université angers

FACULTÉ DE DROIT, D'ÉCONOMIE ET DE GESTION
UNIVERSITÉ D'ANGERS

ECONOMETRIE DES VARIABLES QUALITATIVES

Modèle de Comptage

Réalisé par : BARRY Mamadou Yaya
TANNANI Manal
ZYATE Yassine

Professeur : COMPAIRE Philippe

SOMMAIRE

Introduction 1
I. Description 2
I.1. Principe du jeu 2
I.2. Présentation des données 3
II. La Régression de Poisson 4
III. Interprétations des Résultats (SAS) 5
Conclusion 7
Bibliographie 8
Annexe 9

Introduction

Pour la petite littérature sur le sujet, nous dirons que la modélisation de données de comptage est une problématique très répandue dans divers domaines comme les assurances, la banque, l'économétrie, la médecine ou encore le marketing. Aussi, les méthodes de modélisation adaptées à ce type de données ont été largement explorées dans la littérature. La régression de Poisson est le recours standard dans ce genre de situation, cependant, de nombreuses applications à des cas réels ont mis en évidence la nécessité de trouver des solutions alternatives permettant de gérer les problèmes sur-dispersion et les excès de zéros induits par les mécanismes du phénomène étudié. Parmi les alternatives existantes, les régressions **hurdle** (**Mullahy**, 1986) et zero-inflated (Lambert, 1992) répondent de manière spécifique au problème des zéros en excès tout en gérant la sur-dispersion des données. Les travaux de recherche sur la généralisation de ces modèles ainsi que leurs mises en application sont nombreux. **Consul** et **Famoye** (1992) proposent une régression de Poisson généralisée avec l'introduction d'un nouveau paramètre dans le modèle standard pour modéliser la dispersion. Récemment, **Famoye** et **Singh** (2006) développent une régression de Poisson généralisée zero-inflated pour modéliser les violences domestiques.

Dans le cadre d'un projet d'études, nous mettrons en application un modèle de comptage dans le but de réaliser des prévisions sur la somme totale de boules gagnantes tirées lors d'un jeu loto EuroMillions.

L'EuroMillions[1] est un jeu de loterie à l'échelle européenne avec des tirages tous les mardis et vendredis, c'est à dire un jeu de hasard et d'argent, fonctionnant par répartition d'une partie des sommes misées par les joueurs entre les différents niveaux de gain (17 à 210 millions d'euros).

[1] https://www.lesbonsnumeros.com/euromillions/informations/regles-du-jeu.htm

I. Description

I.1. Principe du jeu

Une grille de jeu EuroMillions valide doit comporter 5 numéros et 2 étoiles.

Le principe consiste à choisir 5 numéros entre 1 et 50 et 2 étoiles numérotées de 1 à 12, dans deux tirages faits par semaine. Ainsi, la probabilité de gagner est 1 sur 140 millions. Le joueur peut utiliser le système Quick Pick (ou Flash en France) au moment de l'achat d'un ticket de participation où les numéros et étoiles vont être choisis au hasard par l'ordinateur enregistrant le pari.

En 2013, le chiffre d'affaires du jeu total a été de 6,6 milliards d'euros (1,5 milliard d'euros en 2004). En 2018 il fut de 15,8 milliards d'euros. C'est un jeu qui peut intégrer la participation de plusieurs personnes.

Pour la grille des numéros des boules, la probabilité de trouver parmi les 5 numéros tirés :

- 5 bons numéros est de 1 chance sur 2 118 760 soit 0,000 047 %
- 4 bons numéros est de 1 chance sur 9 417 soit 0,011 %
- 3 bons numéros est de 1 chance sur 214 soit 0,467 %
- 2 bons numéros est de 1 chance sur 14,93 soit 6,70 %
- 1 bon numéro est de 1 chance sur 2,84 soit 35,2 %
- 0 bon numéro est de 1 chance sur 1,73 soit 57,7 %

Pour la grille des étoiles, la probabilité de trouver parmi les 2 étoiles tirées :

- 2 bonnes étoiles est de 1 chance sur 66 soit 1,52 %
- 1 bonne étoile est de 1 chance sur 3,3 soit 30,3 %
- 0 bonne étoile est de 1 chance sur 1,47 soit 68,18 %

