

Compte-rendu Statistiques – TP1

Statistiques descriptives et visualisation d'information

Université de Tours

Nom, Prénom

10 November 2020

Ceci est un titre de section

Ceci est un sous-titre

Ceci est un document R Markdown. Markdown est un langage permettant de générer facilement des documents HTML, PDF et MS Word.

Quelques commandes préalables pour vous initier. Pour faire des listes de points:

- Ceci est une phrase en gras

On peut mettre des choses en dessous.

- Ceci est une phrase en italique

1. Ceci est une phrase type code avec un numéro devant
2. Pour faire de jolies formules $(a + b)^2 = a^2 + b^2 + 2ab$ ou bien en hors ligne:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} = \frac{\pi^2}{6}$$

Lorsque vous cliquez sur le bouton **Knit** (avec la pelote de laine bleue), un document est généré automatiquement. Essayez !

```
# Ici vous pouvez écrire du code R à exécuter
```

```
X <- 42
```

```
X
```

```
## [1] 42
```

Rendu des questions

Exercice 1

On considère le data frame généré par le code suivant :

```
T <- data.frame(V1=rep(c(1, NA), 3), V2=c(seq(1,5),NA))
T
```

```
##   V1 V2
## 1  1  1
## 2 NA  2
## 3  1  3
## 4 NA  4
## 5  1  5
## 6 NA NA
```

1. Modifier la valeur située ligne 3, colonne 1 de T par la valeur 10.

```
# Insérez ici votre code...
```

2. Dans la colonne 2, remplacer toutes les valeurs ≥ 4 par la valeur 20. On pourra utiliser la commande `ifelse(phi, valT, valF)` qui rend la valeur `valT` si la condition logique `phi` est vérifiée et `valF` sinon.

```
# Insérez ici votre code...
```

3. On peut détecter si une valeur possède la valeur `NA` grâce à la commande `is.na()`.

Remplacer toutes les valeurs `NA` de T par la valeur 0.

```
# Insérez ici votre code...
```

4. Ajouter une nouvelle colonne à T qui est la moyenne des deux colonnes V_1 et V_2 .

```
# Insérez ici votre code...
```

5. Ajouter au dataframe `poke` la nouvelle colonne `base_stats` qui est correspond à la somme des différentes statistiques du pokémon, c.a.d des variables `hp`, `atk`, `...`, `spd`.

```
# Insérez ici votre code...
```

Exercice 2

1. Dresser l'histogramme de la variable `capt_rate`. Cette variable suit-elle une loi normale ? Expliquer votre réponse.

```
# Insérez ici votre code...
```

2. Dresser les graphiques des fréquences cumulées croissantes et la boîte à moustaches de `capt_rate`. Que constatez-vous ? Quelles sont les similitudes entre ces deux graphiques ?

```
# Insérez ici votre code...
```

Exercice 3

Code pour le graphique:

```
# Insérez ici votre code...
```

Exercice 4

1. Quel est le coefficient de corrélation entre les variables `sp_atk` et `base_stats`. Pouvez-vous expliquer pourquoi ce score est élevé ?

Votre réponse à la question...

2. Donner le code R permettant de générer le graphique ci-dessous.

```
# Insérez ici votre code...
```

Exercice 5

Supposons que l'on veuille visualiser, pour chaque type, la proportion de chaque autre type de Pokémon. Par exemple, pour le type `grass`, on souhaite connaître la proportion de pokémons de types `grass` / `t` où `t` est un type quelconque.

Plus formellement, on souhaite visualiser la proportion de chaque combinaison de types (t_1, t_2) .

1. Créer un dataframe doté de deux colonnes `t1` et `t2` où chaque ligne correspond à un pokémon et où `t1` correspond au type primaire du pokémon et `t2` son type secondaire. Dans le cas où `type2 = NA`, `t2` prendra la valeur de `type1`. Par exemple, si on a `type1 = grass` et `type2 = NA`, on affectera à `t2` également la valeur `grass` ce qui correspond à dire que le pokémon est de type “plante” pur.

```
# Insérez ici votre code...
```

2. Donner le code R permettant de générer le graphique des proportions des types secondaires pour chaque type de pokémon.

```
# Insérez ici votre code...
```

Exercice 6

On souhaite vérifier l'hypothèse que les pokémons roses sont davantage genrés comme femelle (`percentage_male < 50`) que les autres pokémons.

