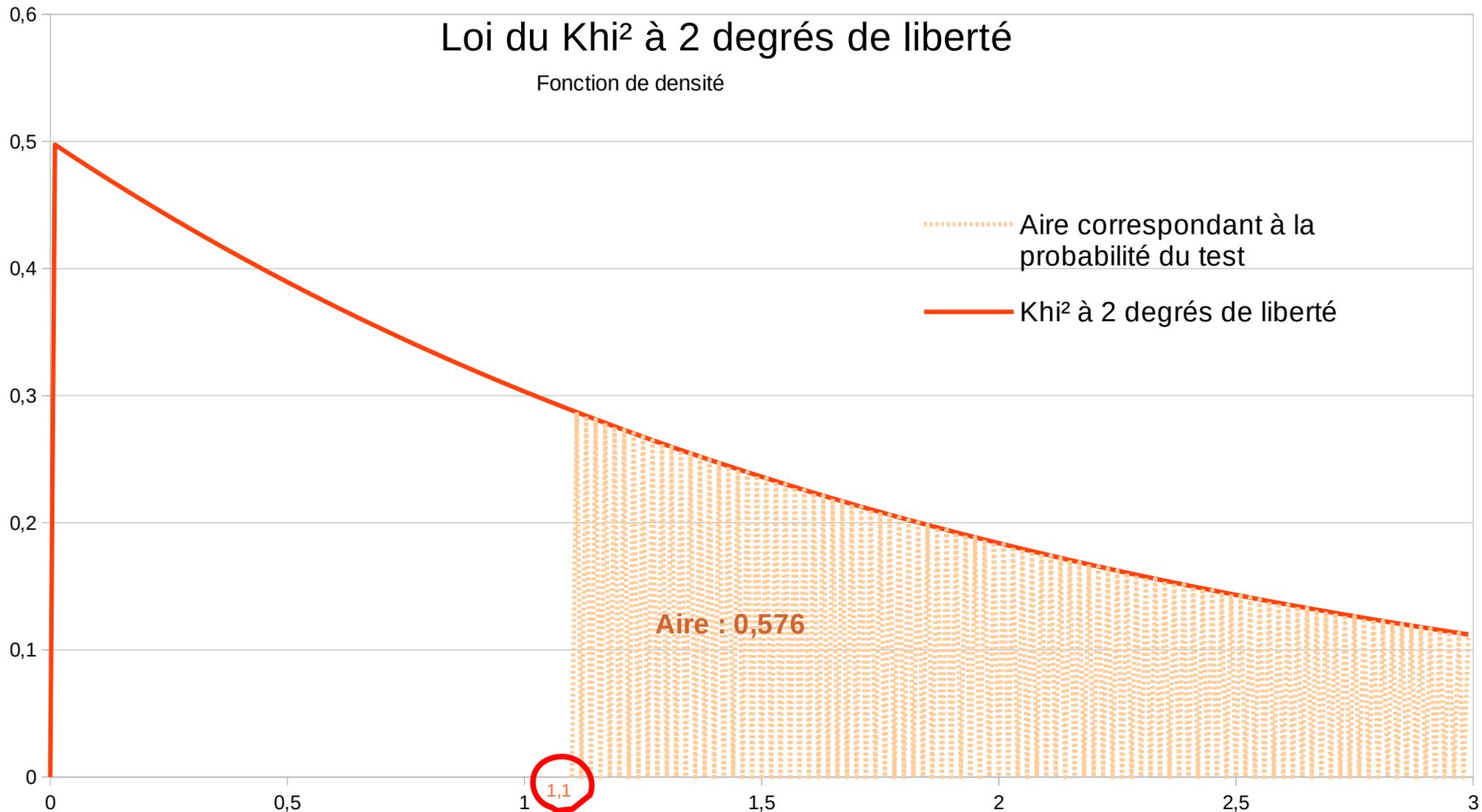
Rappel :
 $\chi^2=1,1$

Tables de la loi du Khi²

Degré de liberté \ Seuil à :	1 %	5 %	10 %
1	6,63489660102121	3,84145882069413	2,70554345409541
2	9,21034037197618	5,99146454710798	4,60517018598809
3	11,3448667301444	7,81472790325118	6,25138863117033
4	13,2767041359876	9,48772903678116	7,77944033973486
5	15,086272469389	11,0704976935164	9,23635689978112
6	16,8118938297709	12,591587243744	10,6446406756684
7	18,4753069065824	14,0671404493402	12,0170366237805
8	20,0902350296632	15,5073130558655	13,3615661365117
9	21,6659943334619	16,9189776046204	14,6836565732598
10	23,2092511589544	18,3070380532752	15,9871791721053
11	24,7249703113183	19,6751375726825	17,2750085175001
12	26,2169673055359	21,0260698174831	18,5493477867032
13	27,688249610457	22,3620324948269	19,8119293071276

