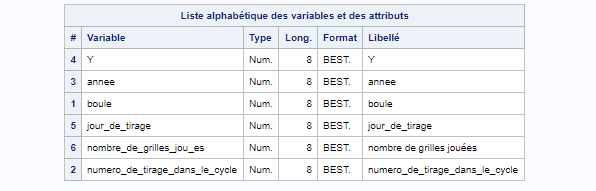
**Codes et interprétations**

Comme énoncé précédemment, nous travaillerons d’abord sur la base de données dans son ensemble, avant de regarder ce qui se passe pour l’année 2019 dans un premier temps, puis pour l’année 2020 dans un second temps. En effet, l’année 2020 étant assez spéciale, nous tenterons de répondre à la question suivante : as t il eu un effet corona virus ?

Tout d’abord, la présente base de données est modifiée avec des variables codés 0 1 et des passages au log dans le but d’avoir des ordres de grandeur plus réalistes .

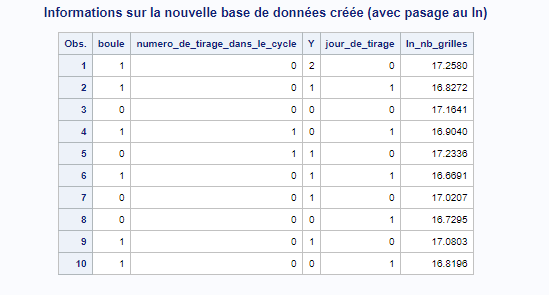
Les variables que nous retenons dans notre base de données sont les jours de tirages, les années de tirages, la parité de la somme des boules de b1 a b5, le nombre de grilles jouées, le numéro de tirage dans le cycle (codés en 0 ou 1) et notre variable endogène Y. Cette variable Y que nous tentons d’expliquer par rapport aux variables précédentes fait référence aux Etoiles. Elle est codée de la façon suivante : si les 2 étoiles sont toutes les 2 inferieures à 9, si elles sont supérieures à 10 ou si l’une est inferieure à 9 et, l’autre supérieur à 10



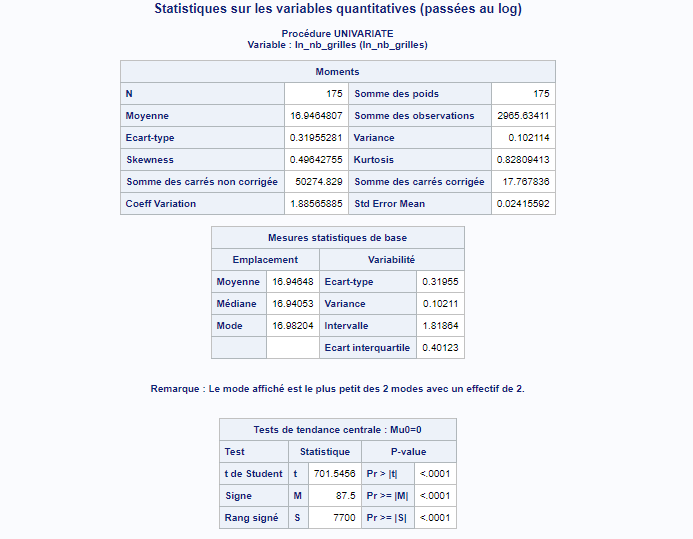
Ainsi, nous arrivons à tirer quelques informations de la variable quantitative (avant le passage au log) de notre base de données :

Ayant qu’une seule variable quantitative à savoir le nombre de grilles jouées, certaines de ses données sont connues à savoir sa moyenne, son écart type, sa variance etc…

