

Analyse statistique des parties d'échecs en ligne

Emmanuel GUEISSAZ

Quentin FOURNIER

2023-05-05

Introduction

Origine

Le jeu de données utilisé dans cette étude statistique provient du site [www.kaggle.com](https://www.kaggle.com/datasets/ulrikthgepedersen/online-chess-games). Vous pouvez le trouver en suivant le lien suivant : <https://www.kaggle.com/datasets/ulrikthgepedersen/online-chess-games>.

Description

Ce jeu de données contient des informations sur une vaste gamme de parties d'échecs en ligne. Le fichier contient les informations de 20058 parties dont l'individu est la partie.

Les échecs sont un jeu de plateau stratégique pour deux joueurs. Il se joue sur un plateau divisé en cases carrées et composé de 64 cases au total, alternant entre les couleurs claires (généralement blanc) et foncées (généralement noir). Chaque joueur dispose d'un ensemble de 16 pièces, comprenant un roi, une dame, deux tours, deux fous, deux cavaliers et huit pions.

L'objectif du jeu est de mettre le roi de l'adversaire en échec et mat, ce qui signifie qu'il est attaqué sans possibilité d'échapper à la capture. Les joueurs déplacent leurs pièces de manière alternée, en utilisant des stratégies tactiques et des mouvements spécifiques pour attaquer les pièces adverses, défendre les leurs et contrôler le plateau.

Chaque pièce a des règles de mouvement spécifiques. Par exemple, la tour se déplace horizontalement ou verticalement sur le plateau, le fou se déplace en diagonale, le cavalier effectue des mouvements en forme de L, tandis que la reine peut se déplacer dans toutes les directions sur le plateau.

Le jeu d'échecs est apprécié à la fois comme un loisir et comme un sport intellectuel. Il exige une réflexion stratégique, la prévision des coups de l'adversaire et la capacité de planifier des séquences de coups. Les échecs sont joués à différents niveaux de compétence, du niveau amateur au niveau professionnel, avec des tournois organisés à l'échelle mondiale.

Dans le cadre de cette analyse statistique des parties d'échecs en ligne, nous allons examiner les données pour comprendre les schémas de jeu, les performances des joueurs et les facteurs qui peuvent influencer les résultats des parties.

Variables quantitatives

Nombre de tours d'une partie

- Cette variable discrète représente le nombre de tours joués lors d'une partie d'échecs en ligne.
- Elle est mesurée en termes de la quantité d'échanges de coups entre les joueurs.
- Cette variable permet d'évaluer la durée et la complexité d'une partie.

Classement du vainqueur

- Cette variable continue indique le classement du vainqueur d'une partie d'échecs en ligne.
- Plus elle est haute, plus le joueur est censé être bon.
- Le classement permet de déterminer la performance des joueurs.

Variables qualitatives

Couleur du vainqueur

- Cette variable représente la couleur (white (blanc) ou black (noir)) du joueur qui remporte la partie.

Partie classée ou non

- Cette variable indique si une partie est classée ou non.
- Une partie classée est généralement enregistrée dans un système de classement officiel, tandis qu'une partie non classée est jouée à des fins de pratique ou de divertissement.