1. Dresser l'histogramme de la variable `percentage_male` et proposer une méthode de discrétisation pour cette variable.

```
# Insérez ici votre code...
```

2. Créer une nouvelle variable `is_pink` qui vaut 1 si le pokémon est rose et 0 sinon.

```
# Insérez ici votre code...
```

3. Proposer une méthode permettant vérifier l'hypothèse puis conclure.

Détaillez ici votre méthode...

```
# Insérez ici votre code...
```

Exercice 7

1. Code pour le graphique:

```
# Insérez ici votre code...
```

Vos conclusions d'analyse et les défauts de visualisation que présentent ce graphique.

2. Quelles analyses peut-on en tirer ? Quel est son défaut majeur ?
3. Reprendre le graphique précédent et faire figurer uniquement les courbes des types `dark` et `fairy`. On ajoutera également un ruban pour générer les écarts-types.

```
# Insérez ici votre code...
```

Exercice 8

1. Quel phénomène selon vous peut expliquer l'apparition de mélanges gaussiens ?

Votre réponse à la question...

2. Déterminer une méthodologie d'analyse, que vous détaillerez, permettant d'expliquer la présence des deux gaussiennes sur le graphique précédent.

Présentez ici votre méthodologie d'analyse

```
# Insérez ici votre code...
```

3. Conclure votre analyse et expliquer à quoi correspond, en terme d'individus, chacune de ces gaussiennes.

Votre réponse à la question...

Exerice 9

```
library("corrplot")
```

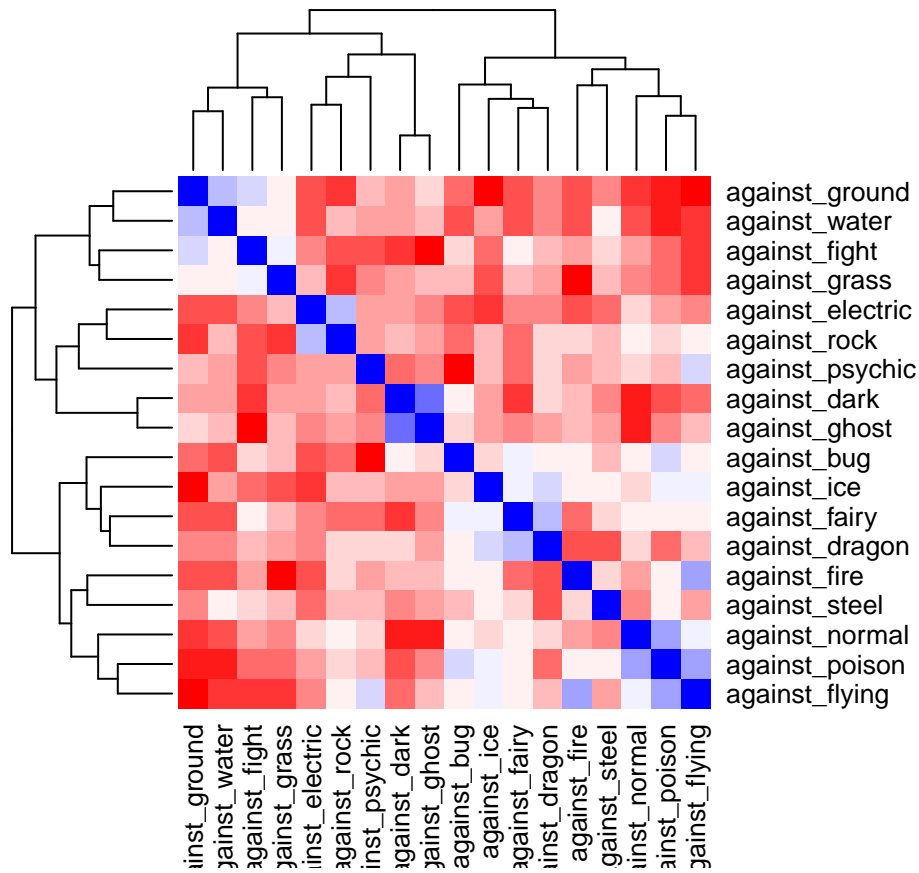
```
## corrplot 0.84 loaded
```

```
source("http://www.sthda.com/upload/rquery_cormat.r")
```

```
num_var <- poke[,24:ncol(poke)]
```

```
col<- colorRampPalette(c("red", "white", "blue"))(20)
```

```
rquery.cormat(num_var, graphType="heatmap", col = col)
```



Donner une analyse du dernier diagramme (carte de chaleur et dendrogramme).

Votre réponse à la question...

Exercice 10

Proposer 3 scénarios d'analyse que vous détaillerez et illustrerez selon votre choix. En particulier, vous préciserez et justifierez :

- Votre scénario d'analyse (but et démarche),
- Les outils statistiques que vous mettrez en oeuvre,
- Votre choix de méthode de visualisation.
- Analyse et conclusion quant à vos observations.

Scénario 1

Scénario 2

Scénario 3