

Examen du 25 mars 2016
Durée 3 heures

Avant de commencer à rédiger vos réponses, lisez avec attention ces consignes: Pour tous les exercices, le code SAS et les sorties associées sont sur les feuilles suivantes. Les modèles (M1)-(M11) 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). Les sorties SAS sont imprimées 4 pages sur une page.

Exercice 1.

Les données pour cet exercice proviennent de l'adresse internet:

<https://datamarket.com/data/list/?q=provider:tsdl>

Le fichier *milk.csv* contient les variables suivantes:

~~id:~~ le numéro de l'observation;

mois: le mois;

quantite: la production de lait de vache mensuelle (en livres par vache).

Les mesures sont mensuelles, entre janvier 1962 et décembre 1975.

On veut modéliser, par un modèle de séries chronologiques la variable *quantite*.

- 1) Donnez le nombre d'observations.
- 2) Justifier (test d'hypothèse) pourquoi on utilise comme modèle pour *quantite* des séries chronologiques? (modèle (M1))
- 3) Pourquoi on a différencié une fois la série donnée par le modèle (M1)? Le modèle obtenu est noté par (M2).
- 4) Pourquoi on a considéré que le modèle (M2) est saisonnier?
- 5) On a désaisonnalisé le modèle (M2) et on a obtenu le modèle (M3). Quel est l'ordre de saisonnalité du modèle (M2)?
- 6) La nouvelle variable obtenue, modélisée par le modèle (M3), est-elle une série chronologique? Est-elle stationnaire? (Appuyez vos réponses, par des tests d'hypothèse et des graphique)
- 7) Pourquoi on a considéré les modèles (M4) et (M5)?
- 8) 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?
- 9) 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?
- 10) Comparez les modèles (M4) et (M5). Lequel vous choisissez?

Pour les exercices 2, 3, 4, les données proviennent du site internet

<http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets>

(Department of Biostatistics, Vanderbilt University, Nashville, Tennessee, USA).

Exercice 2. Cancer de poumon)

Il s'agit d'une étude sur le temps de survie des malades de cancer de poumon. Le fichier *valung.txt* contient des observations pour les variables suivantes: *therapy*, *cell*, *t*, *etat*, *kps*, *diagtime*, *age*, *prior*, avec:

therapy: type de traitement 0: standard 1: Test;

cell: type de cellule cancéreuse: adeno, large, small, squamous ;

t: temps de survie (jusqu'à la mort);

etat: variable qui prend deux valeurs: 0 si le patient est en vie et 1 si le patient est décédé;

kps: la performance (variable numérique) ;

diagtime: le nombre de mois depuis que le cancer a été diagnostiqué;

age: l'âge du patient;

prior: variable qualitative qui prend deux valeurs pour indiquer si la patient a eu un traitement: 0 (non), 1(oui).

1) Par le modèle (M6), on modélise, par une méthode non-paramétrique, la fonction de survie d'un patient malade d'un cancer de poumon. Commentez les résultats.

2) On modélise le risque instantané que le décès survient, fonction des variables: *age*, *diagtime*, *therapy*, *cell*, *prior* par le modèle (M7).

1. Comparez les modèles avec et sans les cinq variables explicatives: tests d'hypothèse et indicateurs quantitatifs.
2. Quelles variables ont une influence sur le risque instantané du décès d'un patient? (tests d'hypothèse) Comment les variables significatives influencent ce risque?
3. Pourquoi on a considéré le modèle (M8)? Comparez les modèles (M7) et (M8) de point de vue critères AIC, BIC, log-vraisemblance. Pour le modèle (M8), quel type de cellules cancéreuses augmentent le risque que le patient décède? Faites une hiérarchie entre les quatre types de cellules concernant le risque de décès.

Exercice 3. (Dystrophie musculaire de Duchenne (DMD))

L'objectif de cette étude est le dépistage des femmes qui donneront naissance à des bébés garçons atteints de DMD, sur la base des marqueurs sériques. Une autre question intéressante est de savoir si l'âge de la mère devrait être prise en compte.

Les deux premiers marqueurs sériques, de la créatine kinase et hémopexine (*ck*, *h*), sont peu coûteux à obtenir, tandis que les deux derniers, la pyruvate kinase et lactate déshydrogénase (*pk*, *ld*), sont plus chers. Donc, il est intéressant de mesurer la contribution de *pk* et *ld* à la prédiction que le bébé soit atteint ou non de DMD.

Le fichier *dmd.txt* contient des observations pour les variables suivantes: *ind*, *hospid*, *age*, *sdate*, *ck*, *h*, *pk*, *ld*, *carrier*, *obsno*, avec:

ind: numéro de l'observation;

hospid: identifiant hôpital ;
age: l'âge de la mère;
sdate: la date quand l'étude a été réalisée;
ck: concentration de la Creatine Kinase ;
h: concentration de la hémopexine;
pk: concentration de la pyruvate kinase;
ld: concentration de la lactate déshydrogénase.
carrier: variable qualitative pour indiquer si bébé est porteur de DMD. Les valeurs de la variable *carrier* sont: 0: non porteur, 1: porteur.
obsno: var qualitative qui ne sert pas à cette étude.

On veut modéliser la probabilité qu'un bébé soit atteint de DMD, en utilisant le modèle (M9).

- 1) Quel type de modèle a-t-on utilisé? Sur combien d'observations a-t-il été considéré? Est-il significatif?
- 2) Donner les variables qui ont une influence sur la probabilité qu'un bébé soit atteint de DMD. Interprétation des estimations des paramètres. Est-ce que la contribution des marqueurs *pk* et *ld* est importante?
- 3) Quelle est la qualité du modèle (M9)?

Exercice 4. (*Pressions*)

Les données ont été mesurées sur des personnes de 8 villages de la République Dominicaine.

