

Examen (DM) du 25 mars 2020, à rendre au plus tard le 27 mars à 23h59.
Les étudiants qui ont un tiers temps peuvent rendre le DM jusqu'au 28 mars à 23h59.
Les réponses doivent être individuelles!

Vous scannez les feuilles avec vos réponses, ou vous les prenez en photo et vous m'envoyez le fichier avec vos réponses à mon adresse mail: gabriela.ciuperca@univ-lyon1.fr

Avant de commencer à rédiger vos réponses, lisez avec attention ces consignes: *Les exercices 1 et 2 ont été réalisés avec SAS, pendant que l'exercice 3 avec R. Les modèles (M1)-(M12) sont spécifiés sur la feuille avec le code, mais aussi sur les sorties associées.*

S'il faut faire des tests d'hypothèse, ils sont à faire pour un seuil $\alpha = 0.05$. Ecrire les hypothèses à tester, les statistiques de test et leurs lois (si elles ont été faites en cours). Si un même type de test d'hypothèse se répète, il faut donner les détails pour un seul exemple, pour les autres cas, il faut donner seulement la conclusion.

Exercice 1.

Les données pour cet exercice proviennent du package *AER* du logiciel R, mais elles ont été traitées avec le logiciel SAS. Le fichier *PepperPrice.txt* contient les prix mensuels moyens en Europe pour le poivre noir et blanc, en dollars américains par tonne. Les mesures débutent en octobre 1973 et se finissent en avril 1996. Plus précisément, on va partager le tableau de données en deux parties:

- *données d'apprentissage:* d'octobre 1973 jusqu'en mai 1995
- *données de test:* de juin 1995 jusqu'en avril 1996

Dans ce qui suit, seulement le prix du poivre noir (noté, *poivreN*) a été modélisé par des séries chronologiques. La variable t représente le numéro de l'observation. Les modèles (M1)-(M5) sont construits sur les données d'apprentissage.

- 1) Justifiez (*test d'hypothèse*) pourquoi on utilise comme modèle pour *poivreN* une série chronologique? (modèle (M1))
- 2) Le modèle (M1) est-il stationnaire? Justification.
- 3) Comment le modèle (M2) est-il obtenu? La variable modélisée par le modèle (M2), est-elle stationnaire? (*Appuyez vos réponses, par des tests d'hypothèse et des graphiques*) Comment la variable modélisée par le modèle (M2) s'écrit fonction de la variable d'origine *poivreN*?
- 4) Pourquoi on a considéré les modèles (M3), (M4) et (M5)?
- 5) Donnez la forme du modèle statistique correspondant au modèle (M3). Donnez les estimations des paramètres de ce modèle. Est-ce que ces paramètres sont significatifs? Est-ce que le modèle (M3) est significatif? Est-ce que les résidus de ce modèle sont un bruit blanc? (*Pour toutes ces questions, justifiez vos réponse par des tests d'hypothèse, si faits en cours*)
- 6) Donnez la forme du modèle statistique correspondant au modèle (M4). Donnez les estimations des paramètres de ce modèle. Est-ce que ces paramètres sont significatifs? Est-ce que le modèle (M4) est significatif? Est-ce que les résidus de ce modèle sont un bruit blanc? (*donnez moins de détails, sachant qu'ils ont été déjà donnés à la question précédente*)
- 7) Donnez la forme du modèle statistique correspondant au modèle (M5). Donnez les estimations des paramètres de ce modèle. Est-ce que ces paramètres sont significatifs? (*donnez moins de détails, sachant qu'elles ont été déjà donnés à la question 5*).
- 8) Pourquoi entre (M3), (M4), (M5) on a choisit (M3)? Justifiez votre réponse.
- 9) On prévoit la variable *poivreN*, des données de test, par le modèle (M3). Comparez les prévisions

avec les vraies valeurs.

Exercice 2.

Les données pour cet exercice proviennent du package *survival* du logiciel R, mais elles ont été traitées avec le logiciel SAS. Trois rats ont été choisis dans chacune des 50 portées. Dans chaque portée, un rat (parmi les trois) a été traité avec un médicament. Ensuite, tous les rats ont été suivis par rapport à l'apparition d'une certaine tumeur.

Le fichier *rats.txt* contient des observations pour les variables suivantes:

litter: la portée (de 1 à 50)

rx: traitement, (1 = médicament, 0 = placebo) ;

time: le temps jusqu'à l'apparition de la tumeur ou du dernier suivi ;

status: statut de l'événement, 1 = tumeur et 0 = pas de tumeur

sex: mâle ou femelle

- 1) Quelle fonction est modélisée par la courbe associée au modèle (M6)? Commentez cette courbe.
- 2) Avec approximation, après combien de jours la probabilité qu'un rat développe une tumeur est de 0.8?
- 3) Qu'est ce qu'on modélise par le modèle (M7)? Ecrivez la forme statistique du modèle (M7). Quelles sont les variables significatives de ce modèle? (*tests d'hypothèse, détails pour une seule variable, pour les autres donnez seulement la conclusion*).
- 4) Pour le modèle (M8) donnez sa forme statistique et les estimations de tous les paramètres. Interprétez les estimations des paramètres. Comparez les modèles avec et sans les variables explicatives (*par tests d'hypothèse et indicateurs quantitatifs*).

Exercice 3.

Les données *coop* de cet exercice proviennent du package *MASS* du logiciel R, et elles ont été traitées avec le logiciel R. Plus précisément il s'agit de sept échantillons qui ont été envoyés à 6 laboratoires, en 3 lots distincts, pour mesurer la concentration d'Analyte. Chaque analyse a été dupliquée.

