

Fraude à l'université

Étude réalisée par les L2, L3 et M1 STAPS

sous la direction de
Christophe Genolini

Septembre 2007 - Janvier 2008

Table des matières

1	Introduction	3
2	Présentation de l'étude	3
3	Analyse univariée	3
3.1	Nettoyage des données	4
3.2	Construction d'un score global de triche	4
3.3	Détail	4
3.3.1	Quel est votre âge ?	4
3.3.2	Vous êtes : Un homme / Une femme ?	5
3.3.3	Votre Niveau d'étude : L2 / L3 / M1 / M2 ?	5
3.3.4	Votre UFR	5
3.3.5	Avez-vous déjà redoublé dans votre scolarité ?	6
3.3.6	Quelle était votre mention au bac ?	6
3.3.7	Copier en regardant la copie ou les brouillons d'un autre	6
3.3.8	Communiquer avec un voisin	7
3.3.9	Échanger des brouillons	7
3.3.10	Fabriquer ou utiliser des antisèches	7
3.3.11	Envoyer et recevoir des SMS	8
3.3.12	Avoir le cours sur les genoux	8
3.3.13	Ne pas rendre sa copie	8
3.3.14	Préparer la salle avant	9
3.3.15	Voler les sujets avant l'examen	9
3.3.16	Autres	9
3.3.17	Lors de votre bac, avez-vous utilisé l'une des techniques décrites ci-dessus ?	10
3.3.18	Quand vous étiez au lycée, avez-vous utilisé l'une des techniques décrites ci-dessus ?	10
3.3.19	Score triche total	10
3.4	Bilan de l'analyse univariée	10

4	Analyse bivariée (L3)	11
4.1	[SCORETRICHETOTAL] vs [AGE]	11
4.1.1	Graphe	11
4.1.2	Diagnostic	12
4.1.3	Test	12
4.2	[SCORETRICHETOTAL] vs [SEXE]	12
4.2.1	Graphe	13
4.2.2	Diagnostic	13
4.2.3	Test	13
4.3	[SCORETRICHETOTAL] vs [NIVEAUDETUDE]	13
4.3.1	Graphe	14
4.3.2	Diagnostic	14
4.3.3	Test	14
4.4	[SCORETRICHETOTAL] vs [UFR]	14
4.4.1	Graphe	15
4.4.2	Diagnostic	15
4.4.3	Test	15
4.5	[SCORETRICHETOTAL] vs [REDOUBLEMENT]	15
4.5.1	Graphe	16
4.5.2	Diagnostic	16
4.5.3	Test	16
4.6	[SCORETRICHETOTAL] vs [MENTIONBAC]	16
4.6.1	Graphe	17
4.6.2	Diagnostic	17
4.6.3	Test	17
4.7	[SCORETRICHETOTAL] vs [BACTRICHE]	17
4.7.1	Graphe	18
4.7.2	Diagnostic	18
4.7.3	Test	18
4.8	[SCORETRICHETOTAL] vs [LYCEETRICHE]	18
4.8.1	Graphe	19
4.8.2	Diagnostic	19
4.8.3	Test	19
4.9	Bilan	19
5	Validation de l'échelle (M1)	20
5.1	α de Cronbach	20
5.2	Unidimensionnalité	20
5.3	Analyse du cercle des corrélations	20
5.4	Comparaison avec d'autres mesures	21
6	Construction d'un modèle (M1)	21
6.1	Selection des variables	21
6.2	Comparaison des modèles	21
6.3	Détail du modèle	22
6.4	Bilan	22

7 Conclusion générale	23
7.1 Réponses aux questions initiales	23
7.2 Limite de l'étude	23

1 Introduction

Dans le cadre de l'enseignement des statistiques, les étudiants STAPS ont réalisé une enquête sur la fraude à l'université. Les L2 ont choisi le sujet, construit le questionnaire et sont allés le faire passer. Un groupe de TD a eu pour consigne de choisir des individus STAPS, l'autre de choisir des individus SEGMI et SJAP. En TD, les L2 ont mené à bien l'analyse univariée, sous Excel. Les L3 ont fait également l'analyse univariée puis ont réalisé un certain nombre de tests, sous le logiciel R. Les M1 ont vérifié la validité de l'échelle établie par les L2, puis ont construit un modèle, sous le logiciel R.

Le présent rapport est une compilation des travaux effectués par les étudiants des différentes années. IL N'A PAS LA PRETENTION D'ÊTRE UN RAPPORT SCIENTIFIQUE (voir la section 7.2 page 23 : "Limites de l'enquête"). Il constitue avant tout le résumé d'un exercice fait par et pour les étudiants, pour que l'enseignement des statistiques soit un peu moins austère...

2 Présentation de l'étude

L'étude porte sur la fraude à l'université. 61 étudiants de L2 ont participé à la construction d'un questionnaire. La moitié d'entre eux a eu pour mission de faire remplir le questionnaire par des étudiants STAPS, l'autre moitié par des étudiants issus des UFR SEGMI et SJAP.

Les objectifs de l'étude étaient de :

1. Comparer le niveau de fraude en STAPS et le niveau de fraude dans d'autres UFR
 - Hypothèse des étudiants et étudiantes : les STAPS trichent plus.
2. Comparer le niveau de fraude selon le sexe.
 - Hypothèse des étudiantes : les garçons trichent plus.
 - Hypothèse des étudiants : il n'y a pas de différence en fonction des sexes.
3. Comparer le niveau de fraude selon le niveau d'étude
 - Hypothèse des étudiants et étudiantes : plus on avance dans les études, moins on fraude.

L'enquête a été réalisée sous forme d'un questionnaire. Outre les questions classiques sur l'âge, le sexe, l'UFR, le département, le niveau d'étude, la mention au bac et le fait d'avoir déjà redoublé ou non, 15 questions ont pour objectif de mesurer la propension à la triche. 9 d'entre elles sont des réponses à l'item et sont utilisées pour construire une échelle globale de fraude (par sommation des items).

Deux (triche au bac et triche lycée) ont pour vocation de vérifier la validité de notre mesure principale.