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

village: le village (qualitatif);
gender: le genre, avec deux valeurs: *Male* et *Female*;
age: l'âge de la personne;
SBP: la pression artérielle systolique;
DBP: Pression sanguine diastolique;

- 1) Pourquoi on a considéré un modèle quantile, noté (M10)? Quel est l'ordre de la quantile considérée? Quelle variable a été modélisée? Quelles sont les variables explicatives?
- 2) Le modèle (M10) est-il significatif? Si oui, quelles sont les variables influentes? 3) Donnez les estimations des paramètres du modèle (M10) et leur interprétation.

Exercice 5. (*growth*)

Dans cet exercice on considère un modèle de croissance, noté par (M11). Les données sont fictives. La variable modélisée est notée par *Y* et la variable explicative est notée par *X*. Le fichier *growth.txt* contient en dehors de ces deux variables, la variable *ind* qui n'est pas utilisée dans le modèle.

- 1) Donnez la forme statistique du modèle (M11). Quels sont les paramètres de ce modèle?
- 2) Sur combien d'observations le modèle a-t-il été estimé?
- 3) Donnez les estimations des paramètres du modèle.
- 4) Le modèle (M11) est-il significatif? Testez si les paramètres du modèles peuvent être égaux (un par un) à zéro.

/* EXERCICE 1 */

```

data milk;
infile "C:\Users\Gabriela.Ciuperca\M2PRO\Outils_stat\examen\an2016\milk.csv" ;
input mois $ quantite;
run;
proc arima data=milk;
    identify var=quantite ; run;      (M1)
    identify var=quantite(1) ; run;   (M2)
    identify var=quantite(1,12) stationarity=(dickey=6); run; (M3)
    estimate p=12; run;               (M4)
    estimate q=12; run;               (M5)
quit;

```

/* EXERCICE 2 */

```

data valung;
infile "C:\Users\Gabriela.Ciuperca\M2PRO\Outils_stat\examen\an2016\valung.txt" ;
input therapy$ cell$ t etat kps diagtime age prior$ ;
run;
proc lifetest data=valung plot=(s);
    time t*etat(0);
run;                                } (M6)
proc phreg data=valung;
class therapy cell prior ;
model t*etat(0)=age diagtime therapy cell prior ;
run;                                } (M7)
proc phreg data=valung;
class therapy cell prior ;
model t*etat(0)=cell;
run;                                } (M8)

```

/* EXERCICE 3 */

```

data dmd;
infile "C:\Users\Gabriela.Ciuperca\M2PRO\Outils_stat\examen\an2016\dmd.txt" ;
input ind hospid age sdate ck h pk ld carrier obsno;
run;
proc logistic data=dmd descending covout;
    model carrier=age ck h pk ld;
    output out=dmd1 p=prev predprob=(individual crossvalidate); run;
run;                                } (M9)
proc freq data=dmd1;
    table _FROM_*_INTO_; run;

```

/* EXERCICE 4 */

```

data pression;
infile "C:\Users\Gabriela.Ciuperca\M2PRO\Outils_stat\examen\an2016\dominicanHTN.txt" expandtabs ;
input Village$ Gender$ Age SBP DBP;
run;
proc univariate data=pression normal;
var SBP; run;
proc quantreg data=pression ;
    class gender;
    model SBP=age gender / quantile=.5 diagnostics ;
    test age gender/wald; /* test du modèle */
    test age / wald; /* test de la variable age */
    test gender / wald; /* test de la variable gender */
run;                                } (M10)

```

/* EXERCICE 5 */

```

data growth;
infile "C:\Users\Gabriela.Ciuperca\M2PRO\Outils_stat\examen\an2016\growth.txt" expandtabs ;
input ind X Y;
run;
proc nlin data=growth method=marquardt;
    model Y=a-exp(-b*x);
    parms a=0 b=1;
run;                                } (M11)

```

Exercice 1

Le Système SAS

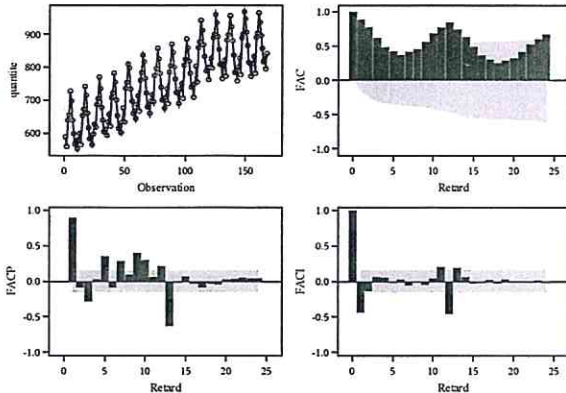
08:47 Friday, March 25, 2016 1

The ARIMA Procedure

Name of Variable = quantite	
Mean of Working Series	754.7083
Standard Deviation	101.8999
Number of Observations	168

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > Chi-2	Autocorrélations					
6	405.02	6	<.0001	0.892	0.778	0.620	0.487	0.428	0.376
12	852.41	12	<.0001	0.415	0.454	0.562	0.687	0.769	0.845
18	1129.76	18	<.0001	0.745	0.638	0.490	0.364	0.306	0.255
24	1415.29	24	<.0001	0.287	0.321	0.417	0.529	0.603	0.673

Trend and Correlation Analysis for quantite



Name of Variable = quantite	
Period(s) of Differencing	1
Mean of Working Series	1.520958
Standard Deviation	45.29647
Number of Observations	167
Observation(s) eliminated by differencing	1

