

# Éducation et Prise de Risque

Analyse Discriminante des Variables Qualitatives

Ali Skalli

15 avril 2022

Thomas Schuler  
Nicolas Moine

Paris 2 Panthéon-Assas  
Master ISF

# Sommaire

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Présentation des données</b>	<b>3</b>
<b>3</b>	<b>Analyse des données</b>	<b>4</b>
3.1	Distribution . . . . .	4
3.2	Distribution entre la variable cible et les variables de contrôles et d'éducation . .	6
<b>4</b>	<b>Modélisation</b>	<b>7</b>
4.1	Présentation du modèle . . . . .	7
4.2	Résultat . . . . .	8
<b>5</b>	<b>Conclusion</b>	<b>10</b>

# 1 Introduction

En 1998 deux chercheurs de l’Insee sont à l’initiative d’un questionnaire devant traiter du sujet suivant : “Patrimoines et Préférences vis à vis du temps et du risque”. Cette recherche fut poursuivie à posteriori par des chercheurs de la Sofres. Elle vise à évaluer, depuis maintenant 2011, le comportement des consommateurs en temps de crise.

La Banque de France dans une étude publiée en février 2022, montre qu’entre 2020 et 2021, les ménages français ont épargné 175 milliards d’euros. La moitié de cette épargne est une conséquence de la crise sanitaire.

Afin d’étudier le comportement des consommateurs, nous nous sommes intéressés à la prise de risque des individus de façon plus globale. La prise de risque d’un individu peut être estimée à travers un certain nombre de critères socio-économiques. Il nous a donc paru important de réduire le périmètre de notre étude. Ainsi, en tant que étudiants, nous avons décidé de nous concentrer sur l’impact de nos propres études sur notre appétit au risque.

A ce stade, il est important de préciser nous abordons la prise de risque selon une dimension économique. Ce faisant, nous abordons directement des sujets comme l’épargne, l’investissement et la spéculation comme définie par l’AMF.

L’éducation et la prise de risque dans la littérature est un thème souvent abordé, notamment en Angleterre. La littérature s’est attachée à décrire le profil d’aversion au risque des étudiants dans la sélection de leur cursus d’étude. La relation de causalité inverse est pourtant plus souvent étudiée par les économistes. C’est notamment dans le papier de Seon Jung qu’on discerne l’effet significatif de l’éducation sur la prise de risque. (Does Education Affect Risk Aversion?: Evidence from the 1973 British Education Reform Seon Jung) L’auteur y affirme que la corrélation est bien négative. En d’autres termes un individu qui étudie plus est moins averse au risque. Il prend donc plus de risque.

Nous chercherons à vérifier ce résultat en France. Pour cela, nous utilisons les données mises à notre disposition et qui tirent leur origine de la base PATER.

## 2 Présentation des données

Nous avons à disposition une base de 330 variables avec 3616 observations. Dans cette base nous retenons 2 variables cibles qui ont exclusivement vocation à étudier la prise de risque que l’agent va prendre sur les marchés financiers. Cela, de deux façons distinctes : C23 : Qui questionne l’arbitrage entre placement à risque et placement non à risque. Il y a 4 réponses possibles. 2200 personnes préfèrent placer leur argent dans un lieu sûr, 880 personnes qui préfèrent prendre un peu de risque, 37 prendre un peu plus de risque et 15 prendre un risque intégrale sur leur portefeuille d’investissement à but exclusivement spéculatif. On a également un proxy de cette variable cette fois ci sur un montant donné avec la variable c25. On observe alors l’échantillon suivant : C25 : On observe dans la même tendance 2300 personnes qui veulent mettre 20 000 euros ou moins dans un actif non risqué, 950 moins de 40 000, 210 avec moins de 60 000 et le reste avec moins de 100 000. Ces variables sont mises en perspective avec le nom : “niveak” qui est en fait une variable qui transcrit l’éducation des individus sondés.

Les variables c23 et les variables c25 étant des proxys l’une de l’autre on s’autorise à construire une variable qu’on intitulera “prise de risque” et qui représente la prise de risque de l’individu sur les marchés de financiers. En parallèle, on décide également de recomposer la variable niveak en une variable qu’on intitulera “etude”. Elle a vocation à s’assurer que chaque catégorie représente un échantillon suffisamment significatif. On regroupe la variable niveak de la façon suivante : 1 : les premières années avant le brevet 2 : les bacs professionnels et les bts. 3 : le début des études supérieures 4: les doctorats et post doctorat.

