



UV101E : Projet Statistiques 12

Utilisation des MOOCs et satisfaction

Rapport

Version 1

F3B - Année scolaire 2017-2018

Groupe 12:

LOURIZ Riahi
AMAROUCH Toufik
MALIKI Rochd
BOUSHAB Youssef
ED-DAHABI Youssef

Encadrants techniques :

BILLOT Romain
COPPIN Gilles



IMT Atlantique

Bretagne-Pays de la Loire
École Mines-Télécom

Table des matières

1. Introduction	2
1.1. Contexte	2
1.2. But du projet	2
2. Représentativité	2
2.1. Représentativité par sexes.....	2
2.2. Représentativité par origines	2
2.3. Représentativité des femmes par sexes et origines :	3
3. Tests statistiques	3
3.1. Tests de satisfaction	3
3.2. Tests du Khi2.....	7
3.3. Tests sur le prix	10
3.4. Tests sur le temps alloué à un cours en moyenne	12
4. ACM :	15

1. Introduction

1.1. Contexte

L'acronyme MOOC (Massive Online Open Course) est utilisé pour désigner les plateformes d'enseignement en ligne qui proposent des cours ouverts et qui peuvent s'adresser à des centaines, des milliers, voire des dizaines de milliers d'étudiants simultanément.

De nombreuses institutions (grandes écoles, universités, grandes entreprises, etc.) investissent dans la création de ce type de cours avec divers objectifs : aider au développement de la formation continue, de la formation des pays en voie de développement, accroître leur visibilité.

La finalité de ce projet n'est pas de proposer une analyse du phénomène MOOC, de ses enjeux, de ses réussites, de ses échecs ou de son avenir, mais d'offrir une vision originale d'un MOOC, au travers d'une base de sondage représentant la population des étudiants - constituée des élèves de l'IMT Atlantique, de leur profil, leur activité et de leurs motivations.

1.2. But du projet

Compte tenu de ce qui précède, l'objectif du projet est de déterminer pour un étudiant une combinaison de facteurs susceptibles d'apporter satisfaction ou non à ce dernier.

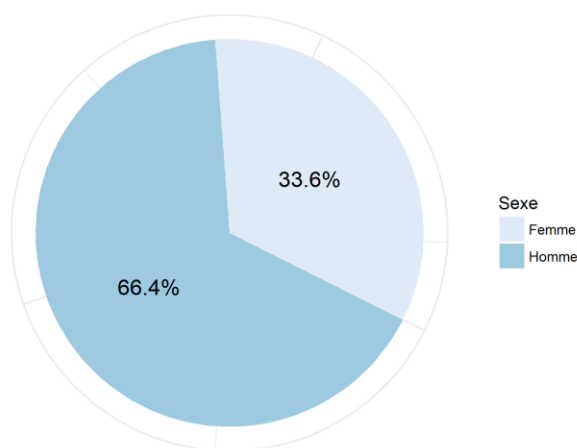
2. Représentativité

2.1. Représentativité par sexes

En se basant sur les données de la source [1], on réalise que le taux de fille à l'IMT-Atlantique est de 25%.

Ainsi, on peut observer la représentativité :

Sexe	Représentativité
Femme	134.4%
Homme	80.54%



Nous constatons donc que la population des filles est surreprésentée par rapport à celles des garçons.

En effet, elle ne représente que 25% des effectifs de l'école alors qu'elle est représentée dans notre échantillon par une proportion de 33.6%. Les garçons sont alors moins intéressés, puisque leur participation ne représente pas la population réelle.

2.2. Représentativité par origines

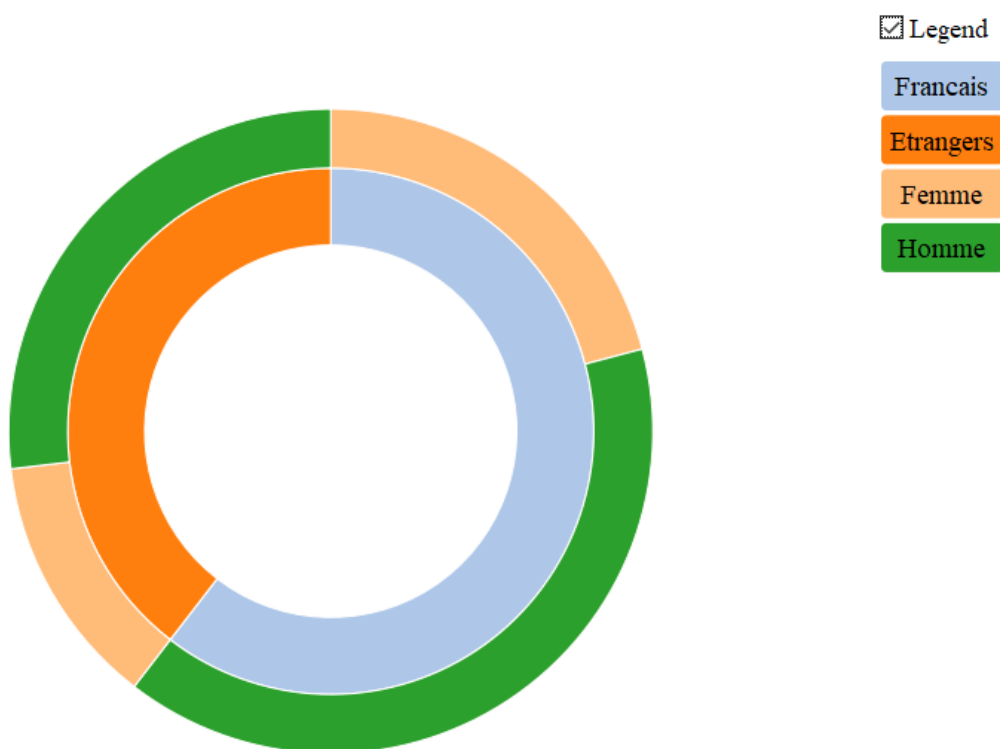
En se basant sur les données de la source [1] on réalise que le taux des étrangers à l'IMT-Atlantique est de 42,2%.

