

EXERCICE2

kpakou

2023-04-26

warning permet de ne pas afficher les messages d'erreurs `eval=TRUE` affiche les resultats **analyse pour une variable quantitative**

```
d.var.quant<-function(var){
  moyenne <- mean(var)#moyenne
  mediane <- median(var)#mediane
  mode <- names(sort(-table(var)))[1]
  cat(paste("La moyenne de 'var' est :", moyenne, "\n"))
  cat(paste("La médiane est :", mediane, "\n"))
  cat(paste("Le mode est :", mode, "\n"))
  #intervalle de confiance
  conf_intervale <- t.test(var)$conf.int
  cat(paste("L'intervalle de confiance à 95% est :", conf_intervale[1], "-", conf_intervale[2], "\n"))
  #Histogramme
  hist(var,main="Hsitogramme ")

  #boxplot
  boxplot(var,main="boxplot ")
  #diagramme en batons
  plot(table(var),main="Diagramme en batons",type="h")
}
```

Analyse pour une variable qualitative

```
d.var.qualit<-function(df_KPAKOU,var){
  # Tendances centrales
  mode <- names(sort(-table(df_KPAKOU[[var]])))[1]

  cat(paste0("Mode: ", mode, "\n\n"))
  #Diagramme en barre
  barplot(table(df_KPAKOU$var), main="Diagramme en barres")
  #diagramme circulaire
  pie(table(df_KPAKOU$var), main="Diagramme circulaire")
  #histogramme
  hist(df_KPAKOU$var, main="Histogramme ")
  #boxplot
  boxplot(df_KPAKOU$var, main="boxplot ")
}
```

Application pour les variables quantitatives

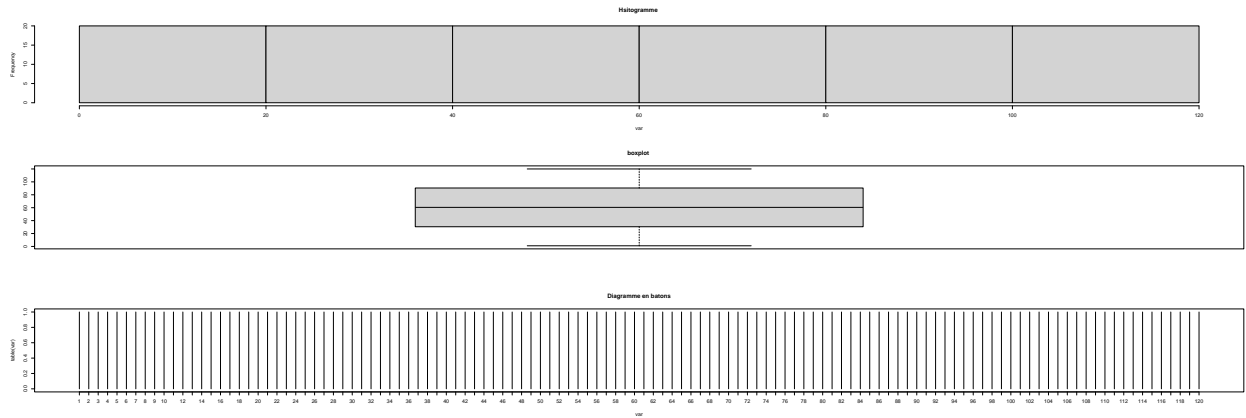
```
apply(df_KPAKOU[,1:5],2,d.var.quant)
```