Loi du Khi^2 à 2 degrés de liberté

Fonction de densité



..... Aire correspondant à la probabilité du test

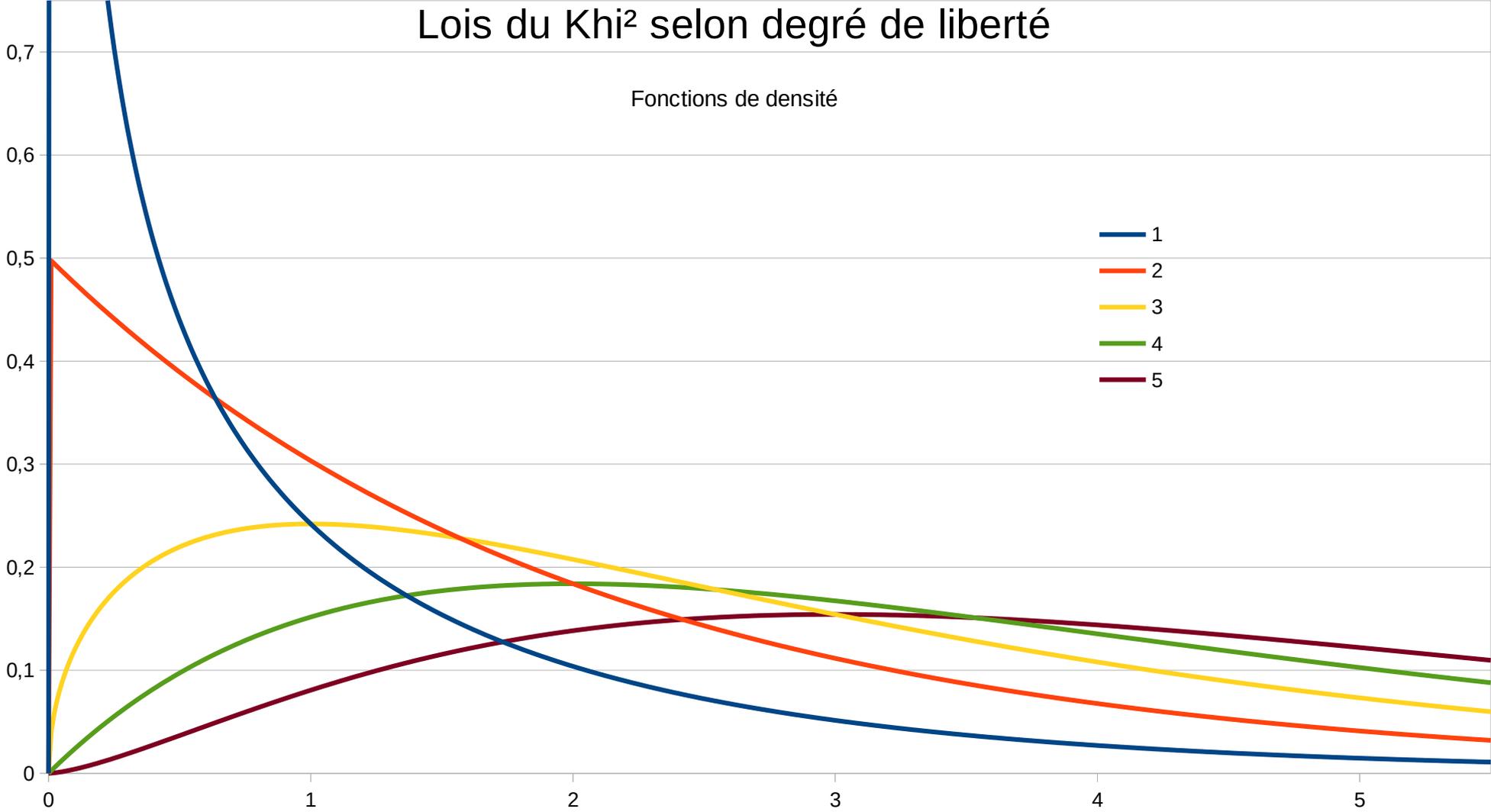
— Khi² à 2 degrés de liberté

Aire : 0,576

1,1

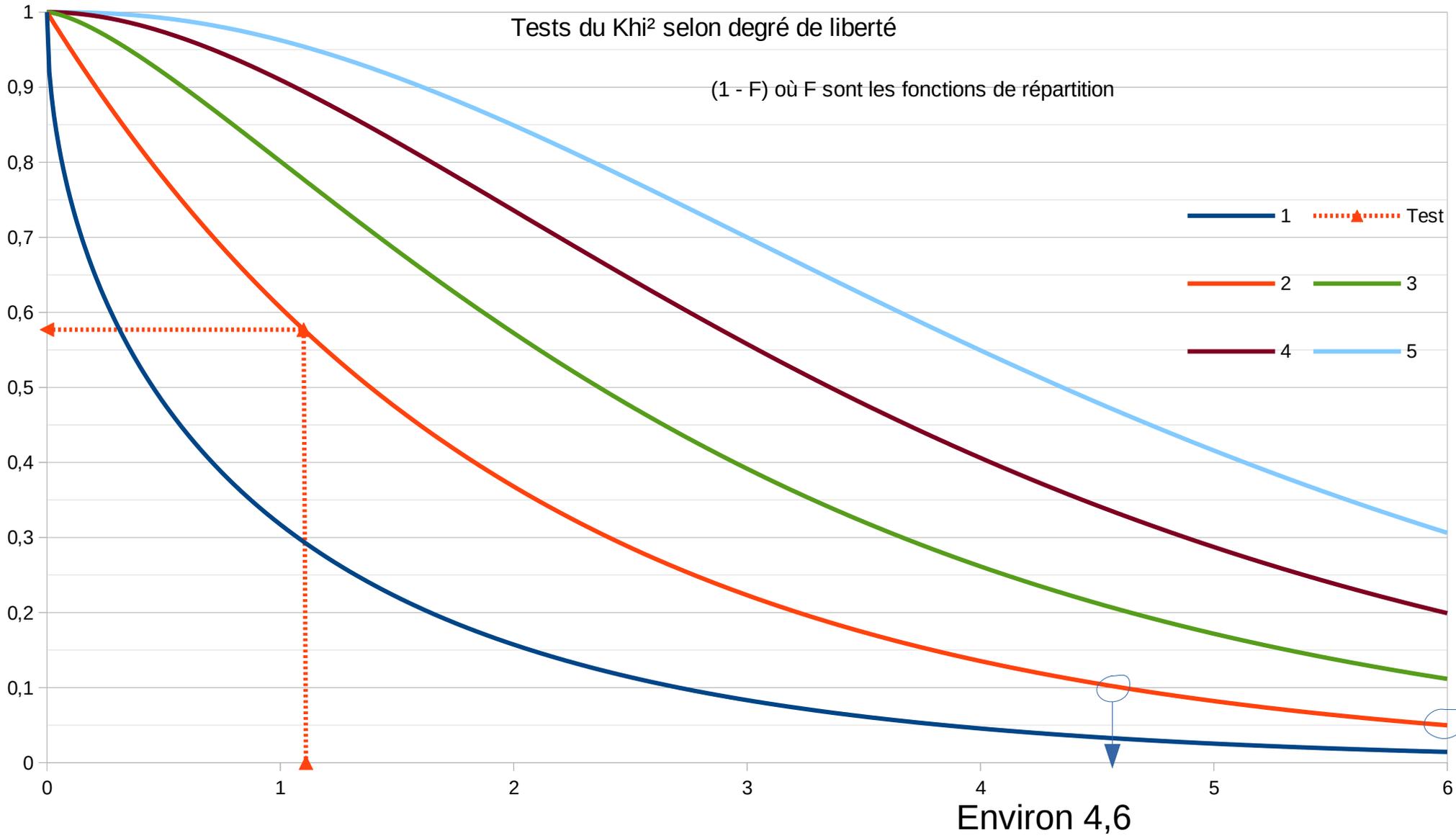
Lois du Khi² selon degré de liberté

Fonctions de densité



Tests du Khi² selon degré de liberté

(1 - F) où F sont les fonctions de répartition



Environ 4,6

Retour rapide sur les indicateurs utilisés à partir d'un tableau croisé

- % ligne, profil des écarts à la moyenne
- Écarts à l'indépendance (comparaison du tableau d'effectif observés au tableau des effectifs théoriques, représentatifs de la situation d'indépendance)
- Écarts pondéré, calcul et interprétation du Khi^2
- Comparaison des écarts observés aux écarts obtenus dans un tableau maximisant des écarts de même sens que ceux observés (PEM)
- Le V de Cramer permet lui de comparer le Khi^2 à une valeur maximale théorique, en se ramenant à une unité en écarts non élevés au carré
- Comment représenter graphiquement les écarts à l'indépendance ?