On peut observer la représentativité :

Origines	Représentativité
France	93.83%
Autre	104.5%

Les effectifs dans ce cas sont bien représentés puisque notre échantillon est presque égal à la population réelle en termes de proportion. Et les étrangers et les français ont répondu au questionnaire avec les proportions attendues.

2.3. Représentativité des femmes par sexes et origines :

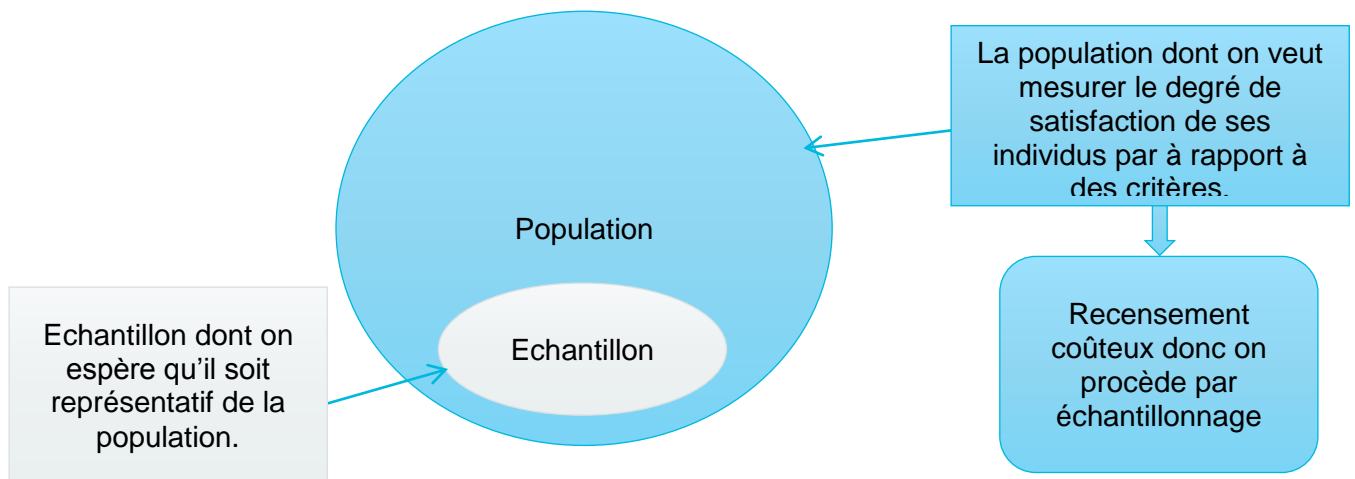


Représentativité	Filles
Françaises	138.8%
Etrangères	128.3%

La sur-représentativité des filles est due à la fois à la participation des filles étrangères et françaises.

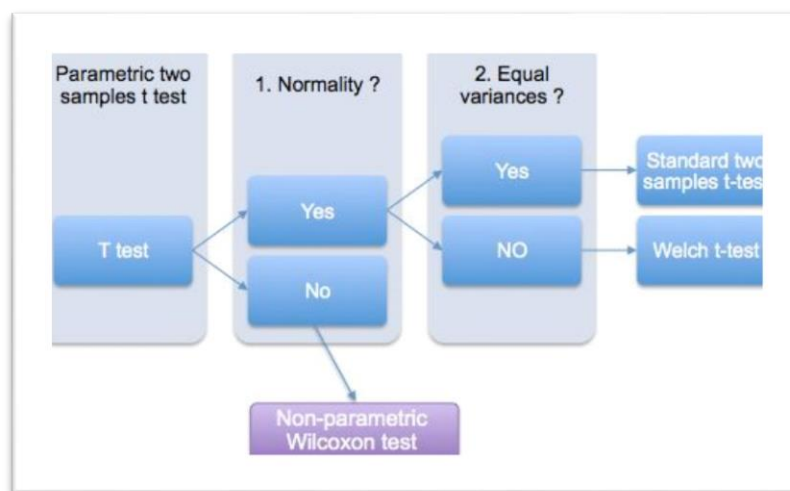
3. Tests statistiques

3.1. Tests de satisfaction



On veut comparer le degré de satisfaction envers les MOOCs en fonction de la catégorie du répondant à notre sondage : le sexe a-t-il une influence sur la satisfaction ? Et qu'en est-il de l'origine du répondant ? De son choix des plateformes ? De la nature des MOOCs suivis ? On va alors effectuer plusieurs comparaisons de deux échantillons en partant toujours d'une hypothèse nulle H_0 qui considère que le degré de satisfaction des deux échantillons est égal (sauf mention de contraire).