La moyenne de 'var' est : 60.5

La médiane est : 60.5

Le mode est : 1

L'intervalle de confiance à 95% est : 54.2123384181789 - 66.7876615818211

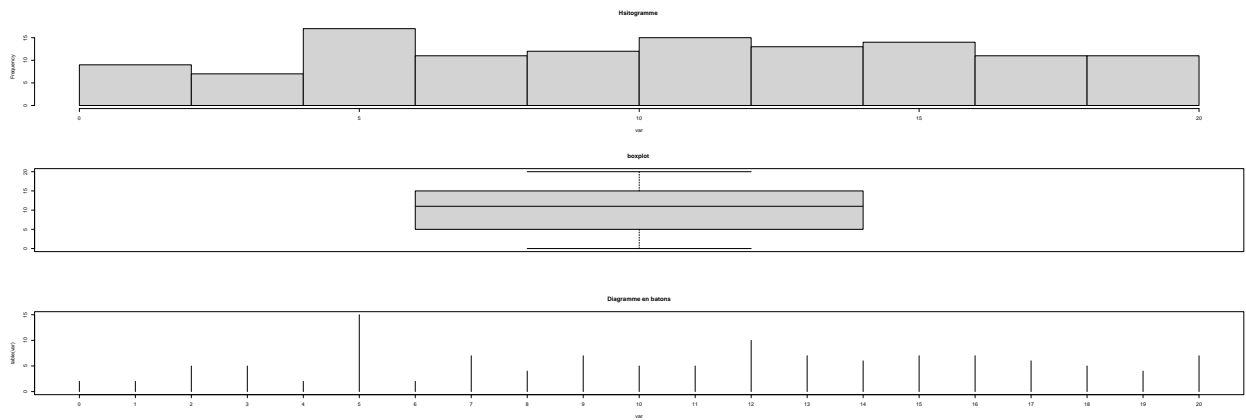


La moyenne de 'var' est : 10.7583333333333

La médiane est : 11

Le mode est : 5

L'intervalle de confiance à 95% est : 9.75150982115038 - 11.7651568455163

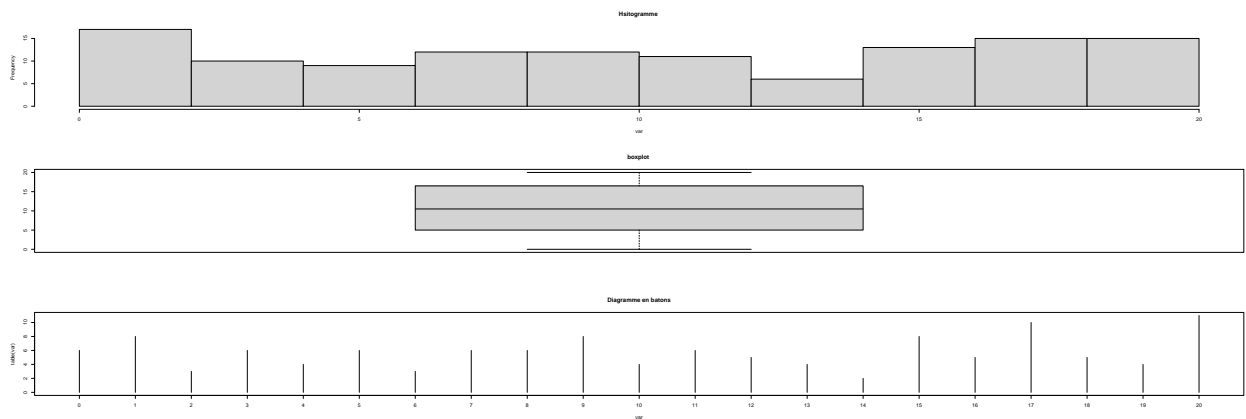


La moyenne de 'var' est : 10.5

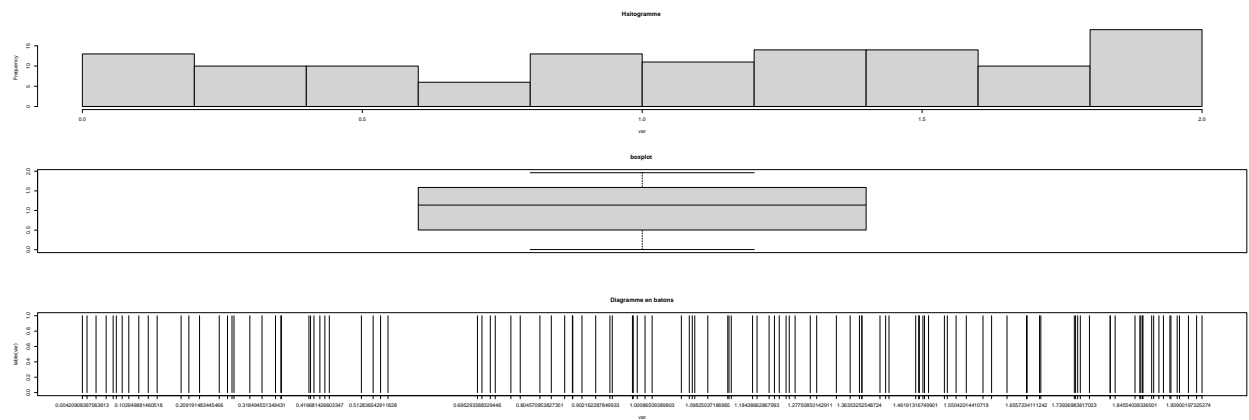
La médiane est : 10.5

Le mode est : 20

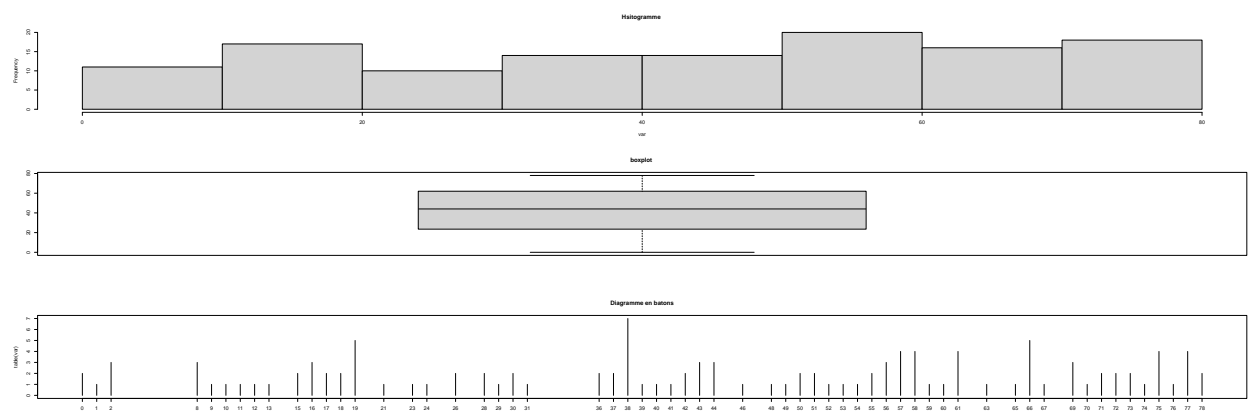
L'intervalle de confiance à 95% est : 9.34722201562451 - 11.6527779843755



La moyenne de 'var' est : 1.07191137475505
 La médiane est : 1.13731596129946
 Le mode est : 0.00420908397063613
 L'intervalle de confiance à 95% est : 0.962475525761773 - 1.18134722374832



La moyenne de 'var' est : 43.7916666666667
 La médiane est : 44
 Le mode est : 38
 L'intervalle de confiance à 95% est : 39.6706996821199 - 47.9126336512134



\$X

```
[1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18
[19] 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
[37] 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
[55] 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72
[73] 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
[91] 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108
[109] 109 110 111 112 113 114 115 116 117 118 119 120
```

\$Note1

```
[1] 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
```

\$Note2

```
[1] 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
```

\$Taille