Les variables du tableau *coop* sont:

Lab: le laboratoire; valeurs: L1, L2, . . . , L6

Spc: le prélèvement; valeurs: S1, S2, . . . , S7.

Bat: le lot, valeurs: B1, B2, B3

Conc: Concentration de l'Analyte en g /kg

La variable *Conc* est modélisée dans cet exercice par quatre modèles: (M9), (M10), (M11), (M12).

- 1) Est-ce que la variable *Conc* suit une loi normale?
- 2) Donnez la forme statistique du modèle (M9). Le modèle (M9) est-il significatif? (*test d'hypothèse*)
- 3) Quelles variables ont une influence sur la variable *Conc*, modélisée par (M9)? (*test d'hypothèse*)
- 4) Donnez la forme statistique du modèle (M11). Il s'agit de quel type de modèle?
- 5) Quelles variables ont une influence sur la variable *Conc*, modélisée par (M11)? (*Justification*)
- 6) Pour le modèle (M12), donnez les estimations des paramètres. Interprétez ces estimations.
- 7) Les modèles (M10) et (M12) modélisent la même variable dépendante en fonction des mêmes variables explicatives. Obtient-on les mêmes estimations des paramètres, donc a-t-on finalement le même modèle? (*Justification*)
- 8) Entre les modèles (M10) et (M12) lequel est plus indiqué pour modéliser la variable *Conc*? Justification.

```

/* EXERCICE 1 */
data ex01;
  infile "PepperPrice.txt" firstobs=1;
  input t poivreN poivreB;
run;
data apprentissage;
  set ex01;
  if t<=260;
data test ;
  set ex01;
  if t>260;
run;

proc arima data=apprentissage;
  identify var=poivreN ; run; (M1)
  identify var=poivreN(1) stationarity=(dickey=6); run; (M2)
  estimate p=1; run; (M3)
  estimate q=1; run; (M4)
  estimate p=1 q=1; run; (M5)
  estimate p=1; run; (M3)
  forecast lead=11 out=prevision; run; /* le tableau "prevision contient la
prévision, avec l'intervalle de confiance */
quit;
data prev;
  set prevision;
  drop residual poivreN STD;
  if _N_>260; /* on considère les obs de test */
data compar;
  merge test prev;
  drop poivreB;
proc print data=compar;
  title "Tableau avec vraies valeurs, prévisions et intervalle de confiance";
run;

/* EXERCICE 2 */
data exo2;
  infile "rats.txt" firstobs=1;
  input litter rx time status;
run;
proc lifetest data=exo2 plot=(s) notable; } (M6)
  title " ";
  time time*status(0);
run;
proc phreg data=exo2; } (M7)
  class rx litter ;
  model time*status(0)=litter rx ;
run;
proc phreg data=exo2; } (M8)
  class rx ;
  model time*status(0)= rx ;
run;

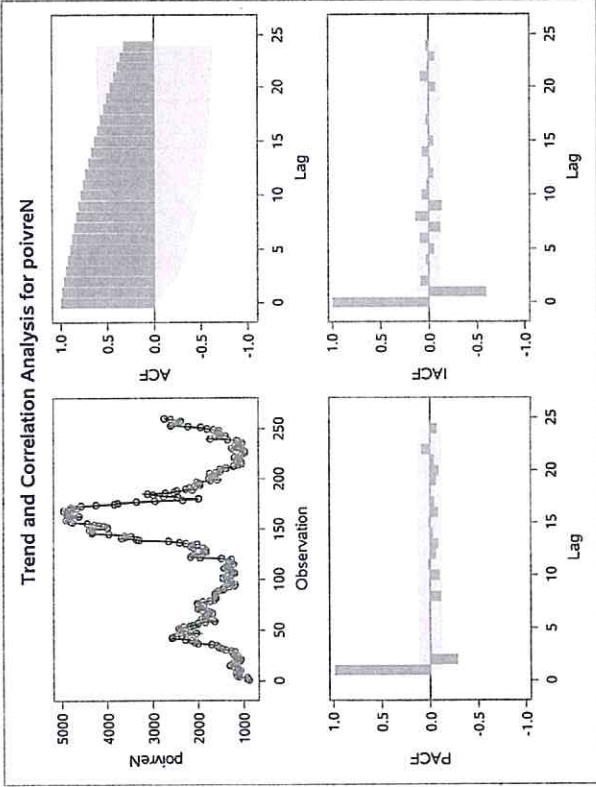
```


Exercise 1

The SAS System
The ARIMA Procedure

Name of Variable = polvreN	
Mean of Working Series	2065.103
Standard Deviation	1069.899
Number of Observations	260

Autocorrelation Check for White Noise				
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations
6	1395.73	6	<.0001	0.988 0.968 0.947 0.925 0.903 0.882
12	2440.06	12	<.0001	0.861 0.838 0.814 0.788 0.761 0.733
18	3096.20	18	<.0001	0.703 0.672 0.642 0.610 0.576 0.543
24	3400.83	24	<.0001	0.509 0.472 0.434 0.398 0.363 0.328



Name of Variable = polvreN	
Period(s) of Differencing	1
Mean of Working Series	7.173552
Standard Deviation	146.0816

