

Projet d'analyse

La perception générale de ses propres attributs et de son efficacité personnelle en mathématiques comme prédicteurs des stratégies préférées pour la résolution de problèmes simples (non spécifiquement mathématiques) issus de la vie quotidienne.

Objectif

L'OCDE (2013, p. 185) relève que, dans le domaine des mathématiques, le sentiment d'efficacité personnelle (*self-efficacy*) est un prédicteur des notes, résolutions de problème, intérêt et choix de carrière relatifs aux mathématiques. De plus, "le degré de confiance des étudiants concernant leur capacité à résoudre des problèmes mathématiques, ainsi que la valeur qu'il accordent aux mathématiques, est un facteur important dans la prédiction ou l'explication du comportement de l'étudiant en regard des mathématiques, par exemple dans le choix des cours ou les décisions de carrière."

Pour cela l'OCDE (2013) précise que "alors que le sentiment d'efficacité personnelle (*self-efficacy*) est défini comme la conviction ou croyance en sa capacité de faire face à certaines tâches, la conception personnelle (*self-concept*) est définie dans un sens plus large comme la perception générale de ses attributs personnels basés sur une auto-évaluation constante."

Prolongeant cette idée, nous chercherons à tester si les comportements prédits ou expliqués par ces variables pouvaient être étendus aux stratégies employées pour la résolution de problèmes simples issus de situations concrètes, par exemple tels que présenté dans le questionnaire PISA 2012 (ST96, ST101, ST104). Il s'agira ainsi de tester si la stratégie préférée pour résoudre ces problèmes (en particulier le choix entre les stratégies de type "penser et raisonner" et "essayer au hasard") peut être inférée en fonction du sentiment d'auto-efficacité.

Question de recherche

La question adressée ici est : est-ce que le sentiment d'efficacité personnelle en mathématiques (capacité à résoudre des tâches précises : *self-efficacy*) et la conception personnelle en mathématiques (perception générale de ses attributs : *self-concept*) peuvent servir de prédicteurs concernant les stratégies préférées pour la résolution de problèmes simples (non spécifiquement mathématiques) rencontrés dans la vie quotidienne ?

Pour répondre à cette question, nous analyserons les données contenues dans le fichier CHE.sav (2012) fourni par l'OCDE concernant la Suisse.

Hypothèses

Les hypothèses testées sont les suivantes :

1. Il existe une relation de dépendance entre efficacité personnelle en mathématiques, conception personnelle en mathématiques d'une part et le choix de stratégie de type "penser et raisonner" ou "essayer au hasard"
2. Il y a une différence significative entre la moyenne du sentiment d'efficacité personnelle en mathématiques des étudiants qui préfèrent une stratégie de type "penser et raisonner" et ceux qui préfèrent une stratégie de type "essayer au"

hasard". Il en est de même pour la moyenne du sentiment de conception personnelle en mathématiques.

3. Les variables "efficacité personnelle en mathématiques" et "conception personnelle en mathématiques" permettent de discriminer les stratégies choisies (catégorisation).

Méthode

Constructs utilisés

Variable indépendante

Conformément aux hypothèses du rapport de l'OCDE (2013) nous considérons les variables du sentiment d'efficacité personnelle en mathématiques (**MATHEFF**) et de la conception personnelle en mathématiques (**SCMAT**). Ces indices sont construits sur la base des questions ST37 Q01-07 pour MATHEFF et ST42 Q02, 04, 06, 07, 09 pour SCMAT (source OCDE (2012)).

Variable dépendante

Le questionnaire PISA 2012 présente trois questions (ST96, ST101 et ST104) concernant respectivement un problème d'envoi de SMS, d'itinéraire dans une ville ou de distributeur de billet inconnu. Pour chacune 4 stratégies de résolution sont présentées : une de type "penser et raisonner", une de type "essayer au hasard", une de type "lire le manuel" et enfin une de type "demander à une connaissance". Chaque question propose au répondant de noter chaque stratégie sur une échelle de Lickert à 4 degrés (inducteur : "Que feriez-vous ?", échelle : "je le ferais certainement", "je le ferais probablement", "je ne le ferais probablement pas" et "je ne le ferais certainement pas").

