

Introduction aux logiciels pour les statistiques

Examen IS2A3 / 2020-2021 / Tous documents autorisés

Le but de l'examen est de vous permettre de montrer vos acquis. Le sujet est volontairement long. Il est donc conseillé de traiter en premier lieu les questions sur lesquelles vous vous sentez le plus à l'aise. Les langages R et Python sont les seuls autorisés pour répondre au sujet. Vous répondrez à chaque question en utilisant au moins une ligne de code et en formulant au moins une phrase de réponse. Recopier une valeur lue visuellement dans la table de données ou l'environnement pour répondre à une question ne rapporte aucun point. Une valeur de la table ne pourra être utilisée qu'à condition de l'avoir trouvée par du code. Pour répondre à une question, vous pouvez admettre le résultat de la question précédente à condition de le mentionner. À la fin de l'examen, vous enregistrerez votre travail dans un fichier zip à votre nom. Vous l'enverrez ensuite à Sofiane.Maazi@ext.polytech-lille.fr. La clarté des commentaires et la qualité du code constitueront un élément important dans l'appréciation de votre travail. Pour chaque question, vous êtes invité(e) à faire figurer toute trace de recherche même infructueuse. Le barème est donné à titre indicatif. L'ensemble des données de l'énoncé sont fictives. Les images sont disponibles sur rainbowdb.canalblog.com.



QCM (1 points)

Indiquez pour chaque affirmation si elle est vraie ou fausse. Une réponse incorrecte n'enlève aucun point. Une réponse correcte rapporte 0.25 points.

1. Le 1^{er} élément d'une liste en Python est indexé par 1.
2. Le module numpy permet de manipuler des tableaux de nombres en Python.
3. L'écriture d'une fonction en R est identifiable par le mot « function ». Il en est de même en Python.
4. Une fonction de Python ou de R qui ne contient pas le mot « return » est appelée procédure. Cela signifie que l'utilisateur ne peut pas récupérer le(s) objet(s) créé(s) en sortie par ladite fonction.



Vous êtes statisticien(ne) à Disneyland Paris. Une nouvelle attraction a été installée dans le parc au 1^{er} Janvier 2018. Vous êtes chargé(e) d'analyser son fonctionnement et son chiffre d'affaires.

Partie 1 : Construire la table de données (7 points)

1. Vérifiez que la table *disney* contient le bon nombre de gentils et de méchants.
2. Écrivez une fonction *Vert(nom)* qui renvoie le nombre de points associés à un nom de personnage Disney si la roue s'arrête sur le vert. Vérifiez ensuite que le nom « Crochet » vaut 72 points.
3. Écrivez une fonction *Jaune(nom)* qui renvoie le nombre de points associés à un nom de personnage Disney si la roue s'arrête sur le jaune. Vérifiez ensuite que le nom « Pinocchio » vaut 34 points.
4. Écrivez une fonction *Bleu(nom)* qui renvoie le nombre de points associés à un nom de personnage Disney si la roue s'arrête sur le bleu. Vérifiez ensuite que le nom « Lady » vaut 12 points.
5. Appliquez l'une des fonctions précédentes sur le vecteur *Personnage* de la table *disney*. Commentez.
6. Écrivez une fonction *g(f,v)* qui applique une fonction *f* sur chacun des éléments d'un vecteur *v* et retourne le vecteur obtenu.
7. En déduire le code à saisir pour construire un data.frame qui associe à chaque personnage le nombre de points qu'il vaut selon les couleurs de la roue.



Partie 2 : Analyse de la table de données (6 points)

1. Importez la table *data* fournie avec l'énoncé.
2. Quelle est la probabilité que Merlin perde sachant que la roue s'arrête sur le vert et que le joueur pioche *Taram* ?
3. Écrivez une fonction *Probabilite(couleur, personnage)* qui renvoie la probabilité théorique que Merlin perde sachant que la roue s'arrête sur *couleur* et que le joueur pioche *personnage*.
4. L'objectif de la question suivante est de simuler une partie d'un joueur contre Merlin et de retrouver la probabilité théorique de la question précédente.
 - (a) Écrivez une fonction *Simulation(couleur, personnageJoueur, personnageMerlin)* qui prend en paramètres la couleur sur laquelle s'arrête la roue ainsi que les noms des personnages du joueur et de Merlin . Elle renvoie 1 si le joueur gagne contre Merlin et 0 sinon.
 - (b) Supposons que pendant une journée les 800 joueurs ont pioché à chaque fois « Maléfique » et que la roue s'est arrêtée sur le vert. Merlin a donc pioché à chaque fois au hasard un personnage gentil de Disney. Simulez les 800 parties et donnez le nombre de fois où Merlin a perdu.
 - (c) En déduire la probabilité empirique que Merlin perde sachant que la roue s'arrête sur le vert et que le joueur pioche « Maléfique » . Comparez avec la probabilité théorique obtenue à la question 2.
5. La table *echantillon* est un échantillon représentatif de l'ensemble des parties qui ont eu lieu en 2018.
 - (a) Combien de fois où la roue s'est-elle arrêtée sur le vert ? le jaune ? le bleu ? Commentez.
 - (b) Quelle est la probabilité empirique que Merlin perde ? Commentez.