Le questionnaire est présenté en dernière page.

3 Analyse univariée

L'analyse univariée a été réalisée par les L2, les L3 et les M1.

3.1 Nettoyage des données

- [AGE] : un étudiant n'a que 17 ans. Le consentement écrit des parents étant nécessaire pour l'inclusion de mineurs dans une enquête, le sujet est exclu.
- [NIVEAUDETUDE] : 24 sujets sont en L1. Cette enquête concerne la fraude à l'université. Au mois de novembre, les L1 n'avaient pas encore présenté d'examen. Ils sont donc exclus de l'étude. Cette variable présente également une valeur aberrante (M3). L'individu n'est pas exclu mais sa variable [NIVEAUDETUDE] est considérée comme manquante.
- [UFR] : Contrairement aux instructions, des étudiants de filières, SPSE et SSA ont été inclus dans l'étude. Comme ils sont en très petit nombre (3 au total), ils sont exclus de la suite de l'analyse.
- Les variables [FILIERE], [TRICHEAUTRELESQUELLES], [TRICHEBACLESQUELLES] et [TRICHELYCEELESQUELLES] sont inexploitable.
- Pas de problème dans les autres variables.

3.2 Construction d'un score global de triche

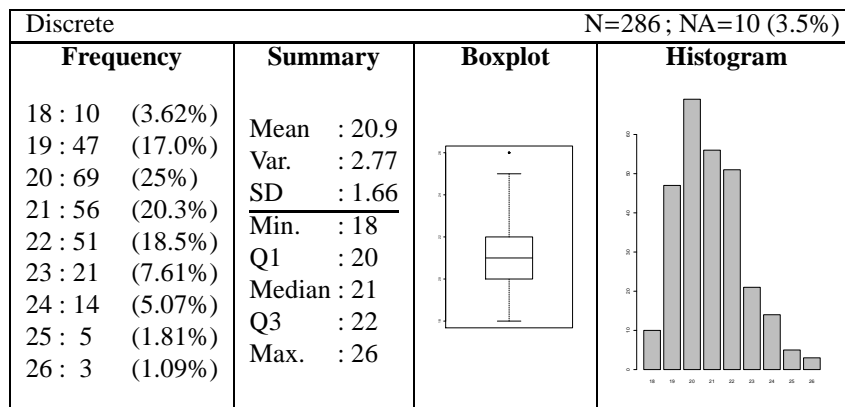
Le score de triche est obtenu par addition des différents items de triche en suivant la convention :

- 0 = *Jamais*
- 1 = *Rarement*
- 2 = *Parfois*
- 3 = *Souvent*
- 4 = *Toujours*

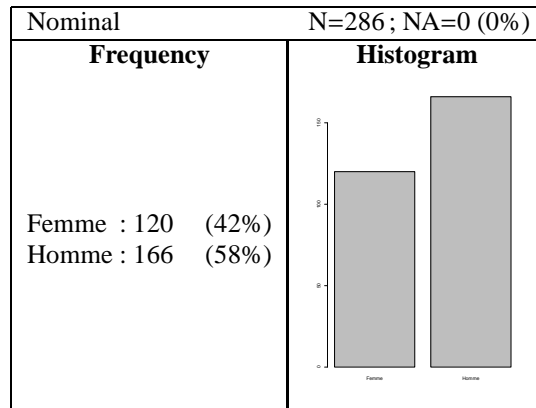
Le score de triche (appelé [SCORETRICHETOTAL]) est donc un score compris entre 0 et 36. Une validation (partielle) de ce score est proposé section 5.

3.3 Détail

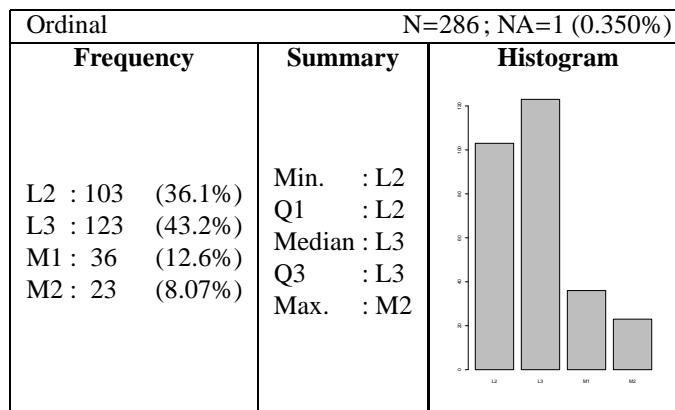
3.3.1 Quel est votre âge ?



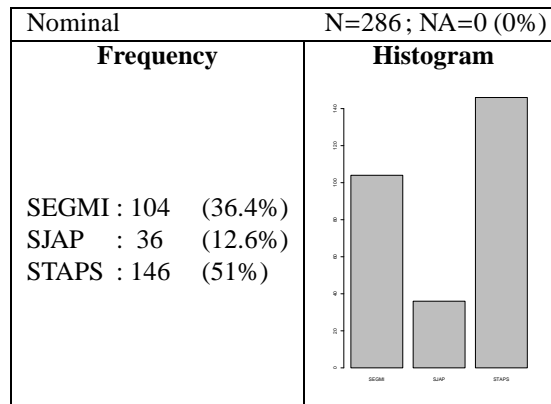
3.3.2 Vous êtes : Un homme / Une femme ?



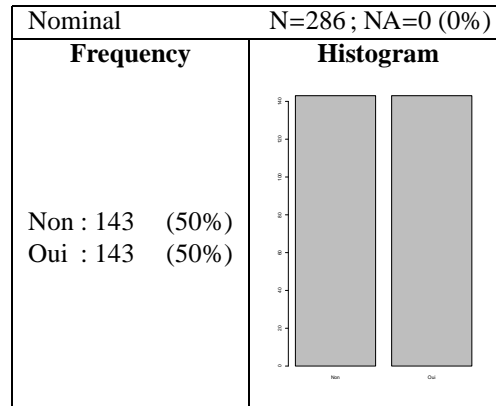
3.3.3 Votre Niveau d'étude : L2 / L3 / M1 / M2 ?