```
[1] 0.004209084 0.012337527 0.028121015 0.045842964 0.058029181 0.063566172
```

```
[7] 0.073864894 0.085425524 0.102949881 0.119437474 0.135021747 0.176903489
[13] 0.190500413 0.209191483 0.243629910 0.258111600 0.265938663 0.269448024
[19] 0.297156844 0.318494551 0.342066376 0.351455190 0.352450551 0.400516671
[25] 0.403350968 0.409307986 0.419681427 0.428381393 0.436180773 0.492316414
[31] 0.512836543 0.525873146 0.538826580 0.695293388 0.703187258 0.717797351
[37] 0.726443563 0.753783489 0.770152287 0.804570954 0.824496007 0.847926646
[43] 0.861370046 0.862096279 0.878126726 0.902162288 0.927048133 0.931169031
[49] 0.966430503 0.967854782 0.974819643 0.988464955 1.000865094 1.051983280
[55] 1.065887922 1.071079296 1.075468013 1.098250372 1.133195996 1.135319420
[61] 1.139312502 1.176576759 1.184386629 1.205470660 1.214799502 1.223245929
[67] 1.235200681 1.240728670 1.250955104 1.277508501 1.288728653 1.323125394
[73] 1.347300126 1.363532525 1.367632870 1.367906984 1.399449387 1.409345749
[79] 1.415178108 1.461913167 1.467137335 1.468295560 1.474550202 1.477257352
[85] 1.484397480 1.512107068 1.517215345 1.532569740 1.550420144 1.579627367
[91] 1.594682019 1.621669005 1.655723411 1.656850279 1.678559897 1.680821900
[97] 1.739369836 1.741604475 1.745294339 1.749883842 1.765820164 1.802122384
[103] 1.802192123 1.810793255 1.845540383 1.854018149 1.855690395 1.858411719
[109] 1.859496333 1.874703831 1.878197953 1.887373002 1.895254736 1.906599790
[115] 1.908404030 1.919318463 1.923315982 1.939001973 1.953314032 1.962794596
```

