
MIC0503V - Examen de Statistique - Corrigé

Exercice 1.

1. Nous enregistrons d'abord les données au format .csv dans le fichier *hémoglobine.csv* puis sous R dans l'objet *hémoglobine*.

```
> hémoglobine<-read.csv2("hémoglobine.csv")  
> hémoglobine
```

| | Femmes | Hommes |
|----|--------|--------|
| 1 | 105 | 141 |
| 2 | 110 | 144 |
| 3 | 112 | 146 |
| 4 | 112 | 148 |
| 5 | 118 | 149 |
| 6 | 119 | 150 |
| 7 | 120 | 150 |
| 8 | 120 | 151 |
| 9 | 125 | 153 |
| 10 | 126 | 153 |
| 11 | 127 | 153 |
| 12 | 128 | 154 |
| 13 | 130 | 155 |
| 14 | 132 | 156 |
| 15 | 133 | 156 |
| 16 | 134 | 160 |
| 17 | 135 | 160 |
| 18 | 138 | 160 |
| 19 | 138 | 163 |
| 20 | 138 | 164 |
| 21 | 138 | 164 |
| 22 | 142 | 165 |
| 23 | 145 | 166 |
| 24 | 148 | 168 |
| 25 | 148 | 168 |
| 26 | 150 | 170 |
| 27 | 151 | 172 |
| 28 | 154 | 172 |
| 29 | 154 | 176 |
| 30 | 158 | 179 |

Pour effectuer le regroupement en classes nous utilisons la fonction *hist* pour chaque groupe femmes et hommes :

```
> Femmes.hist<-hist(hémoglobine$Femmes,breaks=c(104,114,124,134,144,154,164,174,184))  
> Hommes.hist<-hist(hémoglobine$Hommes,breaks=c(104,114,124,134,144,154,164,174,184))
```

Enfin, pour obtenir les effectifs et les fréquences :

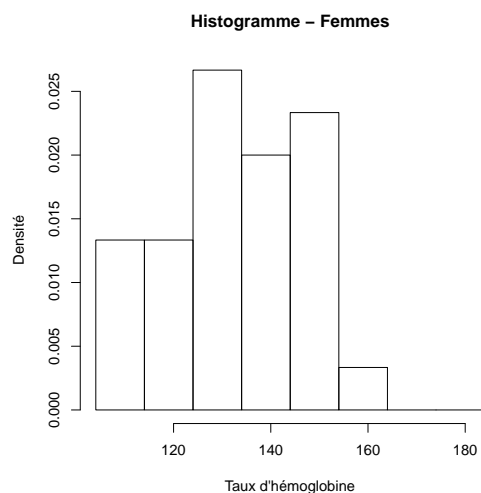
```
> table.Femmes<-round(data.frame(Femmes.hist$mids,Femmes.hist$counts,Femmes.hist$freqs),2)
> colnames(table.Femmes)<-c("centres","effectifs","fréquences")
> table.Femmes
```

| | centres | effectifs | fréquences |
|---|---------|-----------|------------|
| 1 | 109 | 4 | 0.13 |
| 2 | 119 | 4 | 0.13 |
| 3 | 129 | 8 | 0.27 |
| 4 | 139 | 6 | 0.20 |
| 5 | 149 | 7 | 0.23 |
| 6 | 159 | 1 | 0.03 |
| 7 | 169 | 0 | 0.00 |
| 8 | 179 | 0 | 0.00 |

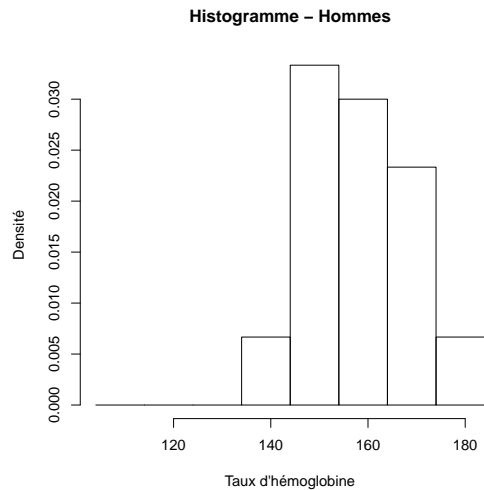
```
> table.Hommes<-round(data.frame(Hommes.hist$mids,Hommes.hist$counts,Hommes.hist$freqs),2)
> colnames(table.Hommes)<-c("centres","effectifs","fréquences")
> table.Hommes
```

| | centres | effectifs | fréquences |
|---|---------|-----------|------------|
| 1 | 109 | 0 | 0.00 |
| 2 | 119 | 0 | 0.00 |
| 3 | 129 | 0 | 0.00 |
| 4 | 139 | 2 | 0.07 |
| 5 | 149 | 10 | 0.33 |
| 6 | 159 | 9 | 0.30 |
| 7 | 169 | 7 | 0.23 |
| 8 | 179 | 2 | 0.07 |

2. `> hist(hémoglobine$Femmes,breaks=c(104,114,124,134,144,154,164,174,184),main="Histogramme - Femmes")`



```
> hist(hémoglobine$Hommes,breaks=c(104,114,124,134,144,154,164,174,184),main="Histogramme - Hommes")
```