Ainsi nous construirons un indice catégoriel (**ThinkVSTry**) basé sur ces trois questions indiquant la préférence du répondant entre les stratégies de type « penser et raisonner » (ST96Q02, ST101Q02 et ST104Q01) et les stratégies de type « essayer au hasard » (ST96Q01, ST101Q05 et ST104Q04). Plus précisément, l'indice prend la valeur 1 si la moyenne des notes attribuées aux stratégies de type "penser et raisonner" est inférieure à la moyenne des notes attribuées aux stratégies de type "essayer au hasard" (montrant ainsi une préférence pour le second type de stratégies). La variable prend la valeur 2 dans le cas contraire (préférence des stratégies de type « penser et raisonner ») et la valeur 0 dans le cas où les deux moyennes sont égales. Ces valeurs (également appelées niveaux par la suite) seront également libellées respectivement « **Try** » (1), « **Think** » (2) et « **Neutral** » (3).

Type d'analyse statistique à mener

Comme il s'agit d'expliquer une variable qualitative à l'aide de deux variables quantitatives, nous devons avoir recours à un test T (pour échantillons indépendants) pour tester l'égalité des moyennes et une régression logistique multinomiale pour montrer le lien de corrélation entre les variables.

Nous voulions initialement également procéder également à une analyse discriminante, cependant le test Box de l'égalité des matrices de covariances rejeta l'hypothèse d'égalité nécessaire pour garantir le bon fonctionnement de ce type d'analyse. Nous avons donc dû renoncer à cette idée.

Résultats

Nous présentons ici nos résultats ainsi que la façon que nous avons eu de les dériver des données, en commençant par expliciter le processus de création de l'indice ThinkVSTry.

Construction des variables

Les variables considérées dans cette étude sont au nombre de trois : MATHEFF, SCMAT et TraceVSTry. Les deux premières sont disponibles par défaut dans le fichier CHE.sav (2012) de l'OCDE, il n'est donc pas besoin de les construire.

La construction de la variable TraceVSTry se déroule en deux étapes selon le code SPSS suivant :

```
COMPUTE TraceVSTry = (ST96Q02+ST101Q02+ST104Q01)/3-  
(ST96Q01+ST101Q05+ST104Q04)/3.  
RECODE TraceVSTry (0=0) (Lowest thru -0.001=1) (0.001 thru  
Highest=2).  
ADD VALUE LABELS ThinkVSTry 0 "Neutral" 1 "Try" 2 "Think".  
EXECUTE.
```

La première instruction calcule la différence entre la moyenne des résultats aux questions de type « penser et raisonner » et la moyenne des résultats aux questions de type « essayer au hasard ». La seconde attribue les valeurs 0, 1 et 2 à la variable selon que la différence des moyennes est nulle, négative ou positive. Enfin la troisième associe aux trois valeurs les libellés « Neutral », « Try » et « Think ». L'entier du code utilisé pour cette recherche est disponible en annexe.

Présentation des données

Les données issues des variables se présentent comme suit :

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type	Skewness	
	Statistiques	Statistiques	Statistiques	Statistiques	Statistiques	Statistiques	Erreur std.
Mathematics Self-Efficacy	7395	-3.75	2.27	.1883	.98126	.152	.028
Mathematics Self-Concept	7385	-2.18	2.26	.1150	1.05026	-.027	.028
ThinkVSTry	7265	.00	2.00	1.0581	.58636	-.010	.029
N valide (liste)	3591						

```
Code: DESCRIPTIVES VARIABLES=MATHEFF SCMAT ThinkVSTry  
/STATISTICS=MEAN STDDEV MIN MAX SKEWNESS.
```

On peut remarquer ici que le nombre des données recueillies est très important (3'591) permettant un traitement statistique performant. De plus les données générales ne semblent pas présenter d'asymétrie (skewness faible) et présentent également un écart-type relativement homogène entre les deux variables indépendantes MATHEFF et SCMAT (entre 0.98 et 1.05).

Le tableau des corrélations ci-dessous montre que notre hypothèse 1 est justifiée, il y a bien un lien de corrélation entre MATHEFF, SCMAT d'une part et ThinkVSTry d'autre part ($p < 0.001$).

