

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

VU Tuan Hung et DO Quoc Khanh

21 juin 2012

1 Introduction

1.1 Contexte de l'étude et objectif de l'analyse

L'étude est effectuée sur différents éléments au sein de la communauté des élèves de Télécom ParisTech, des promotions de 1962 à 2005. Les éléments pris en compte sont : les origines des élèves (leurs nationalités et leurs modes d'admission), leurs parcours (diplômes obtenus), leurs activités professionnelles (types d'entreprises pour lesquelles ils travaillent, leurs responsabilités au sein des entreprises, salaires reçus). L'objectif de l'étude est de mieux comprendre les relations entre ces éléments et les tendances des élèves de Télécom ParisTech dans leurs choix de parcours et de carrières. Un autre point intéressant à éclairer pendant cette étude est l'évolution de ces éléments au cours des 25 promotions de l'école, et ceci est pour mieux comprendre et prédire des grandes tendances de notre temps et de notre avenir.

1.2 Description des données utilisées

Dans le cadre de cette étude, on utilise les données issues de l'enquête *Fonctions et Salaires édition 2009* sur les promotions de 1962 à 2005. L'enquête contient plusieurs questions concernant les origines, les parcours, les diplômes, et les activités après la graduation. La plupart de questions sont à répondre en multiples choix ; les données reçues sont donc représentées en plusieurs variables catégorielles, dont celles qu'on utilise pour les analyses sont : *diplôme*, *promotion*, *admission*, *origine*, *type d'entreprise*, *taille d'entreprise*, *responsabilité*, *activité professionnelle*, *création d'entreprise* et *salaire brut*. Le nombre de variables intéressantes est assez grand, ceci rend difficile de les analyser en une seule fois. On a dû diviser ces variables en deux groupes correspondant à deux thèmes principaux : parcours/origine et création d'entreprise/domaine professionnelle. Le Tableau 1 montre quelques statistiques simples de ces variables. La Figure 1 représente l'histogramme de salaires bruts. La manière de traiter les variables quantitatives *promo* et *salaire brut* sera présentée à la section 1.3.

Summary	Diplôme	Promotion	Admission	Origine	Type d'entreprise
	1 :173	28 : 93	1 :593	FR :1114	1 :1050
	2 : 3	30 : 88	9 :235	HUE : 94	2 : 80
	3 :961	31 : 88	5 :192	UE : 47	3 : 89
	4 : 5	29 : 83	4 :184	NA	4 : 33
	5 : 23	27 : 77	2 : 23	NA	NA's : 3
	6 : 90	26 : 64	(Other) : 27	NA	NA
	NA	(Other) :762	NA's : 1	NA	NA

Création d'entreprise	Taille d'entreprise	Responsabilité	Activité principale	Salaire brut
1 : 115	8 :567	1 :280	2 :213	Min. : 9000
2 :1136	6 :153	2 :199	1 :123	1st Qu. : 52550
NA's : 4	7 :115	3 : 86	11 : 92	Median : 70000
NA	4 :103	4 :168	6 : 85	Mean : 123785
NA	2 : 99	5 :314	4 : 84	3rd Qu. : 102750
NA	(Other) :212	6 :195	(Other) :630	Max. :15000000
NA	NA's : 6	NA's : 13	NA's : 28	NA's :45