3. On utilise la fonction *summary* pour chaque série :

```
> summary(c(hémoglobine$Femmes,hémoglobine$Hommes))
```

| | | | | | |
|-------|---------|--------|-------|---------|-------|
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 105.0 | 133.8 | 149.5 | 145.9 | 158.5 | 179.0 |

```
> summary(hémoglobine$Femmes)
```

| | | | | | |
|-------|---------|--------|-------|---------|-------|
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 105.0 | 121.2 | 133.5 | 132.9 | 144.2 | 158.0 |

```
> summary(hémoglobine$Hommes)
```

| | | | | | |
|-------|---------|--------|-------|---------|-------|
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 141.0 | 151.5 | 158.0 | 158.9 | 165.8 | 179.0 |

On obtient ainsi les moyennes pour chaque série :

Ensemble : 145,9

Femmes : 132,9

Hommes : 158,9.

4. On utilise les résultats de la fonction *hist* pour calculer les moyennes :

Ensemble :

```
> ens.hist<-hist(c(hémoglobine$Femmes,hémoglobine$Hommes),breaks=c(104,114,124,134,144,154,164,174,184))
```

```
> ens.hist$counts
```

```
[1]  4  4  8  8 17 10  7  2
```

```
> moy.ens<-sum(ens.hist$mids*ens.hist$counts)/60
```

```
> moy.ens
```

```
[1] 145.3333
```

Femmes :

```
> moy.F<-sum(Femmes.hist$mids*Femmes.hist$counts)/30
```

```
> moy.F
```

```
[1] 132.6667
```

Hommes :

```
> moy.H<-sum(Hommes.hist$mids*Hommes.hist$counts)/30  
> moy.H
```

```
[1] 158
```

5. En utilisant les résultats de la question 3, on obtient les médianes :

Ensemble : 149,5

Femmes : 133,5

Hommes : 158.

6. En utilisant à nouveau les résultats de la question 3, on obtient les intervalles interquartiles :

Ensemble : [133,8;158,5]

Femmes : [121,2;144,2]

Hommes : [151,5;165,8].

7. Variances et écarts-type :

ensemble :

```
> 59*var(c(hémoglobine$Femmes,hémoglobine$Hommes))/60
```

```
[1] 315.3567
```

```
> sqrt(59*var(c(hémoglobine$Femmes,hémoglobine$Hommes))/60)
```

```
[1] 17.75828
```

Femmes :

```
> 29*var(hémoglobine$Femmes)/30
```

```
[1] 201.2622
```

```
> sqrt(29*var(hémoglobine$Femmes)/30)
```

```
[1] 14.18669
```

Hommes :

```
> 29*var(hémoglobine$Hommes)/30
```

```
[1] 93.18222
```

```
> sqrt(29*var(hémoglobine$Hommes)/30)
```

```
[1] 9.653094
```

8. On utilise la même démarche qu'à la question 4 :

Ensemble :

```
> (sum(ens.hist$mids*ens.hist$mids*ens.hist$counts)/60)-(moy.ens^2)
```

```
[1] 313.2222
```

Femmes :

```
> (sum(Femmes.hist$mids*Femmes.hist$mids*Femmes.hist$counts)/30)-(moy.F^2)
```

```
[1] 196.5556
```

Hommes :

```
> (sum(Hommes.hist$mids*Hommes.hist$mids*Hommes.hist$counts)/30)-(moy.H^2)
```

```
[1] 109
```

Exercice 2.

1. Création du tableau de données :

```
> rhesus<-matrix(c(3570,3825,935,170,630,675,165,30),ncol=4,byrow=T)
> rownames(rhesus)<-c("Rh+", "Rh-")
> colnames(rhesus)<-c("O", "A", "B", "AB")
> rhesus
```

| | O | A | B | AB |
|-----|------|------|-----|-----|
| Rh+ | 3570 | 3825 | 935 | 170 |
| Rh- | 630 | 675 | 165 | 30 |

Calcul des effectifs marginaux :

Lignes :

```
> ni.<-margin.table(rhesus,1)
> ni.
```

| Rh+ | Rh- |
|------|------|
| 8500 | 1500 |

Colonnes :

```
> n.j<-margin.table(rhesus,2)
> n.j
```

| | O | A | B | AB |
|------|------|------|-----|----|
| 4200 | 4500 | 1100 | 200 | |

2. Profils lignes :

```
> rhesus/as.vector(ni.)
```

| | O | A | B | AB |
|-----|------|------|------|------|
| Rh+ | 0.42 | 0.45 | 0.11 | 0.02 |
| Rh- | 0.42 | 0.45 | 0.11 | 0.02 |

On constate que les profils lignes sont identiques : il n'y a donc aucun lien entre le groupe sanguin et le facteur Rhésus (on doit trouver un χ^2 égal à 0).