1

2

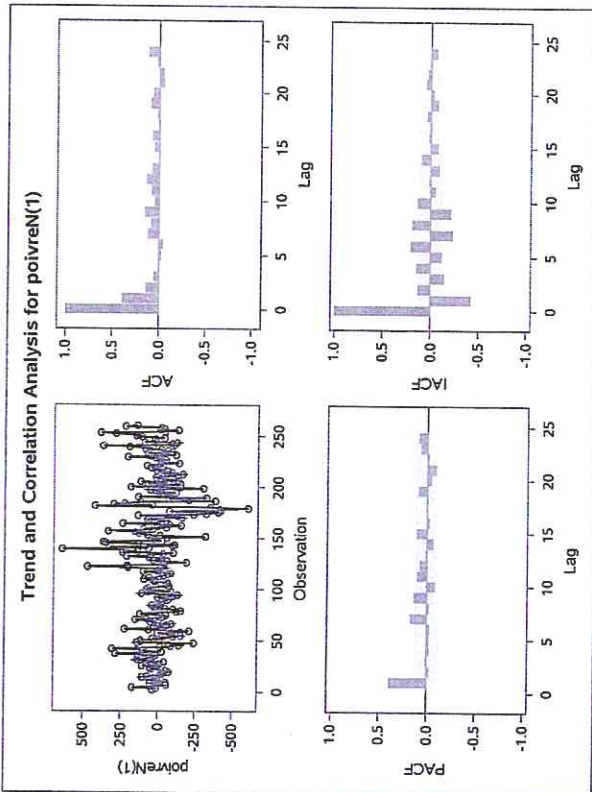
The SAS System
The ARIMA Procedure

Name of Variable = polvreN	
Number of Observations	259
Observation(s) eliminated by differencing	1

Autocorrelation Check for White Noise				
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations
6	47.89	6	<.0001	0.398 0.137 0.055 0.001 -0.027 -0.045
12	66.42	12	<.0001	0.115 0.082 0.151 0.049 0.080 0.130
18	70.43	18	<.0001	0.075 0.007 0.053 0.074 0.004 0.024
24	79.71	24	<.0001	0.093 0.068 -0.051 -0.045 0.023 0.119

Augmented Dickey-Fuller Unit Root Tests						
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	Pr > F
Zero Mean	0	-154.721	0.0001	-10.47	<.0001	
	1	-161.408	0.0001	-8.90	<.0001	
	2	-156.317	0.0001	-7.67	<.0001	
	3	-169.442	0.0001	-7.05	<.0001	
	4	-183.266	0.0001	-6.53	<.0001	
	5	-212.871	0.0001	-6.20	<.0001	
Single Mean	6	-90.9900	<.0001	-4.81	<.0001	
	0	-155.075	0.0001	-10.47	<.0001	54.79 0.0010
	1	-162.033	0.0001	-8.90	<.0001	39.61 0.0010
	2	-157.149	0.0001	-7.68	<.0001	29.46 0.0010
	3	-170.849	0.0001	-7.06	<.0001	24.90 0.0010
	4	-185.644	0.0001	-6.54	<.0001	21.42 0.0010
Trend	5	-216.965	0.0001	-6.21	<.0001	19.27 0.0010
	6	-92.2728	0.0015	-4.82	0.0001	11.61 0.0010
	0	-155.153	0.0001	-10.45	<.0001	54.60 0.0010
	1	-162.245	0.0001	-8.88	<.0001	39.48 0.0010
	2	-157.255	0.0001	-7.65	<.0001	29.34 0.0010
	3	-171.101	0.0001	-7.04	<.0001	24.80 0.0010
	4	-186.314	0.0001	-6.53	<.0001	21.34 0.0010
	5	-217.854	0.0001	-6.19	<.0001	19.20 0.0010
	6	-92.3969	0.0006	-4.80	0.0006	11.57 0.0010

The SAS System
The ARIMA Procedure



Conditional Least Squares Estimation				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
MU	7.63574	13.85916	0.55	0.5921
AR1,1	0.39894	0.05731	6.96	<.0001

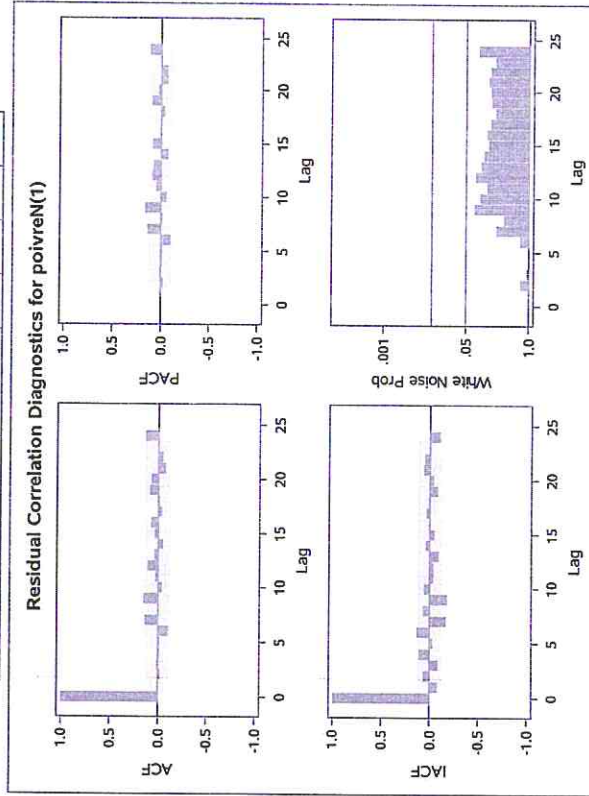
Constant Estimate	4.589519
Variance Estimate	18094.47
Std Error Estimate	134.5157
AIC	3276.073
SBC	3283.187
Number of Residuals	250

Correlations of Parameter Estimates		
Parameter	MU	AR1,1
MU	1.000	0.007
AR1,1	0.007	1.000

