

Sur le fichier de M. Arnaud

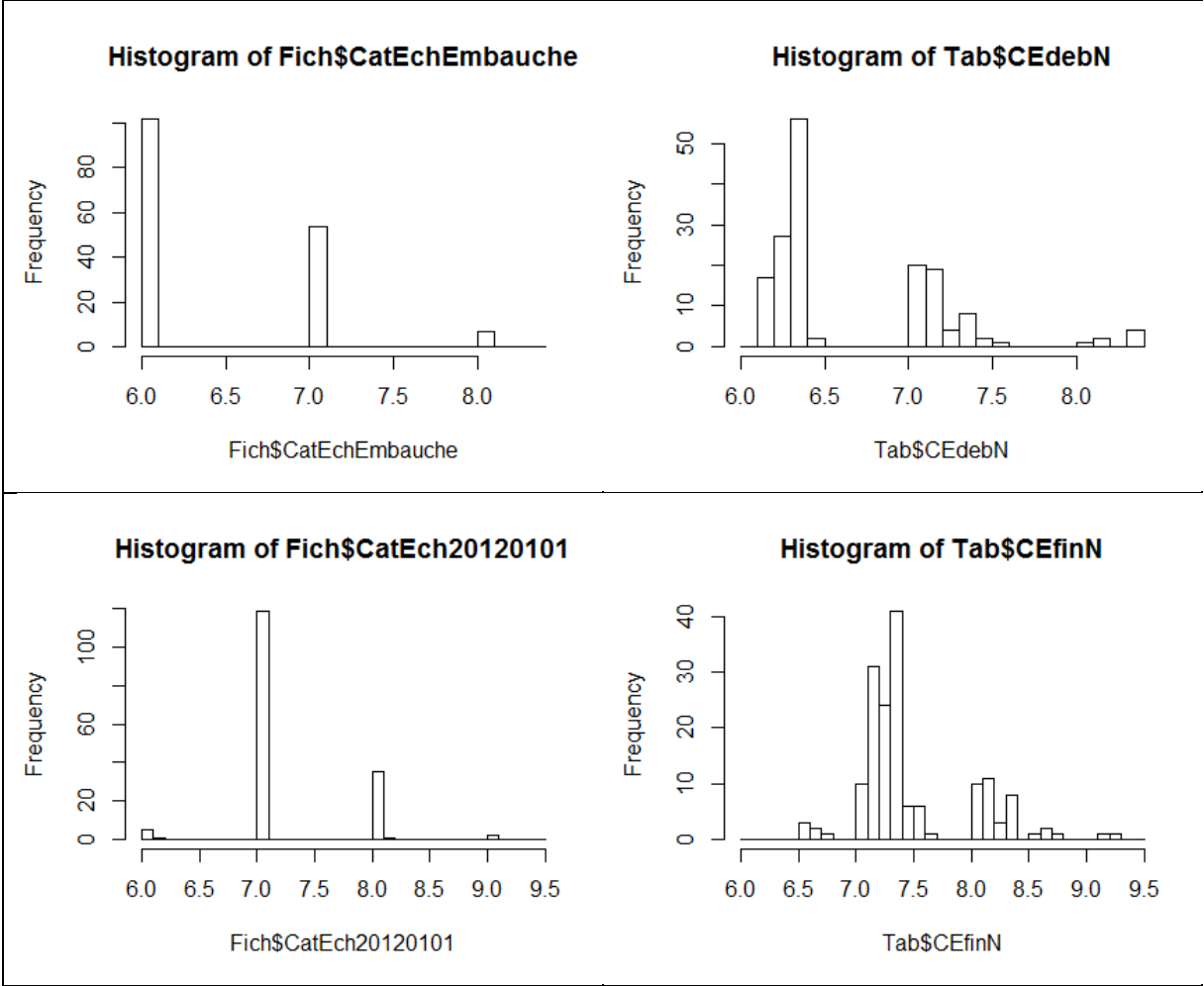
A partir du fichier initial, on construit une table, avec :

- ID : identifiant du salarié ;
- Sexe : sexe du salarié (M ou F) ;
- ExpCirad : variable OUI/NON en fonction d'une expérience au Cirad avant embauche en CDI ;
- AgeDeb : l'âge au recrutement ;
- CEdebN : la catégorie/échelon standardisée (obtention d'une variable continue) de recrutement ;
- CEFinN : la catégorie/échelon standardisée (obtention d'une variable continue) au 01/01/2012 ;
- Vitesse : $(CEfinN - CEdebN) / \text{durée de présence}$;
- NbChangt : nombre de changement de catégorie ;
- TpsChangt1 : temps (année) nécessaire au premier changement de catégorie (temps de présence si NbChangt = 0) ;
- TpsMoyen : si NbChangt = 2 moyenne des temps nécessaires pour changer de catégories et si NbChangt < 2 on prend TpsChangt1.