Nous avons également des variables controles. Ces variables ont vocation à capturer les effets qui pourraient être cachés dans les variables explicatives. On prend par exemple le sexe. Aujourd’hui on retrouve dans la littérature un certain nombre de papiers qui mentionnent la prise de risque chez les hommes et les femmes. Des papiers questionnent d’ailleurs l’impact des crises financières dans des sociétés où plus de femme travaillent dans le monde de la finance (le sexe et la bourse par Danièle Meulders). Cette prise de risque est d’ailleurs généralisée au monde routier. De la même façon, l’âge des individus a aussi un impact sur la prise de risque financier (Vieillessement de la population et prise de risque sur les marchés financiers par Armand Derhy, Abraham Lioui).

Une autre information pertinente à contrôler est le revenu du ménage. En effet on sait que les individus à plus faibles revenus sont amenés, en montant “pur”, à s’exposer plus fortement que les hauts revenus. Ce sont des foyers qui sont à des niveaux d’endettement très élevé. Cela amène à se demander si ce ne sont pas des foyers, qui relativement à leur actif, prennent naturellement plus de risque. Ce raisonnement peut s’inverser. Quelqu’un qui a plus d’argent s’autorise à exposer au risque une somme plus conséquente de son portefeuille. D’autant plus lorsque la question c25 ne raisonne pas en part mais bien en montant d’actifs risqués(ici une enveloppe de 100 000€).

## 3 Analyse des données

### 3.1 Distribution

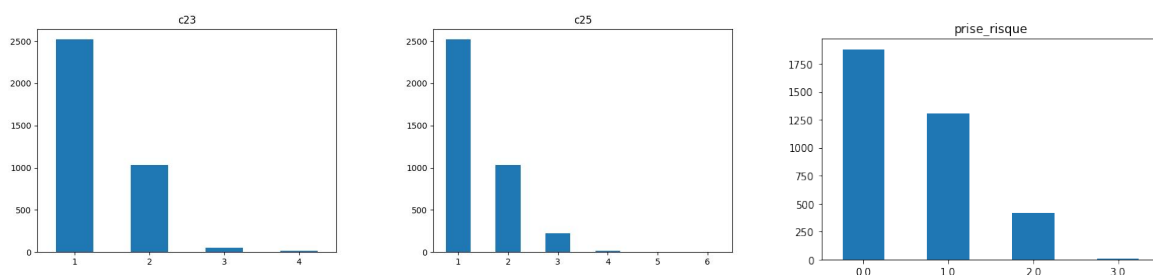


Figure 1: Distribution des Variables liées à la prise de risque

On peut voir ici les 3 variables couvrant la prise de risque. Certaines personnes n’étant pas prêtes à placer la moindre économie dans un placement risqué ont répondu qu’ils pouvaient quand même placer un peu des 100.000€ de gagner dans un placement risqué (l’inverse est aussi

vrai). On remarque cependant que la grande majorité des personnes ne sont prêt à mettre le moindre argent dans un placement risqué. Ce sont les personnes extrêmement averses au risque. A l'inverse, il n'y qu'une infime partie des répondants prêt à mettre quasiment tout leur argent dans un actif risqué, ce sont les personnes qui aiment le risque. Nous avons au milieu un bon nombre de personnes modérément averses au risque et un petit nombre de personne qui aime modérément le risque.