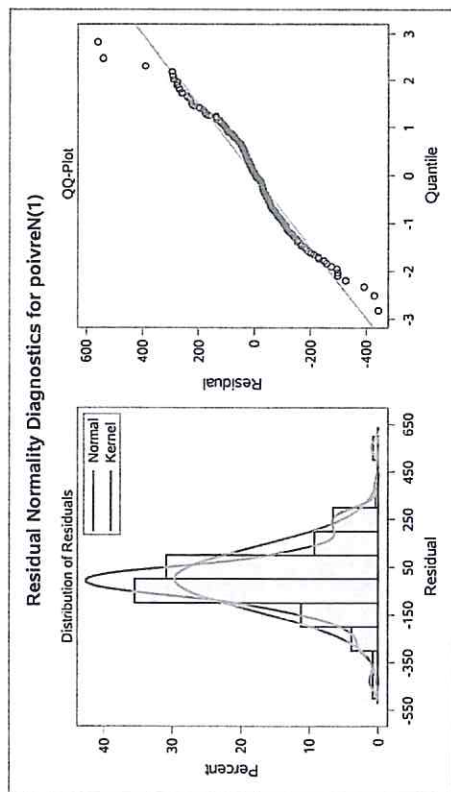
* AIC and SBC do not include log determinant.

The SAS System
The ARIMA Procedure

Autocorrelation Check of Residuals				
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations
6	3.25	5	0.6613	0.009 -0.024 0.011 -0.012 -0.016 -0.105
12	18.07	11	0.0799	0.138 -0.009 0.148 -0.041 0.026 0.107
18	21.57	17	0.2017	0.030 -0.051 0.035 0.074 -0.039 -0.015
24	32.55	23	0.0892	0.085 0.070 -0.077 -0.046 -0.004 0.135
30	42.59	29	0.0496	-0.019 -0.002 0.063 0.056 -0.048 -0.156
36	49.90	35	0.0490	0.010 -0.016 -0.027 -0.115 0.077 0.064
42	67.31	41	0.0059	0.059 -0.101 -0.044 0.015 -0.199 -0.032
48	72.88	47	0.0095	-0.079 0.082 0.022 0.009 -0.012 -0.058



The SAS System
The ARIMA Procedure



Model for variable poivreN	
Estimated Mean	7.635744
Period(s) of Differencing	1

Autoregressive Factors	
Factor 1:	1 - 0.39894 B ¹ (1)

Conditional Least Squares Estimation				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
MU	7.35694	11.55859	0.64	0.5250
MA1,1	-0.37414	0.05788	-6.46	<.0001

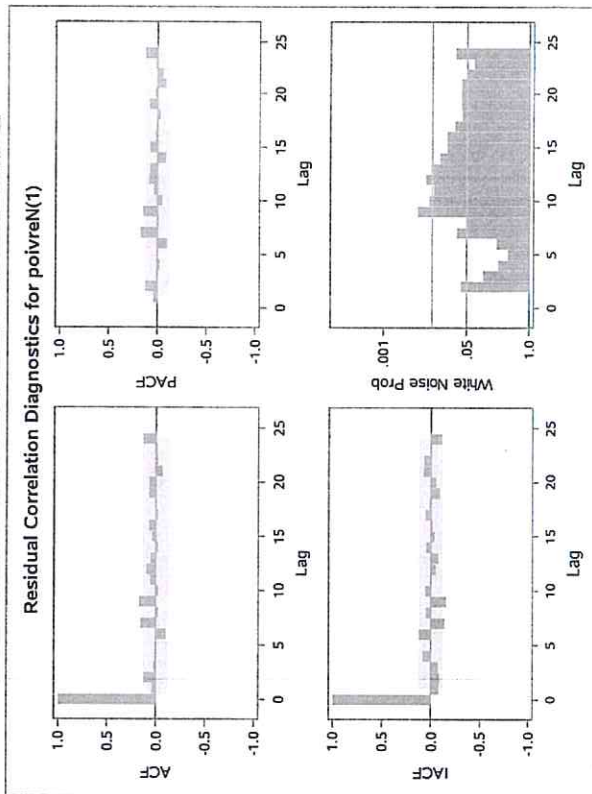
Constant Estimate	7.356939
Variance Estimate	18375.25
Std Error Estimate	135.5554
AIC	3280.061
SBC	3287.175
Number of Residuals	259

* AIC and SBC do not include log determinant.

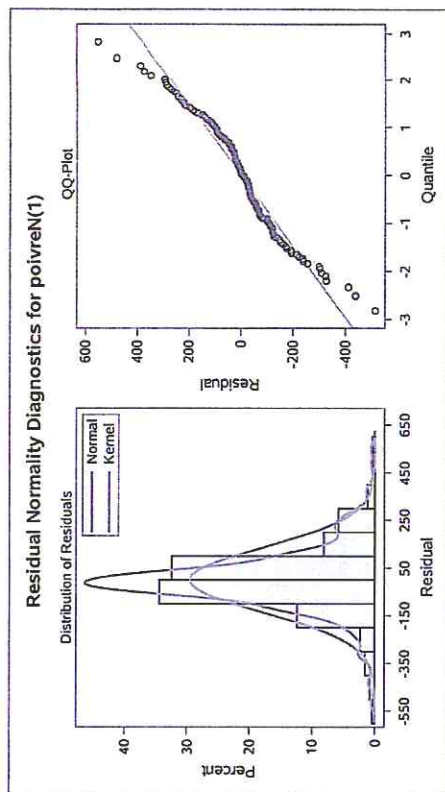
The SAS System
The ARIMA Procedure

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	
MU	1.000	-0.000	
MA1,1	-0.000	1.000	