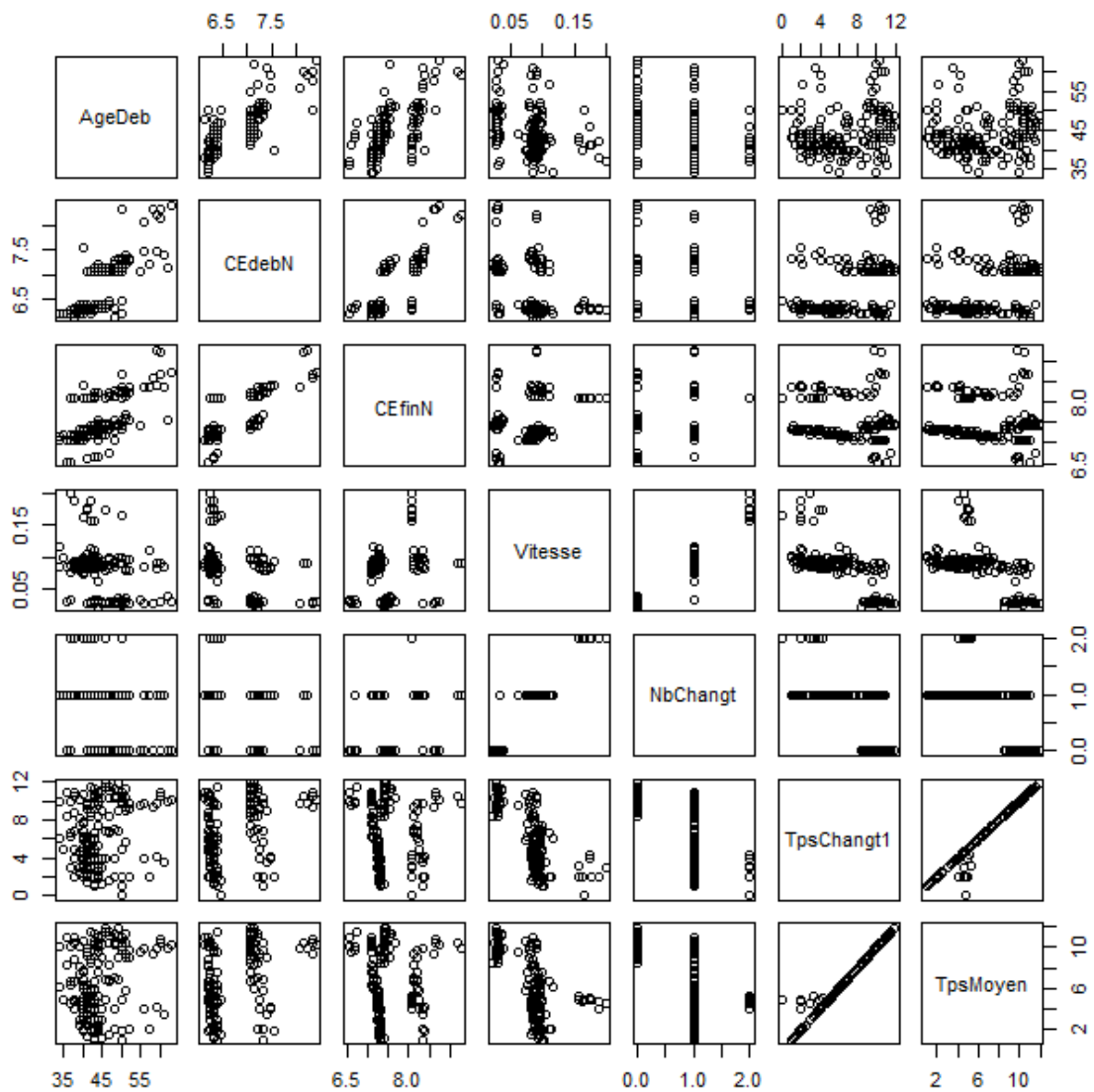
```
> head(Tab)
```

	ID	Sexe	ExpCirad	AgeDeb	CEdebN	CEfinN	Vitesse	NbChangt	TpsChangt1	TpsMoyen
1	1	M	OUI	51	7.267	8.133	0.08375	1	9.337	9.337
2	2	M	NON	46	7.133	7.467	0.02902	0	11.510	11.510
3	3	M	NON	42	7.067	7.400	0.02893	0	11.510	11.510
4	4	M	OUI	59	7.467	8.400	0.08232	1	4.162	4.162
5	5	M	NON	44	6.333	7.333	0.09118	1	1.959	1.959
6	6	M	OUI	43	6.400	7.333	0.08628	1	1.967	1.967

Avec la standardisation des catégories/échelons (chaque échelon représente un 1/15 de sa catégorie), on obtient une variable catégorie/échelon qui tend vers une variable continue.



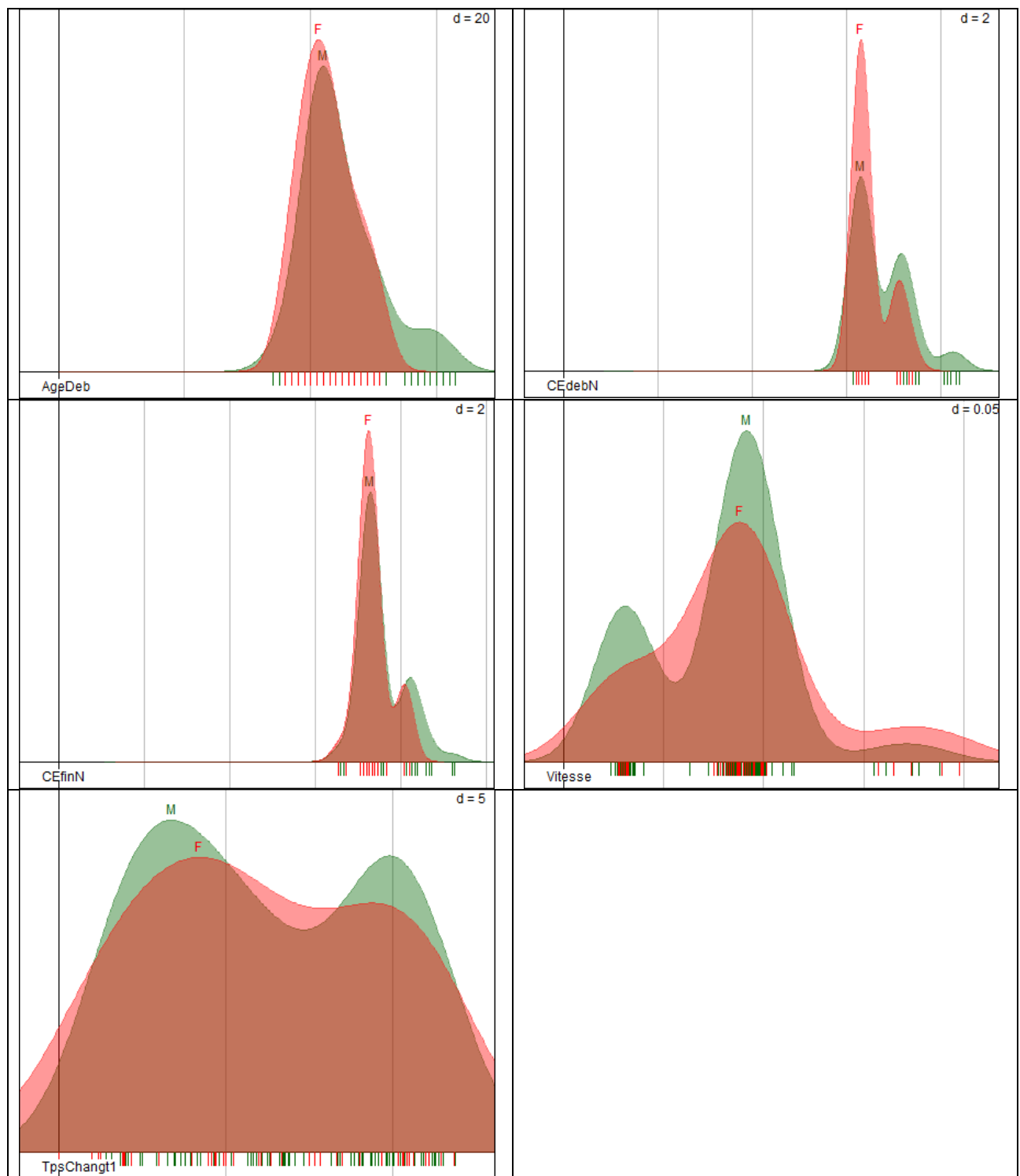
Voilà un aperçu des données



```
> cor(Tab[,4:length(Tab[1,])])
      AgeDeb      CEdebN      CEfinN      Vitesse      NbChangt      TpsChangt1      TpsMoyen
AgeDeb  1.000000  0.788645  0.661939  -0.305132  -0.325095  0.240570  0.237805
CEdebN  0.788645  1.000000  0.744912  -0.475502  -0.521035  0.472794  0.463778
CEfinN  0.661939  0.744912  1.000000  0.217256  0.155404  -0.033739  0.017450
Vitesse -0.305132 -0.475502  0.217256  1.000000  0.962683  -0.738593  -0.654182
NbChangt -0.325095 -0.521035  0.155404  0.962683  1.000000  -0.728266  -0.655513
TpsChangt1 0.240570  0.472794 -0.033739 -0.738593 -0.728266  1.000000  0.982433
TpsMoyen 0.237805  0.463778  0.017450 -0.654182 -0.655513  0.982433  1.000000
```

L'âge d'embauche est corrélé à la catégorie/échelon d'embauche. Celle de fin à celle de début. Le temps de premier de changement est fortement corrélé à celui de second changement (c'est logique car peu de personnes ont changé 2 fois de catégorie, à savoir 10).

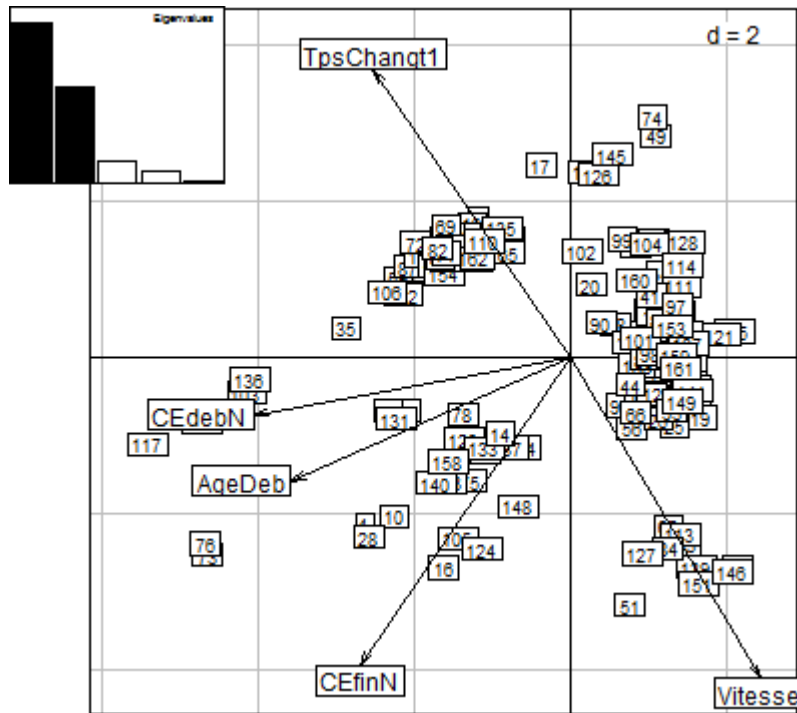
En termes de répartition femme/homme, voilà les distributions pour chacune de ces variables.



