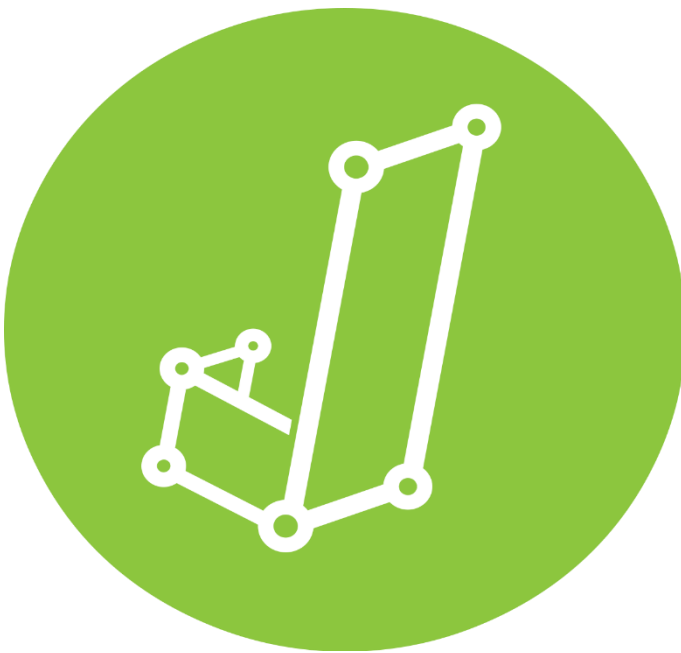


# APPROFONDISSEMENT LOGICIELS

##  tude du logiciel JASP



BARRY Mamadou Yaya

**R alis  par :** TANNANI Manal  
ZYATE Yassine

**Professeur :** M. COMPAIRE Philippe

# SOMMAIRE

Introduction .....	1
I. Étude Économétrique .....	2
II. Résultats sous JASP et R .....	2
II.1. Présentation du nuage de points .....	2
II.2. Estimation des coefficients du modèle par OLS .....	3
II.3. Calcul et Représentation des résidus .....	3
II.4. Graphique de la consommation observée – estimée (sous R) .....	4
II.5. Calcul de l'estimation de la variance des résidus, de la variance de $b$ , de la variance de $a$ et de la covariance entre $a$ et $b$ . .....	5
II.6. Calcul du coefficient $R^2$ .....	5
II.7.1. Test du T-ratio (ou t-test) au seuil $\alpha = (0.05, 0.01 \text{ et } 0.1)$ pour $H_0$ .....	5
II.7.2. Test du T-ratio (ou t-test) au seuil $\alpha = (0.05, 0.01 \text{ et } 0.1)$ pour $H_1$ .....	6
II.8. Test de Durbin-Watson .....	6
II.9. Test de Ljung-Box avec $p=2$ .....	7
II.10. Test d'ARCH avec $p=1$ .....	7
II.11. Prévision sur le passé avec les tests de fiabilité .....	8
II.12. Prévision avec $RD = 300$ .....	8
II.13. Prévision avec $CPC = 400$ .....	9
Conclusion.....	9
Bibliographie .....	10

# Introduction

Dans le cadre de notre cours sur l'approfondissement des logiciels économétriques, nous avons fait le choix de faire une étude de cas en présentant un nouveau logiciel non vu en cours. Mais aussi de mettre en exergue ses limites qui seront compensées par le logiciel R. C'est ainsi que nous avons choisi le logiciel JASP pour son côté pratique afin d'essayer d'effectuer quelques régressions selon l'énoncé du projet.

Le logiciel « JASP » est un programme graphique gratuit doté d'une open-source permettant de faire des analyses statistiques. Il est soutenu par l'Université d'Amsterdam et est conçu pour faciliter l'utilisation aux utilisateurs de SPSS. Il propose des procédures d'analyse standard sous leur forme classique et bayésienne. JASP produit généralement des tableaux de résultats et des tracés de style APA pour soutenir la publication. Il favorise la science ouverte via des Frameworks en Open-source qui intègre les paramètres d'analyse dans les résultats. Le développement de JASP est aujourd'hui soutenu financièrement par plusieurs fonds de recherche et des universités. JASP propose l'inférence fréquentielle et bayésienne sur les mêmes modèles statistiques. Cette inférence fréquentielle utilise des probabilités et des intervalles de confiance pour contrôler les taux d'erreur dans la limite des réplifications parfaites infinies. Quant à l'inférence bayésienne, elle utilise intervalles crédibles et des facteurs bayésiennes afin d'estimer les valeurs des paramètres crédibles et les preuves du modèle compte tenu des données disponibles et des connaissances antérieures.

