

Examen du 29 mars 2018
Durée 2 heures

Avant de commencer à rédiger vos réponses, lisez avec attention ces consignes: Pour tous les exercices, le code SAS (ou R) et les sorties associées sont sur les feuilles suivantes. Les modèles (M1)-(M12) sont spécifiés sur la feuille avec le code SAS, 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).

Exercice 1.

Les données pour cet exercice proviennent du package R *AER*. Le fichier *ChinaIncome.txt* contient des mesures pour le revenu national réel en Chine par section, entre les années 1952 et 1988.

On veut modéliser, en utilisant le logiciel SAS, par un modèle de séries chronologiques la variable *commerce* qui représente le revenu national réel dans le secteur du commerce. La variable *t* représente le numéro de l'observation.

- 0,75 1) Justifiez (test d'hypothèse) pourquoi on utilise comme modèle pour *commerce* des séries chronologiques? (modèle (M1))
- 0,25 2) Le modèle (M1) est-il stationnaire? Justification.
- 0,75 3) Comment le modèle (M2) est-il obtenu? Est-ce que le modèle (M2) est une série chronologique (justification dans une phrase)? Est-ce que le modèle (M2) est stationnaire? (justification dans une phrase)
- 1 4) La variable modélisée par le modèle (M3), est-elle stationnaire? (Appuyez vos réponses, par des tests d'hypothèse et des graphiques) Comment la variable modélisée par le modèle (M3) s'écrit fonction de la variable d'origine *commerce*?
- 0,5 5) Pourquoi on a considéré les modèles (M4), (M5) et (M6)?
- 1,5 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? (Justifiez vos réponse par des tests d'hypothèse, si faits en cours)
- 1,25 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? Est-ce que le modèle (M5) 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)
- 0,75 8) Donnez la forme du modèle statistique correspondant au modèle (M6). 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 6).
- 0,75 9) Quel modèle il faut choisir entre (M4), (M5), (M6)? Justifiez votre réponse.

Exercice 2.

Les données pour cet exercice proviennent du package R *survival* et concernent l'un des premiers essais réussis de chimiothérapie adjuvante pour le cancer du côlon.

Le fichier *colon.txt* contient des observations pour les variables suivantes: *id*, *study*, *rx*, *sex*, *age*, *obstruct*, *perfor*, *adhere*, *nodes*, *status*, *differ*, *extent*, *surg*, *node4*, *time*, *etype*, avec (seulement pour les variables utilisées):

rx: le traitement, avec les valeurs: *Obs(ervation)*, *Lev(amisole)*, *Lev(amisole)+5-FU*

sex: 1=homme; 0=femme ;
age: l'âge ;
obstruct: obstruction du côlon par une tumeur
perfor: perforation du côlon
adhere: adhérence aux organes voisins
nodes: nombre de ganglions lymphatiques avec cancer détectable
time: jours avant l'événement ou la censure
status: statut de la censure
differ: différenciation de la tumeur (1 = bonne, 2 = modéré, 3 = médiocre)
extent: étendue de la propagation locale (1 = sous-muqueuse, 2 = muscle, 3 = séreuse, 4 = structures contiguës)
surg: temps écoulé entre l'intervention chirurgicale et l'observation (0 = court, 1 = long)
node4: plus de 4 ganglions lymphatiques positifs

Les modèles considérés ont été réalisés avec le logiciel SAS.

- 1) Quelle fonction est modélisée par la courbe associée au modèle (M7)? Commentez cette courbe.
 2) Avec approximation, après combien de jours la censure se produit avec une probabilité de 0.6?
 3) Qu'est ce qu'on modélise par le modèle (M8)? Ecrivez la forme statistique du modèle (M8) et répondez aux questions suivantes:

1. comparez les modèles avec et sans les onze variables explicatives (*par tests d'hypothèse et indicateurs quantitatifs*).
 2. quelles sont les variables significatives? (*tests d'hypothèse, détails pour une seule variable, pour les autres donnez seulement la conclusion*).

- 4) Pour le modèle (M9) donnez sa forme statistique et les estimations de tous les paramètres. Interprétez les estimations des paramètres.

Exercice 3.

Les données proviennent de l'adresse internet <http://www.stat.ucla.edu/projects/datasets/ant-explanation.html> et concernent une étude sur les habitudes alimentaires des fourmis de chaume (*Formica planipilis*). L'étude concerne différentes "stratégies" des colonies de fourmis pour optimiser l'équilibre entre la collecte de nourriture et l'exposition au risque. Les variables du fichier *thatch.txt* sont:

COLONY: identifiant de la colonie de fourmis

DISTANCE: la distance jusqu'à l'entrée de la fourmilière (en mètres)

MASS: le poids (en milligrammes) d'une colonie de fourmis

HEADWIDTH: la largeur de bande maximale d'une colonie de fourmi

SIZE: la taille d'une colonie de fourmis. Cinq valeurs sont considérées: "A" pour "< 30", "B" pour "30 - 34", "C" pour "35 - 39", "D" pour "40 - 43" et "E" pour "> 43".