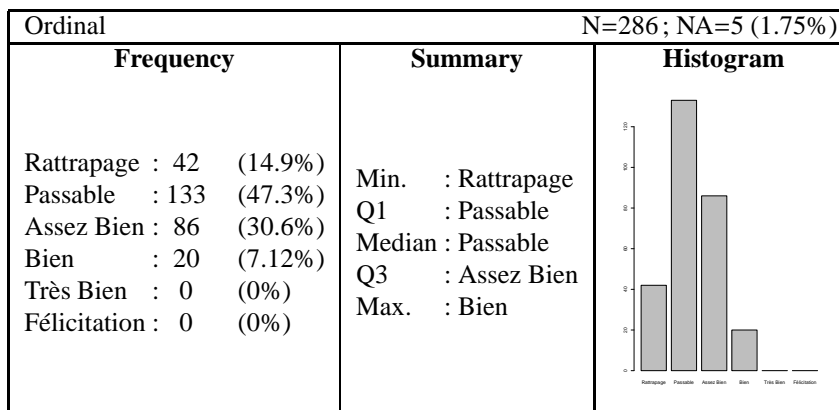
3.3.4 Votre UFR



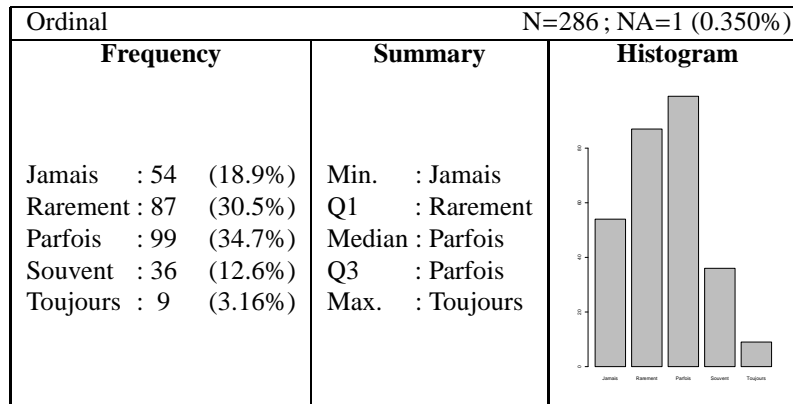
3.3.5 Avez-vous déjà redoublé dans votre scolarité ?



3.3.6 Quelle était votre mention au bac ?



3.3.7 Copier en regardant la copie ou les brouillons d'un autre



3.3.8 Communiquer avec un voisin

Ordinal		N=286 ; NA=0 (0%)	
Frequency	Summary	Histogram	
Jamais : 32 (11.2%)	Min. : Jamais		
Rarement : 81 (28.3%)	Q1 : Rarement		
Parfois : 120 (42%)	Median : Parfois		
Souvent : 40 (14%)	Q3 : Parfois		
Toujours : 13 (4.55%)	Max. : Toujours		

3.3.9 Échanger des brouillons

Ordinal		N=286 ; NA=2 (0.699%)	
Frequency	Summary	Histogram	
Jamais : 122 (43%)	Min. : Jamais		
Rarement : 93 (32.7%)	Q1 : Jamais		
Parfois : 55 (19.4%)	Median : Rarement		
Souvent : 8 (2.82%)	Q3 : Rarement		
Toujours : 6 (2.11%)	Max. : Toujours		

3.3.10 Fabriquer ou utiliser des antisèches

Ordinal		N=286 ; NA=3 (1.05%)	
Frequency	Summary	Histogram	
Jamais : 120 (42.4%)	Min. : Jamais		
Rarement : 72 (25.4%)	Q1 : Jamais		
Parfois : 56 (19.8%)	Median : Rarement		
Souvent : 26 (9.19%)	Q3 : Parfois		
Toujours : 9 (3.18%)	Max. : Toujours		

3.3.11 Envoyer et recevoir des SMS

Ordinal		N=286 ; NA=1 (0.350%)	
Frequency	Summary	Histogram	
Jamais : 245 (86%)	Min. : Jamais		
Rarement : 21 (7.37%)	Q1 : Jamais		
Parfois : 11 (3.86%)	Median : Jamais		
Souvent : 6 (2.11%)	Q3 : Jamais		
Toujours : 2 (0.702%)	Max. : Toujours		

3.3.12 Avoir le cours sur les genoux

Ordinal		N=286 ; NA=2 (0.699%)	
Frequency	Summary	Histogram	
Jamais : 219 (77.1%)	Min. : Jamais		
Rarement : 43 (15.1%)	Q1 : Jamais		
Parfois : 17 (5.99%)	Median : Jamais		
Souvent : 5 (1.76%)	Q3 : Jamais		
Toujours : 0 (0%)	Max. : Souvent		

3.3.13 Ne pas rendre sa copie

Ordinal		N=286 ; NA=1 (0.350%)	
Frequency	Summary	Histogram	
Jamais : 263 (92.3%)	Min. : Jamais		
Rarement : 19 (6.67%)	Q1 : Jamais		
Parfois : 1 (0.351%)	Median : Jamais		
Souvent : 2 (0.702%)	Q3 : Jamais		
Toujours : 0 (0%)	Max. : Souvent		

3.3.14 Préparer la salle avant

Ordinal		N=286 ; NA=1 (0.350%)	
Frequency		Summary	Histogram
Jamais	: 247 (86.7%)	Min. : Jamais	
Rarement	: 22 (7.72%)	Q1 : Jamais	
Parfois	: 13 (4.56%)	Median : Jamais	
Souvent	: 2 (0.702%)	Q3 : Jamais	
Toujours	: 1 (0.351%)	Max. : Toujours	