L'actuel record de gain à l'EuroMillions est de 200 000 000 € remporté en France le 11 décembre 2020. Le précédent record était de 190 000 000 €, un montant remporté quatre fois: le 10 août 2012 au Royaume-Uni (soit 148 656 000 £ au taux de change de l'époque), le 24 octobre 2014 au Portugal (soit 152 000 000 € net après la déduction des 20% d'impôt applicables dans ce pays), le 6 octobre 2017 en Espagne (idem, 152 000 000 € net après la déduction des

20% d'impôt applicables dans ce pays) et le 8 octobre 2019 de nouveau au Royaume-Uni (soit 170 221 000 £ au taux de change de l'époque).

Depuis sa création, et à la date du 2 janvier 2021, l'EuroMillions compte 501 gagnants ont remporté le jackpot à eux tout seul ou qui se sont partagé le jackpot à parts égales recensés à travers toute l'Europe :

- 110 en France
- 110 au Royaume-Uni
- 106 en Espagne
- 78 au Portugal
- 38 en Belgique
- 22 en Suisse
- 18 en Irlande
- 16 en Autriche
- 3 au Luxembourg

I.2. Présentation des données

Sur une base de données de 175 observations et 53 variables, nous allons procéder aux transformations suivantes :

- « Somme_B » pour la somme totale des boules (*variable endogène*).
- « obs » pour le nombre d'observations (175).
- « d_t » pour les dates de tirages.
- « nb_g_f » est le nombre total des 13 rangs de gagnants en France.
- « nb_g_e » de même pour ceux d'Europe.
- « nb_gj » variable recodée pour le nombre de grilles jouées.

Ces trois (3) dernières variables exogènes seront transformées en logarithmes pour réduire la dispersion de données afin d'avoir des résultats significatifs dans notre modèle de régression.

Nous n'avons exposé dans cette description que les variables ayant un rapport direct avec notre phénomène d'étude (seul le cas des boules sera traité).

II. La Régression de Poisson

La régression de Poisson est un modèle de prédiction qui s'applique lorsque la variable cible Y est une variable de comptage (nombre d'apparition d'un évènement durant un laps de temps).

Distribution de la loi de Poisson

$$P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!}$$

Avec :

- Y une variable aléatoire de Poisson.
- y = 0,1,2,…,N ; un entier naturel.
- λ est le paramètre de la loi de Poisson, où E[Y] = λ et V[Y] = λ

Objectif de la Régression de Poisson

Dans notre cas cela implique d'estimer un modèle de la forme suivante :

$$E[Y] = \lambda = exp(cst + \beta X)$$

On estime cst et β par la méthode du maximum de vraisemblance, on en déduit $\hat{\lambda}$:

$$\widehat{E[Y]} = \hat{\lambda} = exp(\widehat{cst} + \hat{\beta}X)$$

NB : le vecteur des variables exogènes X est pris en logarithme (*ln*).

III. Interprétations des Résultats (SAS)

poisson regression

La procédure COUNTREG

Synthèse de l'ajustement de modèle	
Variable dépendante	Somme_B
Nombre d'observations	172
Table	WORK.IMPORT2
Modèle	Poisson
Log-vraisemblance	-796.91195
Gradient absolu maximal	2.2101E-10
Nombre d'itérations	3
Méthode d'optimisation	Newton-Raphson
AIC	1602
SBC	1614

L'algorithme a convergé.

Paramètres estimés					
Paramètre	**DDL**	**Estimation**	**Erreur type**	**Valeur du test t**	**Approx Pr > |t|**
Intercept	1	-1.307965	0.401008	-3.26	0.0011
lnnb_gj	1	1.867975	0.088502	21.11	<.0001
lnnb_g_f	1	-0.468909	0.096357	-4.87	<.0001
lnnb_g_e	1	-1.358055	0.155108	-8.76	<.0001

Les variables exogènes du modèle ont toutes un fort pouvoir explicatif sur notre variable endogène. Elles sont toutes significatives au seuil de 1% (de même au niveau de 5%). Nous pouvons ainsi affirmer que le choix et la transformation des variables furent les bons.