La variable *MASS* est modélisée dans cet exercice par trois modèles: (M10), (M11), (M12). Le logiciel SAS a été utilisé pour les trois modèles. Le modèle (M11) a été traité aussi avec le logiciel R.

- 1) Donnez la forme statistique du modèle (M10). Il s'agit de quel type de modèle? Il a été construit sur combien d'observations?
 2) Quelles variables ont une influence sur la variable *MASS*, modélisée par (M10)?
 3) Donnez toutes les estimations des paramètres du modèle (M10). Interprétez ces estimations.
 4) Donnez la forme statistique du modèle (M11). Il s'agit de quel type de modèle?

- 0,75 5) Quelles variables ont une influence sur la variable *MASS*, modélisée par (M11)? Il y a-t-il des outliers qui ont été pris en compte par ce modèle?
- 0,75 6) Donnez la forme statistique du modèle (M12). Il s'agit de quel type de modèle?
- 0,75 7) Quelles variables ont une influence sur la variable *MASS*, modélisée par (M12)?
- 0,75 8) Les modèles (M11) 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?
- 0,75 9) Entre les modèles (M10), (M11), (M12) lequel est plus indiqué pour modéliser la variable *MASS*? Justification.

```
%%%%%%%%%%%%%% MODELE (M11) AVEC R %%%%%%%%%%%%%%%
library(quantreg)
```

```
thatch=read.table('thatch.txt',header=T,sep=",")
attach(thatch)
size=factor(Size)
m11=rq(Mass~Colony+Distance+Headwidth+Size, tau=0.5)
summary(m11)
```

```
%%%%%%%%%%%%%% LES SORTIES POUR LE MODELE (M11) AVEC R %%%%%%%%%%%%%%%
```

```
Call: rq(formula = Mass ~ Colony + Distance + Headwidth + Size, tau = 0.5)
```

```
tau: [1] 0.5
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-94.72789	9.69261	-9.77321	0.00000
Colony	0.38776	0.14008	2.76803	0.00573
Distance	-0.37415	0.12354	-3.02859	0.00251
Headwidth	4.69388	0.34915	13.44356	0.00000
SizeB	-3.77551	2.26425	-1.66744	0.09569
SizeC	-3.89796	3.69478	-1.05499	0.29164
SizeD	-1.12245	5.20967	-0.21545	0.82945
SizeE	2.46939	6.45855	0.38234	0.70227

/*****

CODE SAS *****/

/* EXERCICE 1 */

```
data exo1;
  infile "ChinaIncome.txt" firstobs=2;
  input agriculture commerce construction industry transport;
  t=_N_;
  run;
  proc arima data=exo1;
    identify var=commerce ; run; (M1)
  identify var=commerce(1) ; run; (M2)
  identify var=commerce(1,1) stationarity=(dickey=6); run; (M3)
  estimate p=1; run; (M4)
  estimate q=1; run; (M5)
  estimate p=1 q=1; run; (M6)
quit;
```

/* EXERCICE 2 */

```
data exo2;
  infile "colon.txt" firstobs=2 dlm=",";
  input id study rx$ sex age obstruct perfor adhere nodes status differ extent surg
  node4 time etype;
  run;
```

```
proc lifetest data=exo2 plot=(s) notable;
  time time*status(0);
  run; } (M7)
```

```
proc phreg data=exo2;
  class rx sex obstruct perfor adhere differ extent surg node4 ;
  model time*status(0)=rx sex age obstruct perfor adhere nodes differ extent surg
  node4 ; } (M8)
  run;
```

```
proc phreg data=exo2;
  class rx sex obstruct perfor adhere differ extent surg node4 ;
  model time*status(0)=rx obstruct adhere nodes differ extent surg node4 ; } (M9)
  run;
```

/* EXERCICE 3 */

```
data exo3;
  infile "thatch.txt" firstobs=2 dlm=",";
  input Colony Distance Mass Headwidth Headwidth1 Size$ ;
  run;
  proc univariate data=exo3 normal;
  var mass ;
  run;
  proc mixed data=exo3 covtest;
  class size;
  model Mass = Colony Distance Size / s outp=prev;
  random Headwidth / g s ;
  run ;
```

```
proc quantreg data=exo3 plots=(reshistogram);
  class size;
  model Mass = Colony Distance Headwidth Size/ quantile=.5 diagnostics;
  test Colony Distance Headwidth Size / wald;
  test Colony / wald;
  test Distance / wald;
  test Headwidth / wald;
  test Size / wald;
  run;
```