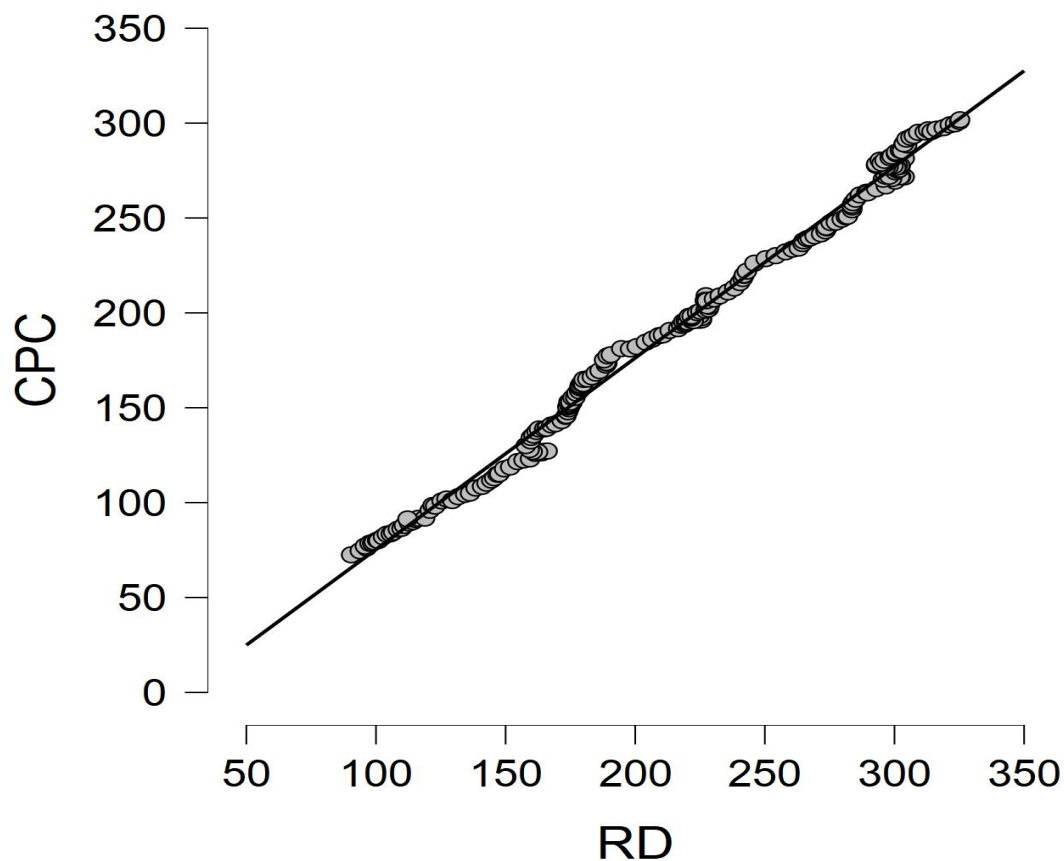
Dans un premier temps, nous répondrons aux questions via le logiciel JASP afin de voir si nos résultats suivent la même logique que ceux obtenus en cours d'approfondissement logiciels économétriques. Ensuite dans un deuxième temps nous essaierons de compléter les questions non réalisables sur JASP par les résultats du logiciel R. Enfin nous nous appuierons sur une conclusion pour présenter les avantages et inconvénients du logiciel JASP.

# I. Étude Économétrique

Pour la réalisation de ce projet, nous disposons d'une base de données trimestrielles de 1963 à 2019. Base à partir de laquelle, nous allons effectuer une étude économétrique visant à expliquer la consommation (« CPC ») en fonction du revenu disponible (« RD »).

## II. Résultats sous JASP et R

### II.1. Présentation du nuage de points



L'équation estimée est :  $CPC_t = -25.503 + 1.009 RD_t$  avec  $\hat{a} = -25,503$  et  $\hat{b} = 1,009$

Ce nuage de points ou diagramme de dispersion représente la distribution de la consommation en fonction du revenu disponible. Nous observons une tendance croissante et linéaire entre la consommation et le revenu. Cela explique bien la corrélation positive (toute augmentation du

revenu disponible entraîne une augmentation moins proportionnelle du niveau de consommation). Il y a également absence de valeurs aberrantes.

