Université d'Orléans
Statistiques
D. Fall, M. Zani

L3 MIAGE
2023-2024

Contrôle terminal, durée 2h

Exercice 1 A la demande de la chambre syndicale des fabricants de produits surgelés, une enquête portant sur les dépenses mensuelles en Euros de produits surgelés chez les ménages dotés d'un frazer (3 étoiles) a été faite. Les résultats de cette enquête sont les suivants:

Classe	Centre de classe	Effectifs
$[0, 20[$	10	25
$[20, 40[$	30	25
$[40, 60[$	50	45
$[60, 80[$	70	60
$[80, 100[$	90	75
$[100, 120[$	110	85
$[120, 140[$	130	75
$[140, 160[$	150	50
$[160, 200[$	180	45
$[200, 300[$	250	15

1. Calculer les indices de centralité et de dispersion (moyenne et écart-type) à partir de ces données classées.
2. Tracer la courbe cumulative des fréquences. Déterminer les quartiles.
3. Déterminer la classe modale par le calcul des densités. Le visualiser en traçant l'histogramme des données (utiliser les classes données dans le tableau).
4. A partir de la forme de l'histogramme, proposer une modélisation de la distribution des dépenses des ménages français en ce qui concerne les produits surgelés. Quelle commande sous R proposez vous pour illustrer cette modélisation sur l'histogramme?

Exercice 2
Pour tester la performance du système de freinage d'une voiture, on la fait rouler jusqu'à atteindre une vitesse x (en mètres par seconde), à laquelle on freine. On mesure alors la distance de freinage y (en mètres). On fait l'expérience pour $n = 8$ vitesses différentes $x_1, \cdots, x_n$ et on mesure les distances de freinage correspondantes $y_1, \cdots, y_n$. Toutes les mesures sont faites avec le même véhicule et le même pilote, et sont supposées indépendantes. On obtient le tableau suivant

x_i : vitesse	5	10	15	20	25	30	35	40
y_i: freinage	3.42	5.96	31.14	41.76	74.54	94.92	133.78	169.16

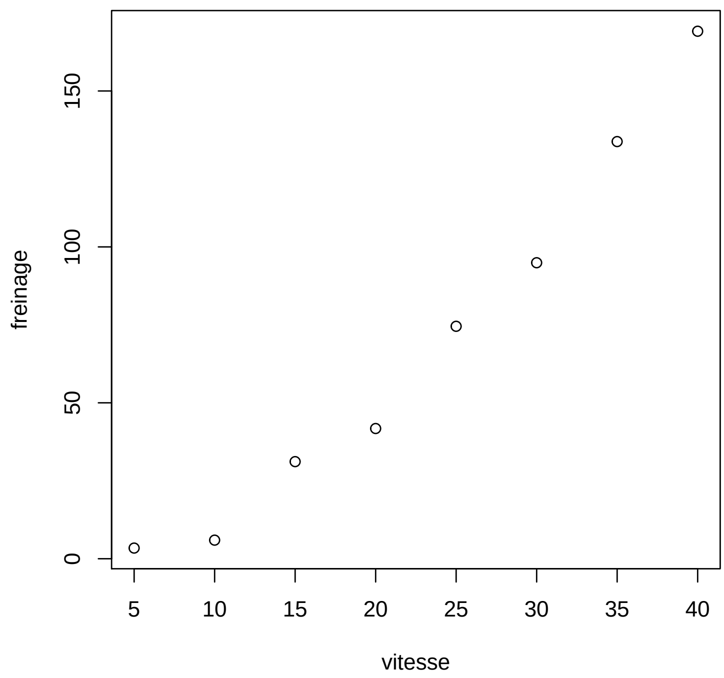
1. Quelle est la nature des deux variables étudiées? Quelle est la variable réponse, la variable explicative?

   Donnez la commande R pour créer le tableau de données ci-dessus.

2. Quelle est la première chose à faire pour étudier ces données bivariées? On exécute la commande suivante:

```
> plot(vitesse,freinage)
```

Qui donne la figure ci-après. Commentez ce graphe.



3. Les commandes suivantes sont exécutées

```
> sum(vitesse)
[1] 180
> sum(freinage)
[1] 554.68
> sum(vitesse*freinage)
[1] 17538.8
> sum(vitesse^2)
[1] 5100
> sum(freinage^2)
[1] 63839.03
```

Calculer la covariance entre les deux variables ainsi que le coefficient de corrélation. Conclusion?

