
THÈMES DES SÉANCES DE TD

Thème n.1: Tableaux statistiques et représentations graphiques.

Thème n.2: Indices de centralité.

Thème n.3: Indices de dispersion et de concentration; Boîtes à moustaches.

Thème n.4: Tableaux de contingence.

Thème n.5: Liaisons entre deux variables (χ^2 , rapport de corrélation).

Thème n.6: Liaison fonctionnelle entre deux variables; Régression; Corrélation linéaire.

Thème n.7: Séries chronologiques.

Thème n.8: Indices élémentaires et indices synthétiques.

FONCTIONNEMENT DES TD, DESCRIPTION DU POLYCOPIÉ ET MODE DE CONTRÔLE

- La 1ère séance de TD se déroulera la semaine du 30 Septembre au 04 Octobre. Le module est prévu sur dix semaines.
- **La présence est obligatoire.**
- Ce polycopié ainsi que la version imprimable des transparents de cours est téléchargeable à l'adresse suivante:
<http://www-ljk.imag.fr/membres/Jean-Francois.Coeurjolly/teaching.html>
L'adresse du précédent site est également en lien depuis le GoogleApps de la Licence d'Economie-Gestion.
- Ce polycopié est décomposé en huit sections différentes chacune correspondant à un thème. Chaque section est composée d'un QCM relié au Cours Magistral ainsi qu'un certain nombre d'exercices. Pour chaque séance, votre travail consiste à préparer le QCM correspondant au thème abordé ainsi que les éventuels exercices que votre enseignant vous aura demandé de préparer. Tout ou partie des préparations pourront être ramassées par l'enseignant de TD, libre de l'intégrer à sa note de contrôle continu. A la fin du polycopié de TD, vous trouverez un mini-guide d'utilisation de la calculatrice (qui ne vous dispense pas de lire la documentation de votre calculatrice) ainsi que les examens de Janvier 2012 et de 2013 dans le format dans lequel il ont été posés.
- L'examen final a lieu en Janvier et compte pour 50% de la note définitive en statistique descriptive. La note de TD compte également (donc) pour 50%. La note de TD (conservée en deuxième session le cas échéant) est une combinaison (choisie par l'enseignant de TD) de notes de tests faits au cours des séances, des préparations ramassées, de la participation aux TD ainsi que de la production d'un rapport de statistique sur un thème choisi par chaque étudiant.

QUELQUES RÉFÉRENCES

- Ce polycopié est largement inspiré du précédent polycopié de TD de Statistique Descriptive rédigé par **Alain Sombardier**.
- **Pierre Bailly**. *Exercices corrigés de statistique descriptive*. L'économie en plus, Presses Universitaires de Grenoble, 1990.
- **Agnès Hamon et Nicolas Jégou**. *Statistique Descriptive - Cours et exercices corrigés*. Pratique de la Statistique, Presses Universitaires de Rennes, 2008.
- **Thomas H. Wonacott et Ronald J. Wonacott**. *Statistique*. Economica, 5ème édition, 1995.
- Plusieurs jeux de données présents dans ce polycopié proviennent de rapports de statistiques d'étudiants de L1.

QUELQUES SITES WEB CONTENANT DE NOMBREUX JEUX DE DONNÉES

- Site web de l'INSEE: www.insee.fr
- Site de la Banque Mondiale: donnees.banquemondiale.org/
- Site web du daily-bourse (où l'on peut entre autres récupérer des séries chronologiques de chiffre d'affaires de grandes entreprises) <http://www.daily-bourse.fr/>
- Site web du Centre National de la Cinématographie <http://www.cnc.fr/web/fr/statistiques/>
- ...

Prénom, Nom et groupe de TD:

QCM RELATIF AU THÈME N.1

**TABLEAUX STATISTIQUES ET REPRÉSENTATIONS
GRAPHIQUES**

Attention: Plusieurs réponses sont parfois possibles.

Questions 1 à 5: Une grande entreprise utilise 5 usines de fabrication de taille différente. Les parts du chiffre d'affaires (CA) pour chacune d'entre elles sont: 30%, 30%, 20%, 15% et 5%.

Question 1: La population statistique étudiée est

- L'ensemble des grandes entreprises Les 5 usines de l'entreprise Le CA La taille des usines

Question 2: Un individu de cette population est

- Un ouvrier Une entreprise Une usine française Une des 5 usines étudiées

Question 3: La variable étudiée est

- Une usine. C'est une variable qualitative.
 Le CA. C'est une variable quantitative discrète.
 Le CA. C'est une variable quantitative continue.
 La part du CA de chaque usine. C'est une variable quantitative continue.

Question 4: L'outil adapté pour représenter cette variable est

- Un diagramme en bâtons. Un tuyau d'orgues. Un camembert.
 Un histogramme (fréquentiel).

Question 5: Si l'on représente la variable part du CA par un secteur sur disque complet, l'angle associé à la seconde entreprise est de

- 108° 54° 1.88 radians 0.94 radians

Questions 6 à 8: La répartition en 1984 du nombre de pièces des résidences principales en France est

Nombre de pièces	1	2	3	4	5	6 et plus
Effectif ($\times 10^3$)	1200	2700	4700	5500	3500	2500

Question 6: Un individu de la population statistique étudiée est

- Une résidence principale. Un nombre de pièces. Un ménage français. Une usine!

Question 7: La variable étudiée est de nature

- qualitative et ordinale. quantitative continue. quantitative discrète. qualitative et nominale.

Question 8: L'outil graphique pour représenter cette variable est

- Un tuyau d'orgues. Un diagramme en bâtons des effectifs.
 Un diagramme des fréquences cumulées. Un histogramme (fréquentiel).

Question 9: Le terme "données individuelles"

- signifie que les observations sont nécessairement quantitatives.
- signifie que les observations sont nécessairement qualitatives.
- signifie que les données ont été regroupées par classes ou groupe de modalités.
- signifie que la variable est observée pour chaque individu de la population.

Question 10: Le terme "données agrégées"

- signifie que les observations sont nécessairement quantitatives.
- signifie que les observations sont nécessairement qualitatives.
- signifie que les données ont été regroupées par classes ou groupe de modalités.
- signifie que les données n'ont pas été observées.

Questions 11 à 14: Sur 200000 appels reçus pour une émission de télévision, 160000 sont reçus entre 14h et 19h, les autres sont reçus entre 19h et 20h.

Question 11: La variable étudiée est

- le nombre d'appels. C'est une variable quantitative discrète.
- le nombre d'appels. C'est une variable qualitative discrète.
- l'heure d'appel. C'est une variable quantitative discrète.
- l'heure d'appel. C'est une variable quantitative continue.

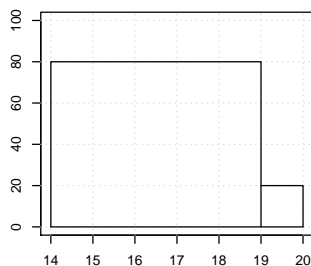
Question 12: Complétez le tableau statistique suivant (les notations sont les mêmes que celles utilisées en cours).

Temps (h)	$n_i (\times 10^3)$	f_i	a_i	d_i
[14, 19[160			
[19, 20]				

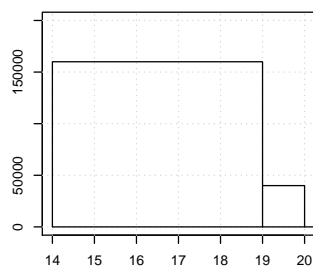
Question 13: L'histogramme fréquentiel correspondant à la variable heure d'appel est (lorsque vous l'aurez trouvé complétez les axes)

- le graphique 1. le graphique 2. le graphique 3. le graphique 4.

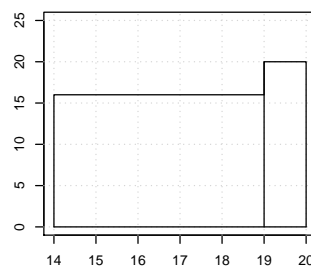
Graphique 1



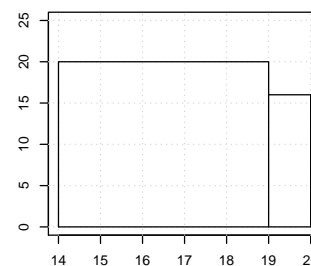
Graphique 2



Graphique 3



Graphique 4



Question 14: L'intervalle d'une heure pour lequel la proportion d'appels a été la plus élevée est

- 14h-15h 18h-19h 18h30-19h30 19h-20h

EXERCICES RELATIFS AU THÈME N.1
TABLEAUX STATISTIQUES ET REPRÉSENTATIONS
GRAPHIQUES

Exercice 1 Le tableau ci-dessous issu du recensement de 2009 présente les effectifs de la population active des 15 ans ou plus par catégorie socio-professionnelle (source INSEE) et par sexe. Les données (légèrement approchées) correspondent à des milliers d'individus.

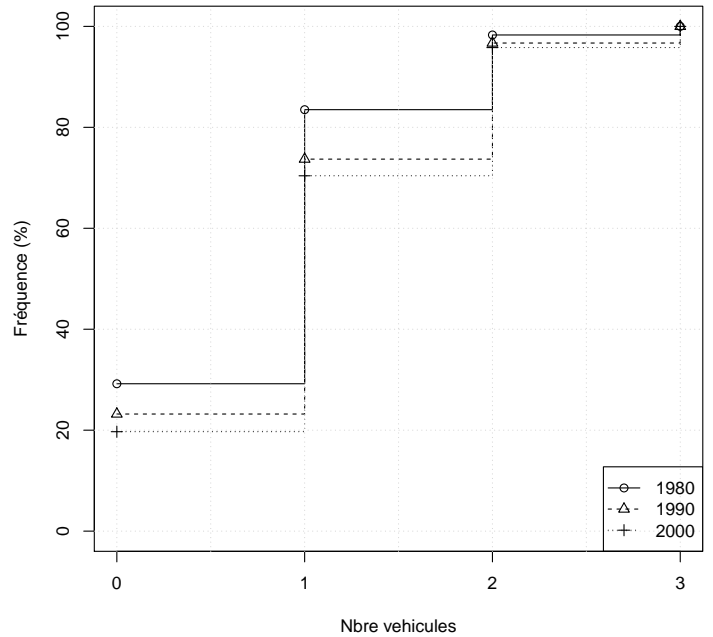
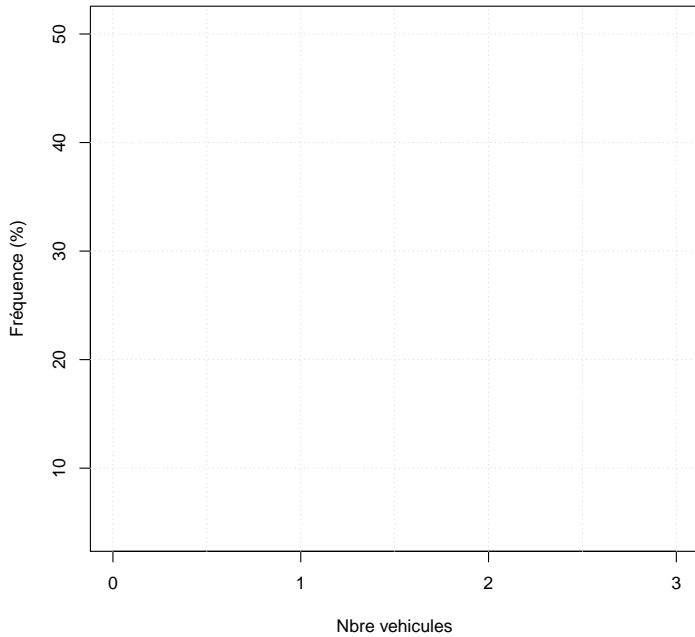
CSP	Hommes	Femmes	Ensemble
Agriculteurs	365	145	510
Artisans	1238	494	1732
Cadres	2750	1746	4496
Professions intermédiaires	3443	3781	7224
Employés	2037	6700	8737
Ouvriers	5719	1365	7084
Total	15552	14231	29783

1. Quels sont la population statistique étudiée, un individu issu de cette population, les variables étudiées ainsi que leur nature?
2. Représentez graphiquement sous la forme d'un secteur sur disque complet (camembert) la répartition de la population active (indépendamment du sexe)?
3. Parmi les employés, quelle est la proportion de femmes?
4. A l'intérieur de chaque secteur représentez la part des hommes et celle des femmes.

Exercice 2 Le tableau ci-dessous représente la distribution en fréquences du nombre de véhicules par ménages français en 1980, 1990, 2000 et 2010 (source INSEE).

Nbre de véhiules	1980	1990	2000	2010
0	29.2%	23.2 %	19.7%	16.5 %
1	54.3%	50.5%	50.7%	47.6%
2	14.8%	23%	25.4%	30.7%
3	1.7%	3.3%	4.2%	5.2%

1. Quels sont la population statistique étudiée, un individu issu de cette population, la variable étudiée ainsi que sa nature?
2. En particulier, représentez sous la forme d'un diagramme en bâtons la distribution du nombre des véhicules en 2010.
3. Sur le second graphique ajoutez la courbe des fréquences cumulées de la distribution du nombre des véhicules en 2010. Commentez.

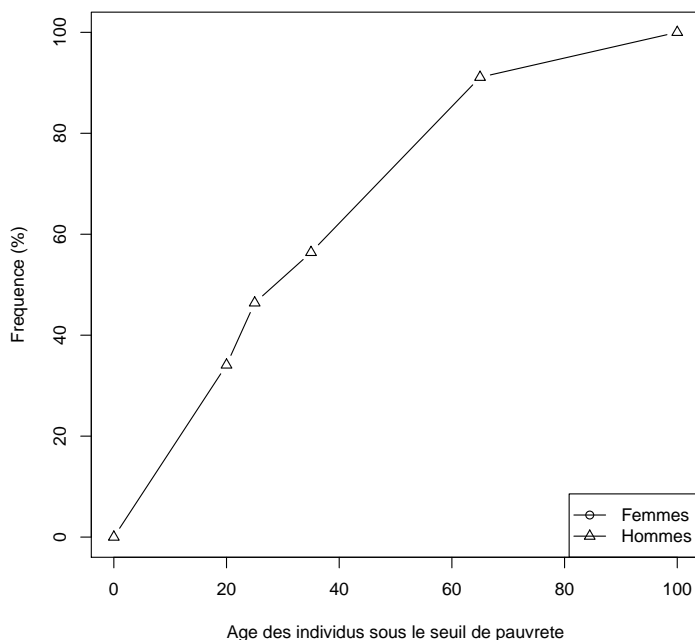
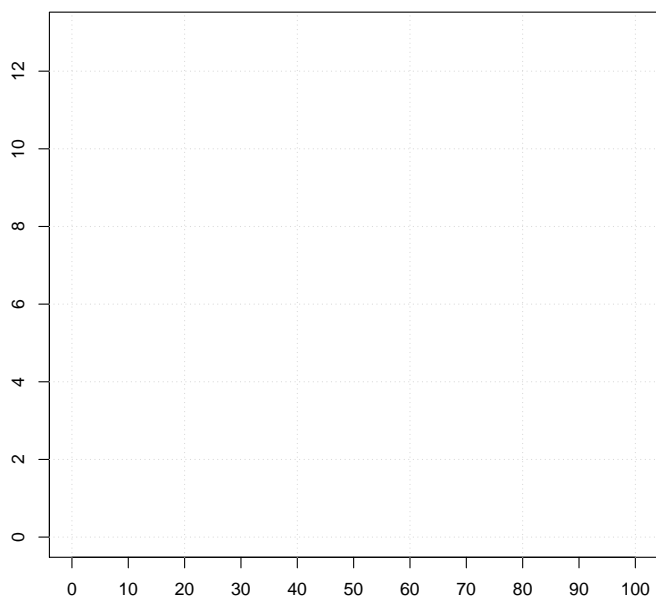


4. Quel est pour chacune des dates, la proportion des ménages ayant au plus 1 véhicule? au moins 2?
5. Pour chacune des dates, déterminez graphiquement le plus grand nombre N tel que moins de 80% des ménages français ont un nombre de véhicules inférieur ou égal à N ?

Exercice 3 Le tableau suivant représente le nombre de personnes (en milliers) sous le seuil de pauvreté (60% du revenu médian de la population) selon leur âge et leur sexe en France en 2010 (source INSEE).