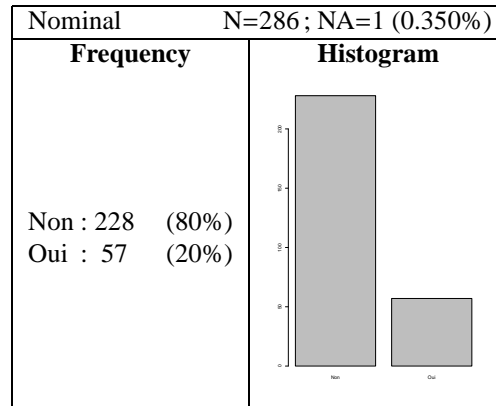
3.3.15 Voler les sujets avant l'examen

Ordinal		N=286 ; NA=0 (0%)	
Frequency		Summary	Histogram
Jamais	: 261 (91.3%)	Min. : Jamais	
Rarement	: 16 (5.59%)	Q1 : Jamais	
Parfois	: 4 (1.40%)	Median : Jamais	
Souvent	: 3 (1.05%)	Q3 : Jamais	
Toujours	: 2 (0.699%)	Max. : Toujours	

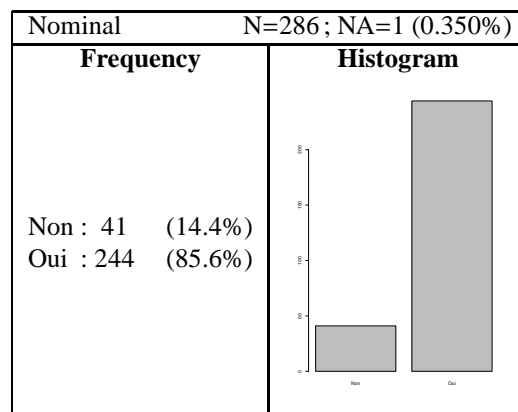
3.3.16 Autres

Ordinal		N=286 ; NA=42 (14.7%)	
Frequency		Summary	Histogram
Jamais	: 229 (93.9%)	Min. : Jamais	
Rarement	: 5 (2.05%)	Q1 : Jamais	
Parfois	: 2 (0.82%)	Median : Jamais	
Souvent	: 2 (0.82%)	Q3 : Jamais	
Toujours	: 6 (2.46%)	Max. : Toujours	

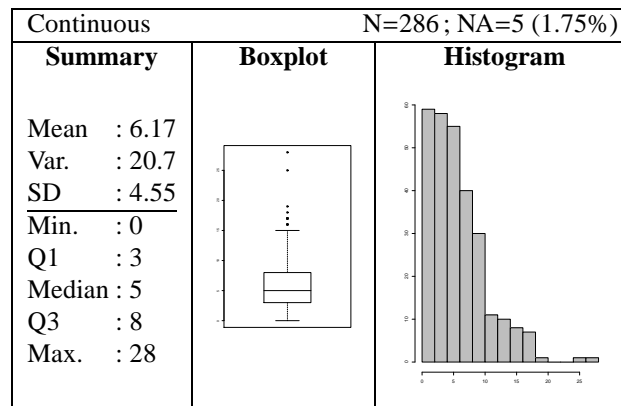
3.3.17 Lors de votre bac, avez-vous utilisé l'une des techniques décrites ci-dessus ?



3.3.18 Quand vous étiez au lycée, avez-vous utilisé l'une des techniques décrites ci-dessus ?



3.3.19 Score triche total



3.4 Bilan de l'analyse univariée

L'échantillon de notre étude est constitué de 314 étudiants.

- [AGE] : Ils ont entre 18 et 26 ans (moyenne 20.9).
- [SEXE] : On trouve plus d'hommes, 166 (58.0%) contre 120 (42.0%) femmes.
- [NIVEAUDETUDE] : 226 (79.3%) des sujets sont en premier cycle, 59 (20.7%) sont en deuxième cycle.
- [UFR] : on compte 146 (51.0%) étudiants issus de STAPS, 104 (36.4%) en provenance de SEGMI et 36 (12.6%) SJAP.
- [REDOUBLEMENT] : 143 (50.0%) des étudiants n'ont jamais redoublé, 143 (50.0%) ont déjà redoublé au moins une fois.
- [MENTIONBAC] : on trouve essentiellement des mentions *Passable* (133 soit 47.3%) puis des *Assez bien* (86 soit 30.6%). A noter, il n'y a aucune mention *Très bien* ou *Félicitation du jury*.
- Les variables de fraudes se divisent en trois groupes :
 - [COPIER] et [COMMUNIQUER] ressemblent plus ou moins à des loi normales : leur médiane est à *Parfois* (modalité centrale). Peu nombreux sont ceux qui n'ont jamais utilisé ces moyens (18.9% et 11.2%).
 - [ECHANGE BROUILLON] et [ANTISECHE] ressemblent plus à des lois de Poisson de paramètre $\lambda = 0.6$ (décroissance quasi linéaire) : un nombre important de personnes répondent *Jamais* (43.0% et 42.4%) mais les autres modalités sont quand même présentes.
 - [SMS], [COURS GENOUX], [GARDER COPIE], [PREPARER SALLE], [VOLER SUJET] et [AUTRE] sont plutôt sur un modèle de ZIP (Zero Inflated value Poisson) ou Poisson de paramètre $\lambda = 0.2$: la modalité *Jamais* regroupe entre 77.1% et 92.3% des étudiants, les autres sont très faibles.
- [BAC TRICHE] : 228 (80%) des étudiants disent ne pas avoir triché au bac.
- [LYCEE TRICHE] : 41 (14.4%) affirment ne pas avoir triché au lycée.
- [SCORE TRICHE TOTAL] : suit une loi de normale censurée à gauche. Le plancher (25% des gens ont un score de 3 ou moins) indique que l'échelle utilisée n'est probablement pas assez sensible pour les faibles score de triche. Par contre, l'absence de scores extrêmes donne une certaine validité aux résultats (personne n'a répondu "Toujours" à toutes les questions...)

Un petit groupe de personnes présente des scores très élevés. Il est possible qu'il s'agisse là de personnes n'ayant pas répondu très sérieusement. Cependant, rien ne permettant de l'affirmer, elles ne sont pas exclues de l'étude.