## II.2. Estimation des coefficients du modèle par OLS

### Model Summary - CPC

Model	R	R <sup>2</sup>	Adjusted R <sup>2</sup>	RMSE	Durbin-Watson		
					Autocorrelation	Statistic	p
Ho	0.997	0.993	0.993	5.519	0.966	0.062	< .001

Note. Null model includes RD

### Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
Ho	(Intercept)	-25.503	1.245		-20.479	< .001
	RD	1.009	0.005	0.997	183.893	< .001

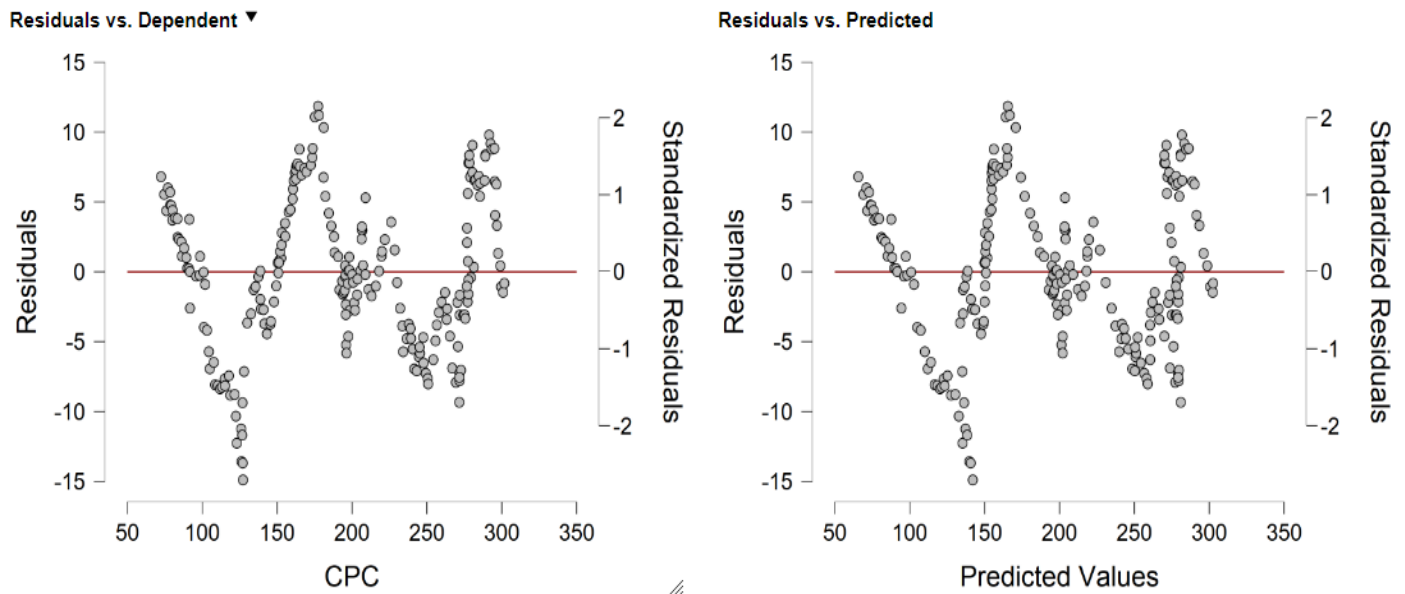
La P-value est inférieure à 1% (de même pour 5%) nous pouvons alors affirmer que le modèle estimé est globalement significatif (les coefficients estimés sont significativement différents de zéro).

Pour l'intercept ( $\hat{\alpha} = -25,503$ ), nous dirons qu'en l'absence d'un revenu disponible, la consommation baisse ceteris paribus (toutes choses étant égales par ailleurs) de 25,503 points. Quant au revenu disponible ( $\hat{\beta} = 1,009$ ), nous dirons que son augmentation d'un point produit ceteris paribus, un effet marginal de l'augmentation du revenu sur la consommation de 1,009 point. Entraînant ainsi une baisse du niveau de consommation à 24,494 (25,503-1,009).

