CY CERGY PARIS UNIVERSITÉ

DU DATA ANALYST

UE INTRODUCTION AUX STATISTIQUES

Devoir final

QUENTIN FOUCHÉ

8 juillet 2022

Table des matières

Liste des figures

Liste des tableaux

1 Description du jeu de données

1.1 Présentation des variables analysées

Le jeu de données analysé dans ce rapport regroupe des réponses à un questionnaire ainsi que des données de logs pour chaque apprenant inscrit à l'une des trois itérations du MOOC Effectuation. Ce jeu de données comprend 17 411 individus et 82 variables. Les sept variables prises en compte dans les analyses sont l'identifiant de l'apprenant, son genre (homme ou femme), l'indice de développement humain de son pays d'origine (IDH), son type (détaillé dans la section 1.2), sa réussite à l'examen (succès ou échec), le nombre total de vidéos qu'il a visionnées (de 0 à 30) et le nombre total de quiz qu'il a réalisés (de 0 à 5) par itération du MOOC (Tableau 1). L'IDH se décline en trois catégories : bas, intermédiaire et très élevé.

Tableau 1. Description des variables analysées. Le nombre de valeurs exclut les données manquantes, exprimées en pourcentage. "ID" : identifiant ; "IDH" : Indice de Développement Humain.

Variable	Type	Nombre de valeurs	Données manquantes (%)
ID apprenant	texte	17411	0
Genre	logique	9099	48
IDH	texte	8969	48
Type d'apprenant	texte	17411	0
Réussite à l'examen	logique	15646	10
Nombre de vidéos vues	numérique	15646	10
Nombre de quiz réalisés	numérique	15646	10

1.2 Classification du type d'apprenant en quatre catégories

Les apprenants ont été classés selon leur type d'utilisation et leur niveau d'investissement dans le MOOC. Quatre catégories ont été distinguées : (1) les "completer" ont passé et réussi l'examen final, (2) les "disengaging" ont réalisé au moins un quiz ou un devoir mais n'ont pas réussi l'examen, (3) les "auditing" n'ont réalisé aucun quiz ou devoir mais ont visionné au moins six vidéos, et (4) les "bystander" n'ont réalisé ni quiz ni devoir et ont visionné moins de six vidéos. Cette classification a permis de différencier 15 646 apprenants. Les 1 765 restants n'ont pas pu être classés pour cause d'information manquante sur leur activité dans le MOOC.

Le Tableau 2 et la Figure 1 présentent le pourcentage d'apprenants en fonction de leur type et de l'itération du MOOC, parmi les apprenants qui ont pu être classés. Le pourcentage d'apprenants de type "auditing" reste très faible comparé aux autres types, quelle que soit l'itération (moins

de 3%). La proportion de completer est quasi-nulle lors de la première itération mais passe à environ 22% lors des itérations suivantes. Cette augmentation est conjointe à la diminution de la proportion de disengaging, qui passe de 58% lors de l'itération 1 à 29% et 25% lors des itérations 2 et 3. Enfin, le pourcentage de bystander reste élevé quelle que soit l'itération du MOOC (entre 39% et 51%).

Tableau 2. Pourcentage d'apprenants en fonction de leur type et de l'itération du MOOC Effectuation (N : nombre d'apprenants par itération).

Type d'apprenant	**Itération n°1** N = 7965	**Itération n°2** N = 3798	**Itération n°3** N = 3883
Completer	0.3	23.1	21.7
Disengaging	58.4	28.8	24.5
Auditing	1.9	2.8	2.8
Bystander	39.4	45.3	51.0

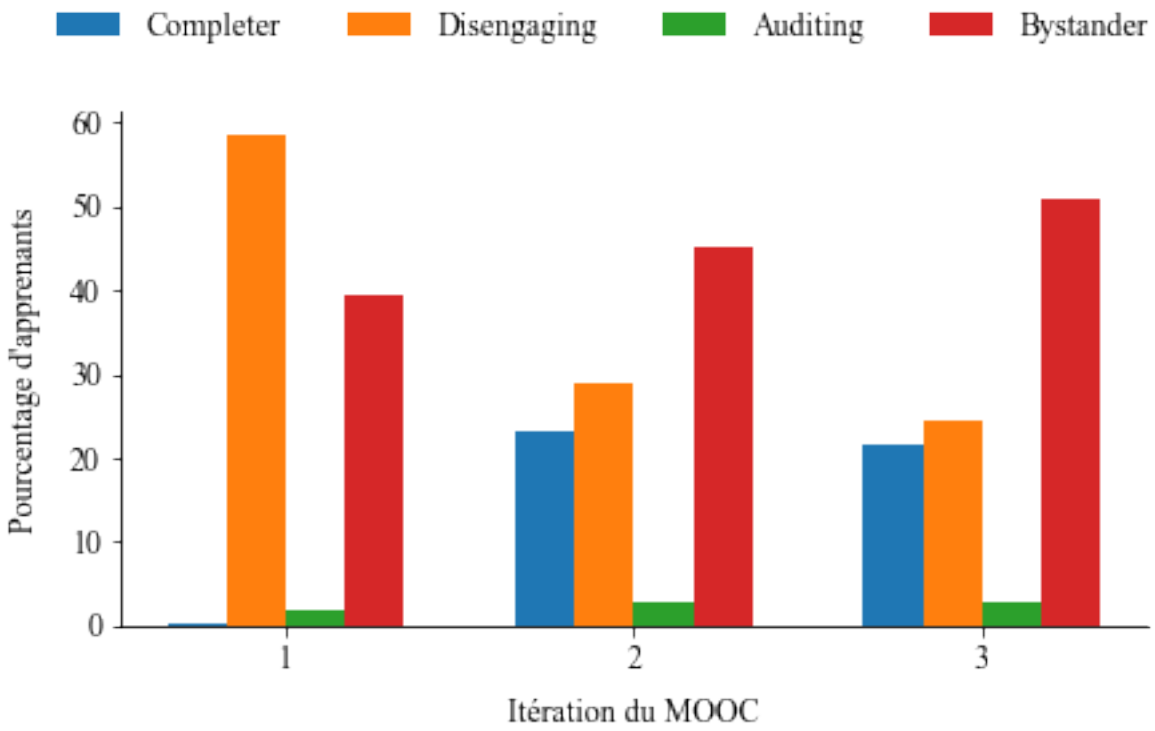

Figure 1. Pourcentage d'apprenants en fonction de leur type et de l'itération du MOOC Effectuation. Les couleurs représentent le type d'apprenant : "Completer" en bleu, "Disengaging" en orange, "Auditing" en vert et "Bystander" en rouge. Les effectifs sont de 7965 apprenants pour l'itération 1, 3798 pour l'itération 2 et 3883 pour l'itération 3.