3. Effectifs théoriques :

```
> efftheor<-round(t(n.j)%*t(ni.)/sum(rhesus)),2)
> efftheor
```

| | O | A | B | AB |
|-----|------|------|-----|-----|
| Rh+ | 3570 | 3825 | 935 | 170 |
| Rh- | 630 | 675 | 165 | 30 |

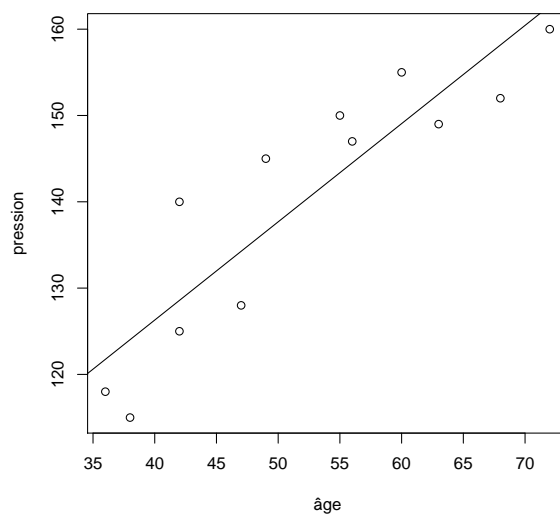
-
4. Le tableau des effectifs théoriques étant identique au tableau initial, on a : $\chi^2 = 0$.

Exercice 3.

1. `> pression<-read.csv2("pression.csv")`
`> pression`

| | âge | pression |
|----|-----|----------|
| 1 | 56 | 147 |
| 2 | 42 | 125 |
| 3 | 72 | 160 |
| 4 | 36 | 118 |
| 5 | 63 | 149 |
| 6 | 47 | 128 |
| 7 | 55 | 150 |
| 8 | 49 | 145 |
| 9 | 38 | 115 |
| 10 | 42 | 140 |
| 11 | 68 | 152 |
| 12 | 60 | 155 |

`> plot(pression)`
`> abline(lm(pression$pression~pression$âge)$coefficients)`



2. Calcul de \bar{X} , σ_X^2 , \bar{Y} et σ_Y^2 :

`> mean(pression$âge)`
[1] 52.33333
`> 11*var(pression$âge)/12`
[1] 129.2222
`> mean(pression$pression)`
[1] 140.3333

```
> 11*var(pression$pression)/12
```

```
[1] 208.3889
```

3. Covariance entre X et Y :

```
> 11*cov(pression$âge,pression$pression)/12
```

```
[1] 147.0556
```

4. Coefficient de corrélation linéaire :

```
> cor(pression$âge,pression$pression)
```

```
[1] 0.8961394
```

Au vu de la valeur du coefficient de corrélation linéaire, proche de 1, et du nuage de points, "étiré", il existe une liaison linéaire forte entre les deux variables.

5. Pour calculer les coefficients de la droite de régression, on utilise la fonction *lm* :

```
> lm(pression$pression~pression$âge)$coefficients
```

```
(Intercept) pression$âge
```

```
80.777730      1.138005
```

6. Estimation de la pression systolique pour une femme de 50 ans :

```
> 80.777730 +1.138005*50
```

```
[1] 137.678
```

Exercice 4.

1. Calculer la moyenne et la variance du pourcentage de temps passé couché pour l'ensemble de tous les individus.

```
> traitement<-read.csv2("traitement.csv")
```

```
> traitement
```

| | Traitement.1 | Traitement.2 | Traitement.3 |
|---|--------------|--------------|--------------|
| 1 | 17.4 | 14.65 | 18.76 |
| 2 | 20.0 | 37.22 | 19.49 |
| 3 | 26.7 | 37.73 | 27.19 |
| 4 | 31.7 | 43.61 | 45.42 |
| 5 | 35.8 | 46.07 | 53.20 |
| 6 | 47.8 | 47.40 | 61.27 |

```
> trait.moy<-mean(c(traitement$Traitement.1,traitement$Traitement.2,traitement$T
```

```
> trait.moy
```

```
[1] 35.07833
```

```
> trait.var<-17*var(c(traitement$Traitement.1,traitement$Traitement.2,traitement$T
```

```
> trait.var
```

```
[1] 180.6152
```

2. Traitement 1 :

```
> moy1<-mean(traitement$Traitement.1)
> moy1
```

```
[1] 29.9
```

```
> 5*var(traitement$Traitement.1)/6
```

```
[1] 103.8267
```

Traitement 2 :

```
> moy2<-mean(traitement$Traitement.2)
> moy2
```

```
[1] 37.78
```

```
> 5*var(traitement$Traitement.2)/6
```

```
[1] 121.7617
```

Traitement 3 :

```
> moy3<-mean(traitement$Traitement.3)
> moy3
```

```
[1] 37.555
```

```
> 5*var(traitement$Traitement.3)/6
```

```
[1] 276.0092
```

3. Variance inter-classes :

```
> varinter<-sum((moy1-trait.moy)^2+(moy2-trait.moy)^2+(moy3-trait.moy)^2)/3
> varinter
```

```
[1] 13.41601
```

```
> sqrt(varinter/trait.var)
```

```
[1] 0.2725427
```

Le rapport de corrélation étant relativement "faible", on peut conclure à une liaison faible entre le traitement et le temps passé couché.