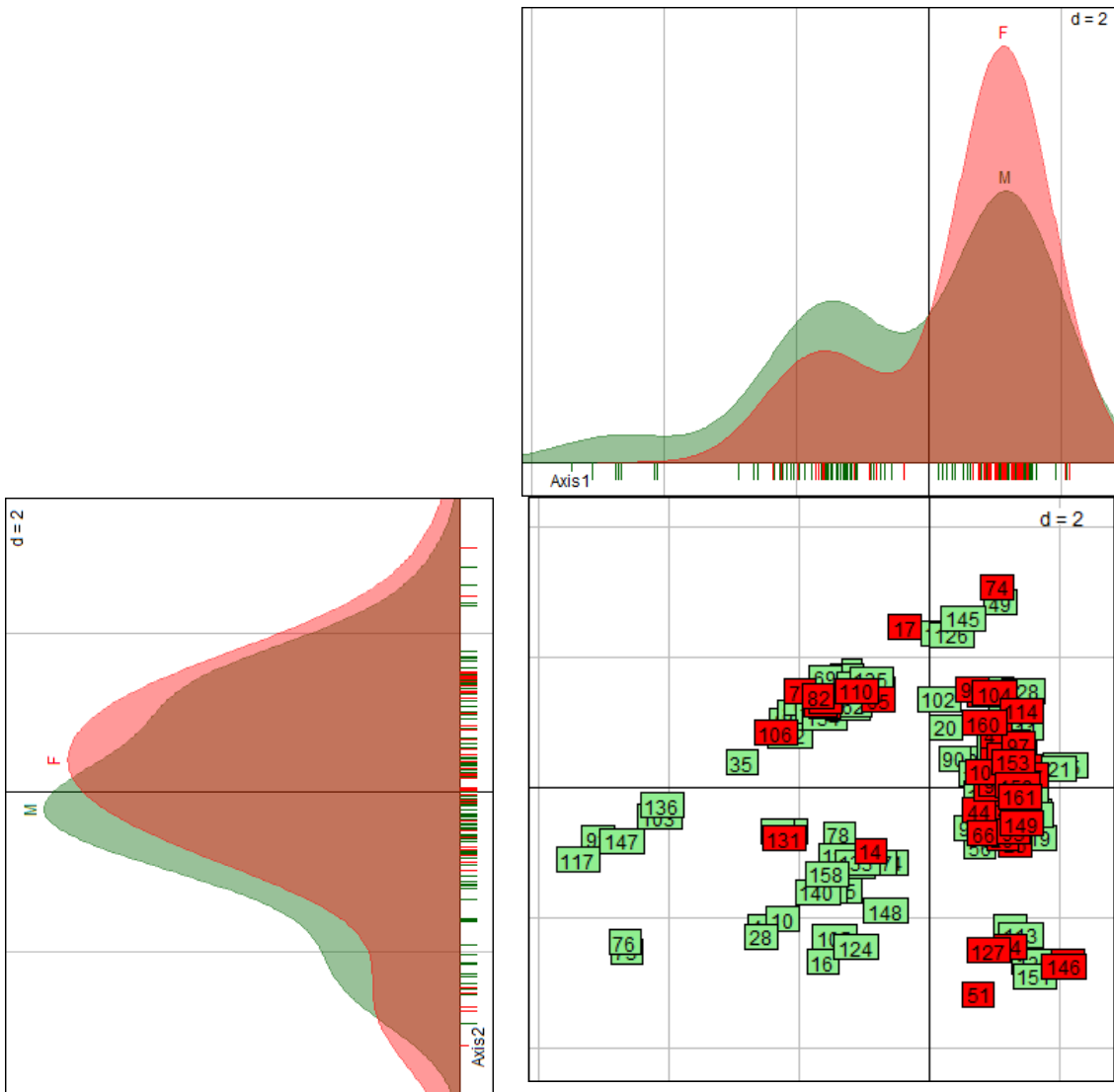
En prenant les variables une par une, on voit qu'on a une différence homme/femme dans l'âge (un groupe d'hommes plus âgés), dans le classement au début (uniquement des hommes dans les catégories 8 (groupe des hommes les plus âgés ?), surreprésentation des femmes en 6 et des hommes en 7), dans le classement fin (distribution équivalente homme/femme sauf dans les plus élevés).

1) Analyse en composante principale, centrée réduite et analyse discriminante

On effectue une ACP centrée réduite, sur AgeDeb, CEdebN, CEFinN, Vitesse, TpsChangt1.

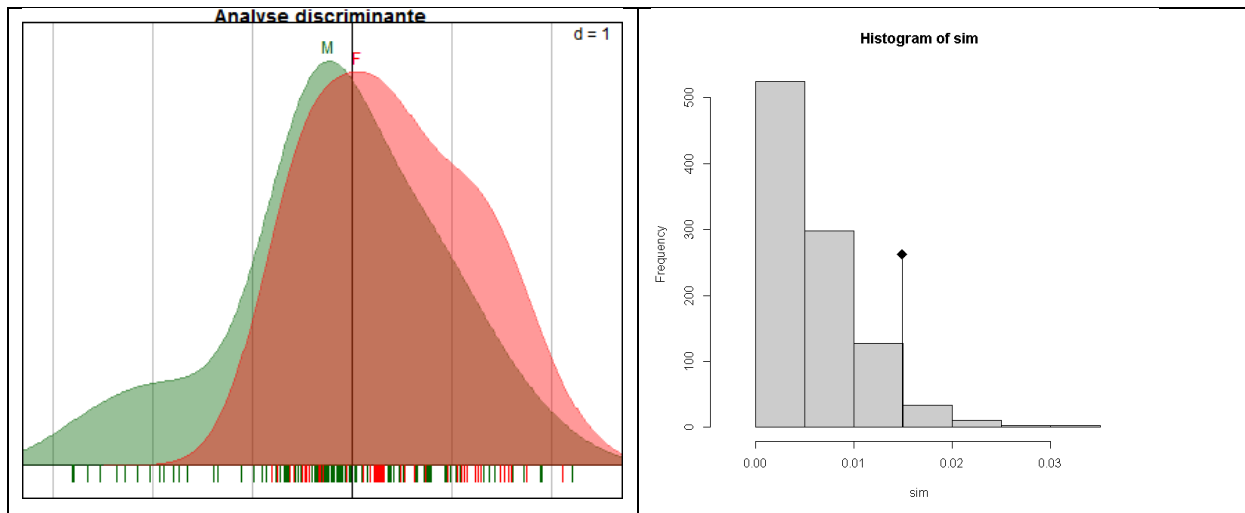


On a une structuration forte (2 axes représentent 89 % de la variance totale). Sur l'axe 1, plus on est à gauche, plus on a été recruté à une Catégorie/échelon élevée. Sur l'axe 2, plus on progresse vite et on finit à une Catégorie/échelon élevée, et moins on passe de temps dans sa première catégorie.