Dans tous les tests ci-dessous, on va procéder comme montré sur la figure suivante [3] :



Hypothèse 1 : Le degré de satisfaction des femmes et des hommes est le même

	mean	sd
Femmes	6.578	1.926
Hommes	6.610	1.987

Test de normalité sur les deux échantillons

- Hommes :

```
shapiro.test(homme_series$satisfaction)
Shapiro-Wilk normality test
data:  homme_series$satisfaction
W = 0.93798, p-value = 0.0009714
```

Donc avec un risque de 10%, on rejette l'hypothèse nulle de normalité pour l'échantillon des hommes.

- Femmes :

```
shapiro.test(femme_series$satisfaction)
Shapiro-Wilk normality test
data:  femme_series$satisfaction
W = 0.92978, p-value = 0.01974
```

Toujours avec un risque de 10% on rejette l'hypothèse nulle de normalité.
Les deux échantillons ne suivent pas un des lois normales. Par conséquent, On va utiliser le test non paramétrique : test de Wilcoxon.

Test de moyenne

```
wilcox.test(femme_series$satisfaction, homme_series$satisfaction)
Wilcoxon rank sum test with continuity correction
data:  femme_series$satisfaction and homme_series$satisfaction
W = 1449.5, p-value = 0.9374
alternative hypothesis: true location shift is not equal to 0
```

Avec un risque de 10%, on accepte H0 et donc il n'y a pas de différence de satisfaction entre les deux échantillons. Sous réserve la conformité de la représentativité de notre échantillon, on peut conclure que le degré de satisfaction des femmes et hommes dans la population est le même.

Hypothèse 2 : Le degré de satisfaction des français est inférieur à celui des étrangers :

	Mean	Sd
Français	6.289	2.029
Etrangers	7.065	1.768

Test de normalité sur les deux échantillons

- Français :

```
shapiro.test(francais$satisfaction)
Shapiro-Wilk normality test
data:  francais$satisfaction
W = 0.94009, p-value = 0.002476
```

- Étrangers :

```
shapiro.test(ettrangers$satisfaction)
Shapiro-Wilk normality test
data:  ettrangers$satisfaction
W = 0.95097, p-value = 0.05113
```

Avec un risque de 10% on rejette l'hypothèse nulle pour les deux échantillons.

Test de moyenne

```
wilcox.test(francais$satisfaction, ettrangers$satisfaction, alternative = 'g')
Wilcoxon rank sum test with continuity correction
data:  francais$satisfaction and ettrangers$satisfaction
```

W = 1261.5, p-value = 0.9706
alternative hypothesis: true location shift is greater than 0

Avec un risque de 10% on accepte H0. Les étrangers sont plus satisfaits par rapport aux français.

Hypothèse 3 : Le degré de satisfaction des répondants qui préfèrent les plateformes avec aide financière est supérieur à celui des autres

	Mean	Sd
Oui (avec aide financière)	7.70	1.45
Non	6.30	1.975

Test de normalité sur les deux échantillons

- Avec aide financière :

```
shapiro.test(aide_oui$satisfaction)
Shapiro-Wilk normality test
data:  aide_oui$satisfaction
W = 0.86242, p-value = 0.003751
```

- Sans aide financière :

```
shapiro.test(aide_non$satisfaction)
Shapiro-Wilk normality test
data:  aide_non$satisfaction
W = 0.95453, p-value = 0.003012
```

Avec un risque de 10% on rejette H0 pour les deux échantillons.

Test de moyenne

```
wilcox.test(aide_oui$satisfaction,aide_non$satisfaction,alternative
='l')
Wilcoxon rank sum test with continuity correction
data:  aide_oui$satisfaction and aide_non$satisfaction
W = 1566.5, p-value = 0.9995
alternative hypothesis: true location shift is less than 0
```

Avec un risque de 10%, on accepte H0. Comme commentaire pour cette hypothèse :
« plus vous payez, moins vous êtes satisfait ». Ceci peut être vrai surtout pour les personnes qui paient beaucoup cher pour un cours qu'ils ont apprécié au début par son intitulé, mais qu'ils n'ont rien appris durant la formation.

Hypothèse 4 : Les sondé qui suivent des MOOCs hors leurs études sont plus satisfaits que ceux qui ne font que des MOOCs dans le cadre de leur étude

	mean	Sd
Cadre des études	6.27	1.45
Hors cadre des études	7.76	1.58

Test de normalité sur les deux échantillons

- Cadre des études :

```
shapiro.test(inclu_oui$satisfaction)
Shapiro-Wilk normality test
data:  inclu_oui$satisfaction
W = 0.94409, p-value = 0.0007357
```

- Hors cadre des études :

```
shapiro.test(inclu_non$satisfaction)
Shapiro-Wilk normality test
data:  inclu_non$satisfaction
W = 0.94114, p-value = 0.1573
```

Avec un risque de 10%, on rejette H_0 pour le premier échantillon. Alors que H_0 est acceptée pour le deuxième échantillon. Mais pour l'exactitude des résultats on va utiliser un test non paramétrique.