```
proc glm data=exo3 ;
  class size;
  model Mass = Colony Distance Headwidth Size/solution;
  run;
```

} (M10)

} (M11)

} (M12)

Exercise 1

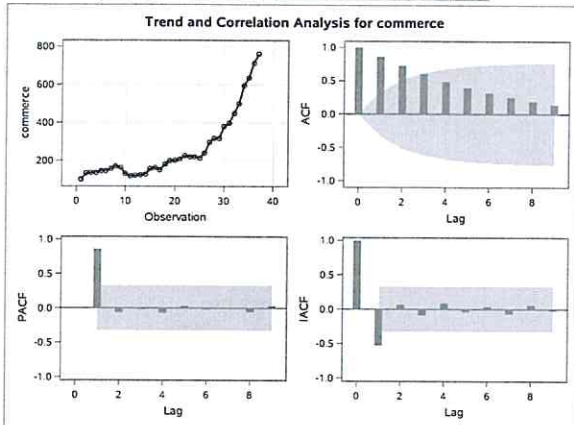
The SAS System

21:38 Tuesday, March 27, 2018

The ARIMA Procedure

Name of Variable = commerce	
Mean of Working Series	260.9865
Standard Deviation	173.8352
Number of Observations	37

Autocorrelation Check for White Noise					
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations	
6	58.47	6	<.0001	0.860	0.723 0.603 0.484 0.393 0.315



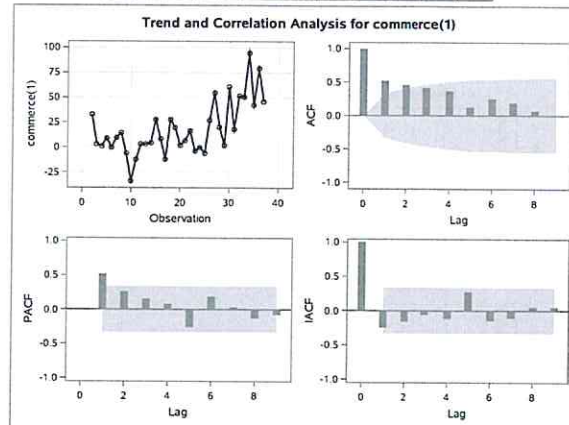
Name of Variable = commerce	
Period(s) of Differencing	1
Mean of Working Series	18.35556
Standard Deviation	26.64602
Number of Observations	36
Observation(s) eliminated by differencing	1

The SAS System

21:38 Tuesday, March 27, 2018

The ARIMA Procedure

Autocorrelation Check for White Noise							
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations			
6	35.47	6	<.0001	0.519	0.458	0.415	0.367 0.127 0.252



Name of Variable = commerce	
Period(s) of Differencing	1,1
Mean of Working Series	0.365714
Standard Deviation	25.98733
Number of Observations	35
Observation(s) eliminated by differencing	2

Autocorrelation Check for White Noise							
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations			
6	21.17	6	0.0017	-0.476	0.024	-0.061	0.263 -0.402 0.236

The SAS System

21:38 Tuesday, March 27, 2018

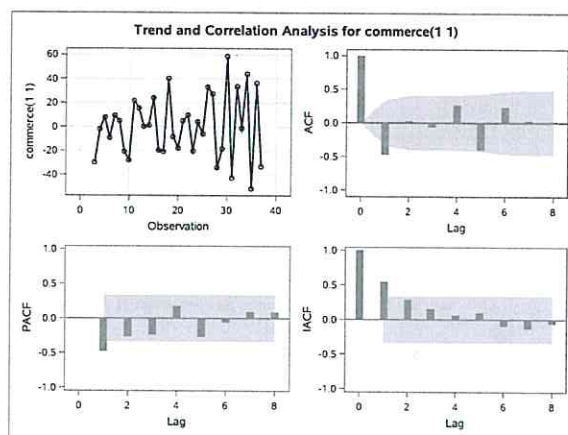
The ARIMA Procedure

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-50.9296	<.0001	-9.86	<.0001		
	1	-88.4460	<.0001	-6.51	<.0001		
	2	721.9639	0.9999	-5.51	<.0001		
	3	-69.8952	<.0001	-3.08	0.0032		
	4	59.9101	0.9999	-3.51	0.0010		
	5	648.3030	0.9999	-2.35	0.0205		
Single Mean	6	118.3467	0.9999	-2.09	0.0366		
	0	-51.0620	0.0002	-9.76	0.0002	47.63	0.0010
	1	-90.4411	0.0002	-6.51	0.0002	21.17	0.0010
	2	384.3754	0.9999	-5.63	0.0002	15.87	0.0010
	3	-118.227	0.0001	-3.26	0.0252	5.33	0.0375
	4	45.7886	0.9999	-3.74	0.0082	7.00	0.0022
Trend	5	67.1335	0.9999	-2.62	0.1005	3.43	0.2232
	6	34.0218	0.9999	-2.49	0.1278	3.14	0.2953
	0	-51.6302	<.0001	-9.67	0.0001	47.08	0.0010
	1	-95.4501	<.0001	-6.49	0.0001	21.15	0.0010
	2	185.5339	0.9999	-5.92	0.0002	17.68	0.0010
	3	-3224.81	0.0001	-3.34	0.0764	5.80	0.0915
	4	29.9588	0.9999	-4.46	0.0070	9.98	0.0010
	5	25.3007	0.9999	-3.37	0.0760	5.67	0.0972
	6	16.1260	0.9999	-3.32	0.0836	5.59	0.1039

The SAS System

21:38 Tuesday, March 27, 2018

The ARIMA Procedure



Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	0.88176	2.65006	0.33	0.7414	0
AR1,1	-0.50140	0.15566	-3.22	0.0029	1

Constant Estimate	1.323884
Variance Estimate	545.1009
Std Error Estimate	23.3474
AIC	321.8003
SBC	324.911
Number of Residuals	35