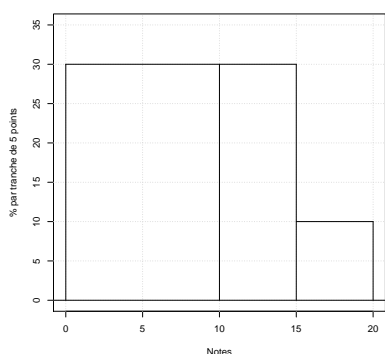
Age	Femmes	Hommes	Ensemble
[0, 18[1327	1338	
[18, 25[564	481	
[25, 35[537	391	
[35, 65[1580	1361	
[65, 100]	689	350	
Total			

1. Quels sont la population statistique étudiée, un individu issu de cette population, les variables étudiées ainsi que leur nature?
2. Complétez le précédent tableau.
3. Représentez dans le premier graphique la distribution de l'âge des individus sous le seuil de la pauvreté indépendamment du sexe. N'oubliez pas de préciser les axes. Commentez.
4. Comment représenter sur ce premier graphique la proportion des individus sous le seuil de pauvreté âgés de plus de 40 ans?



5. Sur le second graphique (ci-dessus), représentez la courbe des fréquences cumulées de l'âge des femmes sous le seuil de pauvreté. Commentez.
6. Estimez graphiquement la proportion de femmes puis d'hommes sous le seuil de pauvreté ayant un âge inférieur à 60 ans.
7. Estimez graphiquement l'âge A des femmes (puis celui des hommes) sous le seuil de pauvreté tel quel 60% des femmes (resp. des hommes) aient un âge inférieur à A .

Exercice 4 Sur 200 notes d'étudiants à un devoir de statistique descriptive, on a observé l'histogramme fréquentiel ci-dessous. Complétez le tableau statistique correspondant.



notes	n_i	f_i	F_i	a_i	d_i
$[0, \dots [$					
$[\dots, \dots [$					
$[\dots, 20]$					

Exercice 5 (Annales 2012 et 2013 - Partie I) Traitez la question n.1 des examens de Janvier 2012 et de Janvier 2013.

Prénom, Nom et groupe de TD:

QCM RELATIF AU THÈME N.2

INDICES DE CENTRALITÉ

Attention: Plusieurs réponses sont parfois possibles.

Question 1: La notation $\sum_{i=1}^{10} i^2$ est égale à

- $1^2 + 2^2 + \dots + 10^2$ $(1 + 2 + \dots + 10)^2$ $1 + 1^2 + \dots + 1^{10}$
 $1 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 8^2 + 9^2 + 10^2$

Question 2: La formule de la moyenne arithmétique d'une variable quantitative continue (avec les notations standard) est

- $\bar{x} = \frac{1}{n} \sum_{i=1}^p n_i c_i$. $(n_1 \times c_1 \times \dots \times n_p \times c_p) / n$. $(n_1 + c_1 \times \dots \times n_p + c_p) / n$. $(n_1 \times c_1 + \dots + n_p \times c_p) / n$.

Question 3: La médiane d'une série statistique quantitative x s'interprète comme

- la valeur partageant la population en deux groupes de même fréquence.
 la valeur d'une modalité ayant la plus grande fréquence.
 la valeur d'une modalité telle que 50% des données soient au delà de cette valeur.
 la moyenne entre les plus petite et grande valeurs de x .

Question 4: Sur 20 lancers d'un dé à 6 faces, on a obtenu 8 fois le 2, 4 fois le 3, 3 fois le 4, 4 fois le 5 et 1 fois le 1. La médiane de cette série statistique est

- égale au mode de la distribution. égale à 3. égale à 3.5. égale à 4.

Question 5: Avec les notations standard, la formule de la médiane d'une variable quantitative continue est

- $Me = x_i + \frac{50\% - F_i}{F_{i+1} - F_i} (x_{i+1} - x_i)$. $Me = x_i + \frac{F_{i+1} - 50\%}{F_{i+1} - F_i} (x_{i+1} - x_i)$. $(x_i + x_{i+1}) / 2$. Autre.

Question 6: Pour une variable qualitative, les indices de centralité ayant du sens sont

- la moyenne arithmétique. le mode. la médiane. l'écart-type.

Question 7: Le mode d'une variable quantitative discrète x est nécessairement

- une valeur observée. une valeur partageant la population en deux groupes égaux.
 une modalité de x ayant la plus grande fréquence. une modalité de x ayant le plus grand effectif.

Question 8: La classe modale d'une variable quantitative continue x

- nécessairement une classe observée. partage la population en deux groupes égaux.
 est la classe ayant la plus grande fréquence. est la classe ayant la plus grande densité.

Question 9: Une entreprise compte deux fois plus d'hommes que de femmes. Le salaire moyen des hommes est de 2000 euros tandis que celui des femmes 1700 euros. Le salaire moyen de cette entreprise est compris entre 1700 et 2000 euros. vaut 1850 euros. vaut 1900 euros. est incalculable.

Question 10: Une entreprise compte 100 salariés. Le salaire mensuel moyen des 99 premiers salariés est de 1000 euros. Le salaire mensuel du 100ème salarié vaut 2000 euros. Le salaire moyen des 100 salariés compris entre 1000 et 2000 euros. vaut 1500 euros. vaut 1010 euros. est incalculable.

Question 11: Le prix d'un produit a augmenté de 20% de 2010 à 2011 puis diminué de 20% de 2011 à 2012. Ce prix a globalement augmenté de 2010 à 2012. Ce prix a globalement diminué de 2010 à 2012. Ce prix est le même en 2010 et 2012. On ne peut pas connaître le prix du produit en 2012.

Question 12: Dans le même contexte que précédemment, ce produit a globalement subit une augmentation de 4% sur deux ans, soit en moyenne exactement 2% par an. augmentation de 4% sur deux ans, soit en moyenne 2.02% par an. diminution de 4% sur deux ans, soit en moyenne exactement 2% par an. diminution de 4% sur deux ans, soit en moyenne 2.02% par an.

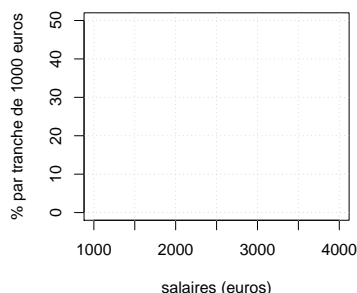
Question 13: Imaginons représenter les salaires d'une entreprise regroupés en trois classes de même amplitude entre 1000 et 4000 euros. Représentez l'histogramme d'une série et complétez le tableau correspondant de telle façon que le mode, la moyenne et la médiane soient toutes trois égales.

salaire	f_i	a_i	d_i
[1000, 2000[1	
[2000, 3000[1	
[3000, 4000[1	



Question 14: Même contexte que la question précédente. Représentez l'histogramme fréquentiel d'une série et complétez le tableau correspondant de telle façon que la moyenne et médiane soient toutes deux égales et différent du mode (ou classe modale).

salaire	f_i	a_i	d_i
[1000, 2000[1	
[2000, 3000[1	
[3000, 4000[1	



<p>EXERCICES RELATIFS AU THÈME N.2</p> <p>INDICES DE CENTRALITÉ</p>

Exercice 6 Reprendre les données de l'exercice 2 (nombre de véhicules par ménage français).

1. Pour chaque date, déterminez le mode.
2. Calculez la moyenne arithmétique.
3. Enfin déterminez par le calcul et graphiquement la médiane.
4. Commentez l'ensemble des résultats.

Exercice 7 On considère ici les données présentées dans l'exercice 3 (seuil de pauvreté).

1. Quelle est la ou les classes modale(s) de l'âge des femmes sous le seuil de pauvreté? Même question pour la population des hommes.
2. Calculez l'âge moyen des femmes sous le seuil de pauvreté?
3. L'âge moyen des individus sous le seuil de pauvreté est de 35.6 (ans). Déterminez avec le minimum de calcul l'âge moyen des hommes sous le seuil de pauvreté?
4. Déterminez l'âge des femmes sous le seuil de pauvreté partageant la population en deux groupes égaux? Peut-on déduire l'âge médian des hommes de celui des femmes et de celui de l'ensemble de la population?
5. Estimez graphiquement Me_H puis comparez l'ensemble des résultats. Que dire des indices de centralité des âges des femmes et hommes sous le seuil de pauvreté?

Exercice 8 1. Le PIB des USA est passé de 9899 à 10007 billions de dollars de 2000 à 2001 (à dollar constant en 2000) tandis que celui de la Chine est passé de 1198 à 1298 billions de \$ sur la même période. Calculez l'évolution annuelle du PIB sur cette période et pour ces deux pays.

2. Complétez alors le tableau ci-dessous détaillant l'évolution annuelle en pourcentage du PIB des USA et de la Chine sur la période 2000-2011 (source banque mondiale) et calculez le taux d'évolution annuel moyen pour les USA et la Chine.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
USA (%)		1.8	2.6	3.5	3.1	2.7	1.9	-0.4	-3.5	3.0	1.7
Chine (%)		9.1	10.0	10.1	5.5	18.9	14.2	9.6	9.2	10.4	9.3

3. Supposons dans cette question et la suivante que l'évolution du PIB dans les années à venir est constante et varie selon le taux d'évolution annuel moyen calculé précédemment. Quelles seraient les valeurs des PIB des USA et de la Chine en 2021?
4. A quelle date le PIB de la Chine dépassera celui des USA?

Exercice 9 *Le CA d'une entreprise a augmenté de 5% pendant deux ans, puis diminué de 4% pendant 3 ans et a réaugmenté de 1.5% pendant la dernière année. Quel est le taux d'évolution moyen du CA sur cette période?*

Exercice 10 *Le tableau suivant issu du site web de l'INSEE donne l'évolution des effectifs et des salaires annuels moyens en 2009 et 2010 pour quatre catégories socio-professionnelles.*

	<i>Effectifs (en milliers)</i>		<i>Salaire annuel moyen (en euros)</i>	
	<i>2009</i>	<i>2010</i>	<i>2009</i>	<i>2010</i>
<i>Cadres supérieurs</i>	<i>4236</i>	<i>4298</i>	<i>45106</i>	<i>46312</i>
<i>Professions intermédiaires</i>	<i>6214</i>	<i>6252</i>	<i>25406</i>	<i>25856</i>
<i>Employés</i>	<i>7469</i>	<i>7415</i>	<i>17867</i>	<i>18116</i>
<i>Ouvriers</i>	<i>5563</i>	<i>5468</i>	<i>18565</i>	<i>18828</i>

- 1. Calculez le salaire moyen pour l'ensemble de la population étudiée en 2009 et 2010. Quelle est l'évolution des salaires moyens de 2009 à 2010? Interprétez.*
- 2. Calculez le salaire moyen en 2010 à effectif fixé en 2009 puis le salaire moyen en 2010 à salaire fixé en 2009? Comment alors expliquer l'évolution obtenue à la question précédente?*

Exercice 11 (Annales des examens 2012 et 2013 - Partie I) *Traitez les questions n.2 à 4 de l'examen de 2012, ainsi que les questions 2(a)-(b) de l'examen de 2013.*

Prénom, Nom et groupe de TD:

QCM RELATIF AU THÈME N.3

INDICES DE DISPERSION ET DE CONCENTRATION

Attention: Plusieurs réponses sont parfois possibles.

Question 1: Plus l'écart-type d'une série statistique est élevé et

- plus la variable est concentrée. plus la variable est dispersée. plus la variance est grande.
 plus la variance est faible.

Question 2: L'écart-type d'une série statistique x représentant le salaire d'une entreprise compris entre 1000 et 2000 euros

- s'exprime en euros. s'exprime en euros². n'a pas d'unité. $\in [1000, 2000]$.

Question 3: La variance d'une série statistique x

- est toujours positive. comprise nécessairement entre 0 et 100%.
 est égale à la racine carrée de l'écart-type est égale au carré de l'écart-type.

Question 4: La variance d'une série statistique x continue (regroupée en classe) est égale

- à $\sum_{i=1}^p n_i(c_i - \bar{x})^2$. à $\bar{x}^2 - \bar{x}^2$. à $(n_1(\bar{x} - c_1)^2 + \dots + n_p(\bar{x} - c_p)^2)/n$.
 au carré de la moyenne moins la moyenne des carrés.

Question 5: Sur 200 notes arrondies au point entre 0 et 20 d'étudiants à un examen de statistiques, on a récupéré les informations suivantes: $\sum_{i=1}^{200} x_i = 2100$ et $\sum_{i=1}^{200} x_i^2 = 24600$. Cochez les affirmations vraies

- $\bar{x} = 10.5$. $\sigma \simeq 3.57$. $Var(x) = -21927$. $Var(x) = 12.75$.

Question 6: Le coefficient de variation

- est égal à \bar{x}/σ et s'exprime en %. est égal à σ/\bar{x} et s'exprime en %.
 permet d'interpréter plus facilement une valeur d'un écart-type.
 est toujours plus petit que 100%.

Question 7: Ecrire dans l'encadré ci-dessous la formule d'un quantile Q_α d'ordre $\alpha \in (0, 1)$ d'une série statistique x représentant une variable continue. On supposera que $Q_\alpha \in [x_i, x_{i+1}[$ et que F_i et F_{i+1} représentent les valeurs des fréquences cumulées en x_i et x_{i+1} .

Question 8: L'intervalle interquartile $[Q_1, Q_3]$ signifie que

- 25% des observations sont supérieures à Q_1 . 75% des observations sont supérieures à Q_3 .
 50% des observations sont concentrées dans cet intervalle. 50% des observations se trouvent à l'extérieur de cet intervalle.

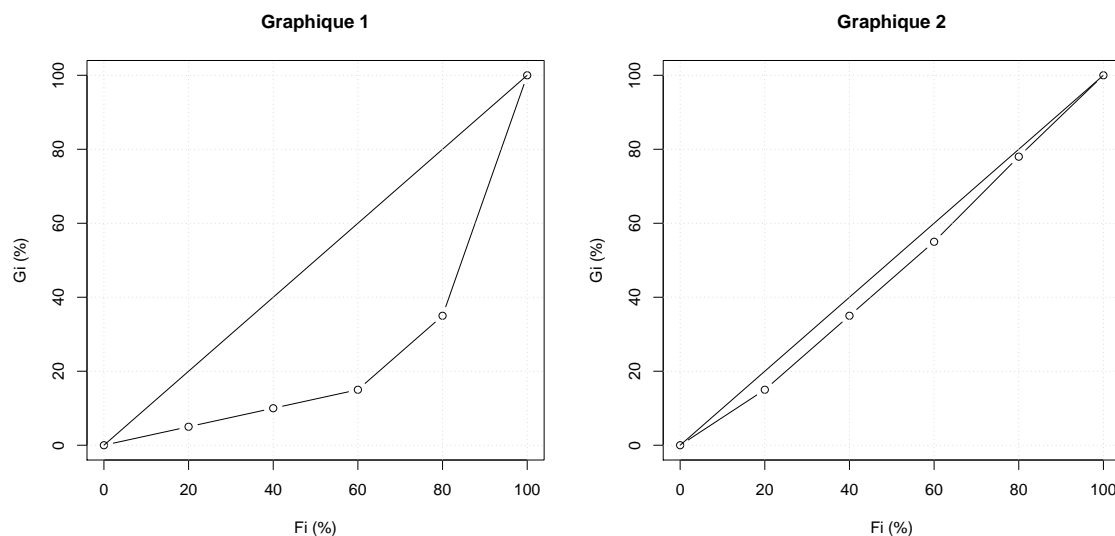
Question 9: On étudie les primes qu'une entreprise distribue à ses employés et employées. L'écart-type total des primes vaut 100 (euros) tandis que la variance intra-population vaut 8000 (euros²).

- La variance inter-population vaut 7900 (euros²). La variance inter-population vaut 2000 (euros²).
- Le sexe explique très peu la valeur des primes distribuées.
- Le sexe explique à 20% la dispersion des primes distribuées.

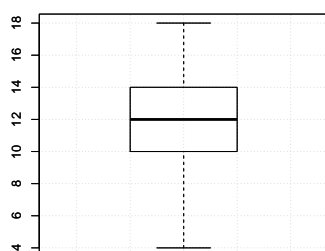
Question 10: Les graphiques ci-dessous représentent les courbes de concentration de deux entreprises. L'une est une multinationale tandis que l'autre est une coopérative. Cochez les affirmations vraies.