2 Chi2 et mosaic plot

Le Tableau 3 représente le nombre d'apprenants en fonction de l'IDH du pays d'origine et du genre, sur l'ensemble des trois itérations du MOOC. Ces deux variables, l'IDH et le genre, sont significativement dépendantes : la répartition des apprenants dans chaque catégorie d'IDH diffère significativement entre les hommes et les femmes (test de chi2 : $X^2 = 179$, ddl = 2, $P < 0.001$).

Tableau 3. Nombre d'apprenants en fonction de l'IDH et du genre, sur l'ensemble des trois itérations du MOOC Effectuation.

	IDH		
	Bas	**Intermédiaire**	**Très élevé**
Hommes	883	432	4716
Femmes	147	233	2546

Les résidus de ce test de chi2 sont représentés dans la Figure 2, qui affiche en rouge les valeurs positives et en bleu les valeurs négatives. Les résidus positifs indiquent une sur-représentation des apprenants dans une catégorie d'IDH chez un genre par rapport à l'autre. A l'inverse, les résidus négatifs indiquent une sous-représentation. Si le résidu d'une catégorie d'IDH est positif pour l'un des deux genres, il sera négatif pour l'autre genre, puisqu'une sur-représentation de l'un des genres indique nécessairement une sous-représentation du second. Cela explique que les blocs de la Figure 2 ne peuvent pas être tous rouges ou tous bleus.

La Figure 2 montre que chez les hommes, la proportion d'apprenants originaires d'un pays ayant un IDH bas est plus élevée que ce qui est observé chez les femmes. Inversement, la proportion d'apprenants venant d'un pays à IDH très élevé est plus élevée chez les femmes que chez les hommes. Cette sous-représentation des femmes chez les apprenants venant d'un pays à IDH bas pourrait être dû à une plus grande difficulté d'accès à du matériel informatique (indispensable pour pouvoir suivre un MOOC) pour les femmes dans ces pays. Elle pourrait aussi refléter une différence plus générale d'accès à l'éducation entre les hommes et les femmes, liée à la situation politique dans ces pays.

3 Modèle linéaire, tests non paramétriques

3.1 Tests de Mann-Whitney

Le nombre de vidéos visionnées par apprenant sur une itération du MOOC est représenté dans la Figure 3 en fonction du genre. Ce nombre est significativement plus élevé chez les femmes que chez les hommes (test de Mann-Whitney : U = 9536047, $P < 0.001$).

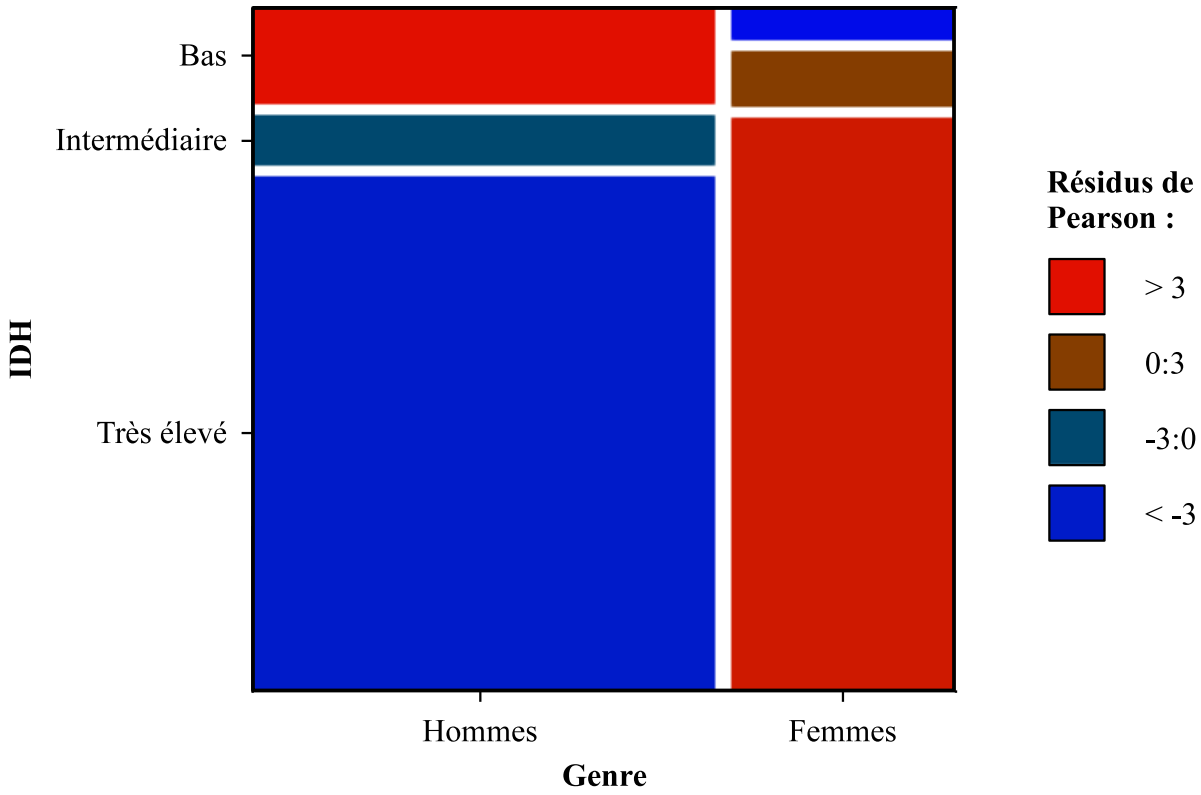

Figure 2. Valeur des résidus du test de chi2 sur l'indépendance entre l'IDH et le genre vis-à-vis du nombre d'apprenants. La hauteur des blocs représente la proportion rela-tive d'apprenants dans chaque catégorie d'IDH, la largeur représente la proportion relative d'hommes et de femmes. Le gradient de couleur indique la valeur des résidus (rouge : valeurs positives ; bleu : valeurs négatives).