Corrélations

		Mathematics Self-Concept	Mathematics Self-Efficacy	ThinkVSTry
Mathematics Self-Concept	Corrélation de Pearson	1	.477**	-.058**
	Sig. (bilatérale)		.000	.001
	N	7385	3691	3596
Mathematics Self-Efficacy	Corrélation de Pearson	.477**	1	-.071**
	Sig. (bilatérale)	.000		.000
	N	3691	7395	7213
ThinkVSTry	Corrélation de Pearson	-.058**	-.071**	1
	Sig. (bilatérale)	.001	.000	
	N	3596	7213	7265

** . La corrélation est significative au niveau 0.01 (bilatéral).

De plus, si l'on répartit les données selon les catégories de ThinkVSTry nous obtenons la répartition suivante :

Rapport

ThinkVSTry		Mathematics Self-Concept	Mathematics Self-Efficacy
Neutral	Moyenne	.0531	.0436
	N	555	1042
	Ecart type	.95626	1.00497
Try	Moyenne	.1424	.3200
	N	2279	4711
	Ecart type	1.01730	.95893
Think	Moyenne	-.1063	-.1160
	N	762	1460
	Ecart type	1.01460	.94797
Total	Moyenne	.0759	.1918
	N	3596	7213
	Ecart type	1.01219	.98041

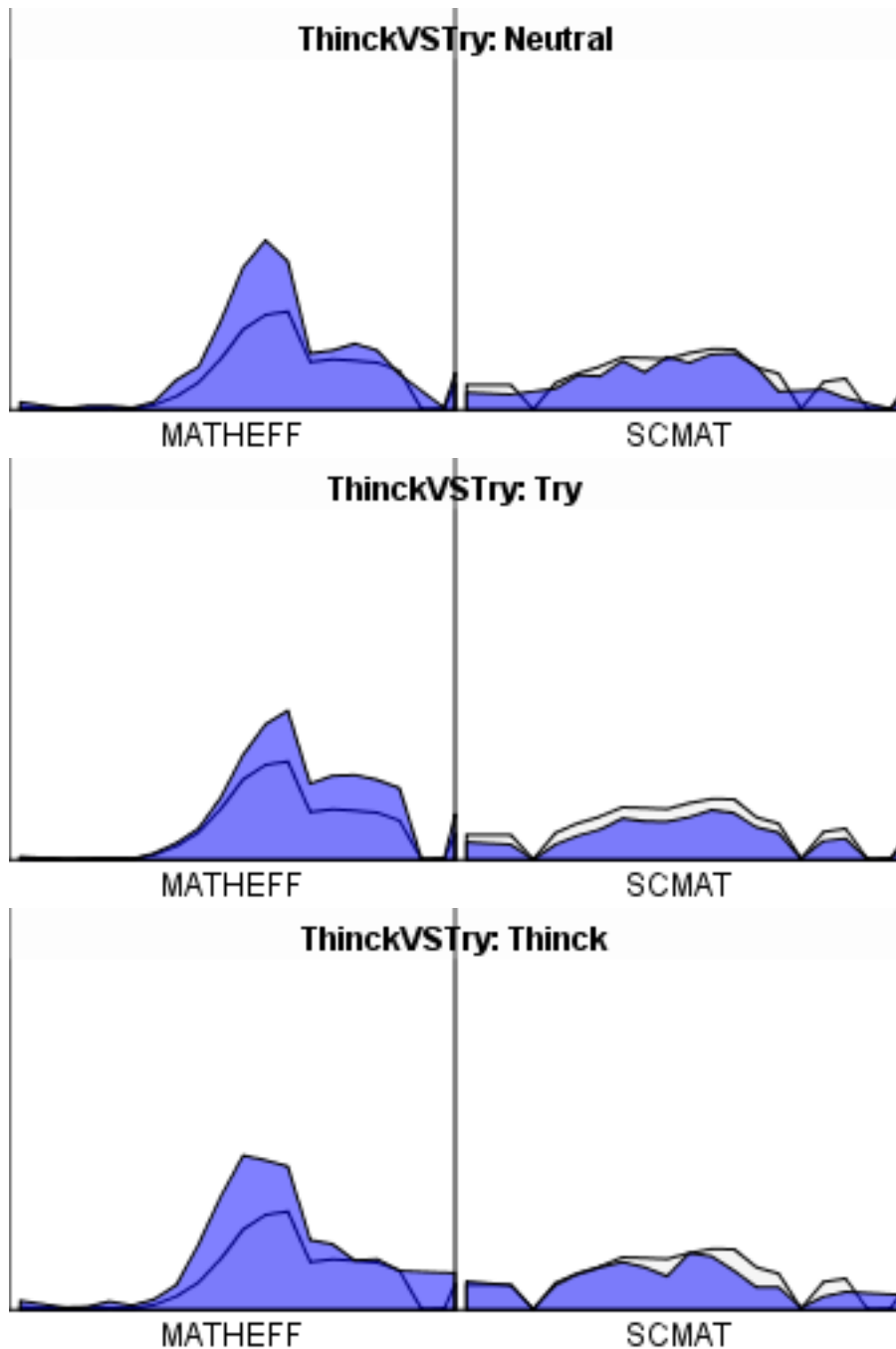
Code: MEANS SCMAT MATHEFF by ThinkVSTry.