- Le graphique 1 montre que les salaires sont très peu concentrés dans cette entreprise.
- Le graphique 1 montre de sérieuses inégalités dans cette entreprise.
- L'indice de Gini associé au graphique 1 vaut à peu près 25%.
- L'indice de Gini associé au graphique 1 vaut à peu près 50%.

- Le graphique 2 montre que les salaires sont bien répartis dans cette entreprise.
- L'indice de Gini associé au graphique 2 est proche de 100%.
- L'indice de Gini associé au graphique 2 est proche de 0%.
- La courbe de Lorenz de la multinationale correspond au graphique 1.



Question 11: Complétez le tableau suivant à partir de la boîte à moustaches (mettre une croix quand l'indice n'est pas calculable).



Indice	Valeur
Etendue	
\bar{x}	
σ	
<i>Me</i>	
$[Q_1, Q_3]$	
D_1	
D_9/D_1	

<p>EXERCICES RELATIFS AU THÈME N.3</p> <p>INDICES DE DISPERSION ET DE CONCENTRATION</p>

Exercice 12 On considère ici les données de l'exercice 3. Complétez le tableau suivant et complétez l'analyse faite dans l'exercice 7 en terme de dispersion des deux séries.

Indice	Femmes	Hommes
Ecart-type σ (ans)		
Coefficient de variation (%)		
$[Q_1, Q_3]$ (ans)		
1er Décile D_1 (ans)		
9ème Décile D_9 (ans)		
Rapport inter-décile D_9/D_1 (%)		

Exercice 13 Une entreprise emploie 30 hommes et 45 femmes, le tableau suivant donne la répartition des salaires par CSP et pour chaque sexe. Les salaires mensuels moyens sont en milliers d'euros.

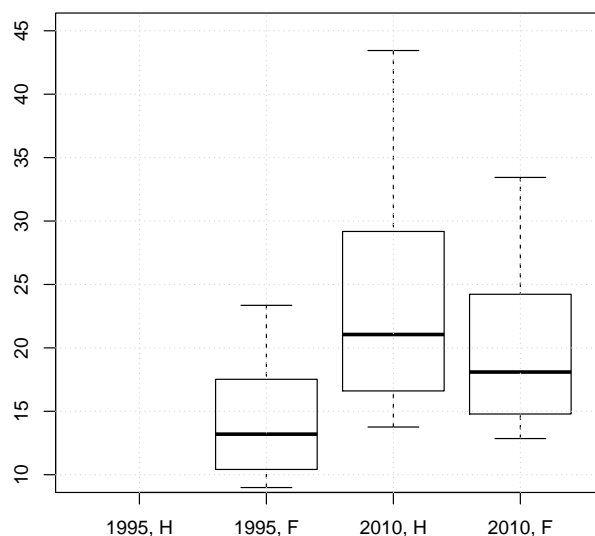
CSP	Hommes		Femmes		Ensemble	
	Effectifs	Salaires	Effectifs	Salaires	Effectifs	Salaires
Ouvriers	10	1	5	0.8	15	
Employés	15	2	38	1.8	53	
Cadres	5	6	2	5	7	

1. Complétez la dernière colonne (attention à la compensation hommes/femmes)
2. Calculez le salaire moyen des hommes noté \bar{x}_h , celui des femmes noté \bar{x}_f ainsi que pour l'ensemble de la population, noté simplement \bar{x} .
3. Calculez la variance totale des salaires, la variance inter-population ainsi que la variance intra-population.
4. Dans quelle mesure le sexe est-il un facteur de dispersion des salaires dans cette entreprise.

Exercice 14 Le tableau suivant détaille la répartition des des salaires nets moyens annuels (en euros) en 1995 et 2010 des hommes et des femmes. (source INSEE)

	D_1	Q_1	Me	Q_3	D_9	$Q_{95\%}$	$Q_{99\%}$
1995 hommes	9721	11844	15448	21753	32713	43526	78112
1995 femmes	8987	10415	13197	17523	23354	28753	46583
2010 hommes	13760	16606	21054	29178	43444	56921	104341
2010 femmes	12850	14786	18095	24226	33438	41506	67181

1. Le graphique ci-dessous représente les boîtes à moustaches des distributions des salaires des Hommes et Femmes en 1995 et 2010. Rajoutez la boîte à moustaches manquante.



2. On se concentre sur les salaires des hommes en 1995. A partir du tableau précédent, complétez le tableau statistique ci-dessous. On bornera la dernière classe à 110000 euros.

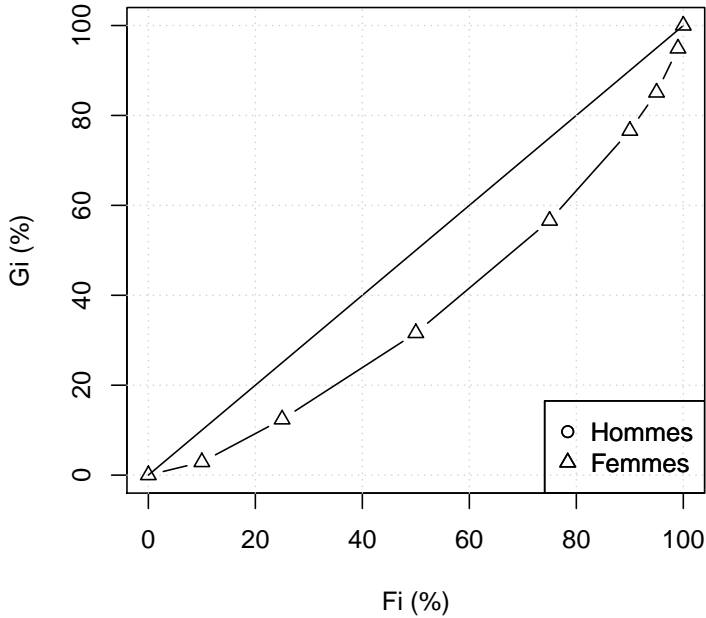
classe des salaires (euros)	c_i (euros)	Salaires		Masse salariale		
		f_i (%)	F_i (%)	$c_i \times f_i$	g_i (%)	G_i (%)
[0, 9721[10%			
[9721, 11844[
[11844, 15448[
[15448, 21753[
[21753, 32713[
[32713, 43526[
[43526, 78112[
[78112, 110000]						
<i>Total:</i>						

3. Complétez le tableau suivant en ajoutant la valeur de la médiale en 1995 et commentez.

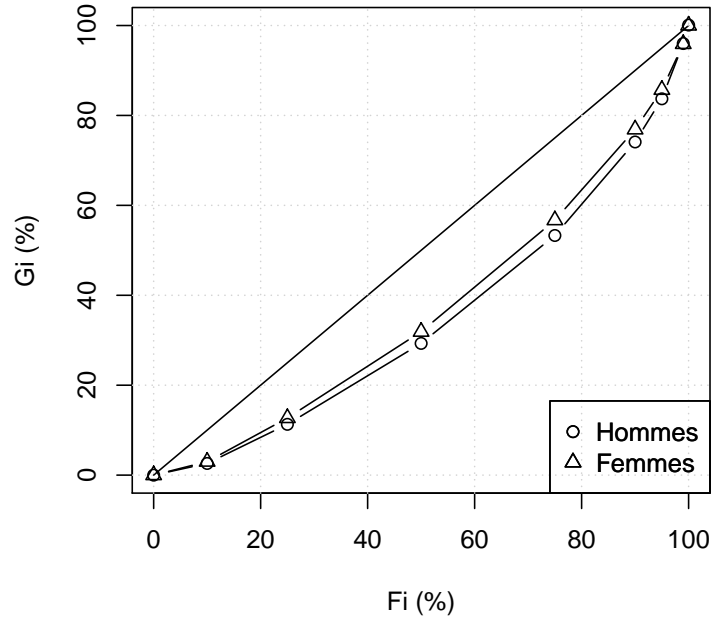
	1995		2010	
	Hommes	Femmes	Hommes	Femmes
Médiale (euros)		16380.9	28061	22569.6

4. Le graphique ci-dessous représente les courbes de concentration des salaires des femmes en 95 ainsi que des salaires des hommes et des femmes en 2010. Ajoutez la courbe de concentration des salaires des hommes en 1995. Commentez.

Concentration des salaires en 1995



Concentration des salaires en 2010



Exercice 15 1. Soit x_i pour $i = 1, \dots, n$ une série statistique quantitative (chaque modalité est observée une et une seule fois). Rappeler les formules de la moyenne et de l'écart-type.

2. Pour tout $i = 1, \dots, n$ on transforme les données initiales par une opération appelée **opération de centrage-réduction** consistant à définir une nouvelle série statistique par la formule

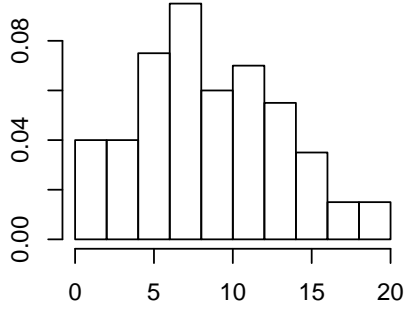
$$y_i = \frac{x_i - \bar{x}}{\sigma_x} \quad \text{pour } i = 1, \dots, n.$$

Montrer que $\bar{y} = 0$ et que $\text{Var}(y) = \sigma_y = 1$.

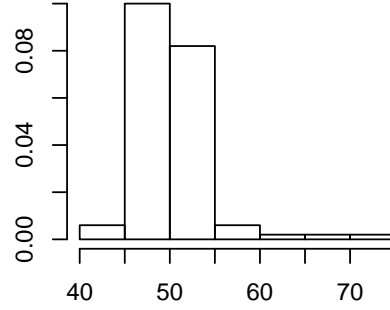
3. Dans quelle unité s'exprime la variable y ?

4. Cette procédure de centrage-réduction est particulièrement utile si on veut comparer deux distributions qui par exemple ne s'expriment pas dans la même unité. A titre d'exemple, on a récupéré les notes de deux groupes d'étudiants. Le Groupe 1 a été noté sur 20, tandis que le Groupe 2 a été noté sur 100. Le graphique ci-dessous représente les histogrammes fréquentiels de ces données ainsi que leurs versions centrées-réduites (les deux graphiques du bas). Quelle information supplémentaire apporte les représentations des versions centrées-réduites des notes des deux groupes?

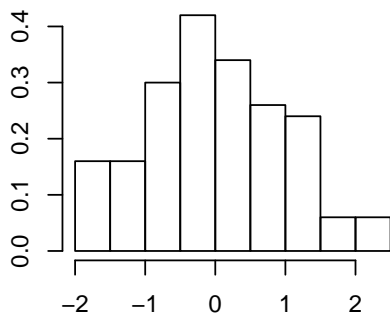
Groupe 1



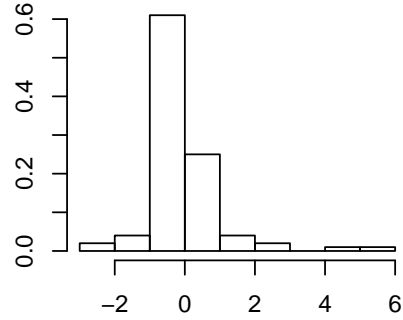
Groupe 2



Groupe 1 - CR



Groupe 2 - CR



Exercice 16 (Annales des examens de 2012 et 2013 - Partie I) Terminez la question 2 de l'examen de 2012 et traitez la question 2(c) et la question 3 de l'examen de 2013.

Prénom, Nom et groupe de TD:

QCM RELATIF AU THÈME N.4
TABLEAUX DE CONTINGENCE

Attention: Plusieurs réponses sont parfois possibles.

Questions 1 à 10: Une étude sur les prêts à la consommation accordés à des jeunes de 18-25 ans dans un certain organisme bancaire au cours d'une année a permis d'obtenir la répartition des prêts selon le montant et le type d'achat suivante:

montant type	véhicule	mobilier	trésorerie	divers	ensemble
[1000, 10000[41	14	24	22	101
[10000, 25000[123	33	15	18	189
[25000, 50000[78	13	1	4	96
[50000, 100000[20	1	1	2	24
Total	262	61	41	46	410

Question 1: A propos des variables

- Les variables sont qualitatives. Les variables sont quantitatives.
 Les variables sont ordinales. Les variables sont de nature différente.

Question 2: A propos des notations

- $n_{12} = 14$ $n_{12} = 123$ $n_{21} = 14$ $n_{21} = 123$
 $n_{3\bullet} = 96$ $n_{\bullet 3} = 41$ $n_{\bullet 4} = 46$ $n_{\bullet 4} = 24$

Question 3: A propos de l'interprétation d'un effectif ou fréquence partielle

- 41 prêts accordés l'ont été pour des montants entre 1000 et 10000 euros et pour l'achat d'un véhicule.
 41 prêts accordés l'ont été pour des montants entre 1000 et 10000 euros ou pour l'achat d'un véhicule.
 41 des prêts entre 1000 et 10000 euros l'ont été pour l'achat d'un véhicule.
 10% prêts accordés l'ont été pour des montants entre 1000 et 10000 euros et pour l'achat d'un véhicule.

Question 4: A propos de l'interprétation d'un effectif ou fréquence partielle

- 61 prêts ont été accordés pour l'achat de mobilier.
 61 prêts ont été accordés pour des montants entre 0 et 100000 euros.
 Environ 11.2% des prêts ont été accordés pour de la trésorerie.
 Environ 23.4% des prêts ont un montant entre 25000 et 50000 euros.

Question 5: Il y a

- exactement 4 distributions conditionnelles du montant sachant le type de prêt.
 exactement 5 distributions conditionnelles du montant sachant le type de prêt.
 autant de distributions conditionnelles du montant sachant le type que le type sachant le montant.
Et ceci est vrai pour toute table de contingence.
 autant de distributions conditionnelles du montant sachant le type qu'il y a de modalités de la variable type.

Question 6: Si on voulait représenter les distributions conditionnelles du montant sachant le type de prêts on utiliserait

- des camemberts. des tuyaux d'orgues. des histogrammes fréquentiels.
 des diagrammes en bâtons.

Question 7: Si on voulait représenter les distributions conditionnelles du type sachant le montant de prêts on utiliserait

- des camemberts. des tuyaux d'orgues. des histogrammes fréquentiels.
 des diagrammes en bâtons.

Question 8: $f_{1|3}$

- $= n_{13}/n_{1\bullet}$ $= n_{31}/n_{1\bullet}$ $= n_{13}/n_{3\bullet}$ $= n_{13}/n_{\bullet 3}$

Question 9: Environ 83.3%

- des prêts accordés entre 50000 et 100000 euros l'ont été pour l'achat d'un véhicule.
 des prêts accordés pour l'achat d'un véhicule ont un montant entre 50000 et 100000.
 des prêts sont accordés pour l'achat d'un véhicule.
 ont un montant entre 10000 et 25000 euros.

Question 10: La distribution conditionnelle du type sachant que le montant du prêt est entre 25000 et 50000 euros est

- 101/410, 189/410, 96/410, 24/410. 262/410, 61/410, 41/410, 46/410.
 78/96, 13/96, 1/96, 4/96. 78/410, 13/410, 1/410, 46/410.

Question 11: Supposons observer une table de contingence associée à deux variables X et Y . On suppose que X et Y sont indépendantes. Cochez les affirmations vraies

- Les lignes de la table de contingence sont identiques.
 Les lignes de la table de contingence sont proportionnelles entre elles.
 Les lignes de la table de contingence sont proportionnelles aux colonnes.
 Les effectifs marginaux de X et Y sont égaux.
 Les distributions conditionnelles de X sachant Y sont toutes les mêmes.
 Les distributions marginales de X et Y sont identiques.
 Les fréquences partielles sont toutes les mêmes.
 Les distributions de Y sachant X sont toutes les mêmes.
 $n_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{n}$
 $f_{i|j} = f_{j|i}$ pour tout i, j
 $f_{i|1} = f_{i|2}$ pour tout $i = 1, \dots, I$
 $f_{ij} = f_{i|j} \times f_{j|i}$

<p>EXERCICES RELATIFS AU THÈME N.4</p> <p>TABLEAUX DE CONTINGENCE</p>

Exercice 17 Reprenez les données de l'exercice 1 (décomposition de la population active).