4 Analyse bivariée (L3)

La variable qui nous intéresse est la variable [SCORETRICHETOTAL]. D'une manière exploratoire, nous allons examiner les variables avec laquelle elle est liée.

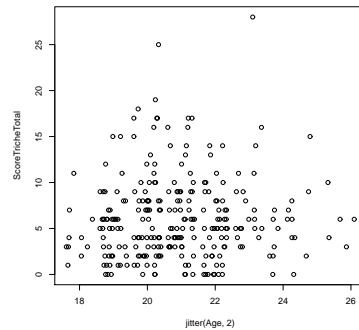
4.1 [SCORETRICHETOTAL] vs [AGE]

Avec 10 modalités différentes, la variable [AGE] peut être considérée comme continue.

- **Représentation graphique** : nuage de points.
- **Diagnostic** : une des deux variables doit suivre une loi normale.
- **Test paramétrique** : r de Pearson
- **Test non paramétrique** : r de Spearman

4.1.1 Graphe

Pour éviter la superposition des points aux coordonnées identiques, une légère dispersion (instruction jitter sous R) est ajoutée à la variable [AGE].



4.1.2 Diagnostic

La variable [AGE] suit approximativement une loi normale, on applique le r de Pearson.

4.1.3 Test

- H_0 : Il n'y a pas de différence selon l'âge, les jeunes et les moins jeunes trichent dans les mêmes proportions.
- r de Pearson : -0.044
- Degrés de liberté : $df = 269$
- p : 0.4699

P est inférieur a 5%, on accepte H_0 :

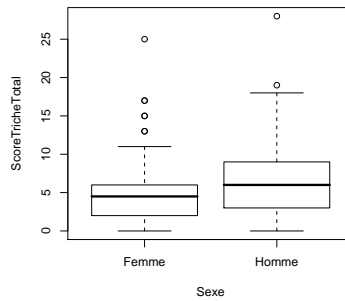
**Il n'y a pas de lien entre les variables [AGE] et [SCORETRICHETOTAL],
les jeunes et moins jeunes trichent dans les mêmes proportions.**

4.2 [SCORETRICHETOTAL] VS [SEXE]

La variable [SEXE] est une variable nominale à deux classes.

- **Représentation graphique** : boîtes à moustaches.
- **Diagnostic** : la variable continue doit être normale ou d'effectif supérieur à 30. De plus, les variances des groupes doivent être égales.
- **Test paramétrique** : T de Student
- **Test non paramétrique** : Test des rangs de Wilcoxon

4.2.1 Graphe



4.2.2 Diagnostic

- **Groupe Homme** moyenne : 6.85 ; variance 21.93
- **Groupe Femme** moyenne : 5.11 ; variance 16.63
- Les variances ne sont pas significativement différentes et les effectifs sont supérieurs à 30 : on applique le T de Student.

4.2.3 Test

- H0 : Il n'y a pas de différence entre les sexes, les hommes et les femmes trichent dans les mêmes proportions.
- test t : 3.2477
- Degrés de liberté : $df = 279$
- $p : 0.0013$

P est inférieur à 5%, on rejette H0 :

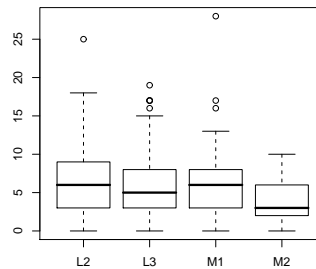
**Il y a un lien entre les variables [SEXE] et [SCORETRICHETOTAL],
les hommes trichent significativement plus que les femmes.**

4.3 [SCORETRICHETOTAL] VS [NIVEAUDETUDE]

La variable [NIVEAUDETUDE] est une nominale ordonnée. Nous choisissons de la traiter comme une nominale.

- **Représentation graphique** : boîtes à moustaches juxtaposées.
- **Diagnostic** : Les variances intra doivent être identiques, la variable continue doit être normale et les effectifs entre les groupes doivent être égaux.
- **Test paramétrique** : ANOVA
- **Test non paramétrique** : Test de Friedman (hors programme)

4.3.1 Graphe



4.3.2 Diagnostic

- **Groupe L2** moyenne : 6.63 ; variance 23.69 ; effectifs : 104
- **Groupe L3** moyenne : 6.03 ; variance 15.47 ; effectifs : 123
- **Groupe M1** moyenne : 6.35 ; variance 34.66 ; effectifs : 36
- **Groupe M2** moyenne : 4.00 ; variance 7.81 ; effectifs : 23

Les conditions de validités ne sont pas remplies. Toutefois, l'ANOVA étant un test résistant aux conditions d'application (et le test de Friedman n'étant pas au programme), nous appliquons tout de même une analyse de variance.

4.3.3 Test

- H0 : Il n'y a pas de différence selon le niveau d'étude, les étudiants en début de cycle et en fin de cycle trichent dans les mêmes proportions.
- F de Fisher : 2.2
- Degrés de liberté : $df = 3 / 277$
- $p : 0.088$

P est inférieur à 5%, on accepte H0 :

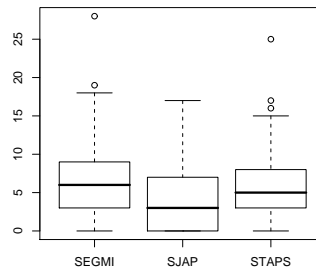
**Il n'y a pas de lien entre les variables [NIVEAUDETUDE] et [SCORETRICHETOTAL],
l'année d'étude n'a pas d'influence sur la triche.
(conclusion à relativiser en raison du non respect des conditions d'application)**

4.4 [SCORETRICHETOTAL] vs [UFR]

La variable [UFR] est une variable nominale à trois classes.

- **Représentation graphique** : boîtes à moustaches juxtaposées.
- **Diagnostic** : Les variances intra doivent être identiques, la variable continue doit être normale et les effectifs entre les groupes doivent être égaux.
- **Test paramétrique** : ANOVA
- **Test non paramétrique** : Test de Friedman (hors programme).