On peut alors remarquer qu'il semble y avoir une différence entre les groupes « Think » « Try » et « Neutral ». En effet les moyennes du groupe « Think » se situent bien en dessous de la moyenne générale, celle du groupe « Try » bien en dessus alors que les moyenne du groupe « Neutral » se situent globalement proche de la moyenne générale. De plus, l'écart-type inter-groupe est très homogène (entre 0.947 et 1.0

Graphique des données brutes

Enfin, si l'on trace les graphiques des deux variables indépendantes selon les catégories de la variable dépendante on obtient les figures suivantes.

Cependant, s'il apparaît une certaine variation dans le profil de la variable MATHEFF, il semble plus difficile de percevoir une variation significative pour la variable SCMAT, laissant à penser que la relation entre SCMAT et ThinkVSTry pourrait être faible ou même inexistante.



```
Code : STATS SUBGROUP PLOTS SUBGROUP=ThinkVSTry VARIABLES=MATHEFF  
SCMAT
```

Test t d'égalité entre les moyennes

Nous avons effectué un test t d'égalité des moyennes pour des échantillons indépendants. Puisque chaque donnée est issue d'un répondant différent, l'hypothèse d'indépendance est justifiée. Nous avons ainsi effectué trois tests pour chaque variable MATHEFF et SCMAT selon les trois différents couples de facteurs possibles de ThinkVSTry. Les résultats sont les suivants.

```
Code: T-TEST GROUPS=ThinkVSTry(0 1)
/MISSING=ANALYSIS
/VARIABLES=SCMAT MATHEFF
/CRITERIA=CI(.95).
```

Statistiques de groupe

	ThinkVSTry	N	Moyenne	Ecart type	Moyenne erreur standard
Mathematics Self-Concept	Neutral	555	.0531	.95626	.04059
	Try	2279	.1424	1.01730	.02131
Mathematics Self-Efficacy	Neutral	1042	.0436	1.00497	.03113
	Try	4711	.3200	.95893	.01397

Test des échantillons indépendants

		Test t pour égalité des moyennes		
		t	ddl	Sig. (bilatéral)
Mathematics Self-Concept	Hypothèse de variances égales	-1.876	2832	.061
	Hypothèse de variances inégales	-1.948	885.110	.052
Mathematics Self-Efficacy	Hypothèse de variances égales	-8.347	5751	.000
	Hypothèse de variances inégales	-8.101	1489.150	.000

```
Code: T-TEST GROUPS=ThinkVSTry(0 2)
/MISSING=ANALYSIS
/VARIABLES=SCMAT MATHEFF
/CRITERIA=CI(.95).
```

Statistiques de groupe

	ThinkVSTry	N	Moyenne	Ecart type	Moyenne erreur standard
Mathematics Self-Concept	Neutral	555	.0531	.95626	.04059
	Think	762	-.1063	1.01460	.03676
Mathematics Self-Efficacy	Neutral	1042	.0436	1.00497	.03113

Think	1460	-.1160	.94797	.02481
-------	------	--------	--------	--------

Test des échantillons indépendants

		Test t pour égalité des moyennes		
		t	ddl	Sig. (bilatéral)
Mathematics Self-Concept	Hypothèse de variances égales	2.884	1315	.004
	Hypothèse de variances inégales	2.911	1231.978	.004
Mathematics Self-Efficacy	Hypothèse de variances égales	4.048	2500	.000
	Hypothèse de variances inégales	4.009	2161.127	.000

```
T-TEST GROUPS=ThinkVSTry(1 2)
/MISSING=ANALYSIS
/VARIABLES=SCMAT MATHEFF
/CRITERIA=CI(.95).
```

Statistiques de groupe

	ThinkVSTry	N	Moyenne	Ecart type	Moyenne erreur standard
Mathematics Self-Concept	Try	2279	.1424	1.01730	.02131
	Think	762	-.1063	1.01460	.03676
Mathematics Self-Efficacy	Try	4711	.3200	.95893	.01397
	Think	1460	-.1160	.94797	.02481

Test des échantillons indépendants

		Test t pour égalité des moyennes		
		t	ddl	Sig. (bilatéral)
Mathematics Self-Concept	Hypothèse de variances égales	5.846	3039	.000
	Hypothèse de variances inégales	5.854	1309.177	.000
Mathematics Self-Efficacy	Hypothèse de variances égales	15.222	6169	.000
	Hypothèse de variances inégales	15.314	2454.609	.000

Ceci confirme notre hypothèse selon laquelle la différence des moyennes concernant la variable MATHEFF est significative pour tous les niveaux de ThinkVSTry ($p < 0.01$). Cependant, il n'en n'est pas de même pour la variable SCMAT dont la différence est marginale entre les niveaux « Neutral » et « Think » ($p = 0.04$) et n'est pas significative entre les niveaux « Neutral » et « Try ».