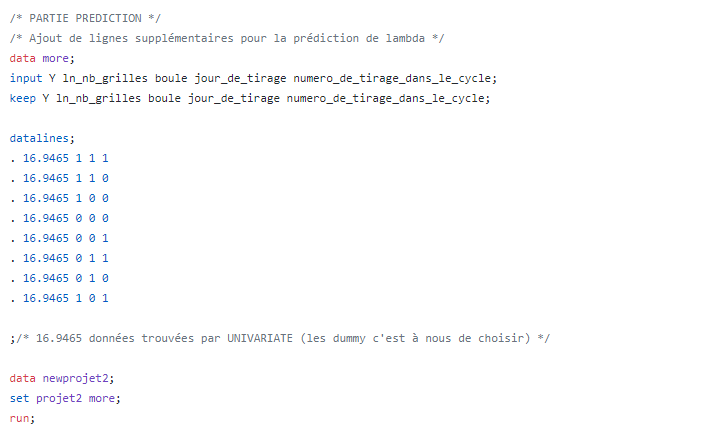
Il est tout à fait normal que ces statistiques changent après le passage au log. Nous nous retrouvons avec des valeurs moins importantes en taille et plus facile à manier comme l’atteste l’image ci-dessous :



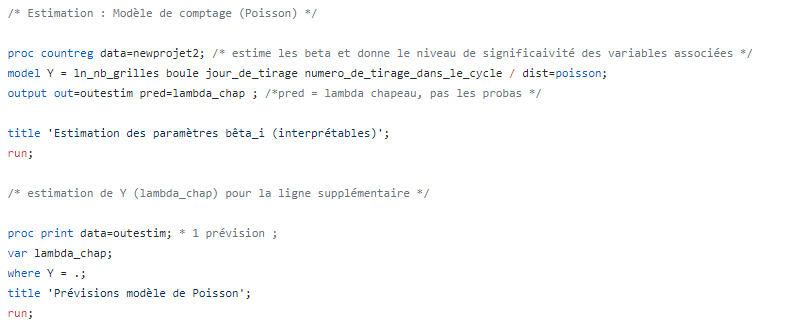
Les nouvelles statistiques après le passage au log sont les suivantes :

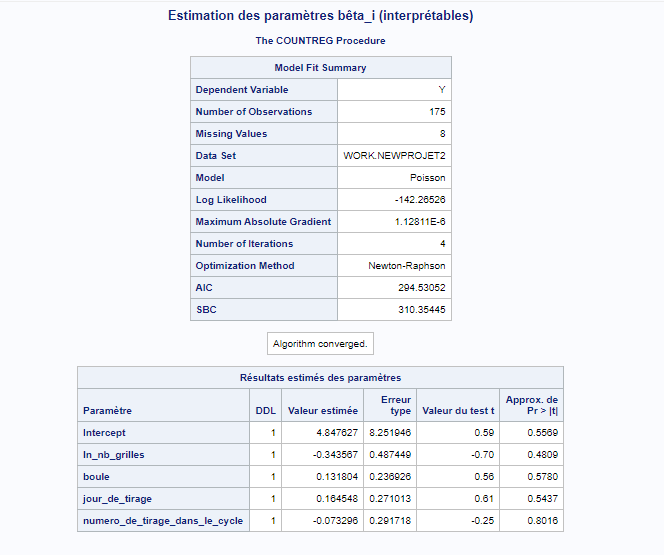


Disposant de notre base de données modifiée, nous pouvons enfin commencer notre prédiction à l’aide de la recherche du lambda dans nos différents cas de figure.



Notre variable Y est codée comme étant le nombre de boules supérieures à 10. En effet, il en va de soi que la probabilité que Y soit égale à 0 fait référence à la probabilité que 0 boule soit supérieure à 10 . Les deux boules sont donc toutes les deux comprises entre les entiers 0 et 10. La probabilité que notre Y soit égale à 1 nous donne la probabilité qu’une boule soit supérieure à 10, la probabilité que Y soit égale à 2 fait référence à la probabilité que les 2 boules soient supérieures à 10. La problématique est restée la même mais elle a juste été reformulée différemment.



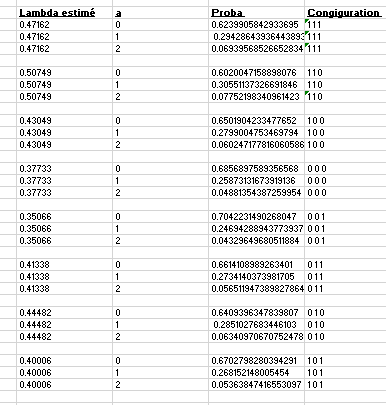


Interprétation : \*Les paramètres ne sont pas du tout significatifs. On s’en doutait un peu car si on avait connaissance de ce qui influe sur l’euro million, le nombre de gagnants aurait fortement augmenté.

