

# TD9 : les étudiants et la ville

## 1 Les données

Les données qui suivent sont issues de l'enquête « les étudiants et la ville » effectuée en 2001 par des étudiants de sociologie sous la direction de S. Denèfle à l'Université François Rabelais de Tours. L'analyse porte sur cinq questions en rapport avec le logement étudiant. L'ensemble des individus statistiques est ici un échantillon de 383 étudiants. Les questions sont les suivantes :

- Habitez-vous (variable `occ`) : seul (modalité `seul`), en colocation (`coloc`), en couple (`couple`), avec les parents (`parents`), non réponse (NR) ?
- Quel type d'habitation occupez-vous (variable `typ`) : cité universitaire (`cite`), studio (`studio`), appartement (`appart`), chambre chez un particulier (`chambre`), autre (`autre`), non réponse (NR) ?
- Si vous vivez en dehors du foyer familial, depuis combien de temps (variable `anc`) : moins de 1 an (`0.1an`), 1 à 3 ans (`1.3ans`), plus de 3 ans (`p3ans`), non applicable (NA), non réponse (NR) ?
- À quelle distance approximative de l'université vivez-vous (variable `dst`) : moins de 1 km (`0.1km`), 1 à 5 km (`1.5km`), plus de 5 km (`p5km`), non réponse (NR) ?
- Quelle est la superficie de votre logement (variable `sur`) : moins de 10 m<sup>2</sup> (`0.10m2`), 10 à 20 m<sup>2</sup> (`10.20m2`), 20 à 30 m<sup>2</sup> (`20.30m2`), plus de 30 m<sup>2</sup> (`p30m2`), non réponse (NR) ?

Dans ce qui suit, on représentera les catégories par le nom de la variable suivi du nom de la modalité, comme par exemple `occ.seul` ou `typ.NR`. Non réponse (NR) correspond à un défaut des données (réponse oubliée ou non fournie), alors que non applicable (NA) est utilisé pour les questions qui n'ont pas de sens pour un individu donné.

Les taux marginaux de réponses aux différentes question (en %) sont reproduit ici

	Poids (%)		Poids (%)		Poids (%)		Poids (%)		Poids (%)
<code>occ.seul</code>	48	<code>typ.cite</code>	11	<code>anc.0.1an</code>	21	<code>dst.0.1km</code>	27	<code>sur.0.10m2</code>	9
<code>occ.coloc</code>	14	<code>typ.studio</code>	28	<code>anc.1.3ans</code>	25	<code>dst.1.5km</code>	50	<code>sur.10.20m2</code>	18
<code>occ.couple</code>	13	<code>typ.appart</code>	30	<code>anc.p3ans</code>	29	<code>dst.p5km</code>	21	<code>sur.20.30m2</code>	25
<code>occ.parents</code>	23	<code>typ.chambre</code>	5	<code>anc.NA</code>	25	<code>dst.NR</code>	2	<code>sur.p30m2</code>	39
<code>occ.NR</code>	1	<code>typ.autre</code>	20	<code>anc.NR</code>	1			<code>sur.NR</code>	9
		<code>typ.NR</code>	6						

On reproduit ci-dessous la partie du tableau de Burt qui correspond aux variables `occ` et `typ` (le tableau total serait trop grand).

	<code>occ.seul</code>	<code>occ.coloc</code>	<code>occ.couple</code>	<code>occ.parents</code>	<code>occ.NR</code>	<code>typ.cite</code>	<code>typ.studio</code>	<code>typ.appart</code>	<code>typ.chambre</code>	<code>typ.autre</code>	<code>typ.NR</code>
<code>occ.seul</code>	185	0	0	0	0	34	90	40	NA	NA	5
<code>occ.coloc</code>	0	53	0	0	0	5	6	32	2	3	5
<code>occ.couple</code>	0	0	50	0	0	2	10	34	0	3	1
<code>occ.parents</code>	0	0	0	90	0	0	1	9	NA	67	8
<code>occ.NR</code>	0	0	0	0	5	0	1	1	0	0	3
<code>typ.cite</code>	34	5	2	0	0	41	0	0	0	0	0
<code>typ.studio</code>	90	6	10	1	1	0	108	0	0	0	0
<code>typ.appart</code>	40	32	34	9	1	0	0	116	0	0	0
<code>typ.chambre</code>	NA	2	0	NA	0	0	0	0	20	0	0
<code>typ.autre</code>	NA	3	3	67	0	0	0	0	0	NA	0
<code>typ.NR</code>	5	5	1	8	3	0	0	0	0	0	22

**Question 1** 7 valeurs sont manquantes (NA) dans le tableau de Burt. Retrouvez les en utilisant les propriétés du tableau.