TABLE 1 – Statistiques simples de Donnée

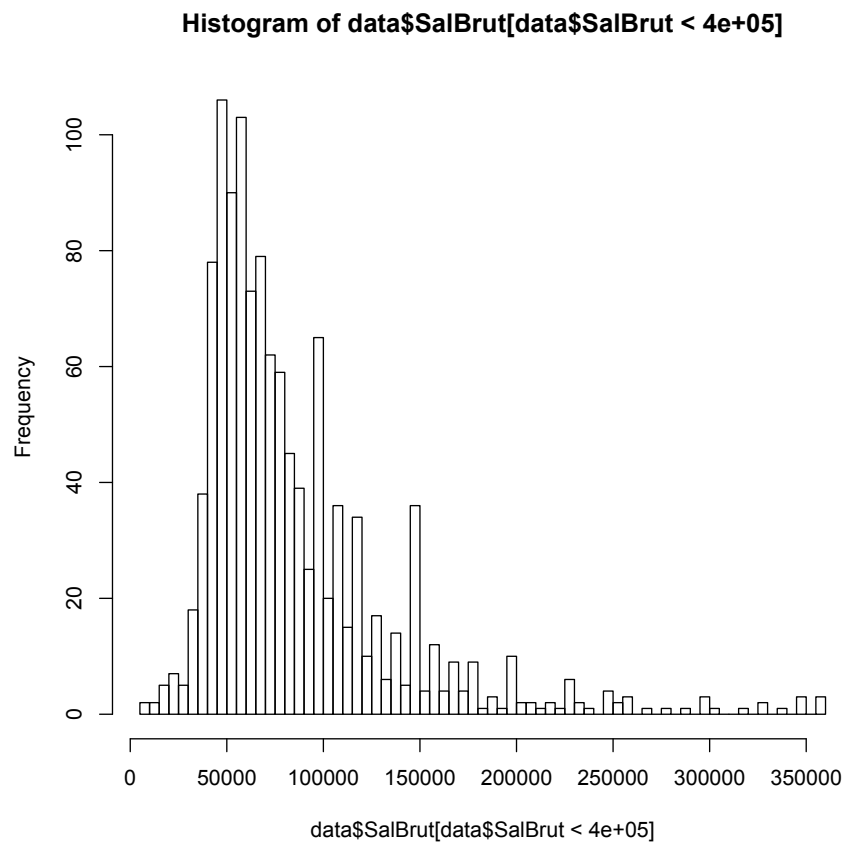


FIGURE 1 – L’histogramme de Salaire brut

1.3 Descriptif des méthodes utilisées

On ne traitera que les variables catégorielles par l'analyse de correspondances multiples. La variable *promotion* peut être traité à la fois quantitativement et qualitativement. Cependant, même si elle est traitée comme une variable catégorielle, elle a beaucoup de catégories par rapport aux autres variables, ce qui pourrait diminuer son impact (L'impact d'une variable catégorielle dépend de son nombre de modalités). Pour résoudre ce problème, on divise les valeurs de cette variable en 4 intervalles (équilibrés en nombre d'individuels) 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 brut*.

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 les remplacer par d'autres valeurs convenables. Dans ce projet, l'analyse de correspondances multiples (ACM) s'effectue avec les tableaux disjonctifs complets créés par la méthode d'imputation. Cette méthode est pour traiter les données manquantes en les remplacer non par les valeurs 0 et 1 (les deux valeurs présentes dans un tableau disjonctif complet) mais par les estimations de probabilités que ces cases vides prennent la valeur 1. D'une manière plus formelle, on traite ces cases comme des variables aléatoires de valeurs discrètes $\{0, 1\}$, et affecte leurs espérances à ces cases vides. Le ACM est ensuite lancé comme avec un tableau disjonctif complet normal.

2 L'analyse sur parcours, origines, entreprises et salaires

Dans cette première partie, on étudie les relations entre les parcours, les origines des élèves, les types d'entreprises dans lesquelles ils travaillent après leur obtention de diplômes, et le niveau de salaire ils pourront recevoir. Les relations entre les admissions et les diplômes qu'ils obtiennent reflètent en quelque sorte la cohérence dans les parcours des élèves, constituant donc un élément important à étudier.

2.1 Liste 1

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

2.1.1 Valeurs propres

Les axes factoriels sont montrés dans le Tableau 2. 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 l'analyse de données. Dans ce qui suit, on utilisera principalement les deux premiers axes.

2.1.2 Résultats relatifs aux modalités

Les résultats relatifs aux modalités sont montrés au Tableau 3, quelques éléments importants sont en gras. Les deux modalités contribuant les plus au premier axe sont *dip₁* (Mastère) et *ad₉* (Autres admissions). Alors qu'au second axe, ce sont les modalités *dip₆* (Ingénieur du corps) et *ad₅* (Polytechnique) qui contribuent les plus. On retient ces modalités qui pourraient caractériser ces axes. Par exemple, la contribution de *dip₆* sur le second axe est la plus positive parmi les modalités de la même variable ; le second axe marquerait probablement l'opposition entre le groupe des élèves d'ingénieur du corps et le reste. On observe ainsi que les variables *FR.UE.HUE* et *salaire brut* jouent les rôles peu importants dans la formation de ces 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 2 – Les axes factoriels de la première liste.

	Crđ Dim1	Crđ 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 3 – Les résultats relatifs aux modalités de la première liste.

2.1.3 Résultats graphiques

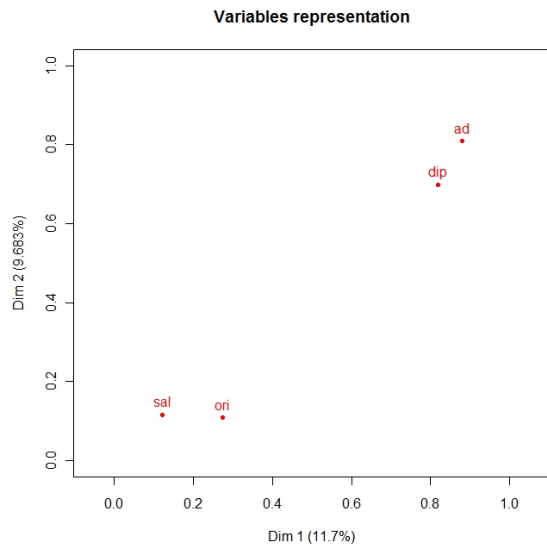


FIGURE 2 – Positionnement des modalités sur deux axes factoriels.

