STATISTIQUE A DEUX VARIABLES

Exercice 1 :
Un hypermarché dispose de 20 caisses. Le tableau ci-dessous donne le temps moyen d'attente à une caisse en fonction du nombre de caisses ouvertes.

Nombre de caisses ouvertes X	3	4	5	6	8	10	12
Temps moyen d'attente (en minutes)	16	12	9,6	7,9	6	4,7	4

1) Construire le nuage de points M_i (x_i ; y_i) correspondant à cette série statistique. Unités graphiques :
   en abscisse : 1 cm pour une caisse ouverte ;
   en ordonnée : 1 cm pour une minute d'attente.
2) Calculer les coordonnées du point moyen G du nuage et le placer sur le graphique.
3) Ajustement affine.
   a) Calculer le coefficient de corrélation linéaire r.
   b) Déterminer, l'équation de la droite de régression (D) de y en x par la méthode des moindres carrés. Tracer la droite (D) sur le graphique. (Marquer les points utilisés pour tracer (D))
   c) Estimer à l'aide d'un calcul utilisant l'équation de la droite (D).
      i) Le nombre de caisses à ouvrir pour que le temps moyen d'attente à une caisse soit de 5 minutes.
      ii) Le temps moyen d'attente à la caisse lorsque 15 caisses sont ouvertes.
      iii) Pensez-vous que, dans le cas de la question ii), l'ajustement affine soit fiable ?

Exercice 2 :
Dans cet exercice, le détail des calculs n'est pas exigé. On donnera les formules utilisées pour répondre aux questions. Les résultats seront donnés à 10^{-2} près par défaut. Le tableau ci-dessous donne l'évolution de la dette des pays du Tiers Monde entre 1978 et 1992 (en milliards de dollars).

Année	1978	1982	1986	1990	1992
Rang de l'année (X_i)	0	4	8	12	14
Dette (Y_i)	383	753	1089	1346	1510

Source : Banque mondiale, FMI, 1993.

1) Le plan est rapporté à un repère orthogonal. Les unités graphiques sont :
   1 cm pour 2 ans, en abscisses ; 1 cm pour 100 milliards de dollars, en ordonnées.
   Représenter le nuage de points (x_i ; y_i) et le point moyen M de cette série.
2) a) Calculer le coefficient de corrélation linéaire de cette série double.
   Un ajustement affine peut-il être envisagé ? Pourquoi ?
   b) Ecrire une équation de la droite de régression (D) de y en x, par la méthode des moindres carrés. Tracer (D).
   c) Estimer, à 1 milliard de dollars près, le montant prévisible de la dette des pays du Tiers Monde en 2000.

Exercice 3 :
Dans une maternité, on a relevé, pour chacune des vingt naissances d'une journée, l'âge x de la mère et le poids y du nouveau-né. Les résultats sont regroupés dans le tableau ci-dessous :

X	22	18	20	20	16	22	26	22	18	22	18	26	20	26	18	18	22	26	22	20
Y	3,2	2,8	3,2	3,6	2,8	2,8	3	2,8	3	3	3,4	3,6	3,2	3	3	3,4	2,8	2,6	3	3,2

1) Présenter ces données dans un tableau à double entrée.
2) Déterminer les séries marginales associées aux caractères x et y.
3) Représenter par des taches le nuage de points associé à cette série.

Exercice 4 :

Pour dater des ossements d'animaux préhistoriques, on utilise couramment la méthode dite du « carbone 14 ». On admet que, tant que l'animal est vivant, la quantité de carbone 14 est égale à 100. Après la mort de l'animal, la quantité de carbone 14 s'amenuise progressivement.
On donne le tableau suivant :

Quantité restante de carbone 14 (q_i)	48	30	23	16	11
Âge des ossements en milliers d'années (y_i)	6	10	12	15	18

1) On pose : $x_i = \ln q_i$, où ln désigne le logarithme népérien.
   a) Présenter dans un tableau la série (x_i ; y_i) en prenant pour (x_i ; y_i) des valeurs approchées à 0,01 près par défaut.
   b) Représenter le nuage de points de coordonnées (x_i ; y_i) associée à cette série double dans le plan muni d'un repère orthogonal tel que :
   - en abscisse, 4 cm représentent une unité ;
   - en ordonnée, 0,5 cm représente 1 unité.
2) Calculer les coordonnées du point moyen G.
3) On admettra que l'on peut réaliser un ajustement affine du nuage précédent par la droite (Δ) passant par le point moyen G et le point A de coordonnées (2 ; 21,2).
   a) Tracer (Δ).
   b) Déterminer une équation de (Δ).
4) On a mis au jour des ossements dont la teneur en carbone 14 est 40. À partir de l'équation de la droite de (Δ) obtenue au 3), estimer à 500 ans près l'âge de ces ossements.

