

Corrigé-type. Initiation au Logiciel R.

M1_Stat.Prob.Approfondies.

Épreuve Finale. Initiation au Logiciel R.

Master1: Statistiques et Probabilités Approfondies. 2020-2021

June 3, 2021

Exercice.

Nous proposons le jeu de données des notes d'une population de 9 étudiants du Master 1 Stat.Proba obtenues dans les 4 modules (Anal.Surv, Anal.Mult, Prob.Approf et Régression), ($E_i, i = 1, \dots, 9$ représentent les 9 étudiants):

| | Anal.Surv | Anal.Mult | Prob.Approf | Régression |
|----|-----------|-----------|-------------|------------|
| E1 | 5.5 | 5 | 6 | 6.5 |
| E2 | 8 | 8 | 8 | 8 |
| E3 | 7 | 6 | 11 | 9.5 |
| E4 | 15 | 15.5 | 14.5 | 14.5 |
| E5 | 12.5 | 12 | 14 | 14.5 |
| E6 | 11 | 10 | 5.5 | 7 |
| E7 | 7 | 5.5 | 11.5 | 14 |
| E8 | 12.5 | 13 | 8.5 | 9.5 |
| E9 | 9 | 9.5 | 12 | 12.5 |

Table 1: Notes des étudiants Master 1.

- Écrire la Table 1 (que l'on notera par X) sous R et renommer respectivement les lignes et les colonnes par E_i et les 4 disciplines citées plus haut.
- Afficher les paramètres de position (Quartiles) et de dispersion (Moyenne) associés à chaque discipline.
- Y a-t-il une liaison entre les disciplines Analyse Multivariée et Régression. Justifier graphiquement.
- Construire la matrice de Variance-Covariance empirique de $X = (X^1, X^2, X^3, X^4)$ matrice des colonnes $X^j, j = 1, \dots, 4$, variables quantitatives dans \mathbb{R}^9 associées aux 4 disciplines citées dans la Table 1. Les éléments $\Gamma_{k,l}$ de Γ sont donnés par $\Gamma_{k,l} = \frac{1}{n} \langle X^k - \mathbb{E}(X^k), X^l - \mathbb{E}(X^l) \rangle$.
- Calculer (sous R) la somme des éléments diagonaux de la matrice Γ .
- Calculer les éléments propres de Γ et en déduire (sous R) la somme des valeurs propres. Que peut-on conclure?

```

> #####Question 1##### (04 pts)
> M=matrix(c(5.5,5,6,6.5,8,8,8,8,7,6,11,9.5,15,15.5,14.5,14.5,12.5,12,14,14.5,
+ 11,10,5.5,7,7,5.5,11.5,14,12.5,13,8.5,9.5,9,9.5,12,12.5),9,4,byrow=TRUE)
> rownames(M)=c("E1", "E2", "E3", "E4", "E5", "E6", "E7", "E8", "E9")
> colnames(M)=c("Anal.Surv", "Anal.Mult", "Prob.Approf", "Régression")
> M
Anal.Surv Anal.Mult Prob.Approf Régression
E1      5.5      5.0      6.0      6.5
E2      8.0      8.0      8.0      8.0
E3      7.0      6.0      11.0     9.5
E4     15.0     15.5     14.5     14.5
E5     12.5     12.0     14.0     14.5
E6     11.0     10.0      5.5      7.0
E7      7.0      5.5     11.5     14.0
E8     12.5     13.0      8.5      9.5
E9      9.0      9.5     12.0     12.5

```

```

> #####Question 2##### (03 pts)
> summary(M)
  Anal.Surv      Anal.Mult      Prob.Approf      Régression
Min.   : 5.500   Min.   : 5.000   Min.   : 5.50   Min.   : 6.50
1st Qu.: 7.000   1st Qu.: 6.000   1st Qu.: 8.00   1st Qu.: 8.00
Median : 9.000   Median : 9.500   Median :11.00   Median : 9.50
Mean   : 9.722   Mean   : 9.389   Mean   :10.11   Mean   :10.67
3rd Qu.:12.500   3rd Qu.:12.000   3rd Qu.:12.00   3rd Qu.:14.00
Max.   :15.000   Max.   :15.500   Max.   :14.50   Max.   :14.50

```

```

> #####Question 3##### (04 pts)
> plot(M[,2],M[,4],xlab="Anal.Mult",ylab="Régression",main="Relation entre l'A
> régression1=lm(M[,4]~M[,2])
> régression1

```

Call:
lm(formula = M[, 4] ~ M[, 2])

