Chapitre 3, Étude de cas : CHAMPA (page 17-18)

Dans une première étape, tester le modèle log-linéaire suivant.

$$\ln Y = \mu + \beta_1 \ln X^{(1)} + \beta_2 \ln X^{(2)} + \beta_3 \ln X^{(3)}.$$

Les résultats sont présentés dans le tableau suivant.

```
Call: lm(formula = lnY ~ lnX1 + lnX2 + lnX3)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.51721    0.74064   0.698   0.4973
lnX1         2.37141    0.12950  18.312 1.15e-10 ***
lnX2        -1.37394    0.24060  -5.711 7.17e-05 ***
lnX3        -0.10799    0.05951  -1.815   0.0927 .

Residual standard error: 0.04165 on 13 degrees of freedom
Multiple R-squared:  0.9925,Adjusted R-squared:  0.9908
F-statistic: 573.9 on 3 and 13 DF,  p-value: 4.644e-14
```

1. Le modèle est il globalement significatif ?

Dès lors que le modèle comporte plusieurs variables explicatives, le test de Fisher permet de voir si le modèle est globalement significatif (i.e. si au moins une variable explicative a une action sur la variable à expliquer. On a ici le p-value de F-statistic 4.644e-14 < 0.0001. Le test global est donc significatif.

2. Quelle est la signification économique des coefficients ? Les coefficients ont-ils un signe conforme à vos attentes ?

Nous sommes devant un modèle log-linéaire. On peut donc le réécrire comme suit $Y = e^{\mu}(X^{(1)})^{\beta_1}(X^{(2)})^{\beta_2}(X^{(3)})^{\beta_3}$. Les coefficient β_1, β_2 et β_3 sont des élasticités.

Plus le revenu augemente, plus la consommation de champagne doit augmenter car le chapagne est un bien supérieur. On s"attend donc à ce que $\hat{\beta}_1 > 0$.

Plus le prix du champagne augmente, plus la consommation de champagne doit diminuer. On s'attend donc à ce que $\hat{\beta}_2 < 0$.

Plus le prix des liqueur et apéritifs augmente, plus la consommation de champagne doit augmenter (hypothèse de substituabilité). $\hat{\beta}_3$ devrait donc être positif.

On retrouve ainsi les coefficients que l'on attendait sauf pour $X^{(3)}$.

3. Une variable ne doit-elle pas être éliminée ? Laquelle ? Pourquoi ?

La variable $X^{(3)}$ doit être éliminée car le p-value du test de Student 0.0927 est supérieur à 0.05.

Dans une seconde étape, tester le modèle log-linéaire suivant.

$$\ln Y = \mu + \beta_1 \ln X^{(1)} + \beta_2 \ln X^{(2)}$$

Les résultats sont présentés dans le tableau suivant.

```
Call: lm(formula = lnY ~ lnX1 + lnX2)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.1073     0.7609   0.141 0.889863
lnX1          2.1986     0.0947  23.216 1.41e-12 ***
lnX2         -1.2191     0.2427  -5.023 0.000186 ***

Residual standard error: 0.04493 on 14 degrees of freedom
Multiple R-squared:  0.9906,Adjusted R-squared:  0.9893
F-statistic: 738.3 on 2 and 14 DF,  p-value: 6.447e-15
```

4. Ce modèle est il statistiquement satisfaisant ?
Oui, car il est globalement significatif, et chaque variable explicative est aussi significative.

5. Testez l'hypothèse $\beta_1 > 1$.
C'est un test unilatéral droit. D'abord on calcule la statistique

$$\hat{t} = \frac{\hat{\beta}_1 - 1}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{2.1986 - 1}{0.0947} \approx 12.66.$$

Ensuite, on cherche la valeur critique dans la table de loi Student t_α avec $\alpha = \mathbb{P}\left(t_{n-k-1} > t_\alpha\right)$ où $n = 17, k = 2$ et $\alpha = 0.05$. On trouve $t_\alpha = 1.761$. Puisque $\hat{t} > t_\alpha$, β_1 est significativement supérieur de 1.

6. Testez l'hypothèse " la baisse des prix du champagne entraîne un progrès de la consommation en volume ". Peut on dire la même chose de la consommation en valeur ?
C'est un test unilatéral gauche avec $H_1 : \beta_2 < 0$. D'abord on calcule la statistique

$$\hat{t} = \frac{\hat{\beta}_2 - 0}{\hat{\sigma}_{\hat{\beta}_2}} = \frac{-1.2191 - 0}{0.2427} \approx -5.02.$$

Ensuite, on cherche la valeur critique dans la table de loi Student t_α avec $\alpha = \mathbb{P}\left(t_{n-k-1} < t_\alpha\right)$ où $n = 17, k = 2$ et $\alpha = 0.05$. On trouve $t_\alpha = -1.761$. Puisque $\hat{t} < t_\alpha$, β_2 est significativement négatif.