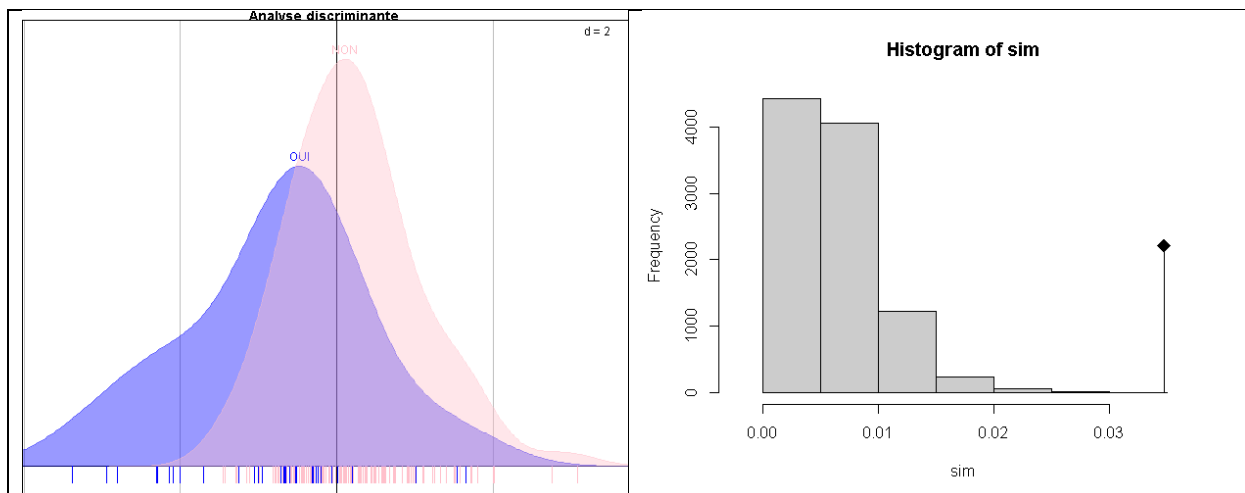


Si on s'intéresse aux hommes (vert) et aux femmes (rouge) dans cette projection, on voit qu'on a des distributions différentes, avec, sur l'axe 1, pas de femmes dans les Catégorie/échelon de recrutement les plus élevées, plus d'hommes en relatif que de femmes dans les catégories 7 que 6, et, sur l'axe 2, une distribution légèrement décalée des hommes par rapport aux femmes.

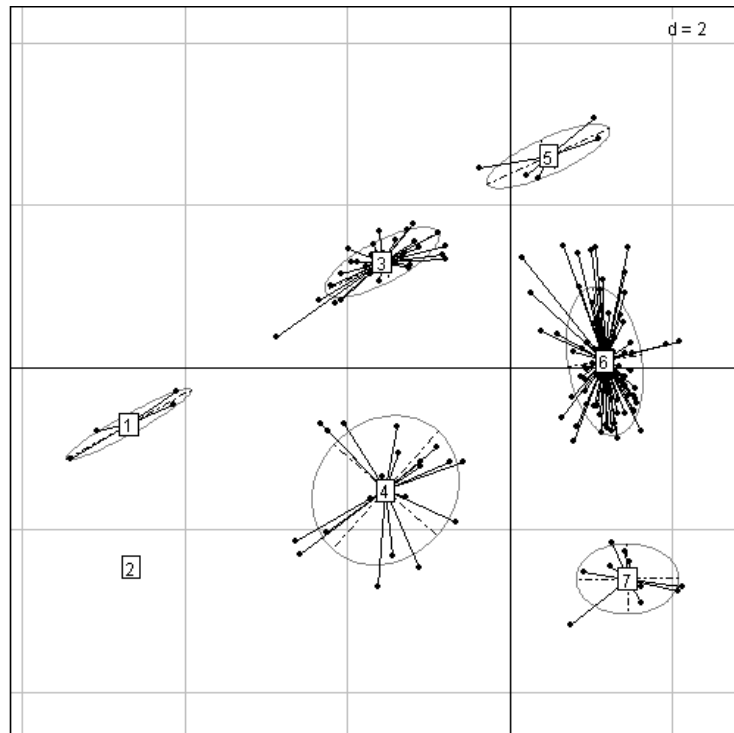
Cette différence de distribution est soutenue par une analyse discriminante linéaire ($p = 0.085$) ou par une analyse discriminante ($p = 0.08$) en limite de détection.



Pour comparaison, on obtient une différence de distribution plus claire, lorsqu'on compare les personnes qui ont eu une expérience Cirad avant leur CDI, versus celles qui n'en ont pas eu par une analyse discriminante linéaire ($p < 0.001$) ou par une analyse discriminante ($p < 0.001$). Il n'y a pas de lien entre le sexe et une pré-expérience au Cirad.

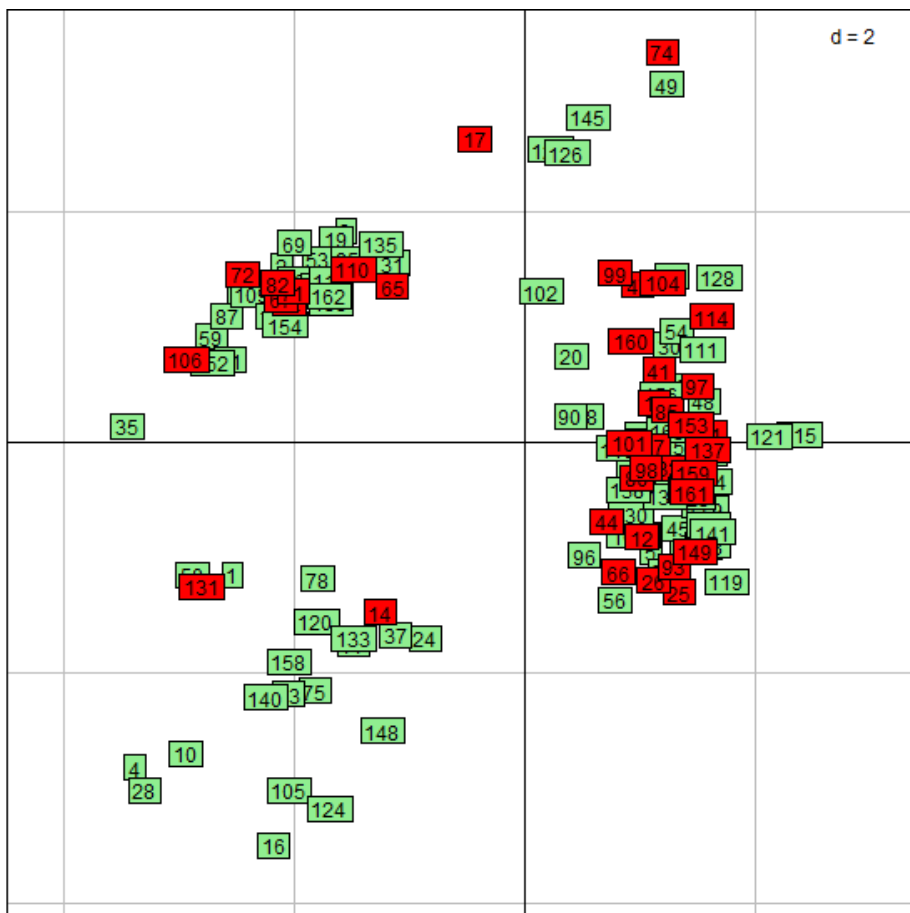
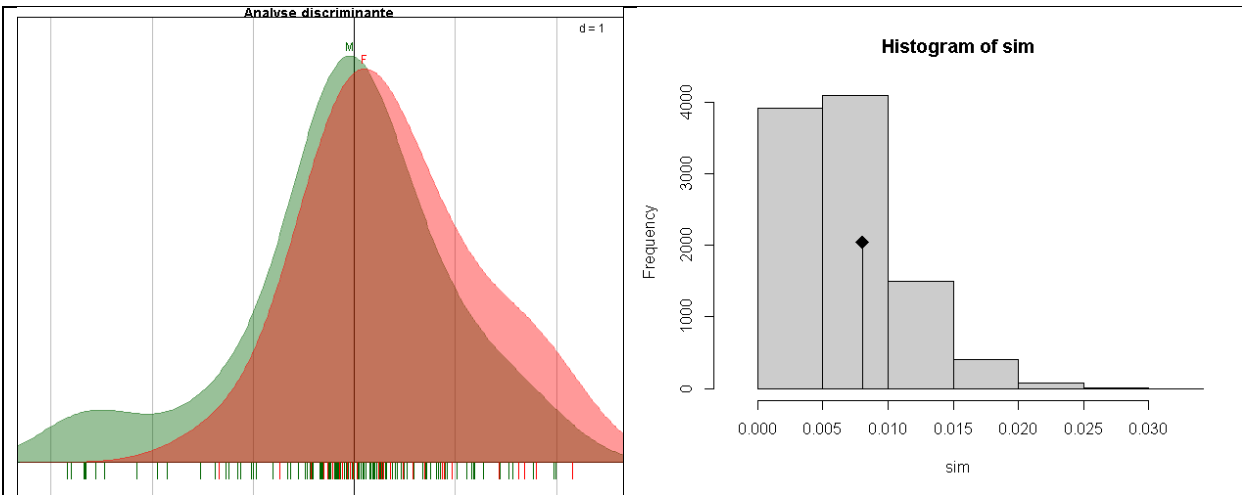