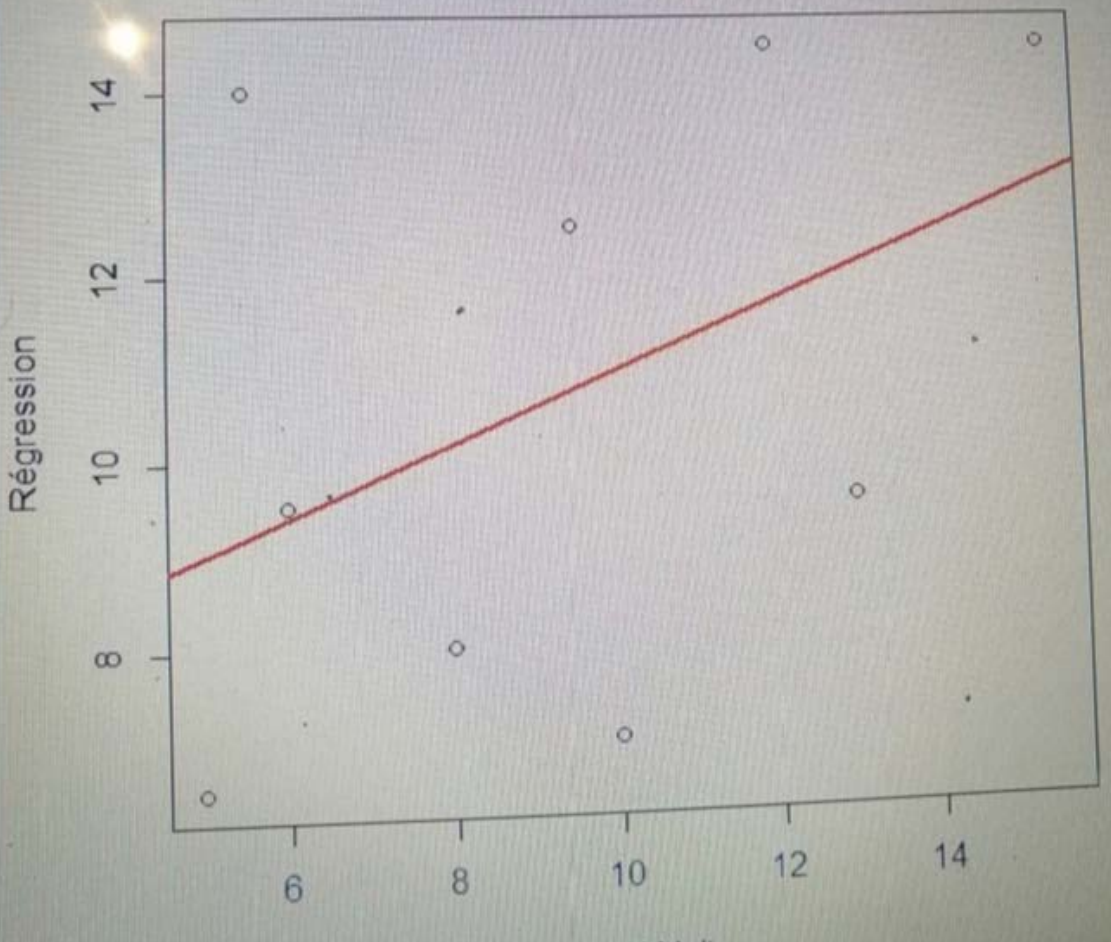
Coefficients:
(Intercept) M[, 2]
 7.0883 0.3811

```

> ### Les disciplines Anal.Mult et Régression ne sont pas en relation, puisque
> ### le nuage de points est aléatoirement dispersé "Corrélation négligeable"
> abline(régression1,col="red",lwd=2)

```

Relation entre l'Anal.Mult et Régression




```

> ### On peut vérifier de la même façon l'existence de la relation entre
> ### L'Anal.Surv et l'Anal.Mult.
> plot(M[,1],M[,2],xlab="Anal.Surv",ylab="Anal.Mult",main="Relation entre l'An
> plot(M[,1],M[,2],xlab="Anal.Surv",ylab="Anal.Mult",main="Relation entre l'An
> régression2=lm(M[,2]~M[,1])
> abline(régression2,col="red",lwd=2)
> ### On remarque l'ajustement du nuages de points par la droite de régression
> ### des coefficients ci-dessous
> régression2