Les coefficients de notre modèle semi-log s'interprètent ainsi :

- une augmentation de 1% du nombre de grilles jouées entraine ceteris paribus (toutes choses étant égales par ailleurs) une augmentation de la somme des boules gagnantes de 1,867975 points ;
- si le nombre de gagnants en France augmente de 1%, la somme totale des boules baisse ceteris paribus, de 0,468909 points ;
- une augmentation de 1% du nombre de gagnants en Europe entraine ceteris paribus, une baisse de 1,358055 points de la somme totale des boules gagnantes tirées ;
- l'intercept (ou constante) : en fin de compte, si aucune des variables exogènes ne subissent d'augmentation, la somme totale des boules gagnantes est, ceteris paribus, susceptible de diminuer de 1,307965 points.

Notre but étant de réaliser des prévisions sur la somme totale de boules gagnantes tirées lors d'un jeu loto EuroMillions, nous allons vérifier si notre modèle tient la route en comparant quelques-unes des prévisions obtenues aux valeurs données par la base de données.

Obs.	Somme_B	d_t	xbeta	phat
7	**120**	22 March 2019	4.79256	**120.610**
28	**152**	4 June 2019	5.02462	**152.113**
42	**95**	23 July 2019	4.55400	**95.0121**
67	**131**	18 October 2019	4.87521	**131.001**
122	**102**	5 May 2020	4.63123	**102.640**
95	**116**	28 January 2020	4.75573	**116.248**

Nous voyons à travers ces quelques observations que le modèle choisi semble effectivement prévoir la somme totale des boules gagnantes obtenues lors du tirage. Le modèle ainsi spécifié est excellent.

Conclusion

En somme, nous retenons suite au travail réalisé dans le cadre du cours d'économétrie sur les variables qualitatives à choix discrets, de même pour tout cours d'économétrie, que l'aboutissement à un bon modèle se trouve au préalable dans la construction et la gestion de données.

Notre estimation par la régression de poisson a permis de prévoir les résultats attendus de notre scénario de jeu sur les boules gagnantes à l'EuroMillions 2019-2020. De ce fait, la véracité du modèle choisi.

Au regard des analyses faites sur la nature des relations de causalité entre variables endogène et exogènes, nous avons pu comprendre leur interdépendance vis-à-vis de ce jeu. Pour ainsi dire, une meilleure prévision dépend fortement de l'agencement entre ces différentes variables. Pour les dates du 22 Mars, 04 Juin, 23 Juillet et 18 octobre 2019, nous avons réalisé des prévisions similaires aux résultats de base. De même pour les dates du 28 Janvier et 05 Mai 2020.

Bibliographie

CONSUL, P. C. et FAMOYE, F. (1992), « Generalized Poisson regression model. », *Communications in Statistics, Theory and Methods*, 21, pp. 89-109.

LIU, W. et CELA, J. (2008), « Count Data Models in SAS. », *Proceedings of SAS Global Forum*, paper 371-2008.

RAKOTOMALALA, R. (2010), « Régression de Poisson - Modèle de comptage. », *Laboratoire ERIC, Unité de Recherche Universitaire*.
http://eric.univlyon2.fr/~ricco/cours/slides/regression_poisson.pdf

SEGUELA, J et SAPORTA, *G.* (2010), « Modèles de comptage appliqués aux décisions de candidature aux offres d'emploi sur le web », *Laboratoire Cédric – CNAM.*
https://cedric.cnam.fr/fichiers/art_2274.pdf