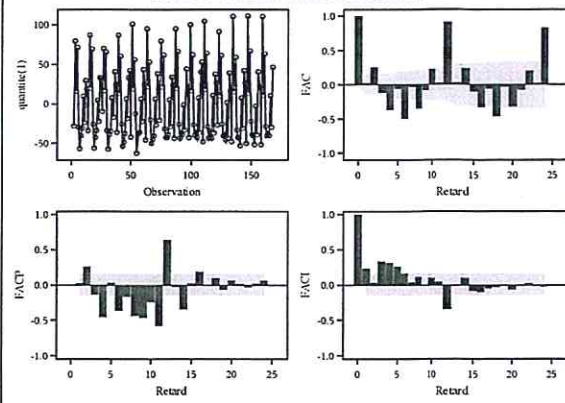
Le Système SAS

08:47 Friday, March 25, 2016 2

The ARIMA Procedure

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > Chi-2	Autocorrélations					
6	80.61	6	<.0001	0.023	0.256	-0.114	-0.359	-0.058	-0.502
12	264.26	12	<.0001	-0.033	-0.347	-0.082	0.233	0.028	0.911
18	338.56	18	<.0001	0.031	0.238	-0.103	-0.330	-0.058	-0.464
24	505.04	24	<.0001	-0.031	-0.320	-0.079	0.208	0.028	0.833

Trend and Correlation Analysis for quantite(1)



Name of Variable = quantite	
Period(s) of Differencing	1,12
Mean of Working Series	0.122581
Standard Deviation	8.911273
Number of Observations	155
Observation(s) eliminated by differencing	13

Le Système SAS

08:47 Friday, March 25, 2016 3

The ARIMA Procedure

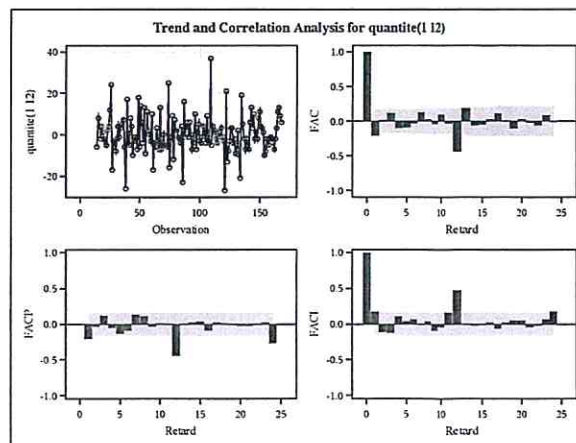
Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > Chi-2	Autocorrélations					
6	12.55	6	0.0507	-0.212	0.014	0.120	-0.100	-0.091	-0.030
12	49.78	12	<.0001	0.127	0.024	-0.043	0.095	-0.034	-0.437
18	59.43	18	<.0001	0.187	-0.066	-0.049	0.029	0.115	-0.006
24	63.97	24	<.0001	-0.112	0.022	-0.025	-0.065	0.082	-0.019

Augmented Dickey-Fuller Unit Root Tests							
Type	Retards	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-186.745	0.0001	-15.35	<.0001		
	1	-197.850	0.0001	-9.83	<.0001		
	2	-133.405	0.0001	-6.62	<.0001		
	3	-164.339	0.0001	-5.97	<.0001		
	4	-505.469	0.0001	-6.06	<.0001		
	5	862.8044	0.9999	-5.92	<.0001		
Single Mean	0	-252.564	0.0001	-4.54	<.0001		
	1	-186.774	0.0001	-15.31	<.0001	117.13	0.0010
	2	-197.923	0.0001	-9.80	<.0001	48.02	0.0010
	3	-133.423	0.0001	-6.60	<.0001	21.81	0.0010
	4	-164.201	0.0001	-5.95	<.0001	17.73	0.0010
	5	-503.744	0.0001	-6.03	<.0001	18.22	0.0010
Trend	0	-186.777	0.0001	-15.25	<.0001	116.36	0.0010
	1	-197.893	0.0001	-9.76	<.0001	47.70	0.0010
	2	-133.097	0.0001	-6.57	<.0001	21.68	0.0010
	3	-163.381	0.0001	-5.91	<.0001	17.60	0.0010
	4	-498.880	0.0001	-5.99	<.0001	18.07	0.0010
	5	867.0961	0.9999	-5.85	<.0001	17.27	0.0010
	6	-247.457	0.0001	-4.46	0.0024	10.13	0.0010

Le Système SAS

08:47 Friday, March 25, 2016 4

The ARIMA Procedure



Conditional Least Squares Estimation				
Paramètre	Valeur estimée	Erreur type	Valeur du test t	Approx. de Pr > t
MU	0.06331	0.38564	0.16	0.8658
AR1,1	-0.22968	0.07468	-3.08	0.0025
AR1,2	0.04229	0.07713	0.55	0.5844
AR1,3	0.13792	0.07818	1.76	0.0799
AR1,4	-0.03101	0.07977	-0.39	0.6950
AR1,5	-0.09563	0.07952	-1.20	0.2311
AR1,6	-0.09202	0.07957	-1.16	0.2494
AR1,7	0.08900	0.07980	1.12	0.2666
AR1,8	-0.08313	0.07987	-1.04	0.2997
AR1,9	0.04261	0.08016	0.53	0.5958
AR1,10	0.01430	0.07960	0.18	0.8577
AR1,11	-0.13974	0.07985	-1.75	0.0823
AR1,12	-0.47580	0.07712	-6.17	<.0001

The ARIMA Procedure

Constant Estimate	0.10476
Variance Estimate	60.16927
Std Error Estimate	7.756885
AIC	1087.353
SBC	1126.918
Number of Residuals	155

* AIC and SBC do not include log determinant.