## II.3. Calcul et Représentation des résidus

### Residuals Statistics

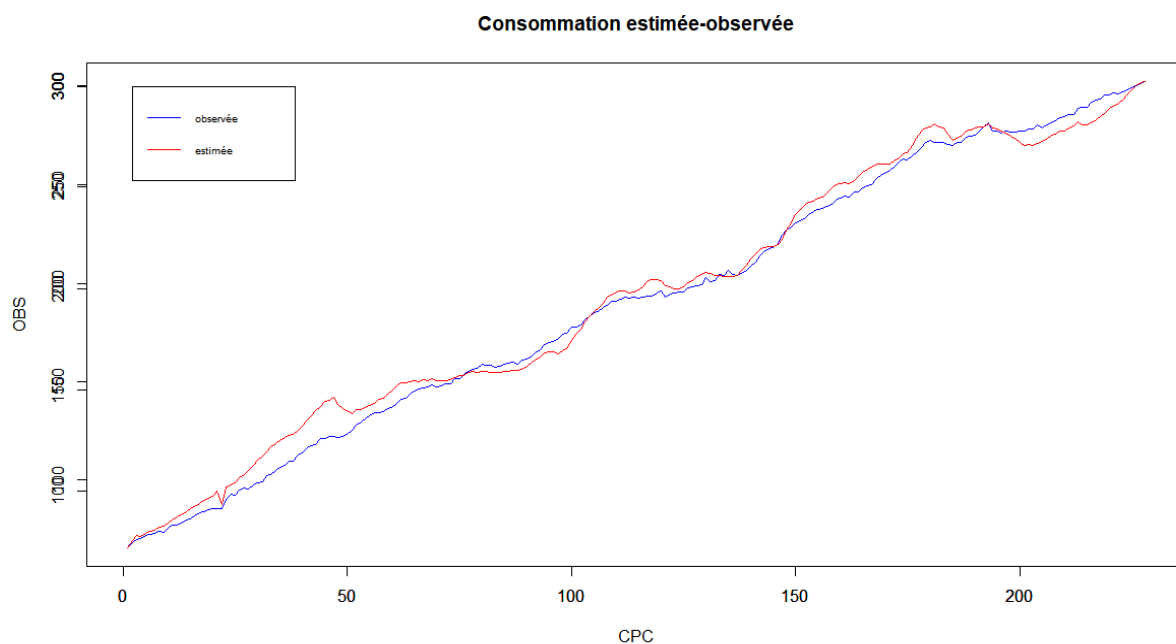
	Minimum	Maximum	Mean	SD	N
Predicted Value	65.719	302.583	193.413	67.367	228
Residual	-14.905	11.845	-9.399e -17	5.507	228
Std. Predicted Value	-1.896	1.621	9.404e -17	1.000	228
Std. Residual	-2.710	2.152	1.708e -4	1.002	228



Par observation graphique, la distribution des résidus (partie non observée par l'équation de régression) ne semble être stationnaire. On peut donc affirmer par-là, qu'il y a absence de bruit blanc.

## II.4. Graphique de la consommation observée – estimée (sous R)

Le logiciel JASP ne permet pas de représenter sur le même graphique la consommation observée et la consommation estimée. Cependant, nous allons réaliser le graphique sous R afin de voir si nous obtenons la même chose que sur SAS.



## II.5. Calcul de l'estimation de la variance des résidus, de la variance de $\hat{b}$ , de la variance de $\hat{a}$ et de la covariance entre $\hat{a}$ et $\hat{b}$ .

				Variance Proportions	
Model	Dimension	Eigenvalue	Condition Index (Intercept)		RD
H <sub>0</sub>	1	1.956	1.000	0.022	0.022
	2	0.044	6.664	0.978	0.978
Coefficients Covariance Matrix					
Model		RD			
H <sub>0</sub>	RD	3.009e -5			

## II.6. Calcul du coefficient R<sup>2</sup>

Le coefficient de détermination (**R<sup>2</sup> = 0.993**) est très élevé. Cela prouve la forte liaison entre les deux variables. C'est-à-dire que le revenu disponible explique 99.3% de la variation de la consommation. La qualité du modèle est ainsi traduite.