Cette différence entre les hommes et les femmes suggère qu'en moyenne les femmes suivent plus assidûment le MOOC que les hommes. Cette interprétation est toutefois à nuancer avec la faible amplitude de la différence : le nombre médian comme le nombre moyen de vidéos vues ne diffèrent que d'une unité entre les deux genres (femmes vs hommes : 12 vs 11 pour la médiane, 14.5 vs 13.5 pour la moyenne).

3.2 Régression linéaire

Le nombre de quiz réalisés par apprenant est significativement corrélé au nombre de vidéos visionnées (Test de corrélation de Spearman : $\rho = 0.80$, $P < 0.001$; Figure 4). Cette corrélation est positive : plus le nombre de vidéos vues est grand, plus le nombre de quiz réalisés est élevé.

L'utilisation d'une régression linéaire pour tester le lien entre ces deux variables est cependant inappropriée. Il s'agit en effet de données de comptage, i.e. des variables quantitatives discrètes, pour lesquelles le modèle d'analyse correct est celui d'une régression de Poisson.

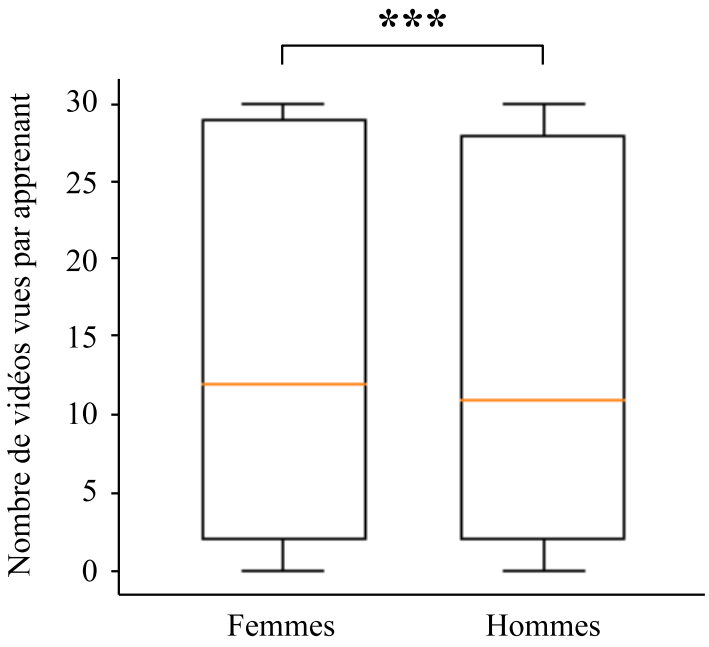

Figure 3. Nombre de vidéos visionnées par apprenant sur une itération du MOOC Effectuation, en fonction du genre. Les boîtes à moustache représentent, du bas vers le haut, la valeur minimale, le premier quartile, la médiane (en orange), le troisième quartile et la valeur maximale. Les effectifs sont de 2990 femmes et 6103 hommes. Test de Mann-Whitney : *** $P < 0.001$.

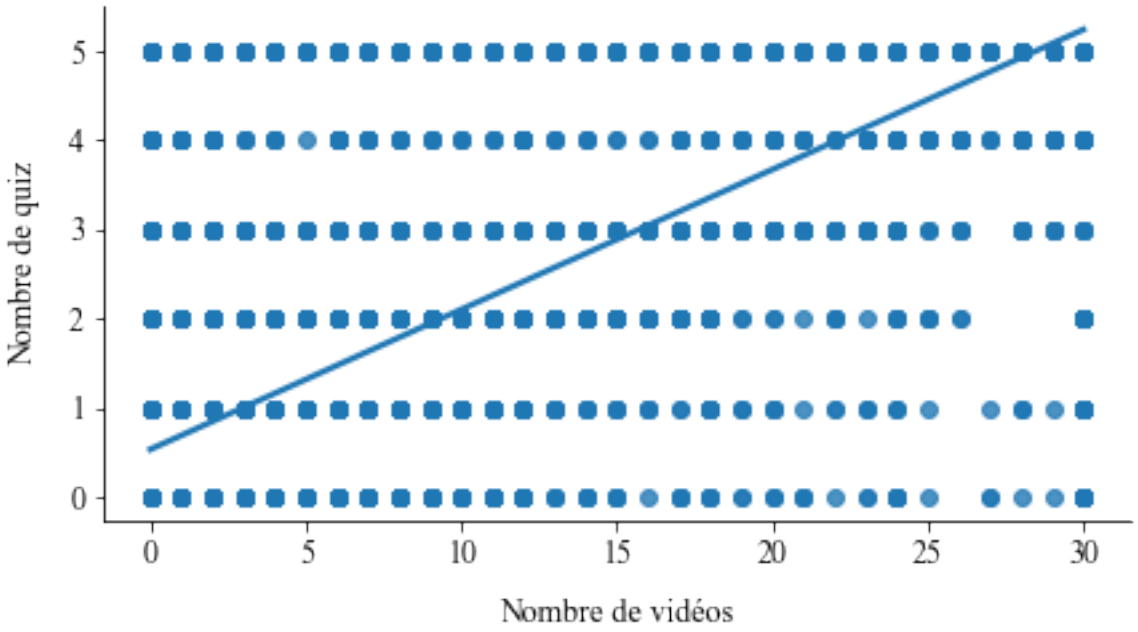

Figure 4. Nombre de quiz réalisés en fonction du nombre de vidéos visionnées par apprenant. La ligne bleue représente la droite de régression ($R^2 = 0.65$).