Certains paramètres sont positifs, d’autres sont négatifs. Leurs signes témoignent de leurs apports positifs ou négatifs sur l’explication de notre variable Y. Cependant, du fait de leurs caractères qualitatifs, ces betas ne sont pas interprétables, nous essaierons simplement d’interpréter les betas de la variable quantitative à, savoir le nombre de grilles. Nous pouvons affirmer de ce fait que l’augmentation de 1% de ce nombre entraine une diminution de 0.343 du nombre de boules supérieures à 10.

De plus, pour chaque configuration et pour un nombre de grille joué en moyenne égale à 16.9465, nous avons un lambda chapeau, lambda chapeau qui nous permettra de calculer la probabilité pour que les étoiles soient en dessous de 10, une au-dessus et les deux au-dessus (a=0 , 1 et 2). Le lambda chapeau apporte de plus une certaine information par rapport à sa proximité aux entiers 0, 1 et 2. Un lambda estimé très proche de 0 nous montre que, en moyenne, les boules sont très proches de 0. Sachant que c’est souvent le cas, nous comprenons ceci comme étant le fait que, très souvent, 0 boule est supérieur à 10 . Ceci est justifiée par une inégale répartition de notre jeu de données avec les quasis ¾ des observations codées comme 0 en ce qui concerne la somme des boules.

Résumons ceci dans le tableau ci-dessous



\*\*\*\*

Tout d’abord , apportons une explication a la configuration . Cette configuration joue sur la logique suivante : boule jour\_de\_tirage numero\_de\_tirage\_dans\_le\_cycle ;

Pour rappel, en ce qui concerne le codage des boules, ce codage prend 0 si la somme des boules (variables exogènes B1 à B5) est pair. Elle est codée 1 sinon.

Le jour de tirage est codé en 0 s’il s’agit d’un Mardi et en 1 s’il s’agit d’un Vendredi.

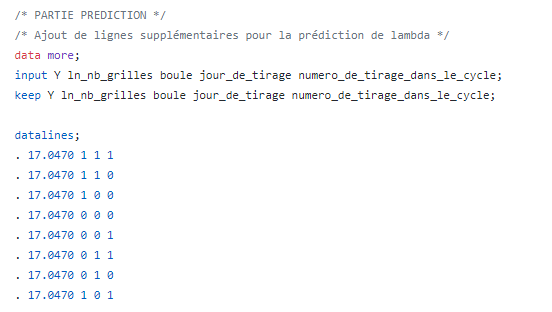
Quant au numéro de tirage dans le cycle, il est codé en 0 si le numéro de tirage est compris entre 1 et 5 et en 1 sinon.

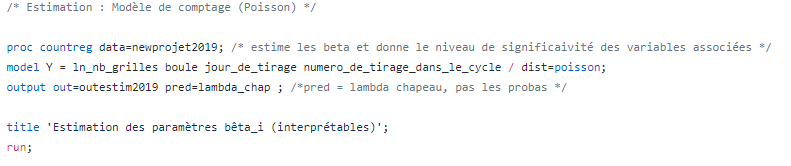
Ce tableau s’interprète comme suit : avec le premier lambda, la probabilité que Y soit égale à 0 sachant la configuration 1 1 1 est 0.623, la probabilité que le Y soit égale a 1 est de 0.069 et la probabilité que le Y soit égale a 2 est très petite, égale à 0. 069. Ces prévisions sont valables pour la base de données dans son ensemble. La probabilité qui nous semble la plus importante est dans le cas du lambda égale à 0.35066 pour la configuration 0 0 1 pour le calcul de Y=0 . Cette probabilité est 0.70. La modalité 0 revient très souvent dans le cas d’une somme des boules paires, pour un tirage effectuer un Mardi et pour un numéro de tirage n’étant pas compris entre 0 et 1. Dans la suite, nous traiterons la prévision suivant l’année 2019 dans un premier temps, puis suivant l’année 2020 dans un second temps.

