

GENESE

AGENDA

IMPACT DU
GENRE

HASARD

QUESTIONS

DONNEES
ET
VARIABLES

MODELES

RESULTATS

AUTRES
CHOIX

CONCLUSION

Biais de genre dans les évaluations comportant des questions fermées avec points négatifs

Catherine Dehon

Université libre de Bruxelles, SBS-EM, ECARES

22 novembre 2023

GENESE DE CETTE RECHERCHE

- Cours de “Probabilité et inférence statistique” avec mon collègue D. Paindaveine : **débat** sur la pertinence des points négatifs dans les examens
- Proposition de **mémoire** sur cette question codirigé avec Ph. Emplit \Rightarrow A. Brogniaux reçoit le prix Philippe Maystadt
- **Publication** basée sur les résultats du mémoire : “Gender bias in closed-ended questions with negative points” par A. Brogniaux, C. Dehon, Ph. Emplit et C. Toma

AGENDA

- Impact du genre dans les évaluations
- Effet du hasard dans les VF/QCM/QRM
- Questions de recherche
- Données et variables
- Modèles économétriques
- Résultats
- Alternatives aux points négatifs
- Recommandations et conclusions

IMPACT DU GENRE DANS LES EVALUATIONS

Littérature sur l'enseignement supérieur : impact du genre à différents niveaux :

- au niveau de la matière
- au niveau du format de la question/de l'examen
- en matière d'aversion au risque

Cette recherche porte uniquement sur le dernier point :

Existe-t-il un biais de genre causé par l'utilisation des points négatifs dans les examens à questions fermées ?

EFFET DU HASARD

Objectif d'une évaluation : vérifier les acquis d'apprentissage (compétences, connaissances). Mais effet du hasard !

Deux types d'évaluation :

- 1 Examen avec 20 (n) questions "Vrai/Faux" ($r = 2$)
- 2 Examen avec 20 (n) QCM avec 5 propositions ($r = 5$)

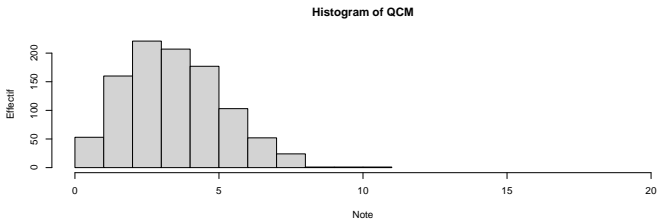
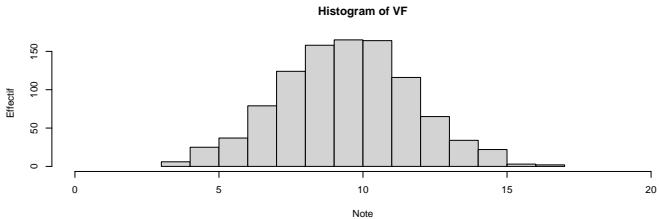
Hypothèse : les étudiants répondent au hasard

- 1 Probabilité réponse correcte pour "Vrai/Faux" : $\frac{1}{2}$ (p)
- 2 Probabilité réponse correcte pour QCM : $\frac{1}{5}$ (p)

Soit X le nombre de réponses correctes. En moyenne, le nombre de réponses correcte est donné par

- 1 $E(X) = np = 20 * \frac{1}{2} = 10$ pour les VF
- 2 $E(X) = np = 20 * \frac{1}{5} = 4$ pour les QCM

Simulations : Notes de 1000 étudiants (1 par réponse correcte)



Responsabilité de certifier les apprentissages n'est absolument pas remplie \Rightarrow introduire une correction pour le hasard

Introduction de points négatifs

20 questions V/F avec pénalisation de 1 point ($\frac{1}{r-1}$) si la réponse est fausse

Hypothèse : les étudiants répondent au hasard aux questions

- Probabilité réponse correcte pour un "Vrai/Faux" : $\frac{1}{2}$ (p)
- En moyenne, l'étudiant répondant au hasard aura 0/20



Possible biais lié à la gestion du risque, donc on n'évalue plus uniquement les acquis d'apprentissage mais également la prise de risque, l'aversion au risque, etc.



Pour éviter cela, UGent (2017) et UCLouvain (2020) ont supprimé les points négatifs en proposant d'autres alternatives.

