Exercices sur le cours d'Analyse de la Variance

Chapitre 4 - Validatation des hypothèses d'une ANOVA à un facteur

Exercice 1

Pour tester la normalité nécessaire pour qu'une ANOVA soit valide, On peut utiliser le test de shapiro-Wilk. Le code ***R*** est le suivant :

```
> shapiro.test(delai)
```

Le résultat est le suivant :

```
        Shapiro-Wilk normality test

data:  delai
W = 0.9496, p-value = 0.1654
```

La probabilité critique est de 0,1654 et dépasse donc le seuil habituel de 5 %. On ne rejette pas significativement l'hypothèse nulle de normalité. L'hypothèse de normalité des délais de guérison en jours est acceptée.

On peut aussi remarquer que l'étude des résidus (faite dans le corrigé dans le fichier chap3c, exercice 2) apporte aussi des renseignements quant à la gaussianité des résidus.

Des tests d'égalité des variances peuvent aussi être menés. Par exemple, avec le test de Levene, le code ***R*** est le suivant :

```
> library(car)
> leveneTest(delai~traitement, data=X)
```

Le résultat est le suivant :

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  4  0.5851 0.6763
      25
```

La probabilité critique est de 0,6763 et dépasse donc le seuil habituel de 5 %. On ne rejette pas significativement l'hypothèse nulle d'homogénéité des variances. L'hypothèse d'égalité des variances est donc acceptée.

On aurait aussi pu faire le test de Bartlett, obtenu via la commande :

```
> bartlett.test(delai~traitement)
```

Le résultat du test est le suivant :

```
        Bartlett test of homogeneity of variances

data:  delai by traitement
Bartlett's K-squared = 2.4197, df = 4, p-value = 0.6591
```

La probabilité critique est de 0,6591 et dépasse donc le seuil habituel de 5 %. On ne rejette pas significativement l'hypothèse nulle d'homogénéité des variances. L'hypothèse d'égalité des variances est donc une fois encore acceptée.

Remarque 0.1 *Si l'hypothèse de normalité avait été violée, on aurait tout de même pu faire un test similaire au test F de l'ANOVA dans ce cas. Il s'agit du test de Kruskal et Wallis. Le code **R** est le suivant :*

```
> kruskal.test(delai~traitement,data=X)
```

Le résultat est le suivant :

```
        Kruskal-Wallis rank sum test

data:  delai by traitement
Kruskal-Wallis chi-squared = 11.01, df = 4, p-value = 0.02645
```

La probabilité critique est de 0,02645 et ne dépasse pas donc le seuil habituel de 5 %. Cela permet de conclure là aussi (voir corrigé de l'exercice 2 du chapitre 3) que les effets d'au moij s deux traitements sont différents.

Exercice 2 à Exercice 6

Les solutions ne sont pas écrites car il s'agit de reproduire, mutatis mutandis, tout ce qui a été détaillé dans l'exercice 1 ci-dessus.

Exercice 7

1. Le code ***R*** pour tracer les boîtes à moustaches est le suivant :

   ```
   > carbu<-rep(1:4,c(6,6,6,6))
   > conso<-c(21,24,25,20,34,17,23,23,32,23,32,15,18,19,28,19,24,14,
   + 20,21,25,15,29,9)
   > carbu<-factor(carbu)
   > don<-data.frame(carbu,conso)
   > moy<-tapply(don$conso,don$carbu,mean)
   > moy
   > ecart<-tapply(don$conso,don$carbu,sd)
   > ecart
   > ecart.g<-sd(don$conso)
   > ecart.g
   > plot(don$carbu,don$conso,col="green")
   > points(1:4,moy,pch="@")
   > abline(h=moy.g)
   ```

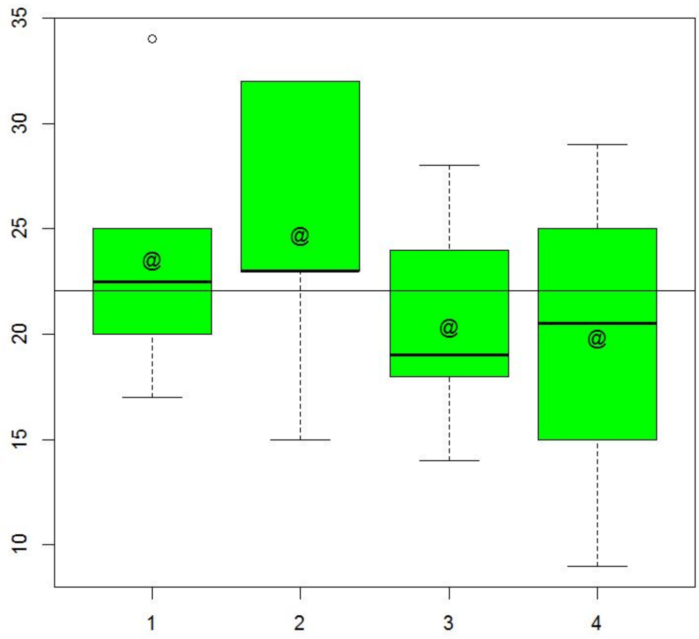

FIGURE 1 – Boîtes à moustaches de consommation par carburateurs

On obtient la figure ci-dessus.

2. La variable à expliquer, *consommation*, est une variable continue. La variable explicative que nous considérons, le type de carburateur, *carburateur*, est qualitative et contrôlée. Le plan qui a été utilisé pour réaliser l'expérience comporte des répétitions, nous pouvons donc essayer de nous servir d'un modèle d'analyse de la variance à un facteur contrôlé.

   Nous introduisons le modèle suivant :

$$Y_{i,j} = \mu + \alpha_i + \varepsilon_{i,j} \qquad , \quad i = 1, \cdots, 4\,,\; j = 1, \cdots, 6\;,$$

   avec la contrainte supplémentaire : $\displaystyle\sum_{i=1}^{4} \alpha_i = 0$,

   où la variable $Y_{i,j}$ est la consommation de la voiture équipée du carburateur i lors du j-ème essai.

   Nous postulons les hypothèses classiques suivantes pour les erreurs :

$$\varepsilon_{i,j} \text{ et } \varepsilon_{k,l} \text{ sont indépendants si } (i,j) \neq (k,l) \text{ avec } 1 \leq i,k \leq 4, \text{ et } 1 \leq j,l \leq 6\,,$$
$$\forall\,(i,j),\; 1 \leq i \leq 4,\; 1 \leq j \leq 6,\; \mathcal{L}(\varepsilon_{i,j}) = \mathcal{N}(0, \sigma^2)\,.$$

   Ce modèle comporte 6 répétitions pour chaque niveau du facteur. Il s'agit donc d'un plan expérimental équilibré.

3. Effectuons un test d'homogénéité des variances des erreurs :

$$H_0 \;:\; \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$$

contre

$$H_1 \;:\; \text{Il existe au moins une variance différente des autres}$$

Ne connaissant pas la loi des erreurs, on peut effectuer un test de levene