Analyses

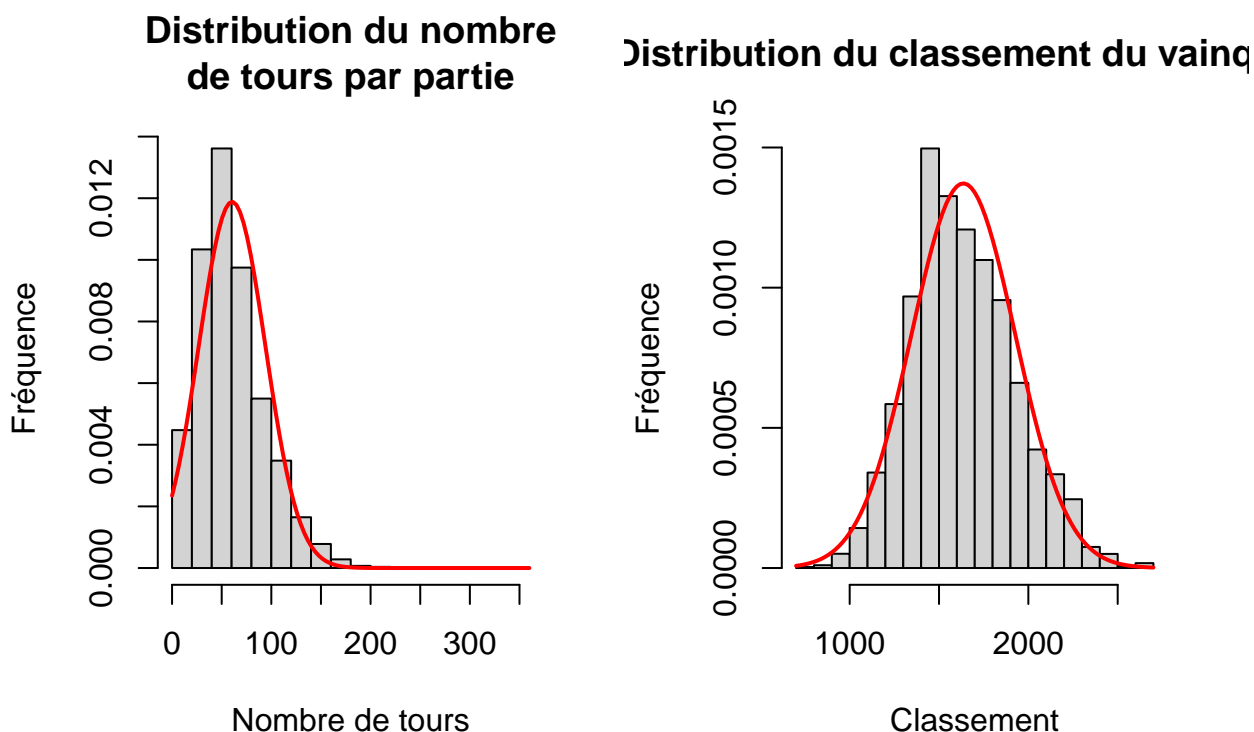
Analyse univariée

Exploration statistique

Les analyses statistiques effectuées sur les parties d'échecs en ligne révèlent les résultats suivants :

- La moyenne du nombre de tours réalisés dans chaque partie est de 60.47 tours.
- La variance du nombre de tours réalisés dans chaque partie est de 1126.98.
- La moyenne des classements des parties est de 1636.35.
- La variance des rangs est de 8.454669×10^4 .

Distribution normale



En examinant l'histogramme du nombre de tours par partie et la courbe rouge représentant la distribution normale, on peut observer une certaine similarité entre les deux. La courbe rouge semble suivre approximativement la forme de l'histogramme, suggérant une possible adéquation à une distribution normale.

Tests de confiance

Les tests de confiance ont été réalisés pour évaluer les intervalles de confiance des variables quantitatives étudiées. Voici les résultats obtenus :

- Nombre de tours : L'intervalle de confiance à 95 % pour le nombre de tours réalisés dans chaque partie est de 60.0013877 à 60.9306095. Cela signifie qu'avec un niveau de confiance de 95 %, on peut estimer que la moyenne réelle du nombre de tours dans la population se situe entre 60.0013877 et 60.9306095.
- Classement du vainqueur : L'intervalle de confiance à 95 % pour le classement du vainqueur est de 1632.3284844 à 1640.37687. Cela signifie qu'avec un niveau de confiance de 95 %, on peut estimer que la moyenne réelle du classement du vainqueur dans la population se situe entre 1632.3284844 et 1640.37687.

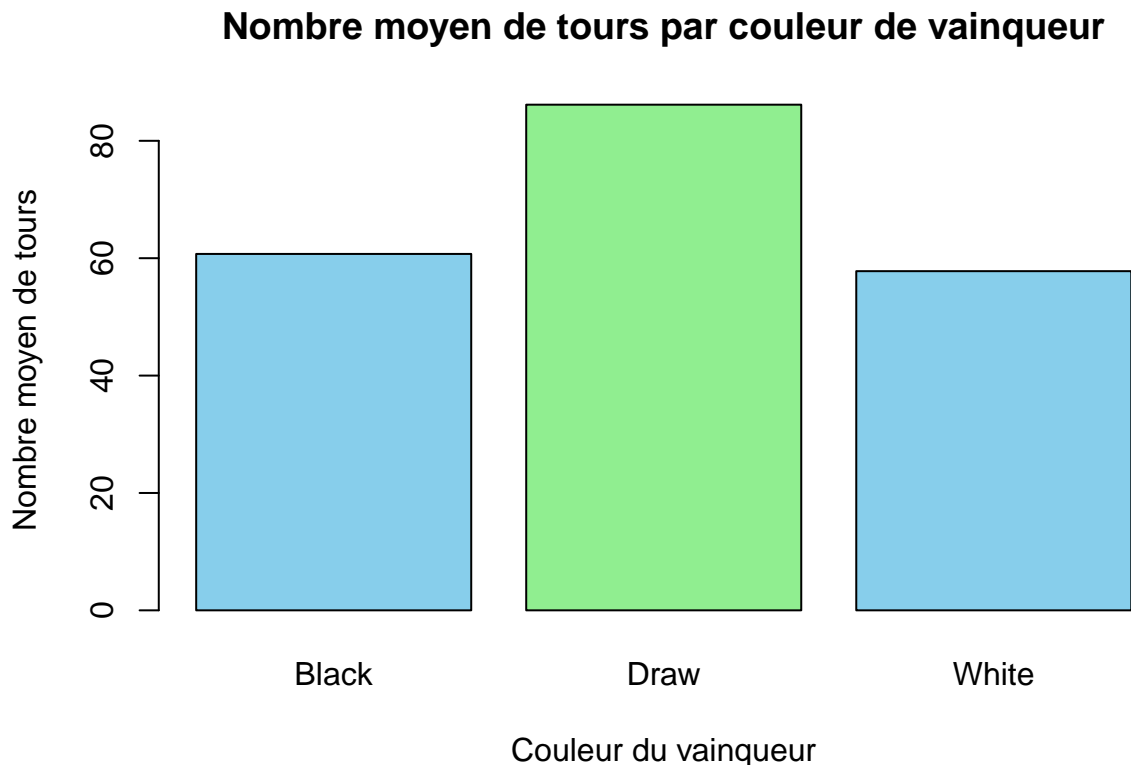
Analyse de la proportion de vainqueur par couleur

