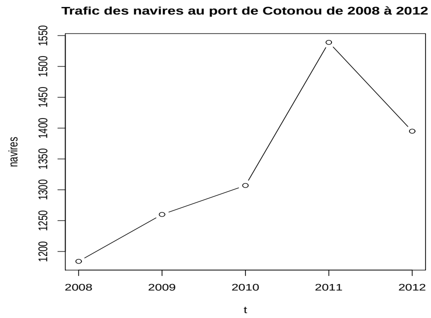
Master 1 MPM **Pôle Lamartine - ULCO**

Statistiques Flux portuaires 23 Octobre 2013 - Contrôle Terminal - Session 1

Durée de l'épreuve : 2h00 Documents interdits. Calculatrice autorisée

Exercice 1 - *Correction* : 10pts

1. (a) 0,5pt On a le graphique suivant :



Après avoir progressé pendant 4 années successives, le trafic est désormais à la baisse.

(b) 2pts Soient T et X_t les variables respectives "Année" et "Nombre de navires". On complète dans un premier temps le tableau de l'énoncé :

						TOTAL
t	2008	2009	2010	2011	2012	10050
t^2	4032064	4036081	4040100	4044121	4048144	20200510
x_t	1184	1260	1307	1539	1395	6685
x_t^2	1401856	1587600	1708249	2368521	1946025	9012251
tx_t	2377472	2531340	2627070	3094929	2806740	13437551

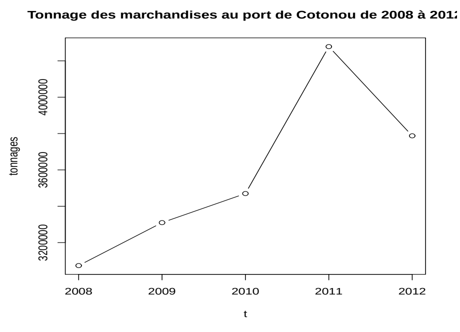
On peut ensuite calculer les paramètres suivants :

- $\overline{t} = \frac{10050}{5} = 2010$,
- $\overline{x_t} = \frac{6685}{5} = 1337$,
- $V(T) = \frac{20200510}{5} - (2010)^2 = 2$,
- $V(X_t) = \frac{9012251}{5} - (1337)^2 = 14881,2$,
- $\text{Cov}(T, X_t) = \frac{13437551}{5} - 2010 \times 1337 = 140,2$,
- $a = \frac{\text{Cov}(T,X_t)}{V(T)} = \frac{140,2}{2} = 70,1$,
- $b = \overline{x_t} - a\overline{t} = 1337 - 70,1 \times 2010 = -139564$.

Finalement, la droite des moindres carrés de X sachant T admet pour équation $D_{X/T} : x_t = 70,1t - 139564$. On en déduit les prévisions de trafic pour les 5 années 2013 à 2017 en remplaçant dans l'équation t par $2013, \ldots, 2017$:

t	2013	2014	2015	2016	2017
x_t	1547,3	1617,4	1687,5	1757,6	1827,7

2. (a) 0,5pt On a le graphique suivant :



(b) $\boxed{\text{2pts}}$ Soit Y_t la variable "Tonnage de marchandises". On complète dans un premier temps le tableau de l'énoncé :

t	2008	2009	2010	2011	2012	TOTAL
						10050
t^2	4032064	4036081	4040100	4044121	4048144	20200510
y_t	3073490	3309890	3469912	4278286	3787511	17919089
y_t^2	$9,446341 \times 10^{12}$	$1,095537 \times 10^{13}$	$1,204029 \times 10^{13}$	$1,830373 \times 10^{13}$	$1,434524 \times 10^{13}$	$6,509097 \times 10^{13}$
ty_t	6171567920	6649569010	6974523120	8603633146	7620472132	36019765328

On peut ensuite calculer les paramètres suivants :
- $\bar{t} = \frac{10050}{5} = 2010$,
- $\overline{y_t} = \frac{17919089}{5} = 3583818$,
- $V(T) = \frac{20200510}{5} - (2010)^2 = 2$,
- $V(Y_t) = \frac{6{,}509097 \times 10^{13}}{5} - (3583818)^2 = 174444486975$,
- $\text{Cov}(T, Y_t) = \frac{36019765328}{5} - 2010 \times 3583818 = 479287,6$,
- $a = \frac{\text{Cov}(T,Y_t)}{V(T)} = \frac{479287{,}6}{2} = 239643,8$,
- $b = \overline{y_t} - a\bar{t} = 3583818 - 239643,8 \times 2010 = -478100220$.