4.4.1 Graphe



4.4.2 Diagnostic

- **Groupe SEGMI** moyenne : 6.58 ; variance : 22.04 ; effectifs 105
- **Groupe SJAP** moyenne : 4.49 ; variance : 29.02 ; effectifs : 36
- **Groupe STAPS** moyenne : 6.19 ; variance : 16.70 ; effectifs : 146

Les conditions de validité ne sont pas remplies en particulier l'égalité des effectifs. Même remarque que précédemment, l'ANOVA étant un test résistant aux conditions d'application (et le test de Friedman n'étant toujours pas au programme), nous appliquons tout de même une analyse de variance.

4.4.3 Test

- Il n'y a pas de différence entre les UFR, les étudiants des différentes filières trichent dans les mêmes proportions.
- F : 2.87
- Degrés de liberté : $df = 2 / 278$
- p : 0.05821

P est supérieur à 5%, on rejette H_0 :

Il n'y a pas de lien entre les variables UFR et ScoreTricheTotal, les étudiants des différentes UFR trichent dans les mêmes proportions. (conclusion à relativiser en raison du non respect des conditions d'application)

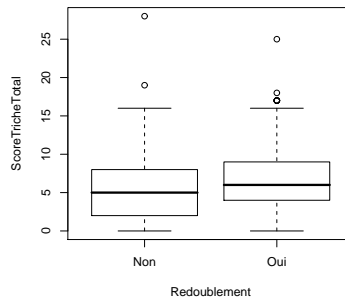
On notera tout de même que le petit p est très proche de 5% et que les étudiants de l'UFR SJAP semblent frauder moins que les autres. Le faible effectif des SJAP et le manque de puissance qu'il implique explique peut-être la non significativité du p.

4.5 [SCORETRICHETOTAL] VS [REDOUBLEMENT]

La variable [REDOUBLEMENT] est une variable nominale à deux classes.

- **Représentation graphique** : boîtes à moustaches.
- **Diagnostic** : la variable continue doit être normale ou d'effectif supérieur à 30. De plus, les variances des groupes doivent être égales.
- **Test paramétrique** : T de Student
- **Test non paramétrique** : Test des rangs de Wilcoxon

4.5.1 Graphe



4.5.2 Diagnostic

- **Groupe Non** moyenne : 5.45 ; variance 18.67
- **Groupe Oui** moyenne : 6.81 ; variance 21.33
- Les variances ne sont pas significativement différentes et les effectifs sont supérieurs à 30 : on applique le T de Student.

4.5.3 Test

- H0 : Il n'y a pas de différence selon le redoublement, ceux qui ont et n'ont pas redoublé trichent dans les mêmes proportions.
- test t : 2.54
- Degrés de liberté : df = 279
- p : 0.011

P est inférieur à 5%, on rejette H0 :

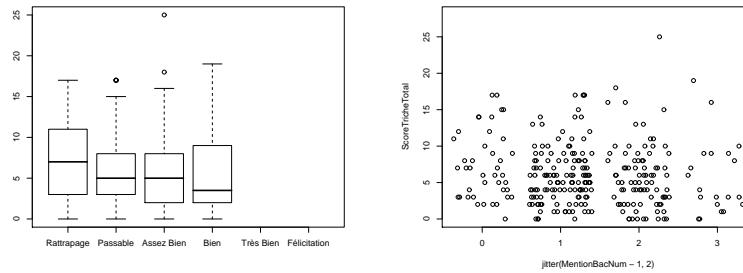
Il y a un lien entre les variables [REDOUBLEMENT] et [SCORETRICHETOTAL], ceux qui ont redoublé dans leur scolarité trichent significativement plus que les autres.

4.6 [SCORETRICHETOTAL] VS [MENTIONBAC]

La variable [MENTIONBAC] est une variable ordonnée à 6 classes. Nous la transformons en variable discrète.

- **Représentation graphique** : boîtes à moustaches et nuage de points.
- **Diagnostic** : une des deux variables doit suivre une loi normale.
- **Test paramétrique** : r de Pearson
- **Test non paramétrique** : r de Spearman

4.6.1 Graphe



4.6.2 Diagnostic

La variable [MENTIONBAC] considérée comme une variable discrète suit approximativement une loi normale, on applique le r de Pearson.

4.6.3 Test

- H_0 : Il n'y a pas de différence selon la mention au bac, ceux qui ont eu des bonnes et des moins bonnes mentions trichent dans les mêmes proportions.
- r de Pearson : -0.097
- Degrés de liberté : $df = 275$
- p : 0.1082

P est supérieur à 5%, on accepte H_0 :

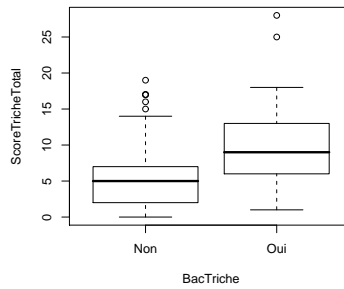
Il n'y a pas de lien entre les variables [MENTIONBAC] et [SCORETRICHETOTAL], la mention au bac est indépendante du score de triche.

4.7 [SCORETRICHETOTAL] vs [BACTRICHE]

La variable [BACTRICHE] est une variable nominale à deux classes.

- **Représentation graphique** : boîtes à moustaches.
- **Diagnostic** : la variable continue doit être normale ou d'effectif supérieur à 30. De plus, les variances des groupes doivent être égales.
- **Test paramétrique** : T de Student
- **Test non paramétrique** : Test des rangs de Wilcoxon

4.7.1 Graphe



4.7.2 Diagnostic

- **Groupe Non** moyenne : 6.85 ; variance 14.26
- **Groupe Oui** moyenne : 9.65 ; variance 30.42
- Les variances sont significativement différentes : on applique un test des rangs de Willcoxon.