\$Age

```
[1] 0 1 2 8 9 10 11 12 13 15 16 17 18 19 21 23 24 26 28 29 30 31 36 37 38
[26] 39 40 41 42 43 44 46 48 49 50 51 52 53 54 55 56 57 58 59 60 61 63 65 66 67
[51] 69 70 71 72 73 74 75 76 77 78
```

Application pour les variables qualitatives

Etude de l'indépendances des variables quantitatives

Test de corrélation de pearson entre Taille et poids

```
cor.test(df_KPAKOU$Age,df_KPAKOU$Taille,method=c("pearson", "kendall", "spearman"))
```

Pearson's product-moment correlation

```
data: df_KPAKOU$Age and df_KPAKOU$Taille
t = -1.2019, df = 118, p-value = 0.2318
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.28362368 0.07066115
sample estimates:
      cor
-0.109973
```

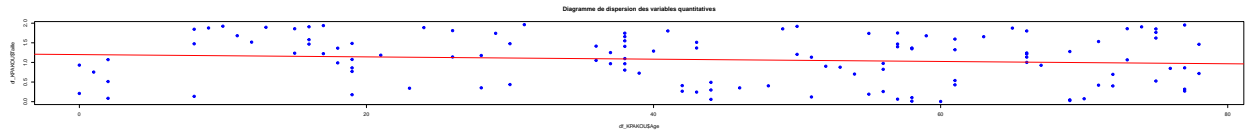
Le coefficient de corrélation est faible(-0.109973) et la p-value supérieur à 0.05 donc le test de pearson n'est significatif *Test de spearman*

```
cor.test(df_KPAKOU$Age,df_KPAKOU$Taille,method="spearman")
```

De même le test de spearman n'est pas significatif

Nuage des points

```
plot(df_KPAKOU$Age,df_KPAKOU$Taille, pch = 19, col = "blue", main = "Diagramme de dispersion des variables quantitatives")
# Ajouter une droite de régression
abline(lm(df_KPAKOU$Taille ~df_KPAKOU$Age), col = "red")
```



Etude de correlation entre deux variables qualitatives

Test de chi2

```
chisq.test(df_KPAKOU$Sexe,df_KPAKOU$Mention)
```

Pearson's Chi-squared test

data: df_KPAKOU\$Sexe and df_KPAKOU\$Mention

X-squared = 120, df = 3, p-value < 2.2e-16

La p-value du test de chi 2 est inferieur à 0.05 donc on peut conclure qu'il y'a correlation entre le sexe et la mention.

Etude correlation entre une variable quantitative et une variable qualitative

*Test t de student

```
t.test(df_KPAKOU$Age ~ df_KPAKOU$Sexe)
```

Welch Two Sample t-test

data: df_KPAKOU\$Age by df_KPAKOU\$Sexe

t = -0.36304, df = 117.04, p-value = 0.7172

alternative hypothesis: true difference in means between group F and group M is not equal to 0

95 percent confidence interval:

-9.790255 6.756922

sample estimates:

mean in group F mean in group M

43.03333 44.55000

La $p - value = 0.7172 > 0.05$,le test de t-Student n'est pas significatif

BOXPLOT

```
boxplot(df_KPAKOU$Age ~ df_KPAKOU$Sexe)
```