Les résultats de l'analyse indiquent que l'intervalle de confiance à 95% pour la proportion de vainqueurs blancs est de 0.49 à 0.51, tandis que l'intervalle de confiance à 95% pour la proportion de vainqueurs noirs est de 0.45 à 0.46. On en déduit donc que les blancs gagnent plus souvent les parties d'échec en ligne.

Analyse multivarié

Nombres de tours par rapport à la couleur du vainqueur

```
mean_turns <- aggregate(turns ~ winner, data = chess_games_datas, FUN = mean)
barplot(mean_turns$turns, names.arg = mean_turns$winner, xlab = "Couleur du vainqueur",
        ylab = "Nombre moyen de tours", main = "Nombre moyen de tours par couleur de vainqueur",
        col = c("skyblue", "lightgreen"))
```

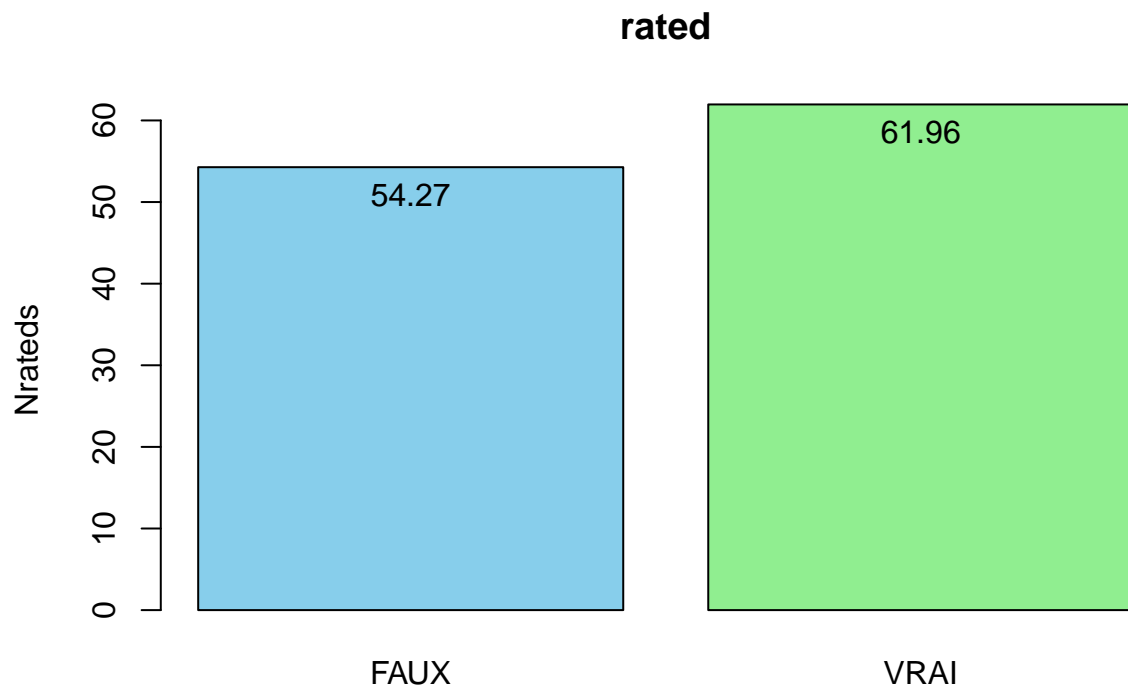


```
b_turns <- aggregate(turns ~ rated, data = chess_games_datas, FUN = mean)
# Arrondir les valeurs à deux chiffres après la virgule
b_turns$turns <- round(b_turns$turns, 2)

barplot(b_turns$turns, names.arg = b_turns$rated, xlab = "rated",
        ylab = "Nrateds", main = "rated",
        col = c("skyblue", "lightgreen"))

# Calculer la position du texte à l'intérieur de chaque barre
text_pos <- barplot(b_turns$turns, plot = FALSE)

# Ajouter le texte arrondi à l'intérieur des barres
text(x = text_pos, y = b_turns$turns, labels = b_turns$turns, pos = 1, offset = 0.5)
```



```

rated
c_turns <- aggregate(winClassement ~ winner, data = chess_games_datas, FUN = mean)
barplot(c_turns$winClassement, names.arg = c_turns$winner, xlab = "xxxxxx",
        ylab = "xxxxxx", main = "Nxxxxxx",
        col = c("skyblue", "lightgreen"))

```