QUESTIONS DE RECHERCHE

Le comportement par rapport à l'abstention des étudiants et étudiantes est-il différent dans le contexte des VF et QCM à points négatifs pour un cours de mathématique ?

Si oui, comment quantifier :

- 1 Probabilité d'abstention et facteurs d'abstention
- 2 Perte moyenne de points attendue pour les étudiantes par rapport aux étudiants

DONNEES ET VARIABLES

Données : 11 examens de MATH-S102 « Mathématiques pour la Gestion : Analyse et Algèbre Linéaire » en BA1 INGE (sessions 2017-18, 2018-19 et janvier 2019-20)

Acad. year	Period	Q1/Q2	Number students	Female students
2017-18	January	Q1	573	39%
	June	Q1	119	45%
	June	Q2	397	39%
	August	Q1	253	44%
	August	Q2	281	42%
2018-19	January	Q1	515	40%
	June	Q1	121	47%
	June	Q2	347	41%
	August	Q1	218	42%
	August	Q2	209	42%
2019-20	January	Q1	653	38%

Composition de l'examen :

- Question ouverte théorie (25%)
- Question ouverte exercice (25%)
- 7 VF : +2 si réponse correcte -1 si incorrecte (16,7%)
- 7 QCM : +4 si réponse correcte -1 si incorrecte (33,3%)

Variables

- Type de question (QCM vs VF)
- Niveau de difficulté de la question
- Moyenne de PAE
- Session d'examen

Mean score Jan. 19-20	TFAs			MCQs		
	Easy	Med.	Diff.	Easy	Med.	Diff.
Female students	0.98	0.67	0.33	2.06	1.20	0.61
Male students	1.01	0.78	0.35	2.00	1.19	0.70

MODELES ECONOMETRIQUES

Utilité du sujet i ($i = 1, \dots, n$) pour chaque choix j (Abstention ($ABST$), Correct answer (CA), Wrong answer (WA)) :

$$U_{i,j} = V_{i,j} + \varepsilon_{i,j},$$

où $V_{i,j} = \alpha_j + X_j' \beta + Z_i' \gamma$ est mesurée sur base de facteurs connus X et Z et $\varepsilon_{i,j}$ est le terme d'erreur.

Probabilité $P_{i,j}$ qu'un étudiant i choisisse l'alternative j :

$$\begin{aligned} P_{i,j} &= P(U_{i,j} > U_{i,k} && \forall k = 1, \dots, J \text{ et } k \neq j) \\ &= P(\varepsilon_{i,k} - \varepsilon_{i,j} \leq V_{i,j} - V_{i,k} && \forall k = 1, \dots, J \text{ et } k \neq j). \end{aligned}$$

Sous l'hypothèse que $\varepsilon_{i,1}, \dots, \varepsilon_{i,J}$ sont i.i.d selon une loi Gumbel à valeur extrême de type 1 (McFadden, 1974) :

$$P_{i,j} = \frac{e^{V_{i,j}}}{\sum_{k=1}^J e^{V_{i,k}}}.$$

⇒ Modèle multinomial logit basé sur l'hypothèse "independence of irrelevant alternatives" (IIA)

Mais cette hypothèse semble irréaliste dans notre situation où certaines alternatives peuvent être plus similaires que d'autres

GENESE

AGENDA

IMPACT DU
GENRE

HASARD

QUESTIONS

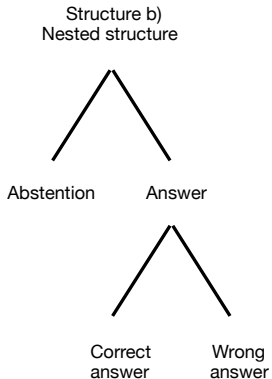
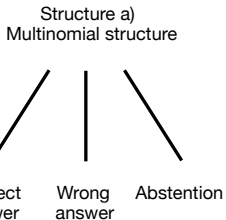
DONNEES
ET
VARIABLES

MODELES

RESULTATS

AUTRES
CHOIX

CONCLUSION



Modèle nested logit qui permet à certaines alternatives de partager une composante commune dans leur terme d'erreur :

$$U_{i,ABST} = V_{i,ABST} + \varepsilon_{i,ABST}$$

$$U_{i,CA} = V_{i,A} + V_{i,CA} + \varepsilon_{i,A} + \varepsilon_{i,CA}$$

$$U_{i,WA} = V_{i,A} + V_{i,WA} + \varepsilon_{i,A} + \varepsilon_{i,WA}$$

Estimation des probabilités (full information maximum-likelihood estimator) :