Partie 3 : Chiffre d'affaires et innovation (6 points)

1. Le service marketing de Disneyland Paris vous transmet la table de données CA. Vous êtes chargé(e) de l'étudier en vous appuyant sur des indicateurs statistiques vus en cours. L'usage d'autres indicateurs est apprécié dès lors qu'ils s'avèrent pertinents. Vous rédigerez une synthèse compréhensible par un collègue non statisticien. Vous l'illustrerez au moins un graphique. Vous mentionnerez également les informations complémentaires que vous souhaitez obtenir du service marketing. Vous expliquerez alors à quelles fins vous les utiliseriez pour compléter votre analyse.
2. Le service innovation de Disneyland propose un nouveau jeu intitulé « Fantasia Words ». Le joueur et Mickey choisissent chacun un personnage de Disney différent. Il doit contenir au moins 6 lettres. Mickey tire au hasard sans remise une lettre d'un sac dont la composition est donnée par le tableau ci-dessous. Chacun barre la lettre dans le nom de son personnage Disney. Si la lettre apparaît plusieurs fois dans le nom, elle ne peut être barrée qu'une seule fois. Mickey tire ensuite une nouvelle lettre et on réitère l'opération jusqu'à ce que toutes les lettres du nom d'un des personnages soient barrées. Si le premier à barrer toutes les lettres de son personnage est le joueur, il remporte un bon d'achat de 20 euros. Sinon il perd 20 euros qu'il doit remettre à Mickey.
 - (a) Écrire une fonction *Simulation_tirage(nom)* qui renvoie le nombre de lettres à tirer du sac pour barrer toutes les lettres du nom de personnage passé en paramètre.
 - (b) Écrire une fonction *Simulation_Fantasia(nom_Mickey,nom_Joueur)* qui simule le jeu Fantasia Words. Elle prend en paramètre le nom du personnage Disney de chacun des joueurs. Elle renvoie le nom du personnage vainqueur et le nombre de lettres qu'il a été nécessaire de tirer du sac.
 - (c) Testez les deux fonctions précédentes à l'aide de paramètres bien choisis. Commentez.
 - (d) Le jeu est-il équitable ?

Composition du sac

Lettres	Exemplaires
A	9
I	8
N O R S T U	6
L	5
D M	3
B C P F G H V	2
J Q K W X Y Z	1
E	10

Note de lecture : Les lettres N, O, R, S, T et U sont chacune présentes 6 fois dans le sac.



Annexes Partie 1

Flyer de l'attraction

La maison de Merlin : l'attraction qui fait sensation !



Bienvenu dans la maison de Merlin ! Ne misez rien et tentez de remporter 15 euros ! Faites tourner une roue tricolore puis regardez devant vous sur la table. Le sac blanc contient 15 noms de gentils personnages de Disney. Le sac noir contient 15 noms de méchants personnages de Disney. Choisissez un sac et piochez un nom. Merlin pioche un nom dans l'autre sac. La couleur sur laquelle s'est arrêtée la roue donne le nombre de points que vaut chaque lettre de l'alphabet. Le nombre de points associé au nom de chaque personnage pioché est la somme des points associés à chacune des lettres qui le composent.

Si la roue s'arrête sur le :

- vert : la valeur d'une lettre est donnée par sa position dans l'alphabet.
- jaune : les lettres M, I, C, K, E, Y valent 6 points et les autres lettres valent 2 points.
- bleu : les voyelles valent 6 points et les consonnes valent 2 points.



Si vous obtenez strictement plus de points que Merlin, vous remportez un bon d'achat de 15 euros valable à la boutique de Disneyland Paris !



Descriptif de la table disney

disney

Variables :

- **ID** : identifiant disney du personnage
- **Personnage** : nom du personnage
- **Type** : indique si le personnage est gentil (G) ou méchant (M)

Annexes Partie 2

Descriptif de la table data

data	Variables :
	<ul style="list-style-type: none">● ID : identifiant disney du personnage● Personnage : nom du personnage● Type : indique si le personnage est gentil (G) ou méchant (M)● Vert : points que vaut le personnage si la roue s'arrête sur le vert.● Jaune : points que vaut le personnage si la roue s'arrête sur le jaune.● Bleu : points que vaut le personnage si la roue s'arrête sur le bleu.

Descriptif de la table echantillon

echantillon	Variables :
	<ul style="list-style-type: none">● ID : identifiant disney du personnage pioché par le joueur● Points : points marqués par le joueur● Victoire : indique si le joueur a gagné (1) ou perdu (0) contre Merlin.

Annexe Partie 3

Descriptif de la table CA

CA	Variables :
	<ul style="list-style-type: none">● jour : date du jour au format « jour- mois »● maison_2018 : chiffre d'affaires de la maison de Merlin en 2018 (en k€)● boutique_2018 : chiffre d'affaires de la boutique souvenir de Disney-land Paris en 2018 (en k€)● boutique_2017 : chiffre d'affaires de la boutique souvenir de Disney-land Paris en 2017 (en k€)