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	7.17	5	0.2081	0.041	0.121	0.020	-0.008	0.008	-0.101
12	25.60	11	0.0074	0.158	-0.022	0.171	-0.024	0.059	0.097
18	28.43	17	0.0402	0.053	-0.023	0.039	0.070	-0.021	0.007
24	37.38	23	0.0297	0.071	0.067	-0.067	-0.019	-0.013	0.130
30	46.25	29	0.0222	-0.009	0.023	0.056	0.034	-0.042	-0.153
36	53.33	35	0.0243	-0.003	-0.056	-0.019	-0.113	0.078	0.034
42	73.25	41	0.0014	0.057	-0.093	-0.067	-0.009	-0.217	-0.028
48	80.22	47	0.0018	-0.106	0.074	0.006	0.009	-0.006	-0.072



The SAS System
The ARIMA Procedure



Model for variable poivreN	
Estimated Mean	7.356939
Period(s) of Differencing	1

Moving Average Factors	
Factor 1:	$1 + 0.37414 B^{(1)}$

Conditional Least Squares Estimation				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
MU	7.57742	13.56278	0.56	0.5769
MA1,1	-0.06328	0.15697	-0.40	0.6872
AR1,1	0.34558	0.14781	2.34	0.0202

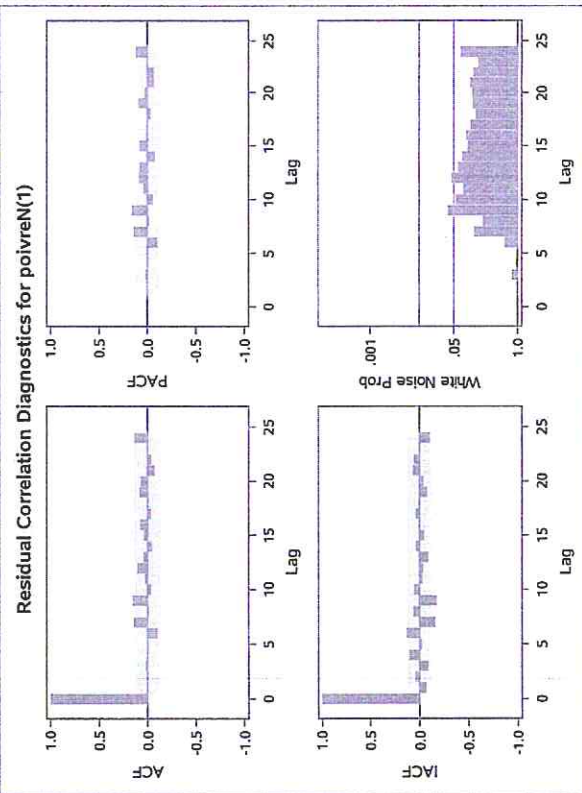
Constant Estimate	4.95881
Variance Estimate	18154.36
Std Error Estimate	134.7381
AIC	3277.919
SBC	3288.59
Number of Residuals	259

* AIC and SBC do not include log determinant.

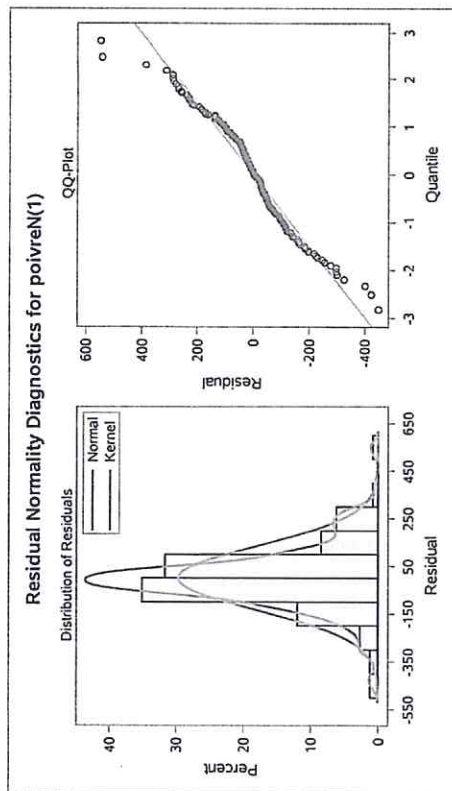
The SAS System
The ARIMA Procedure

Correlations of Parameter Estimates				
Parameter	MU	MA1,1	AR1,1	
MU	1.000	0.010	0.011	
MA1,1	0.010	1.000	0.918	
AR1,1	0.011	0.918	1.000	

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.10	4	0.5416	-0.000	-0.003	0.018	-0.011	-0.010	-0.105
12	18.58	10	0.0460	0.142	-0.015	0.152	-0.040	0.030	0.106
18	21.98	16	0.1439	0.039	-0.048	0.036	0.074	-0.037	-0.012
24	32.54	22	0.0686	0.082	0.069	-0.075	-0.041	-0.006	0.135
30	42.23	28	0.0412	-0.019	0.000	0.061	0.052	-0.046	-0.155
36	49.32	34	0.0433	0.011	-0.022	-0.024	-0.115	0.077	0.057
42	67.00	40	0.0047	0.059	-0.101	-0.046	0.015	-0.201	-0.027
48	72.57	46	0.0075	-0.084	0.080	0.018	0.009	-0.011	-0.061



The SAS System
The ARIMA Procedure



Model for variable poivreN	
Estimated Mean	7.577424
Period(s) of Differencing	
1	
Autoregressive Factors	
Factor 1:	1 - 0.34558 B**(1)
Moving Average Factors	
Factor 1:	1 + 0.06328 B**(1)

Conditional Least Squares Estimation				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
MU	7.63574	13.65916	0.55	0.5821
AR1,1	0.39894	0.05731	6.96	<.0001

Constant Estimate	4.589519
Variance Estimate	18094.47
Std Error Estimate	134.5157
AIC	3276.073
SBC	3283.187
Number of Residuals	259

* AIC and SBC do not include log determinant.