3.3 ANOVA

3.3.1 Effet de l'IDH et du genre sur le nombre de vidéos vues, sans prise en compte de l'interaction

La Figure 5 représente le nombre moyen de vidéos vues par apprenant en fonction du genre et de l'IDH du pays d'origine. D'après ce graphique, il apparaît que les apprenants venant d'un pays à IDH très élevé ont regardé environ 1.5 plus de vidéos que ceux venant d'un pays à IDH intermédiaire, ces derniers ayant eux-mêmes regardé 1.5 à 2 fois plus de vidéos que les apprenants issus d'un pays à IDH bas. Afin de tester l'effet de ces deux variables (genre et IDH) sur le nombre de vidéos vues, une ANOVA à deux facteurs a été réalisée, en ne considérant dans un premier temps que l'effet des variables prises séparément (i.e. sans interaction).

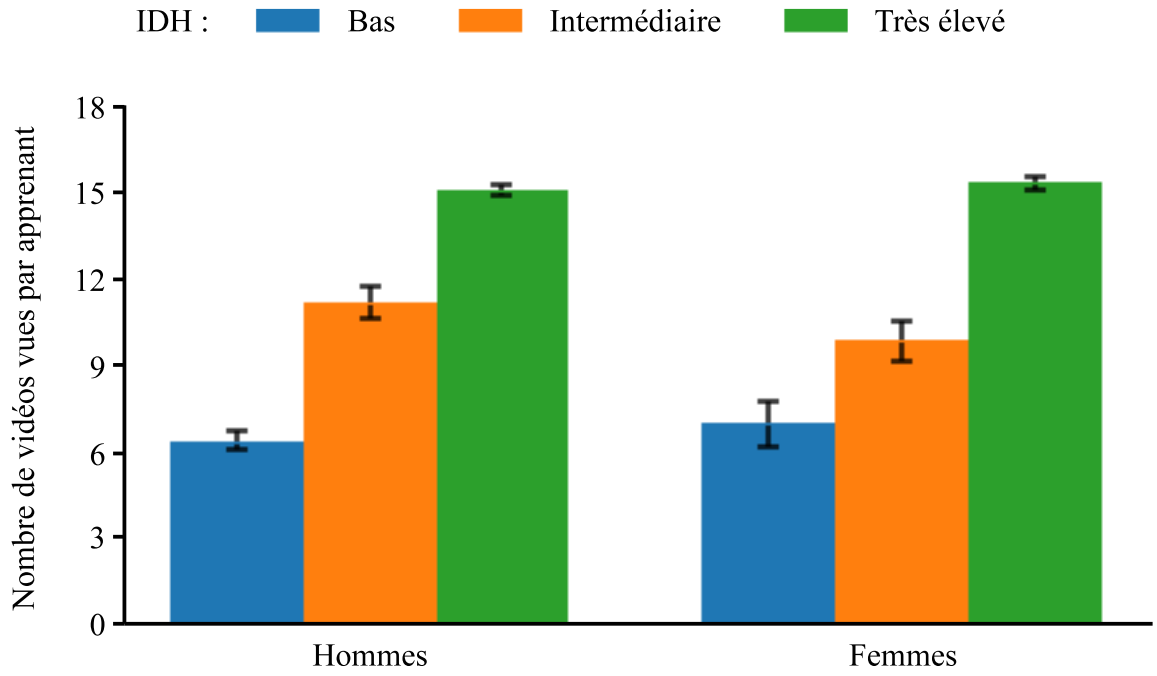

Figure 5. Nombre moyen de vidéos visionnées par apprenant en fonction du genre et de l'IDH ("Bas" en bleu, "Intermédiaire" en orange et "Très élevé" en vert). Les barres d'erreurs représentent l'erreur standard à la moyenne. Les effectifs de chaque catégorie sont donnés dans le Tableau 3.

Le Tableau 4 résume les résultats de cette ANOVA. Il en ressort que l'IDH a un effet significatif sur le nombre de vidéos vues ($F(2,8947) = 284$, $P < 0.001$), contrairement au genre ($F(1,8947) = 4.41$, $P = 0.51$). Le Tableau 5 donne un aperçu des effets associés à la modalité des apprenants hommes venant d'un pays à IDH bas. D'après les valeurs des coefficients, les apprenants venant d'un pays à IDH intermédiaire regardent en moyenne quatre vidéos de plus que ceux venant d'un pays à IDH bas, et environ neuf de plus pour les apprenants issus d'un pays à IDH très élevé.

Tableau 4. Table d'ANOVA testant l'effet de l'IDH et du genre sur le nombre de vidéos vues par apprenant : description des paramètres. L'interaction entre l'IDH et le genre n'est pas prise en compte. "ddl" = degrés de liberté.

	Somme des carrés	**ddl**	**F**	***P*(>F)**
IDH	7.42e+4	2	2.84e+2	1.40e-120
Genre	5.74e+1	1	4.41e-1	5.07e-1
Résidus	1.17e+6	8947		

Tableau 5. Table d'ANOVA testant l'effet de l'IDH et du genre sur le nombre de vidéos vues par apprenant : estimation des effets associés à la modalité "hommes issus de pays à IDH bas". L'interaction entre l'IDH et le genre n'est pas prise en compte.

	Coefficient	**Erreur standard**	**t**	***P*(>|t|)**
Intercept	6.43	0.36	17.99	< 0.001
IDH Intermédiaire	4.25	0.57	7.45	< 0.001
IDH Très élevé	8.71	0.38	22.69	< 0.001
Femmes	0.17	0.26	0.66	0.51

Les degrés de liberté (ddl) associés à la variance inter-groupe sont de deux pour l'IDH et de un pour le genre (Tableau 4). Ces valeurs sont obtenues selon la formule suivante, où k est le nombre de groupes comparés :

$$ddl = k - 1$$

Pour l'IDH, trois modalités sont comparées (IDH bas , intermédiaire et très élevé), d'où un ddl de deux, et seulement deux modalités sont comparées pour le genre (hommes et femmes), d'où un ddl de un.

