Evaluation des stocks de carbone. Corrigé

L'estimation des stocks de carbone en forêt tropicale repose sur l'estimation de la biomasse aérienne des arbres (AGB : Above Ground Biomass). Pour estimer cette grandeur, on utilise des relations allométriques qui lient la valeur AGB au diamètre D, à la hauteur H et à la densité des arbres ρ. Des arguments géométriques simples suggèrent que la grandeur AGB pour un arbre de diamètre D doit être proportionnelle au produit de la densité spécifique ρ de l'arbre par la surface basale du tronc ($\pi * D^2/4$) et la hauteur H, ce qui conduit à une expression générique

$$AGB = F \times \rho \times \left(\frac{\pi D^2}{4} \times H\right), \tag{1}$$

où F est un coefficient de proportionalité inconnu.

Cette relation suppose que le tronc est comparable à un cylindre de densité homogène. Pour alléger cette hypothèse, la relation allométrique la plus couramment suggérée, notamment dans [**?**], est~:

$$AGB = F \times \left(\rho \times \left(\frac{\pi D^2}{4} \times H\right)\right)^{\beta}, \tag{2}$$

où β est un coefficient à estimer.

Pour estimer la quantité de biomasse contenue dans une parcelle plantée d'arbres de la même espèce, on mesure pour chaque arbre, son diamètre, sa hauteur et sa densité et on mesure après séchage sa biomasse au-dessus du sol. Nous disposons de 60 mesures d'arbres.

Le diamètre, comme la hauteur sont exprimés en mètres. Nous avons également ajouté les logarithmes de ces variables : `DLog`$=\log(D)$, `HLog`$=\log(H)$, `rhoLog`$=\log(\rho)$, `AGBLog`$=\log(AGB)$. Un extrait des données est fourni par

```
head(AGBParcelle, n=5)
```

```
##      D  H  rho  AGB        DLog     HLog     rhoLog    AGBLog
## 1 0.90 53 0.46 2.41 -0.10536052 3.970292 -0.7765288 0.8796267
## 2 0.86 53 0.57 2.48 -0.15082289 3.970292 -0.5621189 0.9082586
## 3 0.64 32 0.59 1.15 -0.44628710 3.465736 -0.5276327 0.1397619
## 4 0.77 49 0.33 1.35 -0.26136476 3.891820 -1.1086626 0.3001046
## 5 0.97 66 0.54 3.35 -0.03045921 4.189655 -0.6161861 1.2089603
```

1. On se propose de modéliser le logarithme de AGB (`AGBLog`) en fonction du logarithme du diamètre (`DLog`), du logarithme de la hauteur (`HLog`) et du logarithme de la densité (`rhoLog`). Écrire le modèle linéaire associé $\mathcal{M}_1$ en spécifiant toutes les hypothèses.

On prend le log de l'equation (2):

$$\log AGB = \log F + \beta \log\left(\frac{\pi}{4}\right) + 2\beta \log D + \beta \log H + \beta \log \rho$$

On écrit le modèle suivant :

$$Y_i = \log AGB_i = \underbrace{\log F + \beta \log\left(\frac{\pi}{4}\right)}_{\beta_0} + \underbrace{2\beta}_{\beta_1} \log D + \underbrace{\beta}_{\beta_2} \log H_i + \underbrace{\beta}_{\beta_3} \log \rho_i + \varepsilon_i$$

2. Écrire ce même modèle sous forme matricielle, en donnant les dimensions des différents objets considérés.

   Immediat

3. Si la relation (1) est vérifiée, quelles seront les valeurs attendues pour les paramètres du modèle $\mathcal{M}_1$?

   On attend $\beta = 1$ d'où
   $$\beta_1 = 2\beta = 2, \quad \beta_2 = 1, , \quad \beta_3 = 1$$

   On ajuste le modèle $\mathcal{M}_1$ et on obtient les sorties suivantes:

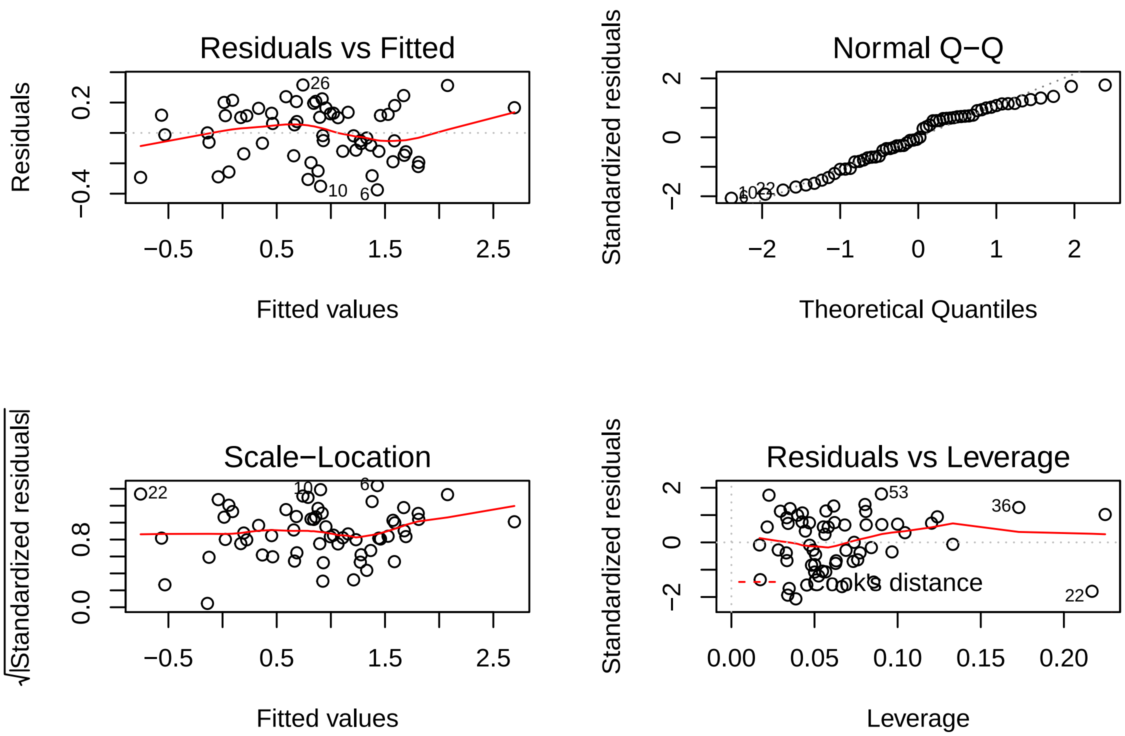
```
M1 <- lm(AGBLog~DLog+HLog+rhoLog, data = AGBParcelle )
summary(M1)
```

```
##
## Call:
## lm(formula = AGBLog ~ DLog + HLog + rhoLog, data = AGBParcelle)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37482 -0.12748 -0.00588  0.13141  0.31588
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.16282    1.05758  -0.154   0.8782
## DLog         2.36365    0.33010   7.160 1.88e-09 ***
## HLog         0.43195    0.25319   1.706   0.0935 .
## rhoLog       0.48035    0.07981   6.019 1.42e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1849 on 56 degrees of freedom
## Multiple R-squared:  0.9354, Adjusted R-squared:  0.9319
## F-statistic: 270.3 on 3 and 56 DF,  p-value: < 2.2e-16
```

```
M0 <- lm(AGBLog~1, data = AGBParcelle )
anova(M0,M1)
```

```
## Analysis of Variance Table
##
## Model 1: AGBLog ~ 1
## Model 2: AGBLog ~ DLog + HLog + rhoLog
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     59 29.6366
## 2     56  1.9146  3    27.722 270.28 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow=c(2,2))
plot(M1)
```



4. Les hypothèses du modèle linéaire sont-elles vérifiées ? Préciser quel graphique est utilisé pour chaque hypothèse.

> Indépendance des observations : dues au protocole expérimental. Résidus centrés car graphe en haut à gauche : pas de tendance. Résidus gaussiens car points alignés sur la droite du Normal Q-Q. Et modèle homoscedastique car pas de grosse tendance dans le graphe en bas à gauche. En outre, pas de point abhérrant à noter sur le graphe des Residuals versus Leverage.

5. Donner les hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$ du test du modèle complet. Donner l'expression de la statistique de test et sa loi sous $\mathcal{H}_0$. On spécifiera les degrés de liberté. Donner la réalisation de la statistique de test sur ces donnée ainsi que la p-value. Conclure.

- $\mathcal{H}_0 : \beta_1 = \beta_2 = \beta_3 = 0,$ versus $\mathcal{H}_1 : \beta_1 \neq 0,$ or $\beta_2 \neq 0$ or $\beta_3 \neq 0$.
-
$$F = \frac{(\mathrm{SCR}_0 - \mathrm{SCR}\ )/(p+1-1)}{\mathrm{SCR}\ /(n-(p+1))} \sim \mathcal{F}_{p,n-(p+1)}$$
avec
$$\mathrm{SCR}_0 = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \mathrm{SCT}$$
.
$$\mathrm{SCR}\ = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$
- On retrouve bien $p = 3, n-(p+1) = 60-4 = 56$. Réalisation $= 270.3$ et p-value $< 2.2e{-}16$ donc on rejette $\mathcal{H}_0$.

6. SCT $= 29.6366$, SCR $= 1.9146$ et SCM $=$ SCT $-$ SCR $= 27.722$.

7. Donner l'expression de l'estimateur de la variance résiduelle et la valeur de cette estimation.

$\hat{\sigma}^2 = \frac{SCT}{n-(p+1)}$. Attention : ici on donne l'écart type estimé. Il faut en prendre le carré $0.1849^2 = 0.034$

8. Quelle est l'hypothèse $\mathcal{H}_0$ du test proposé sur la ligne Dlog de la commande summary ? Comment l'interprète-t-on ? Quelle est la loi de la statistique de test ?

On test $\beta_1 = 2\beta = 0$. Stat de test :
$$T = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 (X'X)_{11}}} \sim \mathcal{T}_{60-(3+1)}$$
On rejette l'hypothèse que $\beta = 0$.

9. Que pouvez-vous dire de l'ajustement du modèle aux données ?

Le R^2 est très haut. Donc la formule semble correcte.

10. Quel test pourriez-vous proposer pour tester si l'effet du diamètre est bien celui prévu par la relation allométrique de l'équation (1) ?

Il faudrait tester :
$$\beta_1 = 2\beta_2 = 2\beta_3 = 1$$
.