Test de moyenne

```
wilcox.test(inclu_non$satisfaction, inclu_oui$satisfaction, alternative
= 'l')
Wilcoxon rank sum test with continuity correction
data:  inclu_non$satisfaction and inclu_oui$satisfaction
W = 1613.5, p-value = 0.9996
alternative hypothesis: true location shift is less than 0
```

Avec un risque de 10%, on accepte H_0 . Ceci peut être expliqué par le fait que les MOOCs dans le cadre des études ne sont pas forcément ce que l'étudiant apprécie le plus. Moi personnellement, j'ai fait un MOOC dans le cadre des études ici à IMT Atlantique dans le domaine du réseau que je n'apprécie pas beaucoup. Par conséquent, je n'ai rien appris et je n'ai été pas satisfait à la fin de ce MOOC.

Conclusion : À l'issue de ces tests statistiques, on peut dire que le degré de satisfaction n'est pas le même pour tous les groupes de notre échantillon et par suite de la population (grâce à la représentativité). Au respect des résultats ci-dessus, nous proposons pour les écoles, qui imposent des MOOCs dans le cadre des études, de laisser le choix à l'étudiant pour le MOOC qu'il apprécie le plus.

3.2 Tests du Khi2

Nous allons étudier ci-dessous l'influence de plusieurs éléments de notre questionnaire sur la terminaison des MOOCs d'un groupe de personnes les ayant suivis.

Pour chaque hypothèse, les résultats sont regroupés dans une table de contingence comme indiqué sur le tableau ci-dessous.

Le test du Khi2 examinera si les lignes et colonnes de la table de contingence sont significativement associées, statistiquement.

Hypothèse nulle (H_0) : les variables des lignes et des colonnes de la table de contingence sont indépendantes. Hypothèse alternative (H_1) : les variables des lignes et des colonnes sont dépendantes.

Hypothèse 5 : La terminaison des MOOCs ne dépend pas du genre.

Catégorie	Femme	Homme	Total
Suivi et réussi	24	49	73
Suivi mais pas terminé	14	28	42
Total	38	77	115

Sur R, le test du Khi deux donne une valeur de p-value=1 et avec un risque de 10%, on garde l'hypothèse nulle et donc conclure que les deux variables catégorielles sont indépendantes.

```
Chisq.test(data_suivi$a_suivi,data_suivi$sexe)
Pearson's Chi-squared test with Yate's continuity correction
data: data_suivi$a_suivi and data_suivi$sexe
X-squared = 0, df = 1, p-value = 1
```

Hypothèse 6 : La terminaison des MOOCs ne dépend pas de la formation suivie

Catégorie	FIG	FIL	FIP	MSc	Master recherche	Master spécialisé	Total
Suivi et réussi	61	2	0	6	1	3	73
Suivi mais pas terminé	36	2	3	1	0	0	42
Total	97	4	3	7	1	3	115

Avec un risque de 10%, on rejette H0 pour la formation suivie.

```
Chisq.test(data_suivi$a_suivi,data_suivi$formation)
Pearson's Chi-squared test
data: data_suivi$a_suivi and data_suivi$formation
X-squared = 9.3367, df = 5, p-value = 0.09637
```

Hypothèse 7 : La terminaison des MOOCs ne dépend pas du background du répondant

Catégorie	école	DUT	Prépa	Université	Total
Suivi et réussi	4	2	59	8	73
Suivi mais pas terminé	0	2	36	4	42
Total	4	4	95	12	115

Toujours avec un risque de 10% on accepte H0, donc les colonnes et les lignes sont indépendantes.

```
Chisq.test(data_suivi$a_suivi,data_suivi$background)
Pearson's Chi-squared test
data: data_suivi$a_suivi and data_suivi$background
X-squared = 2.7447, df = 3, p-value = 0.4327
```

Hypothèse 8 : La terminaison des MOOCs ne dépend pas de la plateforme utilisée

```
Chisq.test(data_suivi$a_suivi,data_suivi$plateformes_utilisees)
Pearson's Chi-squared test
data: data_suivi$a_suivi and data_suivi$plateformes_utilisees
X-squared = 28.772, df = 29, p-value = 0.477
```

Même conclusion.

Hypothèse 5 : La terminaison des MOOCs ne dépend pas de périodes de suivi de MOOCs

Catégorie	Vacances	Etudes	Les deux	Total
Suivi et réussi	6	33	34	73
Suivi mais pas terminé	3	15	24	42
Total	9	48	58	115

Toujours avec un risque de 10% on accepte H0, donc les colonnes et les lignes sont indépendantes.

```
Chisq.test(data_suivi$a_suivi,data_suivi$vacances_etude)
Pearson's Chi-squared test
data: data_suivi$a_suivi and data_suivi$vacances-etude
X-squared = 1.2052, df = 2, p-value = 0.5474
```

Il est aussi pertinent de s'intéresser à l'influence de la motivation de la personne sur la terminaison ou pas des MOOCs. Pour cela nous effectuons trois autres hypothèses.