Évaluer la force des écarts (2)

le Φ et le V de Cramer

- On a vu que le χ^2 (alias χ^2) était proportionnel à l'effectif du tableau (pour un tableau présentant *la même structure*) et qu'il tendait à lui être inférieur
- Ceci résulte du fait que le χ^2 est d'autant plus grand que les écarts sont significatifs (non dus au hasard)*
- Deux éléments sont pris en compte pour rejeter le hasard
- (1) la force des écarts à l'indépendance compte tenu des marges du tableau
- (2) le fait d'avoir un échantillon suffisamment important
- Si on veut éliminer la relation de proportionnalité avec l'effectif, il est naturel de diviser le χ^2 par l'effectif du tableau. On obtient un indicateur appelé Φ^2 , homogène à des écarts au carré. Le calcul de la racine carrée du Φ^2 , notée Φ , permet de se ramener à des écarts
- Dans une logique similaire au calcul du PEM, on peut montrer que le χ^2 ne peut jamais dépasser la valeur $(k-1)*n$ k étant le minimum entre le nombre de lignes et de colonnes et n l'effectif du tableau
- On en déduit le V de Cramer, défini comme la racine carrée du quotient entre le χ^2 et ce majorant
- Le majorant ainsi utilisé ne dépend pas de la structure des écarts observé

Application numérique

- $\chi^2 = 1,1$ ici

$$\Phi^2 = \frac{\chi^2}{n} = \frac{1,1}{897} \approx 0,00123$$

$$\Phi = \sqrt{\frac{\chi^2}{n}} \approx 0,035$$

$$V = \sqrt{\frac{\chi^2}{n \cdot (k-1)}} = \Phi \approx 0,035$$

Rappel : PEM = 5,6 %

k=2 car il y a deux modalités en ligne

Récapitulatif : indicateurs, formules

$$\% \text{ ligne} = \frac{\text{effectif case}}{\text{marge ligne}} \text{ (en \%)}$$

$$PEM = \frac{\text{écarts positifs (observés)}}{\text{écarts positif (EM)}} \text{ (en \%)}$$

$$\text{écart à la moyenne} = \frac{\% \text{ ligne case} - \% \text{ ligne moyen}}$$

$$\chi^2 = \text{somme des écarts pondérés du tableau}$$

$$\text{effectif théorique} = \frac{\text{produit des marges}}{\text{effectif total}}$$

$$\Phi = \sqrt{\frac{\chi^2}{n}}$$

$$\text{écart à l'indépendance} = \frac{\text{effectif observé} - \text{effectif théorique}}$$

$$V = \sqrt{\frac{\chi^2}{n \cdot (k - 1)}}$$

$$\text{contribution au } \chi^2 = \frac{\text{écart pondéré}}{\text{effectif théorique}} = \frac{[\text{écart à l'indépendance}]^2}{\text{effectif théorique}}$$

Récapitulatif : indicateurs, formules

$$p_{ij|i} = \frac{n_{ij}}{n_{i.}} \text{ (en \%)}$$

$$\begin{array}{l} \text{écart à} \\ \text{la moyenne} \end{array} = \frac{n_{ij}}{n_{i.}} - \frac{n_{.j}}{n}$$

$$\begin{array}{l} \text{effectif} \\ \text{théorique} \end{array} = \tilde{n}_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

$$\begin{array}{l} \text{écart à} \\ \text{l'indépendance} \end{array} = n_{ij} - \tilde{n}_{ij}$$

$$\begin{array}{l} \text{contribution} \\ \text{au } \chi^2 \end{array} = \begin{array}{l} \text{écart} \\ \text{pondéré} \end{array} = \frac{[n_{ij} - \tilde{n}_{ij}]^2}{\tilde{n}_{ij}}$$

$$PEM = \frac{\sum_{i,j|n_{ij} > \tilde{n}_{ij}} (n_{ij} - \tilde{n}_{ij})}{\sum_{i,j|n_{ij}^{EM} > \tilde{n}_{ij}} (n_{ij}^{EM} - \tilde{n}_{ij})} \text{ (en \%)}$$

$$\chi^2 = \sum_{i=1}^l \sum_{j=1}^c \frac{[n_{ij} - \tilde{n}_{ij}]^2}{\tilde{n}_{ij}}$$

$$\Phi = \sqrt{\frac{\chi^2}{n}}$$

$$V = \sqrt{\frac{\chi^2}{n \cdot (k-1)}}$$

Un autre exemple

- Nous avons vu que les femmes privilégiaient un peu plus souvent que les hommes la famille, mais que cette différence est faible, au point de ne pas être significative
- Qu'en est-il des personnes dont nous avons constaté qu'elles étaient mobiles géographiquement pour des raisons liées au travail...
- ... déclarent-elles plus souvent privilégier leur carrière professionnelle ?
- ... Les écarts observés sont-ils forts ?
- ... Sont-ils significatifs ?