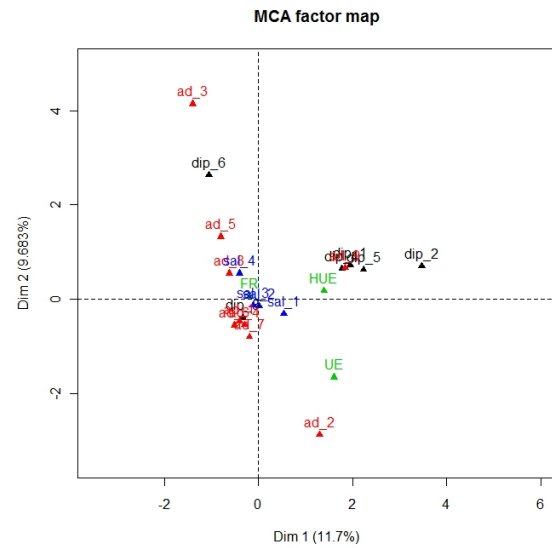


FIGURE 3 – Positionnement des modalités sur deux axes factoriels.

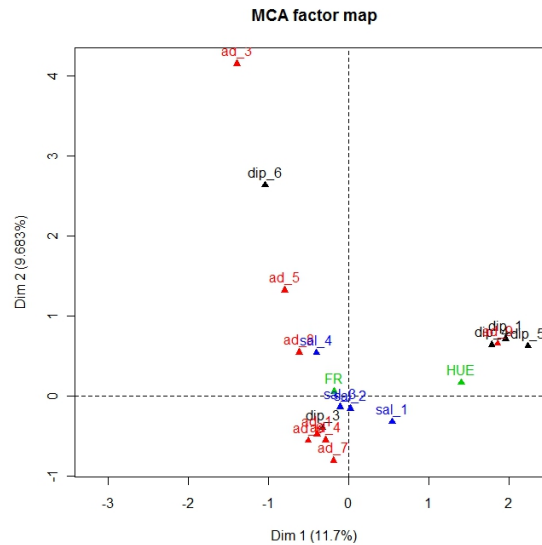


FIGURE 4 – Positionnement des modalités dans le groupe des élèves français.

La figure 2 montre qu'il y a une forte corrélation entre *diplôme* et *admission*, alors que les autres relations sont moins remarquables. Le premier axe est caractérisé par l'opposition entre *dip₂* (Master of Sciences) et *dip₆* (Ingénieur du corps), entre *ad₉* (Autres admissions), *ad₂* (Double diplôme) et *ad₃* (Ec.Nal.Sup). Le deuxième axe marque une forte opposition entre *ad₃* (Ec.Nal.Sup) et *ad₂* (Double diplôme), entre *dip₆* (Ingénieur du corps) et *dip₃* (Ingénieur). En gros, c'est l'opposition entre les élèves qui font un parcours traditionnel d'ingénieur français (y compris les ingénieurs du corps) et les élèves intégrant l'école des autres universités européens pour obtenir leur double diplôme. En regardant la figure 3, on peut partitionner les modalités (et donc les élèves) en 3 groupes. Le premier groupe contient *ad₉* (Autres admissions), *dip₂* (Master of Sciences), *dip₅* (Docteur), *dip₁* (Mastère), *dip₄* (DEA STN) et *HUE*. Ce sont

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

les élèves hors européens qui ont une tendance de faire des parcours plutôt scientifiques. Le deuxième groupe contient ad_2 (Double diplôme) et UE correspondant aux 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.

En considérant de plus près le groupe des élèves français (Figure 4), on constate que les modalités ad_1 (Prépa), ad_4 (Maitrise 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) correspondent aux 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.

2.2 Liste 2

Avec la première liste, on remarque une forte corrélation entre *diplôme* et *admission*. Dans la deuxième liste, on prendre la variable *admission* comme le représentant de ces deux variables (donc le représentant de l'élément *parcours des élèves*). On y ajoute la variable *promo*, et deux autres reflétant les carrières des élèves, *Type d'entreprise* et *salaire brut*. Les deux variables quantitatives *salaire brut* et *promo* sont factorisées en 4 modalités par variable selon les indications à la section 1.3.

2.2.1 Valeurs propres

Les 17 axes factoriels sont montrés au Tableau 5. La contribution des axes à la variance totale descend plutôt lentement.

	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 5 – Les résultats relatifs aux modalités de la deuxième liste.