Finalement, la droite des moindres carrés de Y_t sachant T admet pour équation $D_{Y_t/T} : y_t = 239643,8t - 478100220$. On en déduit les prévisions de trafic pour les 5 années 2013 à 2017 en remplaçant dans l'équation t par $2013, \ldots, 2017$:

t	2013	2014	2015	2016	2017
y_t	4302749	4542393	4782037	5021681	5261325

3. (a) – $\boxed{\text{1pt}}$ Soit t le rang du trimestre considéré, Y_t la variable "Tonnages trimestriels" et y_t les valeurs atteintes par Y_t. Calculons le trend par la méthode des moyennes mobiles, avec $k = 4$ par exemple, et la formule

$$\tilde{y}_t = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}.$$

On a le tableau suivant :

Année \ Trimestre	1	2	3	4
2008	734909,4	776078,14	776062,5	794860,75
2009	813062,13	823813,88	826111,75	826134,38
2010	841777,38	861757,5	870757,13	897654,75
2011	976544,38	1050693,6	1071944,13	1063206,63
2012	1013109,38	960500	995020	1070760,4

– $\boxed{\text{1,5pt}}$ Calculons ensuite le trend par la méthode des moindres carrés. On dresse le tableau suivant :

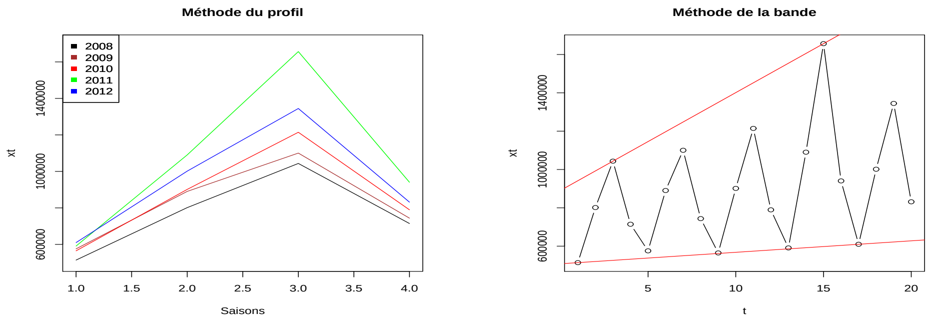
t	y_t	t^2	y_t^2	ty_t	t	y_t	t^2	y_t^2	ty_t
1	514034	1	$2,642310 \times 10^{11}$	514034	11	1214389	121	$1,474741 \times 10^{12}$	13358279
2	801456	4	$6,423317 \times 10^{11}$	1602912	12	789466	144	$6,232566 \times 10^{11}$	9473592
3	1043567	9	$1,089032 \times 10^{12}$	3130701	13	590901	169	$3,491640 \times 10^{11}$	7681713
4	714433	16	$5,104145 \times 10^{11}$	2857732	14	1090337	196	$1,188835 \times 10^{12}$	15264718
5	575554	25	$3,312624 \times 10^{11}$	2877770	15	1656558	225	$2,744184 \times 10^{12}$	24848370
6	890332	36	$7,926911 \times 10^{11}$	541992	16	940490	256	$8,845214 \times 10^{11}$	15047840
7	1100312	49	$1,210686 \times 10^{12}$	7702184	17	609882	289	$3,719561 \times 10^{11}$	10367994
8	743702	64	$5,530927 \times 10^{11}$	5949616	18	1001456	324	$1,002914 \times 10^{12}$	18026208
9	564668	81	$3,188500 \times 10^{11}$	5082012	19	1344661	361	$1,808113 \times 10^{12}$	25548559
10	901389	100	$8,125021 \times 10^{11}$	9013890	20	831512	400	$6,914122 \times 10^{11}$	16630240
				Totaux	210	17919099	2870	$1,766419 \times 10^{13}$	200320356

On peut ensuite calculer les paramètres suivants :

- $\bar{t} = \frac{210}{20} = 10,5$,
- $\overline{y_t} = \frac{17919099}{20} = 895954,9$,
- $V(T) = \frac{2870}{20} - (10,5)^2 = 33,25$,
- $V(Y_t) = \frac{1,687966 \times 10^{13}}{20} - (855954,9)^2 = 80474297227$,
- $\mathrm{Cov}(T, Y_t) = \frac{200320356}{20} - 895954,9 \times 10,5 = 608490,8$,
- $a = \frac{\mathrm{Cov}(T,Y_t)}{V(T)} = \frac{608490,8}{33,25} = 18300,48$,
- $b = \overline{y_t} - a\bar{t} = 895954,9 - 18300,48 \times 10,5 = 703799,9$.