Hypothèse 9 : La terminaison des MOOCs ne dépend pas de la motivation envers les certificats

Catégorie	D'accord	Désaccord	Neutre	Tout à fait d'accord	Tout à fait en désaccord	Total
Suivi et terminé	20	15	9	21	8	73
Suivi mais pas terminé	7	10	8	9	8	42
Total	27	25	17	30	16	115

Sur R, le test de Khi deux donne une valeur de p-value=0.3984 et avec un risque de 10%, on garde l'hypothèse pour la motivation envers les certificats.

```
Chisq.test(data_suivi$a_suivi,data_suivi$motivation_certificat)
Pearson's Chi-squared test
data: data_suivi$a_suivi and data_suivi$motivation_certificat
X-squared = 4.0563, df = 4, p-value = 0.3984
```

Hypothèse 10 : La terminaison des MOOCs ne dépend pas de la recherche d'emploi

Catégorie	D'accord	Désaccord	Neutre	Tout à fait d'accord	Tout à fait en désaccord	Total
Suivi et terminé	17	22	20	2	12	73
Suivi mais pas terminé	16	8	12	4	12	42
Total	33	30	32	6	24	115

```
Chisq.test(data_suivi$a_suivi,data_suivi$trouver_emploi)
Pearson's Chi-squared test
data: data_suivi$a_suivi and data_suivi$trouver_emploi
X-squared = 6.5827, df = 4, p-value = 0.1597
```

Hypothèse 11: La terminaison des MOOCs ne dépend pas de vouloir performer dans ses projets de travail

Catégorie	D'accord	Désaccord	Neutre	Tout à fait d'accord	Tout à fait en désaccord	Total
Suivi et terminé	33	1	13	25	1	73
Suivi mais pas terminé	18	3	9	12	0	42

Sur R, le test de Khi deux donne une valeur de p-value=0.461 et avec un risque de 10%, on garde H0 pour la performance dans les projets de travail.

```
Chisq.test(data_suivi$a_suivi,data_suivi$performer_travail_projet)
Pearson's Chi-squared test
data: data_suivi$a_suivi and data_suivi$performer_travail_projet
X-squared = 3.6126, df = 4, p-value = 0.461
```

Tableau récapitulatif :

Elément traité dans le questionnaire	Terminaison de MOOCs
Genre	Indépendance
Formation suivie	Dépendance
Background du répondant	Indépendance
Plateforme utilisée	Indépendance
Période de suivi de MOOCs	Indépendance
Motivation envers les certificats	Indépendance
Trouver un emploi	Indépendance
Besoin de performer dans ses projets de travail	Indépendance

On peut discuter de la significativité des tests ; par exemple dans la table de contingence de la formation suivie, il y a une forte association entre la colonne FIP et la rangée "Suivi et terminé" ce qui n'est pas forcément le cas puisqu'il y a en réalité des FIP ayant suivi et terminé des MOOCs, ce qui peut fausser les résultats.

3.3 Tests sur le prix

Dans cette partie nous allons voir les différents facteurs qui influencent le prix que les élèves sont prêts à payer pour un MOOCs : nous allons faire une comparaison entre la formation suivie par les individus, de la présence ou non de l'aide financière.

Dans toute la partie suivante nous posons l'hypothèse H0 : le prix payé par les deux catégories est le même

Hypothèse 12 : le prix est le même avec et sans aide financière

	mean	Sd
Oui	21.8	27.4
Non	15.97	23.73

On effectue un test de normalité sur les deux échantillons :

Oui :

```
Shapiro-Wilk normality test
W = 0.77187, p-value = 0.0003386
```

On remarque que pour un risque de 10% la distribution n'est pas normale.

Non :

```
Shapiro-Wilk normality test
W = 0.68885, p-value = 2.102e-11
```

De même la distribution ici n'est pas normale.

Les deux distributions ne sont pas normales donc on procède à un test de wilcoxon le résultat est le suivant :

Wilcoxon rank sum test with continuity correction
W = 636, p-value = 0.2578

Les deux échantillons sont relativement différents.

Hypothèse 13: le prix est le même pour les FIG et les MSC

	mean	Sd
FIG	15.68	22.39
MSC	16.66	22.50

On effectue un test de normalité sur les deux échantillons.

FIG :

Shapiro-Wilk normality test
W = 0.70773, p-value = 4.347e-11

On remarque que pour un risque de 10% la distribution n'est pas normale.

MSC :

De même la distribution ici n'est pas normale.

Les deux distributions ne sont pas normales donc on procède à un test de wilcoxon le résultat est le suivant :

Wilcoxon rank sum test with continuity correction
W = 224, p-value = 0.9075

Les deux échantillons sont relativement différents.

Hypothèse 13 : le prix est le même pour les FIG et les FIP

	mean	Sd
FIG	15.68	22.39
FIP	11.045	12.79

On effectue un test de normalité sur les deux échantillons.

FIP :

La distribution pour les FIP n'est pas normale.

Les deux distributions ne sont pas normales donc on procède à un test de wilcoxon le résultat est le suivant :

Wilcoxon rank sum test with continuity correction
W = 63.5, p-value = 0.6807

Les deux échantillons sont relativement différents.

3.4 Tests sur le temps alloué à un cours en moyenne

Quelles sont les populations ayant passé plus de temps sur les MOOCs?

On souhaite comparer le temps alloué en moyenne pour un MOOCs en fonction de la catégorie du lecteur. Le sexe a-t-il une influence sur ce temps ? Et qu'en est-il de la nationalité du sondé ? Ou encore de l'année d'étude de l'élève ?