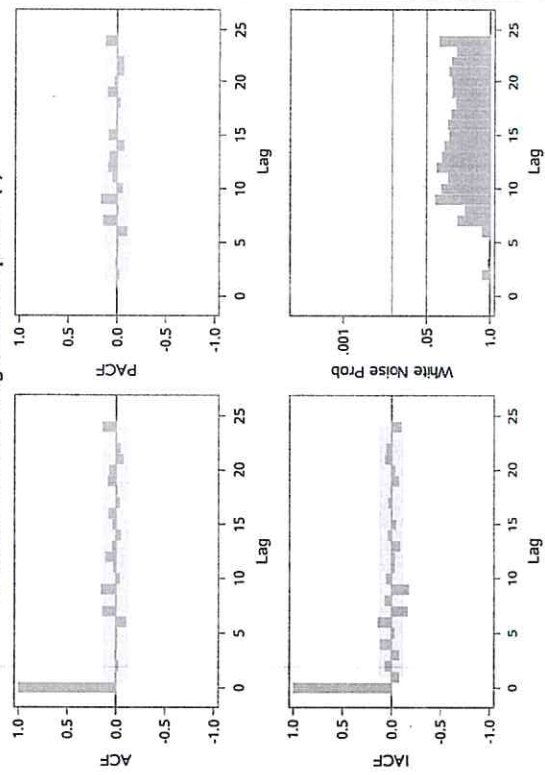
9

The SAS System
The ARIMA Procedure

Correlations of Parameter Estimates			
Parameter	MU	AR1,1	
MU	1.000	0.007	
AR1,1	0.007	1.000	

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.25	5	0.6613	0.009	-0.024	0.011	-0.012	-0.016	-0.105
12	18.07	11	0.0799	0.138	-0.009	0.148	-0.041	0.026	0.107
18	21.57	17	0.2017	0.038	-0.051	0.035	0.074	-0.039	-0.015
24	32.55	23	0.0892	0.085	0.070	-0.077	-0.046	-0.004	0.135
30	42.59	29	0.0496	-0.019	-0.002	0.063	0.056	-0.048	-0.156
36	49.90	35	0.0490	0.010	-0.016	-0.027	-0.115	0.077	0.064
42	67.31	41	0.0059	0.059	-0.101	-0.044	0.015	-0.199	-0.032
48	72.88	47	0.0095	-0.079	0.082	0.022	0.009	-0.012	-0.058

Residual Correlation Diagnostics for poivreN(1)

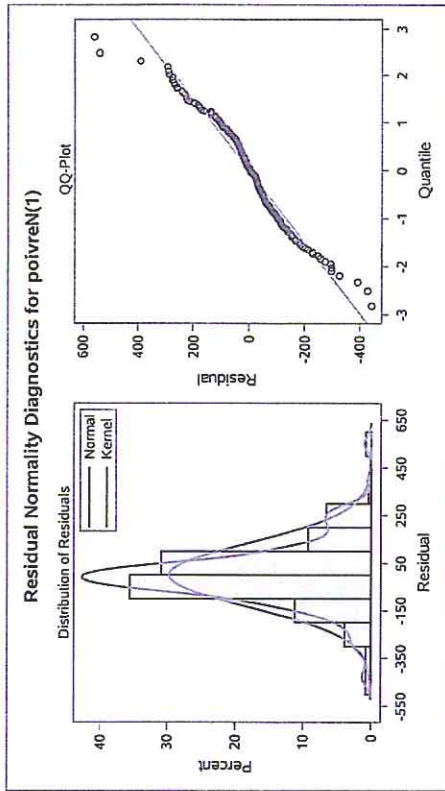


(M3)

10

(M3)

The SAS System
The ARIMA Procedure



Model for variable poivreN	
Estimated Mean	7.635744
Period(s) of Differencing	1
Autoregressive Factors	
Factor 1:	1 - 0.39894 B ¹ (1)

Forecasts for variable poivreN			
Obs	Forecast	Std Error	95% Confidence Limits
261	2803.2394	134.5157	2539.5935 3066.8853
262	2832.2599	231.3139	2376.8931 3285.6268
263	2848.4270	312.1434	2236.6371 3460.2168
264	2859.4662	380.8071	2113.0981 3605.8344
265	2868.4598	440.5595	2004.9791 3731.9405
266	2876.6372	493.7362	1908.9321 3844.3423
267	2884.4891	541.9411	1822.3040 3946.6741
268	2892.2110	586.2778	1743.1276 4041.2944
269	2899.8811	627.5204	1669.9638 4129.7985
270	2907.5306	666.2262	1601.7512 4213.3100
271	2915.1718	702.8080	1537.6935 4292.6502