1. Calculez les fréquences partielles f_{ij} . Interprétez au moins une valeur.
2. Calculez les fréquences marginales de la variable CSP (X) $f_{i\bullet}$ puis celles de la variable sexe (Y) $f_{\bullet j}$. Interprétez.
3. Calculez les fréquences conditionnelles de la CSP sachant le sexe $f_{i|j}$. Interprétez au moins une valeur. Comment représenter graphiquement ces distributions conditionnelles?
4. Calculez les fréquences conditionnelles du sexe sachant la CSP $f_{j|i}$.
5. Concluez sur l'existence ou non d'une liaison entre les deux caractères.

Exercice 18 L'exercice porte sur une enquête sur un échantillon de 1000 personnes à propos de l'impact de la publicité diffusée à la télévision sur l'achat d'un produit à la télévision. Complétez la table de contingence suivante en fonction des indications ci-après. Peut-on dire qu'il y a un impact de la publicité sur l'achat du produit.

Achat Publicité	Oui	Non	Total
Oui	n_{11}	n_{12}	$n_{1\bullet}$
Non	n_{21}	n_{22}	$n_{2\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	1000

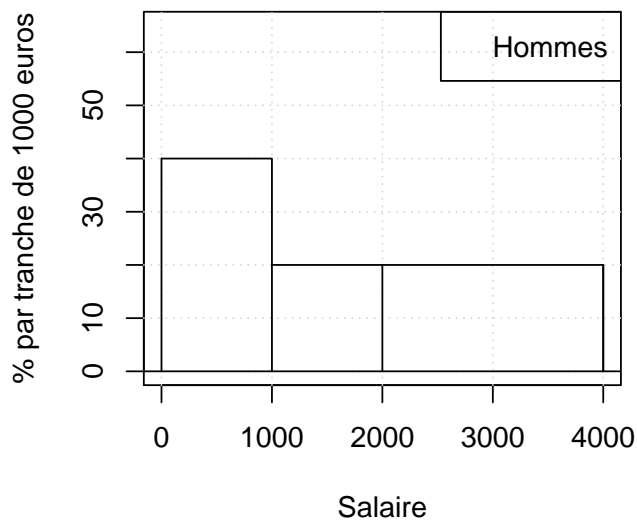
1. 25% des personnes interrogées ont acheté le produit.
2. Parmi les gens ayant vu la publicité 20% d'entre eux ont acheté le produit.
3. Parmi les gens ayant acheté le produit 60% n'ont pas vu la publicité.

Exercice 19 Complétez les tables de contingence suivantes sachant que les deux variables X et Y sont indépendantes.

X Y	0	20	30	40	$n_{i\bullet}$
100					40
200					60
$n_{\bullet j}$	10	30	40	20	100

X Y	10	20	0	$n_{i\bullet}$
60	4	6	1	
70		12		
80			5	
$n_{\bullet j}$				

Exercice 20 Dans une entreprise comprenant 300 employés dont 100 femmes, on a représenté les distributions conditionnelles du salaire selon le sexe. A partir des deux graphiques que vous commencerez par comparer, complétez la table de contingence correspondante (attention: le salaire est une variable continue!).



<i>Salaire</i> <i>Sexe</i>	<i>Hommes</i>	<i>Femmes</i>	<i>Ensemble</i>
[0, 1000[
[1000, 2000[
[2000, 4000[
<i>Total</i>		100	300

Prénom, Nom et groupe de TD:

QCM RELATIF AU THÈME N.5

LAISON ENTRE DEUX VARIABLES
(χ^2 , RAPPORT DE CORRÉLATION)

Attention: Plusieurs réponses sont parfois possibles.

Question 1: Le χ^2 et le coefficient de Cramèr peuvent se calculer lorsqu'on dispose d'une table de contingence associée à deux variables X et Y

de même nature. nécessairement qualitatives. nécessairement quantitatives. de nature indifférente.

Question 2: Soit n_{ij} et n'_{ij} respectivement les effectifs et les effectifs sous hypothèse d'indépendance de la table de contingence. La formule du χ^2 est

$\sum_{i,j} \frac{(n_{ij}-n'_{ij})^2}{n_{ij}}$ $\sum_{i,j} \frac{(n_{ij}-n'_{ij})^2}{n'_{ij}}$ $\sum_{i,j} \left(\frac{n_{ij}-n'_{ij}}{n_{ij}}\right)^2$ $\sum_{i,j} \left(\frac{n_{ij}-n'_{ij}}{n'_{ij}}\right)^2$

Question 3: Si les variables X et Y sont indépendantes

le $\chi^2 = 0$ le χ^2 vaut 1. le coefficient de Cramèr vaut 0. $n_{ij} = n'_{ij}$ pour toute case.

Question : Si X (resp. Y) est totalement dépendante de Y (resp. X) alors

le $\chi^2 = \chi^2_{\max}$ le χ^2_{\max} vaut 1. le coefficient de Cramèr vaut 100%. le χ^2 vaut 1.

Question 4: Plus le χ^2 est élevé et

plus les variables sont indépendantes. plus la variance est grande.
 plus les variables sont dépendantes l'une de l'autre. plus le coefficient de corrélation est élevé.

Question 5: Le tableau suivant représente le croisement de deux variables réponses de 200 étudiants aux questions "avez-vous travaillé régulièrement au cours du semestre en statistiques" et "avez-vous réussi l'examen?" Complétez la table de contingence puis cochez les affirmations vraies

Travail	Réussite	Oui	Non	Total
	Oui	60	10	
	Non	30	100	
	Total			200

$n_{11} = 31.5$ $n'_{11} = 31.5$ $n'_{12} = 38.5$ $n'_{21} = 38.5$

$\chi^2 = 0$. $\chi^2 \simeq 72.1$ $\chi^2 \simeq 1.8$ $\chi^2 \simeq 130$

$\chi^2_{\max} = 0$. $\chi^2_{\max} = 100$. $\chi^2_{\max} = 200$. $\chi^2_{\max} = 400$.

$C \simeq 72.1\%$. $C \simeq 84.3\%$. $C \simeq 36.05\%$. $C \simeq 59.6\%$.

Question 6: La variance inter-population est

- la variance des moyennes conditionnelles. la moyenne des variances conditionnelles.
- toujours inférieure à la variance totale. parfois supérieure à la variance totale.

Question 7: La variance intra-population est

- la variance des moyennes conditionnelles. la moyenne des variances conditionnelles.
- toujours inférieure à la variance totale. parfois supérieure à la variance totale.

Question 8: Le rapport de corrélation de X en Y peut se calculer

- lorsque les deux variables sont de même nature.
- uniquement lorsque les deux variables sont de nature différente.
- lorsque la variable X est quantitative.
- lorsque la variable Y est quantitative.

Question 9: Plus le rapport de corrélation de X en Y est

- faible et plus les variables X et Y sont indépendantes.
- faible et plus les variables X et Y sont dépendantes.
- élevée et plus la part de variance expliquée par Y est élevée.
- plus les moyennes conditionnelles de X sachant Y forment une série dispersée.

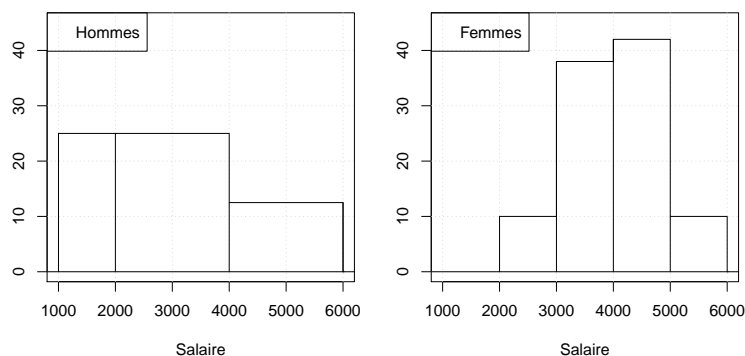
Question 10: Soient X la variable salaire et Y la variable sexe. Supposons que sur la population étudiée, la moyenne des salaires des hommes et femmes soient les mêmes. Alors

- les variables X et Y sont nécessairement indépendantes. la variance inter-population est nulle.
- la variance intra-population est nulle. le rapport de corrélation de X en Y est nul.

Question 11: Dans une population contenant autant d'hommes que de femmes, on a observé $\bar{x}_H = 2000, \bar{x}_F = 1000, \sigma_H = \sigma_F = 500$ (euros).

- $Var(x) = 250000$ (euros²) Var. Inter=250000 (euros²)
- Var. Intra. = 250000 (euros²) $\eta_{X|Y}^2 = 90\%$

Question 12: Le graphique suivant représente la distribution des salaires mensuels d'une entreprise comprenant à peu près autant d'hommes que de femmes. Le rapport de corrélation est



- égal à 0. proche de 0. proche de 1. égal à 1.

EXERCICES RELATIFS AU THÈME N.5

LAISON ENTRE DEUX VARIABLES (χ^2 , RAPPORT DE CORRÉLATION)

Exercice 21 Le tableau suivant présente une étude sur le nombre de nuitées en hôtel pour des français et étrangers en fonction du nombre d'étoiles de l'hôtel. La question sous-jacente est l'existence d'une distinction français-étrangers au niveau des concepts d'hôtellerie de tourisme en France.

Nbre d'étoiles	Origine	Français	Etrangers	Total
0 étoile		42	7	49
1 étoile		12	1	13
2 étoiles		93	28	121
3 étoiles		47	36	83
4 étoiles et plus		13	19	32
Total		207	91	298

- Calculer les distributions conditionnelles du nombre d'étoiles pour les touristes français et étrangers. Vous semble-t-il qu'il y ait une dépendance importante entre la qualité de l'hôtel choisi et l'origine des touristes?
- Calculez les effectifs théoriques sous hypothèse d'indépendance, le χ^2 , le χ^2_{\max} ainsi que le coefficient de Cramèr C . Poursuivez l'analyse de la précédente question.
- Pour essayer de comprendre quelle couple de modalités apporte la plus grande contribution, calculez les contributions de chaque case au χ^2 , c'est-à-dire $\frac{(n_{ij} - n'_{ij})^2 / n'_{ij}}{\chi^2}$ à exprimer en pourcentage. Interprétez.

Exercice 22 Le tableau suivant présente une enquête sur l'emploi en 2011 (dans un pays fictif) où l'on met en relation le type de contrat de travail et l'âge des travailleurs. Pour simplifier les données sont exprimées en milliers d'individus et approchées au millier près.

Type contrat	Âge	[15, 24[[25, 49[[50, 65[Total
Temps complet		200	14800	6200	21200
Temps partiel		2100	1800	500	4400
Total		2300	16600	6700	25600

- Calculez les distributions conditionnelles de l'âge des travailleurs selon le type de contrat. Ces deux variables sont-elles indépendantes?
- Complétez le tableau suivant en calculant les moyennes et écart-types conditionnels des âges des chômeurs selon le type de contrat, ainsi que les moyenne et écart-type de l'ensemble de la population.

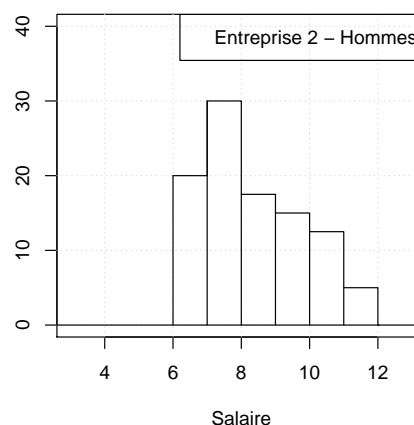
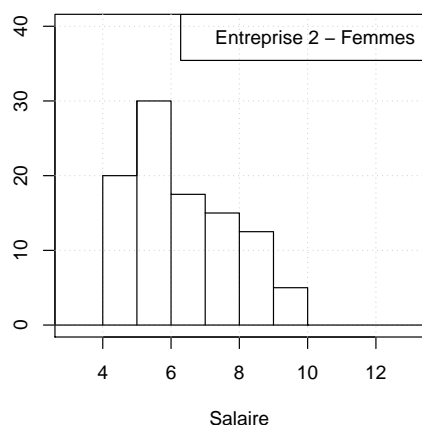
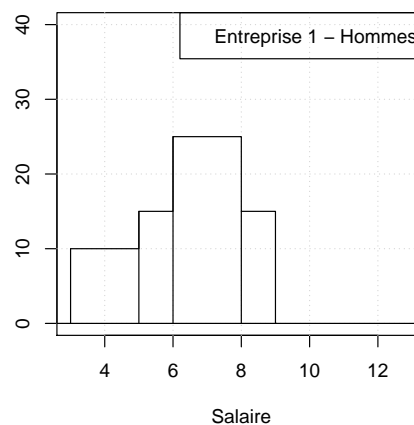
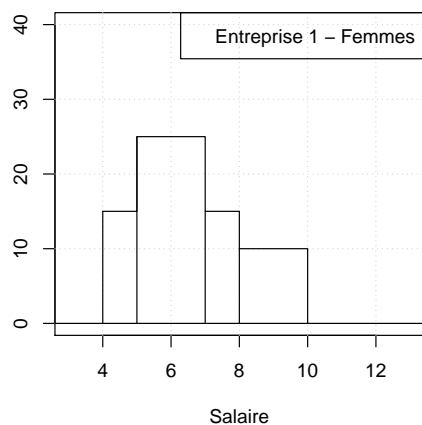
	Temps complet	Temps partiel	Ensemble
Effectif	21200	4400	25600
Moyenne			
Ecart-type			

3. A l'aide du tableau précédent, calculez les variances inter-population et intra-population. En déduire la part de variance de l'âge des travailleurs expliquée par le type de contrats. Interprétez.
4. Les précédentes données étaient fictives. Les données réelles ainsi que le tableau résumé sont maintenant présentées. Recalculez le rapport de corrélation de l'âge en le type de contrat (surtout en utilisant le second tableau), comparez avec les précédents résultats et interprétez.

Type contrat Âge	[15, 24[[25, 49[[50, 65[Total
Temps complet	1722	13989	5445	21156
Temps partiel	497	2704	1421	4622
Total	2219	16693	6866	25778

	Temps complet	Temps partiel	Ensemble
Effectif	21156	4622	25778
Moyenne	40.9	41.4	41
Ecart-type	10.9	11.9	11.1

Exercice 23 Une population est décrite à l'aide de deux variables: X le salaire (unité indifférente) et Y le sexe. Dans chacune des deux entreprises, indiquez si le rapport de corrélation vous paraît un indicateur adéquat pour caractériser cette liaison.



Exercice 24 Traitez les questions 4 à 6 de la partie I de l'examen de 2012 ainsi que la question 4 de l'examen de 2013.

Prénom, Nom et groupe de TD:

QCM RELATIF AU THÈME N.6
LIAISON FONCTIONNELLE ENTRE DEUX VARIABLES;
RÉGRESSION; CORRÉLATION LINÉAIRE

Attention: Plusieurs réponses sont parfois possibles.

Question 1: Le coefficient de corrélation linéaire entre deux variables x et y mesure

- le niveau de dépendance entre x et y . le niveau d'indépendance entre x et y .
 la part de variance expliquée par un modèle linéaire. le niveau de dépendance linéaire entre x et y .

Question 2: Le coefficient de corrélation linéaire est

- d'autant plus faible que les variables sont indépendantes.
 proche de 1 ou -1 si le modèle linéaire est très adapté.
 nécessairement positif.
 nécessairement plus grand que 50%.

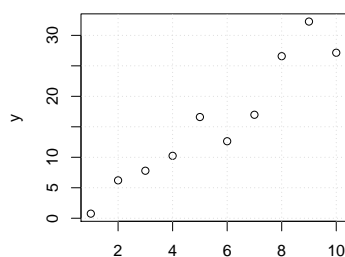
Question 3: Si le coefficient de corrélation linéaire entre deux variables x et y est nul

- les deux droites de régression ont une pente nulle.
 l'une des deux droites a une pente nulle.
 les variables sont nécessairement indépendantes.
 peuvent tout de même être liées fonctionnellement.