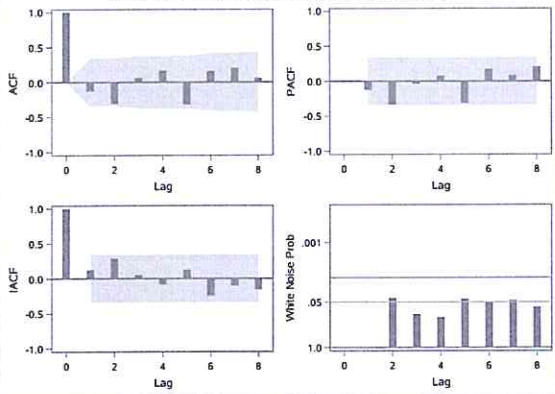
* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates		
Parameter	MU	AR1,1
MU	1.000	-0.018
AR1,1	-0.018	1.000

5

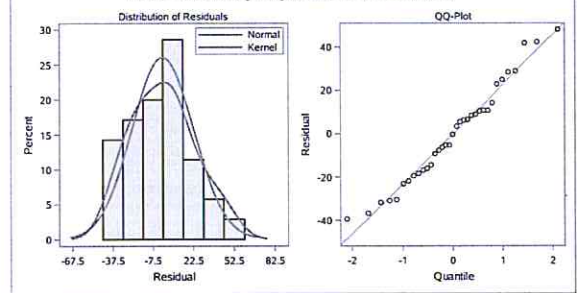
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	11.03	5	0.0508	-0.125	-0.308	0.057	0.164	-0.319	0.154
12	17.36	11	0.0976	0.203	0.057	-0.156	-0.195	0.003	0.139
18	24.20	17	0.1140	-0.090	0.085	0.135	-0.101	-0.206	0.115
24	33.18	23	0.0780	0.156	-0.045	0.029	0.039	-0.151	-0.187

Residual Correlation Diagnostics for commerce(1 1)



6

Residual Normality Diagnostics for commerce(1 1)



Model for variable commerce	
Estimated Mean	0.881764
Period(s) of Differencing	1,1
Autoregressive Factors	
Factor 1:	1 - 0.5014 B ¹ (1)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	1.09721	1.41989	0.77	0.4452	0
MA1,1	0.63720	0.13442	4.74	<.0001	1

Constant Estimate	1.097212
Variance Estimate	489.9468
Std Error Estimate	22.13474
AIC	318.0667
SBC	321.1774
Number of Residuals	35

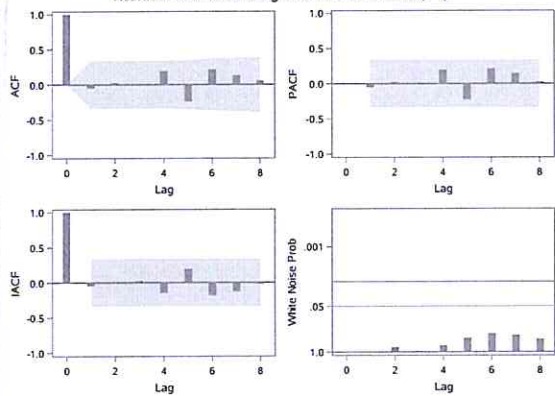
* AIC and SBC do not include log determinant.

7

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	
MU	1.000	-0.071	
MA1,1	-0.071	1.000	

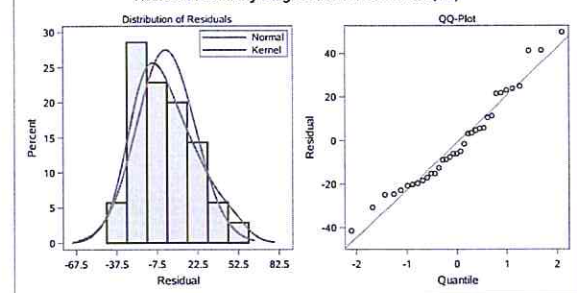
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.12	5	0.2943	-0.053	0.022	0.001	0.191	-0.235	0.217
12	10.30	11	0.5037	0.129	0.055	-0.122	-0.169	-0.068	0.119
18	13.85	17	0.6775	-0.054	0.105	0.085	-0.056	-0.137	0.093
24	23.80	23	0.4152	0.119	-0.022	0.026	-0.051	-0.167	-0.214

Residual Correlation Diagnostics for commerce(1 1)



8

Residual Normality Diagnostics for commerce(1 1)



Model for variable commerce	
Estimated Mean	1.097212
Period(s) of Differencing	1,1
Moving Average Factors	
Factor 1:	1 - 0.6372 B ¹ (1)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	1.10239	1.49814	0.74	0.4672	0
MA1,1	0.57367	0.22699	2.53	0.0166	1
AR1,1	-0.11946	0.27932	-0.43	0.6717	1

Constant Estimate	1.23409
Variance Estimate	502.0247
Std Error Estimate	22.40591
AIC	319.842
SBC	324.508
Number of Residuals	35

* AIC and SBC do not include log determinant.

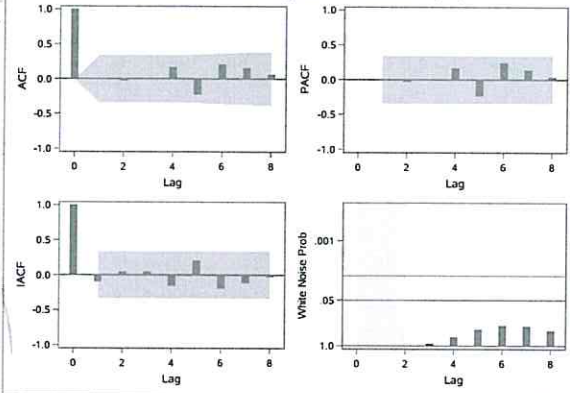
The ARIMA Procedure

Correlations of Parameter Estimates				
Parameter	MU	MA1,1	AR1,1	
MU	1.000	-0.082	-0.071	
MA1,1	-0.082	1.000	0.770	
AR1,1	-0.071	0.770	1.000	

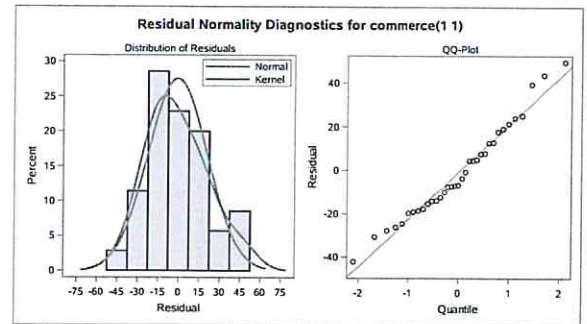
Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations							
6	5.23	4	0.2648	-0.003	-0.029	-0.009	0.162	-0.226	0.206		
12	10.61	10	0.3888	0.155	0.061	-0.145	-0.196	-0.065	0.120		
18	14.52	16	0.5602	-0.038	0.110	0.091	-0.068	-0.147	0.094		
24	25.53	22	0.2725	0.136	-0.002	0.026	-0.051	-0.181	-0.219		

Residual Correlation Diagnostics for commerce(1 1)