Bien qu'avoir des moyennes différentes selon les niveaux de la variable expliquée ne soit pas une hypothèse nécessaire pour la régression logistique ou l'analyse discriminante, il se pourrait que ce manque de distinction entre les niveaux « Neutral » et « Try » diminue l'impact de la variable SCMAT sur ThinkVSTry.

Régression logistique multinomiale

La régression multinomiale permet d'effectuer l'équivalent d'une régression logistique pour une variable dépendante ayant plus de 2 niveaux. Il s'agit donc d'évaluer le vecteur de coefficients β_j qui gouvernent les probabilités qu'une observation i appartienne à la classe j :

$$P(y_i = j) = \frac{\exp(x_i \cdot \beta_j)}{\sum_j \exp(x_i \cdot \beta_j)}$$

Régression multinominale

La commande

```
NOMREG ThinkVSTry (BASE=LAST ORDER=ASCENDING) WITH MATHEFF SCMAT
  /CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20)
LCONVERGE(0) PCONVERGE(0.000001) SINGULAR(0.00000001)
 /MODEL
  /STEPWISE=PIN(.05) POUT(0.1) MINEFFECT(0) RULE(SINGLE)
ENTRYMETHOD(LR) REMOVALMETHOD(LR)
 /INTERCEPT=INCLUDE
 /PRINT=CLASSTABLE FIT PARAMETER SUMMARY LRT CPS STEP MFI IC.
```

donne les résultats suivants.

Récapitulatif de traitement des observations

		N	Pourcentage marginal
ThinkVSTry	Neutral	554	15.4%
	Try	2278	63.4%
	Think	759	21.1%
Valide		3591	100.0%
Manquant		7638	
Total		11229	
Sous-population		476 ^a	

a. La variable dépendante possède une seule valeur observée dans les sous-populations 229 (48.1%).

Informations sur l'ajustement du modèle

Modèle	Critères d'ajustement du modèle			Tests du rapport de vraisemblance		
	AIC	BIC	Log de vraisemblance - 2	Khi-deux	ddl	Sig.
Constante uniquement	2245.440	2257.813	2241.440			
Final	2127.970	2165.087	2115.970	125.470	4	.000

Qualité d'ajustement

	Khi-deux	ddl	Sig.
Pearson	1106.504	946	.000
Déviante	1114.048	946	.000

Le test est effectué sur un total de 3'591 données et nous permet de conclure que le modèle général est significatif ($p < 0.001$ dans le test du Khi-deux au deuxième et troisième tableau).

Tests du rapport de vraisemblance

Effet	Critères d'ajustement du modèle			Tests du rapport de vraisemblance		
	AIC de modèle réduit	BIC de modèle réduit	Log de vraisemblance - 2 du modèle réduit	Khi-deux	ddl	Sig.
Constante	3331.503	3356.248	3323.503	1207.533	2	.000
MATHEFF	2214.196	2238.941	2206.196	90.226	2	.000
SCMAT	2125.393	2150.138	2117.393	1.423	2	.491

La statistique du khi-deux représente la différence entre le modèle final et un modèle réduit dans les logs de vraisemblance -2. Le modèle réduit est formé en omettant un effet du modèle final. L'hypothèse nulle stipule que tous les paramètres de cet effet sont égaux à 0.

Estimations des paramètres

ThinkVSTry ^a		B	Erreur standard	Wald	ddl	Sig.	Exp(B)
Neutral	Constante	-.307	.056	29.868	1	.000	
	MATHEFF	.197	.067	8.723	1	.003	1.218
	SCMAT	.074	.064	1.365	1	.243	1.077
Try	Constante	1.059	.043	613.344	1	.000	
	MATHEFF	.458	.052	78.170	1	.000	1.580
	SCMAT	.042	.048	.740	1	.390	1.043

a. La catégorie de référence est : Think.

