

Rapport du projet: Analyse de fonctions et salaires des 25 promotions

VU Tuan Hung et DO Quoc Khanh

18 juin 2012

1 Introduction

2 Parcours, origines, entreprises et salaires

Dans cette première partie, on étudie les relations entre les parcours et les origines des élèves avec les types d'entreprises dans lesquelles ils travaillent après leur obtention de diplômes, et le niveau de salaire ils pourront recevoir. Les relations entre l'admission et les diplômes qu'ils obtiennent ainsi que la cohérence des parcours sont aussi à étudier.

2.1 Traitements préliminaires des données

Dans cette étude de données, on ne traitera que les variables catégorielles par l'analyse de correspondances multiples. Il y a 6 variables à analyser dans cette partie : *diplôme*, *promotion*, *admission*, *FR.UE.HUE*, *type d'entreprise*, et *salaire brute*. La variable *promotion* peut considérer comme quantitative ou qualitative. Cependant, même si elle est traitée comme une variable catégorielles, elle a beaucoup de catégories par rapport aux autres variables (ce qui pourrait diminuer son impact). Pour résoudre ce problème, on divise les promotions en 4 intervalles équilibrés (en nombre d'éléments) en utilisant la fonction de *quantile*, et transforme donc la variable *promotion* en une variable qualitative de 4 modalités. On effectue les mêmes démarches pour *salaire brute*.

Il y a des cases vides (ou erronées) dans les données. La raison est plutôt qu'on n'a pas pu obtenir de bonnes informations et n'est pas que les interviewés voulaient répondre autrement. On souhaite donc remplacer ces valeurs nulles. Dans ce projet, l'analyse de correspondances multiples (ACM) s'effectue avec les tableaux disjonctifs complets créés par la méthode d'imputation afin de traiter avec les données manquantes.

2.2 Liste 1

On considère quatre variables : *diplôme*, *admission*, *FR.UE.HUE* et *salaire brute*.

2.2.1 Valeurs propres

Les axes factoriels sont montrés dans le Tableau 1. On constate que la contribution des axes à la variance totale descend plutôt lentement, ce qui rend difficile le choix de nombre d'axes utilisés pour analyser les données. Dans ce qui suit, on utilisera principalement les deux premiers axes.

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.523	11.7	11.7
dim 2	0.433	9.68	21.4
dim 3	0.41	9.16	30.5
dim 4	0.294	6.58	37.1
dim 5	0.267	5.98	43.1
dim 6	0.259	5.8	48.9
dim 7	0.255	5.69	54.6
dim 8	0.25	5.59	60.2
dim 9	0.25	5.59	65.8
dim 10	0.25	5.59	71.4
dim 11	0.25	5.59	77
dim 12	0.226	5.06	82
dim 13	0.224	5	87
dim 14	0.203	4.54	91.5
dim 15	0.186	4.17	95.7
dim 16	0.092	2.06	97.8
dim 17	0.0793	1.77	99.5
dim 18	0.0208	0.465	100

TABLE 1 – Les axes factoriels de la première liste.

2.2.2 Résultats relatifs aux modalités

Les résultats relatifs aux modalités sont montrés au Tableau 2, quelques éléments importants sont en gras. Les deux modalités contribuant les plus au premier axe sont dip_1 (Mastère) et ad_9 (Autres admissions). Alors qu’au second axe, ce sont les modalités dip_6 (Ingénieur du corps) et ad_5 (Polytechnique) qui contribuent les plus. On retient ces modalités, car les axes pourraient être caractérisés par elles. Par exemple, la coordination de dip_6 sur le second axe est la plus positive parmi les modalités de la même variable. On observe ainsi que les variables FR.UE.HUE et *salaire brute* jouent les rôles peu importants dans la formation de ces axes.