Annexe

Obs.	Somme_B	d_t	xbeta	phat
1	91	1 March 2019	4.72299	112.505
2	117	5 March 2019	4.93493	139.064
3	118	8 March 2019	4.72865	113.143
4	75	12 March 2019	4.35199	77.633
5	130	15 March 2019	4.94020	139.798
6	97	19 March 2019	4.87295	130.706
7	120	22 March 2019	4.79256	120.610
8	144	26 March 2019	4.71095	111.158
9	141	29 March 2019	4.81289	123.087
10	89	2 April 2019	4.81180	122.953
11	92	5 April 2019	4.62021	101.515
12	147	9 April 2019	4.94474	140.434
13	146	12 April 2019	5.14466	171.513
14	117	16 April 2019	4.81985	123.947
15	137	19 April 2019	4.93774	139.454
16	154	23 April 2019	4.65894	105.524
17	179	26 April 2019	5.12188	167.650
18	111	30 April 2019	4.67058	106.759
19	101	3 May 2019	4.72176	112.366
20	132	7 May 2019	4.84039	126.518
21	125	10 May 2019	4.88298	132.023
22	116	14 May 2019	4.63915	103.457
23	174	17 May 2019	5.04192	154.768
24	185	21 May 2019	5.21681	184.345
25	123	24 May 2019	4.74779	115.329
26	103	28 May 2019	4.50044	90.056
27	109	31 May 2019	4.88381	132.133
28	152	4 June 2019	5.02462	152.113
29	75	7 June 2019	4.76947	117.857
30	179	11 June 2019	5.03261	153.333
31	114	14 June 2019	4.72487	112.716

Obs.	Somme_B	d_t	xbeta	phat
32	157	18 June 2019	4.78493	119.693
33	86	21 June 2019	4.50743	90.688
34	139	25 June 2019	4.94929	141.075
35	102	28 June 2019	4.84327	126.884
36	148	2 July 2019	4.79967	121.470
37	107	5 July 2019	4.85984	129.003
38	145	9 July 2019	4.97672	144.998
39	164	12 July 2019	5.13314	169.548
40	78	16 July 2019	4.59974	99.459
41	98	19 July 2019	4.56994	96.538
42	95	23 July 2019	4.55400	95.012
43	113	26 July 2019	4.84563	127.184
44	125	30 July 2019	4.69576	109.482
45	162	2 August 2019	4.72000	112.169
46	128	6 August 2019	4.84553	127.171
47	134	9 August 2019	5.14476	171.531
48	146	13 August 2019	4.83538	125.887
49	125	16 August 2019	4.55669	95.267
50	126	20 August 2019	4.91013	135.657
51	143	23 August 2019	4.93127	138.555
52	159	27 August 2019	5.03170	153.194
53	192	30 August 2019	5.05611	156.979
54	139	3 September 2019	4.82601	124.712
55	133	6 September 2019	4.79777	121.240
56	129	10 September 2019	4.83319	125.612
57	129	13 September 2019	4.75406	116.055
58	191	17 September 2019	5.10342	164.584
59	142	20 September 2019	4.85883	128.873
60	172	24 September 2019	5.13302	169.529
61	191	27 September 2019	5.17724	177.193
62	117	1 October 2019	4.97815	145.206
63	135	4 October 2019	4.91623	136.487
64	125	8 October 2019	4.77852	118.929
65	122	11 October 2019	4.82086	124.072

Obs.	Somme_B	d_t	xbeta	phat
66	128	15 October 2019	4.95045	141.238
67	131	18 October 2019	4.87521	131.001
68	119	22 October 2019	4.95554	141.960
69	200	25 October 2019	5.24500	189.615
70	153	29 October 2019	4.91250	135.979
71	153	1 November 2019	5.01860	151.200
72	129	5 November 2019	4.88383	132.136
73	113	8 November 2019	4.68536	108.349
74	109	12 November 2019	4.84623	127.260
75	102	15 November 2019	4.84759	127.433
76	105	19 November 2019	4.74945	115.521
77	138	22 November 2019	5.01993	151.401
78	114	29 November 2019	4.65829	105.456
79	167	3 December 2019	5.13278	169.488
80	116	6 December 2019	4.90043	134.348
81	99	10 December 2019	4.64194	103.745
82	98	13 December 2019	4.57367	96.899
83	94	17 December 2019	4.57589	97.115
84	128	20 December 2019	4.91137	135.826
85	88	24 December 2019	4.60060	99.544
86	91	27 December 2019	4.71511	111.621
87	179	31 December 2019	5.10355	164.606
88	123	3 January 2020	4.96269	142.978
89	105	7 January 2020	4.75346	115.985
90	145	10 January 2020	4.89938	134.206
91	158	14 January 2020	4.92331	137.456
92	120	17 January 2020	4.67572	107.310
93	127	21 January 2020	4.86436	129.588
94	46	24 January 2020	4.48542	88.714
95	116	28 January 2020	4.75573	116.248
96	159	4 February 2020	4.97219	144.343
97	106	7 February 2020	4.63342	102.865
98	171	11 February 2020	5.07842	160.521
99	150	14 February 2020	5.01969	151.365