- $P_{i,ABST}$
- $P_{i,A}$
- $P_{i,CA|A}$
- $P_{i,WA|A}$
- $P_{i,CA} = P_{i,A} * P_{i,CA|A}$
- $P_{i,WA} = P_{i,A} * P_{i,WA|A}$

RESULTATS

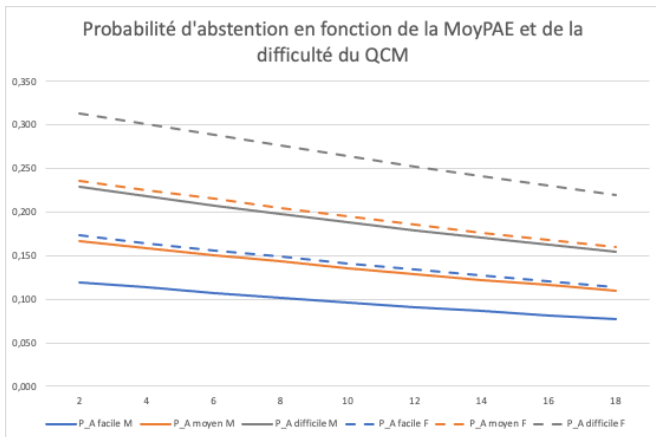
Examen de janvier 2019-20 (Q1)

Estimation du modèle nested logit

Variable	Coeff.	SE	p-value	[95% Conf. Int.]	
Upper-level (ABST, A)					
Gender	0.433	0.096	0.000	[0.246	0.621]
Level of difficulty	0.391	0.029	0.000	[0.334	0.447]
Relative loss	2.025	0.191	0.000	[1.650	2.399]
AvgPAE	-0.030	0.011	0.008	[-0.053	-0.008]
Lower-level (CA, WA)					
Grading	0.144	0.016	0.000	[0.112	0.176]

- Biais de genre : variable “Gender” significative et positive
⇒ une étudiante s’abstient plus qu’un étudiant
- L’abstention augmente avec la difficulté de la question, et diminue en fonction de la “qualité globale” de l’étudiant

Examen de janvier 2019-20 (Q1)



Examens sur les 11 sessions

Estimation du modèle nested logit

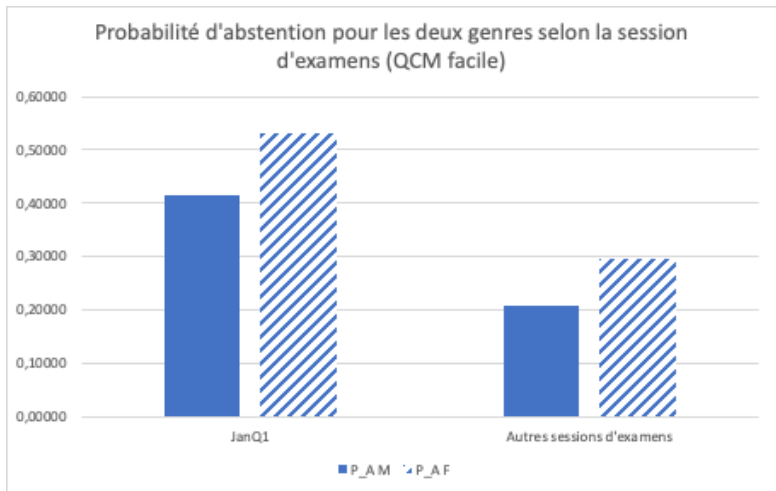
Variable	Coeff.	SE	p-value	[95% Conf. Int.]	
Upper-level (ABST, A)					
Gender	0.471	0.181	0.009	[0.117	0.825]
Level of difficulty	0.101	0.048	0.036	[0.007	0.196]
Relative loss	1.927	0.504	0.000	[0.939	2.915]
JanQ1	1.001	0.481	0.037	[0.058	1.944]
Lower-level (CA, WA)					
Grading	-0.25	0.090	0.005	[-0.426	-0.074]