4. Les lignes de commande suivantes

```
> reg=lm(freinage~vitesse)
> summary(reg)
```

donnent le résultat ci-dessous

```
Call:
lm(formula = freinage ~ vitesse)

Residuals:
    Min      1Q  Median      3Q     Max
-15.531  -7.766  -2.609   7.048  18.393

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -39.0614    10.1113  -3.863  0.00833 **
vitesse       4.8176     0.4005  12.030    2e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.98 on 6 degrees of freedom
Multiple R-squared:  0.9602,Adjusted R-squared:  0.9536
F-statistic: 144.7 on 1 and 6 DF,  p-value: 2.002e-05
```

Commenter ces résultats. Les variables sont elles significativement liées par une relation linéaire? Si oui, calculer la distance de freinage pour une vitesse $\mathbf{x}=50$, et donner la commande R correspondante.

5. On cherche une autre forme de modélisation pour expliquer cette distribution de points. On pose

```
>z=sqrt(freinage)
```

Expliquer pourquoi on choisit cette nouvelle variable.

6. On réalise avec $\mathbf{z}$ les commandes :

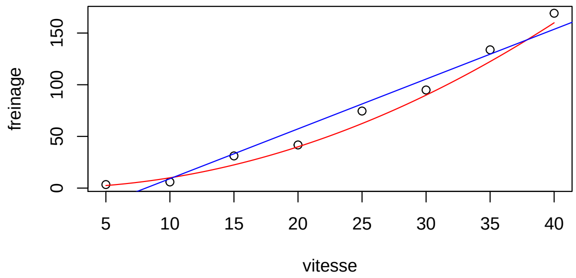
```
> cor(vitesse, z)
[1] 0.99391
> reg1=lm(z~vitesse)
> summary(reg1)

Call:
lm(formula = z ~ vitesse)

Residuals:
     Min       1Q   Median       3Q      Max
-0.85044 -0.14649 -0.04338  0.25365  0.64117

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.003044   0.376538  -0.008    0.994
vitesse      0.329480   0.014913  22.093 5.62e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4832 on 6 degrees of freedom
Multiple R-squared:  0.9879,Adjusted R-squared:  0.9858
F-statistic: 488.1 on 1 and 6 DF,  p-value: 5.621e-07
```

Expliquer ce qu'apporte cette nouvelle modélisation.

7. On effectue les commandes

```
> plot(vitesse, freinage)
> abline(reg,col='blue')
> x0=seq(5,40,by=0.01)
> points(x0,0.1*x0^2-0.00198*x0+0.000009,type='l',col='red')
```

qui donne la Figure 2. Expliquer ces différents graphes.



8. Avec cette nouvelle modélisation, prédire le freinage pour une vitesse **x**=50 et comparer au résultat obtenu en question 4. Quelle prédiction vous semble la plus réaliste?

Exercice 3

La SNCF annonce que sur l'année écoulée, en moyenne 90% de ses trains en banlieue parisienne sont à l'heure. Un usager utilisant ces trains quotidiennement veut tester ce fait sur son trajet. Il collecte donc des données pendant 3 mois, soit $n = 62$ jours ouvrés: son train est arrivé à l'heure 51 fois et en retard 11 fois . L'usager s'intéresse au taux p de trains (sur cet itinéraire et à cette heure) qui sont en retard. On note X_i la variable aléatoire telle que $X_i = 1$ si le train est arrivé en retard au jour i, et $X_i = 0$ sinon. On note par N la variable aléatoire désignant le nombre de trains arrivés en retard sur les n.

1. Quelle est la loi des X_i? Quelle est la loi de N? Justifier votre réponse.
2. Construire un estimateur de p par la méthode des moments.
3. Donner une estimation de p. Pourquoi l'usager peut, à juste titre, penser que le chiffre indiqué par la SNCF est incorrect?
4. Construire un intervalle de confiance pour p, avec un niveau de confiance de 95%.

Valeurs de quantiles et résultats pouvant vous être utiles:

```
> qnorm(.98)        > qnorm(.995)
[1] 2.053749       [1] 2.575829
> qnorm(.95)        > qnorm(.99)
[1] 1.644854        [1] 2.326348
> qnorm(.9)        > qnorm(.03)
[1] 1.281552       [1]  -1.880794
> qnorm(.97)       > qnorm(.99)
[1] 1.880794       [1] 2.326348
> qnorm(.975)      > qt(.99,2999)
[1] 1.959964       [1] 2.364606
> qnorm(.995)      > qt(.98,2999)
[1] 2.575829       [1] 2.081162
```