Le test du rapport de vraisemblance ci-dessus montre que la relation entre les variables indépendantes et la variable dépendante est significative pour MATHEFF ($p < 0.001$). Comme l'erreur standard est faible pour toutes les variables, le test de Wald nous permet d'accepter l'hypothèse d'une relation entre MATHEFF et ThinkVSTry ($p < 0.003$).

Cependant cette hypothèse est rejetée pour SCMAT ($p = 0.493$ et $p > 0.243$ dans le tableau ci-dessus). Le test ne nous permet donc pas d'accepter l'hypothèse selon laquelle SCMAT permet de différencier entre les niveaux « Neutral », « Try » et « Think ».

Interférence possibles entre les niveaux « Neutral » et « Try »

Nous avons alors considéré la possibilité que l'absence d'une différence de moyenne significative pour SCMAT selon les niveaux « Neutral » et « Try » soit responsable du rejet de l'hypothèse lors du test du maximum de vraisemblance. Pour tester cette hypothèse, nous avons alors filtré les observations pour ne conserver que celles correspondant aux niveaux « Try » et « Think » par la commande

```
USE ALL.
COMPUTE filter_$=(ThinkVSTry =1 OR ThinkVSTry=2).
VARIABLE LABELS filter_$ 'ThinkVSTry =1 OR ThinkVSTry=2 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```

Puis nous avons procédé à une régression logistique binaire (puisque ThinkVSTry ne contient plus que 2 niveaux) pas à pas descendante (selon le critère du maximum de vraisemblance) pour vérifier si la variable SCMAT continuait d'être rejetée du modèle. Les résultats sont les suivants.

```
Code: LOGISTIC REGRESSION VARIABLES ThinkVSTry
/METHOD=BSTEP(LR) MATHEFF SCMAT
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Variables de l'équation

		B	E.S	Wald	ddl	Sig.	Exp(B)
Pas 1 ^a	MATHEFF	-.470	.053	78.572	1	.000	.625
	SCMAT	-.033	.048	.454	1	.501	.968
	Constante	-1.058	.043	611.651	1	.000	.347
Pas 2 ^a	MATHEFF	-.487	.047	107.944	1	.000	.615
	Constante	-1.057	.043	611.616	1	.000	.347

a. Introduction des variables au pas 1 : MATHEFF, SCMAT.

Variables absentes de l'équation

			Score	ddl	Sig.
Pas 2 ^a	Variables	SCMAT	.454	1	.501
	Statistiques générales		.454	1	.501

a. Elimination des variables au pas 2 : SCMAT.

La variable SCMAT est donc bel et bien retirée dans tous les cas du modèle. Nous devons donc conclure que, selon les observations à notre disposition, la variable SCMAT ne nous permet pas d'inférer l'appartenance à la catégorie « Try » ou la catégorie « Think ».

Enfin, en regardant le tableau de classification effectuée lors de la régression multinomiale (ci-dessous), il apparaît que la quasi-totalité des observations (99.6%) a été classée dans la catégorie « Try ».

Classification

Observé	Prévisions			
	Neutral	Try	Think	Pourcentage correct
Neutral	0	548	6	0.0%
Try	0	2270	8	99.6%
Think	0	743	16	2.1%
Pourcentage global	0.0%	99.2%	0.8%	63.7%

De cela, il nous semble légitime d'émettre un doute conséquent sur la véracité de l'hypothèse selon laquelle la variable MATHEFF permettrait de discriminer les niveaux de la variable ThinkVSTry. Cependant, en l'absence de critère quantitatif permettant de tester la signification de ces résultats, nous ne pouvons conclure de manière décisive.

Discussion

Dans cette recherche, nous voulions tester les hypothèses suivantes :

- 1) Il existe une relation de dépendance entre efficacité personnelle en mathématiques, conception personnelle en mathématiques d'une part et le choix de stratégie de type "penser et raisonner" ou "essayer au hasard"
- 2) Il y a une différence significative entre la moyenne du sentiment d'efficacité personnelle en mathématiques des étudiants qui préfèrent une stratégie de type "penser et raisonner" et ceux qui préfèrent une stratégie de type "essayer au hasard". Il en est de même pour la moyenne du sentiment de conception personnelle en mathématiques.
- 3) Les variables "efficacité personnelle en mathématiques" et "conception personnelle en mathématiques" permettent de discriminer les stratégies choisies (catégorisation).

Notre analyse a partiellement confirmé nos hypothèses. En effet, si l'hypothèse 1 a été acceptée pour les deux variables indépendantes, il n'en va pas de même pour les hypothèses 2 et 3.