2.2.3 Résultats graphiques

La figure 1 montre qu’il y a une forte corrélation entre *diplôme* et *admission*, alors que les autres variables sont moins remarquables. Le premier axe est caractérisé par l’opposition entre dip_2 (Master of Sciences) et dip_6 (Ingénieur du corps), entre ad_9 (Autres admissions), ad_2 (Double diplôme) et ad_3 (Ec.Nal.Sup). Le deuxième axe marque une forte opposition entre ad_3 (Ec.Nal.Sup) et ad_2 (Double diplôme), entre dip_6 (Ingénieur du corps) et dip_3 (Ingénieur). En gros, c’est l’opposition entre les élèves qui font un parcours traditionnel d’ingénieur français et les élèves intégrant l’école des autres universités européens pour obtenir leur double diplôme. En regardant la figure 2, on peut partitionner les modalités en 3 groupes. Le premier groupe contient ad_9 (Autres admissions), dip_2 (Master of Sciences), dip_5 (Docteur), dip_1 (Mastère), dip_4 (DEA STN) et HUE. Ce sont les élèves hors européens qui font des parcours plutôt scientifiques. Le deuxième groupe contient ad_2 (Double diplôme) et UE, qui sont les élèves européens des programmes de coopération entre universités. Les dernier groupe contient les élèves français avec des parcours d’ingénieur traditionnels (ENS, Polytechnique, préparatoire etc.). Alors que la relation entre *admission* et *diplôme* est assez évidente, la relation entre ces variables et les nationalités est un nouveau point intéressant.

	Crd Dim1	Crd Dim2	Contrib Dim1	Contrib Dim2	Cos2 Dim1	Cos2 Dim2
dip 1	1.95	0.716	25.1	4.08	0.609	0.082
dip 2	3.47	0.698	1.37	0.0672	0.0288	0.00117
dip 3	-0.327	-0.396	3.91	6.94	0.349	0.513
dip 4	1.78	0.64	0.602	0.0943	0.0127	0.00164
dip 5	2.23	0.627	4.37	0.417	0.0932	0.00735
dip 6	-1.05	2.63	3.76	28.7	0.0847	0.536
ad 1	-0.399	-0.475	3.59	6.16	0.143	0.202
ad 2	1.30	-2.88	1.48	8.77	0.0315	0.155
ad 3	-1.40	4.15	0.82	8.72	0.0173	0.152
ad 4	-0.288	-0.549	0.581	2.55	0.0143	0.0519
ad 5	-0.801	1.32	4.7	15.4	0.116	0.316
ad 6	-0.507	-0.557	0.0684	0.1	0.00144	0.00174
ad 7	-0.19	-0.805	0.00689	0.149	0.000145	0.00259
ad 8	-0.623	0.546	0.0591	0.0549	0.00124	0.000954
ad 9	1.85	0.66	30.8	4.71	0.793	0.101
FR	-0.186	0.0558	1.46	0.160	0.273	0.0246
HUE	1.4	0.171	7	0.126	0.159	0.00237
UE	1.61	-1.67	4.62	6	0.100	0.108
sal 1	0.537	-0.320	3.47	1.48	0.101	0.0356
sal 2	0.0171	-0.155	0.00315	0.313	8.83e-05	0.00727
sal 3	-0.112	-0.137	0.150	0.269	0.00432	0.00643
sal 4	-0.403	0.542	2.14	4.67	0.0638	0.115

TABLE 2 – Les résultats relatifs aux modalités de la première liste.

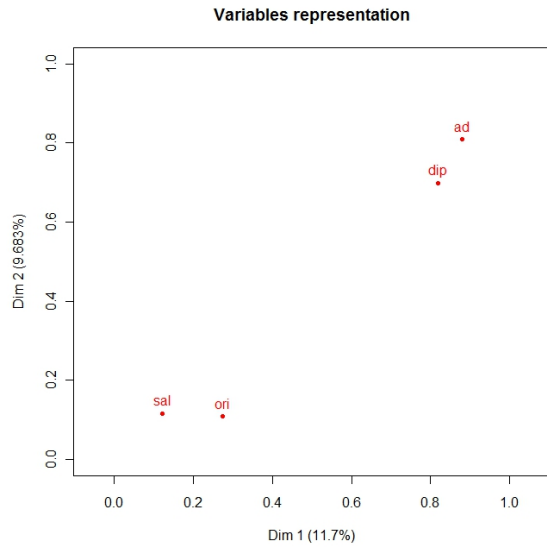


FIGURE 1 – Positionnement des modalités sur deux axes factoriels.