3.3.2 Prise en compte de l'interaction entre IDH et genre

Si le nombre de vidéos vues par apprenant n'est pas significativement affecté par la seule variable genre, il est possible qu'il soit influencé par l'interaction entre le genre et l'IDH. Autrement dit, l'effet de l'IDH sur le nombre de vidéos vues pourrait varier en fonction du genre. Pour étudier cette hypothèse, l'ANOVA a été réalisée à nouveau en prenant en compte l'interaction entre IDH et genre.

Les Tableaux 6 et 7 présentent les résultats de cette nouvelle ANOVA. D'après le Tableau 6, l'interaction entre l'IDH et le genre n'a pas d'effet significatif sur le nombre de vidéos vues par

apprenant (F(2,8945) = 1.5, $P = 0.22$). Le Tableau 7 confirme cette observation : ni la modalité incluant les femmes venant de pays à IDH intermédiaire, ni celle des femmes venant de pays à IDH très élevé, ne diffèrent significativement de la modalité "hommes venant d'un pays à IDH bas" ($P = 0.16$ et 0.77, respectivement).

Tableau 6. Table d'ANOVA testant l'effet de l'IDH et du genre sur le nombre de vidéos vues par apprenant, interaction prise en compte : description des paramètres. "ddl" = degrés de liberté.

	Somme des carrés	**ddl**	**F**	***P*(>F)**
IDH	7.42e+4	2	2.85e+2	1.36e-120
Genre	5.74e+1	1	4.41e-1	5.07e-1
IDH : Genre	3.90e+2	2	1.50	2.24e-1
Résidus	1.16e+6	8945		

Tableau 7. Table d'ANOVA testant l'effet de l'IDH et du genre sur le nombre de vidéos vues par apprenant, interaction prise en compte : estimation des effets associés à la modalité "hommes issus de pays à IDH bas".

	Coefficient	**Erreur standard**	**t**	***P*(>|t|)**
Intercept	6.37	0.38	16.60	< 0.001
IDH Intermédiaire	4.84	0.67	7.22	< 0.001
IDH Très élevé	8.73	0.42	20.86	< 0.001
Femmes	0.59	1.02	0.58	0.57
IDH Intermédiaire : Femmes	-1.93	1.38	-1.40	0.16
IDH Très élevé : Femmes	-0.31	1.06	-0.29	0.77

Les résultats de cette ANOVA ne permettent donc pas de mettre en évidence une différence de comportement entre hommes et femmes dans le visionnage des vidéos du MOOC Effectuation. Ce résultat est contraire à celui du test de Mann-Whitney réalisé en section 3.1, qui montrait un nombre de vidéos vues légèrement mais significativement plus élevé chez les femmes que chez les hommes. Les résultats de l'ANOVA sont toutefois plus robustes, car ils rendent visible l'effet de l'IDH. Cet effet a pu apporter un biais dans le résultat du test de Mann-Whitney car les effectifs des apprenants par catégorie d'IDH ne sont pas les mêmes selon le genre. Comme l'ont montré les résidus du test de Chi2 réalisés en section 2, les femmes venant d'un pays à IDH bas sont sous-représentées par rapport aux hommes de cette catégorie, et celles venant de pays à IDH très élevé sont sur-représentées. Ces différences de représentation ont vraisemblablement biaisé le nombre moyen de vidéos vues par les femmes vers une valeur plus élevée que celle qui aurait été observée, si les effectifs des femmes avaient été identiques à ceux des hommes dans chaque catégorie d'IDH.

Les différences de nombre de vidéos vues selon l'IDH suggèrent un niveau d'implication plus élevé chez les apprenants issus des pays riches par rapport aux pays pauvres. Elles pourraient toutefois simplement signifier une différence de pratique. Par exemple, les apprenants des pays pauvres réalisent peut-être davantage de quiz ou investissent plus de temps dans les devoirs que les apprenants des pays riches, des hypothèses qui pourraient être testées directement avec le présent jeu de données. Par ailleurs, les apprenants des pays à IDH bas ont peut-être une connexion Internet de moins bonne qualité en moyenne, rendant plus compliqué la lecture des vidéos et incitant plutôt à lire directement leur retranscription écrite. La barrière de la langue pourrait aussi avoir une influence, en particulier si la part d'apprenants non-francophones est plus grande dans les pays pauvres.

4 Régression logistique

4.1 Présenter des odd-ratios

Dans cette section, l'effet des variables socio-démographiques (genre et IDH) a été étudié sur le résultat de l'examen final, une variable booléenne indiquant soit un succès soit un échec. Une régression logistique a été appliquée, en prenant en compte l'ensemble des données des trois itérations du MOOC. L'interaction entre le genre et l'IDH n'a pas été étudiée. Les odds-ratios issus de cette régression sont présentés dans le Tableau 8 et la Figure 6.

Tableau 8. Table d'odds-ratios : régression logistique entre la réalisation de l'examen final et le genre et l'IDH. L'interaction entre le genre et l'IDH n'est pas prise en compte dans la régression. "Réf." indique la modalité de référence.

	Odd-ratios	***P*(>|z|)**	**2.5%**	**97.5%**
Hommes	Réf.			
Femmes	1.12	4.72e-2*	1.00	1.26
IDH Bas	Réf.			
IDH Intermédiaire	1.12	4.16e-1	0.85	1.47
IDH Très élevé	1.37	7.86e-4***	1.14	1.65

Les résultats indiquent que les femmes ont davantage réussi l'examen que les hommes (OR = 1.12, $P = 0.047$). Concernant l'IDH, les apprenants issus des pays à indice très élevé ont eu plus de succès que ceux des pays à indice faible (OR = 1.37, $P < 0.001$). Aucune différence significative n'a été observée entre les apprenants venant de pays à IDH intérmédiaire par rapport à ceux venant de pays à IDH bas (OR = 1.12, $P = 0.42$; Tableau 8, Figure 6).