The ARIMA Procedure

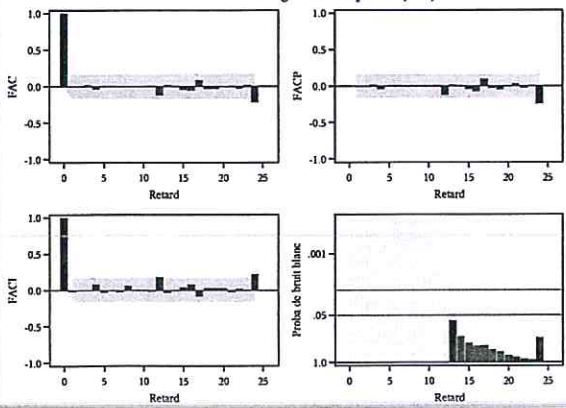
Correlations of Parameter Estimates									
Parameter	MU	AR1,1	AR1,2	AR1,3	AR1,4	AR1,5	AR1,6	AR1,7	AR1,8
MU	1.000	0.000	0.007	0.021	0.034	0.041	0.039	0.032	0.028
AR1,1	0.000	1.000	0.229	-0.015	-0.117	0.094	0.163	0.056	-0.155
AR1,2	0.007	0.229	1.000	0.232	-0.025	-0.087	0.125	0.163	0.016
AR1,3	0.021	-0.015	0.232	1.000	0.253	-0.022	-0.100	0.111	0.163
AR1,4	0.034	-0.117	-0.025	0.253	1.000	0.245	-0.047	-0.116	0.123
AR1,5	0.041	0.094	-0.087	-0.022	0.245	1.000	0.237	-0.039	-0.112
AR1,6	0.039	0.163	0.125	-0.100	-0.047	0.237	1.000	0.258	-0.038
AR1,7	0.032	0.056	0.163	0.111	-0.116	-0.039	0.258	1.000	0.242
AR1,8	0.028	-0.155	0.016	0.163	0.123	-0.112	-0.038	0.242	1.000
AR1,9	0.026	-0.111	-0.176	0.018	0.171	0.122	-0.118	-0.047	0.247
AR1,10	0.022	0.016	-0.109	-0.186	0.009	0.155	0.106	-0.107	-0.029
AR1,11	0.014	-0.024	0.002	-0.119	-0.191	0.004	0.156	0.109	-0.103
AR1,12	0.006	-0.000	-0.032	-0.004	-0.134	-0.165	0.055	0.152	0.072

Correlations of Parameter Estimates				
Parameter	AR1,9	AR1,10	AR1,11	AR1,12
MU	0.026	0.022	0.014	0.006
AR1,1	-0.111	0.016	-0.024	-0.000
AR1,2	-0.176	-0.109	0.002	-0.032
AR1,3	0.018	-0.186	-0.119	-0.004
AR1,4	0.171	0.009	-0.191	-0.134
AR1,5	0.122	0.155	0.004	-0.165
AR1,6	-0.118	0.106	0.156	0.055
AR1,7	-0.047	-0.107	0.109	0.152
AR1,8	0.247	-0.029	-0.103	0.072
AR1,9	1.000	0.256	-0.027	-0.126
AR1,10	0.256	1.000	0.261	0.011
AR1,11	-0.027	0.261	1.000	0.273
AR1,12	-0.126	0.011	0.273	1.000

The ARIMA Procedure

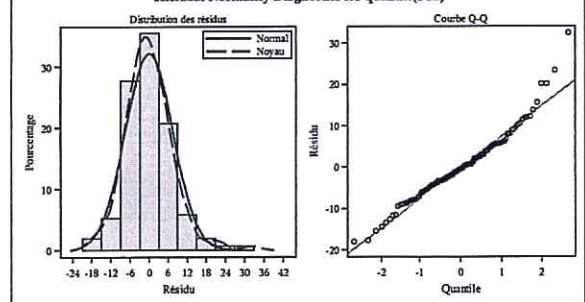
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > Chi2	Autocorrélations					
6	.	0	.	-0.012	0.004	0.019	-0.043	0.016	0.012
12	.	0	.	-0.011	-0.003	0.003	0.004	0.010	-0.125
18	5.93	6	0.4311	0.023	0.011	-0.051	-0.063	0.085	-0.036
24	15.70	12	0.2054	-0.039	0.009	0.012	-0.024	0.020	-0.223
30	19.74	18	0.3478	0.084	-0.075	-0.045	0.069	-0.035	-0.026

Residual Correlation Diagnostics for quantile(1 12)



The ARIMA Procedure

Residual Normality Diagnostics for quantile(1 12)



Model for variable quantile	
Estimated Mean	0.063313
Period(s) of Differencing	1,12

Autoregressive Factors

$r = 1 + 0.22968 B^{**}(1) - 0.04229 B^{**}(2) - 0.13792 B^{**}(3) + 0.03101 B^{**}(4) + 0.09563 B^{**}(5) + 0.09202 B^{**}(6) - 0.089 B^{**}(7) - 0.08313 B^{**}(8) - 0.04261 B^{**}(9) - 0.0143 B^{**}(10) - 0.13974 B^{**}(11) + 0.4758 B^{**}(12)$

: The model defined by the new estimates is unstable. The iteration process has been terminated.

: Estimates may not have converged.

ARIMA Estimation Optimization Summary	
Estimation Method	Conditional Least Squares
Parameters Estimated	13
Termination Criteria	Maximum Relative Change in Estimates
Iteration Stopping Value	0.001
Criteria Value	17.97002
Maximum Absolute Value of Gradient	4955.967
R-Square Change from Last Iteration	0.342517
Objective Function	Sum of Squared Residuals
Objective Function Value	8734.301
Marquardt's Lambda Coefficient	1E-6
Numerical Derivative Perturbation Delta	0.001

The ARIMA Procedure

ARIMA Estimation Optimization Summary	
Iterations	4
Warning Message	Estimates may not have converged.