- Les étudiantes s'abstiennent toujours significativement plus que les étudiants
- La variable "JanQ1" est significative et positive : possibilité d'un processus d'apprentissage

Probabilité d'abstention pour les deux genres selon la session d'examens (niveau facile)

Abstention probability	Session de janvier		Autres sessions	
	QCM	VF	QCM	VF
Female	0.532	0.257	0.295	0.113
Male	0.415	0.178	0.207	0.074

- Probabilités estimées d'abstention plus élevées chez les étudiantes quelque soit le type de question, ou la session
- Probabilités estimées d'abstention plus faibles pour les VF par rapport au QCM
- Probabilités estimées d'abstention pour la session de janvier nettement plus élevées que pour les autres sessions



Estimation de la différence en termes de points

Exemple : QCM facile, janvier avec une perte relative de 25% :

$$E[\textit{points}] = 0 * P_{ABST} + P_A(4 * P_{(CA|A)} - 1 * P_{(WA|A)}).$$

Sur base des modèles économétriques, on obtient les estimations suivantes :

$$\hat{E}[\textit{points}|\textit{Etudiante}] = 0,45\textit{points}$$

$$\hat{E}[\textit{points}|\textit{Etudiant}] = 0,57\textit{points}$$

Estimation de perte pour une étudiante de 0,12 points (par rapport à un étudiant) sur une question qui en vaut 4, soit une **perte moyenne de 3%** des points

La perte de point dépend des différents facteurs : difficulté de la question, type de question, type de session, etc.

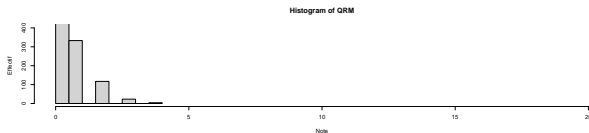
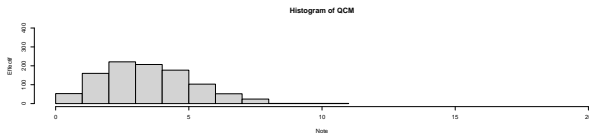
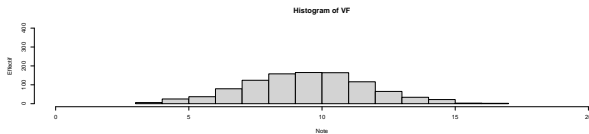
Abstention probability	Male students	Female students	Expected mean loss
JanQ1 - MCQ - Easy question	0.415	0.532	2.93%
JanQ1 - MCQ - Medium-level difficulty	0.440	0.556	2.90%
JanQ1 - MCQ - Difficult question	0.465	0.582	2.93%
JanQ1 - TF - Easy question	0.178	0.257	3.95%
JanQ1 - TF - Medium-level difficulty	0.193	0.277	4.20%
JanQ1 - TF - Difficult question	0.209	0.298	4.45%

⇒ Perte moyenne liée au genre pour la session de janvier entre 2,9% et 4.45%

Dans la littérature, Baldiga (2014) estime la perte moyenne de 2% sur un examen d'entrée à l'université

ALTERNATIVES AUX POINTS NEGATIFS

Utilisation des QRM



Standard setting

Intuition cas simple : 20 Vrai/Faux

Règle : Obtenir au minimum **15 bonnes réponses** sur les 20 pour réussir l'examen

Pourquoi ? En moyenne, 10 réponses correctes à cause du hasard
Pour les 10 questions restantes, répondre correctement à au moins 50% des réponses \Rightarrow 15 réponses correctes est équivalent à la réussite et donc à la note de 10/20

Il existe d'autres alternatives dans la littérature : poids en fonction de la certitude, etc.

CONCLUSION

- Limite : analyse d'un cours et un public spécifique
- Un biais de genre n'est pas systématiquement détecté sur les examens de mathématique analysés.
- Lorsque ce biais de genre existe, les étudiantes s'abstiennent toujours plus que les étudiants
- Dans le cadre de cette recherche estimation de la perte moyenne de 3% à 4,5% des points pour les étudiantes

RECOMMANDATION

- Principe de prudence : essayer de limiter l'utilisation des points négatifs dans les questions fermées
- Utiliser d'autres méthodes pour neutraliser l'effet du hasard
- Mettre en place des tests préparatoires avec un format similaire à l'examen