Ces résultats diffèrent de ceux de l'ANOVA précédente, qui testait l'effet des mêmes variables socio-démographiques (genre et IDH) sur le nombre de vidéos vues (Tableaux 4 et 5). Une

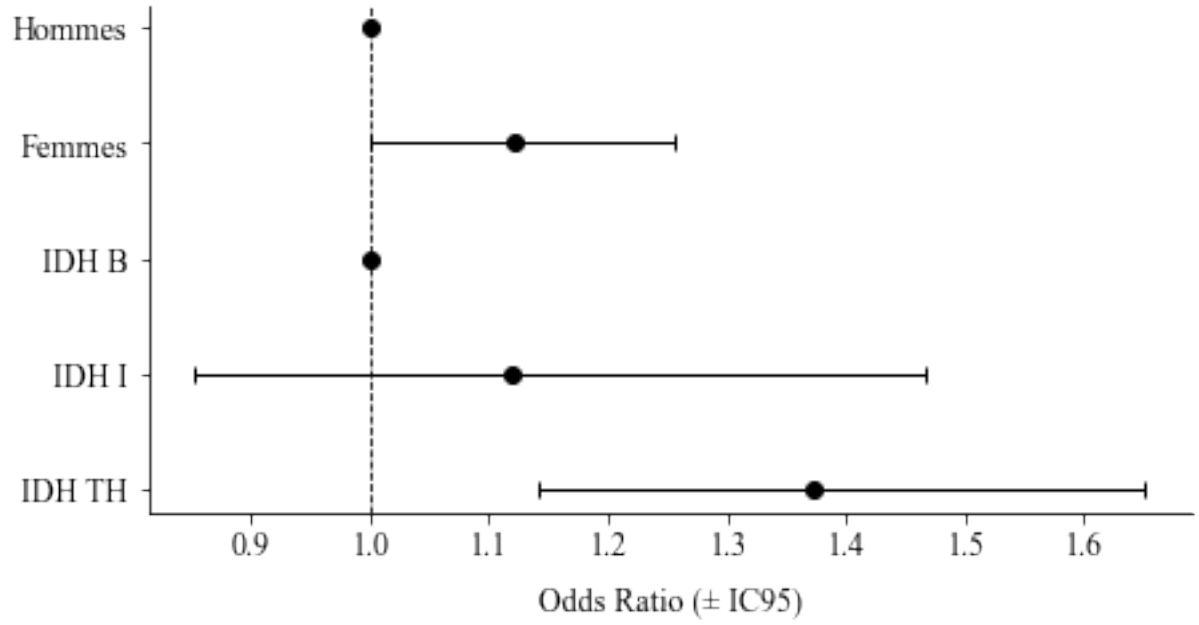

Figure 6. Odds-ratios (± IC95) issus de la régression logistique entre la réalisation de l'examen final et le genre et l'IDH ("B" : bas ; "I" : intermédiaire ; "TH" : très élevé). Les modalités "Hommes" et "IDH Bas" sont utilisées comme références.

première différence se trouve au niveau du genre : les femmes ont davantage réussi l'examen, mais sans regarder significativement plus de vidéos que les hommes. Cette meilleure réussite à l'examen pourrait être liée à d'autres facteurs, tels que l'investissement dans la réalisation des quiz ou la participation aux discussions dans le forum. La seconde différence concerne l'IDH : les apprenants venant des pays à IDH intermédiaire ont vu davantage de vidéos, mais sans avoir plus de succès que les ceux des pays à IDH bas. Un lien positif entre nombre de vidéos vues et réussite à l'examen est cependant observé chez les apprenants des pays riches par rapport à ceux des pays pauvres. L'assiduité dans le suivi des vidéos pourrait donc jouer un rôle dans la réussite à l'examen, mais sans être nécessaire, les apprenants préférant peut-être d'autres voies d'apprentissage en fonction de leur organisation et du temps qu'ils peuvent allouer au MOOC.

Les odds-ratios présentés dans cette analyse donnent une information sur la différence de réussite à l'examen entre la modalité de référence et les autres, mais ils ne constituent pas pour autant des risques relatifs, i.e. des parts de chances de réussir davantage l'examen selon qu'on se situe dans une catégorie plutôt qu'une autre. Ils peuvent s'en rapprocher lorsque l'évènement considéré est suffisamment rare, ce qui n'est pas le cas ici (la proportion d'apprenants ayant réussi l'examen, sur l'ensemble des trois itérations du MOOC, étant de 11%).

4.2 Données de comptage et loi de Poisson

Cette section revient sur l'analyse du nombre de vidéos vues par apprenant en fonction du genre et de l'IDH. La première analyse vise à évaluer l'adéquation du modèle linéaire créé dans la section 3.3, utilisé pour étudier le lien entre ces variables. Ce modèle ne peut être utilisé que si la distribution du nombre de vidéos vues suit une loi normale. Cette distribution est représentée

dans la Figure 7 : loin de suivre une loi normale, elle semble davantage se rapprocher d'une loi de Poisson, avec un nombre très élevé de valeurs proches de zéro.

