

TP1 : 1 échantillon

2024-09-24

Fonctions R utiles pour ce TP :

Dans ce TP, vous aurez besoin des fonctions suivantes pour réaliser les tests statistiques :

- **t.test()** : Effectue un test de Student pour comparer une moyenne observée à une moyenne théorique.
- **power.t.test()** : Calcule la puissance d'un test de Student
- **binom.test()** : Réalise un test exact sur le paramètre d'une loi de Bernoulli (utile pour tester une proportion).
- **prop.test()** : Effectue un test approximatif (via une approximation normale) sur le paramètre d'une loi de Bernoulli.

Pour réaliser les tests “à la main”, vous aurez également besoin des fonctions suivantes :

- **pt(x, n)** : Renvoie la probabilité qu'une variable de Student à n degrés de liberté soit inférieure à $t = x$. Utile pour calculer la p-value lors d'un test t.
- **qt(alpha, n)** : Renvoie le quantile α pour une loi de Student à n degrés de liberté. Utile pour trouver la valeur critique dans un test t.

Conseil :

N'hésitez pas à consulter l'aide de ces fonctions avec `?t.test`, `?power.t.test`, `?binom.test`, etc., pour en savoir plus sur leur utilisation et leurs options.

Ligue 1:

Nous avons une base de données nommée `ligue1_2425` qui donne les résultats des matchs de la saison 2024-2025 de la Ligue 1. À ce jour, 5 journées ont eu lieu (chaque équipe a donc joué 5 matchs).

La base de données contient 4 colonnes :

- `Dom` (l'équipe jouant à domicile),
- `Dom_but` (nombre de buts marqués par l'équipe à domicile),
- `Ext` (l'équipe jouant à l'extérieur),
- `Ext_but` (nombre de buts marqués par l'équipe à l'extérieur).

Nombre de but (PSG)

L'entraîneur du PSG pense que son équipe est devenue significativement meilleure que la saison dernière et qu'à partir du début de cette saison (août 2024), elle marquera en moyenne plus de buts par match de Ligue 1.

L'année dernière, le PSG avait marqué en moyenne 2.38 buts par match. Pour l'instant, nous n'avons que les résultats des 5 premières journées de Ligue 1, ce qui correspond à 5 matchs du PSG.

Soit X_i le nombre de buts marqués par le PSG pendant le match de la journée i . Nous supposons que les X_i sont i.i.d (indépendants et identiquement distribués), avec $E(X_i) = \mu$ et $V(X_i) = \sigma^2$.

Mieux que l'année dernière ?

1. Représentez graphiquement le nombre de buts marqués par le PSG au cours de ses matchs de Ligue 1.

```
# lire le fichier de données
load(file = "ligue1_2425.Rdata")

# Pour récupérer le nombre de but du PSG :

but_psg <- ligue1_2425 %>%
  filter(Ext == "PSG" | Dom == "PSG") %>%
  mutate(but_psg = case_when(Dom == "PSG" ~ Dom_but,
                             TRUE ~ Ext_but)) %>%
  pull(but_psg) %>% as.numeric()
```

2. Comme nous n'avons qu'un échantillon de taille $n = 5$, quelle hypothèse supplémentaire devons-nous faire sur les X_i pour pouvoir effectuer un des tests statistiques vu en cours ?
3. En supposant ce qu'il faut (réponse à la question 2), mettez en place un test au niveau de significativité de 5 % pour tester si la moyenne des buts par match du PSG diffère de celle de l'année dernière (2.38 buts). Effectuez ce test en utilisant la fonction adéquate et en le faisant étape par étape (en décrivant la zone de rejet, calculant la p-valeur etc)
4. Pouvons-nous conclure que l'équipe n'est pas meilleure ? Que faudrait-il analyser en plus avant d'arriver à cette conclusion ?

Puissance du test

L'entraîneur pense que la vraie nouvelle moyenne de buts est 3.5 et non 2.38.

5. S'il a raison, quelle est la puissance du test pour détecter cette différence ? Autrement dit, quelle est la probabilité de mettre en évidence, avec nos données, une différence de 1.12 entre la moyenne de l'année dernière et l'espérance actuelle des buts marqués par le PSG ?
6. Actuellement, nous n'avons que 5 matchs, mais d'autres matchs seront joués au fur et à mesure. Comment la puissance du test évolue-t-elle avec l'augmentation du nombre de matchs disponibles (augmente, diminue, reste inchangée) ?
7. Combien de matchs du PSG (après août 2024) faudrait-il pour mettre en évidence cette différence (de 1.12) avec une puissance de 90 % ?
8. Tracez une courbe représentant la puissance du test en fonction du nombre de matchs disponibles.
9. L'entraîneur pense finalement que la vraie nouvelle moyenne de buts est de 3, et non 3.5. Cela change-t-il la puissance du test ? Si oui, comment ? Recalculez-la.
10. Tracez la puissance du test en fonction de l'écart entre l'ancienne moyenne et la nouvelle moyenne. Par exemple, si la nouvelle moyenne est de 3, l'écart est de 0.62; si elle est de 3.5, l'écart est de 1.12 etc. Calculer là pour des écarts allant de 0.2 à 4.

Proba de gagner (PSG)

Finalement, l'entraîneur réfléchit. Ce qui l'intéresse réellement, c'est la probabilité de gagner un match. L'année dernière, le PSG a gagné 64% de ses matchs. L'entraîneur souhaite montrer, avec un risque de 5 %, que cette proportion est significativement plus élevée depuis août 2024. Réalisez le test.

Ligue 1: nombre de but (Toulouse)

L'entraîneur de Toulouse pense que son équipe marque significativement moins de but que l'année dernière. L'année dernière l'équipe marquait 1.23 buts par match en moyenne. En supposant la normalité du nombre

de buts marqués par match, réalisez un test au niveau 5% pour répondre à l'entraîneur.

L'entraîneur pense que la nouvelle espérance de but de l'équipe de Toulouse est de 0.9. Calculer la puissance du test pour mettre en évidence une telle différence.