Lecture rapide

% ligne	Priorités entre carrière et vie familiale			
Mobilité (résumée)	1:Privilégie sa carrière professionnelle	2:Accorde une égale importance...	3:Privilégie sa famille	Ensemble
0:Non mobile	11,6%	10,1%	78,3%	100,0%
1:Mobile	24,0%	14,7%	61,2%	100,0%
Ensemble	13,4%	10,8%	75,8%	100,0%

écarts à la moyenne	Priorités entre carrière et vie familiale		
Mobilité (résumée)	1:Privilégie sa carrière professionnelle	2:Accorde une égale importance...	3:Privilégie sa famille
0:Non mobile	-1,8 pts	-0,7 pts	+2,4 pts
1:Mobile	+10,7 pts	+3,9 pts	-14,6 pts

Effectifs observés	Priorités entre carrière et vie familiale			
Mobilité (résumée)	1:Privilégie sa carrière professionnelle	2:Accorde une égale importance...	3:Privilégie sa famille	Ensemble
0:Non mobile	89	78	602	769
1:Mobile	31	19	79	129
Ensemble	120	97	681	898

Effectifs théoriques	Priorités entre carrière et vie familiale			
Mobilité (résumée)	1:Privilégie sa carrière professionnelle	2:Accorde une égale importance...	3:Privilégie sa famille	Ensemble
0:Non mobile	102,8	83,1	583,2	769
1:Mobile	17,2	13,9	97,8	129
Ensemble	120	97	681	898

Écarts à l'indépendnce	Priorités entre carrière et vie familiale			
Mobilité (résumée)	1:Privilégie sa carrière professionnelle	2:Accorde une égale importance...	3:Privilégie sa famille	Ensemble
0:Non mobile	-13,8	-5,1	18,8	0
1:Mobile	13,8	5,1	-18,8	0
Ensemble	0	0	0	0

Pourcentage de l'écart maximal

Somme des écarts + (observés) $\approx 37,65$

Effectifs / écarts max	Priorités entre carrière et vie familiale			
Mobilité (résumée)	1:Privilégie sa carrière professionnelle	2:Accorde une égale importance...	3:Privilégie sa famille	Ensemble
0:Non mobile	0	88	681	769
1:Mobile	120	9	0	129
Ensemble	120	97	681	898

Écarts / PEM	Priorités entre carrière et vie familiale			
Mobilité (résumée)	1:Privilégie sa carrière professionnelle	2:Accorde une égale importance...	3:Privilégie sa famille	Ensemble
0:Non mobile	-102,8	4,9	97,8	0,0
1:Mobile	102,8	-4,9	-97,8	0,0
Ensemble	0,0	0,0	0,0	0,0

Somme des écarts + (max) $\approx 205,52$ et PEM $\approx 18,3 \%$

Khi², V de Cramer, utilisation

Écarts pondérés	Priorités entre carrière et vie familiale			
	1:Privilégie sa carrière professionnelle	2:Accorde une égale importance...	3:Privilégie sa famille	Ensemble
0:Non mobile	1,84	0,31	0,61	2,76
1:Mobile	10,99	1,84	3,62	16,45
Ensemble	12,83	2,15	4,23	19,21

Khi ²	19,21
Prob(Khi ² >19,21)	6,73582356299111E-05
Phi ²	0,021393063210503
Phi	0,146263677003223
V de Cramer	0,146263677003223