```

```

Call:
lm(formula = M[, 2] ~ M[, 1])

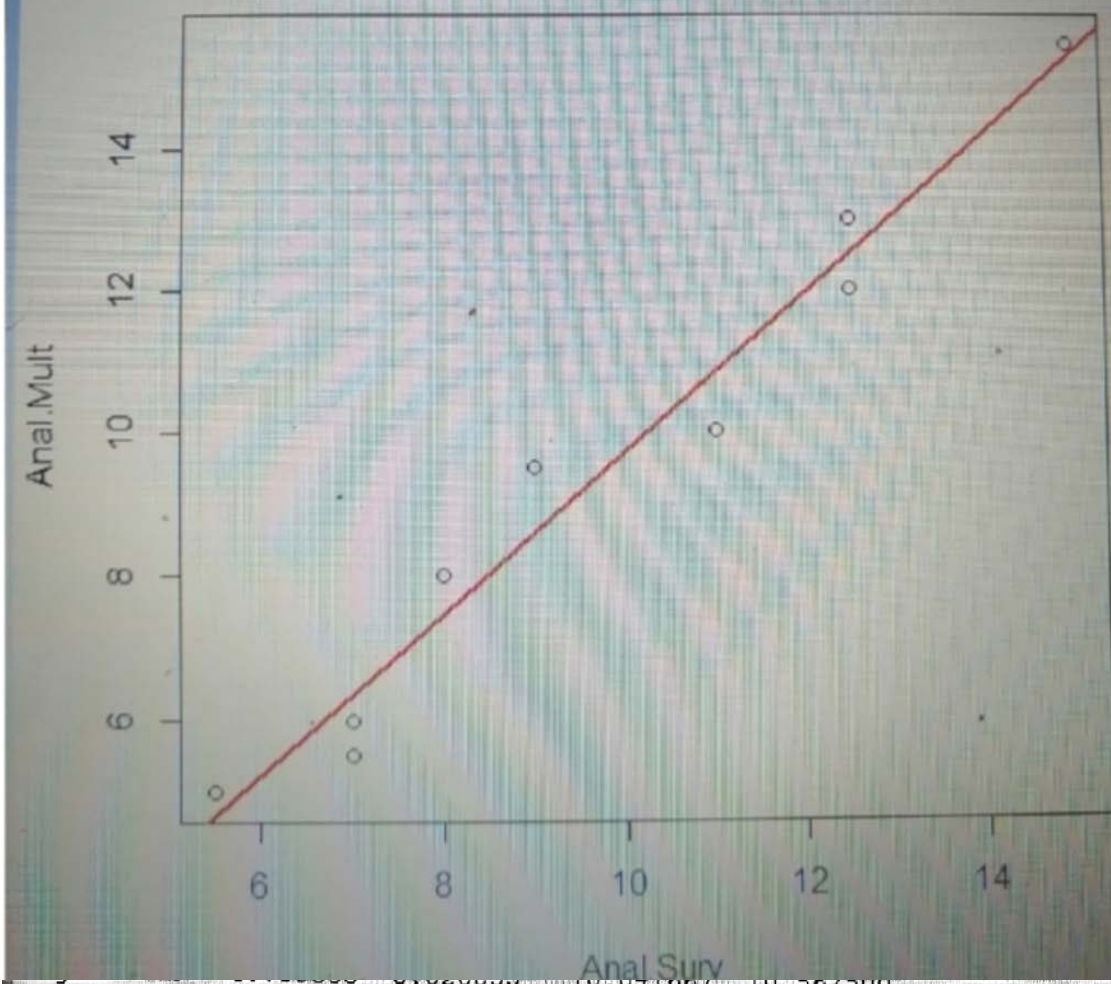
```

```

Coefficients:
(Intercept)      M[, 1]
    -1.523         1.122

```

Relation entre l'Anal.Mult et l'Anal.Surv



```

> #####Question 4 ##### (03pts)
> Gamma=matrix(NA,4,4)#### Matrice de Variance-Covariances corrigée
> for(i in 1:4){
+ for (j in 1:4)
+ {Gamma[i,j]=(1/9)*(M[,i]-mean(M[,i]))*(M[,j]-mean(M[,j]))}
> Gamma

```

| | [,1] | [,2] | [,3] | [,4] |
|------|-----------|-----------|-----------|-----------|
| [1,] | 10.131944 | 11.371528 | 4.628472 | 4.739583 |
| [2,] | 11.371528 | 13.173611 | 5.138889 | 5.020833 |
| [3,] | 4.628472 | 5.138889 | 10.736111 | 10.041667 |
| [4,] | 4.739583 | 5.020833 | 10.041667 | 10.562500 |

```

> #### Pour construire la matrice de Var-Cov, on utilise le code R sui
> cov(M)

```

| | Anal.Surv | Anal.Mult | Prob.Approf | Régression |
|-------------|-----------|-----------|-------------|------------|
| Anal.Surv | 10.131944 | 11.371528 | 4.628472 | 4.739583 |
| Anal.Mult | 11.371528 | 13.173611 | 5.138889 | 5.020833 |
| Prob.Approf | 4.628472 | 5.138889 | 10.736111 | 10.041667 |
| Régression | 4.739583 | 5.020833 | 10.041667 | 10.562500 |

```
#####Question5#####(02points)
sum(diag(Gamma))#### Somme des Variances
] 44.60417
#####Question6#####(04points)
res=eigen(Gamma)#####Calcul des éléments propres de Gamma
#####Valeurs Propres De Gamma=res$values#####
res$values
] 31.7534617 12.0674054 0.6361523 0.1471473
sum(res$values)#####Somme des Valeurs propres de Gamma
] 44.60417
###Conclusion: On voit bien que la somme des valeurs propres somme des
###Variances= Inertie totale = 44.60417
```