Obs.	Somme_B	d_t	xbeta	phat
100	162	18 February 2020	5.07381	159.783
101	137	21 February 2020	4.71944	112.106
102	92	25 February 2020	4.68124	107.903
103	84	28 February 2020	4.51938	91.778
104	134	3 March 2020	4.84059	126.544
105	187	6 March 2020	5.20882	182.879
106	144	10 March 2020	4.86284	129.392
107	98	13 March 2020	4.58061	97.574
108	56	17 March 2020	4.42075	83.159
109	106	20 March 2020	4.78391	119.571
110	108	24 March 2020	4.53836	93.538
111	139	27 March 2020	4.80576	122.213
112	119	31 March 2020	4.63004	102.518
113	165	3 April 2020	4.93806	139.500
114	104	7 April 2020	4.80880	122.585
115	102	10 April 2020	4.57650	97.174
116	144	14 April 2020	5.09923	163.896
117	158	17 April 2020	4.93601	139.213
118	86	21 April 2020	4.32451	75.529
119	74	24 April 2020	4.36749	78.845
120	124	28 April 2020	4.64089	103.637
121	169	1 May 2020	4.91733	136.638
122	102	5 May 2020	4.63123	102.640
123	136	8 May 2020	4.83368	125.673
124	130	12 May 2020	4.77525	118.540
125	139	15 May 2020	4.77321	118.299
126	142	19 May 2020	4.96945	143.948
127	102	22 May 2020	4.61764	101.255
128	75	26 May 2020	4.49168	89.271
129	88	29 May 2020	4.57548	97.074
130	119	2 June 2020	4.69280	109.158
131	94	5 June 2020	4.54515	94.175
132	129	9 June 2020	4.83298	125.584
133	156	12 June 2020	5.08167	161.042

Obs.	Somme_B	d_t	xbeta	phat
134	80	16 June 2020	4.53342	93.076
135	105	19 June 2020	4.70524	110.525
136	128	23 June 2020	4.83767	126.175
137	70	26 June 2020	4.58321	97.828
138	88	30 June 2020	4.68971	108.822
139	123	3 July 2020	4.73789	114.192
140	125	7 July 2020	4.77441	118.441
141	123	10 July 2020	4.59494	98.982
142	147	14 July 2020	4.94189	140.035
143	133	17 July 2020	4.70678	110.695
144	124	21 July 2020	5.15765	173.756
145	122	24 July 2020	4.70818	110.851
146	88	28 July 2020	4.51359	91.248
147	170	31 July 2020	5.00743	149.520
148	65	4 August 2020	4.67225	106.938
149	123	7 August 2020	5.02090	151.548
150	127	11 August 2020	4.80424	122.027
151	129	14 August 2020	4.83004	125.216
152	111	18 August 2020	4.77808	118.876
153	154	21 August 2020	4.92481	137.664
154	131	25 August 2020	4.84359	126.924
155	83	28 August 2020	4.58436	97.940
156	101	1 September 2020	4.85874	128.862
157	140	4 September 2020	4.92889	138.226
158	118	8 September 2020	5.03307	153.403
159	202	11 September 2020	5.19999	181.271
160	129	15 September 2020	4.82060	124.039
161	101	18 September 2020	4.73886	114.304
162	148	22 September 2020	4.94558	140.552
163	158	25 September 2020	4.82086	124.072
164	119	29 September 2020	4.92501	137.690
165	118	2 October 2020	4.78244	119.395
166	149	6 October 2020	5.13187	169.333
167	152	9 October 2020	4.94407	140.340

Obs.	Somme_B	d_t	xbeta	phat
168	144	13 October 2020	5.05631	157.010
169	176	16 October 2020	5.03459	153.636
170	105	20 October 2020	4.85362	128.204
171	88	23 October 2020	4.51189	91.094
172	132	27 October 2020	4.78693	119.933