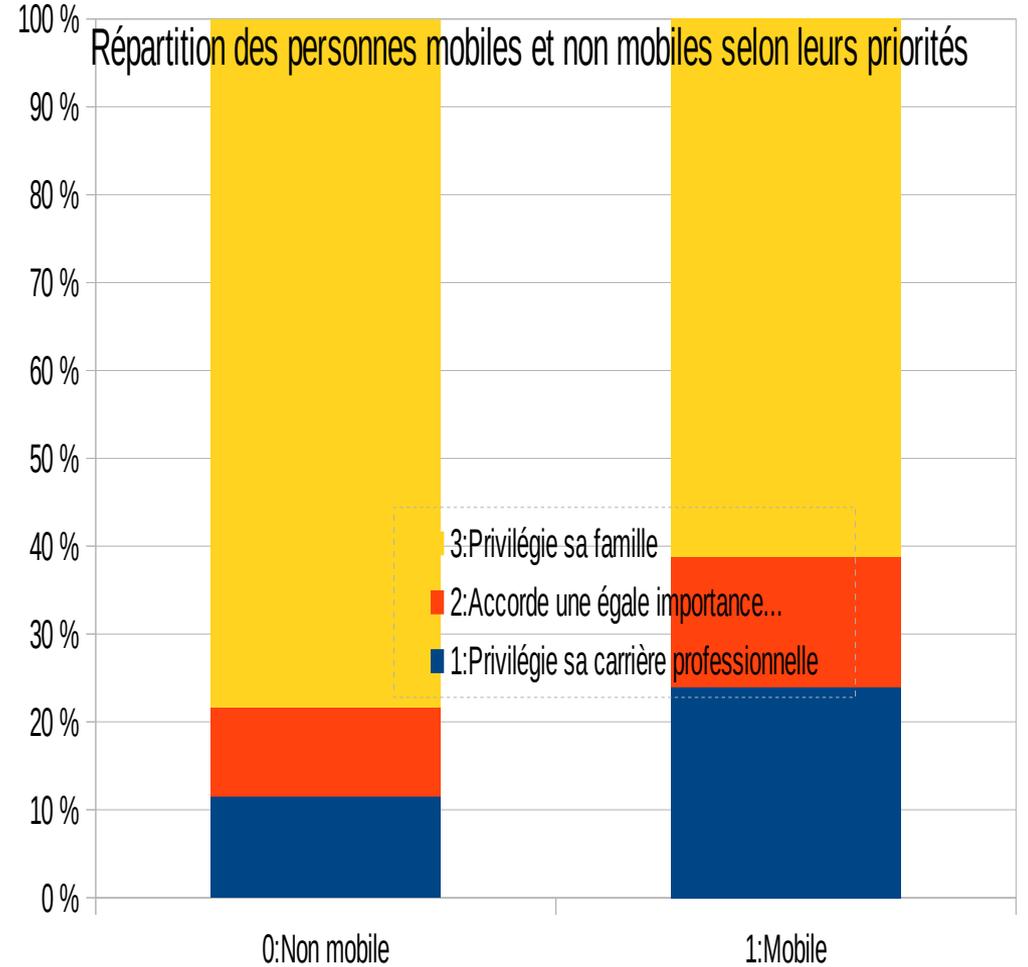
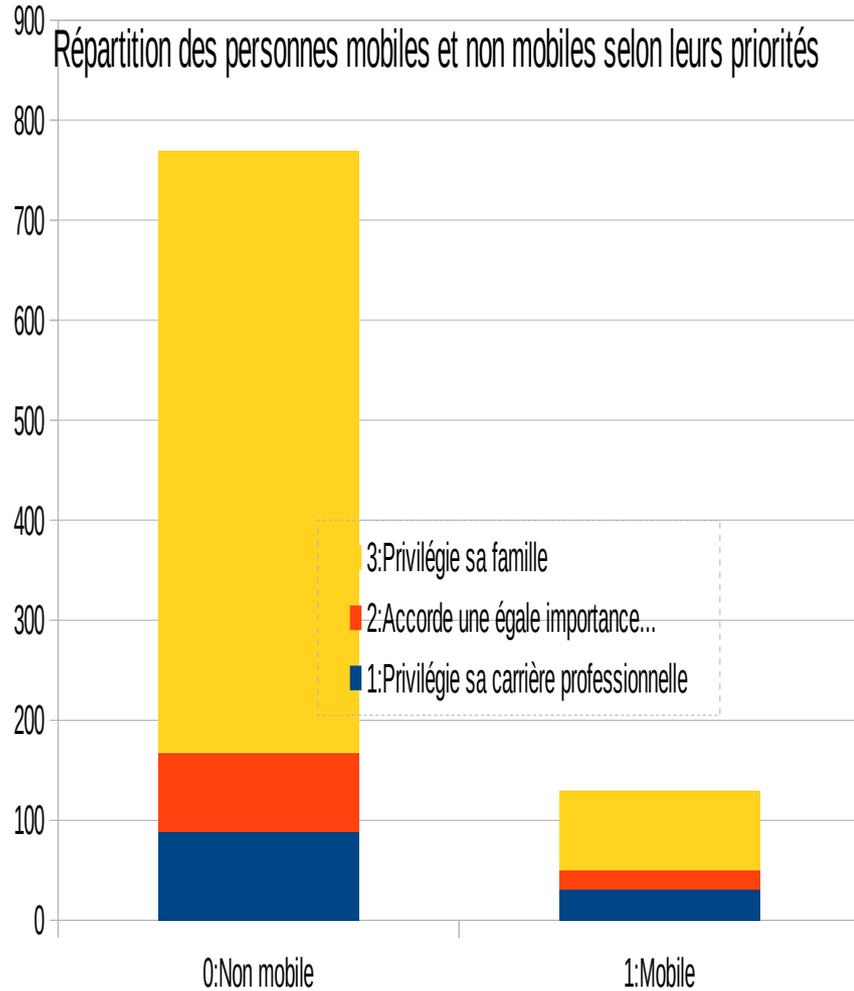
Degré de liberté du tableau : 2