Question 4: Le coefficient de détermination R^2 entre deux variables x et y vaut

- $\hat{a} \times \hat{a}'$. $\frac{\hat{a}}{\hat{a}'}$. $\frac{Cov(x,y)}{\sigma_x \sigma_y}$. $\frac{Cov(x,y)^2}{Var(x)Var(y)}$.

Question 5: Au vu du graphique ci-dessous



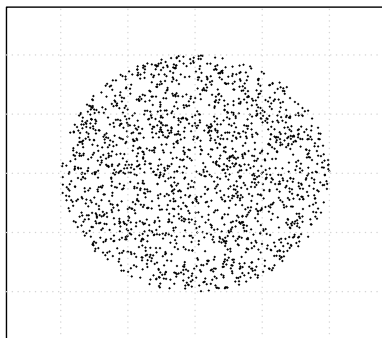
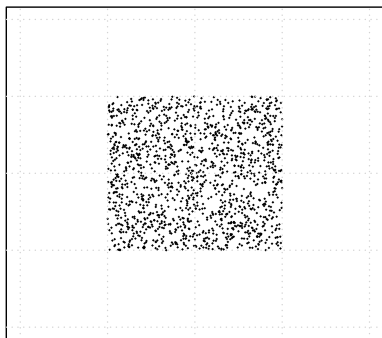
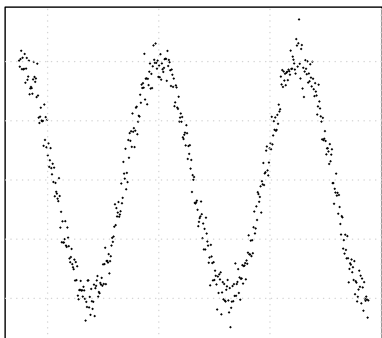
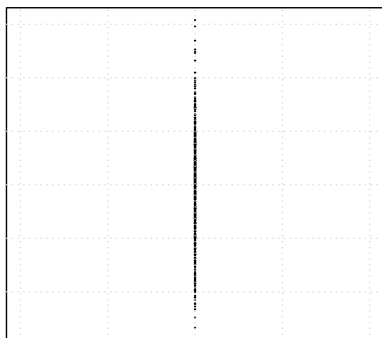
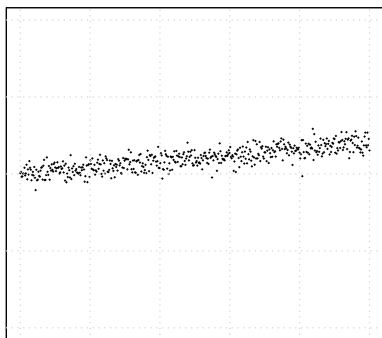
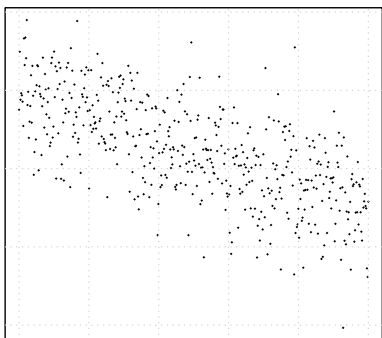
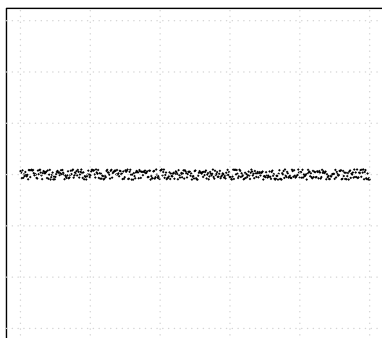
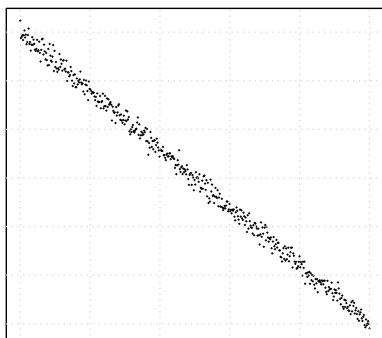
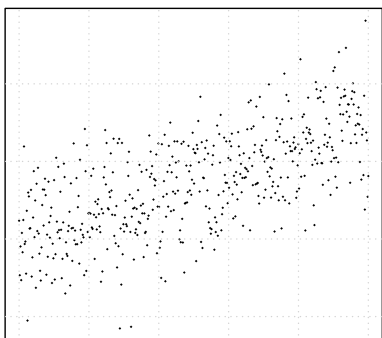
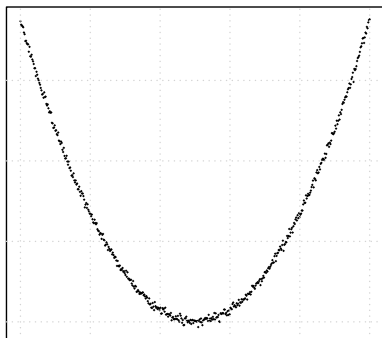
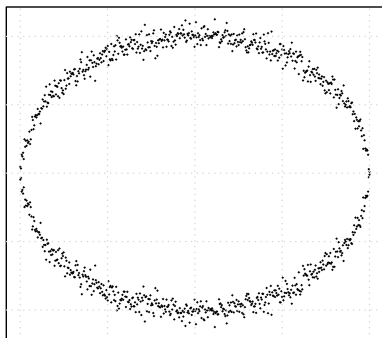
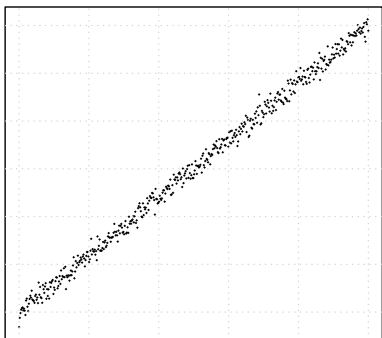
- $R < 0$ $R > 0$ $\hat{a}' > 0$ $|R| > 90\%$

Question 6: Les données associées au graphique précédent permettent de calculer $\bar{x} = 5.5, \bar{y} \simeq 18.75, \sigma_x \simeq 3, \sigma_y \simeq 6.8, Cov(x, y) \simeq 17.9$. On a alors

- $y \simeq -1.6x + 0.4$. $y \simeq 1.99x + 7.8$. $x \simeq -1.6y + 0.4$. $x \simeq 1.99y + 7.8$.

Question 7: Les 12 graphiques ci-dessous représentent 12 nuages de points (x_i, y_i) différents. Pour chacun de ces nuages, les coefficients de corrélation ont été évalués. Huit d'entre eux sont inférieurs

à 5%. Les autres valeurs sont: -99.7%, -64.5%, 68.9%, 82.5% et 99.8%. Associez à chaque graphique l'une des valeurs possibles (en indiquant " $< 5\%$ " lorsqu'elle est inférieure à 5%). Indiquez par un "A" si le modèle linéaire vous semble un modèle adéquat et/ou par un "L" s'il existe selon vous une liaison fonctionnelle entre x et y .

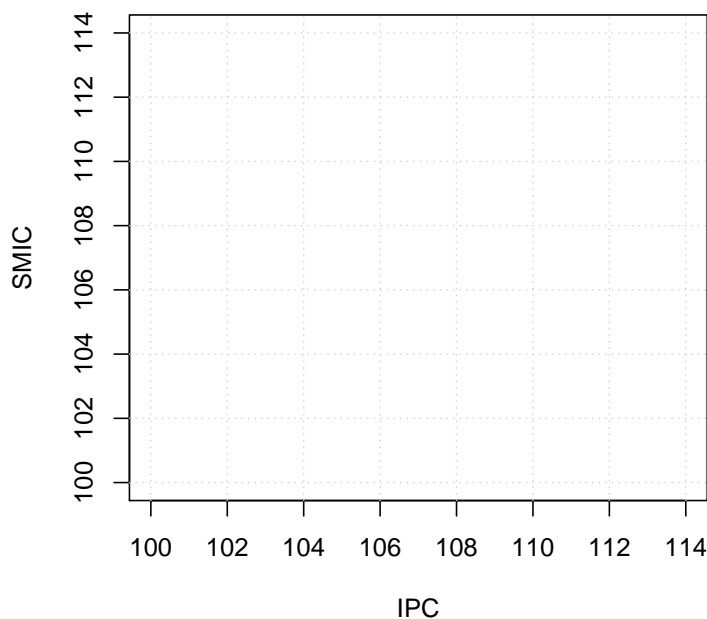


EXERCICES RELATIFS AU THÈME N.6
LIAISON FONCTIONELLE ENTRE DEUX VARIABLES;
RÉGRESSION; CORRÉLATION LINÉAIRE

Exercice 25 Le tableau suivant présente entre 2006 et 2010, l'évolution des indices de prix à la consommation (IPC) et de l'évolution des indices des smic (SMIC). Ces indices sont en base 100 en 2005.

Date	2006	2007	2008	2009	2010	2011	2012	2013
IPC	102.3	102.0	105.2	106.0	107.3	109.4	112.2	113.8
SMIC	101.6	103.6	104.9	108.2	108.2	109.8	112.4	114.1

1. Sur le graphique ci-dessous, représentez le nuage de points montrant la relation entre les variables IPC et SMIC. Quels sont vos commentaires?



2. Complétez le tableau suivant

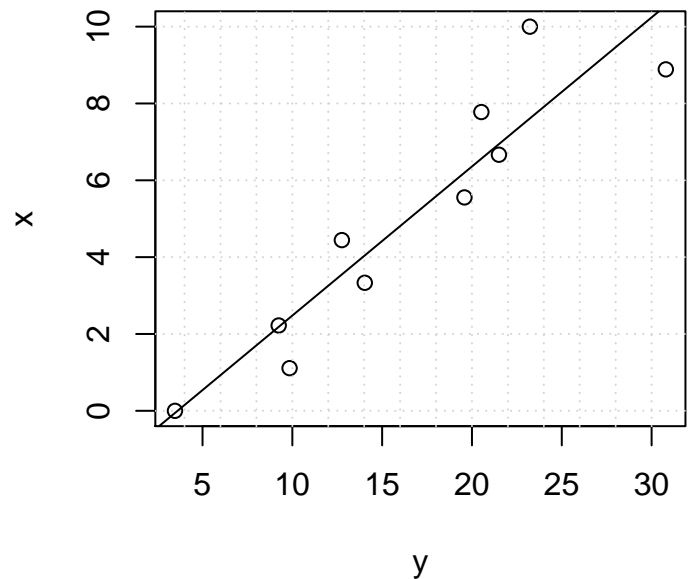
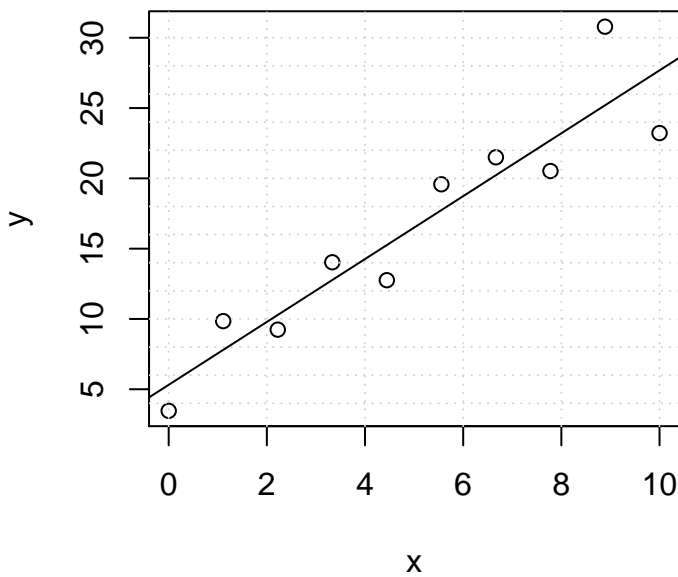
Caractéristique	\overline{IPC}	σ_{IPC}	\overline{SMIC}	σ_{SMIC}	$\overline{IPC \times SMIC}$	$Cov(IPC, SMIC)$
Valeur						

3. Calculez le coefficient de détermination linéaire en utilisant le tableau précédent ainsi qu'en utilisant les fonctions internes de la calculatrice.
4. Après l'avoir justifié, calculez la droite de régression de la variable SMIC (variable y) en fonction de la variable IPC (variable x) à partir du tableau précédent. Vérifiez en utilisant plus directement les fonctions internes de la calculatrice.

5. Rajoutez cette droite sur le graphique précédent.

6. Lorsque les prix auront augmenté de 20% par rapport en 2005, quelle évolution du SMIC pourra-t-on prévoir?

Exercice 26 Les deux figures ci-dessous représentent les nuages de points (x_i, y_i) et (y_i, x_i) ainsi que les droites de régression (estimées par moindres carrés). Uniquement à partir de ces graphiques, donnez une valeur approximative du coefficient de corrélation linéaire entre x et y . Sachant que $\bar{x} = 5$, évaluez approximativement \bar{y} .



(les vrais résultats sont $R \simeq 93.1\%$ et $\bar{y} \simeq 16.5$).

Exercice 27 Le tableau suivant fournit pour quelques années les évolutions des investissements (*Inv*) et des revenus (*Rev*) en millions d'euros d'une certaine multinationale.

<i>Inv</i>	360	480	500	530	580	620	570	640	720	760
<i>Rev</i>	3200	3400	3450	3600	3800	4000	4200	4400	4600	4800

1. A l'aide de la calculatrice, complétez le tableau suivant après l'avoir justifié.

	Modèle $Inv = a \times Rev + b$		Modèle $Rev = a' \times Inv + b'$	
R	\hat{a}	\hat{b}	\hat{a}'	\hat{b}'

2. Donnez une estimation des investissements si le revenu est de 5200 (millions d'euros) ainsi que du revenu nécessaire à des investissements de 600 (millions d'euros).

Prénom, Nom et groupe de TD:

<p>QCM RELATIF AU THÈME N.7</p> <p>SÉRIES CHRONOLOGIQUES</p>
--

Attention: Plusieurs réponses sont parfois possibles.

Question 1: Une série chronologique est

- une distribution statistique. les modalités d'une variable qualitative. une suite de valeurs quantitatives dépendant du temps. Autre.

Question 2: D'un point de vue général, une série chronologique est en général composée

- d'une tendance d'une composante saisonnière d'une tendance et d'une composante saisonnière (et d'autres effets qu'on espère faibles) Autre

Question 3: La série des moyennes mobiles d'ordre 4 à l'instant t a pour formule

- $\frac{y(t)+y(t+1)+y(t+3)+y(t+4)}{4}$ $\frac{0.5y(t-2)+y(t-1)+y(t)+y(t+1)+0.5y(t+2)}{4}$
 $\frac{1}{2} \left(\frac{y(t-2)+y(t-1)+y(t)+y(t+1)}{4} + \frac{y(t-1)+y(t)+y(t+1)+y(t+2)}{4} \right)$ $\frac{y(t-1)+y(t)+y(t+1)}{4}$

Question 4: L'ordre d'une moyenne mobile correspond (généralement) à

- 4 si les données sont observées mensuellement. 2 si les données sont observées trimestriellement.
 12 si les données sont observées mensuellement. 4 si les données sont observées trimestriellement.

Question 5: La tendance d'une série est généralement définie comme

- la régression linéaire de la série y en fonction du temps.
 la régression linéaire de la série des moyennes mobiles en fonction du temps.
 la série des moyennes mobiles.
 la régression linéaire de la série corrigée des variations saisonnières en fonction du temps.

Question 6: Pour estimer la tendance d'une série chronologique, il

- est indispensable de choisir un modèle (par exemple additif ou multiplicatif).
 est parfois nécessaire de choisir un modèle pour la série.
 n'est pas nécessaire de choisir un modèle pour la série.
 Autre.

Question 7: Pour estimer la fonction de saisonnalité puis les coefficients saisonniers d'une série chronologique, il

- est indispensable de choisir un modèle (par exemple additif ou multiplicatif).
 est parfois nécessaire de choisir un modèle pour la série.
 n'est pas nécessaire de choisir un modèle pour la série.
 Autre.