Finalement, la droite des moindres carrés de Y_t sachant T admet pour équation $D_{Y_t/T} : y_t = 18300,48t + 703799,9$.

(b) $\boxed{\text{1,5pt}}$ Calculons les coefficients saisonniers à l'aide des moyennes mobiles. Il faut d'abord déterminer le modèle de la série (t, Y_t).



On rappelle que pour faire cette détermination (modèle additif, multiplicatif) graphiquement, on peut par exemple superposer les saisons représentées par des droites de profil sur un même graphique (méthode du profil). Si ces droites sont parallèles, le modèle est additif, autrement le modèle est multiplicatif. Sur le graphique de notre exemple (ci-dessus à gauche), les droites de profil ne sont pas parallèles pour toutes les saisons, le modèle est multiplicatif. Ceci est confirmé en utilisant une autre méthode graphique : la méthode de la bande, dont le principe est le suivant : on fait un graphique représentant la série chronologique, puis on trace une droite passant respectivement par les minima et par les maxima de chaque saison. Si ces deux droites sont parallèles, nous sommes en présence d'un modèle additif. Dans le cas contraire, cest un modèle multiplicatif. On retrouve bien la conclusion précédente grâce à la figure ci-dessus à droite.

On détermine donc les rapports à la tendance $r_t = \frac{y_t}{\tilde{y}_t}$ où les $\tilde{y}_t$ sont les moyennes mobiles. On a

Année \ Trimestre	1	2	3	4
2008	0,6994522	1,0327001	1,3446945	0,8988153
2009	0,7078844	1,0807441	1,3319167	0,9002192
2010	0,6708044	1,0459892	1,3946357	0,8794762
2011	0,6050939	1,0377307	1,5453772	0,8845788
2012	0,6019903	1,0426403	1,3513909	0,7765622

On en déduit les coefficients saisonniers :

- $s_1 = \frac{0,6994522+0,7078844+0,6708044+0,6050939+0,6019903}{5} = 0,657045$,
- $s_2 = \frac{1,0327001+1,0807441+1,0459892+1,0377307+1,0426403}{5} = 1,047961$,
- $s_3 = \frac{1,3446945+1,3319167+1,3946357+1,5453772+1,3513909}{5} = 1,393603$,
- $s_4 = \frac{0,8988153+0,9002192+0,8794762+0,8845788+0,7765622}{5} = 0,8679303$.

Comme $\bar{s} = \frac{0,657045+1,047961+1,393603+0,8679303}{4} = 0,9916348 \neq 1$, on rectifie les coefficients saisonniers :

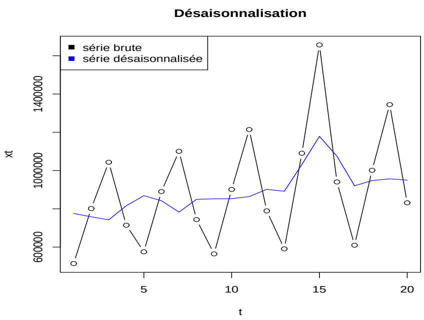
- $s_1^\star = \frac{s_1}{\bar{s}} = \frac{0,657045}{0,9916348} = 0,6625877$,

- $s_2^\star = \frac{s_2}{\bar{s}} = \frac{1{,}047961}{0{,}9916348} = 1,056801$,
- $s_3^\star = \frac{s_3}{\bar{s}} = \frac{1{,}393603}{0{,}9916348} = 1,405359$,
- $s_4^\star = \frac{s_4}{\bar{s}} = \frac{0{,}8679303}{0{,}9916348} = 0,875252$.