2.2.2 Les résultats relatifs aux modalités

Les résultats relatifs aux modalités sont montrés au Tableau 5. 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 brut* 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 concentrait aux parcours, alors dans cette deuxième, on va insister sur les activités professionnelles, dans les relations plus ou moins avec les éléments qu'on a analysés via la première liste.

2.2.3 Résultats graphiques

Le positionnement des variables sur les deux premiers axes est montré par la Figure 5. 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 6) va montrer davantage cette relation.

Sur la Figure 6, c'est difficile de partitionner les modalités en groupes comme ce qu'on a fait à la première liste. Le premier axe marque (même si pas vraiment clair) l'opposition entre 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 (Maîtrise de Sciences), ad_1 (Prépa), ad_6 (Arts et Métiers), et ad_9

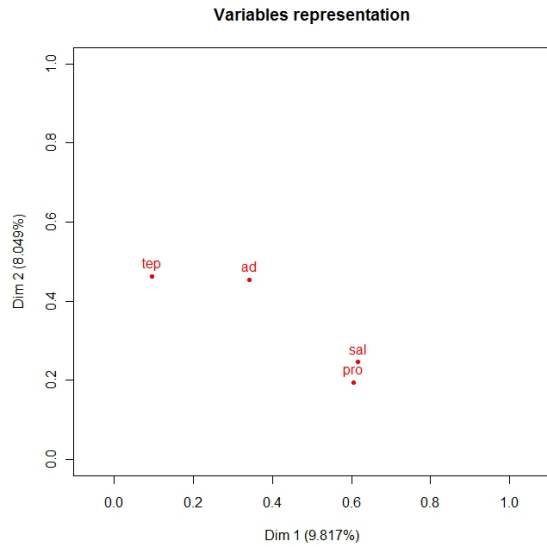


FIGURE 5 – Positionnement des modalités sur deux axes factoriels pour la deuxième liste.

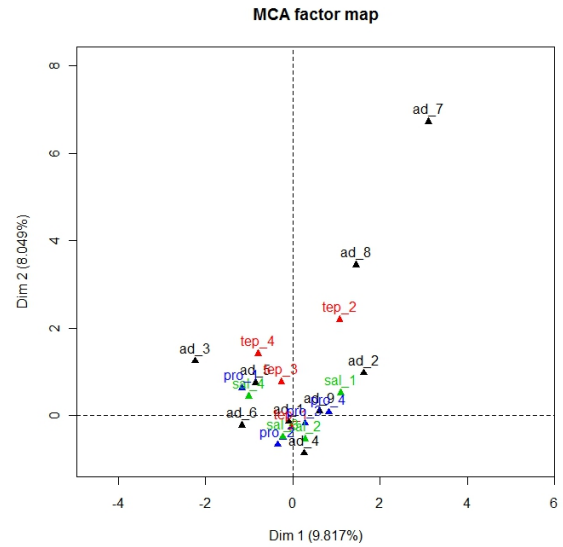


FIGURE 6 – Positionnement des modalités sur deux axes factoriels pour la deuxième liste.

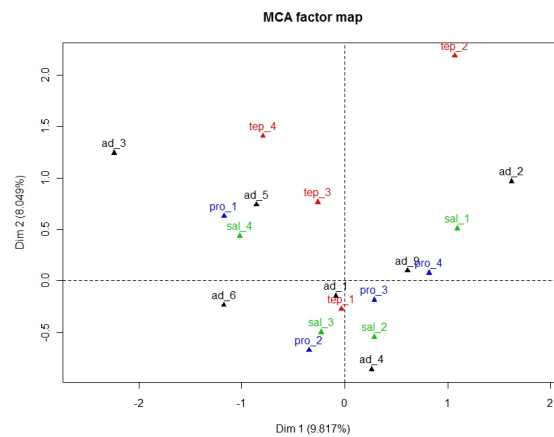


FIGURE 7 – Positionnement des modalités de plus près.

(Autres admissions). Les entreprises privées correspondent à 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 brut* 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 (Maîtrise 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 introduites ces dernières années.

3 L'analyse sur activités professionnelles, entreprises, responsabilités et salaires

Dans cette deuxième partie, on étudie les relations entre le domaine d'activité, la responsabilité au sein de l'entreprise, la taille de l'entreprise et le niveau de salaire exprimée chez les élèves de Télécom ParisTech. Aussi, les relations entre les types d'entreprises, la création d'entreprise et la responsabilité sont à étudier.

3.1 Liste 1

On considère quatre variables : *Taille d'entreprise*, *Responsabilités*, *Activités professionnelles* et *salaire brut*

3.1.1 Valeurs propres

Le Tableau 6 montre les 32 axes factoriels. On constate que la contribution des axes à la variance totale descend plutôt lentement.