Conditional Least Squares Estimation					
Paramètre	Valeur estimée	Erreur type	Valeur du test t	Approx. de Pr > t	Retard
MU	-0.08995	0.52193	-0.17	0.8634	0
MA1,1	0.20206	0.08233	2.45	0.0153	1
MA1,2	-0.0095908	0.08328	-0.12	0.9085	2
MA1,3	-0.05675	0.08372	-0.68	0.4990	3
MA1,4	0.10603	0.08404	1.26	0.2091	4
MA1,5	-0.0016823	0.08368	-0.02	0.9840	5
MA1,6	0.12137	0.08333	1.46	0.1475	6
MA1,7	-0.07539	0.08319	-0.91	0.3664	7
MA1,8	0.02557	0.08392	0.30	0.7610	8
MA1,9	0.01878	0.08388	0.22	0.8231	9
MA1,10	0.02638	0.08377	0.31	0.7533	10
MA1,11	0.12284	0.08418	1.46	0.1467	11
MA1,12	0.52037	0.08213	6.34	<.0001	12

Constant Estimate	-0.08995
Variance Estimate	61.50916
Std Error Estimate	7.842778
AIC	1090.767
SBC	1130.332
Number of Residuals	155

* AIC and SBC do not include log determinant.

à prendre en compte pour la question b) de l'exercice

The ARIMA Procedure

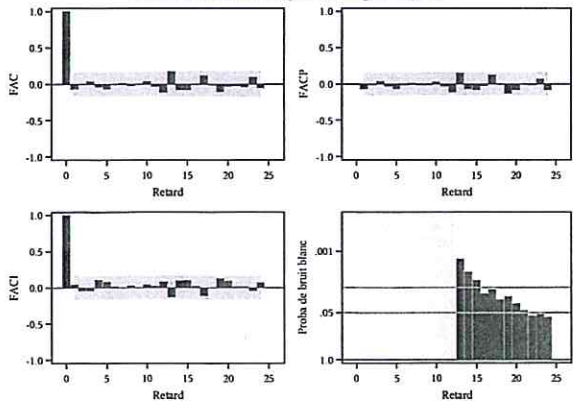
Correlations of Parameter Estimates									
Parameter	MU	MA1,1	MA1,2	MA1,3	MA1,4	MA1,5	MA1,6	MA1,7	MA1,8
MU	1.000	0.354	0.303	0.314	0.323	0.293	0.281	0.265	0.283
MA1,1	0.354	1.000	-0.245	0.130	0.125	0.008	0.102	-0.075	0.152
MA1,2	0.303	-0.245	1.000	-0.266	0.107	0.129	0.008	0.101	-0.096
MA1,3	0.314	0.130	-0.266	1.000	-0.259	0.100	0.127	0.000	0.108
MA1,4	0.323	0.125	0.107	-0.259	1.000	-0.269	0.102	0.118	0.013
MA1,5	0.293	0.008	0.129	0.100	-0.269	1.000	-0.284	0.103	0.108
MA1,6	0.281	0.102	0.008	0.127	0.102	-0.284	1.000	-0.256	0.107
MA1,7	0.265	-0.075	0.101	0.000	0.118	0.103	-0.256	1.000	-0.292
MA1,8	0.283	0.152	-0.096	0.108	0.013	0.108	0.107	-0.292	1.000
MA1,9	0.292	0.070	0.127	-0.090	0.110	0.008	0.111	0.097	-0.283
MA1,10	0.281	0.073	0.068	0.124	-0.089	0.100	0.003	0.110	0.101
MA1,11	0.266	0.077	0.067	0.073	0.116	-0.094	0.084	0.004	0.102
MA1,12	0.265	-0.202	0.067	0.057	0.061	0.130	-0.082	0.074	-0.007

Correlations of Parameter Estimates				
Parameter	MA1,9	MA1,10	MA1,11	MA1,12
MU	0.292	0.281	0.266	0.285
MA1,1	0.070	0.073	0.077	-0.202
MA1,2	0.127	0.068	0.067	0.067
MA1,3	-0.090	0.124	0.073	0.057
MA1,4	0.110	-0.089	0.116	0.061
MA1,5	0.008	0.100	-0.094	0.130
MA1,6	0.111	0.003	0.084	-0.082
MA1,7	0.097	0.110	0.004	0.074
MA1,8	-0.283	0.101	0.102	-0.007
MA1,9	1.000	-0.283	0.104	0.083
MA1,10	-0.283	1.000	-0.296	0.114
MA1,11	0.104	-0.296	1.000	-0.239
MA1,12	0.083	0.114	-0.239	1.000

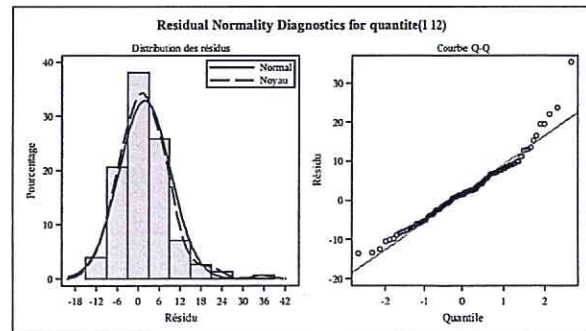
The ARIMA Procedure

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > Chi-2	Autocorrélations					
6	.	0	.	0.001	0.048	0.094	0.017	-0.007	0.046
12	.	0	.	0.069	0.030	0.050	0.092	0.024	-0.051
18	19.43	6	0.0035	0.212	-0.026	-0.021	0.066	0.169	0.037
24	23.74	12	0.0220	-0.046	0.021	0.028	0.011	0.141	-0.008
30	26.65	18	0.0857	0.102	0.022	-0.007	0.047	0.018	0.043

Residual Correlation Diagnostics for quantile(1 12)



The ARIMA Procedure



Model for variable quantile	
Estimated Mean	-0.08995
Period(s) of Differencing	1,12

