

L'objectif de ce document est de présenter une très courte introduction au logiciel R (via l'interface RStudio), de sorte que des étudiants découvrant R puissent en quelques heures se familiariser avec ce logiciel et être opérationnels par la suite pour réaliser des exercices de travaux pratiques accompagnant un cours de Probabilités et de Statistiques.

Ce TP veut vous permettre de réaliser des graphes de Statistiques descriptives simples et d'utiliser l'aide du logiciel RStudio sur des modèles simples.

Les fichiers de données sont joints au TP.

Exercice 1

Le but de cet exercice est d'adapter le dimensionnement des cabines de camions en fonctions de la taille du conducteur.

Sur les 80 conducteurs étudiés, la taille en cm est présentée dans le tableau suivant :

Taille en cm des conducteurs							
151	163	165	168	170	173	175	179
152	163	165	168	171	173	175	179
157	164	166	169	172	174	176	180
159	164	166	169	172	174	176	180
160	164	167	169	172	174	177	181
160	164	167	169	172	174	177	181
160	164	167	169	173	174	177	181
161	164	168	169	173	175	178	181
161	165	168	170	173	175	178	186
162	165	168	170	173	175	179	187

Pour l'ensemble des questions suivantes, donner les commandes R permettant de répondre aux questions suivantes.

1. Quelle est la population statistique observée ? La taille de l'échantillon ?
2. Quel est le caractère statistique observé et sa nature ?
3. Dépouiller les données suivant une distribution des fréquences dont la limite inférieure est 150 cm (inclus) et dont la largeur de chaque classe est de 5. On indiquera les fréquences de chaque classe et on représentera graphiquement l'histogramme associé à cette distribution.
4. Tracer la courbe de la fonction de répartition empirique (diagramme des fréquences cumulées) de la distribution de fréquences de la question précédente.
5. Calculer la médiane, l'intervalle interquartile et interdécile.
6. Si on élimine les 5% plus grands et les 5% les plus petits des conducteurs, quel intervalle de taille standard retiendra-t-on ?

Exercice 2

On s'intéresse à une maladie des feuilles des bourgeonnements en forme de pépins sur les feuilles. On observe la distribution des bourgeonnements par feuille.

Les résultats sont les suivants :

Nombre de bourgeonnements par feuille	0	1	2	3	4	5	6	7	8	9	10
Nombre de feuilles contaminées	482	133	46	24	6	5	2	1	0	1	1

Pour l'ensemble des questions suivantes, donner les commandes R permettant de répondre aux questions suivantes.

1. Quelle est la population statistique observée ? La taille de l'échantillon ?
2. Quel est le caractère statistique observé et sa nature ?
3. Représenter graphiquement les données ?
4. Calculer la moyenne et l'écart-type corrigé empirique du nombre de bourgeonnements par feuille.

Exercice 3

Le tableau ci-dessous donne pour 4 professions le nombre de diplômés entre 1970 et 1984.

Profession	Nombre de diplômés
Médecins	106759
Pharmaciens	43924
Dentistes	25965
Sage-femmes	8215
Total	184863

Pour l'ensemble des questions suivantes, donner les commandes R permettant de répondre aux questions suivantes.

1. Quelle est la population statistique observée ? La taille de l'échantillon ?
2. Quel est le caractère statistique observé et sa nature ?
3. Représenter graphiquement les données avec un diagramme circulaire.

Exercice 4

Le tableau suivant donne la répartition des exploitations agricoles suivant leur SAU (surface agricole utilisée) en France pour les années 1979 et 2010.

Superficie (en ha)	Nombre d'exploitations en milliers	
	Année 1979	Année 2010
[0;20[766,7	211,4
[20;50[347,3	87,9
[50;100[114,1	97,4
[100;200[29	72,6
[200;250[5,5	20,6

Pour l'ensemble des questions suivantes, donner les commandes R permettant de répondre aux questions suivantes.

1. Quelles sont les populations observées ?
2. Quel est le caractère statistique observé, sa nature ?
3. Tracer les histogrammes des deux distributions statistiques et donner les commandes R permettant de les obtenir.
4. Réaliser un box-plot comparatif en approchant les données de chaque classe par leur milieu et commenter l'évolution.

Exercice 5

Des ingénieurs s'intéressent à la durée de vie de composants de 3 types.

Composant bleu	3,4	3,5	3,8	3,9	4	4,1	4,1	4,1	4,2	4,2	4,4	4,4	4,5				
Composant jaune	4,1	4,5	4,5	4,5	4,6	4,6	4,8	4,9	5	5	5,1	5,2	5,3	5,3	5,4	5,6	5,9
Composant vert	4,8	5,1	5,2	5,3	5,3	5,4	5,5	5,5	5,7	5,7	6,1	6,2	6,7	6,7	6,7		

Pour l'ensemble des questions suivantes, donner les commandes R permettant de répondre aux questions suivantes.

1. Quelles sont les populations observées ?
2. Quel est le caractère statistique observé, sa nature ?
3. Calculer les moyennes et les écart-types corrigés du temps écoulé pour chaque composant ?
4. Tracer le diagramme des moyennes avec des barres d'erreurs d'un écart-type. Commenter.
5. Réaliser un box-plot comparatif des 3 composants.