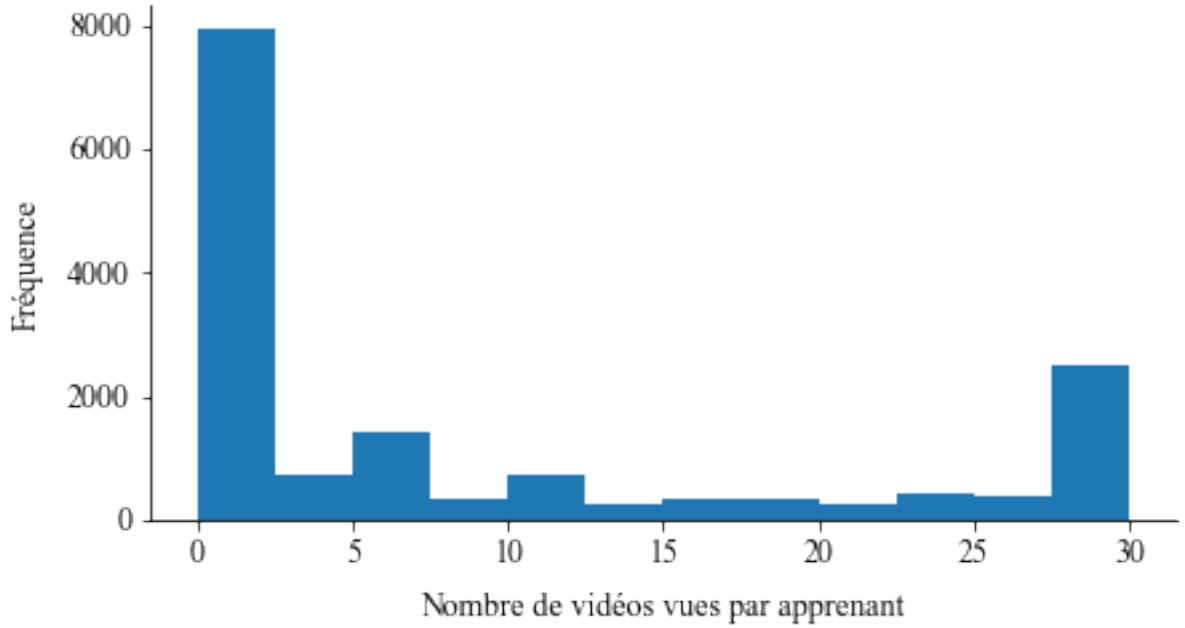

Figure 7. Distribution du nombre de vidéos vues par apprenant.

La Figure 8 présente quatre graphiques qui permettent d'analyser plus en détail la normalité de la distribution du nombre de vidéos vues, ainsi que l'homoscédasticité (i.e. l'égalité des variances) des résidus du modèle linéaire. Le premier graphique présente les résidus du modèle de régression en fonction des valeurs ajustées (Figure 8A). Si la variance des résidus est égale entre les groupes d'apprenants, les résidus devraient été dispersés de manière égale dans chaque groupe, ce qui est effectivement observé dans ce graphe. La condition d'homoscédasticité semble donc respectée dans ce modèle. Toutefois, les résidus sont distribués majoritairement au-dessus de la droite "y = 0" pour les faibles valeurs ajustées, ce qui suggère un écart à la normalité. En effet, si le nombre de vidéos vues était normalement distribuée, la dispersion des résidus devraient être homogène de part et d'autre de la droite, ce qui n'est pas le cas.

Le second graphique représente les quantiles observés en fonction des quantiles théoriques estimés par le modèle (Figure 8B). Si le nombre de vidéos vues était distribué normalement, les quantiles observés et théoriques devraient avoir quasiment les mêmes valeurs et donc être très proches à la droite "y = x". Or ici, les quantiles observés sont supérieurs aux quantiles théoriques pour les faibles valeurs, et inversement pour les valeurs plus élevées. Ces observations confirment que le nombre de vidéos vues n'est pas normalement distribué. Les valeurs basses sont sous-estimées par le modèle, tandis que les valeurs élevées sont surestimées.

Le troisième graphique, qui représente la racine carrée de la valeur absolue des résidus standardisés en fonction des valeurs ajustées, apporte une information similaire aux deux précédents : en condition de normalité, la taille des résidus standardisés aurait dû être indépendante des valeurs ajustées, alors qu'elle tend ici à diminuer quand les valeurs augmentent (Figure 8C).

Enfin, le quatrième graphique représente les résidus standardisés en fonction du leverage (terme anglais signifiant "levier"), qui donne une mesure de l'influence de chaque valeur des variables explicatives (ici, chaque catégorie socio-démographique) sur les coefficients de régression du modèle (Figure 8D). Plus le leverage est grand, plus l'influence sur les coefficients est grande, donc plus ceux-ci seraient modifiés si la catégorie socio-démographique correspondante était supprimée. Ici, le leverage le plus élevé est associé aux femmes venant de pays à IDH bas.