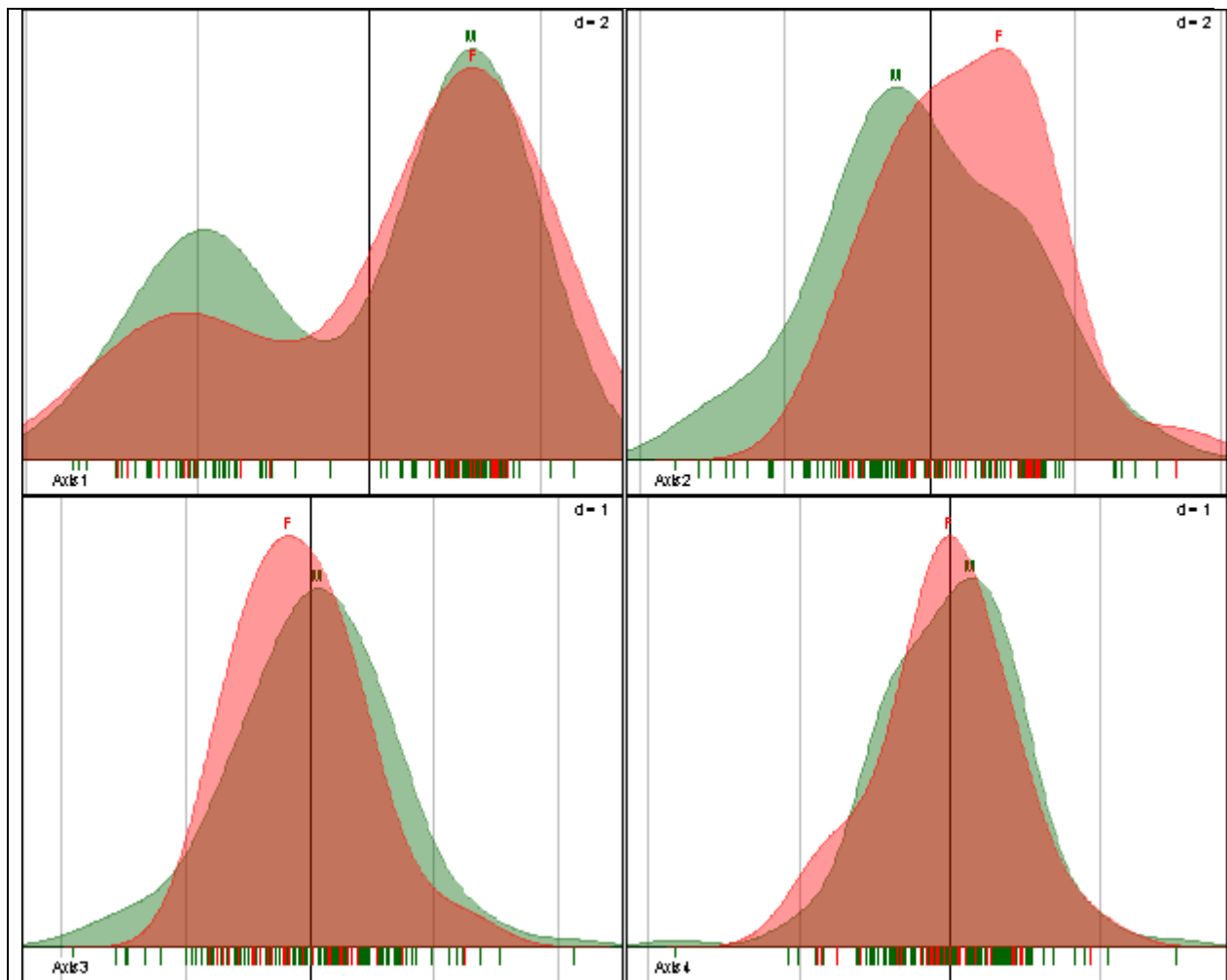
On s'intéresse maintenant aux caractéristiques des 7 groupes identifiables sur l'ACP :



- Les groupes 1 et 2 sont composés d'hommes recrutés en 8 (âge ~ 47 ans), qui ont 1 ou 0 changement de catégorie sur la période ;
- Le groupe 3 est composé d'hommes et de femmes recrutés en 7 (âge ~ 30 ans), qui n'ont pas changé de catégorie sur la période ;
- Le groupe 4 est composé d'hommes et de femmes recrutés en 7 (âge ~ 31 ans), qui ont changé de catégorie une fois sur la période. Les femmes largement sous-représentées restent, en moyenne, plus longtemps en catégorie 7 ;
- Le groupe 5 est composé d'hommes et de femmes recrutés en 6 (âge ~ 26 ans), qui n'ont pas changé de catégorie, sauf un homme, ce qui a une influence sur le temps de premier changement de catégorie ;
- Le groupe 6 est composé d'hommes et de femmes recrutés en 6 (âge ~ 25 ans), qui ont changé une fois de catégorie ;
- Le groupe 7 est composé d'hommes et de femmes recrutés en 6 (âge ~ 28 ans) et qui ont changé 2 fois de catégorie.

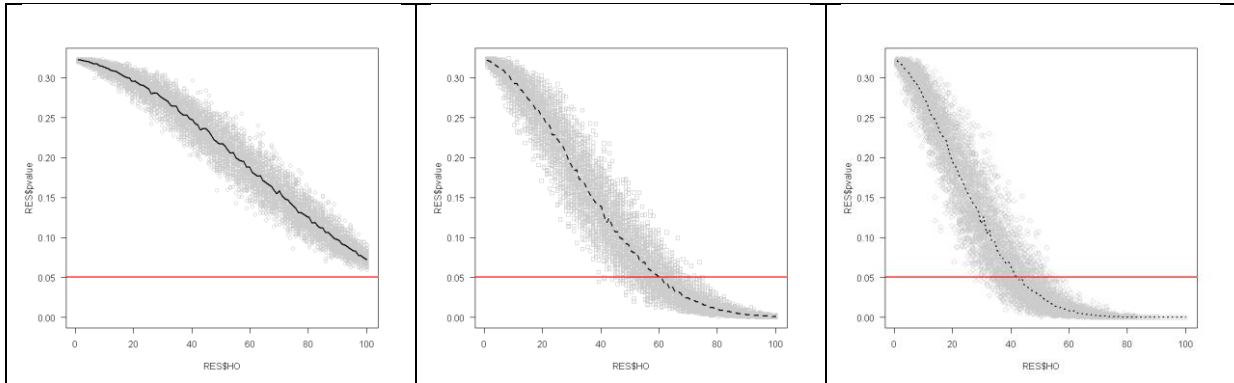
On se place dans les groupes homogènes, à savoir les groupes 3, 4, 5 et 6. Dans ce cas, les analyses discriminantes ne mettent pas en évidence de différence entre les distributions des hommes et des femmes (analyse discriminante linéaire, $p = 0.32$ et analyse discriminante, $p = 0.32$), même si il y a un léger décalage sur l'axe 2, qui compte pour 35 % de l'inertie totale et qui est lié principalement à la vitesse de progression et au temps du premier changement de catégorie.