Pour avoir la consommation de champagne en valeur, il suffit de faire le produit de la consommation de champagne en volume par les prix. Soit la consommation de chmpagne en valeur $= YX^{(2)} = e^\mu (X^{(1)})^{\beta_1} (X^{(2)})^{\beta_2 + 1}$. Ainsi la consommation de champagne en valeur augmente quand le prix du champagne diminue si $\beta_2 + 1 < 0$. On teste donc cette fois ci $H_0 : \beta_2 = -1$ contre $H_1 : \beta_2 < -1$. C'est un test unilatéral gauche. D'abord on calcule la statistique

$$\hat{t} = \frac{\hat{\beta}_2 - (-1)}{\hat{\sigma}_{\hat{\beta}_2}} = \frac{-1.2191 + 1}{0.2427} \approx -0.90.$$

Ensuite, on cherche la valeur critique dans la table de loi Student t_α avec $\alpha = \mathbb{P}\left(t_{n-k-1} < t_\alpha\right)$ où $n = 17, k = 2$ et $\alpha = 0.05$. On trouve $t_\alpha = -1.761$. Puisque $\hat{t} > t_\alpha$, on ne peut pas rejeter l'hypothèse nulle, β_2 n'est pas significativement plus petit que -1.

7. Présentez de manière synthétique les résultats obtenus.
Dans la régression descendante, on procède par éliminations successives des variables non significatives. C'est la méthode que nous venons juste d'utiliser.

Étape 1 : On part du modèle avec toutes les variables exogènes.

Étape 2 : On écarte la moins significative des variables non-significatives. C'est-à-dire on teste la significativité de toutes les variables. Si elle est établie, alors ARRÊT. Si non, on sort la variable dont le $\hat{t}$ est le plus faible.

Étape 3 : On recommence Étape 2 jusqu'à l'obtention d'un modèle qui ne contienne que des variables significatives.

8. La régression ascendante donne t il les mêmes résultats que la régression descendante ? Qu'en concluez-vous ?

La régression ascendante procède dans l'autre sens : par sélection successives.

Étape 1 : On part de rien. On sélectionne la variable qui contribue le mieux à expliquer la variable Y, c.a.d. celle dont le coefficient de corrélation est le plus élevé. On vérifie avec un test de Fisher si l'apport d'explication de cette variable est significatif. Si non, alors ARRÊT.

Étape 2 : On sélectionne parmi les variables restantes, celle qui contribuerait le mieux à expliquer le résidu estimé, c.a.d. celle dont le coefficient de détermination partielle est le plus élevé. On vérifie par un test du Fisher partiel que le supplément d'explication est significatif. Si ce n'est pas le cas, alors ARRÊT.

Étape 3 : On recommence Étape 2 jusqu'à la situation où aucune variable restante n'apporte un supplément significatif d'explication des variations de Y.

Les régressions ascendante et descendante peuvent ne pas donner les mêmes résultats. Lorsque l'on obtient les mêmes conclusions avec les deux méthodes, alors cela augmente le degré de confiance que l'on peut avoir dans les résultats obtenus.

Chapitre 4, Exercice (page 22)

Notons $F_{M_n}(y)$ la fonction de répartition (f.d.r.) de M_n, $F_{X_1}(y)$ la f.d.r. de X_1. On a

$$\mathbb{P}(M_n \leq y) = \mathbb{P}(\max(X_1, \ldots, X_n) \leq y) = \mathbb{P}(X_1 \leq y, \ldots, X_n \leq y) = (\mathbb{P}(X_1 \leq y))^n.$$

C.a.d.

$$F_{M_n}(y) = (F_{X_1}(y))^n. \tag{1}$$

<u>**Weibull**</u>

D'après (1), on a

$$F_{M_n}(y) = (\Psi_\alpha(y))^n = (e^{-(-y)^\alpha})^n = e^{-n(-y)^\alpha}, \quad y \leq 0. \tag{2}$$

La f.d.r. de $n^{-1/\alpha}X$ est, pour $y \leq 0$,

$$F_{n^{-1/\alpha}X}(y) = \mathbb{P}(n^{-1/\alpha}X \leq y) = \mathbb{P}(X \leq n^{1/\alpha}y) = \Psi_\alpha(n^{1/\alpha}y) = e^{-n(-y)^\alpha}. \tag{3}$$

Les inégalités (2) et (3) impliquent $M_n \stackrel{\mathcal{L}}{=} n^{-1/\alpha}X$.

<u>**Gumbel**</u>

D'après (1), on a

$$F_{M_n}(y) = (\Lambda(y))^n = (e^{-e^{-y}})^n = e^{-ne^{-y}}. \tag{4}$$

La f.d.r. de $X + \log(n)$ est

$$F_{X+\log(n)}(y) = \mathbb{P}(X + \log(n) \leq y) = \mathbb{P}(X \leq y - \log(n)) = \Lambda(y - \log(n)) = e^{-ne^{-y}}. \tag{5}$$

Les inégalités (4) et (5) impliquent $M_n \overset{\mathcal{L}}{=} X + \log(n)$.

<u>**Fréchet**</u>
D'après (1), on a

$$F_{M_n}(y) = (\Phi_\alpha(y))^n = (e^{-y^{-\alpha}})^n = e^{-ny^{-\alpha}}, \quad y > 0. \tag{6}$$

La f.d.r. de $n^{1/\alpha}X$ est, pour $y > 0$,

$$F_{n^{1/\alpha}X}(y) = \mathbb{P}(n^{1/\alpha}X \leq y) = \mathbb{P}(X \leq n^{-1/\alpha}y) = \Phi_\alpha(n^{1/\alpha}y) = e^{-ny^{-\alpha}}. \tag{7}$$

Les inégalités (6) et (7) impliquent $M_n \overset{\mathcal{L}}{=} n^{1/\alpha}X$.