(c) $\boxed{\text{1pt}}$ Afin de calculer les valeurs corrigées des variations saisonnières, il suffit de diviser chaque y_t par son coefficient saisonnier adéquat. On obtient le tableau :

Année \ Trimestre	1	2	3	4
2008	775797,7	758379,3	742562,6	816259,8
2009	868645,8	842478,4	782940,2	849700,4
2010	852216,2	852941,1	864113,0	901987,1
2011	891808,0	1031733,5	1178743,7	1074536,2
2012	920454,8	947629,7	956809,6	950025,8

On peut représenter les deux séries brute et désaisonnalisée sur un même graphique :



Exercice 2 *Correction* : $\boxed{\text{4pts}}$

1. $\boxed{\text{1pt}}$
   - *Le port de Canton est en 2005 le troisième port chinois en terme de trafic de marchandises avec 250 millions de tonnes de trafic, soit une progression de plus de* 16% *par rapport à l'année précédente.*
     Les deux premières informations se lisent directement dans le tableau 1. Ensuite, la progression se calcule par la formule $\frac{250-213}{213} \times 100 = 17,37\%$ qui est bien supérieure à 16%.
     Il est également le 7ème port chinois pour le trafic de conteneurs avec près de 4,6 millions d'EVP. Malgré sa 18ème place au classement mondial pour le trafic de conteneurs il reste essentiellement un port de transport de marchandises en vrac.
     Ces informations ne sont pas présentes dans le tableau 1.
   - *Avec 5,2 millions d'EVP en 2005, le port de Ningbo se place au 15ème rang mondial au classement des ports de conteneurs.*
     Ces informations ne sont pas présentes dans le tableau 1.
     En terme de trafic de marchandises, Ningbo est le second port chinois derrière Shanghai avec 268 millions de tonnes de trafic en 2005, contre 226 millions de tonnes en 2004. Il reste pour l'instant dans l'ombre de son concurrent régional, le port de Shanghai.
     Ces informations se lisent directement dans le tableau 1.
2. On complète le tableau initial :

PORT	x_{2004}	x^2_{2004}	x_{2005}	x^2_{2005}
Shanghai	379	143641	449	201601
Ningbo	228	51984	269	72361
Canton	213	45369	250	62500
Tianjin	209	43681	240	57600
Qingdao	163	26569	188	35344
Qinhuangdao	150	22500	169	28561
Dalian	143	20449	170	28900
Shenzhen	135	18225	153	23409
Total	1620	372418	1888	510276

- $\boxed{\text{1pt}}$ Calculs pour 2004 :
  - L'étendue est égale à $379 - 135 = 244$.
  - Les 8 modalités sont toutes des modes (chaque modalité est représentée une et une seule fois).
  - La série étant constituée de 8 valeurs rangées par ordre de grandeur décroissant, la médiane est la 1/2 somme de la 4ème et la 5ème valeurs. On a donc Mé= $\frac{209+163}{2} = 186$. 50% des valeurs de la série (ordonnée) sont donc inférieures - ou supérieures - à cette valeur.
  - $\overline{x}_{2004} = \frac{1620}{8} = 202,5$.
- $\boxed{\text{1pt}}$ Calculs pour 2005 :
  - L'étendue est égale à $449 - 153 = 296$.
  - Les 8 modalités sont toutes des modes (chaque modalité est représentée une et une seule fois).
  - On a Mé= $\frac{240+188}{2} = 214$. 50% des valeurs de la série (ordonnée) sont donc inférieures - ou supérieures - à cette valeur.
  - $\overline{x}_{2005} = \frac{1888}{8} = 236$.

3. $\boxed{\text{1pt}}$
   - Calculs pour 2004 : on a $V(X) = \frac{372418}{8} - (202,5)^2 = 5546$ donc $\sigma(X) = 74,47$, ce qui signifie que théoriquement, l'intervalle $[202,5 - 74,47; 202,5 + 74,47] = [128,03; 276,97]$ contient 68% des valeurs de la série portant sur 2004.
   - Calculs pour 2005 : on a $V(X) = \frac{510276}{8} - (236)^2 = 8088,5$ donc $\sigma(X) = 89,94$, ce qui signifie que théoriquement, l'intervalle $[236 - 89,94; 236 + 89,94] = [146,06; 325,94]$ contient 68% des valeurs de la série portant sur 2005.

Exercice 3 *Correction* : $\boxed{\text{6pts}}$

1. $\boxed{\text{0,5pt}}$ La porte de débarquement est un caractère :
   (a) qualitatif nominal
   (b) $\boxed{\text{qualitatif ordinal}}$
   (c) quantitatif discret
   (d) quantitatif continu.
2. $\boxed{\text{0,5pt}}$ Le tableau ci-dessous donne le nombre d'entrées dans le Port de Nantes relevé quotidiennement par la Direction au cours du mois de juin de l'année 2013.