3.1.2 Résultats relatifs aux modalités

Les résultats relatifs aux modalités sont montrés au Tableau 7. Les modalités contribuant les plus au premier axe sont ta_1 , ta_2 , ta_8 (Taille d'Entreprise), re_3 (Responsabilité). Attention les notations : sal_1 sont les salaires les plus basses et sal_4 sont les plus hautes ; ta_1 sont les entreprises les plus petite alors que ta_8 sont les plus grandes. Pour le deuxième axe, ce sont re_2 , re_4 , sal_1 , sal_4 qui y contribuent les plus. C'est-à-dire les catégories se différenciant les plus parmi les modalités des variables *responsabilité* (sur le premier axe) et *salaire brut* (sur le deuxième axe). On va le vérifier avec l'analyse des résultats graphiques suivante.

3.1.3 Résultats graphiques

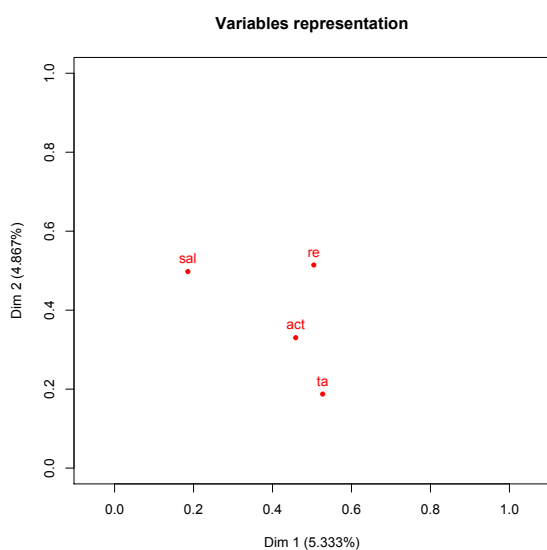


FIGURE 8 – Positionnement des modalités sur deux axes factoriels pour la première liste.

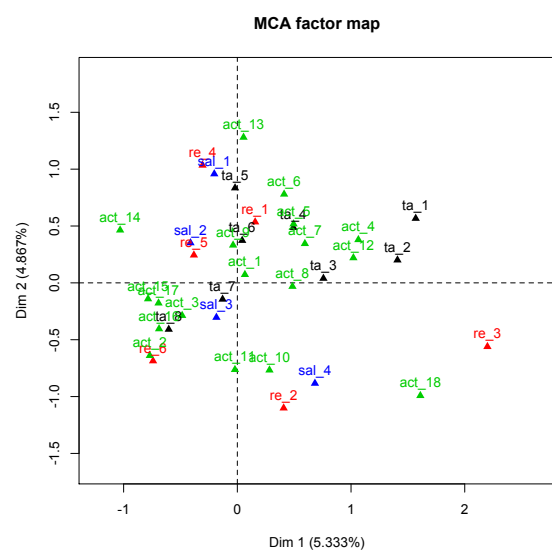


FIGURE 9 – Positionnement des modalités sur deux axes factoriels pour la première liste.

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.419	5.33	5.33
dim 2	0.383	4.87	10.2
dim 3	0.327	4.17	14.4
dim 4	0.307	3.91	18.3
dim 5	0.302	3.85	22.1
dim 6	0.293	3.73	25.9
dim 7	0.285	3.63	29.5
dim 8	0.279	3.55	33
dim 9	0.274	3.48	36.5
dim 10	0.269	3.42	39.9
dim 11	0.267	3.4	43.3
dim 12	0.259	3.29	46.6
dim 13	0.253	3.21	49.8
dim 14	0.248	3.16	53
dim 15	0.245	3.12	56.1
dim 16	0.244	3.11	59.2
dim 17	0.244	3.1	62.3
dim 18	0.242	3.08	65.4
dim 19	0.238	3.03	68.4
dim 20	0.236	3	71.4
dim 21	0.228	2.91	74.3
dim 22	0.223	2.84	77.2
dim 23	0.22	2.8	80
dim 24	0.213	2.71	82.7
dim 25	0.207	2.64	85.3
dim 26	0.205	2.61	87.9
dim 27	0.191	2.43	90.4
dim 28	0.189	2.41	92.8
dim 29	0.181	2.31	95.1
dim 30	0.157	2	97.1
dim 31	0.125	1.59	98.7
dim 32	0.104	1.32	100

TABLE 6 – Les axes factoriels de la première liste.

