

Projet – Producteur de jeu télé : Consignes

- Groupes de 2 à 4
- Un fichier .rmd (+ vos fichiers .R si vous en utilisez) et son export format pdf, html ou docx. Pour les démonstrations analytiques un scan ou photo lisible de votre démonstration.
- Vous choisirez une seed spécifique pour générer certaines de vos données. Respectez là! Sinon nous remarquerons tout de suite la copie bête sur 1 autre groupe et la note sera automatiquement divisée par 2 pour les 2 groupes.
- Commentez vos codes le plus possible et représentez graphiquement la distribution de vos données/paramètres (density plot ou histogramme)
- Répondez aux questions avec des phrases.
- 4 points / slide et 2 points pour le bonus. Jusqu'à 2 points de malus si les consignes ne sont pas respectées.
- Merci d'envoyer vos questions à benjamin.duputel@exystat.com **ET** jeremie.lespinasse@exystat.com

Projet - Exercice 1: Juste Prix

- On souhaite créer un nouveau concept de jeu télé pour une grande chaîne. Pour gagner gros il faut réussir plusieurs challenges.
 - Lors d'une première étape vous devez deviner le prix d'un gain en essai. Vous cherchez à estimer la probabilité de trouver. Vous allez dans un premier temps définir une vraisemblance Géométrique (nombre de participants nécessaires avant premier succès.) $P(X=k) = p(1-p)^k$
 - Avec p la probabilité cherchée et k le nombre de participants avant le premier succès.
 - Vous vous attendez à ce que parmi 15 participants 4 trouve du premier coup.
- Proposer un a priori beta. (1point)
- Résoudre analytiquement le calcul du posterior. (1point)
- On observe un premier succès au bout de 6 participants.
- Quelle est la probabilité à posteriori de p ? (1point)
- $P(p|Data > 0.5)$? (1point) **en fixant la seed**

Projet – Exercice 1 : Juste Prix

- Finalement après observation de ce premier succès, les réussites s'enchainent... sur les 20 suivants la moitié réussissent. Vous observez le vecteur succès/échec (0,1,1,0,0,1,0,1,0,1,1,1,0,1,0,0,1,0,1,0).
- La loi géométrique n'est donc plus appropriée. Mais l'information n'est pas perdue. Utiliser le posterior précédemment trouvé pour p comme nouveau prior et proposer une loi pour les données des 20 participants. (1point)
- Résoudre analytiquement le calcul du nouveau posterior. Quelle est la moyenne à posteriori de p ? (2points)
- $P(p|Data > 0.5)$? (1point) **en fixant la seed**

Projet – Exercice 2: La Fosse

- On décide ensuite de proposer une épreuve physique:
 - Les participants doivent sauter une fosse dont on cherche à déterminer la taille. On veut que la probabilité de passer cette épreuve soit de 35% au plus. Si l'épreuve n'est pas passée on perd les éventuels gains de la première.
 - On pense qu'en moyenne un individu saute 110cm en longueur IC95[90,130]. Proposer un a priori normal. (1point)
 - On observe un échantillon test de 50 participants: Les données $N(120,3)$. Générer les données **en fixant la seed** (1 point)
- Quelle est la nouvelle distribution a posteriori de la moyenne? (1 point)
- On sait que la variance des mesures (longueur des sauts) est fixe et vaut 3. Quelle taille devrait faire la fosse pour s'attendre à une probabilité de réussir de 35% au plus ? (1point)

Projet – Exercice 2: La Fosse

- On observe un échantillon test de 450 participants (chargez le fichier data.Rdata : dataset « data_jump »). On remarque que la longueur du saut de chaque personne est dépendante de son âge, son poids et sa taille.
 - Utilisez Stan pour expliquer la longueur du saut par ces variables prédictives en utilisant une régression linéaire simple (1,5point)
 - Choisissez des priors non informatifs adaptés (0,5point)
- Vérifier la convergence et la fiabilité de vos chaînes MCMC (1 point)
- Reporter les résultats obtenus pour les différents paramètres de votre modèle (moyenne, intervalle de crédibilité, plot de la distribution a posteriori) (1 point)

BONUS (+2 points)

- Juste avant l'épreuve vous vous rendez compte que les participants du jour ne sont pas représentatifs de votre échantillon test. Utilisez les données du dataset « participants » et ajoutez un bloc « generated quantities » à votre code pour prédire la longueur des sauts de vos participants. (1,5 point)
- Quelle taille doit faire la fosse pour s'attendre à ce que 35% réussissent ? (0,5 point)

Projet – Exercice 3: La Cible

- Lors de la dernière épreuve, on devra toucher une cible un maximum de fois en tirant à 100m de distance. On voudra que la cible soit touchée au moins 4 fois pour que l'épreuve soit gagnée.
- On pense qu'en moyenne les participants ne toucheront la cible qu'à 2 reprises.
- On observe un échantillon de 50 participants. Les données suivent une loi de poisson de paramètre 3: $rpois(50,3)$. Générer les données en **fixant la seed**. (0,5 point)
- Combien de participants ont réussi l'épreuve? (1point)
- En utilisant un a priori conjugué et en fixant sa variance à 2. Donner la moyenne à posteriori du paramètre recherché. Analytiquement ou en utilisant Stan. (1,5 point)
- Générer des prédictions **en fixant la seed** suivant votre distribution à posteriori. Combien de succès prédisiez-vous maintenant? (1point)