Date	1/06	2/06	3/06	4/06	5/06	6/06	7/06	8/06
n_i	9	2	5	6	3	0	2	5

   La population étudiée est :
   (a) le nombre de bateaux entrant quotidiennement dans le port de Nantes
   (b) $\boxed{\text{le mois de juin de l'année 2013}}$
   (c) le port de Nantes
   (d) la flotte du port de Nantes.
3. $\boxed{\text{0,5pt}}$ Pour un caractère quantitatif discret, la fonction de répartition :
   (a) est représentée par un diagramme en bâtons
   (b) est connue pour les seules extrémités de classes
   (c) $\boxed{\text{est discontinue en chaque modalité } x_j,\ j = 1, \dots, k}$
   (d) prend la valeur 0 pour tout x supérieur à la plus grande valeur possible x_k.
4. $\boxed{\text{0,5pt}}$ Par définition, un histogramme se construit en associant à chaque valeur possible $x \in C_j = [e_j - 1; e_j[,\ j = 1, \dots, k$:
   (a) l'effectif n_j
   (b) la fréquence f_j
   (c) la fréquence cumulée F_j

(d) $\boxed{\text{la fréquence moyenne } f_j/a_j \text{ où } a_j \text{ est l'amplitude de } C_j.}$

5. $\boxed{\text{0,5pt}}$ Si la moyenne des salaires des manutentionnaires du Port Autonome de Dunkerque est à 1200 euros :
   (a) $\boxed{\text{les salariés, s'ils gagnaient tous le même salaire, percevraient tous 1200 euros}}$
   (b) la moitié des salariés gagnent moins de 1200 euros
   (c) 1200 euros est le salaire le plus fréquent
   (d) l'écart entre le plus petit salaire et le plus grand est de 1200 euros.
6. $\boxed{\text{0,5pt}}$ La moyenne des différences à la moyenne :
   (a) est d'autant plus grande que la série est dispersée
   (b) est d'autant plus petite que la série est concentrée
   (c) $\boxed{\text{est toujours égale à 0}}$
   (d) est égale à 0 uniquement si la série est symétrique.
7. $\boxed{\text{0,5pt}}$ La série $5, 7, 12, 15, 22, 25$:
   (a) n'admet pas de médiane
   (b) $\boxed{\text{admet pour médiane } 13,5}$
   (c) admet pour médiane $14,33$
   (d) admet pour intervalle médian $[12; 15]$.
8. $\boxed{\text{0,5pt}}$ Si 2 variables x et y sont en relation linéaire $y = ax + b$:
   (a) $V(y) = aV(x) + b$
   (b) $V(y) = aV(x)$
   (c) $V(y) = a^2V(x) + b$
   (d) $\boxed{V(y) = a^2V(x).}$
9. $\boxed{\text{0,5pt}}$ Dans le cas d'une distribution des salaires, le quantile x_α donne :
   (a) le pourcentage de salariés qui gagnent moins que x_α
   (b) $\boxed{\text{le niveau de salaire en deçà duquel on trouve } \alpha\% \text{ des individus}}$
   (c) le pourcentage de salariés compris dans un intervalle centré sur la moyenne de $\pm\alpha$
   (d) le pourcentage de salariés compris dans un intervalle centré sur la moyenne d'amplitude α.
10. $\boxed{\text{0,5pt}}$ Si la médiane est inférieure à la moyenne :
   (a) la série est asymétrique à droite
   (b) la série est asymétrique à gauche
   (c) il y a plus d'individus à droite de la moyenne
   (d) $\boxed{\text{il y a plus d'individus à gauche de la médiane.}}$
11. $\boxed{\text{0,5pt}}$ L'indice de Gini est d'autant plus grand que :
   (a) la courbe de concentration est proche de la première bissectrice
   (b) $\boxed{\text{la courbe de concentration est éloignée de la première bissectrice}}$
   (c) la distribution est proche d'une distribution égalitaire
   (d) la moyenne est grande et l'écart-type est petit.
12. $\boxed{\text{0,5pt}}$ Dans un tableau de contingence dénombrant N individus, la fréquence $f_{j.}$ vérifie :
   (a) $\boxed{f_{j.} = n_{j.}/N}$
   (b) $f_{j.} = n_{j.}/n_{.i}$
   (c) $\boxed{f_{j.} = \sum_i f_{ji}}$
   (d) $\boxed{f_{j.} = \frac{1}{n}\sum_i n_{ji}.}$