Moving Average Factors	
1 - 0.20206 B**(1) + 0.00959 B**(2) + 0.05675 B**(3) - 0.10603 B**(4) + 0.00168 B**(5) - 0.12137 B**(6) + 0.07539 B**(7) - 0.02557 B**(8) - 0.01878 B**(9) - 0.02638 B**(10) - 0.12284 B**(11) - 0.52037 B**(12)	

Procédure LIFETEST

Exercice 2

Le Système SAS

08:47 Friday, March 25, 2016 13

Procédure LIFETEST

Product-Limit Survival Estimates					
t	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.000	1.0000	0	0	0	137
1.000				1	136
1.000	0.9854	0.0146	0.0102	2	135
2.000	0.9781	0.0219	0.0125	3	134
3.000	0.9708	0.0292	0.0144	4	133
4.000	0.9635	0.0365	0.0160	5	132
7.000				6	131
7.000				7	130
7.000	0.9416	0.0584	0.0200	8	129
8.000				9	128
8.000				10	127
8.000				11	126
8.000	0.9124	0.0876	0.0242	12	125
10.000				13	124
10.000	0.8978	0.1022	0.0259	14	123
11.000	0.8905	0.1095	0.0267	15	122
12.000				16	121
12.000	0.8759	0.1241	0.0282	17	120
13.000				18	119
13.000	0.8613	0.1387	0.0295	19	118
15.000				20	117
15.000	0.8467	0.1533	0.0308	21	116
16.000	0.8394	0.1606	0.0314	22	115
18.000				23	114
18.000				24	113
18.000	0.8175	0.1825	0.0330	25	112
19.000				26	111
19.000	0.8029	0.1971	0.0340	27	110
20.000				28	109
20.000	0.7883	0.2117	0.0349	29	108
21.000				30	107
21.000	0.7737	0.2263	0.0357	31	106
22.000	0.7664	0.2336	0.0361	32	105

Le Système SAS

08:47 Friday, March 25, 2016 15

Procédure LIFETEST

Product-Limit Survival Estimates					
t	Survival	Failure	Survival Standard Error	Number Failed	Number Left
63.000	0.5235	0.4765	0.0428	65	71
72.000	0.5161	0.4839	0.0428	66	70
73.000	0.5087	0.4913	0.0428	67	69
80.000				68	68
80.000	0.4940	0.5060	0.0429	69	67
82.000	0.4866	0.5134	0.0428	70	66
83.000	*			70	65
84.000	0.4791	0.5209	0.0428	71	64
87.000	0.4716	0.5284	0.0428	72	63
87.000	*			72	62
90.000	0.4640	0.5360	0.0428	73	61
92.000	0.4564	0.5436	0.0428	74	60
95.000				75	59
95.000	0.4412	0.5588	0.0427	76	58
97.000	*			76	57
99.000				77	56
99.000	0.4257	0.5743	0.0426	78	55
100.000	0.4180	0.5820	0.0425	79	54
100.000	*			79	53
103.000	0.4101	0.5899	0.0424	80	52
103.000	*			80	51
105.000	0.4021	0.5979	0.0423	81	50
110.000	0.3940	0.6060	0.0422	82	49
111.000				83	48
111.000	0.3779	0.6221	0.0420	84	47
112.000	0.3699	0.6301	0.0419	85	46
117.000				86	45
117.000	0.3538	0.6462	0.0416	87	44
118.000	0.3458	0.6542	0.0414	88	43
122.000	0.3377	0.6623	0.0412	89	42
123.000	*			89	41
126.000	0.3295	0.6705	0.0410	90	40
132.000	0.3213	0.6787	0.0408	91	39

Le Système SAS

08:47 Friday, March 25, 2016 14

Procédure LIFETEST

Product-Limit Survival Estimates					
t	Survival	Failure	Survival Standard Error	Number Failed	Number Left
24.000				33	104
24.000	0.7518	0.2482	0.0369	34	103
25.000				35	102
25.000				36	101
25.000	0.7299	0.2701	0.0379	37	100
25.000	*			37	99
27.000	0.7226	0.2774	0.0383	38	98
29.000	0.7152	0.2848	0.0386	39	97
30.000				40	96
30.000	0.7004	0.2996	0.0392	41	95
31.000				42	94
31.000	0.6857	0.3143	0.0397	43	93
33.000	0.6783	0.3217	0.0400	44	92
35.000	0.6709	0.3291	0.0402	45	91
36.000	0.6636	0.3364	0.0404	46	90
42.000	0.6562	0.3438	0.0406	47	89
43.000	0.6488	0.3512	0.0408	48	88
44.000	0.6415	0.3585	0.0410	49	87
45.000	0.6341	0.3659	0.0412	50	86
48.000	0.6267	0.3733	0.0414	51	85
49.000	0.6193	0.3807	0.0416	52	84
51.000				53	83
51.000				54	82
51.000	0.5972	0.4028	0.0420	55	81
52.000				56	80
52.000				57	79
52.000	0.5751	0.4249	0.0423	58	78
53.000	0.5677	0.4323	0.0424	59	77
54.000				60	76
54.000	0.5530	0.4470	0.0426	61	75
56.000	0.5456	0.4544	0.0427	62	74
59.000	0.5382	0.4618	0.0427	63	73
61.000	0.5309	0.4691	0.0428	64	72

Procédure LIFETEST

Product-Limit Survival Estimates					
t	Survival	Failure	Survival Standard Error	Number Failed	Number Left
411.000	0.0450	0.9550	0.0194	123	5
467.000	0.0360	0.9640	0.0175	124	4
553.000	0.0270	0.9730	0.0153	125	3
587.000	0.0180	0.9820	0.0126	126	2
991.000	0.00900	0.9910	0.00894	127	1
999.000	0	1.0000		128	0

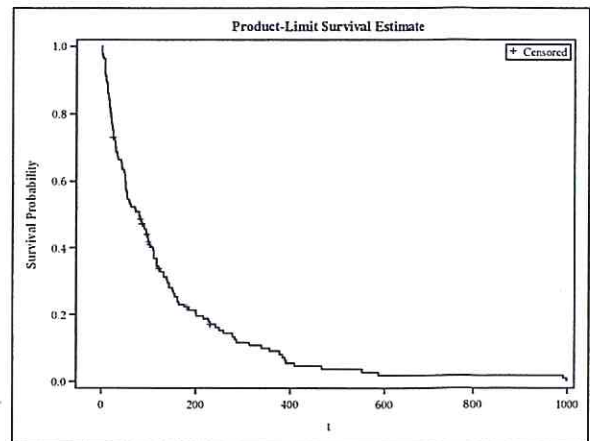
Note: The marked survival times are censored observations.