	Crđ Dim1	Crđ Dim2	Contrib Dim1	Contrib Dim2	Cos2 Dim1	Cos2 Dim2
ta 1	1.57	0.577	9.77	1.45	0.176	0.0239
ta 2	1.41	0.2	9.38	0.206	0.171	0.00344
ta 3	0.757	0.0383	1.15	0.00322	0.0199	5.11e-05
ta 4	0.49	0.482	1.18	1.25	0.0217	0.021
ta 5	-0.0226	0.83	0.00213	3.14	3.87e-05	0.052
ta 6	0.0448	0.375	0.0147	1.13	0.000282	0.0197
ta 7	-0.129	-0.146	0.0914	0.129	0.00169	0.00218
ta 8	-0.603	-0.409	9.85	4.96	0.304	0.14
re 1	0.159	0.537	0.339	4.26	0.00742	0.0852
re 2	0.407	-1.11	1.58	12.8	0.0318	0.235
re 3	2.2	-0.561	20.1	1.43	0.368	0.0239
re 4	-0.309	1.03	0.77	9.45	0.0151	0.169
re 5	-0.383	0.243	2.21	0.977	0.0499	0.0202
re 6	-0.74	-0.683	5.1	4.77	0.102	0.0872
act 1	0.063	0.0692	0.0237	0.0314	0.000452	0.000546
act 2	-0.77	-0.639	6.12	4.61	0.126	0.087
act 3	-0.482	-0.285	0.327	0.125	0.00574	0.00201
act 4	1.06	0.375	4.63	0.633	0.0855	0.0107
act 5	0.494	0.515	0.87	1.03	0.0159	0.0172
act 6	0.417	0.787	0.722	2.82	0.0134	0.0475
act 7	0.596	0.351	0.92	0.349	0.0166	0.00574
act 8	0.485	-0.0337	0.39	0.00206	0.00689	3.32e-05
act 9	-0.04	0.33	0.00403	0.301	7.21e-05	0.00491
act 10	0.288	-0.757	0.109	0.826	0.00191	0.0133
act 11	-0.0161	-0.756	0.00116	2.8	2.15e-05	0.0475
act 12	1.02	0.219	3.56	0.179	0.0649	0.00299
act 13	0.0558	1.28	0.00941	5.41	0.00017	0.0892
act 14	-1.03	0.464	2.36	0.525	0.0419	0.0085
act 15	-0.788	-0.146	1.65	0.0623	0.0295	0.00102
act 16	-0.695	-0.414	0.934	0.364	0.0165	0.00586
act 17	-0.698	-0.182	1.48	0.111	0.0267	0.00181
act 18	1.61	-0.985	3.31	1.36	0.0585	0.0219
sal 1	-0.201	0.959	0.607	15.1	0.0141	0.322
sal 2	-0.419	0.344	2.35	1.73	0.0523	0.0352
sal 3	-0.176	-0.295	0.459	1.41	0.0106	0.0297
sal 4	0.681	-0.889	7.64	14.3	0.183	0.312

TABLE 7 – Les résultats relatifs aux modalités de la premiere liste.

La Figure 8 montre le positionnement des variables sur les deux premiers axes. On constate qu'il n'y a pas forte corrélation entre les variables. Néanmoins, on peut constater une légère corrélation entre le domaine professionnel, la taille d'entreprise et la responsabilité. Une possible conclusion est que, pour chaque domaine et chaque taille d'entreprise, les élèves de Télécom ParisTech occupent les responsabilités et les salaires typiques correspondantes.

Sur la Figure 9, on peut partitionner les modalités en groupes. En gros, le premier axe est caractérisé par l'opposition entre ta_1 (petite entreprises) et ta_8 (grande entreprises). Le deuxième axe marque l'opposition entre sal_1 et sal_4 (salaire brut). Les différents domaines d'activités sont distribués selon les tailles d'entreprise. Sur le premier axe, on peut voir les quatre principaux groupes dépendant des salaires bruts :

- Le premier groupe contient $re_{4,1}$ (Ingénieur d'Études/Experts), ta_5 (250-499 salariés) et sal_1 . Les activités professionnelles liées sont Recherche/Enseignement et Conseil en Technologie. Ce sont les élèves qui s'occupent des postes d'ingénieur ou de chercheur dans les entreprises moyennes de conseil, ou bien dans les établissements éducationnels, et dont le niveau de salaire est plutôt bas.
- Le deuxième groupe contient re_5 (Chef de Projet), ta_6 (500-1999 salariés) et sal_2 . Les activités professionnelles liées sont Industrie des télécommunications, Editeur de logiciels et Industries électroniques. Ce sont les élèves dans le secteur d'électronique, de télécommunication ou d'informatique, qui s'occupent des postes de chef de projet dans les entreprises assez grandes, et reçoivent donc les salaires moyennes.
- Le troisième groupe contient re_6 (Responsable de Service), $ta_{7,8}$ (de 2000 salariés et plus) et sal_3 . Les activités professionnelles liées sont Opérateur de télécommunications, Constructeur informatique, Industries automobiles et des transport, Energie, Aéronautique, espace. Les domaines d'activités sont plutôt traditionnels et conservatoires, dont le coût d'introduction aux marchés est important, ceci ne permet que les grandes et très grandes entreprises d'exercer leurs activités. Les salaires dans ces entreprises sont supérieures à la moyenne.
- Le quatrième groupe contient re_2 (Directeur d'unité) et sal_4 . Les activités professionnelles liées sont Banque - Assurance et autre services. Le secteur financier présente des postes typiques de type Directeur d'unité dont la responsabilité est haute accompagnant à de très hauts salaires (et primes). Ce sont eux qui reçoivent les montants de salaires les plus importants.