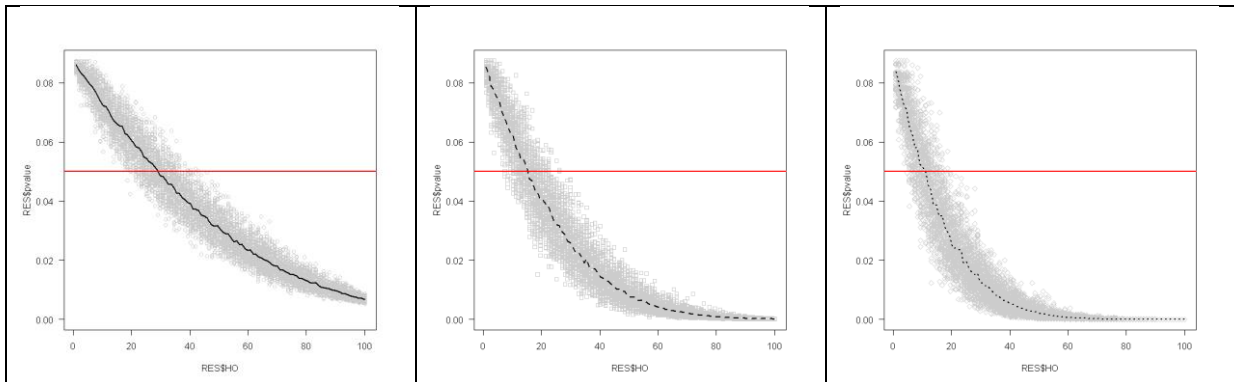


	Total	Groupe 1			Groupe 2			Groupe 3			Groupe 4			Groupe 5			Groupe 6			Groupe 7		
<i>Nombre</i>	<i>N = 163</i>	<i>N = 5</i>			<i>N = 2</i>			<i>N = 33</i>			<i>N = 21</i>			<i>N = 6</i>			<i>N = 86</i>			<i>N = 10</i>		
% F	27.0 %	0 %			0 %			24.2 %			9.5 %			33.3 %			31.4 %			50.0 %		
% ExpCirad	22.7 %	0 %			0 %			15.2 %			33.3 %			16.7 %			25.6 %			20.0 %		
Hommes		<i>Moy</i>	<i>Min</i>	<i>Max</i>	<i>Moy</i>	<i>Min</i>	<i>Max</i>	<i>Moy</i>	<i>Min</i>	<i>Max</i>	<i>Moy</i>	<i>Min</i>	<i>Max</i>	<i>Moy</i>	<i>Min</i>	<i>Max</i>	<i>Moy</i>	<i>Min</i>	<i>Max</i>	<i>Moy</i>	<i>Min</i>	<i>Max</i>
Age		46.8	39	52	47.5	47	48	36.2	30	52	39.2	30	50	31.3	27	34	30.7	24	38	30.4	29	32
CEdebN		8.29	8.07	8.40	8.17	8.13	8.20	7.12	7.07	7.27	7.28	7.07	7.53	6.30	6.20	6.33	6.30	6.13	6.47	6.32	6.27	6.40
CEfinN		8.60	8.33	8.73	9.23	9.20	9.27	7.43	7.33	7.60	8.25	8.13	8.40	6.62	6.53	6.67	7.24	7.07	7.40	8.07	8.07	8.07
Vitesse		0.030	0.027	0.033	0.091	0.090	0.091	0.031	0.024	0.040	0.093	0.080	0.110	0.032	0.028	0.035	0.091	0.063	0.115	0.171	0.155	0.188
NbreChangt		0.00	0	0	1.00	1	1	0.00	0	0	1.00	1	1	0.25	0	1	1.00	1	1	2.00	2	2
TpsChangt1		10.2	9.4	10.8	10.2	9.8	10.5	10.3	8.4	11.9	5.3	1.0	9.8	9.8	9.6	10.0	4.5	1.2	10.9	2.7	2.0	4.2
Femmes								<i>Moy</i>	<i>Min</i>	<i>Max</i>	<i>Moy</i>	<i>Min</i>	<i>Max</i>	<i>Moy</i>	<i>Min</i>	<i>Max</i>	<i>Moy</i>	<i>Min</i>	<i>Max</i>	<i>Moy</i>	<i>Min</i>	<i>Max</i>
Age								36.3	32	40	36.5	32	41	30.0	25	35	30.4	26	38	32.6	27	40
CEdebN								7.13	7.07	7.33	7.23	7.07	7.40	6.30	6.20	6.40	6.31	6.20	6.40	6.33	6.27	6.47
CEfinN								7.44	7.33	7.67	8.20	8.20	8.20	6.63	6.53	6.73	7.21	7.07	7.33	8.07	8.07	8.07
Vitesse								0.031	0.027	0.032	0.089	0.080	0.099	0.030	0.029	0.030	0.089	0.075	0.102	0.177	0.158	0.198
NbreChangt								0.00	0	0	1.00	1	1	0.00	0	0	1.00	1	1	2.00	2	2
TpsChangt1								10.3	8.5	11.8	8.3	7.5	9.0	11.2	11.0	11.4	5.1	1.0	10.6	2.2	0.0	4.0

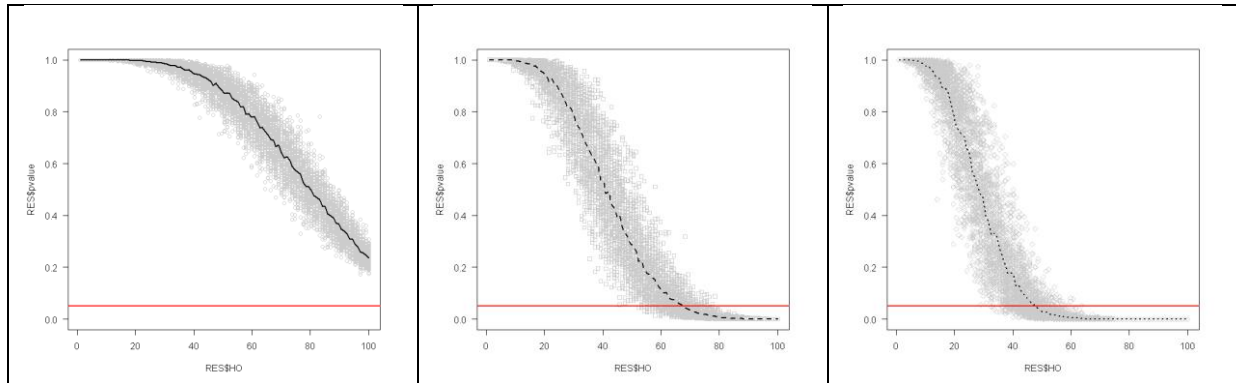
On peut aussi réaliser une analyse de sensibilité en utilisant l'analyse discriminante linéaire. En rajoutant un (gauche), deux (milieu), ou trois (droite) échelon(s) à un pourcentage d'hommes (100 réplicas par pourcentage). La sensibilité est faible, puisqu'on s'intéresse ici à différents facteurs (échelon début, fin, vitesse, temps de changement de catégorie, âge).



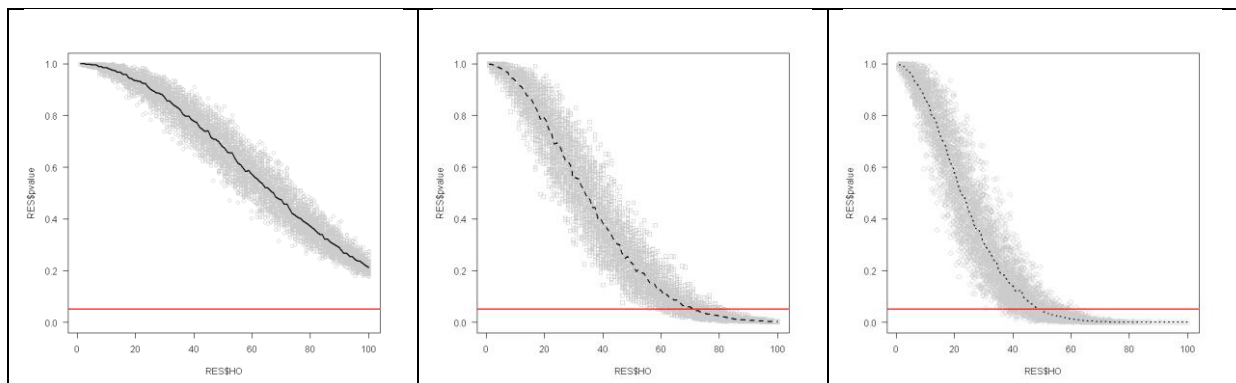
Si on réalise la même analyse uniquement sur l'échelon de fin et la vitesse, alors l'analyse est plus sensible. On détecterait 1 échelon supplémentaire pour 30 % d'hommes, 2 échelons supplémentaires pour moins de 20 % d'hommes et 3 échelons supplémentaires pour 15 % d'hommes.