```
> library(car)
> leveneTest(conso~carbu, data=don)
```

Cela donne :

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3   0.194 0.8992
      20
```

La probabilité critique valant 0,8992, on ne rejette pas l'hypothèse H_0. On ne rejette pas significativement l'hypothèse nulle d'homogénéité des variances. L'hypothèse d'égalité des variances est donc acceptée.

On s'intéresse maintenant à la normalité de $Y_{i,j}$. Un test de Shapiro-Wilk peut être mené :

```
> shapiro.test(conso)
```

Le résultat est le suivant :

```
        Shapiro-Wilk normality test

data:  conso
W = 0.9804, p-value = 0.9037
```

La probabilité critique valant 0,9037, on ne rejette pas l'hypothèse H_0. Le test n'est donc pas significatif au niveau $\alpha = 5\,\%$, et nous ne rejettons pas l'hypothèse H_0.

Comme l'hypothèse de normalité des erreurs n'a pas été rejetée (voir ci-dessus), on peut s'intéresser à nouveau à l'hypothèse d'homogénéité des variances portant sur les erreurs :

$$H_0 \;:\; \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$$

contre

$$H_1 \;:\; \text{Il existe au moins une variance différente des autres}$$

Il est maintenant possible d'utiliser le test paramétrique de Bartlett puisque ses conditions d'application, la normalité des erreurs, sont vérifiées.

```
> bartlett.test(conso~carbu)
```

Les résultats sont les suivants :

```
        Bartlett test of homogeneity of variances

data:  conso by carbu
Bartlett's K-squared = 0.6503, df = 3, p-value = 0.8848
```

Le test n'est pas significatif car 0,8848 est bien supérieur à $\alpha = 5\,\%$. Nous ne pouvons rejeter l'hypothèse nulle H_0 à ce seuil. Les résultats sont cohérents avec ceux obtenus précédemment.

4. Déterminons si le facteur *carburateur* a un effet sur la *consommation*. Nous testons donc les hypothèses :

$$H_0 \ : \ \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$

contre

$$H_1 \ : \ \text{Il existe } i_0 \in \{1, 2, 3, 4\} \text{ tel que } \alpha_{i_0} \neq 0\,.$$

Le code ***R*** est le suivant :

```
> modele<-aov(conso~carbu,data=don)
> summary(modele)
```

Le tableau d'analyse de variance est le suivant :

```
            Df Sum Sq Mean Sq F value Pr(>F)
carbu        3  100.8   33.61   0.888  0.464
Residuals   20  757.0   37.85
```

La probabilité critique associée à la statistique de Fisher est de 0,464. Elle est bien supérieure à $\alpha = 5\,\%$; le test n'est donc pas significatif à ce seuil et on peut ne peut rejeter l'hypothèse nulle H_0 d'absence d'effet du facteur *carburateur* sur la *consommation*. Donc, au niveau $\alpha = 5\,\%$, le facteur *carburateur* n'a pas d'effet significatif sur la *consommation*.

Les estimations des coefficients du modèle sont données via le code ***R*** suivant :

```
> library(FactoMineR)
> AovSum(conso~carbu,data=don)
```

On obtient alors directement via cette commande, et le tableau de l'analyse de la variance et les estimations des coefficients :

```
$Ftest
              SS df     MS F value Pr(>F)
carbu     100.83  3 33.611   0.888 0.4643
Residuals 757.00 20 37.850

F test
             Estimate Std. Error    t value     Pr(>|t|)
(Intercept) 22.083333   1.255820 17.5847949 1.240826e-13
carbu - 1    1.416667   2.175144  0.6512980 5.222681e-01
carbu - 2    2.583333   2.175144  1.1876610 2.488811e-01
carbu - 3   -1.750000   2.175144 -0.8045446 4.305431e-01
carbu - 4   -2.250000   2.175144 -1.0344144 3.132890e-01
```

Nous remarquons que le carburateur A_4 nous fait consommer le moins de carburant. Néanmoins, puisque nous n'avons pu rejeter H_0 d'absence d'effet du facteur *carburateur* sur la *consommation* au seuil $\alpha = 5\,\%$, ce comportement n'est pas significativement différent des autres à ce seuil.