Christopher PHILET

Statistique descriptive : Etude descriptives de données

Fiche TP A : Réflexion sur l'indépendance en statistique descriptive

* **9.1**

snee <-read.table("http://www.biostatisticien.eu/

springeR/snee74.txt»,header=TRUE)

* **9.2**

head(snee)

tail(snee)

* Il y a 592 individus et 3 variables qualitatives
* **9.3**

attach(snee)

names(snee)

class(cheveux)

levels(cheveux) # Modalités de cheveux

class(yeux)

levels(yeux) # Modalités de yeux

class(sexe)

levels(sexe) # Modalités du sexe

* **9.4**
* Etude de la variable cheveux

table(cheveux) # effectifs

round(table(cheveux)/length(cheveux)\*100,2) # fréquences en pourcentages

names(which.max(table(cheveux))) # Mode de la variable

barplot(sort(table(cheveux),TRUE),col=c("yellow2","tan4","black","tan1"))

* Etude de la variable yeux

table(yeux) # effectifs

round(table(yeux)/length(yeux)\*100,2) # fréquences en pourcentages

names(which.max(table(yeux))) # Mode de la variable

barplot(sort(table(yeux),TRUE),col=c("blue","brown","tan3","green"))

* Etude de la variable sexe

table(sexe) # effectifs

round(table(sexe)/length(sexe)\*100,2) # fréquences en pourcentages

names(which.max(table(sexe))) # Mode de la variable

barplot(sort(table(sexe),TRUE),col=c("pink","blue"))

* **9.5**

yeuxcheveux <- table(yeux,cheveux) # Table de contingence

* **9.6**

fchev <- margin.table(yeuxcheveux,2)/592 # Profils colonnes

fchev

* **9.7**

nbleus <- margin.table(yeuxcheveux,1)[1] # Nombre d'individus ayant les yeux bleux

nbleus

* **9.8**

round(fchev\*nbleus)

* **9.9**

margeX <- margin.table(yeuxcheveux,1)

tab.ind1 <- margeX%\*%t(fchev)

round(tab.ind1)

* **9.10**

fyeu <- margin.table(yeuxcheveux,1)/592 # Profils lignes

fyeu

* **9.11**

nblonds <- margin.table(yeuxcheveux,2)[1] # Nombre d'individus ayant les cheveux blonds

nblonds

* **9.12**

round(fyeu\*nblonds)

* **9.13**

margeY <- margin.table(yeuxcheveux,2)

tab.ind2 <- fyeu%\*%t(margeY)

round(tab.ind2)

* **9.14**

all.equal(tab.ind1,tab.ind2)

Les deux tableaux sont identiques, ce qui est rassurant puisque s'intéresser à l’indépendance des yeux et des cheveux est équivalent à s’intéresser à l’indépendance des cheveux et des yeux.

* **9.15**

(yeuxcheveux-tab.ind1)^2

* **9.16**

tab.contr <- (yeuxcheveux-tab.ind1)^2/tab.ind1

tab.contr

* **9.17**

khi2 <- sum(tab.contr)

khi2

Phi2 <- khi2/sum(yeuxcheveux)

Phi2

C <- sqrt(khi2/(sum(yeuxcheveux)+khi2))

C

V2 <- Phi2/(min(dim(yeuxcheveux))-1)

V2

Tous ces indicateurs ne sont pas nuls, il y a donc une certaine forme de dépendance

* **9.18**

prop.table(yeuxcheveux,1) # distributions conditionnelles de cheveux sachant yeux = (bleu, ou marron ou ...).

Les lignes ne sont pas égales

prop.table(yeuxcheveux,2) # distributions conditionnelles de yeux sachant cheveux = (blond ou ...).

* les colonnes ne sont pas égales
* Ceci confirme la dépendance (d'un point de vue statistique descriptive) entre les deux caractères
* **9.19**

table(cheveux,sexe)

plot(cheveux~sexe)

plot(table(cheveux,sexe))

* **9.20**

table(yeux,sexe)

plot(yeux~sexe)

plot(table(yeux,sexe))

* 9.21