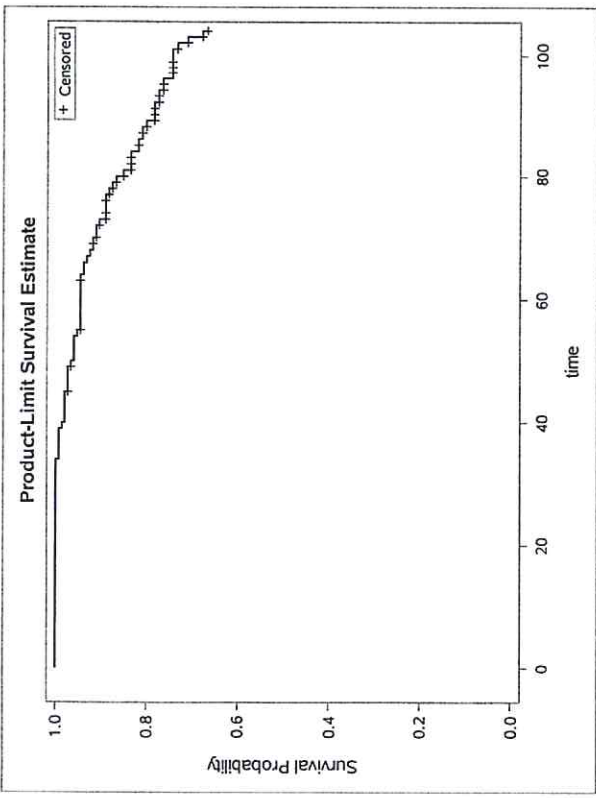
Tableau avec vraies valeurs, prévisions et intervalle de confiance

Obs	t	poivreN	FORECAST	L95	U95
1	261	2625	2803.24	2539.59	3066.89
2	262	2644	2832.26	2376.80	3285.63
3	263	2667	2848.43	2236.64	3460.22
4	264	2542	2859.47	2113.10	3605.83
5	265	2431	2868.46	2004.98	3731.94
6	266	2063	2876.64	1908.93	3844.34
7	267	2136	2884.49	1822.30	3946.67
8	268	2200	2892.21	1743.13	4041.29
9	269	2244	2899.88	1669.96	4129.80
10	270	2294	2907.53	1601.75	4213.31
11	271	2100	2915.17	1537.69	4292.65

Exercise 2

13

The LIFETEST Procedure



Summary of the Number of Censored and Uncensored Values			
Total	Failed	Censored	Percent Censored
150	40	110	73.33

14

The PHREG Procedure

Model Information			
Data Set	WORKEX02		
Dependent Variable	time		
Censoring Variable	status		
Censoring Value(s)	0		
Ties Handling	BRESLOW		

Number of Observations Read	150
Number of Observations Used	150

(117)

The PHREG Procedure

296

(M2)

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	86.6192	50	0.0010
Score	120.0226	50	<.0001
Wald	30.4661	50	0.9868

Type 3 Tests			Pr > ChiSq
Effect	DF	Wald Chi-Square	
litter	49	25.6426	0.9976
rx	1	8.8446	0.0029

Analysis of Maximum Likelihood Estimates							
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Inter	1	0.30222	1.4708	0.0455	0.8311	1.353	Inter 1
Inter	2	-16.8350	4262	0.0000	0.9968	0.000	Inter 2
Inter	3	-16.82895	4159	0.0000	0.9968	0.000	Inter 3
Inter	4	-16.58394	5612	0.0000	0.9960	0.000	Inter 4
Inter	5	-16.69069	4625	0.0000	0.9971	0.000	Inter 5
Inter	6	1.04298	1.23381	0.7146	0.3979	2.838	Inter 6
Inter	7	0.80933	1.22539	0.4362	0.5090	2.246	Inter 7
Inter	8	0.06537	1.41814	0.0022	0.9627	1.093	Inter 8
Inter	9	-16.66203	5724	0.0000	0.9977	0.000	Inter 9
Inter	10	0.94208	1.39897	0.4346	0.5087	2.565	Inter 10
Inter	11	0.51923	1.42671	0.1325	0.7159	1.681	Inter 11
Inter	12	-16.76856	6082	0.0000	0.9978	0.000	Inter 12
Inter	13	4.30360	1.36904	9.9816	0.0017	73.965	Inter 13
Inter	14	0.20901	1.41608	0.0218	0.8827	1.232	Inter 14
Inter	15	-16.69968	4619	0.0000	0.9971	0.000	Inter 15
Inter	16	-16.69228	5689	0.0000	0.9977	0.000	Inter 16
Inter	17	-16.92374	4832	0.0000	0.9972	0.000	Inter 17
Inter	18	-16.89550	4697	0.0000	0.9971	0.000	Inter 18
Inter	19	-16.67691	5374	0.0000	0.9977	0.000	Inter 19
Inter	20	1.74297	1.26278	1.9669	0.1608	5.714	Inter 20
Inter	21	-16.82895	4159	0.0000	0.9968	0.000	Inter 21
Inter	22	-16.82895	4159	0.0000	0.9968	0.000	Inter 22
Inter	23	0.23684	1.41592	0.0280	0.8672	1.267	Inter 23

47

48

The PHREG Procedure

Model Information	
Data Set	WORK.EXO2
Dependent Variable	time
Censoring Variable	status
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	150
Number of Observations Used	150

Class Level Information		
Class	Value	Design Variables
rx	0	1
	1	0

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
150	40	110	73.33

Convergence Status	
Convergence criterion (GCONV=1E-8) satisfied.	

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	371.559	363.690
AIC	371.559	365.690
SBC	371.559	367.379

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	7.8692	1	0.0050
Score	8.5546	1	0.0034
Wald	8.0087	1	0.0047

The PHREG Procedure

Type 3 Tests			
Effect	DF	Chi-Square	Pr > ChiSq
rx	1	8.0087	0.0047

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
rx	0	-0.89823	0.31740	8.0087	0.0047
				Hazard Ratio	Label
				0.407	rx 0

(M8)

(M8)