3.2 Liste 2

On considère quatre variables : *Taille d'entreprise*, *Responsabilités*, *création d'entreprise* et *Salaire brut*

3.2.1 Valeurs propres

Le Tableau 8 montre les 16 axes factoriels de données. Dans ce qui suit, on utilisera principalement les deux premiers axes.

3.2.2 Résultats relatifs aux modalités

Les résultats relatifs aux modalités sont montrés au Tableau 9. Les modalités contribuant les plus au premier axe sont ta_1 (taille d'entreprise), cr_1 (création d'entreprise), re_3 (responsabilité). Pour le deuxième axe, ce sont re_2 (responsabilité), sal_1 , sal_4 (salaire brut) qui y contribuent les plus.

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.484	12.3	12.3
dim 2	0.352	8.91	21.2
dim 3	0.301	7.62	28.8
dim 4	0.276	6.99	35.8
dim 5	0.269	6.8	42.6
dim 6	0.255	6.46	49
dim 7	0.251	6.36	55.4
dim 8	0.248	6.28	61.7
dim 9	0.24	6.08	67.8
dim 10	0.235	5.95	73.7
dim 11	0.234	5.92	79.6
dim 12	0.213	5.38	85
dim 13	0.212	5.35	90.3
dim 14	0.171	4.32	94.7
dim 15	0.118	2.98	97.6
dim 16	0.0932	2.36	100

TABLE 8 – Les axes factoriels de la deuxième liste.

	Crđ.Dim.1	Crđ.Dim.2	Contrib.Dim.1	Contrib.Dim.2	Cos2.Dim.1	Cos2.Dim.2
ta 1	2.65	1.05	24.1	5.18	0.501	0.0785
ta 2	0.822	-0.0107	2.76	0.000645	0.0584	9.91e-06
ta 3	0.136	-0.0386	0.0323	0.00356	0.00065	5.21e-05
ta 4	-0.17	-0.0906	0.124	0.0481	0.00262	0.000743
ta 5	-0.184	0.786	0.122	3.05	0.00254	0.0465
ta 6	-0.328	-0.0122	0.682	0.0013	0.0151	2.1e-05
ta 7	-0.315	-0.319	0.471	0.663	0.0101	0.0103
ta 8	-0.329	-0.185	2.54	1.1	0.0906	0.0286
cr 1	2.62	0.466	32.5	1.41	0.694	0.022
cr 2	-0.265	-0.0471	3.28	0.143	0.694	0.022
re 1	0.186	0.476	0.403	3.63	0.0102	0.0669
re 2	0.106	-1.53	0.0936	26.5	0.00218	0.449
re 3	2.38	-0.394	20.4	0.768	0.43	0.0118
re 4	-0.291	1.1	0.593	11.6	0.0134	0.191
re 5	-0.438	0.382	2.5	2.61	0.0655	0.0497
re 6	-0.478	-0.516	1.84	2.96	0.0426	0.0498
sal 1	-0.105	0.984	0.142	17.3	0.00381	0.338
sal 2	-0.44	0.406	2.23	2.61	0.0574	0.0489
sal 3	-0.155	-0.259	0.31	1.19	0.00827	0.0231
sal 4	0.59	-0.99	4.97	19.2	0.138	0.387

TABLE 9 – Les résultats relatifs aux modalités de la deuxième liste.

3.2.3 Résultats graphiques

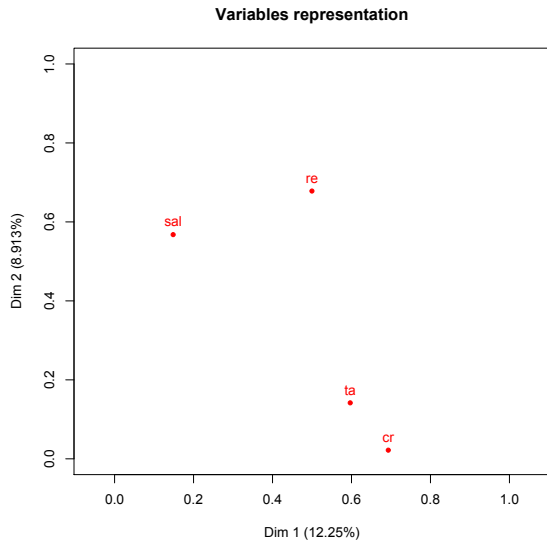


FIGURE 10 – Positionnement des modalités sur deux axes factoriels pour la deuxième liste.