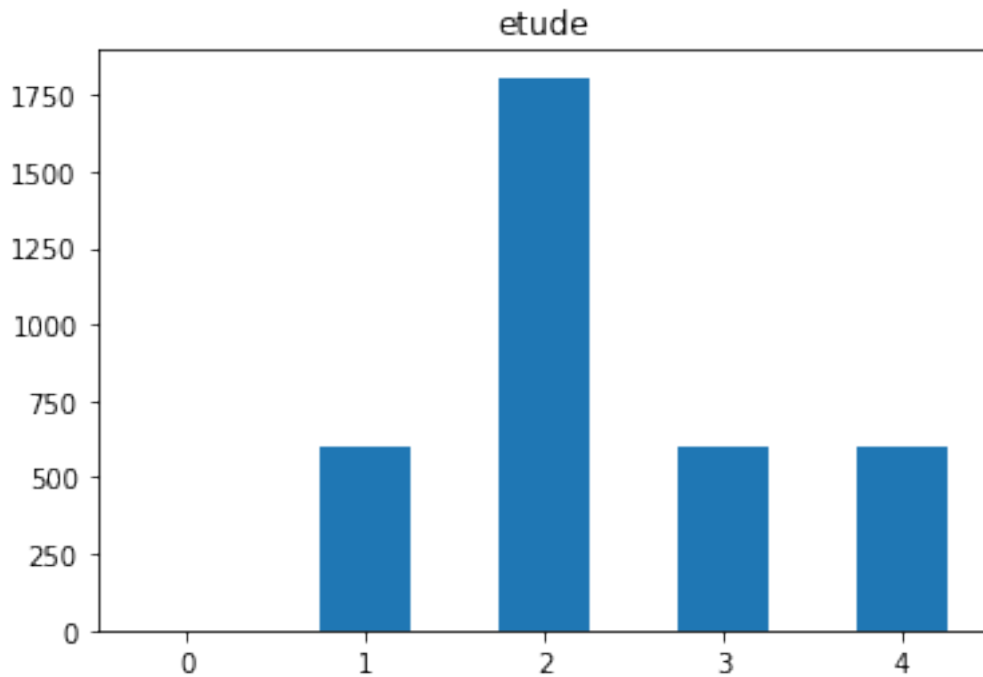


Figure 2: Distribution du niveau d'éducation

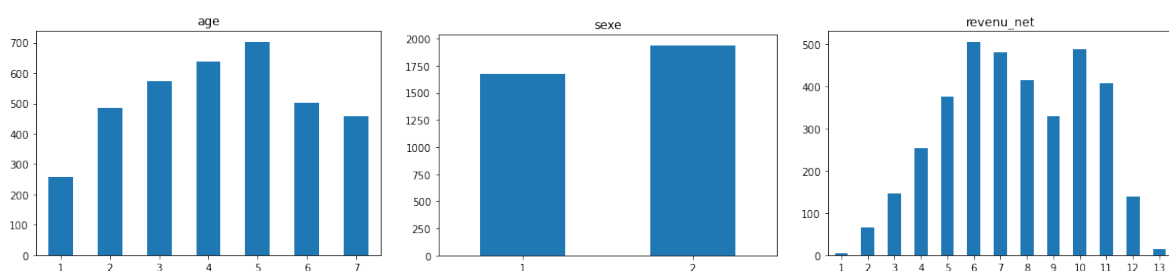


Figure 3: Distribution des Variables de contrôle

On regarde ensuite les distributions des variables de contrôle. On peut remarquer que les variables *age* et *revenu\_net* sont plutôt bien distribuées sur toute la population. La majorité des personnes sont comprises entre 25 et 64ans (2 à 5) avec aussi une grande représentation pour les retraités. Le revenu net lui se situe entre 1200 et 2700 avec une belle représentation des personnes entre 2700 et 3800. Le sexe est plutôt représentatif de la population française, avec un peu plus de femmes que d'hommes.

### 3.2 Distribution entre la variable cible et les variables de contrôles et d'éducation

etude	0	1	2	3	4	Total
prise de risque						
0	1	376	1020	257	226	1880
1	0	208	728	312	332	1580
2	0	17	57	35	44	153
3	0	1	2	0	0	3
Total	1	602	1807	604	602	3616

Figure 4: Education et la prise de risque

On remarque que dans les personnes ayant un niveau d'étude académique, ayant un brevet, un bac général, professionnel ou alors des formations professionnelles sont plus nombreux à avoir une prise de risque nulle. A l'inverse, pour les études supérieures, il y a plus de monde qui ont une aversion au risque faible. On remarque aussi que seuls les personnes ayant un niveau académique ont un gout total pour le risque.

age	1	2	3	4	5	6	7	Total
prise de risque								
0	121	226	263	308	370	308	284	1880
1	131	242	282	290	303	174	158	1580
2	7	17	28	39	28	19	15	153
3	0	0	0	1	2	0	0	3
Total	259	485	573	638	703	501	457	3616