2019

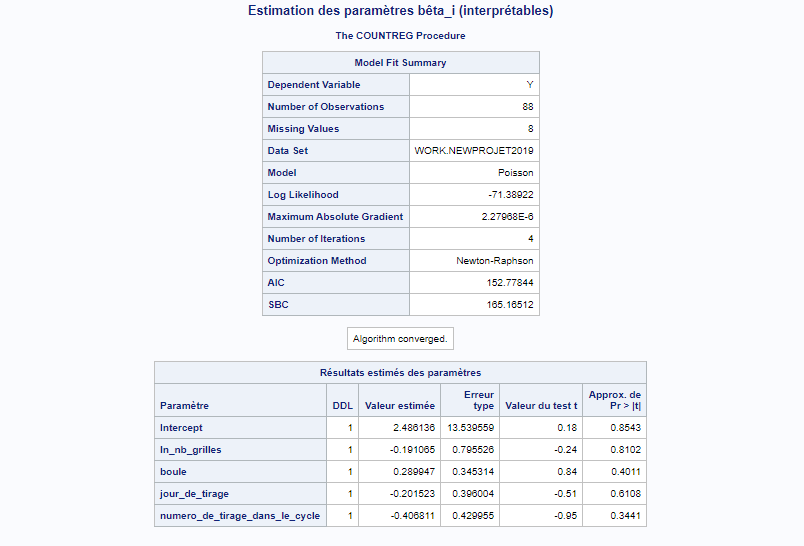


Nous restons toujours avec la base au format log car comme dit précédemment, elle est plus facile à manier. Le tableau ci-dessus nous témoigne des statistiques de la variable quantitative (nombre de grilles jouées) . Notre étude utilisera encore une fois essentiellement, la moyenne, qui est de 17.0470 grilles jouées. Ainsi, à l’aide de cette donnée, nous pouvons commencer nos prédictions pour une configuration donnée.



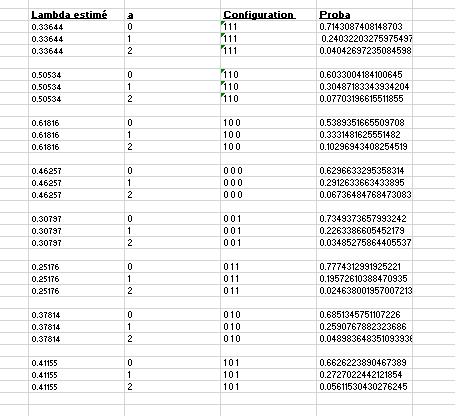


Comme précédemment, l’algorithme a convergé . L’estimation des bêtas est donné dans le tableau ci-dessous. Les betas des variables qualitatives ne sont toujours pas interprétables. Nous interprèterons uniquement le beta de la variable quantitative après avoir discuté de la significativité des variables explicatives.



Le critère AIC est passé de 294 pour une étude sur les 2 années, a 152 pour une étude uniquement sur l’année 2019. Sous la volonté d’une minimisation du critère AIC, il est clair que le modèle « fit » mieux sur l’année 2019 que sur les 2 années combinées. Les coefficients des variables explicatives restent non significatifs. Nous pouvons dire que notre modèle fait fi du fait que l’année soit précisée ou pas. La conclusion reste la même : « Si on connaissait quelles variables jouent sur l’euromillion , le nombre gagnant aurait énormément augmenté » . Nous pouvons dire que l’augmentation de 1% du nombre e grilles joué entraine une diminution de 0.19 du nombre de boules supérieurs à 10.

Par la suite, la prédiction d’une configuration nous permet de d’avoir un certain lambda : un lambda estimé . Ce lambda estimé nous permet de calculer une proba suivant a=0 1 ou 2. Tout ceci est résumé dans le tableau ci-dessous.