### II.7.1. Test du T-ratio (ou t-test) au seuil $\alpha = (0.05, 0.01 \text{ et } 0.1)$ pour H<sub>0</sub>

#### Coefficients

						95% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower Upper
H <sub>0</sub>	(Intercept)	-25.503	1.245		-20.479	< .001	-27.957 -23.049
	RD	1.009	0.005	0.997	183.893	< .001	0.998 1.020

\*  $\alpha = 0.05$

#### Coefficients

						99% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower Upper
H <sub>0</sub>	(Intercept)	-25.503	1.245		-20.479	< .001	-28.738 -22.268
	RD	1.009	0.005	0.997	183.893	< .001	0.994 1.023

\*  $\alpha = 0.01$

#### Coefficients

						90% CI	
Model		Unstandardized	Standard Error	Standardized	t	p	Lower Upper
H <sub>0</sub>	(Intercept)	-25.503	1.245		-20.479	< .001	-27.559 -23.446
	RD	1.009	0.005	0.997	183.893	< .001	1.000 1.018

\*  $\alpha = 0.1$

Les T-ratios sont significatifs dans les 3 cas car leurs p-values sont inférieures à 1%.

## II.7.2. Test du T-ratio (ou t-test) au seuil $\alpha = (0.05, 0.01 \text{ et } 0.1)$ pour $H_1$

### Coefficients

				t	p	95% CI	
Model		Unstandardized	Standard Error			Lower	Upper
H <sub>1</sub>	(Intercept)	193.413	4.476	43.208	< .001	184.593	202.234
	RD	1.009	0.005	0.997	183.893 < .001	0.998	1.020

\*  $\alpha = 0.05$

### Coefficients

				t	p	99% CI	
Model		Unstandardized	Standard Error			Lower	Upper
H <sub>1</sub>	(Intercept)	193.413	4.476	43.208	< .001	181.785	205.041
	RD	1.009	0.005	0.997	183.893 < .001	0.994	1.023

\*  $\alpha = 0.01$

### Coefficients

				t	p	90% CI	
Model		Unstandardized	Standard Error			Lower	Upper
H <sub>1</sub>	(Intercept)	193.413	4.476	43.208	< .001	186.020	200.806
	RD	1.009	0.005	0.997	183.893 < .001	1.000	1.018

\*  $\alpha = 0.1$

Les T-ratios sont aussi significatifs dans les 3 cas car leurs p-values sont inférieures à 1%.

## II.8. Test de Durbin-Watson

Le test de Durbin-Watson (1950) permet de détecter une autocorrélation des erreurs d'ordre un. Il repose sur l'estimation d'un modèle autorégressif de premier ordre pour les résidus estimés. La statistique de Durbin-Watson est comprise entre 0 et 4. L'hypothèse nulle d'absence d'autocorrélation des erreurs est acceptée lorsque la valeur de cette statistique est proche de 2. Des valeurs critiques au seuil de 5% (resp.  $d_1$  et  $d_2$ , avec  $d_1 > d_2$ ) ont été tabulées. L'interprétation du test de Durbin et Watson est alors la suivante :

- Si la valeur calculée de la statistique DW est inférieure à la valeur tabulée  $d_1$  alors il existe une autocorrélation positive (ou  $p > 0$ ).
- Si la valeur calculée de la statistique DW est comprise entre  $d_2$  et  $4 - d_2$ , il n'est pas possible de rejeter l'hypothèse nulle d'absence d'autocorrélation des résidus. Cet intervalle est autrement dit l'intervalle pour il n'existe pas d'autocorrélation des erreurs.



- Si la valeur calculée de la statistique DW est supérieure à la valeur tabulée 4-d1 alors il existe une autocorrélation négative (ou  $p < 0$ ).

#### Model Summary - CPC

Model	R	R <sup>2</sup>	Adjusted R <sup>2</sup>	RMSE	Durbin-Watson		
					Autocorrelation	Statistic	p
H <sub>0</sub>	0.997	0.993	0.993	5.519	0.966	0.062	< .001

Note. Null model includes RD

DW = 0.062 (inférieure à 2). Donc il y a bel et bien présence d'autocorrélation des résidus.

## II.9. Test de Ljung-Box avec $p=2$

N'existant pas sous JASP, nous le réalisons sur R. Ce test est une autre façon de tester l'autocorrélation des résidus.

#### Box-Ljung test

```
data: RD
x-squared = 443.71, df = 2, p-value < 2.2e-16
```

La p-value est très largement significative au seuil 1% avec  $p=2$ . Donc il y a là aussi, présence d'autocorrélation des individus.