Figure 5: Age et prise de risque

Il n'y a pas de grandes différences ici, à part le fait que les tranches d'âges entre 25 et 34 ont tendance à parier à peine plus sur le risque que les seniors qui eux sont en grande majorité totalement averse au risque.

sexe	1	2	Total
prise de risque			
0	776	1104	1880
1	805	775	1580
2	93	60	153
3	3	0	3
	1677	1939	3616

Figure 6: Sexe et prise de risque

On remarque ici une différence assez significative entre hommes et femmes, les hommes ayant beaucoup plus tendance à aimer le risque que les femmes.

## 4 Modélisation

Nous traitons ici de variable catégorielles ordonnées, appelé aussi séquentielles. En effet si on reprend la nature de nos variables dépendantes on note bien que plus un individu a un profil risquophile plus sa part d'actif risqué augmente et donc plus le "numéro" de sa classe est élevé allant ainsi de 1 à 6 pour la variable c23. Lorsqu'on traite de variables séquentielles comme ici on favorisera un modèle qu'on appelle modèle logit ordonné ou probit en fonction de la distribution. Il s'agit d'une décomposition du modèle logit (respectivement probit) dont on va maximiser comme dans un modèle logit classique la log vraisemblance. Dans le cadre d'un modèle logit on cherchera donc à maximiser la fonction de log vraisemblance suivante :

### 4.1 Présentation du modèle

Nous traitons ici de variable catégorielles ordonnées, appelé aussi séquentielles. En effet si on reprend la nature de nos variables dépendantes on note bien que plus un indivi a un profil risquophile plus sa part d'actif risqué augmente et donc plus le "numéro" de sa classe est élevé allant ainsi de 1 à 6 pour la variable c23.

Lorsqu'on traite de variables séquentielles comme ici on favorisera un modèle qu'on appelle modèle logit ordonné ou probit en fonction de la distribution.

Il s'agit d'une décomposition du modèle logit (respectivement probit) dont on va maximiser comme dans un modèle logit classique la log vraisemblance.

Dans le cadre d'un modèle logit on cherchera donc à maximiser la fonction de log vraisemblance suivante :

$$\log L = \sum_{i=1}^n y_i \log \phi(\beta' x_i) + \sum_{i=1}^n (1 - y_i) \log(1 - \phi(\beta' x_i))$$

On a alors la condition de première ordre tel que  $\delta \log L / \delta \beta = 0$  tel que :

$$S(\beta) = - \sum_{i=1}^n \frac{[y_i - \phi(\beta' x_i)]}{\phi(\beta' x_i)(1 - \phi(\beta' x_i))} \varphi(\beta' x_i) = 0$$

On vérifie bien que ces équations sont non linéaires en  $\beta$  et que la résolution se fait par algorithmes numériques.

On a alors une matrice d'information tel que :

$$I(\beta) = E\left(\frac{\delta \log L}{\delta \beta \delta \beta'}\right) = \sum_{i=1}^n \frac{[\varphi(\beta' x_i)]^2}{\phi(\beta' x_i)(1 - \phi(\beta' x_i))} x_i x_i'$$

Lorsqu'on détermine une estimation finale  $\beta^*$  la matrice de variance covariance asymptotique est donc  $[I(\beta^*)]^{-1}$ .

Si on était dans le cas d'un modèle logit on aurait alors la fonction de log vraisemblance suivante :

$$\log L = \beta' t + \sum_{i=1}^n \log(1 + \exp(\beta' x_i))$$

Avec l'optimisation suivante :

$$S(\beta) = - \sum_{i=1}^n \frac{\exp(\beta' x_i)}{(1 + \exp(\beta' x_i))} x_i + t = 0 \text{ qui converse la non linéarité avec le } \beta \text{ et donc toujours l'algorithme numérique}$$

On a cette fois la matrice d'information suivante :

$$I(\beta) = E\left(\frac{\delta \log L}{\delta \beta \delta \beta'}\right)$$

L'estimation de  $\beta^*$  faite, si on veut estimer la  $i$ ème observation alors on a la probabilité  $p^*_i$  suivante :

$$p^*_i = \frac{\exp(\beta'^* x_i)}{1 + \exp(\beta'^* x_i)}$$

En reprenant notre modèle ordonné ici avec notre variable en 4 valeurs on a alors :

$$P_j = F(\alpha_j - \beta'x) - F(\alpha_{j-1} - \beta'x), j=1, \dots, m-1$$

$$P_m = 1 - F(\alpha_{m-1} - \beta'x)$$

C'est donc ces informations que nous retiendrons pour notre modèle logit.