On peut aussi réaliser la même analyse de sensibilité, mais en partant de 2 populations identiques (celles des hommes dupliquées) : ajout d'un (gauche), deux (milieu), ou trois (droite) échelon(s) à un pourcentage d'hommes (100 réplicas par pourcentage). La sensibilité est du même ordre qu'avec les 2 populations initiales (hommes/femmes) en partant bien sûr de 1 pour les premières *p-values* (populations identiques).



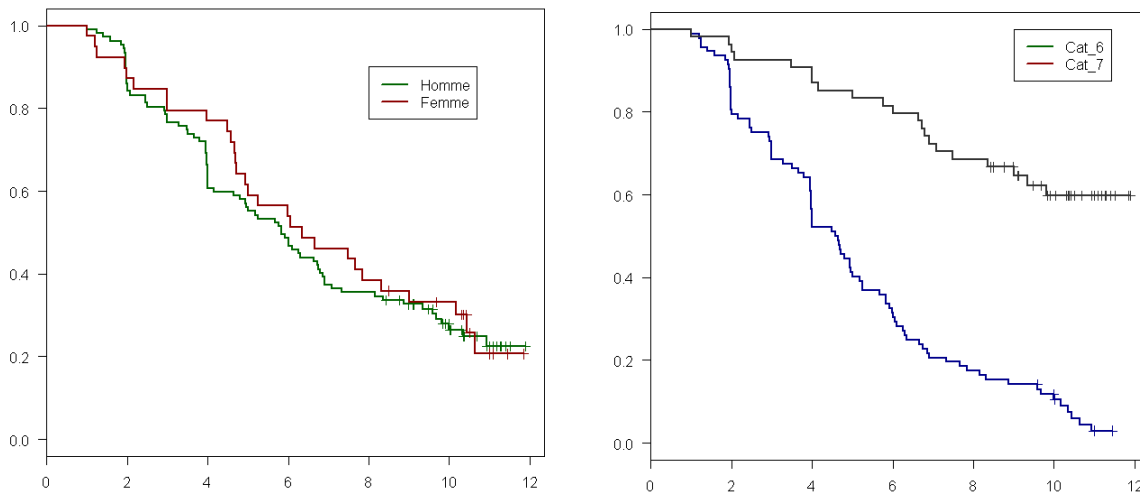
Si on réalise la même analyse uniquement sur l'échelon de fin et la vitesse, alors l'analyse n'est pas particulièrement plus sensible.



2) Analyse de survie non paramétrique

On effectue une analyse de survie non paramétrique sur TpsChangt1 et NbChangt, en ne prenant en compte que les groupes 3 à 6.

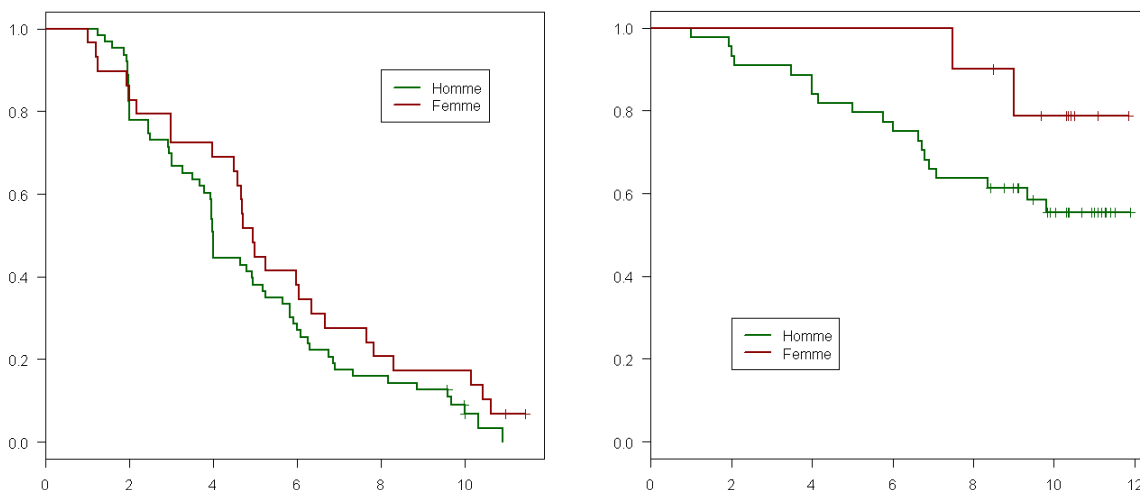
On utilise la méthode de Kaplan-Meier.



On voit que le temps de « survie » des hommes est un peu plus court que celui des femmes. Aucune différence entre hommes et femmes ($p = 0.64$ pour le test de Mantel-Haenszel et $p = 0.50$ pour le test de Gehan-Wilcoxon modifié par Peto).

En revanche, la catégorie d'origine influence fortement le temps de premier changement de catégorie ($p < 0.001$), le temps de « survie » dans la catégorie est beaucoup plus court pour les catégories 6 que 7.

On s'intéresse donc au temps de « survie » par sexe indépendamment dans les catégories d'embauche 6 (figure à gauche) et 7 (figure à droite).

