

Projet - MST

Nicolas Brunel, Anastase Charantonis, Christian Kahindo

19/04/2019

*Ce projet est à déposer sur exam.ensiie.fr **avant** minuit le 26/05/2019.*

CONSIGNES

1. Vous devez effectuer OBLIGATOIREMENT votre rapport en BINOME.
2. Le rendu doit être un document .pdf de 5 à 10 pages, limites strictes. Il ne peut pas contenir des annexes.
3. La notation tiendra également compte de la qualité de rédaction ainsi que de la qualité et de la pertinence des interprétations des graphiques.
4. Le rapport doit avoir comme titre : Binome_Numero_Nom1_Nom2.pdf

Des jeux de cartes à collectionner (https://fr.wikipedia.org/wiki/Jeu_de_cartes_à_collectionner) existent depuis plus de 25 ans (par exemple Magic the Gathering). Dans ces jeux, les joueurs rassemblent des paquets de cartes à partir de leurs collections de cartes, et affrontent d'autres joueurs qui ont fait de même. Un marché de cartes existe pour la plus grande partie de ces jeux, et deux paquets peuvent avoir un coût très différent.

La valeur des cartes constituant un jeu peut avoir un impact sur la performance de celui-ci en tournoi. Nous étudions ici les proportions de victoires (en %) obtenues dans un tournoi de haut niveau, en fonction du prix total du jeu d'un joueur (en Euros), de sa tranche d'âge (jeune : 0, adulte : 1, âgé : 2) et de son sexe (homme : 0 femme : 1).

Des pourcentages de victoire supérieurs à 100% ou inférieurs à 0% correspondent à des irrégularités et doivent être éliminés.

Le but de ce projet est de réaliser une étude en utilisant ce qui a été vu durant le module de modélisation statistique. Le projet comprend plusieurs parties : statistiques descriptives, estimation, intervalle de confiance et test.

Quand vous avez votre binôme, allez sur [<https://preview.tinyurl.com/yy5hcpmv>]. Vous y trouverez un document de type spreadsheet à remplir avec les noms de vos binômes. Inscrivez votre nom dans les cases adaptées, puis téléchargez la base de données correspondante à votre binôme, présente dans le dossier BD ici : [<https://pydio.ensiie.fr/public/764b34>].

1. Statistiques Descriptives

- (a) Tracer les boîtes à moustache du pourcentage de victoires pour les hommes et pour les femmes sur un même graphique. Que remarquez-vous ?
- (b) Tracer le pourcentage de victoires en fonction de la tranche d'âge et du prix du jeu. Que remarquez-vous ?
- (c) Tracer l'histogramme du prix des jeux. Quelles distributions le prix des jeux pourrait-il suivre (justifiez) ?
- (d) Tracer l'histogramme du pourcentage de victoires. Quelles distributions le pourcentage de victoire pourrait-il suivre (justifiez) ?

2. Estimation

On considère que le prix du jeu suit une loi exponentielle translatée $\mathcal{E}(\lambda, x_0)$. Dans cette partie nous cherchons à estimer λ et x_0 .

- (a) Vérifier que les observations suivent une loi exponentielle translatée.
- (b) Ecrire la vraisemblance (ou la log vraisemblance) du modèle.
- (c) Déterminer les estimateurs du maximum de vraisemblance des paramètres λ et x_0 . Tracer la log-vraisemblance de l'échantillon en fonction de la valeur du paramètre λ .
- (d) Donner une estimation du paramètre λ et la faire apparaître sur le graphique.

On suppose que le pourcentage de victoire d'un joueur suit une loi Normale $\mathcal{N}(40 + \frac{P}{10}, \sigma^2)$, où P est le prix de son jeu. On a observé le prix du jeu de chaque joueur.

- (e) Ecrire la vraisemblance (ou la log-vraisemblance) théorique du modèle du pourcentage de victoire.
- (f) Déterminer le maximum de vraisemblance du paramètre σ . Tracer la log-vraisemblance de l'échantillon en fonction de la valeur du paramètre σ .

3. Intervalles de confiance

- (a) Ecrire l'expression théorique de l'intervalle de confiance asymptotique pour le paramètre λ en le justifiant.
- (b) Implémenter ces formules sous R, en faisant varier $\alpha \in \{50\%, \dots, 95\%\}$.

4. Tests

- (a) Est-ce que le pourcentage de victoire des hommes est plus élevé que 50% ? Et celui des femmes ?
- (b) Nous souhaitons comparer le pourcentage de victoire obtenue par les femmes et les hommes. Proposez un test et l'implémenter sous R. Que concluez-vous ?
- (c) Nous souhaitons examiner si le pourcentage de victoire des jeux chers (dont la valeur est supérieure au quantile 70%) sont issus d'une distribution normale avec un écart-type $\sigma = 5$. Quelle est la valeur de alpha (à 0.01 près) à partir de laquelle nous ne pouvons pas rejeter cette hypothèse ?