Exercice 3 (R)

```
library(MASS)
library(quantreg)
library(car)
### EXERCICE 3

data(coop)
attach(coop)
shapiro.test(Conc)
cat("Modèle par LS\n")
Lab=factor(Lab); Spc=factor(Spc); Bat=factor(Bat);
m0=lm(Conc~Lab+Spc+Bat)
summary(m0)
cat("\n ANOVA DE TYPE III\n")
print(Anova(m0, type="III"))
summary(m10)
predict_LS=predict(m10);

cat("Modèle par la méthode quantile\n")
m1=rq(Conc~Lab+Spc+Bat, tau=0.5)
summary(m1, se="iid"); #
m2=rq(Conc~Lab+Spc, tau=0.5)
summary(m2)
predict_Q=predict(m2)

cat("Comparaison des deux méthodes\n")
median(abs(Conc-predict_LS)); median(abs(Conc-predict_Q));

##### SORTIES EXERCICE 3 #####
```

Shapiro-Wilk normality test

data: Conc

W = 0.641, p-value < 2.2e-16

Modèle par LS

Call:

lm(formula = Conc ~ Lab + Spc + Bat)

Residuals:

Min	1Q	Median	3Q	Max
-1.1075	-0.1532	-0.0179	0.1389	1.9354

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.149444	0.0847624	1.763	0.0792
LabL2	0.3735714	0.0784748	4.760	3.36e-06 ***
LabL3	0.0007143	0.0784748	0.009	0.9927
LabL4	0.7802381	0.0784748	9.943	< 2e-16 ***
LabL5	0.332857	0.0784748	4.324	2.26e-05 ***
LabL6	0.4738095	0.0784748	6.038	5.96e-09 ***
SpcS2	-0.1422222	0.0847624	-1.678	0.0947
SpcS3	0.5688889	0.0847624	6.712	1.40e-10 ***
SpcS4	0.1338889	0.0847624	1.580	0.1155
SpcS5	7.2533333	0.0847624	85.572	< 2e-16 ***
SpcS6	1.2777778	0.0847624	15.075	< 2e-16 ***
SpcS7	0.8025000	0.0847624	9.468	< 2e-16 ***
BatB2	0.0879762	0.0554900	1.585	0.1142
BatB3	0.0040476	0.0554900	0.073	0.9419

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3506 on 238 degrees of freedom

Multiple R-squared: 0.98, Adjusted R-squared: 0.9789

F-statistic: 894.9 on 13 and 238 DF, p-value: < 2.2e-16

ANOVA DE TYPE III

Anova Table (Type III tests)

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	0.40	1	3.1085	0.07917
Lab	18.59	5	28.7537	< 2e-16 ***
Spc	1485.52	6	1914.4654	< 2e-16 ***
Bat	0.41	2	1.6022	0.20362
Residuals	30.78	238		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:

lm(formula = Conc ~ Lab + Spc)

Residuals:

Min	1Q	Median	3Q	Max
-1.13417	-0.15438	-0.02046	0.13970	1.99274

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.1801190	0.0786714	2.290	0.0229 *
LabL2	0.3735714	0.0786714	4.749	3.53e-06 ***
LabL3	0.0007143	0.0786714	0.009	0.9928
LabL4	0.7802381	0.0786714	9.918	< 2e-16 ***
LabL5	0.332857	0.0786714	4.313	2.36e-05 ***
LabL6	0.4738095	0.0786714	6.023	6.40e-09 ***
SpcS2	-0.1422222	0.0849749	-1.674	0.0955
SpcS3	0.5688889	0.0849749	6.695	1.52e-10 ***
SpcS4	0.1338889	0.0849749	1.576	0.1164
SpcS5	7.2533333	0.0849749	85.359	< 2e-16 ***
SpcS6	1.2777778	0.0849749	15.037	< 2e-16 ***
SpcS7	0.8025000	0.0849749	9.444	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3605 on 240 degrees of freedom

Multiple R-squared: 0.9797, Adjusted R-squared: 0.9788

F-statistic: 1052 on 11 and 240 DF, p-value: < 2.2e-16

Modèle par la méthode quantile

Call: rq(formula = Conc ~ Lab + Spc + Bat, tau = 0.5)

tau: [1] 0.5

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.25500	0.03825	6.66701	0.00000
LabL2	0.15000	0.03541	4.23599	0.00003
LabL3	0.01000	0.03541	0.28240	0.77788
LabL4	0.64500	0.03541	18.21478	0.00000
LabL5	0.20000	0.03541	5.64799	0.00000
LabL6	0.27500	0.03541	7.76599	0.00000
SpcS2	-0.18500	0.03825	-4.83685	0.00000
SpcS3	0.56500	0.03825	14.77200	0.00000
SpcS4	0.23000	0.03825	6.01338	0.00000
SpcS5	7.09500	0.03825	185.49966	0.00000
SpcS6	1.26000	0.03825	32.94286	0.00000
SpcS7	0.79500	0.03825	20.78537	0.00000
BatB2	0.04000	0.02504	1.59749	0.11148
BatB3	0.00500	0.02504	0.19969	0.84190

Call: rq(formula = Conc ~ Lab + Spc, tau = 0.5)

tau: [1] 0.5

(M9)

(M10)

(M11)

(M12)

(mm.)

Coefficients:

	coefficients	lower bd	upper bd
(Intercept)	0.23000	0.14217	0.30567
LabL2	0.19000	0.11217	0.26783
LabL3	0.03000	-0.08350	0.11350
LabL4	0.67000	0.61217	0.92134
LabL5	0.23000	0.13217	0.33567
LabL6	0.28000	0.15650	0.38783
SpCS2	-0.18000	-0.29503	-0.06497
SpCS3	0.61000	0.45486	0.68503
SpCS4	0.24000	0.13497	0.29503
SpCS5	7.12000	6.74497	7.37006
SpCS6	1.29000	1.18497	1.35503
SpCS7	0.83000	0.71497	0.96006

Comparaison des deux méthodes

[1] 0.1471825
[1] 0.11