L'hypothèse 2 (et donc *a fortiori* l'hypothèse 3) a été rejetée en ce qui concerne la variable "conception personnelle en mathématiques", qui semble ne pas permettre de distinguer entre les différentes préférences de stratégie. Enfin, il apparaît que la variable "efficacité personnelle en mathématiques" permet bien de discriminer entre les étudiants qui préfèrent une stratégie de type "penser et raisonner" et ceux qui préfèrent une stratégie de type "essayer au hasard". Cependant, la qualité de cette discrimination nous semble discutable et ainsi nous avons tendance à rejeter l'hypothèse 3 dans ce cas.

En conclusion nous pouvons dire que, si "le degré de confiance des étudiants concernant leur capacité à résoudre des problèmes mathématiques, ainsi que la valeur qu'il accordent aux mathématiques, est un facteur important dans la prédiction ou l'explication du comportement de l'étudiant en regard des mathématiques, par exemple dans le choix des

cours ou les décisions de carrière” (OCDE, 2013) il n’en va pas de même en ce qui concerne la résolution de problèmes directement issus de leur vie quotidienne. Dans ce dernier cas, il semble que la grande majorité des étudiants aient recours à une stratégie de type « essayer au hasard jusqu’à ce que cela fonctionne » et non à une stratégie basée sur la pensée et la raison.

En outre, s’il semble bien y avoir une relation entre ces variables, il ne semble pas découler de cela un comportement général en regard des trois problèmes proposés dans le questionnaire PISA. Il pourrait être possible que cela soit dû à la différence entre les trois problèmes proposés : écrire un message avec son téléphone portable, trouver un itinéraire en voiture ou savoir utiliser une machine à vendre des billets de train. Il est possible que ces tâches ne demandent pas à être résolues de la même manière ou qu’elles soient trop « simples » pour demander réellement aux étudiants un effort de réflexion, biaisant ainsi la relation entre les variables considérées dans cette recherche.

Références

Références PISA

OCDE (2012). *Codebook for PISA 2012 Main Study Student Questionnaire - MAIN DATABASE*. disponible en ligne à l’adresse suivante :
<http://tecfalms.unige.ch/moodle/mod/url/view.php?id=11557>

OCDE. (2013). *PISA 2012 Assessment and Analytical Framework Mathematics, Reading, Science, Problem Solving and Financial Literacy: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD Publishing. doi : dx.doi.org/10.1787/9789264190511 Disponible à l’adresse suivante :
<http://browse.oecdbookshop.org/oecd/pdfs/free/9813011e.pdf>

CHE.sav (2012) Fichier SPSS fourni par l’OCDE. Récupéré sur
<http://tecfa.unige.ch/guides/methodo/PISA/>

Références SPSS

IDRE (2014). *SPSS Learning Module How can I analyze a subset of my data?* [Webpage]. UCLA (Los-Angeles). Récupéré sur
http://www.ats.ucla.edu/stat/spss/modules/subset_analyze.htm

Schneider, D. (2006). *Balises de méthodologie pour la recherche en sciences sociales*. Slides du cours – TECFA – UNIGE (Genève). Récupéré sur
<http://tecfa.unige.ch/guides/methodo/IDHEAP/methodes.book.pdf>

Schwab, J. (2012a). *Multinomial Logistic Regression – Basic Relationships*. Slides du cours SW388R7. University of Texas (Austin). Récupéré sur
http://www.utexas.edu/courses/schwab/sw388r7/SolvingProblems/DiscriminantAnalysis_BasicRelationships.ppt

Schwab, J. (2012b). *Discriminant Analysis – Basic Relationships*. Slides du cours SW388R7. University of Texas (Austin). Récupéré sur
http://www.utexas.edu/courses/schwab/sw388r7/SolvingProblems/MultinomialLogisticRegression_BasicRelationships.ppt