\*\*\*\*

\*\*\*\*

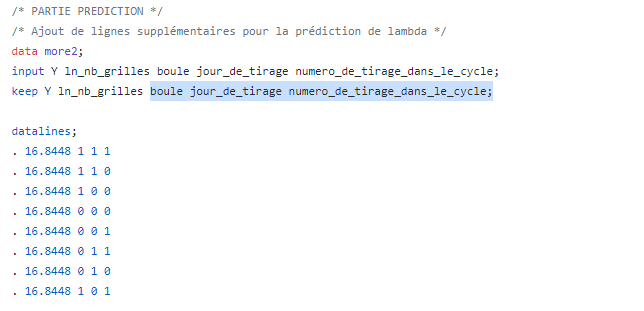
Les 2 cas de figures qui nous intéressent sont les cas 1 0 0 pour un Y égal à 2 avec un lambda estimé à 0.616, et le cas 0 1 1 pour un Y égal a 0 pour un lambda estimé à 0.25. Dans le premier cas de figure, la probabilité que les deux boules soient supérieures à 10, sachant que la somme des boules est impaire, que le tirage ait lieu un Vendredi et que le numéro de tirage soit compris entre 1 et 5, cette probabilité est relativement élevée au regard de tous les autres cas de figure où nous souhaitons avoir toutes les 2 boules supérieurs à 10. Cette proba est de 0.10 pratiquement alors que nous avions l’habitude d’avoir des probas de l’ordre de 0.07. Dans le deuxième cas de figure, la proba qu’aucune des 2 boules soient supérieures à 10, pour un lambda estimé à 0.25, sachant que la somme des boules soient paire, pour un Vendredi comme jour de tirage, avec un numéro de tirage n’étant pas compris entre 1 et 5, cette proba est très élevé et est égale a 0.77. En d’autres termes, sous ces hypothèses, nous sommes sures a 80% qu’aucune des boules ne sera supérieurs a 10

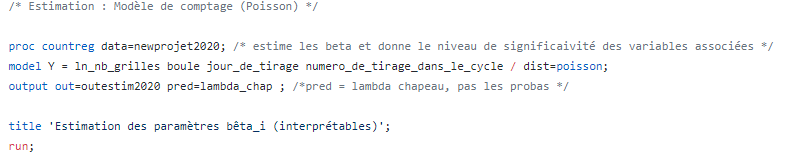
2020

Malheureusement, 2020 n’a pas été une année comme les autres. Ainsi, sous la contrainte du COVID-19, il est clair qu’un certain nombre de données statistiques soit affecté .

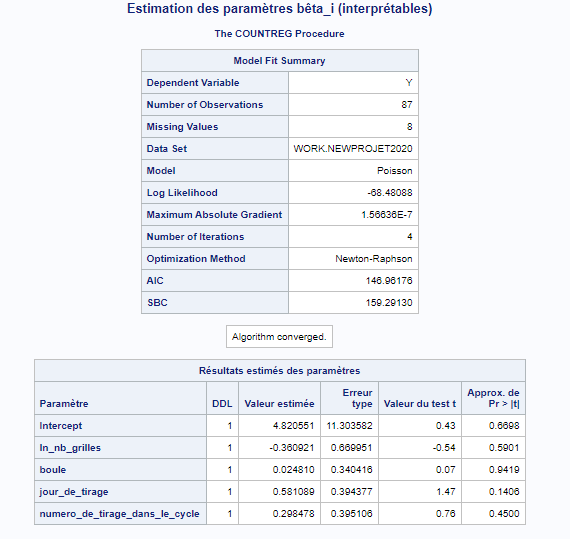


Comme dit ci-dessus, en raison du confinement et de la fermeture de certains points de ventes relatifs à l’euromillion , le nombre de grilles vendus en moyenne baisse . Notre étude utilisera encore une fois essentiellement, la moyenne, qui est passé de 17.0470 à 16.844 grilles jouées. Comme précédemment, nous avons :