Question 8: La fonction de saisonnalité $S(t)$ d'un modèle additif se calcule par la formule

- $S(t) = y(t) + T(t)$ $S(t) = y(t) - T(t)$ $S(t) = y(t) \times T(t)$ $S(t) = y(t)/T(t)$

Question 9: La fonction de saisonnalité $S(t)$ d'un modèle multiplicatif se calcule par la formule

- $S(t) = y(t) + T(t)$ $S(t) = y(t) - T(t)$ $S(t) = y(t) \times T(t)$ $S(t) = y(t)/T(t)$

Question 10: Supposons que la série est observée trimestriellement. Les coefficients saisonniers S_j

- sont obtenus en faisant la somme des $S(t)$ par année.
 sont obtenus en faisant la moyenne des $S(t)$ trimestre par trimestre.
 sont obtenus en faisant la moyenne des $S(t)$ année par année.
 Autre.

Question 11: Toujours en utilisant les notations du cours, on a

- $S'_j = S_j/\overline{S_j}$ si le modèle est additif.
 $S'_j = S_j - \overline{S_j}$ si le modèle est additif.
 $S'_j = S_j/\overline{S_j}$ si le modèle est multiplicatif.
 $S'_j = S_j - \overline{S_j}$ si le modèle est additif.

Question 12: La série C.V.S. corrigée des variations saisonnières permet

- de comparer la tendance et l'effet saisonnier.
 de pouvoir interpréter les valeurs de la série chronologique indépendamment de l'effet saisonnier.
 de pouvoir calculer la fonction de tendance.
 de pouvoir décider si l'on choisit un modèle additif ou multiplicatif.

Question 13: La série C.V.S. notée $y^*(t)$ se calcule à tout instant par la formule (en abusant un peu des notations)

- $y^*(t) = y(t)/S'_j$ si le modèle est additif.
 $y^*(t) = y(t) - S'_j$ si le modèle est additif.
 $y^*(t) = y(t)/S'_j$ si le modèle est multiplicatif.
 $y^*(t) = y(t) \times S'_j$ si le modèle est additif.

Question 14: Supposons avoir observé une série chronologique trimestriellement pendant 3 ans (de 2010 à 2012), avoir calculé la tendance par régression linéaire de la série des moyennes mobiles d'ordre 4 et les coefficients saisonniers en supposant un modèle additif. Soient \hat{a} et \hat{b} les estimations des paramètres de la régression définissant $T(t)$.

- $\overline{S_j} = 0$ $\overline{S_j} = 1$ $\overline{S'_j} = 0$ $\overline{S'_j} = 1$
 $\hat{y}(14) = (14\hat{a} + \hat{b}) + S'_1$ $\hat{y}(14) = (14\hat{a} + \hat{b}) \times S'_1$
 $\hat{y}(14) = (14\hat{a} + \hat{b}) + S'_2$ $\hat{y}(14) = (4\hat{a} + \hat{b}) + S'_2$
 $y^*(16) = y(16)/S'_4$ $y^*(16) = y(16) - S'_4$ $y^*(16) = y(16)/S'_2$ $y^*(16) = y(16) - S'_2$

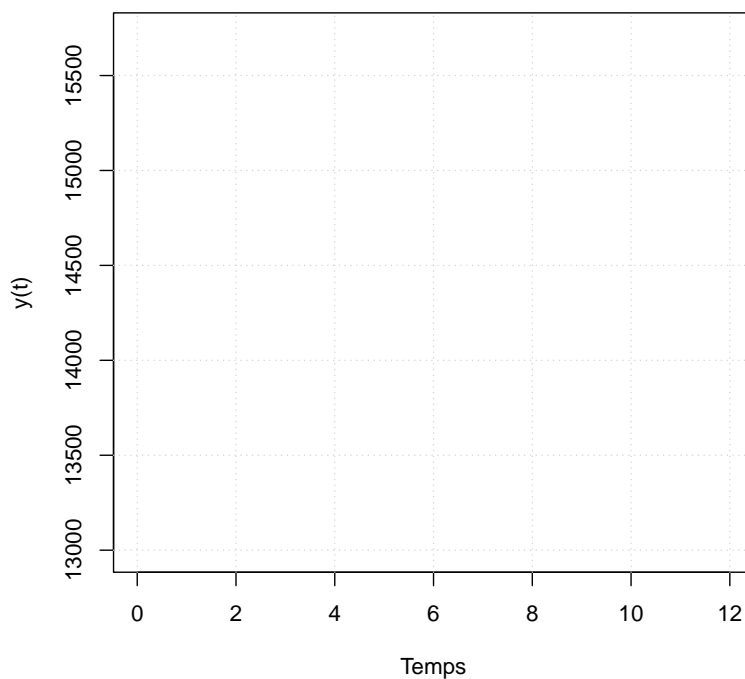
EXERCICES RELATIFS AU THÈME N.7

SÉRIES CHRONOLOGIQUES

Exercice 28 L'exercice a pour objectif l'étude du chiffre d'affaires trimestriel (en milliers d'euros) de l'entreprise Peugeot présenté dans le tableau ci-dessous.

Trimestre	Année	2010	2011
I		13986	15414
II		14408	15721
III		12993	13450
IV		14674	15237

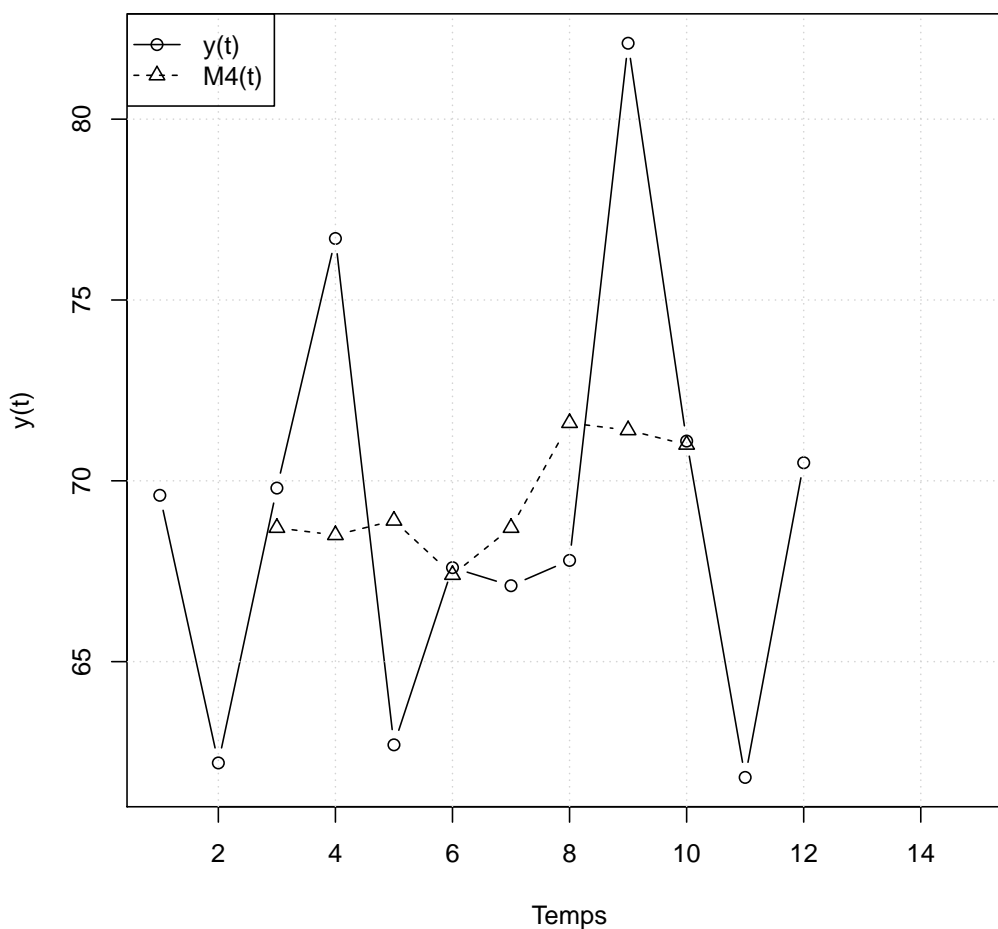
1. Représentez cette série chronologique sur le graphique ci-dessous. On adoptera un modèle additif pour modéliser cette série chronologique. Cela vous semble-t-il cohérent?



Date	Temps	$y(t)$	$M_4(t)$	$T(t)$	$S(t)$	S_j	S'_j
2010,I	1	13986	×				
II	2	14408	×				
III	3	12993	14193.8				
IV	4	14674	14536.6				
2011,I	5	15414	14757.6				
II	6	15721	??				
III	7	13450	×				
IV	8	15237	×				

2. Complétez la 3ème colonne du tableau précédent en calculant la moyenne mobile d'ordre 4 (ordre que vous justifierez) manquante. Représentez la série des moyennes mobiles sur le graphique.
3. Calculez la tendance $T(t)$ par un modèle de régression linéaire (que vous justifierez en calculant le coefficient de corrélation) de la série M_4 en fonction du temps. Représentez la droite de régression sur le graphique.
4. Complétez les trois dernières colonnes du tableau visant à calculer les coefficients saisonniers.
5. Quel chiffre d'affaires l'entreprise Peugeot pouvait-elle espérer les trois premiers trimestres de 2012? Représentez ces valeurs sur le graphique ainsi que les vraies valeurs observées 14289, 15264 et 12931 (milliers d'euros). Quels sont vos commentaires?

Exercice 29 On s'intéresse dans cet exercice à la fréquentation trimestrielle des salles de cinéma sur la période 2010-2012. Les données sont exprimées en milliers de spectateurs. Les tableau et graphique suivants présentent les données ainsi que la série des moyennes mobiles.



Date	Temps	$y(t)$	$M_4(t)$	$T(t)$	$S(t)$	S_j	S'_j	$y^*(t)$
2010,I	1	69.6	×					
II	2	62.2	×					
III	3	69.8	??					
IV	4	76.7	68.5					
2011,I	5	62.7	68.9					
II	6	67.6	67.4					
III	7	67.1	68.7					
IV	8	67.8	71.6					
2012,I	9	82.1	71.4					
II	10	71.1	71.0					
III	11	61.8	×					
IV	12	70.5	×					

1. Suivant la méthodologie de l'exercice précédent, complétez le tableau en supposant que la série chronologique étudiée peut être modélisée par un modèle multiplicatif (modèle que vous prendrez le soin de justifier). Prenez deux décimales pour le calcul des coefficients saisonniers. La dernière colonne correspond à la série C.V.S. (corrigée des variations saisonnières). Représentez les tendances et série C.V.S. sur le graphique.
2. Estimer le nombre de spectateurs ainsi que leur version C.V.S. des trois premiers trimestres de 2013. Reportez ces valeurs sur le graphique (le nombre d'entrées en millions de spectateurs a été évaluée par le CNC à 64.83 pour le premier trimestre 2013).

Exercice 30 Traitez les questions associées à la Partie II de l'examen de 2012 ainsi que les questions 5 (a)-(d) de l'examen de 2013.

Prénom, Nom et groupe de TD:

QCM RELATIF AU THÈME N.8
INDICES ÉLÉMENTAIRES ET INDICES SYNTHÉTIQUES

Attention: Plusieurs réponses sont parfois possibles.

Question 1: L'indice d'une valeur V à la date t base 100 à la date 0 se calcule par la formule

- $\frac{V_0}{V_t}$. $\frac{V_t}{V_0}$. $\frac{V_0}{V_t} \times 100$. $\frac{V_t}{V_0} \times 100$.

Question 2: L'indice du prix d'un produit en 2013 base 100 en 2000 vaut 135. Cela signifie que

- le prix a augmenté de 135% de 2000 à 2013. le prix a augmenté de 35% de 2000 à 2013.
 le prix a augmenté de 135% de 2013 à 2010. le prix a augmenté de 35% de 2013 à 2000.

Question 3: Le C.A. d'une entreprise a augmenté de 20% de 2010 à 2011 puis diminué de 16.67% de 2011 à 2012. L'indice du C.A. en 2012 base 100 en 2010 vaut

- 66.67. 100. 140. 333.4.

Question 4: Le prix d'un produit a augmenté de 20% de 2010 à 2012 tandis que les quantités ont diminué de 20% sur la même période. Les recettes de ce produit sur cette période

- n'ont pas évolué. ont augmenté de 4%. ont diminué de 4%. Autre.

Questions 5 à 11: Imaginons disposer d'un panier de produits dont on connaît les prix et quantités à deux dates 0 et t .

Question 5: Les indices synthétiques de Laspeyres, Paasche et Fisher permettent

- de déterminer l'influence des prix ou des quantités dans l'évolution globale des recettes.
 de déterminer l'influence de chaque produit.
 de pouvoir comprendre l'évolution des prix (ou quantités) à quantités fixées (resp. à prix fixés) à une certaine date.
 Autre.

Question 6: L'indice de Laspeyres des prix à la date t base 100 à la date 0 se calcule par la formule

- $L_{t/0}^P = \frac{\sum_i P_i(t)Q_i(t)}{\sum_i P_i(t)Q_i(0)} \times 100$ $L_{t/0}^P = \frac{\sum_i P_i(t)Q_i(0)}{\sum_i P_i(0)Q_i(0)} \times 100$ $L_{t/0}^P = \frac{\sum_i P_i(t)Q_i(t)}{\sum_i P_i(0)Q_i(t)} \times 100$
 $L_{t/0}^P = \frac{\sum_i P_i(0)Q_i(t)}{\sum_i P_i(0)Q_i(0)} \times 100$

Question 7: L'indice de Laspeyres des quantités à la date t base 100 à la date 0 se calcule par la formule

- $L_{t/0}^Q = \frac{\sum_i P_i(t)Q_i(t)}{\sum_i P_i(t)Q_i(0)} \times 100$ $L_{t/0}^Q = \frac{\sum_i P_i(t)Q_i(0)}{\sum_i P_i(0)Q_i(0)} \times 100$ $L_{t/0}^Q = \frac{\sum_i P_i(t)Q_i(t)}{\sum_i P_i(0)Q_i(t)} \times 100$
 $L_{t/0}^Q = \frac{\sum_i P_i(0)Q_i(t)}{\sum_i P_i(0)Q_i(0)} \times 100$

Question 8: L'indice de Paasche des prix à la date t base 100 à la date 0 se calcule par la formule

- $P_{t/0}^P = \frac{\sum_i P_i(t)Q_i(t)}{\sum_i P_i(t)Q_i(0)} \times 100$ $P_{t/0}^P = \frac{\sum_i P_i(t)Q_i(0)}{\sum_i P_i(0)Q_i(0)} \times 100$ $P_{t/0}^P = \frac{\sum_i P_i(t)Q_i(t)}{\sum_i P_i(0)Q_i(t)} \times 100$
 $P_{t/0}^P = \frac{\sum_i P_i(0)Q_i(t)}{\sum_i P_i(0)Q_i(0)} \times 100$

Question 9: L'indice de Paasche des quantités à la date t base 100 à la date 0 se calcule par la formule

$$\begin{aligned} \square P_{t/0}^Q &= \frac{\sum_i P_i(t)Q_i(t)}{\sum_i P_i(t)Q_i(0)} \times 100 & \square P_{t/0}^Q &= \frac{\sum_i P_i(t)Q_i(0)}{\sum_i P_i(0)Q_i(0)} \times 100 & \square P_{t/0}^Q &= \frac{\sum_i P_i(t)Q_i(t)}{\sum_i P_i(0)Q_i(t)} \times 100 \\ \square P_{t/0}^Q &= \frac{\sum_i P_i(0)Q_i(t)}{\sum_i P_i(0)Q_i(0)} \times 100 \end{aligned}$$

Question 10: L'indice de Fisher des prix à la date t base 100 à la date 0 se détermine par

$$\begin{aligned} \square F_{t/0}^P &= \frac{1}{2}(L_{t/0}^P + P_{t/0}^P). & \square & \text{la moyenne géométrique des indices de Laspeyres et Paasche des prix.} \\ \square F_{t/0}^P &= \sqrt{L_{t/0}^P \times P_{t/0}^P}. & \square & \text{la moyenne arithmétique des indices de Laspeyres et Paasche des prix.} \end{aligned}$$

Question 11: Pour calculer l'indice des volumes globaux (ici recettes) on utilise la formule

$$\begin{aligned} \square I_{t/0}^V &= \frac{\sum_i P_i(t)Q_i(t)}{\sum_i P_i(0)Q_i(0)} \times 100 & \square I_{t/0}^V &= \frac{\sum_i P_i(t)Q_i(0)}{\sum_i P_i(0)Q_i(0)} \times 100 & \square I_{t/0}^V &= \frac{\sum_i P_i(t)Q_i(t)}{\sum_i P_i(0)Q_i(t)} \times 100 \\ \square I_{t/0}^V &= \frac{\sum_i P_i(0)Q_i(t)}{\sum_i P_i(0)Q_i(0)} \times 100 \end{aligned}$$

ou encore

$$\square I_{t/0}^V = L_{t/0}^P \times L_{t/0}^Q \times \frac{1}{100} \quad \square I_{t/0}^V = L_{t/0}^P \times P_{t/0}^Q \times \frac{1}{100} \quad \square I_{t/0}^V = L_{t/0}^Q \times P_{t/0}^Q \times \frac{1}{100} \quad \square I_{t/0}^V = F_{t/0}^P \times F_{t/0}^Q \times \frac{1}{100}$$

Questions 12 à 15: L'évolution des prix et des quantités de 4 produits entre 2010 et 2011 a donné les résultats suivants

Indice	Variable	Prix	Quantités
	$L_{t/0}^\bullet$	120	80
	$P_{t/0}^\bullet$	135	90

Question 12: L'indice des volumes globaux vaut

$$\square 100 \quad \square 107.5 \quad \square 108 \quad \square 85$$

Question 13: Les indices de Fisher des indices des prix et quantités valent

$$\square F_{t/0}^P = 127.5, F_{t/0}^Q = 85. \quad \square F_{t/0}^P = 127.3, F_{t/0}^Q = 84.9. \quad \square F_{t/0}^P = 100, F_{t/0}^Q = 107.5. \quad \square \text{Autre.}$$

Question 14: Globalement, on note

- une augmentation des prix et une augmentation des quantités. une augmentation des prix et une diminution des quantités.
 une diminution des prix et une augmentation des quantités. une diminution des prix et une diminution des quantités.