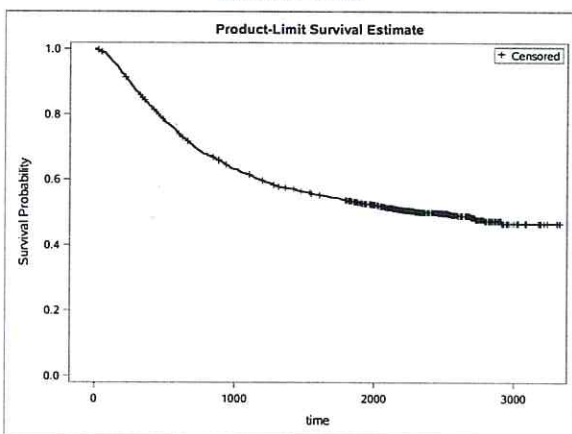
The ARIMA Procedure



Model for variable commerce	
Estimated Mean	1.102394
Period(s) of Differencing	1,1
Autoregressive Factors	
Factor 1:	1 - 0.11946 B ⁽¹⁾
Moving Average Factors	
Factor 1:	1 - 0.57367 B ⁽¹⁾

Exercise 2

The LIFETEST Procedure



Summary of the Number of Censored and Uncensored Values			
Total	Failed	Censored	Percent Censored
1858	920	938	50.48

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	1858
Number of Observations Used	1776

Class Level Information		
Class	Value	Design Variables
rx	"Lev"	1 0
	"Leve5FU"	0 1
	"Obs"	0 0
sex	0	1
	1	0
obstruct	0	1
	1	0
perfor	0	1
	1	0
adhere	0	1
	1	0
differ	1	1 0
	2	0 1
	3	0 0
extent	1	1 0 0
	2	0 1 0
	3	0 0 1
	4	0 0 0
surg	0	1
	1	0
node4	0	1
	1	0

The PHREG Procedure

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
1776	876	900	50.68

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	12500.593	12231.865
AIC	12500.593	12261.865
SBC	12500.593	12333.495

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	268.7278	15	<.0001
Score	302.2283	15	<.0001
Wald	280.3178	15	<.0001

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
rx	2	28.8616	<.0001
sex	1	0.6142	0.4332
age	1	0.6890	0.4065
obstruct	1	6.5165	0.0107
perfor	1	0.3469	0.5559
adhere	1	3.3525	0.0671
nodes	1	14.1279	0.0002
differ	2	8.6956	0.0129
extent	3	28.5416	<.0001
surg	1	10.0665	0.0015
node4	1	38.0858	<.0001

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
rx	1	-0.04841	0.07991	0.3670	0.5447	0.953	rx "Lev"
rx	1	-0.43599	0.08628	25.5369	<.0001	0.647	rx "Lev+5FU"
sex	1	0.05364	0.06844	0.6142	0.4332	1.055	sex 0
age	1	0.00242	0.00291	0.6890	0.4065	1.002	
obstruct	1	-0.21673	0.08490	6.5165	0.0107	0.805	obstruct 0
perfor	1	-0.10934	0.18564	0.3469	0.5559	0.896	perfor 0
adhere	1	-0.17093	0.09335	3.3525	0.0671	0.843	adhere 0
nodes	1	0.04027	0.01071	14.1279	0.0002	1.041	
differ	1	-0.18265	0.13821	1.7463	0.1863	0.833	differ 1
differ	2	-0.26122	0.08901	8.6131	0.0033	0.770	differ 2
extent	1	-1.27100	0.40955	9.6309	0.0019	0.281	extent 1
extent	2	-0.89291	0.19243	21.5303	<.0001	0.409	extent 2
extent	3	-0.38865	0.15279	6.4699	0.0110	0.678	extent 3
surg	1	-0.23616	0.07443	10.0665	0.0015	0.790	surg 0
node4	1	-0.61993	0.10045	38.0858	<.0001	0.538	node4 0

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	1858
Number of Observations Used	1776

Class Level Information

Class	Value	Design Variables
rx	"Lev"	1 0
	"Lev+5FU"	0 1
	"Obs"	0 0
obstruct	0	1
	1	0
adhere	0	1
	1	0
differ	1	1 0
	2	0 1
	3	0 0
extent	1	1 0 0
	2	0 1 0
	3	0 0 1
	4	0 0 0
surg	0	1
	1	0
node4	0	1
	1	0

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
1776	876	900	50.68

The PHREG Procedure