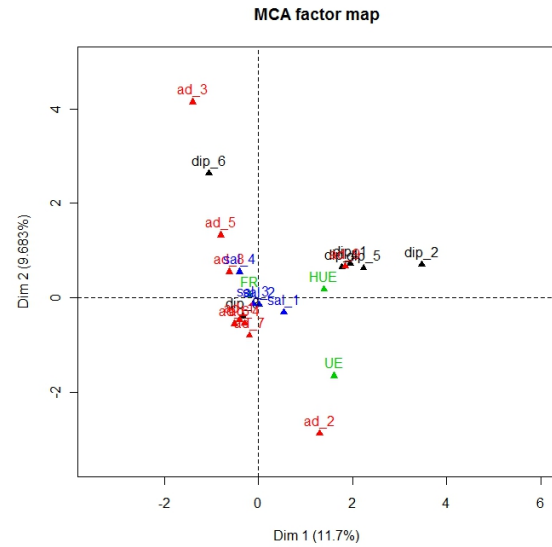


FIGURE 2 – Positionnement des modalités sur deux axes factoriels.

En considérant de plus près le groupe des élèves français (Figure 3), on constate que les modalités ad_1 (Prépa), ad_4 (Maîtrise de sciences), ad_6 (Arts et Métiers) et ad_7 (Militaire) sont très proches. Ce groupe de modalités, avec dip_3 (Ingénieur) et salaires moyennes (sal_2 et sal_3) indique les élèves du parcours typique d'ingénieur de Télécom ParisTech. Les élèves ENS, Polytechnique et Ingénieur de corps (ad_3 , ad_5 et dip_6) sont mis un peu à part. La faible variance des modalités de la variable *salaire* montre que les parcours, les diplômes et les origines ne reflètent pas bien le montant de salaire ; quoi qu'ils obtiennent comme diplôme, les salaires varient quand même fortement au sein de chaque groupe d'élèves.

2.3 Liste 2

Avec la première liste, on remarque une forte corrélation entre *diplôme* et *admission*. Dans la deuxième liste, on prend la variable *admission* comme le représentant de ces deux variables. On y ajoute la variable *promo*, et deux autres variables reflétant les carrières des élèves, *Type d'entreprise* et *salaire brute*. Les deux variables quantitatives *salaire brute* et *promo* sont factorisées en 4 modalités par variable selon les indications à la partie 2.1.

2.3.1 Valeurs propres

Les 17 axes factoriels sont montrés au Tableau 4. La contribution des axes à la variance totale descend plutôt lentement.

2.3.2 Les résultats relatifs aux modalités

Les résultats relatifs aux modalités sont montrés au Tableau 4. Les modalités contribuant les plus au premier axe sont sal_1 , sal_4 , pro_1 , pro_4 , c'est-à-dire les catégories se différenciant les plus parmi les modalités des variables *salaire brute* et *promo*. (Attention les notations : sal_1 sont les salaires les plus basses et sal_4 sont les plus hautes ; pro_1 sont les élèves les plus anciens alors que pro_4 sont les plus jeunes). Pour le deuxième axe, ce sont ad_7 (Militaire) et tep_2 (Administration) qui y contribuent les plus. On peut constater que, si à la première liste, on se

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.414	9.82	9.82
dim 2	0.34	8.05	17.9
dim 3	0.316	7.47	25.3
dim 4	0.271	6.42	31.8
dim 5	0.266	6.3	38.1
dim 6	0.261	6.17	44.2
dim 7	0.255	6.04	50.3
dim 8	0.253	6	56.3
dim 9	0.248	5.86	62.1
dim 10	0.240	5.7	67.8
dim 11	0.233	5.53	73.4
dim 12	0.227	5.37	78.7
dim 13	0.220	5.2	84
dim 14	0.203	4.82	88.8
dim 15	0.194	4.59	93.3
dim 16	0.167	3.96	97.3
dim 17	0.114	2.69	100

TABLE 3 – Les résultats relatifs aux modalités de la deuxième liste.