4.7.3 Test

- H_0 : Il n'y a pas de différence selon le fait d'avoir triché au bac, ceux qui ont et n'ont pas triché au bac trichent dans les mêmes proportions.
- test W : 3085.5
- p : 6.17e-09

P est inférieur à 5%, on rejette H_0 :

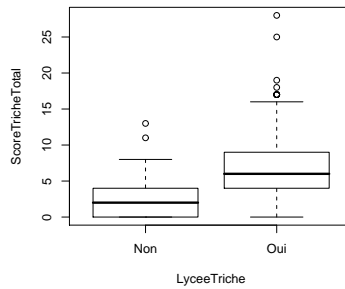
Il y a un lien entre les variables [TRICHEBAC] et [SCORETRICHETOTAL], ceux qui ont triché au bac trichent significativement plus que les autres à l'université.

4.8 [SCORETRICHETOTAL] vs [LYCEETRICHE]

La variable [LYCEETRICHE] est une variable nominale à deux classes.

- **Représentation graphique** : boîtes à moustaches.
- **Diagnostic** : la variable continue doit être normale ou d'effectif supérieur à 30. De plus, les variances des groupes doivent être égales.
- **Test paramétrique** : T de Student
- **Test non paramétrique** : Test des rangs de Wilcoxon

4.8.1 Graphe



4.8.2 Diagnostic

- **Groupe Non** moyenne : 2.80 ; variance 20.13
- **Groupe Oui** moyenne : 6.69 ; variance 9.16
- Les variances sont significativement différentes : on applique un test des rangs de Wilcoxon.

4.8.3 Test

- H_0 : Il n'y a pas de différence selon le fait d'avoir triché au lycée, ceux qui ont et n'ont pas triché au lycée trichent dans les mêmes proportions à l'université.
- test W : 2070
- p : 2.764e-09

P est inférieur à 5%, on rejette H_0 :

**Il y a un lien entre les variables [LYCEE_BAC] et [SCORE_TRICHE_TOTAL],
ceux qui ont triché au lycée trichent significativement plus que les autres à l'université.**

4.9 Bilan

Naturellement, au vu du nombre important de tests (nombre qui entraîne l'augmentation du risque de première espèce bien au-delà du classique 5%), les résultats obtenus dans l'analyse bivariée ne sont valables qu'à titre exploratoire.

- L'âge n'a pas d'influence sur le score de triche.
- Les hommes trichent significativement plus que les femmes.
- Le niveau d'étude n'a pas d'influence sur le score de triche.
- L'UFR n'a pas d'influence.
- Ceux qui ont déjà redoublé dans leur scolarité ont un score de triche plus élevé.
- La mention au bac n'a pas d'influence sur la triche.
- Le fait d'avoir triché au bac a une forte influence.
- Le fait d'avoir triché au lycée a également une forte influence.

A noter toutefois, il est possible que les étudiants en SJAP soient moins fraudeurs que les autres mais qu'un manque de puissance nous empêche de le détecter.

De plus, les résultats concernant l'UFR et le niveau d'étude sont à relativiser, les conditions d'application des tests n'étant pas vraiment vérifiées.

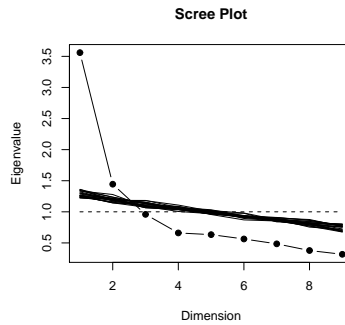
5 Validation de l'échelle (M1)

5.1 α de Cronbach

L' α de Cronbach mesure la cohérence interne de l'échelle. On trouve $\alpha = 0.80$.

5.2 Unidimensionnalité

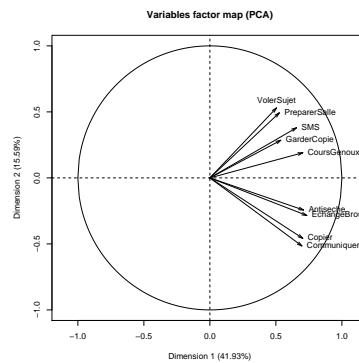
Pour tester l'unidimensionnalité, nous traçons le diagramme des valeurs propres.



Il y a clairement une dimension principale et une secondaire. Notre échelle semble donc raisonnablement adaptée à ce à quoi elle se destine.

5.3 Analyse du cercle des corrélations

Le nombre de dimensions (deux et non une seule) nous indique déjà que le score triche n'est pas parfait puisqu'il n'est pas unidimensionnel, mais bidimensionnel (il mesure deux types de pratiques et non une seule). L'examen du cercle des corrélations va nous permettre d'en savoir plus :



Le score de triche est bien lié à la première composante principale. Par contre, les pratiques se divisent en deux blocs relativement indépendants (présence d'un angle droit) : D'un coté, les méthodes classiques que l'on pourrait (hélas) appeler "communes" les plus souvent employées : communiquer, copier, préparer des antisèches, échanger des brouillons. De l'autre, les méthodes plus rares (plus graves ?) : avoir le cours sur les genoux, SMS, préparer la salle, voler le sujet, garder sa copie.

La première composante correspond donc au fait de tricher : plus les individus sont graphiquement à droite, plus leur score de triche est important. La seconde composante sépare les étudiants ayant recours aux méthodes classiques de triche (en bas) des méthodes autres pouvant être considérées comme plus graves (en haut).

5.4 Comparaison avec d'autres mesures

Nous ne disposons pas à proprement parler de réelle mesure de fraude sur l'échantillon que nous considérons. Cependant, les variables [TRICHEBAC] et [TRICHELYCEE] peuvent être considérées comme des indicateurs potentiels d'une propension à la fraude.

L'analyse bivariée donne :

- [SCORETOTALTRICHE] vs [BACTRICHE]
 - Moyenne des *Non* : 6.85 ; Moyenne des *Oui* : 9.65
 - Différence entre les *Oui* et les *Non* : 2.80
 - Test des rangs de Wilcoxon 3085.5 ; petit p significatif ($p=6.17e-09$)
- [SCORETOTALTRICHE] vs [LYCEETRICHE]
 - Moyenne des *Non* : 2.80 ; Moyenne des *Oui* : 6.69
 - Différence de score selon les groupes : 3.89
 - Test des rangs de Wilcoxon 2070, petit p significatif ($p=2.764e-09$)

Dans un cas comme dans l'autre, il y a un lien très fort entre le score de triche mesuré par notre échelle et les indicateurs de fraude.