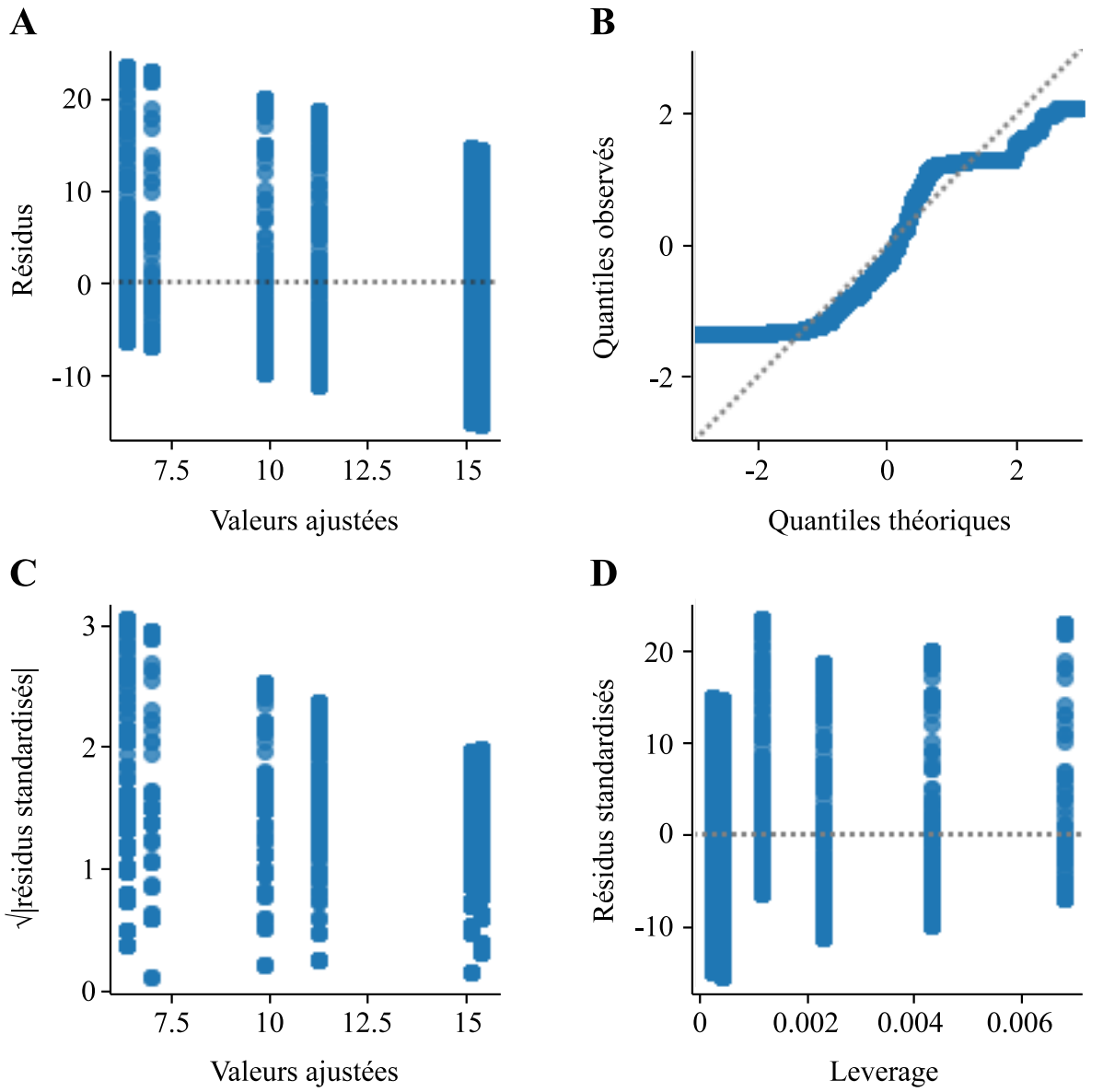

Figure 8. Diagnostic du modèle linéaire sur le nombre de vidéos vues par apprenant en fonction du genre et de l'IDH. **A.** Résidus en fonction des valeurs ajustées. La ligne en pointillé représente la droite "y = 0". **B.** Quantiles observés en fonction des quantiles théoriques. La ligne en pointillé représente la droite "y = x". **C.** Racine carrée de la valeur absolue des résidus standardisés en fonction des valeurs ajustées. **D.** Résidus standardisés en fonction du leverage. La ligne en pointillé représente la droite "y = 0".

L'analyse de la Figure 8 permet de conclure que le nombre de vidéos vues n'est pas normalement distribué, et donc que le modèle linéaire n'est pas adapté pour étudié l'effet du genre et de l'IDH sur cette variable. Un modèle plus adéquat est celui de la régression de Poisson. L'analyse du nombre de vidéos vues a été de nouveau réalisée avec ce modèle, dont les résultats sont présentés dans le Tableau 9.

Tableau 9. Régression de Poisson entre le nombre de vidéos vues par apprenant et le genre et l'IDH : estimation des effets associés à la modalité "hommes issus de pays à IDH bas".

	Coefficient	**Erreur standard**	**z**	***P*(>|z|)**
Intercept	1.85	1.3e-2	138.95	< 0.001
IDH Intermédiaire	0.56	2.0e-2	28.82	< 0.001
IDH Très élevé	0.86	1.4e-2	62.31	< 0.001
Femmes	0.09	3.4e-2	2.59	0.010
IDH Intermédiaire : Femmes	-0.22	4.2e-2	-5.09	< 0.001
IDH Très élevé : Femmes	-0.07	3.5e-2	-2.01	0.044

La statistique du chi2 de Pearson associée au modèle est de 86 200. Cette valeur est bien au-dessus de un, ce qui montre une faible adéquation du modèle aux données, malgré l'utilisation de la loi de Poisson. Ce modèle met néanmoins en évidence un effet significatif du genre, de l'IDH, ainsi que de l'interaction entre les deux (Tableau 9). Chez les hommes, le nombre de vidéos vues est significativement plus élevé lorsque les apprenant viennent d'un pays à IDH intermédiaire (coefficient = 0.56, $P < 0.001$) ou élevé (coefficient = 0.86, $P < 0.001$) plutôt que d'un pays à IDH bas. Par rapport aux hommes, les femmes ont regardé significativement plus de vidéos (coefficient = 0.09, $P = 0.01$). Par ailleurs, l'influence de l'IDH sur le nombre de vidéos diffère selon le genre : l'augmentation du nombre de vidéos vues est moins marquée chez les femmes venant de pays à IDH intermédiaire (coefficient = -0.22, $P < 0.001$) ou très élevé (coefficient = -0.07, $P = 0.044$) par rapport à celles venant de pays à faible IDH (Tableau 9, Figure 5).

Ces résultats amènent à modifier les conclusions de l'ANOVA réalisée en section 3.3.2, qui n'avait mis en évidence ni d'effet du genre, ni d'effet de l'interaction entre genre et IDH (Tableau 7). Le nombre plus élevé de vidéos regardées chez les femmes suggère des différences de pratique suivant le genre, ou traduit peut-être un niveau de motivation supérieur chez ces dernières. Étant donné l'effectif inférieur des femmes par rapport aux hommes parmi les apprenants, en particulier le faible nombre de femmes issues de pays à faible IDH, il se pourrait que les difficultés d'accès que rencontreraient les femmes dans les pays les moins développés les incitent, lorsqu'elles sont surmontées, à s'investir davantage dans le suivi du MOOC. L'analyse des réponses aux questionnaires pourrait apporter des indices supplémentaires sur les motivations des apprenants, ce qui permettrait de confirmer ou non cette hypothèse.