Seuil (1 %) = 9,21

Table de la loi du Khi² (extrait)

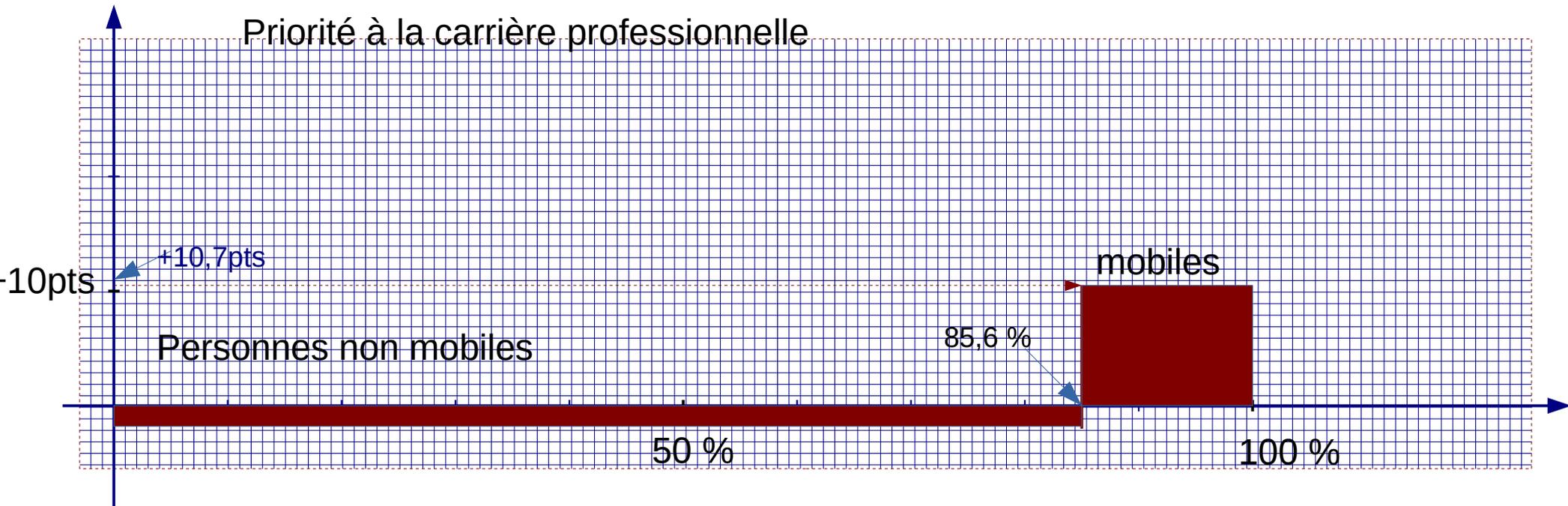
DDL	Seuil de...	1%	5%	10%
1		6,63489660102121	3,84145882069413	2,70554345409541
2		9,21034037197618	5,99146454710798	4,60517018598809
3		11,3448667301444	7,81472790325118	6,25138863117033
4		13,2767041359876	9,48772903678116	7,77944033973486
5		15,086272469389	11,0704976935164	9,23635689978112
6		16,8118938297709	12,591587243744	10,6446406756684
7		18,4753069065824	14,0671404493402	12,0170366237805
8		20,0902350296632	15,5073130558655	13,3615661365117
9		21,6659943334619	16,9189776046204	14,6836565732598
10		23,2092511589544	18,3070380532752	15,9871791721053
11		24,7249703113183	19,6751375726825	17,2750085175001
12		26,2169673055359	21,0260698174831	18,5493477867032