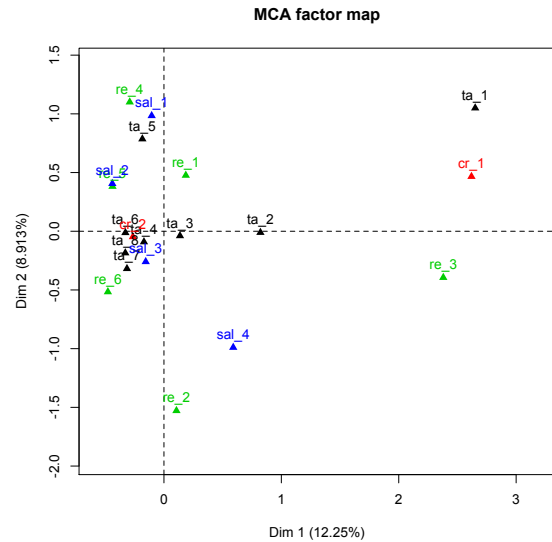


FIGURE 11 – Positionnement des modalités sur deux axes factoriels pour la deuxième liste.

Le positionnement des variables sur les deux premiers axes est montré par la Figure 10. On peut voir qu'il y a une forte corrélation entre *taille d'entreprise* et *création d'entreprise*. C'est raisonnable étant donné que la majorité des élèves ne peuvent pas créer immédiatement des grandes entreprises, mais des petites entreprises privées.

D'une même manière décrites à la section 3.1, et sur la Figure 9, on peut partitionner les modalités en des groupes typiques. En gros, le premier axe est caractérisé par l'opposition entre cr_1 (crée une entreprise) et cr_2 (n'avoir pas une entreprise). Le deuxième axe marque l'opposition entre sal_1 et sal_4 (salaire brut). On peut indiquer deux groupes principales, selon le premier axe :

- Le premier groupe contient cr_1 (crée une entreprise), ta_1 (mois de 10 salariés) et re_3 (Directeur Général). C'est le monde de start-up et d'entrepreneurs. On voit aussi que ces modalités se trouvent au dessus du premier axe (sauf le re_3 , les directeurs), indiquant que les salaires de ces jeunes entrepreneurs sont un peu moins importants (comme on a présenté en dessus, le deuxième axe marque l'opposition entre le sal_1 (bas salaires) et sal_4 (hauts salaires)).
- Le deuxième groupe contient cr_2 (n'avoir pas une entreprise), $ta_{6,7,8}$ (grandes entreprises) et $sal_{2,3}$. Ce sont les élèves intégrant dans les grandes entreprises, s'occupant de postes variés et recevant des niveaux moyens de salaire.

4 Conclusions

Au cours de cette étude, on a analysé les différents types de relations entre les éléments concernant les origines, les parcours, et les activités professionnelles après la graduation des élèves de Télécom ParisTech des 25 promotions de 1962 à 2005. Ces relations sont complexes ; on a donc dû appliquer certaines méthodes spéciales pour les illustrer, comme par exemple diviser les variables en petites listes, appliquer la méthode d'imputation pour traiter les données manquantes.

Les résultats obtenus sont les différents constats qu'on a présentés en dessus. On a montré la corrélation étroite entre les nationalités, les modes d'admission et les diplômes obtenus par les élèves après l'école. D'une manière générale, on peut diviser les élèves en trois groupes suivant leurs nationalités : français, UE et hors UE. Chaque groupe possède ses propres modes d'admission et ses préférences en terme de diplôme. Mais peu importe le diplôme obtenu et les origines, le niveau de salaire ne varie pas beaucoup.

Alors que le montant de salaire ne dépend pas énormément aux diplômes (sauf quelques exceptions, par exemple le groupe de polytechniciens travaillant dans le secteur public et recevant de hauts revenus), ce montant dépend fortement de l'ancienneté dans l'entreprise et des éléments professionnels. On a découvert les relations étroites entre ces derniers : pour chaque domaine d'activité, on a des entreprises de taille typique, des postes typiques occupés par les élèves, et les niveaux de salaire correspondants. Les autres relations concernant l'opposition privé/public et start-up/non start-up ont aussi été remarquées. Finalement, les voies d'admission et les parcours ont évolué au cours des ces 25 promotions, illustré par l'écart entre les premières promotions et les autres plus jeunes.