Statistiques descriptives pour variable temps t

Quartile Estimates				
Percent	Point Estimate	95% Confidence Interval		
		Transform	(Lower)	(Upper)
75	162.000	LOGLOG	132.000	231.000
50	80.000	LOGLOG	52.000	100.000
25	25.000	LOGLOG	18.000	33.000

Mean	Standard Error
132.777	15.368

Procédure LIFETEST



Summary of the Number of Censored and Uncensored Values			
Total	Failed	Censored	Percent Censored
137	128	9	6.57

The PHREG Procedure

Model Information	
Data Set	WORK.VALUNG
Dependent Variable	t
Censoring Variable	etat
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	137
Number of Observations Used	137

The PHREG Procedure

Class Level Information				
Class	Value	Design Variables		
therapy	standard	1		
	test	0		
cell	Adeno	1	0	0
	Large	0	1	0
	Small	0	0	1
	Squamous	0	0	0
prior	no	1		
	yes	0		

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
137	128	9	6.57

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

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	1011.768	984.975
AIC	1011.768	998.975
SBC	1011.768	1018.939

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > Chi-2
Likelihood Ratio	26.7930	7	0.0004
Score	27.7147	7	0.0002
Wald	26.1883	7	0.0005

The PHREG Procedure

Type 3 Tests				
Effect	DF	Wald Chi-Square	Pr > Chi-2	
age	1	0.2580	0.6115	
diagtime	1	1.1612	0.2812	
therapy	1	0.6647	0.4149	
cell	3	22.8236	<.0001	
prior	1	0.0837	0.7724	

Analysis of Maximum Likelihood Estimates							
Parameter	DDL	Parameter Estimate	Standard Error	Chi-Square	Pr > Chi-2	Hazard Ratio	Label
age	1	0.00490	0.00965	0.2580	0.6115	1.005	
diagtime	1	0.01002	0.00930	1.1612	0.2812	1.010	
therapy	standard	-0.16440	0.20164	0.6647	0.4149	0.848	therapy standard
cell	Adeno	1.18212	0.29903	15.6272	<.0001	3.261	cell Adeno
	Large	0.28892	0.28532	1.0254	0.3112	1.335	cell Large
	Small	1.03434	0.27776	13.8672	0.0002	2.813	cell Small
prior	no	0.06701	0.23166	0.0837	0.7724	1.069	prior no

The PHREG Procedure

Model Information	
Data Set	WORK.VALUNG
Dependent Variable	t
Censoring Variable	etat
Censoring Value(s)	0
Ties Handling	BRESLOW

Number of Observations Read	137
Number of Observations Used	137

Class Level Information				
Class	Value	Design Variables		
cell	Adeno	1	0	0
	Large	0	1	0
	Small	0	0	1
	Squamous	0	0	0

The PHREG Procedure

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
137	128	9	6.57

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

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	1011.768	987.197
AIC	1011.768	993.197
SBC	1011.768	1001.753

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > Chi-2
Likelihood Ratio	24.5708	3	<.0001
Score	25.2041	3	<.0001
Wald	23.8213	3	<.0001

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > Chi-2
cell	3	23.8213	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter	DDL	Parameter Estimate	Standard Error	Chi-Square	Pr > Chi-2	Hazard Ratio
cell	Adeno	1	1.14137	0.29285	15.1904	<.0001
cell	Large	1	0.23018	0.27731	0.6890	0.4065
cell	Small	1	0.99643	0.25355	15.4442	<.0001

Exercice 3

Procédure LOGISTIC

Model Information	
Table	WORK.DMD
Variable de réponse	carrier
Nombre de niveaux de réponse	2
Modèle	logit binaire
Technique d'optimisation	Score de Fisher

Nombre d'observations lues	209
Nombre d'observations utili	194

Response Profile		
Ordered Value	carrier	Total Frequency
1	1	67
2	0	127

La probabilité modélisée est carrier=1.

Note: 15 observations were deleted due to missing values for the response or explanatory variables.

Model Convergence Status	
Critère de convergence (GCONV=1E-8) respecté	

Model Fit Statistics		
Criterion	Intercept Only	Constante et Covariables
AIC	252.077	96.038
SC	255.344	115.646
-2 Log	250.077	84.038

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DDL	Pr > Chi-2
Rapport de vrais	166.0383	5	<.0001
Score	108.0405	5	<.0001
Wald	33.4505	5	<.0001

Procédure LOGISTIC

Analysis of Maximum Likelihood Estimates					
Parameter	DDL	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-2
Intercept	1	-19.5489	3.3059	34.9666	<.0001
age	1	0.1478	0.0447	10.9066	0.0010
ck	1	0.0467	0.0148	9.8952	0.0017
h	1	0.0826	0.0272	9.1923	0.0024
pk	1	0.1163	0.0482	5.8261	0.0158
ld	1	0.0125	0.00574	4.7412	0.0294

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
age	1.159	1.062	1.266
ck	1.048	1.018	1.079
h	1.086	1.030	1.146
pk	1.123	1.022	1.235
ld	1.013	1.001	1.024