	Crđ Dim1	Crđ Dim2	Contrib Dim1	Contrib Dim2	Cos2 Dim1	Cos2 Dim2
ad 1	-0.088	-0.149	0.221	0.774	0.00695	0.0200
ad 2	1.62	0.965	2.91	1.26	0.0491	0.0174
ad 3	-2.24	1.24	2.65	0.994	0.0443	0.0136
ad 4	0.259	-0.862	0.594	8.03	0.0115	0.128
ad 5	-0.857	0.741	6.78	6.19	0.133	0.0993
ad 6	-1.17	-0.235	0.465	0.0226	0.00775	0.000310
ad 7	3.11	6.72	2.33	13.2	0.0388	0.181
ad 8	1.45	3.43	0.403	2.77	0.00671	0.0378
ad 9	0.61	0.0994	4.2	0.136	0.0857	0.00228
tep 1	-0.0341	-0.276	0.0589	4.68	0.00607	0.396
tep 2	1.07	2.19	4.4	22.5	0.0782	0.328
tep 3	-0.263	0.761	0.297	3.03	0.00532	0.0445
tep 4	-0.795	1.41	1.01	3.84	0.0172	0.0538
sal 1	1.09	0.507	18	4.75	0.414	0.0894
sal 2	0.285	-0.547	1.10	4.94	0.0244	0.0897
sal 3	-0.228	-0.501	0.781	4.59	0.0178	0.0859
sal 4	-1.02	0.434	17.3	3.83	0.41	0.0744
pro 1	-1.17	0.628	19.7	6.89	0.428	0.123
pro 2	-0.346	-0.675	1.47	6.83	0.0307	0.117
pro 3	0.285	-0.188	0.995	0.526	0.0207	0.00897
pro 4	0.817	0.0745	14.4	0.145	0.369	0.00307

TABLE 4 – Les résultats relatifs aux modalités de la deuxième liste.

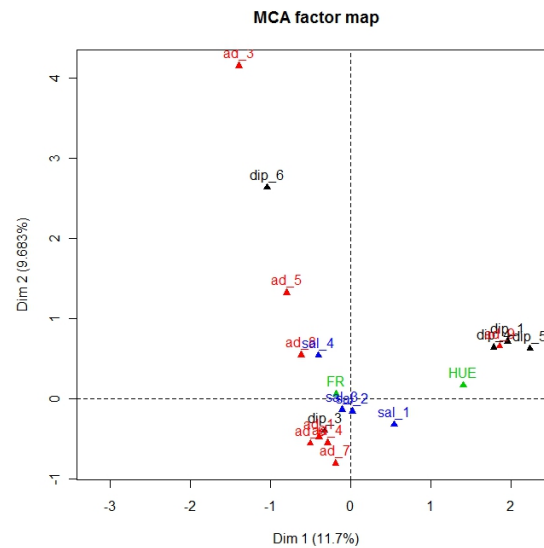


FIGURE 3 – Positionnement des modalités dans le groupe des élèves français.

concentrait aux parcours, alors dans cette deuxième, on va insister sur les types d'entreprises et les salaires, dans les relations plus ou moins avec les admissions.

2.3.3 Résultats graphiques

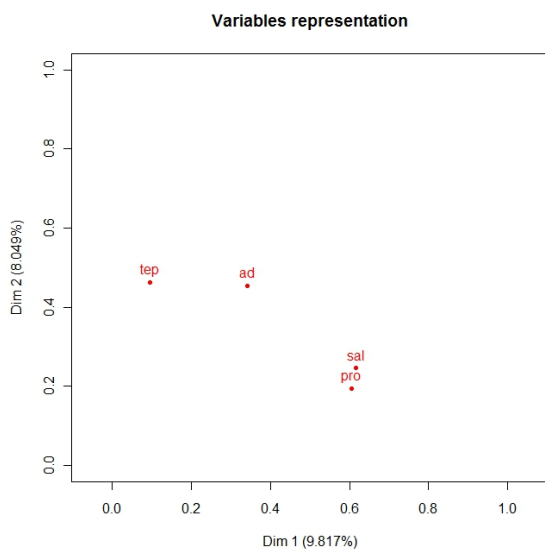


FIGURE 4 – Positionnement des modalités sur deux axes factoriels pour la deuxième liste.

Le positionnement des variables sur les deux premiers axes est montré par la Figure 4. On constate qu'il y a une forte corrélation entre les salaires et les promotions, c'est-à-dire que les salaires sont identifiées le plus par l'ancienneté des employés dans leurs entreprises. L'analyse sur la figure des modalités (la Figure 5) va montrer davantage cette relation.