Annexe - Code SPSS de l'analyse

```
* ## Computation des variables (ce code permet de conserver les étape de calcul).
```

```
*-----.
```

```
*Code plus direct pour le même résultat (mais moins lisible).
```

```
COMPUTE ThinkVSTry=(ST96Q02+ST101Q02+ST104Q01)/3-  
(ST96Q01+ST101Q05+ST104Q04)/3.
```

```
*Recodage de la variable en fonction des catégorie.
```

```
RECODE ThinkVSTry (0=0) (Lowest thru -0.001=1) (0.001 thru Highest=2).
```

```
ADD value labels ThinkVSTry 0 "Neutral" 1 "Try" 2 "Think".
```

```
EXECUTE.
```

```
*****.
```

```
* ## Skewness assymétrie.
```

```
* pas véritablement d'asymétrie marquée, donc les données ne sont pas déséquilibrées.
```

```
* malgré la présence d'une grande différence dans le nombre d'observations pour VARTThinkVSTry.
```

```
DESCRIPTIVES VARIABLES=MATHEFF SCMAT ThinkVSTry  
  /STATISTICS=MEAN STDDEV MIN MAX SKEWNESS.
```

```
* tableau récapitulatif des sous-groupes.
```

```
MEANS SCMAT MATHEFF by ThinkVSTry.
```

```
* ## Graphiques des sous-groupes pour voir les différences en fonction de VARTracVSTry.
```

```
STATS SUBGROUP PLOTS SUBGROUP=ThinkVSTry VARIABLES=MATHEFF SCMAT  
 /OPTIONS XSIZE=1.75 YSIZE=1.75 YSCALE=90 ALLDATACOLOR=whitesmoke  
 SUBGROUPCOLOR=blue TRANSPARENCY=50  
 ALLDATAPATTERN=solid SUBGROUPPATTERN=solid BINCOUNT=20 SMOOTHPROP=.05  
 MISSING=VARIABLEWISE HISTOGRAM=AREA.
```

```
*****.
```

```
* ## Tests T pour égalité des moyennes entre les couples de groupes.
```

```
* groupe Neutral et Try.
```

```
T-TEST GROUPS=ThinkVSTry(0 1)
```

```
  /MISSING=ANALYSIS
```

```
  /VARIABLES=SCMAT MATHEFF
```

```
  /CRITERIA=CI(.95).
```

```
* groupe Neutral et Think.
```

```
T-TEST GROUPS=ThinkVSTry(0 2)
```

```
  /MISSING=ANALYSIS
```

```
  /VARIABLES=SCMAT MATHEFF
```

```
  /CRITERIA=CI(.95).
```

```
* groupe Try et Think.
```

```
T-TEST GROUPS=ThinkVSTry(1 2)
```

```
  /MISSING=ANALYSIS
```

```
  /VARIABLES=SCMAT MATHEFF
```

```
  /CRITERIA=CI(.95).
```

```

*****.

* ## Regression nominale pour une Régression logistique avec 3 catégories.

* Générale.
* classification 43.7% (by chance =  $.3^2 + .28^2 + .42^3 = .3448$  --> 43.1% est le seuil.
NOMREG ThinkVSTry (BASE=LAST ORDER=ASCENDING) WITH MATHEFF SCMAT
  /CRITERIA CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20) LCONVERGE(0)
PCONVERGE(0.000001) SINGULAR(0.00000001)
  /MODEL
  /STEPWISE=PIN(.05) POUT(0.1) MINEFFECT(0) RULE(SINGLE) ENTRYMETHOD(LR)
REMOVALMETHOD(LR)
  /INTERCEPT=INCLUDE
  /PRINT=CLASSTABLE FIT PARAMETER SUMMARY LRT CPS STEP MFI IC.

*****.

* Regression sur seulement les 1 et 2.
USE ALL.
COMPUTE filter_$=(ThinkVSTry =1 OR ThinkVSTry=2).
VARIABLE LABELS filter_$ 'ThinkVSTry =1 OR ThinkVSTry=2 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

* Générale.
LOGISTIC REGRESSION VARIABLES ThinkVSTry
  /METHOD=BSTEP(LR) MATHEFF SCMAT
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

FILTER OFF.
USE ALL.

*****.

* ## Analyse discriminante.
*On calcule les probabilités a priori selon les effectifs car ils sont très différents.

* probabilité d'être juste par classement aléatoire =  $\sum_i \pi_i^2 = .154^2 + .634^2 + .211^2 = 0.470193$ .
* un gain de 25% par rapport à cela représente : 58.8% seuil.

* Générale.
* marche a 63.6% alors qu'en classant tout dans le groupe 1 on a déjà 63.4% de juste...
DISCRIMINANT
  /GROUPS=ThinkVSTry(0 2)
  /VARIABLES=MATHEFF
  /ANALYSIS ALL
  /PRIORS SIZE
  /STATISTICS=MEAN STDDEV UNIVF BOXM TABLE CROSSVALID
  /PLOT=COMBINED
  /CLASSIFY=NONMISSING POOLED.

```