Filles Vs Garçons :

```
ddply(temps_sexe, ~sexe, summarise, mean=mean(temps_alloue_cours), sd=sd(temps_alloue_cours))
```

Sexe	mean	Sd
Femme	9.781250	2.445989
Homme	2.445989	4.031412

Les variances ne sont pas égales, donc allons voir si les distributions sont normales ou non ?

Pour les hommes

```
shapiro.test(hommes$temps_alloue_cours)
Shapiro-Wilk normality test
data:  hommes$temps_alloue_cours
W = 0.95853, p-value = 0.01713
```

Pour cette p-value et risque de 10%, on conclut que La distribution du temps alloué en moyenne par cours des hommes n'est pas normale.

Pour les femmes

```
shapiro.test(femmes$temps_alloue_cours)
Shapiro-Wilk normality test
data:  femmes$temps_alloue_cours
W = 0.88893, p-value = 0.003275
```

Même conclusion qu'avant. Donc on doit utiliser un test de Wilcoxon qui est non paramétrique :

```
wilcox.test(femmes$temps_alloue_cours, hommes$temps_alloue_cours)
Wilcoxon rank sum test with continuity correction
data:  femmes$temps_alloue_cours and hommes$temps_alloue_cours
W = 1200, p-value = 0.8235
```

Avec cette p-value et pour un risque de 10 %, on accepte H0 donc pas de différence de temps alloué entre les hommes et les femmes.

Nationalité Français Vs Etrangers :

```
ddply(temps_nat, ~Fr_vs_Etranger, summarise, mean=mean(temps_alloue_cours), sd=sd(temps_alloue_cours))
```

FR_vs_Etranger	mean	Sd
Etranger	9.809524	3.637462
Français	9.476190	3.618310

Les variances sont égales, donc allons voir si les distributions sont normales ou non ? Pour les français :

```
shapiro.test(fra$temps_alloue_cours)
Shapiro-Wilk normality test
data:  fra$temps_alloue_cours
W = 0.95913, p-value = 0.03515
```

Pour cette p-value, on conclut que La distribution du temps alloué en moyenne par cours des français n'est pas tout à fait normale et vu la courbe qui est proche d'une droite. On procède de même pour les Etrangers :

```
shapiro.test(etr$temps_alloue_cours)
Shapiro-Wilk normality test
data:  etr$temps_alloue_cours
W = 0.95443, p-value = 0.09314
```

Pour cette p-value et un risque de 10%, on conclut que La distribution du temps alloué en moyenne par cours des étrangers est presque normale.

On effectue un test de Student :

```
t.test(etr$temps_alloue_cours, fra$temps_alloue_cours)
Welch Two Sample t-test
data:  etr$temps_alloue_cours and fra$temps_alloue_cours
t = 0.46099, df = 87.698, p-value = 0.6459
```

Conclusion : D'après ce test, on garde H0 donc on conclut qu'il n'y a pas de différence de temps alloué entre les français et les étrangers.

Formation FIG vs FIP vs FIL vs Master of science vs Master Spé :

```
temps_form<-suivi[,c("temps_alloue_cours", "formation")]
ddply(temps_form, ~formation, summarise, mean=mean(temps_alloue_cours), sd
=sd(temps_alloue_cours))
```

Formation	mean	Sd
FIG	9.482759	3.601880
FIL	10.000000	3.829708
FIP	9.666667	5.033223
MSc	11.428571	2.992053
Master spé	9.333333	5.507571

Les variances ne sont pas égales, voyons les distributions :

```
shapiro.test(fig$temps_alloue_cours)
Shapiro-Wilk normality test
data:  fig$temps_alloue_cours
W = 0.95658, p-value = 0.005292
```

La distribution du temps n'est pas normale pour les FIG

```
shapiro.test(fip$temps_alloue_cours)
Shapiro-Wilk normality test
data:  fip$temps_alloue_cours
W = 0.98684, p-value = 0.7804
```

La distribution du temps est normale pour les FIP

```
shapiro.test(fil$temps_alloue_cours)
Shapiro-Wilk normality test
data:  fil$temps_alloue_cours
W = 0.86337, p-value = 0.2725
```

La distribution du temps est normale pour les FIL

```
shapiro.test(msc$temps_alloue_cours)
Shapiro-Wilk normality test
data:  msc$temps_alloue_cours
W = 0.90214, p-value = 0.3441
```

La distribution du temps est normale pour les Master of science

```
shapiro.test(mssp$temps_alloue_cours)
Shapiro-Wilk normality test
data:  mssp$temps_alloue_cours
W = 0.99725, p-value = 0.8999
```

La distribution du temps est normale pour les Master spécialisé.