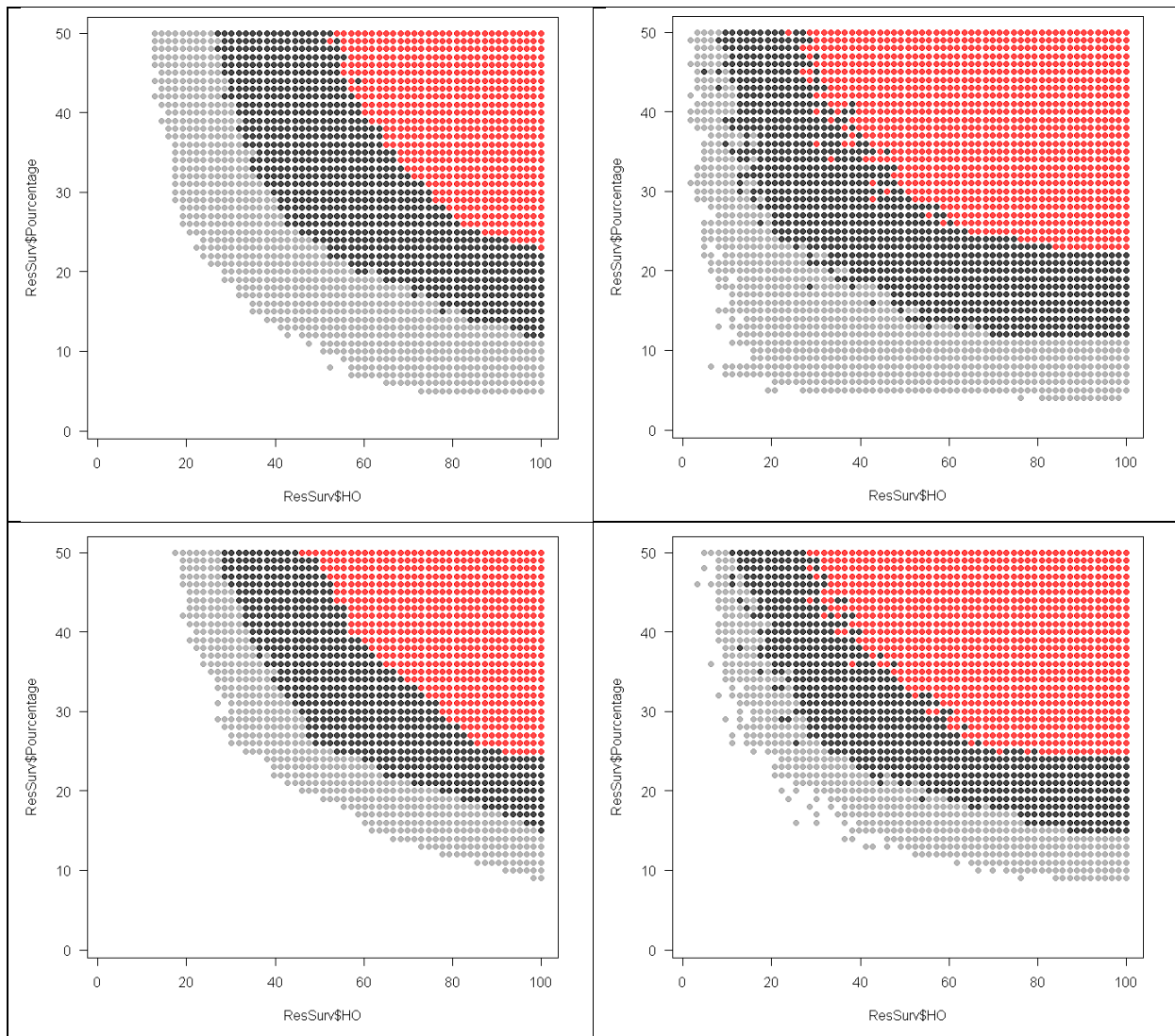


Les hommes restent un peu moins longtemps que les femmes en catégorie 6, même si la différence n'est pas « statistiquement significative » ($p = 0,16$ pour le test de Mantel-Haenszel). On voit en revanche que les hommes passent plus facilement et rapidement de 7 en 8, même si la différence n'est pas « statistiquement significative » ($p = 0,12$ pour le test de Gehan-Wilcoxon modifié par

Peto), sans doute en partie parce que une proportion importante de cette cohorte reste en catégorie 7.

On peut aussi faire une analyse de sensibilité de la méthode. On se place en catégorie 6. On a 60 hommes (sur 63) qui ont obtenu un changement de catégorie. Parmi eux, on tire au sort (100 répliques) certains hommes, pour qui on réduit le temps de premier changement d'une certaine proportion.

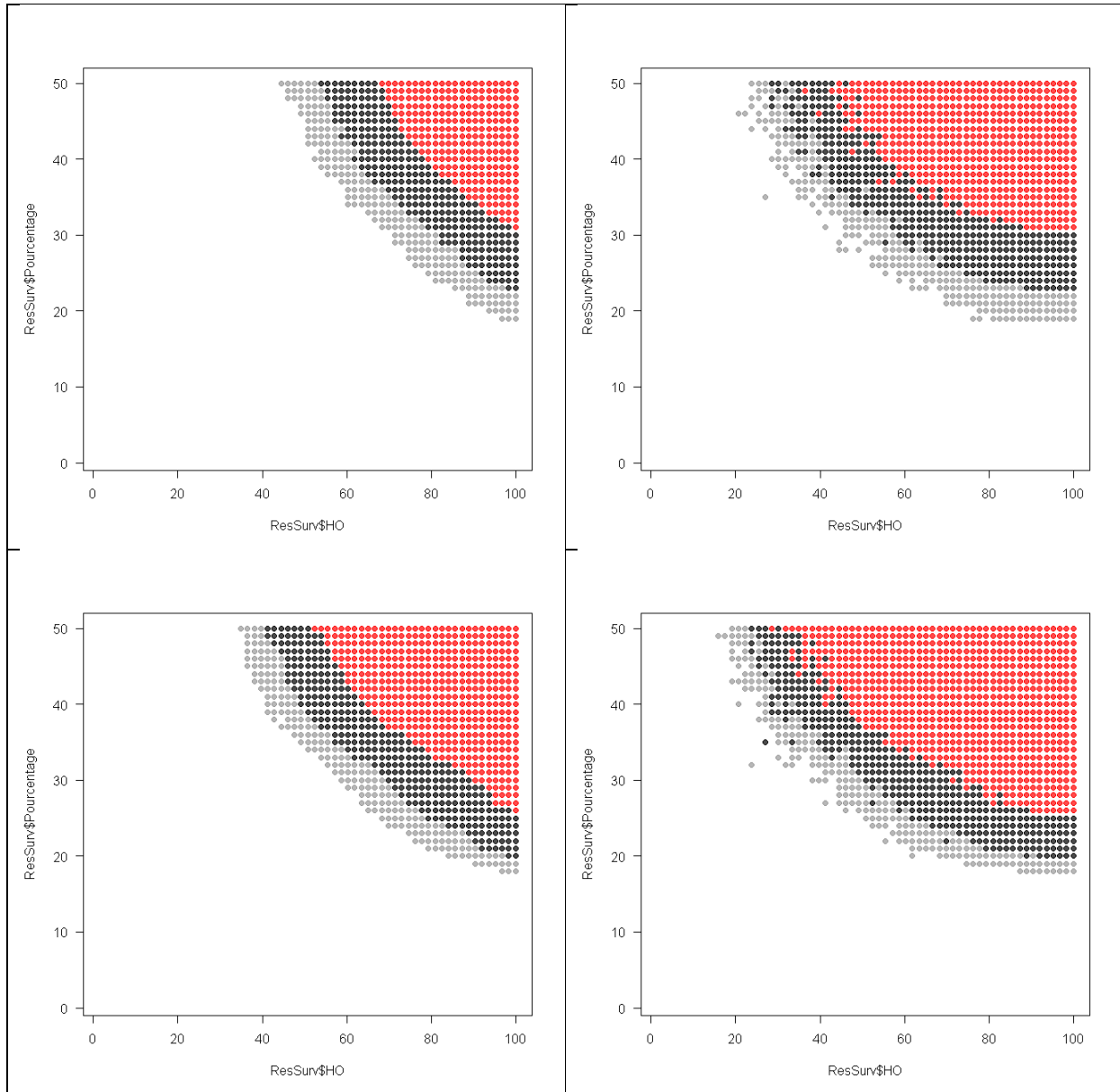
Voici une figure représentant la p-value moyenne (gauche) et minimale (droite) pour le test de Mantel-Haenszel (haut) et pour le test de Gehan-Wilcoxon modifié par Peto (bas), en fonction du pourcentage d'hommes à qui on a réduit le temps de premier changement de catégorie, et en fonction de la réduction de ce temps, avec en blanc $p > 0,10$, en gris $0,05 < p < 0,10$, en noir $0,001 < p < 0,05$ et en rouge $p < 0,01$.



On a un seuil de sensibilité, qui correspond à par exemple diviser par 3 le temps de premier changement de catégorie pour 40 % des hommes.

On peut refaire l'analyse de sensibilité de la méthode, avec deux populations identiques, celle des hommes en catégorie 6 dupliquée.

Voici la même figure que ci-dessus : p-value moyenne (gauche) et minimale (droite) pour le test de Mantel-Haenszel (haut) et pour le test de Gehan-Wilcoxon modifié par Peto (bas), avec en blanc $p > 0,10$, en gris $0,05 < p < 0,10$, en noir $0,001 < p < 0,05$ et en rouge $p < 0,01$.



On a des seuils de sensibilité beaucoup plus élevés que dans le cas d'une comparaison des populations réelles.

CONCLUSION

Les différences observées dans les données globales sont faibles entre femmes et hommes, dès lors qu'on sépare la population en fonction de la catégorie de recrutement :

	Sexe	CatDeb	AgeDeb	CEdebN	CEfinN	Vitesse	NbChangt	TpsChangt1
1	M	6	41.35294	6.302882	7.261794	0.09372824	1.0294118	4.635412
2	F	6	41.35294	6.309794	7.302029	0.09837088	1.0882353	5.038882
3	M	7	48.29545	7.186409	7.786273	0.05776091	0.4318182	8.153114
4	F	7	47.20000	7.146700	7.593400	0.04240300	0.2000000	9.915000
5	M	8	58.00000	8.257000	8.781000	0.04748000	0.2857143	10.159571

Seuls des hommes sont recrutés en catégorie 8, il s'agit des individus les plus âgés de la population.

Au sein des catégories 6, les données moyennes sont équivalentes avec un temps de premier changement de catégorie plus important pour les femmes, mais différence qui n'est pas « statistiquement significative » avec une analyse de survie.

Au sein des catégories 7, les données moyennes montrent un nombre de changement de catégorie bien plus élevé et un temps de premier changement de catégorie plus faible chez les hommes, sans que ces différences ne soient « statistiquement significatives » avec une analyse de survie.