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	12500.593	12233.458
AIC	12500.593	12257.458
SBC	12500.593	12314.763

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	267.1346	12	<.0001
Score	300.2471	12	<.0001
Wald	278.4360	12	<.0001

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
rx	2	28.2838	<.0001
obstruct	1	6.7239	0.0095
adhere	1	4.2290	0.0397
nodes	1	13.8670	0.0002
differ	2	8.8021	0.0123
extent	3	28.9957	<.0001
surg	1	10.1735	0.0014
node4	1	38.6846	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
rx	1	-0.04774	0.07984	0.3576	0.5499	0.953	rx "Lev"
rx	1	-0.43104	0.08619	25.0084	<.0001	0.650	rx "Lev+5FU"
obstruct	1	-0.21697	0.08367	6.7239	0.0095	0.805	obstruct 0
adhere	1	-0.18830	0.09157	4.2290	0.0397	0.828	adhere 0
nodes	1	0.03962	0.01064	13.8670	0.0002	1.040	
differ	1	-0.17817	0.13786	1.6703	0.1962	0.837	differ 1
differ	2	-0.26235	0.08898	8.6920	0.0032	0.769	differ 2
extent	1	-1.27862	0.40912	9.7675	0.0018	0.278	extent 1

Exercise 3

The SAS System

21:38 Tuesday, March 27, 2018

17

The PHREG Procedure

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
extent	2	1	-0.90203	0.19190	22.0937	<.0001
extent	3	1	-0.39698	0.15252	6.7748	0.0092
surg	0	1	-0.23705	0.07432	10.1735	0.0014
node4	0	1	-0.61938	0.09958	38.6846	<.0001

The SAS System

21:38 Tuesday, March 27, 2018

18

The UNIVARIATE Procedure
Variable: Mass

Moments			
N	1195	Sum Weights	1195
Mean	95.23264	Sum Observations	113803
Std Deviation	28.02696	Variance	785.51032
Skewness	0.0160721	Kurtosis	-0.1665407
Uncorrected SS	11775659	Corrected SS	93789.327
Coeff Variation	29.4299919	Std Error Mean	0.81075941

Basic Statistical Measures			
Location		Variability	
Mean	95.23264	Std Deviation	28.02696
Median	96.00000	Variance	785.51032
Mode	90.00000	Range	166.00000
		Interquartile Range	38.00000

Tests for Location: Mu0=0			
Test	Statistic	p Value	
Student's t	t	Pr > t	<.0001
Sign	M	Pr >= M	<.0001
Signed Rank	S	Pr >= S	<.0001

Tests for Normality			
Test	Statistic	p Value	
Shapiro-Wilk	W	Pr < W	0.0317
Kolmogorov-Smirnov	D	Pr > D	<0.0100
Cramer-von Mises	W-Sq	Pr > W-Sq	0.0809
Anderson-Darling	A-Sq	Pr > A-Sq	0.0733

Quantiles (Definition 5)	
Level	Quantile
100% Max	189
99%	161
95%	142
90%	130
75% Q3	114
50% Median	96
25% Q1	76

The SAS System

21:38 Tuesday, March 27, 2018

19

The UNIVARIATE Procedure
Variable: Mass

Quantiles (Definition 5)	
Level	Quantile
10%	58
5%	47
1%	34
0% Min	23

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
23	613	168	782
25	697	172	795
28	1149	176	806
28	970	189	1103
28	207	189	1111

The SAS System

21:38 Tuesday, March 27, 2018

20

The Mixed Procedure

Model Information	
Data Set	WORK.EXO3
Dependent Variable	Mass
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information		