Tests de wilcoxon sur le temps alloué pour les étudiants selon formation :

```
wilcox.test(fig$temps_alloue_cours,fip$temps_alloue_cours)
Wilcoxon rank sum test with continuity correction
data:  fig$temps_alloue_cours and fip$temps_alloue_cours
W = 132, p-value = 0.9817
```

Avec un risque de 10%, on accepte H0, les deux formations (FIG et FIP) ont le même temps alloué au MOOC.

```
wilcox.test(fip$temps_alloue_cours,fil$temps_alloue_cours)
Wilcoxon rank sum test with continuity correction
data:  fip$temps_alloue_cours and fil$temps_alloue_cours
W = 6, p-value = 1
```

Avec un risque de 5%, on accepte H0, les deux formations (FIP et FIL) ont le même temps alloué au MOOC.

```
wilcox.test(fig$temps_alloue_cours,fil$temps_alloue_cours)
Wilcoxon rank sum test with continuity correction
data:  fig$temps_alloue_cours and fil$temps_alloue_cours
W = 152, p-value = 0.6714
```

Avec un risque de 10%, on accepte H0, les deux formations (FIG et FIL) ont le même temps alloué au MOOC.

```
wilcox.test(fig$temps_alloue_cours,msc$temps_alloue_cours)
Wilcoxon rank sum test with continuity correction
data:  fig$temps_alloue_cours and msc$temps_alloue_cours
W = 193, p-value = 0.1038
```

Avec un risque de 10%, on accepte H0, les deux formations (FIG et Master of science) ont le même temps alloué au MOOC.

```
wilcox.test(fip$temps_alloue_cours,msc$temps_alloue_cours)
Wilcoxon rank sum test with continuity correction
```

```
data: fip$temps_alloue_cours and msc$temps_alloue_cours
W = 8, p-value = 0.6445
```

Avec un risque de 10%, on accepte H0, les deux formations (FIP et Master of science) ont le même temps alloué au MOOC.

```
wilcox.test(msc$temps_alloue_cours,mssp$temps_alloue_cours)
data: msc$temps_alloue_cours and mssp$temps_alloue_cours
W = 13, p-value = 0.6445
```

Avec un risque de 10%, on accepte H0, les deux formations (MSc et Master spé) ont le même temps alloué au MOOC.

```
wilcox.test(fig$temps_alloue_cours,mssp$temps_alloue_cours)
Wilcoxon rank sum test with continuity correction
data: fig$temps_alloue_cours and mssp$temps_alloue_cours
W = 138.5, p-value = 0.8636
```

Avec un risque de 10%, on accepte H0, les deux formations (FIG et Master spé) ont le même temps alloué au MOOC.

Conclusion : pas de différences entre les différentes formations en ce qui concerne le temps alloué à un MOOC en moyenne.

4. ACM

L'Analyse des Correspondances Multiples (ACM ou MCA pour Multiple Correspondence Analysis) est une extension de l'analyse factorielle des correspondances pour résumer et visualiser un tableau de données contenant plus de deux variables catégorielles. On peut aussi la considérer comme une généralisation de l'analyse en composantes principales lorsque les variables à analyser sont catégorielles plutôt que quantitatives (source STHDA, Abdi and Williams 2010).

L'étude faite ci-dessous est le fruit d'une autre ACM faite en dehors du rapport. En effet, une analyse d'ACM sera pertinente si on arrive à trouver les variables et individus pertinents. Pour en savoir plus sur la première étude voir annexes.

L'objectif de cette étude est d'identifier des groupes de personnes qui partagent un profil similaire envers les cours en ligne.

Les données que lesquelles on va faire l'ACM sont :

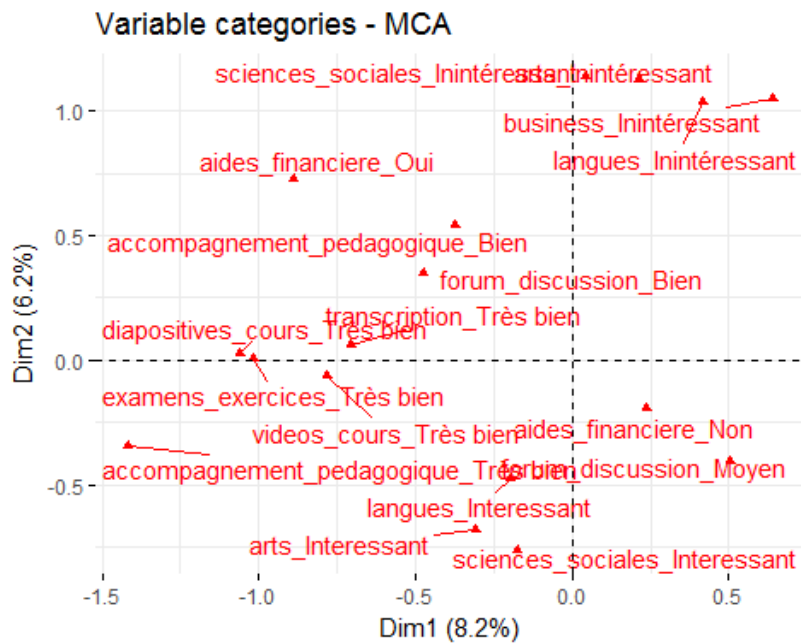
- Individus actifs (lignes 1:115 sauf 81): individus qui sont utilisés dans l'ACM.
- Variables actives (toutes les colonnes sauf les colonnes numériques+ plateforme_utilisee+ a_suivi+sexe, est_etudiant, formation, niveau_actuel): variables utilisées dans l'ACM.
- Variables
- supplémentaires (sexe, age, satisfaction, nbr_certificats, temps_alloue_cours): elles ne participent pas à l'ACM. Les coordonnées de ces variables seront prédites.