Association of Predicted Probabilities and Observed Responses			
Pourcentage concordant	96.2	D de Somers	0.925
Pourcentage discordant	3.7	Gamma	0.926
Pourcentage lié	0.1	Tau-a	0.420
Paires	8509	c	0.962

The FREQ Procedure

Table de FROM par INTO			
FROM (Valeur formée de la réponse observée)	INTO (Valeur formée de la réponse prédite)		
	0	1	Total
0	122 62.89 96.06 90.37	5 2.58 3.94 8.47	127 65.46
1	13 6.70 19.40 9.63	54 27.84 80.60 91.53	67 34.54
Total	135 69.59	59 30.41	194 100.00

Fréquence manquante = 15

The UNIVARIATE Procedure
Variable: SBP

Moments		
N	381	Sum Weights
Mean	132.994751	Sum Observations
Std Deviation	25.7157534	Variance
Skewness	0.79561281	Kurtosis
Uncorrected SS	6990271	Corrected SS
Coeff Variation	19.3359161	Std Error Mean

Basic Statistical Measures			
Emplacement	Variabilité		
Mean	132.9948	Std Deviation	25.71575
Median	130.0000	Variance	661.29997
Mode	120.0000	Range	156.00000
		Interquartile Range	32.00000

Tests for Location: Mu0=0			
Test	Statistique	P-value	
Student's t	t	100.948	Pr > t <.0001
Sign	M	190.5	Pr >= M <.0001
Signed Rank	S	36385.5	Pr >= S <.0001

Exercice 4

The UNIVARIATE Procedure
Variable: SBP

Tests de normalité			
Test	Statistique	Pr < W	P-value
Shapiro-Wilk	W	0.95608	Pr < W < 0.0001
Kolmogorov-Smirnov	D	0.118533	Pr > D < 0.0100
Cramer-von Mises	W-Sq	0.887516	Pr > W-Sq < 0.0050
Anderson-Darling	A-Sq	5.083379	Pr > A-Sq < 0.0050

Quantiles (Definition 5)	
Quantile	Valeur estimée
100% Max	236
99%	204
95%	180
90%	170
75% Q3	150
50% Median	130
25% Q1	118
10%	105
5%	98
1%	90
0% Min	80

Observations extrêmes			
Le plus bas		Le plus haut	
Valeur	Obs	Valeur	Obs
80	220	200	111
80	202	204	174
90	345	210	248
90	338	210	259
90	321	236	226

The QUANTREG Procedure

Model Information	
Data Set	WORK.FRESSION
Dependent Variable	SBP
Number of Independent Variables	2
Number of Continuous Independent Variables	1
Number of Class Independent Variables	1
Number of Observations	381
Optimization Algorithm	Simplex
Method for Confidence Limits	Inv_Rank

Number of Observations Read	381
Number of Observations Used	381

Class Level Information		
Name	Levels	Values
Gender	2	Female Male

Parameter Information		
Parameter	Effect	Gender
Intercept	Intercept	
Age	Age	
GenderFemale	Gender	Female
GenderMale	Gender	Male

Summary Statistics						
Variable	Q1	Median	Q3	Mean	Standard Deviation	MAD
Age	38.0000	47.0000	59.5000	47.9711	15.1112	14.8260
SBP	116.5	130.0	150.0	133.0	25.7158	26.6868

Quantile and Objective Function	
Quantile	0.5
Objective Function	3524.6667
Predicted Value at Mean	125.6474

(M10)

The QUANTREG Procedure

Parameter Estimates				
Parameter	DDL	Valeur estimée	95% Confidence Limits	
Intercept	1	93.6667	85.1770	105.8377
Age	1	0.6667	0.4928	0.8699
Gender Female	1	6.3333	-1.4259	12.3652
Gender Male	0	0.0000	0.0000	0.0000

Diagnostics		
Obs	Standardized Residual	Outlier
9	3.6637	*
15	3.2805	*
64	3.3571	*
82	3.4644	*
138	3.2192	*
174	3.3724	*
226	4.7061	*
248	3.4644	*
259	3.1119	*
338	-3.1272	*

Diagnostics Summary		
Observation Type	Proportion	Cutoff
Outlier	0.0262	3.0000

Test 1 Results				
Test	Test Statistic	DDL	Khi-2	Pr > Khi-2
Wald	39.0957	2	39.10	<.0001

Test 2 Results				
Test	Test Statistic	DDL	Khi-2	Pr > Khi-2
Wald	39.0898	1	39.09	<.0001

(M10)

The QUANTREG Procedure

Test 3 Results				
Test	Test Statistic	DDL	Khi-2	Pr > Khi-2
Wald	3.3116	1	3.31	0.0688

The NLIN Procedure
Dependent Variable Y
Method: Marquardt

Iterative Phase		
Iter	a	Sum of Squares
0	0.10000	383.5
1	0.5204	302.2
2	0.5228	302.2
3	0.5228	302.2

NOTE: Convergence criterion met.

Estimation Summary	
Method	Marquardt
Iterations	3
R	2.029E-6
PPC(b)	3.006E-6
RPC(b)	0.000081
Object	2.845E-9
Objective	302.1938
Observations Read	300
Observations Used	300
Observations Missing	0

Source	DDL	Somme des carrés	Moyenne quadratique	F Value	Approx Pr > F
Model	1	101.2	101.2	99.84	<.0001
Error	298	302.2	1.0141		
Corrected Total	299	403.4			

Exercise 5

(M11)

The NLIN Procedure

Parameter	Valeur estimée	Approx Std Error	Approximate 95% Confidence Limits	
a	0.5228	0.0587	0.4072	0.6384
b	0.9037	0.0776	0.7511	1.0564

Approximate Correlation Matrix		
	a	b
a	1.0000000	-0.1421505
b	-0.1421505	1.0000000