Class	Levels	Values
Size	5	A B C D E

Dimensions	
Covariance Parameters	2
Columns in X	8
Columns in Z	1
Subjects	1
Max Obs per Subject	1195

Number of Observations	
Number of Observations Read	1195
Number of Observations Used	1195
Number of Observations Not Used	0

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	9993.12840661	
1	1	9795.90362896	0.00000000

Convergence criteria met.

Estimated G Matrix		
Row	Effect	Col1
1	Headwidth	27.9583

(M10)

The Mixed Procedure

Covariance Parameter Estimates				
Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
Headwidth	27.9583	39.7180	0.70	0.2407
Residual	213.94	8.7816	24.36	<.0001

Fit Statistics	
-2 Res Log Likelihood	9795.9
AIC (Smaller is Better)	9799.9
AICC (Smaller is Better)	9799.9
BIC (Smaller is Better)	9795.9

Solution for Fixed Effects						
Effect	Size	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		-116.41	16.2069	1187	-7.18	<.0001
Colony		0.4854	0.1349	1187	3.60	0.0003
Distance		-0.6116	0.1223	1187	-5.00	<.0001
Size	A	5.4649	6.8006	1187	0.80	0.4218
Size	B	0.3939	4.9567	1187	0.08	0.9367
Size	C	-2.5772	3.0900	1187	-0.83	0.4044
Size	D	-1.9021	1.6939	1187	-1.12	0.2617
Size	E	0				

Solution for Random Effects					
Effect	Estimate	Std Err	DF	t Value	Pr > t
Headwidth	5.2756	0.3549	1187	14.86	<.0001

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Colony	1	1187	12.94	0.0003
Distance	1	1187	24.99	<.0001
Size	4	1187	3.37	0.0093

The QUANTREG Procedure

Model Information	
Data Set	WORK.EX03
Dependent Variable	Mass
Number of Independent Variables	4
Number of Continuous Independent Variables	3
Number of Class Independent Variables	1
Number of Observations	1195
Optimization Algorithm	Simplex
Method for Confidence Limits	Inv_Rank

Number of Observations Read	1195
Number of Observations Used	1195

Class Level Information		
Name	Levels	Values
Size	5	A B C D E

Parameter Information		
Parameter	Effect	Size
Intercept	Intercept	
Colony	Colony	
Distance	Distance	
Headwidth	Headwidth	
SizeA	Size	A
SizeB	Size	B
SizeC	Size	C
SizeD	Size	D
SizeE	Size	E

Summary Statistics						
Variable	Q1	Median	Q3	Mean	Standard Deviation	MAD
Colony	3.0000	6.0000	9.0000	6.0937	3.1881	4.4478
Distance	1.0000	4.0000	7.0000	4.7356	3.5286	4.4478
Headwidth	38.0000	41.0000	44.0000	40.3230	4.6332	4.4478
Mass	76.0000	96.0000	114.0	95.2326	28.0270	28.1694

The QUANTREG Procedure

Quantile Level and Objective Function	
Quantile Level	0.5
Objective Function	6432.1939
Predicted Value at Mean	97.6038

Parameter Estimates				
Parameter	DF	Estimate	95% Confidence Limits	
Intercept	1	-92.2585	-128.3550	-59.6579
Colony	1	0.3878	0.0552	0.6146
Distance	1	-0.3741	-0.6241	-0.1762
Headwidth	1	4.6939	3.9900	5.5149