On a cette fois une log vraisemblance tel que :

$$\log L = \sum_{i=1}^j \sum_{j=1}^m Z_{ij} \log(\phi(\alpha_j - \beta'x_i) - \phi(\alpha_{j-1} - \beta'x_i))$$

On optimiser alors en maximisant cette log-vraisemblance pour obtenir le delta pour récupérer les maximums et les paramètres.

## 4.2 Résultat

Ainsi, dans ce modèle, nous utilisons les variables de sexe, de revenu net, d'âge pour que notre régression soit bien ajuster, afin de regarder comment influe l'éducation sur le comportement d'un individu face à une prise de risque potentielle. Nous allons utilisé comme expliqué un modèle Logit Ordonné.



<b>Dep. Variable:</b>	prise_risque	<b>Log-Likelihood:</b>	-2954.2
<b>Model:</b>	OrderedModel	<b>AIC:</b>	5922.
<b>Method:</b>	Maximum Likelihood	<b>BIC:</b>	5966.
<b>Date:</b>	Sat, 30 Apr 2022		
<b>Time:</b>	00:15:04		
<b>No. Observations:</b>	3616		
<b>Df Residuals:</b>	3609		
<b>Df Model:</b>	7		

	coef	std err	z	P> z	[0.025	0.975]
<b>etude</b>	0.2912	0.041	7.186	0.000	0.212	0.371
<b>revenu_net</b>	0.0508	0.014	3.589	0.000	0.023	0.079
<b>sexe</b>	-0.4101	0.067	-6.103	0.000	-0.542	-0.278
<b>age</b>	-0.0454	0.020	-2.251	0.024	-0.085	-0.006
<b>0/1</b>	0.3237	0.202	1.605	0.109	-0.072	0.719
<b>1/2</b>	1.1346	0.027	42.607	0.000	1.082	1.187
<b>2/3</b>	1.3873	0.143	9.706	0.000	1.107	1.667

Figure 7: Régression Logit Ordonné

Ainsi, on obtient un maximum de vraisemblance de -2956.8. Nous remarquons que nos coefficients sont tous significatifs à part le coefficient de passage entre la classe 1 et la classe 2. Ainsi, nous remarquons que le plus le revenu net augmente, plus nous avons de chance de monter d'une classe de prise de risque. A l'inverse, la prise de risque diminue durant quand on vieillit. En effet, plus on approche de la retraite, plus on a tendance à épargner et prendre le moins de risque possible, car nous savons que nous allons connaître une perte de salaire et donc de pouvoir d'achat. Nous pouvons aussi remarquer que le fait de passer de la catégorie homme à la catégorie femme, plus nous avons de chance d'être dans une classe en dessous. Ainsi, les hommes ont tendance à prendre plus de risque que les femmes. Enfin, nous pouvons confirmer nos analyses précédentes. L'éducation a belle et bien un effet sur la prise de risque. Nous avons plus de chance de passer d'une classe au dessus de risque quand on passe un cycle de plus dans nos études. Cela est sûrement dû à la capacité d'analyse et de réflexion sur le monde que les étudiants découvrent au fil de leurs études. Cependant, il y a un effet qui n'est pas visible ici, les personnes qui ont le plus fait d'études ne sont pourtant pas très appétant à tout miser sur le risque.

## 5 Conclusion

Pour conclure, nous pouvons dire en effet que le fait d'avoir des études plus longues et plus hautes, fait changer notre rapport au risque. Nous prenons en effet plus de risque, car nos connaissances et nos compétences de compréhension ont augmenté durant toutes ces années. Le revenu, le sexe ainsi que l'âge, contrôlent bien l'effet de l'éducation. Il pourrait être aussi intéressant de voir quels sujets de formations font augmenter, plus que d'autre ce goût pour le risque