Cela augmente la confiance que l'on peut avoir dans l'échelle.

6 Construction d'un modèle (M1)

6.1 Sélection des variables

- [SEXE] et [REDOUBLEMENT] : ces variables sont significativement associées à [SCORETRICHETOTAL], elles sont conservées pour le modèle.
- [UFR] est à la limite de la significativité, elle est également conservée.
- [BACTRICHE] et [LYCEETRICHE] sont significativement associées à [SCORETRICHETOTAL] mais ces variables sont plus utiles pour vérifier la validité de l'échelle que pour l'expliquer. Elles ne sont donc pas incluses dans le modèle.
- [AGE] et [NIVEAUDETUDE] sont fortement corrélées entre elles. L'une est donc incluse, mais pas l'autre pour éviter une colinéarité. Les deux possibilités seront testées.
- La [MENTIONBAC] peut être incluse ou non, les deux possibilités seront testées.

Au final, nous allons donc comparer 4 modèles. (~ est le symbole signifiant "est expliqué par")

- ScoreTricheTotal~Sexe+Redoublement+UFR+NiveauDEtude+MentionBac
- ScoreTricheTotal~Sexe+Redoublement+UFR+Age+MentionBac
- ScoreTricheTotal~Sexe+Redoublement+UFR+NiveauDEtude
- ScoreTricheTotal~Sexe+Redoublement+UFR+Age

6.2 Comparaison des modèles

Nous sélectionnons le modèle selon la proportion de variance qu'il explique, donc en fonction du r^2 ajusté.

- NiveauDEtude+MentionBac : $r^2 = 0.080$

- Age+MentionBac : $r^2 = 0.056$
- NiveauDEtude : $r^2 = 0.074$
- Age : $r^2 = 0.051$

Le premier modèle semble donc être le plus pertinent. On constate que la variance expliquée par le modèle est relativement faible. Cela pourrait être amélioré en incluant les variables [BACTRICHE] et [LYCEETRICHE], mais expliquer le fait qu'un étudiant triche par le fait qu'il trichait avant ne nous paraît pas vraiment pertinent. Nous conservons donc un modèle "moins explicatif", mais "plus informatif".

6.3 Détail du modèle

Le modèle sélectionné est donc :

ScoreTricheTotal ~ Sexe + Redoublement + UFR + NiveauDEtude + MentionBac

Le résultat de la régression linéaire est le suivant :

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	5.08628	0.61617	8.255	7.25e-15	***
SexeHomme	1.44602	0.53283	2.714	0.00709	**
RedoublementOui	1.22412	0.52487	2.332	0.02044	*
UFRSJAP	-1.97386	0.88103	-2.240	0.02589	*
UFRSTAPS	-0.39193	0.62312	-0.629	0.52990	
NiveauDEtude.L	-1.66244	0.69439	-2.394	0.01736	*
NiveauDEtude.Q	-0.64440	0.66175	-0.974	0.33105	
NiveauDEtude.C	-1.24343	0.69930	-1.778	0.07653	.
MentionBac.L	-0.57143	0.80668	-0.708	0.47934	
MentionBac.Q	1.07252	0.65546	1.636	0.10297	
MentionBac.C	-0.01764	0.46448	-0.038	0.96973	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

6.4 Bilan

Le modèle présente les variables qui ont un impact sur le score de triche après ajustement sur les autres variables¹

- Le [SEXE] a une influence sur le score de triche. En moyenne, le score de triche des garçons est supérieur de 1.4 points à celui des filles.
- Ceux qui ont redoublé ont un score de triche supérieur de 1.26 points de plus que ceux qui n'ont pas redoublé.
- Les étudiants en SJAP trichent moins, leur score est en moyenne inférieur de -1.93 point à celui des SEGMI et -1.58 point à celui des STAPS.
- Concernant le niveau d'étude et la mention au bac (variables ordonnées), aucune conclusion ne peut être tirée, il faudrait préalablement faire un test global (hors programme). Mais vu la faiblesse des valeurs individuelles, il y a peu de chance que ces variables soient significatives.

¹Ce type d'ajustement permet d'éviter les facteurs de confusion. Par exemple, s'il est avéré que les femmes trichent moins, un UFR dont la population serait essentiellement féminine présenterait l'apparence d'un UFR où l'on triche peu. Cette illusion constituerait un biais. L'ajustement de la variable UFR par la variable Sexe permet de comparer les UFR à sexe constant et donc corrige ce biais potentiel.

7 Conclusion générale

7.1 Réponses aux questions initiales

En se basant sur les résultats apportés par le modèle (résultats les plus fiables puisqu'ils prennent en compte les ajustements sur les autres variables), on peut répondre à nos questions initiales :

- Les STAPS ne trichent pas plus que les autres. Par contre, les SJAP trichent moins.
- Les hommes trichent plus que les femmes.
- Le niveau d'étude n'a pas d'impact sur la fraude.

7.2 Limite de l'étude

Même s'ils présentent une certaine cohérence, ces résultats sont toutefois à relativiser. En particulier :

- La fraude étant illégale, la désirabilité sociale peut avoir biaisé les résultats.
- Les étudiants enquêteurs ne sont pas des professionnels de l'enquête et les conditions de passation des questionnaires n'étaient parfois pas optimales.
- Aucun contrôle n'a été fait concernant l'échantillonnage ou le choix des sujets.
- Enfin, le choix de faire cette enquête a été une décision collective. Les étudiants de L2 ont eu à voter² pour choisir entre un projet "intéressant mais demandant un investissement" et utiliser une base de donnée "moins intéressante, mais déjà construite". Néanmoins, certains d'entre eux ont pu considérer cette enquête comme une contrainte et ont pu avoir la tentation de remplir eux-même les questionnaires ou de les faire remplir par leurs copains.

²Vote a main levé hors présence enseignante