

# Template for preparing your research report submission to PNAS using RMarkdown

Pénélope Robert<sup>a</sup>, Édouard Nadon-Beaumier<sup>a</sup>, Alexis Matte<sup>a</sup>, and Nadia Tardy<sup>a</sup>

<sup>a</sup>Université de Sherbrooke, Département de bio, 2500 Boulevard de l'Université, Sherbrooke, Québec, G1V 0A9

This manuscript was compiled on April 29, 2022

**Please provide an abstract of no more than 250 words in a single paragraph. Abstracts should explain to the general reader the major contributions of the article. References in the abstract must be cited in full within the abstract itself and cited in the text.**

La notion du réseau écologique d'un écosystème est souvent déformé et confondu avec celle du réseau trophique. Nous avons régulièrement tendance à imaginer les interactions entre les différents espèces avec une belle pyramide qui retrace les réseaux trophiques et la chaîne alimentaire en plaçant le plus important prédateur au sommet du prisme. En réalité, un réseau écologique inclut toutes les multitudes d'interactions possibles entre les espèces comme le mutualisme, la compétition, le commensalisme et le parasitisme. Évidemment, les réseaux trophiques ont une grande influence sur les réseaux écologiques, mais ce dernier ne devrait pas être représenté par une pyramide. Il devrait plutôt être décrit comme une toile ("web") avec les espèces clés de l'écosystème au centre, c'est-à-dire les espèces qui ont le plus grand nombre de connexions interspécifiques. C'est dans ce contexte que nous avons voulu comparer le réseau de collaboration entre les étudiants en écologie et de celles des réseaux écologiques. Afin d'approfondir notre réflexion, nous avons voulu vérifier si le nombre de collaborations par étudiants de BIO500 sont semblables ou bien s'il y a des "étudiants clés," c'est-à-dire qui possèdent le plus de collaborations différentes au même titre qu'une espèce clé.

Avant même de se pencher sur nos questions, il a fallu que chacun des étudiants de BIO500 recense l'entièreté des étudiants avec qui il ou elle a collaboré dans un travail d'équipe dans leur parcours académique en biologie à l'Université de Sherbrooke. Une fois les données recueillies et traitées, nous avons toutes les informations en main pour commencer notre analyse. Nous avons en premier lieu émis une hypothèse. Selon nous, le nombre de collaboration par personne devrait être environ semblable d'un étudiant à l'autre puisque le parcours académique et les cours sont relativement les mêmes. Nous nous attendons donc à ce que le nombre de collaborations par étudiant suive une distribution normale. Si ce n'est pas le cas, nous nous pencherons sur l'hypothèse inverse des "étudiants clés." L'objectif sera de les identifier et de déterminer la cause de leur haut nombre de collaborateurs.

Nous avons en premier lieu une analyse visuelle et graphique et en deuxième lieu une analyse statistique de notre distribution. Les deux analyses, comme les traitements de données ci-haut, ont été effectués sur le logiciel R version 4.0.3. L'analyse visuelle est représentée par un histogramme, voici 4 résultats possibles ainsi que leur signification:

*Une courbe symétrique et mince*

C'est à quoi nous nous attendions. La variance entre le nombre de collaboration entre les élèves de BIO500 est faible

et très concentré autour de la moyenne. Les étudiants ont donc eu un parcours académique semblable.

*Une courbe symétrique et large*

Cette courbe corrobore également notre hypothèse: une distribution normale. Par contre, la variation entre les élèves est plus grande qu'attendu. Il y a une petite divergence entre les parcours, mais qui sans débalancer les données.

*Une courbe asymétrique à gauche*

Ce genre de distribution non normal indique que la majorité des élèves se retrouve malgré tout proche de la moyenne. Cependant, un certain nombre non négligeable d'élèves ont plus que collaboré et qu'aucun ou très peu d'élèves ont peu collaboré. Cette courbe nous indique qu'il y a plusieurs "étudiants clés" plus ou moins importants.

*Une courbe asymétrique avec des données aberrantes vers la droite*

On observe une distribution normale, mais avec ce qui semble être des données aberrantes. Cela représente précisément l'autre hypothèse "d'étudiants clés" peu nombreux et très importants. Leur divergence distinctive pourrait signifier une grande différence de leur parcours.

L'analyse statistique est effectuée par le test de Shapiro-Wilk dans lequel les hypothèses sont les suivantes:

- Hypothèse nulle (H0): les données ont une distribution normale
- Hypothèse alternative (H1): les données n'ont pas une distribution normale

La commande `shapiro.test()` est facile d'utilisation et sa sortie nous informe sur la valeur de W et le p-value. C'est seulement le p-value qui nous intéresse. Si le p-value est plus petit que 0.05, on rejette l'hypothèse nulle: les données sont non normales. Si le p-value est plus grand que 0,05, on ne rejette pas l'hypothèse nulle: les données ont donc une distribution normale.

Significance Statement