Question 15: A propos de l'interprétation de l'évolution des recettes:

- L'augmentation de 8% des recettes est due à une augmentation des prix plus importante que la diminution des quantités.
 L'augmentation de 8% des recettes est due à une diminution des prix plus importante que l'augmentation des quantités.
 La diminution de 8% des recettes est due à une augmentation des prix plus importante que la diminution des quantités.
 La diminution de 8% des recettes est due à une diminution des prix plus importante que l'augmentation des quantités.

<p>EXERCICES RELATIFS AU THÈME N.8</p> <p>INDICES ÉLÉMENTAIRES ET INDICES SYNTHÉTIQUES</p>
--

Exercice 31 On s'intéresse ici à l'évolution de la fréquentation des salles de cinéma français et le prix du billet de cinéma.

1. Les recettes en millions d'euros sont passées de 824 à 1305.6 de 2000 à 2012. Quelle est l'évolution des recettes au guichet de 2000 à 2012?
2. La fréquentation des salles de cinéma a augmenté de 36% de 1990 à 2000 et de 66.9% de 1990 à 2012. Quelle a été l'évolution de la fréquentation de 2000 à 2012?
3. En déduire l'évolution du prix (moyen) du billet de 2000 à 2012 puis le prix du billet moyen en 2012 sachant qu'en 2000 il valait 5.39 euros.
4. Évaluez les recettes en 2012 et l'évolution des recettes de 2010 à 2012 au prix moyen du billet de 2000.
5. Enfin, le nombre de séances de cinéma a diminué de 26.5% de 2012 à 2000. Quelle est l'évolution de ce nombre de séances de 2000 à 2012?

Exercice 32 Continuons dans le cinéma! Le tableau suivant présente en 2000 et 2012 le nombre de films en exploitation par type (français, américains, européens, autres) ainsi que le nombre moyen d'entrées par film en milliers d'entrées.

Type	2000		2012	
	Nbre de films	Entrées moy. par film	Nbre de films	Entrées moy. par film
Films français	1690	27.9	2669	30.8
Films américains	1377	75	1658	52.4
Films européens	897	11.8	1332	20.3
Autres films	409	11.5	873	8.6

Pour simplifier, notez $F_i(t)$ la variable nombre de films et $E_i(t)$ la variable nombre moyen d'entrées par film pour $i = 1, \dots, 4$ (car 4 type de films) à l'instant t (2000 ou 2012). Enfin la variable $G_i(t)$ désignera le nombre total d'entrées par type de films à l'instant t .

1. Calculez l'indice des volumes (nombres totaux d'entrées) en 2012 base 100 en 2000.
2. Calculez les indices de Laspeyres du nombre de films puis du nombre moyen d'entrées par film en 2012 base 100 en 2000.
3. En déduire les indices de Paasche du nombre de films puis du nombre moyen d'entrées par film en 2012 base 100 en 2000 puis ceux de Fisher.
4. Commentez.

Exercice 33 Le directeur commercial d'une chaîne cherche à apprécier l'évolution de ses ventes annuelles de moulins à café. Pour limiter les calculs, on ne retiendra ici que trois modèles de moulin à café. Les prix sont ici en euros et les quantités en milliers.

Modèle	2011		2012	
	Prix	Quantités	Prix	Quantités
Modèle A	25	146	45	210
Modèle B	20	220	42.5	180
Modèle C	50	39	79	100

Calculez l'indice des volumes globaux, complétez le tableau (les notations sont celles du cours) ci-dessous et commentez

	Prix	Quantités
$L_{12/11}^\bullet$		
$P_{12/11}^\bullet$		
$F_{12/11}^\bullet$		

Exercice 34 Traitez la partie III de l'examen de 2012 et les questions 6 (a)-(b) de l'examen de 2013.

ANNEXE: MINI-GUIDE CALCULATRICE

Petit guide de l'utilisation de la calculatrice pour les statistiques descriptives

Jean-François Coeurjolly & Alain Sombardier

September 28, 2011

L'objectif de cette petite note consiste à vous assister dans l'utilisation de la calculatrice pour les cours, travaux dirigés et examens. Il est impossible d'être exhaustif sur l'ensemble des calculatrices disponibles sur le marché. Nous considérons ici deux modèles particuliers assez courants. D'autres modèles ou versions de ces modèles fonctionnent à peu près dans le même esprit. Et nous vous invitons de toute façon à consulter votre manuel d'utilisation.

Pour illustrer cette documentation, on considèrera la série statistique correspondant à l'exercice 2 du TD1 suivante:

Modalité (x_i)	0	1	2	3	4	5
Effectif n_i	12	31	40	11	4	2

Voici selon le modèle, les étapes à suivre pour obtenir des caractéristiques d'une série statistique unidimensionnelle.

1 Modèle Casio (graph 35+)

- A l'allumage de Menu STAT
- Vous arrivez au tableau statistique. Rentrez le tableau statistique correspondant à l'exemple en mettant en L1 les modalités et en L2 les effectifs.
- Onglet CALC du menu STAT.
- Avant de demander des caractéristiques univariées avec l'onglet 1-VAR, il est ESSENTIEL de bien préciser la manière dont ont été rentrées les données. Ceci se fait dans l'onglet SETUP.
- Vous pouvez revenir au menu supérieur et demandez des caractéristiques univariées 1-VAR. Vous pouvez alors vérifier entre autres que la moyenne notée \bar{x} fait 1.7 et la $\sum x_i$ notée Σx sur la calculatrice fait 170.

Remarques importantes:

- si on avait rentré en L1 les effectifs (ou fréquence) et en L2 les modalités, il suffit de fixer 1-VAR XList à L2 et 1-VAR Freq à L1.

- Comment effectuer des statistiques descriptives lorsque chaque modalité n'est observée qu'une seule fois (ce qui est le cas lorsque l'on observe des variables continues par exemple qui ne sont pas encore regroupées en classe)? Deux solutions: la première pas très jolie consiste à mettre les modalités en L1 par exemple et des 1 en L2; la seconde solution consiste dans l'onglet SETUP à fixer 1-VAR Freq à 1. A titre de vérification si vous faites cela pour l'exemple considéré la moyenne devient 2.5 (ce qui est bien la moyenne de la série 0,1,2,3,4,5 pour laquelle chaque modalité ne serait observée qu'une seule fois.).

2 Modèle Texas TI 82 Stats_fr

- A l'allumage de Menu STAT
- Onglet EDIT puis 1:Edite
- Vous arrivez au tableau statistique. Rentrez le tableau statistique correspondant à l'exemple en mettant en L1 les modalités et en L2 les effectifs.
- Onglet CALC puis Stats 1 Var
- Pour obtenir des résumés statistiques unidimensionnels, il suffit alors de taper 1-VAR L1,L2. Vous pouvez alors vérifier entre autres que la moyenne notée \bar{x} fait 1.7 et la $\sum x_i$ notée Σx sur la calculatrice fait 170.

Remarques importantes:

- si on avait rentré en L1 les effectifs (ou fréquence) et en L2 les modalités, il suffit de taper 1-VAR L2,L1.
- Comment effectuer des statistiques descriptives lorsque chaque modalité n'est observée qu'une seule fois (ce qui est le cas lorsque l'on observe des variables continues par exemple qui ne sont pas encore regroupées en classe)? Si les modalités sont rentrées en L1 par exemple, il suffit de taper 1-VAR L1. Autrement dit par défaut la liste des effectifs est fixée à 1 pour toutes les modalités. A titre de vérification si vous faites cela pour l'exemple considéré la moyenne devient 2.5 (ce qui est bien la moyenne de la série 0,1,2,3,4,5 pour laquelle chaque modalité ne serait observée qu'une seule fois.).

3 Compléments et précisions communs aux deux modèles

- les calculs de quantiles ne correspondent pas (sauf exemples particuliers) aux formules utilisées dans ce cours (formules basées sur l'interpolation linéaire).
- Une fois que vous avez compris comment on éditait les listes et comment on paramétrait les listes à utiliser (SETUP), il est aisé d'utiliser la calculatrice pour effectuer des statistiques descriptives bi-dimensionnelles (2-VAR) ainsi que des régression linéaires.
- Les formules d'écart-type (1-VAR et 2-VAR) sont à prendre avec précaution: deux écart-type sont en général fournis

– écart-type sur l'échantillon défini par

$$Sx = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

– écart-type sur la population défini par

$$\sigma x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

La subtilité entre ces deux quantités (qui ne diffèrent que d'un rapport de $\sqrt{(n-1)/n}$) sera développée en L2 dans le cours de Statistiques Inférentielles. L'écart-type tel qu'il est défini dans ce cours de statistiques descriptives correspond à l'**écart-type sur la population**.

Enfin cette documentation n'a absolument rien d'exhaustif sur les capacités de votre calculatrice à produire des représentations graphiques ou calculs statistiques plus avancés. Une lecture du manuel d'utilisation n'est pas à négliger!!

4 Remerciements

Merci à Céline Pougeol et Kamel Soualah (groupe B7) pour leur aide.

ANNEXE: ANNALES DES EXAMENS
DE STATISTIQUE DESCRIPTIVE DE
JANVIER 2012 ET 2013

Examen Statistique descriptive

1ère Session: janvier 2012

Durée: 3 heures

Tous documents interdits

Numéro d'étudiant:

Contexte général. Dans le pays Ecoland, le Ministère de l'Enseignement Supérieur s'interroge sur les résultats des étudiants inscrits en premier cycle. Pour cela il a demandé aux services de scolarité plusieurs renseignements et les tableaux suivants indiquent les réponses. Un échantillon de 500 étudiants, très représentatif de l'ensemble des étudiants, a été sélectionné et on a regroupé leurs résultats à l'épreuve M.I.S.E. (Mathématiques Informatique Statistique Economie) dans le tableau 1. Le tableau 2 indique leur répartition selon la note obtenue à l'épreuve MISE et le type de Baccalauréat obtenu. Les tableaux 3 et 4 donnent l'évolution des notes moyennes de l'épreuve MISE depuis 2008; les étudiants d'Ecoland sont évalués chaque fin de trimestre. Enfin l'Université test a instauré un tutorat pour aider les étudiants en difficulté depuis deux ans et le tableau 5 indique les bénéficiaires de ce tutorat ainsi que les moyens (en heures) mis à leur service. Le Ministre souhaiterait que vous l'aidiez dans la rédaction de son rapport. Pour cela vous répondrez aux différentes questions.

Tableau 1: notes MISE

Notes (sur 20)	Effectif
[0, 4[50
[4, 6[65
[6, 8[70
[8, 10[70
[10, 12[105
[12, 16[80
16 et plus	60
Total	500

Tableau 2: type de Baccalauréat et notes MISE

Notes \ Bac	Général	Technologique	Etranger	Total
[0, 6[25	30	60	115
[6, 10[55	35	50	140
[10, 12[90	15	0	105
[12, 20]	90	40	10	140
Total	260	120	120	500

Tableau 3: Evolution de la moyenne trimestrielle des notes MISE depuis 2008

Année \ Trimestre	I	II	III	IV
2008	7.6	6.2	11.8	9
2009	7.8	6.4	12.2	9.4
2010	8.2	7	13.2	9.6
2011	8.4	7.6	×	×

Tableau 4: Moyennes mobiles des moyennes MISE

Année \ Trimestre	I	II	III	IV
2008	×	×	8.675	8.725
2009	8.8	8.9	9	9.125
2010	9.325	9.475	9.525	9.625
2011	×	×		

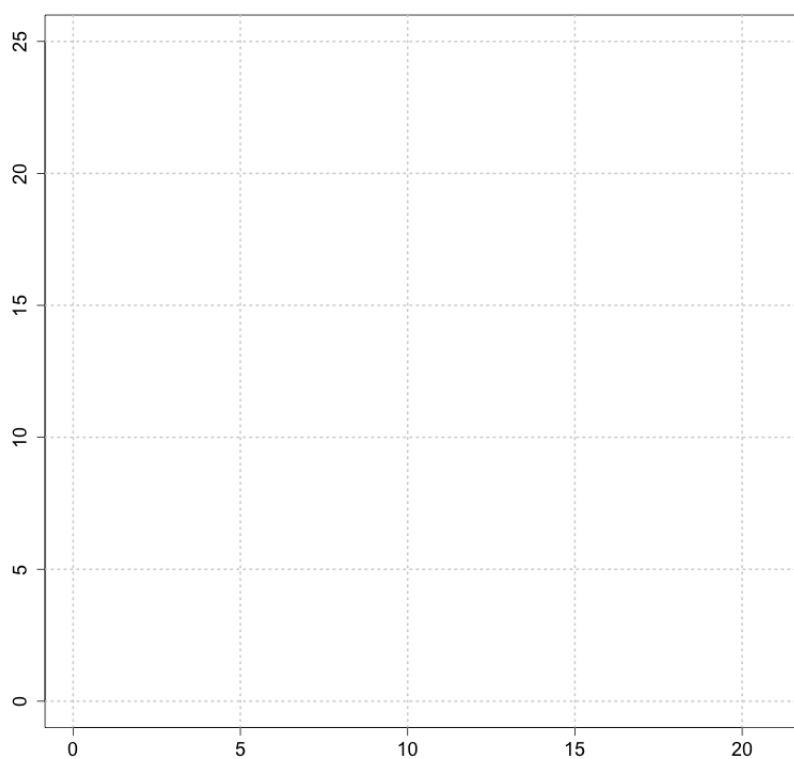
Tableau 5: Tutorat 2009 et 2010

Matière	2009		2010	
	Nbre étud.	Heures par étud.	Nbre étud.	Heures par étud.
Mathématiques	80	8	100	6
Informatique	100	4	60	5
Statistique	80	8	100	10
Economie	120	12	100	10

Consignes générales. Les trois parties peuvent être traitées de manière indépendante. Le barème est donné à titre indicatif. Toute réponse doit être détaillée: si vous utilisez les fonctions internes de la calculatrice, précisez brièvement le mode opératoire. Dans les autres situations, précisez succinctement le calcul. Les résultats doivent être présentés à 10^{-2} près, commentés et interprétés.

PARTIE I: ANALYSE DES TABLEAUX 1 ET 2 (10 POINTS)

1. Représenter graphiquement la distribution de la variable notes MISE (tableau 1) sur le graphique ci-dessous. Détaillez tous les calculs (page suivante) conduisant à la représentation et précisez les axes.