Sur la Figure 5, c'est difficile de partitionner les modalités en groupes comme ce qu'on a fait avec la première liste. Le premier axe marque (même si pas vraiment clair) l'opposition entre

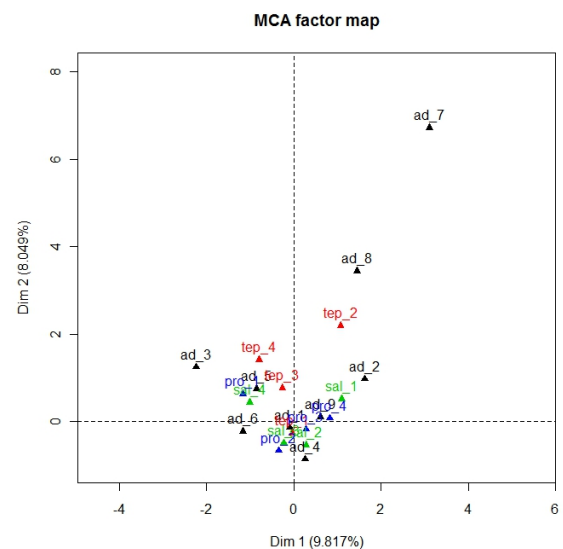


FIGURE 5 – Positionnement des modalités sur deux axes factoriels pour la deuxième liste.

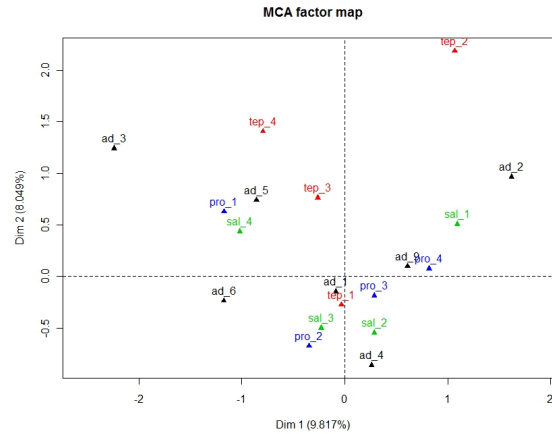


FIGURE 6 – Positionnement des modalités de plus près.

ad_7 (Militaire), ad_8 (Formation Prof), ad_2 (Double diplôme) avec ad_3 (Ec.Nat.Sup), ad_5 (Polytechnique) et ad_6 (Arts et Métiers). Le premier groupe correspond à tep_2 (Administration), et sal_1 (basses salaires), alors que le deuxième correspond à tep_4 (Libérale), sal_4 (haute salaire). Le deuxième axe est caractérisé par l'opposition privé-public. On a des admissions qui préfèrent les entreprises privées : ad_4 (Maitrise de Sciences), ad_1 (Prépa), ad_6 (Arts et Métiers), et ad_9 (Autres admissions). Les entreprises privées s'occupent des salaires moyennes, alors que dans le domaine public, les salaires peuvent être inférieures ou supérieures.

On va regarder cette figure de plus près pour découvrir les regroupements importants de modalités. La corrélation entre *salaire brute* et *promo* s'exprime par les distances proches entre les modalités de salaire et celle de promotion, par exemple, entre pro_1 et sal_4 , pro_2 et sal_3 . On constate aussi un groupe de salaires moyennes contenant ad_1 (Prépa), ad_4 (Maitrise de Sciences), sal_2 , sal_3 , pro_2 , pro_3 , tep_1 (Privé). Ce sont les élèves typiques de Télécom ParisTech qui préfèrent les entreprises privées. Il y a un autre groupe des élèves polytechniciens : ad_5 , pro_1 , sal_4 et tep_3 ; ils sont caractérisés par l'ancienneté, haute salaires, et entreprises publiques. En terme de promotions, les premières (pro_1) s'écarte des autres, marquant un changement ou une réforme probable au niveau d'admissions et de carrières. La modalité ad_9 (Autres admissions) concerne particulièrement pro_4 , indiquant que ces nouvelles admissions viennent juste d'être additionnées ces dernières années.

3 Activités professionnelles, entreprises, responsabilités et salaires

4 Conclusions