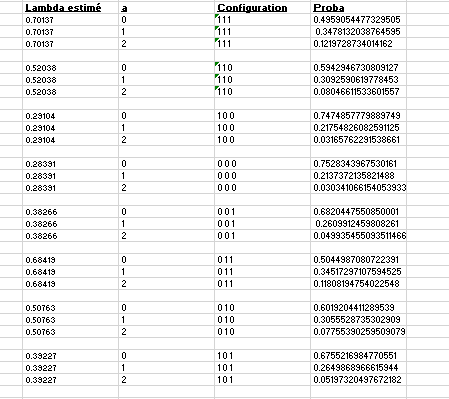


L’estimation des bêtas s’illustre dans le tableau ci-dessous. Les betas des variables qualitatives restent non interprétables. Nous interprèterons uniquement le beta de la variable quantitative après avoir discuté de la significativité des variables explicatives.



Le critère AIC est passé de 152 pour une étude uniquement sur l’année 2019, à 146 pour l’année 2020. Le modèle « fit » mieux sur l’année 2020 comparé aux cas précédents. Les coefficients des variables explicatives restent non significatifs, ce qui ne nous surprend plus. De plus, l’augmentation de 1% du nombre de grilles joué entraine une diminution de 0.36 du nombre de boules supérieurs à 10.

Notre tableau qui résume nos lambdas calculés suivant une certaine configuration et les probabilités associées, se présentent ci-dessous. Nous tenterons de commenter certaines probabilités qui se distinguent de par leurs grandeurs.



\*\*\*

\*\*\*

2 cas de figures se distinguent. Le premier cas de figure est celui d’un lambda estimé a 0.70, pour une probabilité que 2 boules soient supérieurs a 10, suivant une configuration 1 1 1, égale à 0.12 . Cette proba est la plus forte qu’on puisse avoir pour un tel nombre de boule et est donc maximale que dans la configuration 1 1 1 , c’est-à-dire une somme des boules paires , Vendredi comme jour de tirage et un numéro de tirage n’étant pas compris entre 1 et 5 . Le deuxième cas de figure correspond au lambda estimé a 0.29 pour une configuration 1 0 0 pour une proba qu’aucune boule soit supérieur à 10 . Cette proba est égale a 0.74. Ce qui veut dire que, si le jour de tirage est un vendredi ,que la somme des boules est impaires et que le numéro de tirage n’est pas compris entre 1 et 5 , nous avons environ 75% de chance qu’aucune boule soit supérieurs a 10.

CONCLUSION : ( petite ébauche, à finir ultérieurement )

Nous aurions pu jouer sur le nombre de grilles joué maximale ou minimale

Le modèle de poisson peut être utilisé dans le passé et le futur. En effet, lors de son introduction en cours, nous l’avions utilisé pour prédire le nombre de médailles connaissant certains paramètres. Adopter cette approche pour notre jeu de donnée n’était pas séduisant. C’est pourquoi nous avions préféré jouer sur la carte futuriste de ce modèle.

Pour répondre à notre problématique, nous avons remarqué que le nombre d’observations est quasi la même. Une différence d’observation aurait pu biaiser notre étude . Donc les différences entre les prédictions sur l’année 2019 et l’année 2020 ne peuvent être lié qu’a un phénomène particulier a 2020. Nous avons pensé à cette pandémie et ceci peut également justifier la baisse du nombre de grille joué .•

***A rajouter :***

\*Dire peut-être pourquoi les proba pour a=0 sont les plus fortes (done)

\*Rappeler les configurations pour aider le lecteur (done)

\*Faire pour 2020 (done)

\*et y introduire effet corona (done)

\*Rappeler qu’en classe, on joue sur le passé ,on faisait une Verif alors que là,on prévoit le turfu

\*Dire qu’au lieu de jouer sur la valeur moyenne du nombre de grilles joué, on peut jouer sur le nombre de grille max ou min par exemple

\*Parler de la minimisation du critère AIC (comparer les critères AIC et BIC de 2019 et 2020 ) et dire le quel fit mieux notre jeu de données (done)

\*Faire les mêmes screens pour 2020 et 2019 en y rajoutant les moyennes , nbres de grilles joués (done)