## II.10. Test d'ARCH avec $p=1$

Le test ARCH consiste à effectuer une régression autorégressive des résidus carrés sur  $q$  retard avec les hypothèses suivantes :

$H_0$  : homoscedasticité et  $\alpha_1 = \dots \alpha_q = 0$

$H_1$  : hétéroscedasticité et il y a au moins un coefficient  $\alpha_i$  significativement différent de 0.

#### ARCH LM-test; Null hypothesis: no ARCH effects

```
data: RD
chi-squared = 226.88, df = 1, p-value < 2.2e-16
```

## II.11. Pr vision sur le pass  avec les tests de fiabilit 

	fit	lwr	upr
1	65.71856	64.17224	67.26489
2	69.10890	67.59463	70.62317
3	71.73264	70.24304	73.22224
4	71.07847	69.58273	72.57421
5	72.89377	71.41505	74.37249
6	73.92384	72.45475	75.39293
7	74.56203	73.09889	76.02516
8	75.85373	74.40263	77.30483
9	76.33854	74.89195	77.78514
10	78.13326	76.70331	79.56321
11	79.61913	78.20291	81.03535
12	81.16546	79.76348	82.56743
13	82.09693	80.70350	83.49035
14	83.88811	82.51109	85.26514
15	85.33680	83.97298	86.70062
16	86.25992	84.90449	87.61535
17	88.29800	86.96102	89.63498
18	89.74318	88.41922	91.06714
19	90.85805	89.54410	92.17200
20	91.71912	90.41287	93.02536
21	94.44555	93.16357	95.72753
22	87.66039	86.31765	89.00314
23	96.32138	95.05599	97.58678
24	97.41623	96.16046	98.67200
25	98.40490	97.15779	99.65201
26	100.94070	99.71567	102.16572
27	102.83444	101.62577	104.04311
28	105.07615	103.88669	106.26562
29	107.34223	106.17200	108.51247
30	110.10543	108.95839	111.25247

Pr vision sur les 30 premiers trimestres.

## II.12. Pr vision avec $RD = 300$

	fit	lwr	upr
1	277.1061	266.1693	288.0428

On constate que pour un niveau de revenu disponible  gal   300, la consommation pr vue est de 227.1061.

## II.13. Pr vision avec CPC = 400

	fit	lwr	upr
1	420.4748	409.4864	431.4632

Pour une consommation future de 400, le revenu disponible   cet effet est estim    420.4748.

## Conclusion

En somme, JASP est un logiciel en cours de r alisation et con u pour Data Mining et quelques fonctions en rapport avec la r gression lin aire simple et multiple : d'o  l'absence de certaines fonctionnalit s (tests). Par contre il est possible de faire des mod les ANOVA dans les analyses de base mais pas sur des tests sp cifiques. Nous avons,   travers cet outil, obtenu les m mes r sultats en ce qui concerne les analyses effectu es sur SAS et R.

Un des avantages dont il dispose, est son interface simple qui ne n cessite aucun effort de compr hension. Et parmi ces inconv nients, nous dirons qu'il ne permet pas de r aliser des transformations de donn es (telles que le logarithme des variables), les analyses statistiques approfondies, les illustrations graphiques et tant d'autres disponibles sur les autres logiciels (tels que SAS, STATA, R et PYTHON. Cependant il propose plusieurs m thodes statistiques et des r sultats clairs selon l'utilisation et les sp cialistes de la statistique.

# Bibliographie

FIELD, A.P., MILES, J., & FIELD, Z. (2012), « Discovering statistics using R. », *London : Sage*.

LJUNG, G. M. and BOX, G. E. P. (1978), « On a measure of lack of fit in time series models ». *Biometrika*, 65, pp.297–303.

MOORE, D.S., MCCABE, G.P., & CRAIG, B.A. (2012), « Introduction to the practice of statistics », (7th ed.). *New York, NY : W.H. Freeman and Company*.

SELLKE, T., BAYARRI, M. J., & BERGER, J. O. (2001), « Calibration of p values for testing precise null hypotheses. », *The American Statistician*, 55(1), pp.62-71.

STEVENS, J.P. (2009), « Applied multivariate statistics for the social sciences », (5th ed.). *New York, NY : Routledge*.