Réponse _____

Fin réponse

2. Calculer la moyenne et l'écart-type des notes.

Réponse _____

_____ Fin réponse

3. Calculer la note médiane de l'ensemble des étudiants étudiés.

Réponse _____

_____ Fin réponse

4. Calculer la moyenne des notes MISE pour les bacheliers de type général, ainsi que de l'ensemble des étudiants selon le regroupement effectué dans le tableau 2.

Réponse _____

_____ Fin réponse

5. On donne les informations suivantes:

- moyenne des notes MISE pour les bacheliers de type technologique : 9.79
- moyenne des notes MISE pour les bacheliers de type étranger : 6.17

Calculer le rapport de corrélation de la moyenne des notes MISE en fonction de leur baccalauréat.

Réponse _____

_____ Fin réponse

6. On souhaite à présent, savoir si les étudiants sont égaux devant l'épreuve MISE selon leur baccalauréat. Calculer le V^2 de Cramer.

Réponse _____

_____ Fin réponse

7. Faire un commentaire synthétique des résultats obtenus (en 10 lignes maximum).

Réponse _____

_____ Fin réponse

PARTIE II: ANALYSE DES TABLEAUX 3 ET 4 (4 POINTS)
--

1. Préciser l'ordre des moyennes mobiles utilisé pour construire le tableau 4.

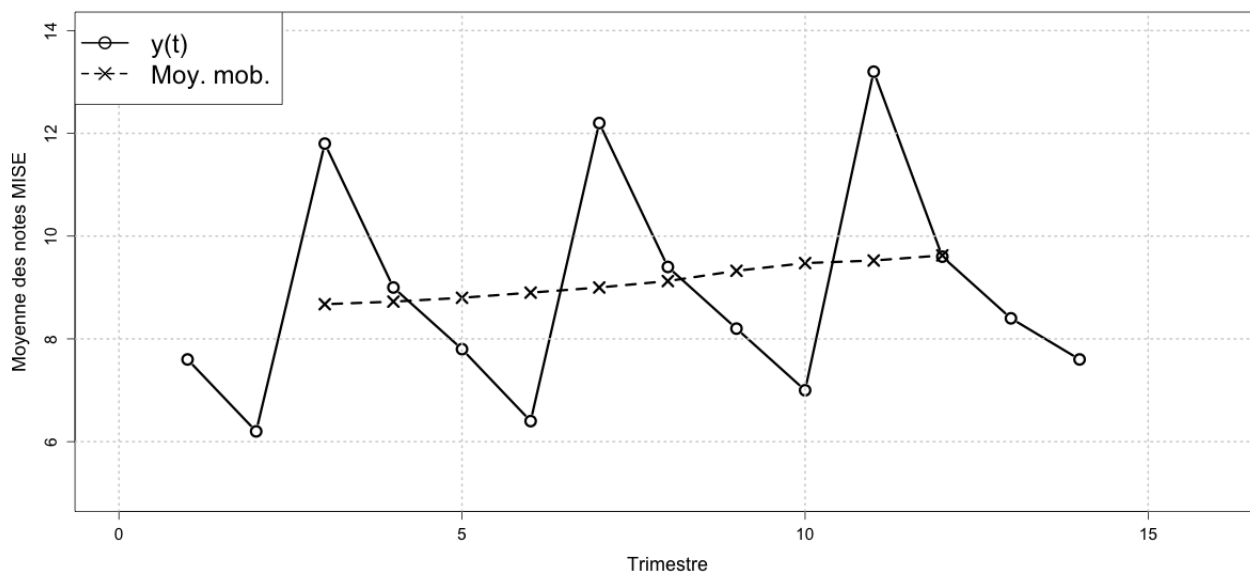
Réponse _____

_____ Fin réponse

2. Calculer une valeur justifiant qu'il semble judicieux de faire un ajustement linéaire des moyennes mobiles. Effectuer l'ajustement linéaire que vous ajouterez au graphique ci-après représentant les séries chronologiques issues des tableaux 3 et 4.

Réponse _____

Fin réponse



3. On utilise un modèle multiplicatif. Justifier ce choix et prédire les moyennes des notes MISE aux deux derniers trimestres 2011. Reportez ces valeurs ainsi que la tendance calculée à la question précédente sur la figure.

Réponse _____

_____ Fin réponse

PARTIE III: ANALYSE DU TABLEAU 5 (4 POINTS)

1. Calculer l'indice des heures de tutorat en 2010 base 100 en 2009 selon les méthodes de Laspeyres et de Paasche.

Réponse _____

_____ Fin réponse

2. Calculer l'indice de la dépense globale en heures du tutorat et en déduire les indices de Laspeyres et de Paasche des effectifs étudiants.

Réponse _____

_____ Fin réponse



Libellé de la matière : Statistiques Descriptives.

Date de l'épreuve : 16/01/2013.

Nom de l'enseignant : Alain Sombardier et Jean-François Coeurjolly

Durée de l'épreuve : 3 heures

Documents autorisés : aucun document autorisé.

Calculatrice autorisée : oui

Numéro d'étudiant :

Le Ministère de l'Agriculture doit préparer les prochaines discussions européennes afin de définir la PAC pour les années 2014-2020. Dans un premier temps, il souhaite connaître la distribution des exploitations du territoire national en fonction de leur surface agricole utilisée (SAU). Pour ce faire, il dispose des données 2010 fournies par l'INSEE (tableau 1). Il désire ensuite mieux connaître les exploitations qui font du bétail et du maïs afin de déterminer les budgets qu'il serait susceptible de demander à Bruxelles afin de soutenir la branche bovine. Il se procure une étude test sur un échantillon de 500 exploitations de l'Ouest (tableau 2). Il est regardant sur le revenu des exploitants agricoles. Pour cela un relevé du revenu net par actif agricole sur les années 2009, 2010 et 2011 a été effectué (tableaux 3 et 4). Enfin, un des problèmes actuels rencontrés par les producteurs agricoles est celui de la politique de prix. Trois produits retiennent tout particulièrement l'attention du Ministère : le maïs, le fourrage et le bétail bovin. Les relevés du prix par tonne (en €) et de l'ensemble des valeurs vendues (en €) sont regroupés dans le tableau 5.

Tableau 1: répartition des exploitations par SAU

Superficie (en ha)	Nbre d'exploitations (en milliers)
[0, 20[210
[20, 50[88
[50, 100[100
[100, 200[82
[200, 360[20
Total	500

Tableau 2: bovins (en milliers de têtes) et maïs (en milliers de tonnes)

maïs \ bovins	[0, 1[[1, 5[[5, 20[Total
[6, 10[140	30	5	175
[10, 12[100	90	5	195
[12, 16]	60	30	40	130
Total	300	150	50	500

Tableau 3: Evolution du revenu net par actif (base 100 en 2010)

Année \ Trim.	I	II	III	IV
2009	78.8	86.2	76.9	83.1
2010	81.4	92.2	98.2	99
2011	101.3	105.5	100	101.2

Tableau 4: Moyennes mobiles du revenu net par actif

Année \ Trim.	I	II	III	IV
2009	×	×	?	82.65
2010	86.06	90.71	95.19	99.34
2011	101.22	101.73	×	×

Tableau 5: maïs, fourrage et bétail

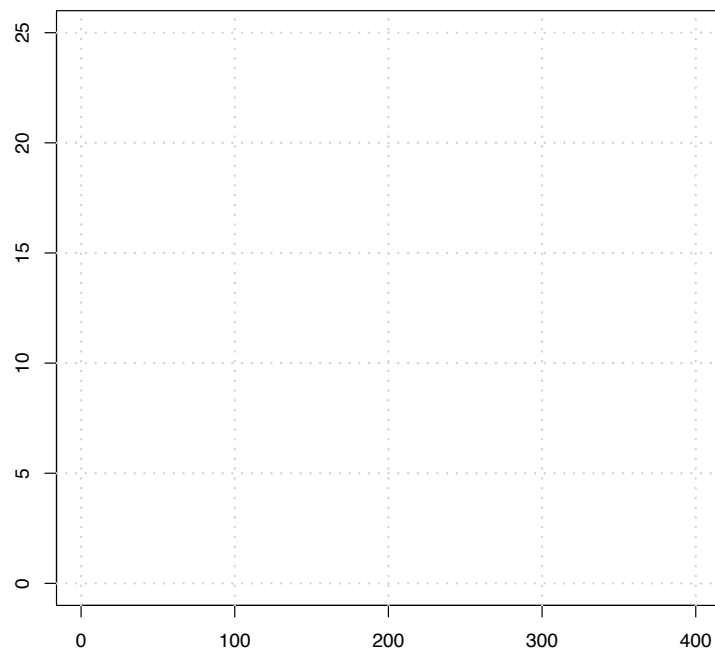
	2010		2011	
	Prix par tonne (en €)	Valeurs (en 10 ⁹ €)	Prix par tonne (en €)	Valeurs (en 10 ⁹ €)
maïs	252	10.9	265	11.8
fourrage	110	7.3	130	7.7
bétail	3500	10.1	3780	10.9

Consignes générales. Les trois parties peuvent être traitées de manière indépendante. Le barème est donné à titre indicatif. Toute réponse doit être détaillée: si vous utilisez les fonctions internes de la calculatrice, précisez brièvement le mode opératoire. Dans les autres situations, précisez succinctement le calcul. Les résultats doivent être présentés à 10^{-2} près, commentés et interprétés.

1. [2pts] Représentez graphiquement la distribution de la variable SAU (n'oubliez pas de préciser les axes). Vous présenterez ci-dessous les calculs nécessaires à la construction de ce graphique sous la forme d'un tableau.

Réponse _____

Fin réponse



2. (a) **[0.5pt]** Déterminez la classe modale de la variable SAU.

Réponse _____

_____ Fin réponse

- (b) **[1.5pt]** Déterminez la superficie médiane d'un terrain agricole, ainsi que les 1er et 9ème déciles associés.

Réponse _____

_____ Fin réponse

- (c) **[1.5pt]** Déterminez la superficie moyenne d'un terrain et l'écart type associé.

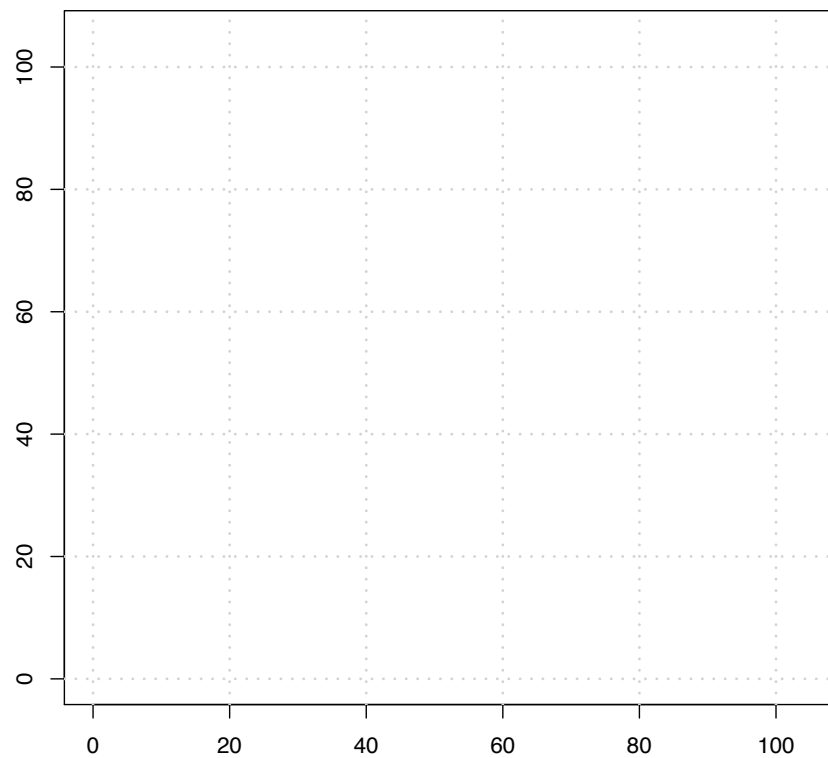
Réponse _____

_____ Fin réponse

3. [2pts] Sur le graphique ci-dessous, représentez la courbe de concentration de la SAU (indication : calculez la variable masse de la SAU). Vous présenterez les calculs nécessaires à la construction de ce graphique sous la forme d'un tableau. Que dire du niveau de concentration de la SAU?

Réponse _____

Fin réponse



4. [2pts] Calculez le V^2 de Cramer entre les variables quantité de maïs et nombre de bovins.
Réponse _____

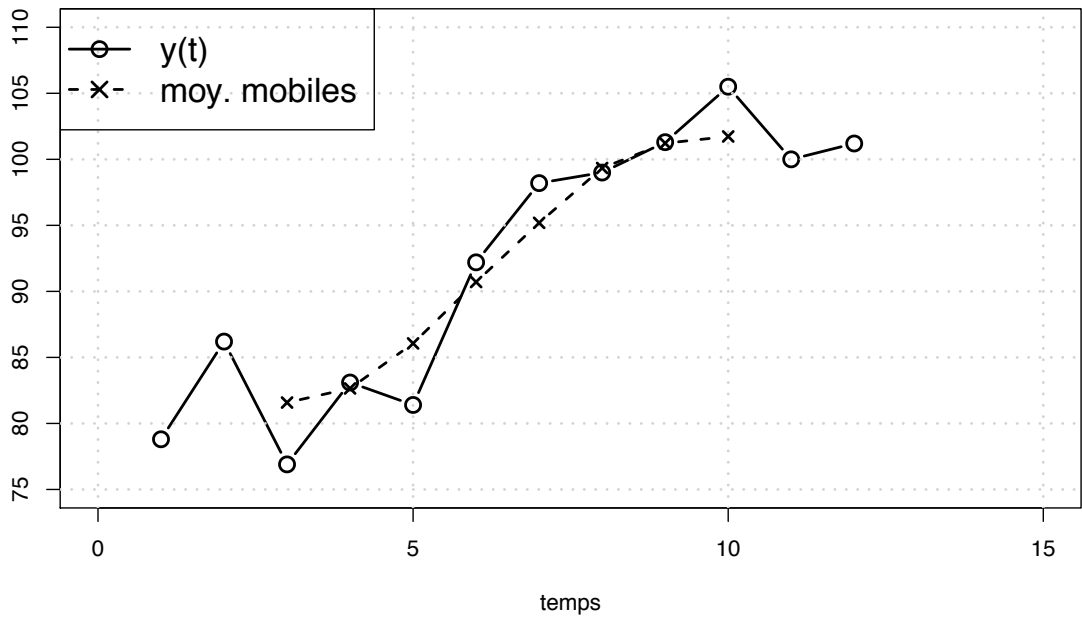
_____ Fin réponse

5. (a) [0.5pt] Calculez la moyenne mobile manquante dans le tableau 5 (précisez auparavant l'ordre de la moyenne mobile utilisée).
Réponse _____

_____ Fin réponse

- (b) [0.5pt] La série chronologique, ainsi que la série des moyennes mobiles (exceptée la 1ère valeur que vous ajouterez au graphique) sont représentées dans la figure ci-après. Quel modèle vous semble le mieux adapté?
Réponse _____

_____ Fin réponse



- (c) [1.5 pt] Après l'avoir justifié, effectuez un ajustement linéaire de la série des moyennes mobiles. Représentez la droite de régression sur le graphique.
 Réponse _____

Fin réponse

- (d) [1.5] Calculez les coefficients saisonniers et déterminez le revenu net prévu pour les 2 premiers trimestres 2013. Représentez ces prévisions sur le graphique.
Réponse _____

_____ Fin réponse

6. (a) [2pts] Calculez l'évolution des prix moyens entre 2010 et 2011 en utilisant les méthodes de Laspeyres et de Paasche (*indication*: il pourra être utile de calculer au préalable les quantités en tonnes de produits vendus pour chacun des trois produits en 2010 et 2011).
Réponse _____

_____ Fin réponse

- (b) [1.5pt] Commentez : précisez notamment si l'évolution de la valeur totale de la production est due à l'augmentation des prix ou plutôt au changement de répartition des volumes produits.

Réponse _____

_____ Fin réponse

7. [5pts] Faites un commentaire synthétique pour préparer les discussions à Bruxelles.

Réponse _____

_____ Fin réponse