

TD 1. Application & utilisation des fonctions statistiques implantées dans R



Statistiques descriptives et inférentielles

Le fichier Excel représente la taille de 140 (type A) et 32 (type B) graminées. Les fichiers « graminA.txt » et « graminB.txt » sont les équivalents en format txt.

Ouvrir le fichier « R pour les statophobes.pdf » puis, tout en suivant la lecture du pdf, réaliser les étapes suivantes :

1. Indiquer à R le répertoire dans lequel se trouvent les données (c'est-à-dire les fichiers « graminA.txt » et « graminB.txt »). Fichier/Changer le répertoire courant...
2. Ouvrir les fichiers « graminA.txt » et « graminB.txt » sous les noms PA et PB (commande *read.table*).

```
PA <-read.table("graminA.txt", sep="\t",dec=",",header=TRUE)
```

```
PB <-read.table("graminB.txt", sep="\t",dec=",",header=TRUE)
```

Pour obtenir des informations sur la fonction *read.table* a tout moment : `?read.table`

3. Calculer la moyenne, la variance, l'écart type, la longueur, l'erreur standard des EL de PA (commandes *mean*, *var*, *sd*, *length*).
4. Créer une table par type (commande *table*).
5. Créer un sommaire du tableau PA (commande *summary*).
6. Dessiner un graphique de comparaison de EL pour PA et PB (commande *stripchart*). Profitez-en pour changer le nom des axes : "mes différents types" en X et "hauteur des gramin en cm" en Y, (utiliser l'aide : `?stripchart`).

7. Représenter sous forme d'histogrammes la distribution de la variable EL pour PA (commande *hist*). Modifier le nombre de classes, et tracer la densité de probabilité (?*hist* pour plus d'information).
8. Tracer $EL=f(HL)$ pour PA (commande *plot*). Modifier le type de symbole , leur taille et leur couleur. Ajouter le nom des axes et un nom pour le graphique.
9. Ajouter un point aux coordonnées (40,35) avec un symbole, une couleur et une taille différente (commande *points*)
10. Compléter le diagramme avec les moyennes des X et les moyennes des Y (commande *abline*).
11. Dessiner le camembert pour les sous-types de PA (commande *pie*).
12. Dessiner la distribution des sous-types sous la forme d'un diagramme en baton (commande *barplot*). Ajouter le nom des axes et un titre.
13. Calculer l'intervalle de confiance de la moyenne des EL de PA (commande *t.test*).
14. Calculer l'intervalle de confiance de la médiane des EL de PA pour les 10 premiers individus (commande *wilcox.test*).
15. Calculer l'intervalle de confiance de la moyenne des EL de PA pour les 10 premiers individus par bootstrap (commande *sample*).
16. Comparer graphiquement les EL de PA et de PB (commande *boxplot*).
17. Comparer à l'aide d'un test statistique les EL de PA et de PB (commande *t.test*).
18. Comparer à l'aide d'un test statistique les EL et ELbis de PA sachant qu'il s'agit de la même mesure par deux opérateurs différents (commande *t.test*).
19. Comparer à l'aide d'un test statistique les EL et ELbis de PA sachant qu'il s'agit de la même mesure par deux opérateurs différents pour les seuls 10 premiers échantillons (commande *wilcox.test*).
20. Comparer à l'aide d'un test statistique les EL de PA et de PB pour les 10 premiers échantillons de chaque groupe (commande *wilcox.test*).
21. Comparer le pourcentage du sous-type G dans PA puis PB face à la valeur théorique 70% (commande *binom.test*).
22. Comparer les distributions théoriques des sous-types de PA face aux valeurs théoriques 60%, 20%, 20% (commande *chisq.test*).
23. Comparer entre elles les distributions des sous-types recensés dans PA et PB (commande *chisq.test* et *fisher.test*).
24. Calculer le coefficient de corrélation entre EL et HL pour PA (commande *cor*).
25. Est-il significatif (commande *cor.test*) ?
26. Déterminer les paramètres de la régression au sens des moindres carrés (commande *lm*).
27. Utiliser l'équivalent non-paramétrique sans faire l'hypothèse de la bi-normalité des données (commande *cor.test, option method= « spearman »*).