**Question 2** Est-il exact de dire que « la proportion des gens seuls qui vivent en appartement est plus faible que la proportion des couples qui vivent en appartement » ?

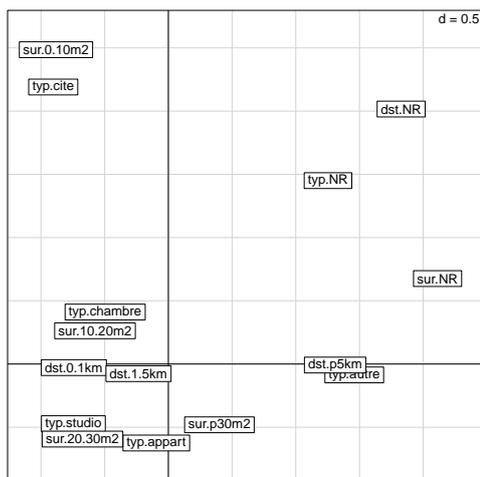
**Question 3** Expliquer pourquoi les catégories `occ.NR`, `anc.NR` et `dst.NR` risquent de poser des problèmes dans une ACM. Quelles pourraient être les solutions envisageables ?

## 2 Analyse des correspondances multiples

On réalise une analyse des correspondances multiples (ACM) uniquement sur les variables `typ`, `dst` et `sur`, qui décrivent les caractéristiques des logements. Les 6 premières valeurs propres sont

[1] 0.6990 0.5737 0.5254 0.3892 0.3558 0.3359

On donne ci-dessous pour le premier plan factoriel : la projection des catégories actives, les coordonnées de ces catégories sur chaque axe factoriel, leur contribution aux axes (en %), ainsi que la qualité de leur représentation par le plan (en % encore).



	Comp1	Comp2	Axis1	Axis2	Axis1	Axis1:2
typ.cite	-0.90	2.19	4.2	29.8	9.8	67.3
typ.studio	-0.75	-0.47	7.5	3.6	21.8	30.6
typ.appart	-0.09	-0.62	0.1	6.9	0.4	17.3
typ.chambre	-0.50	0.41	0.6	0.5	1.4	2.3
typ.autre	1.46	-0.09	20.1	0.1	52.7	52.9
typ.NR	1.25	1.45	4.3	7.0	9.6	22.4
dst.0.1km	-0.74	-0.03	7.1	0.0	20.2	20.3
dst.1.5km	-0.23	-0.08	1.3	0.2	5.5	6.0
dst.p5km	1.31	-0.01	17.1	0.0	45.2	45.2
dst.NR	1.83	2.02	3.7	5.6	8.0	17.8
sur.0.10m2	-0.88	2.49	3.4	32.8	7.8	70.0
sur.10.20m2	-0.58	0.26	2.8	0.7	7.2	8.6
sur.20.30m2	-0.67	-0.59	5.4	5.1	15.0	26.6
sur.p30m2	0.40	-0.48	3.0	5.3	10.3	25.2
sur.NR	2.11	0.68	19.5	2.4	44.9	49.5

Les variables *occ* et *anc*, qui décrivent les occupants des logements, sont utilisées comme variables supplémentaires qualitatives. On calcule les coordonnées et les valeurs test suivantes pour leurs catégories sur les deux premiers axes factoriels

	Comp1	Comp2	Comp1	Comp2
occ.seul	-0.55	0.14	-8.61	2.19
occ.coloc	-0.02	-0.21	-0.15	-1.59
occ.couple	-0.06	-0.36	-0.42	-2.63
occ.parents	1.13	0.00	11.44	0.02
occ.NR	0.77	0.61	1.74	1.36
anc.0.1an	-0.48	0.18	-4.55	1.73
anc.1.3ans	-0.32	-0.02	-3.30	-0.19
anc.p3ans	-0.33	-0.20	-3.77	-2.26
anc.NA	1.06	0.06	11.00	0.62
anc.NR	1.60	1.12	2.77	1.94

**Question 4** Calculer la proportion d'inertie expliquée si l'on conserve les deux premiers axes. Commentez la qualité.

**Question 5** Quelles sont les catégories qui définissent les deux premiers axes factoriels ? (on précisera sur quel(s) critère(s) on se fonde).

**Question 6** À quoi correspond la qualité de la représentation d'une catégorie par un sous espace vectoriel ? Que peut on dire ici de la qualité de la représentation des catégories par le premier plan factoriel ? Pouvait-on s'y attendre ?

**Question 7** Quelles sont les catégories supplémentaires qui sont liées aux axes ? On précisera les conditions sous lesquelles une valeur test permet de conclure.

**Question 8** Comment peut-on interpréter les axes ?