Nous avons éliminé l'individu 81 car il fausse l'étude. Ceci a été prouvé lors de la première étude. C'était un individu aberrant, chose qui a perturbé fortement les résultats.

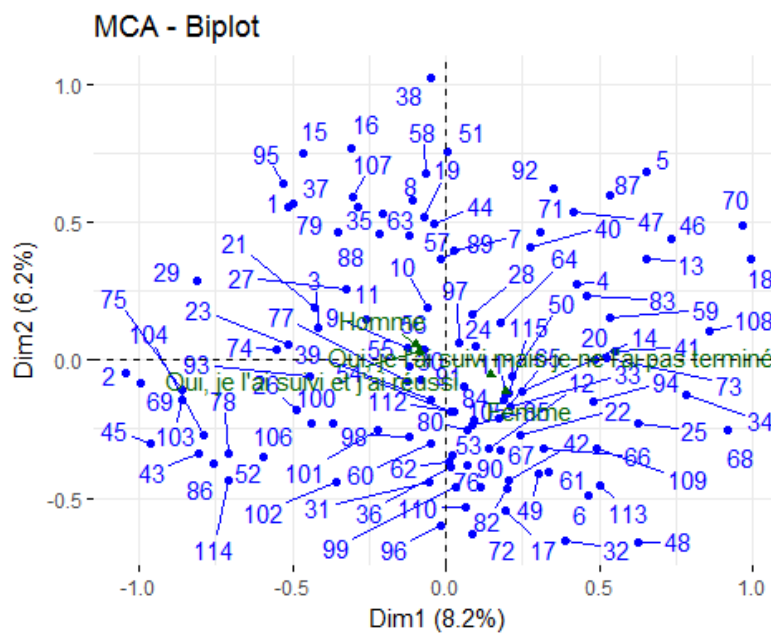
Résultats

Les valeurs propres

On affiche les 10 premières valeurs propres :



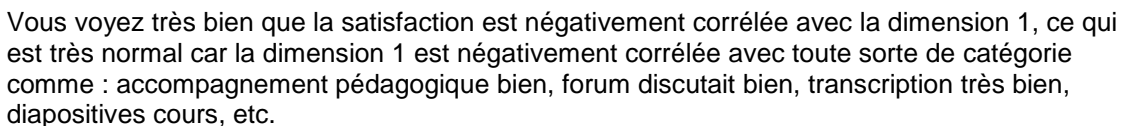
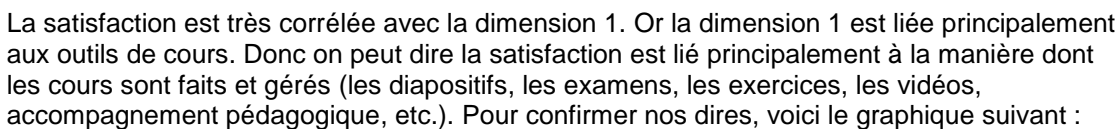
Individus et variables qualitatives supplémentaires



Analyse et critiques de deux graphiques ci-dessus : - Les sondés ayant suivi et réussi leurs MOOCs sont généralement des hommes qui ont trouvé les outils de cours intéressants. - Les sondés ayant suivi mais pas réussi leurs MOOCs sont généralement des femmes qui ont trouvé les outils de cours inintéressants. .

Revenons maintenant à l'interprétation des variables supplémentaires.

Variables actives et variables supplémentaires quantitatives



18

5. Conclusion

6. Bibliographie

[1] <http://www.letudiant.fr/palmares/palmares-des-ecoles-d-ingenieurs/telecombretagne-brest.html>

[2] <http://www.sthda.com/french/articles/38-methodes-des-composantes-principales-dans-r-guide-pratique/75-acm-analyse-des-correspondances-multiples-avec-r-l-essentiel/>

[3] <http://www.sthda.com/french/wiki/test-de-student-est-il-toujours-correct-de-comparer-des-moyennes>

7. Annexes

7.1 Organisation du questionnaire

7.2 Analyse intégrale ACM

OUR WORLDWIDE PARTNERS UNIVERSITIES - DOUBLE DEGREE AGREEMENTS

3 CAMPUS, 1 SITE



IMT Atlantique Bretagne-Pays de la Loire – <http://www.imt-atlantique.fr/>

Campus de Brest

Technopôle Brest-Iroise
CS 83818
29238 Brest Cedex 3
France
T +33 (0)2 29 00 11 11
F +33 (0)2 29 00 10 00

Campus de Nantes

4, rue Alfred Kastler
CS 20722
44307 Nantes Cedex 3
France
T +33 (0)2 51 85 81 00
F +33 (0)2 99 12 70 08

Campus de Rennes

2, rue de la Châtaigneraie
CS 17607
35576 Cesson Sévigné Cedex
France
T +33 (0)2 99 12 70 00
F +33 (0)2 51 85 81 99

Site de Toulouse

10, avenue Édouard Belin
BP 44004
31028 Toulouse Cedex 04
France
T +33 (0)5 61 33 83 65



IMT Atlantique

Bretagne-Pays de la Loire
École Mines-Télécom

© IMT Atlantique, Année
Imprimé à IMT Atlantique
Dépôt légal : Mois Année
ISSN 2556-5060