Size	A	-2.4694	-14.7184	13.9382
Size	B	-6.3333	-16.2139	3.3745
Size	C	-6.3673	-13.2462	-0.2083
Size	D	-3.5918	-7.2541	0.0471
Size	E	0.0000	0.0000	0.0000

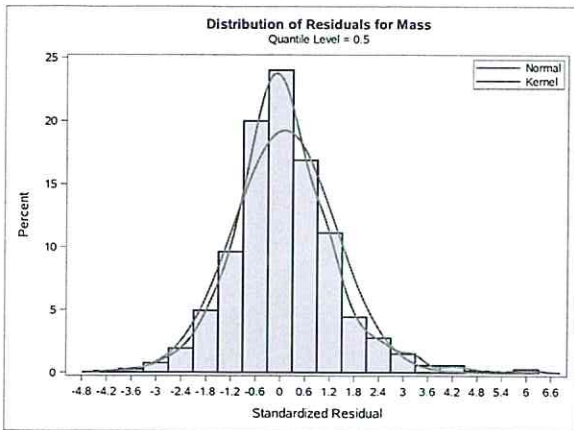
Diagnostics		
Obs	Standardized Residual	Outlier
24	3.2667	*
31	4.3434	*
37	4.1735	*
54	3.0447	*
124	-3.4453	*
170	-4.0799	*
291	3.4123	*
300	3.3066	*
334	-3.1887	*
396	3.0066	*
480	3.1540	*
492	3.2389	*
546	4.1198	*
575	-3.2823	*
584	3.2910	*
602	-3.0274	*
603	-3.5684	*

The QUANTREG Procedure

Diagnostics		
Obs	Standardized Residual	Outlier
614	3.9186	*
627	5.1913	*
674	3.0257	*
705	4.3348	*
736	3.1505	*
765	3.3308	*
773	3.3638	*
795	6.1086	*
797	3.1869	*
843	3.0066	*
883	4.6053	*
885	3.4158	*
886	6.0167	*
889	3.9776	*
985	3.3095	*
992	3.3407	*
1079	3.2748	*
1103	5.9666	*
1111	4.5891	*
1116	3.0979	*
1117	3.2684	*
1165	-3.7470	*
1172	4.3244	*

Diagnostics Summary		
Observation Type	Proportion	Cutoff
Outlier	0.0335	3.0000

25



Test 1 Results				
Test	Test Statistic	DF	Chi-Square	Pr > ChiSq
Wald	5023.0323	7	5023.03	<.0001

Test 2 Results				
Test	Test Statistic	DF	Chi-Square	Pr > ChiSq
Wald	7.6620	1	7.66	0.0056

Test 3 Results				
Test	Test Statistic	DF	Chi-Square	Pr > ChiSq
Wald	9.1723	1	9.17	0.0025

Test 4 Results				
Test	Test Statistic	DF	Chi-Square	Pr > ChiSq
Wald	180.7294	1	180.73	<.0001

26

Test 5 Results				
Test	Test Statistic	DF	Chi-Square	Pr > ChiSq
Wald	27.1533	4	27.15	<.0001

{M11}

27

Class Level Information		
Class	Levels	Values
Size	5	A B C D E

Number of Observations Read	1195
Number of Observations Used	1195

{M12}

28

Dependent Variable: Mass

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	683957.5926	97708.2275	456.72	<.0001
Error	1187	253941.7346	213.9357		
Corrected Total	1194	937899.3272			

R-Square	Coeff Var	Root MSE	Mass Mean
0.729244	15.35675	14.62654	95.23264

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Colony	1	1292.8973	1292.8973	6.04	0.0141
Distance	1	548.8440	548.8440	2.57	0.1095
Headwidth	1	679221.9606	679221.9606	3174.89	<.0001
Size	4	2893.8908	723.4727	3.38	0.0092

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Colony	1	2780.41506	2780.41506	13.00	0.0003
Distance	1	5349.09073	5349.09073	25.00	<.0001
Headwidth	1	47479.43875	47479.43875	221.93	<.0001
Size	4	2893.89080	723.47270	3.38	0.0092

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-117.5016216	16.24333569	-7.23	<.0001
Colony	0.4864042	0.13492245	3.61	0.0003
Distance	-0.6117242	0.12233668	-5.00	<.0001
Headwidth	5.2995129	0.35573357	14.90	<.0001
Size A	5.8869430	6.81370905	0.86	0.3878
Size B	0.7082185	4.96669712	0.14	0.8866
Size C	-2.3856590	3.09593767	-0.77	0.4411
Size D	-1.8140572	1.69616169	-1.07	0.2851
Size E	0.0000000			

{M12}