Représentation graphique des écarts



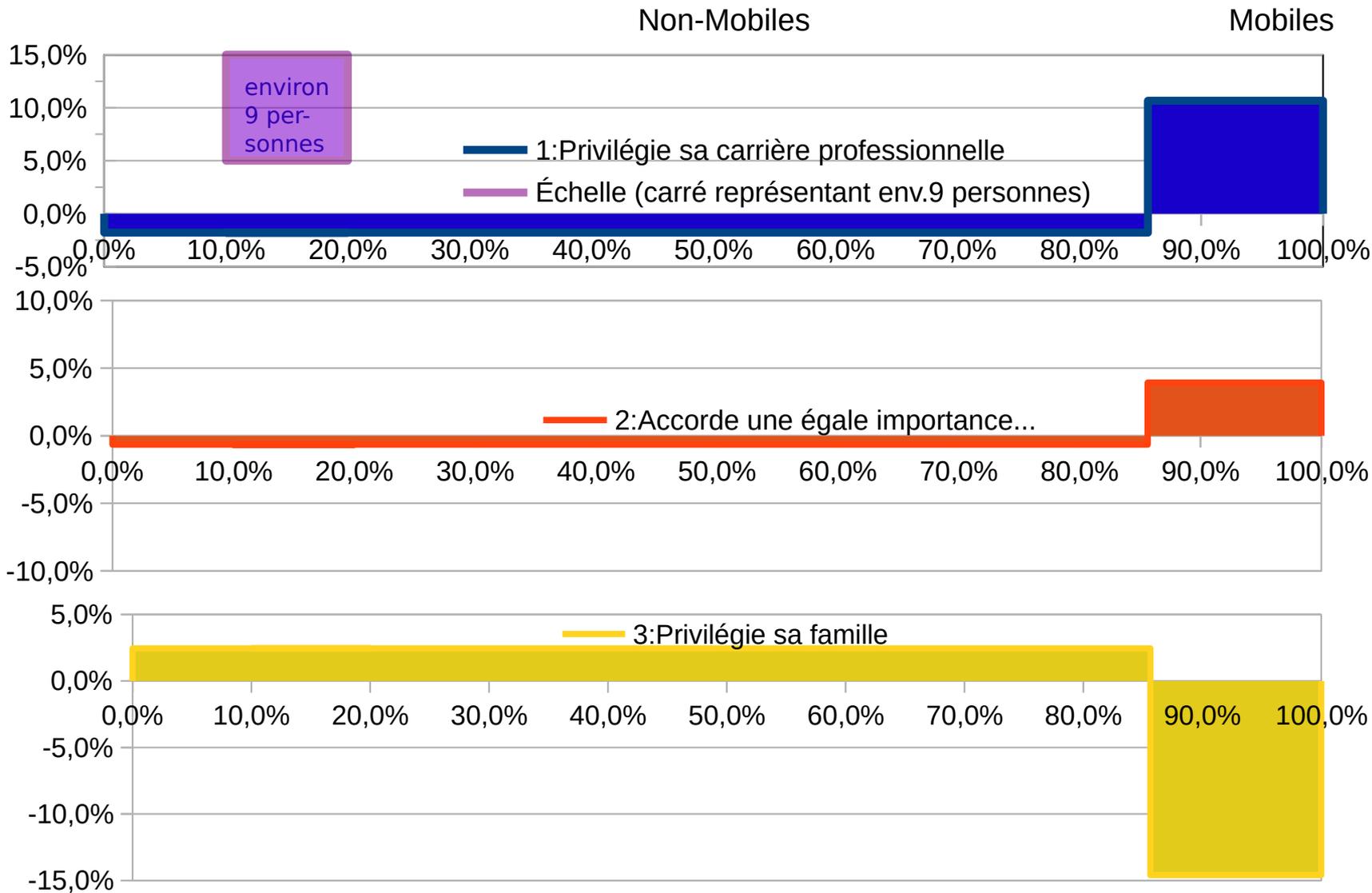
Écarts à la moyenne

Pour la représentation graphique	1:Privilégie sa carrière professionnelle	2:Accorde une égale importance...	3:Privilégie sa famille	Profil colonne
0:Non mobile	-1,8 pts	-0,7 pts	+2,4 pts	85,6%
1:Mobile	+10,7 pts	+3,9 pts	-14,6 pts	14,4%
Ensemble				100,0%



Représentation

graphique des écarts



Commentaire

- Un peu moins d'un quart des personnes mobiles privilégient leur carrière professionnelle et un peu moins de 3 sur 20 lui accordent une égale importance qu'à la famille
- Les proportions sont de l'ordre d'une personne non mobile sur 10 ; ces différences sont relativement fortes et tout à fait significatives ; cependant, 6 personnes mobiles sur 10 privilégient la famille en France en 2007
- Il serait plus aisé d'interpréter ces résultats en supposant que les dispositions vis à vis de la famille et du travail se construisent avant d'être confronté à une opportunité ou une contrainte de travail nécessitant une mobilité géographique
- Cependant, on peut aussi considérer que la réponse à la question sur les priorités peut amener les enquêté·es à constater qu'au moins dans l'immédiat, leurs impératifs professionnels ont eu tendance à primer sur la famille et notamment sur la construction d'une famille ou encore sur les impératifs d'une vie de couple, notamment la cohabitation