Exercice 5 :

Le tableau statistique ci-dessous donne le degré de salinité Y_i du Lac Rose pendant le i^{ème} mois de pluie, noté X_i.

X_i	0	1	2	3	4
Y_i	4,26	3,4	2,01	1,16	1,01

Dans ce qui suit il faudra rappeler chaque formule le cas échéant, avant de faire les calculs.
On donnera les valeurs approchées par excès des résultats à 10^{-3} près.

1) a) Déterminer le coefficient de corrélation linéaire de cette série (X, Y) et interpréter le résultat.
   b) Quelle est l'équation de la droite de régression de Y en X
   c) Cette équation permet-elle d'estimer le degré de salinité du lac au 6^{ième} mois de pluie ? Le cas échéant, justifier la réponse.
2) On pose $Z = \ln(Y - 1)$.
   a) Donner le tableau correspondant à la série (X, Z). Les résultats seront arrondis au millième près.
   b) Donner le coefficient de corrélation linéaire de cette série (X, Z).
   c) Donner l'équation de la droite de régression de Z en X, puis exprimer Y en fonction de X.
   d) Utiliser cette équation pour répondre à la question 1) c).

Exercice 6 :

Une entreprise a mis au point un nouveau produit et cherche à en fixer le prix de vente. Une enquête est réalisée auprès des clients potentiels ; les résultats sont donnés dans le tableau suivant où yi représente le nombre d'exemplaires du produit que les clients sont disposés acheter si le prix de vente exprimé en milliers de francs est xi.

xi	60	80	100	120	140	160	180	200
yi	952	805	630	522	510	324	205	84

On appelle x la variable statistique dont les valeurs sont xi et y celle dont les valeurs sont les yi.

1) Calculer le coefficient de corrélation linéaire de y et x.
La valeur trouvée justifie-t-elle la recherche d'un ajustement linéaire ?
2) Déterminer l'équation de la droite de régression de y en x.
3) Les frais de conception du produit se sont élevés à 28 millions de francs.
Le prix de fabrication de chaque produit est de 25.000 francs.
a) Déduire de la précédente question que le bénéfice z en fonction du prix de vente x est donné par l'égalité : $z = -5{,}95x^2 + 1426{,}25x - 59937{,}5$ où x et z sont exprimés en milliers de francs.
b) Déterminer le prix de vente x permettant de réaliser un bénéfice maximal et calculer ce bénéfice.
N.B : Prendre 2 chiffres après la virgule sans arrondir.
Rappel : Bénéfice = Prix de vente – Prix de revient.

Exercice 7 :

1) (X,Y) est une série statistique double. Soit (D_1) la droite de régression de Y en X.
Soit (D_2) la droite de régression de X en Y. On suppose que : $(D_1) : y = ax + b$ et $(D_2) : x = a'y + b'$.
Soit r le coefficient de corrélation linéaire entre X et Y. Etablir que : $r^2 = aa'$.
2) Dans une entreprise une étude simultanée portant sur deux caractères X et Y donnent les résultats suivants :
- la droite de régression de Y en X a pour équation : $2{,}4x - y = 0$;
- la droite de régression de X en Y a pour équation : $3{,}5y - 9x + 24 = 0$.
a) Calculer le coefficient de corrélation linéaire entre X et Y, sachant que leur covariance est positive.
b) Calculer la moyenne de chacun des caractères X et Y.

Exercice 8 :

Le tableau suivant indique la distribution de 50 logements en fonction de leur nombre X de pièces principales et leur surface Y en cm^2.

X \ Y	30	50	70	90	120
1	1	0	0	0	0
2	1	2	2	0	0
3	0	1	6	6	0
4	0	0	2	16	5
5	0	0	0	4	4

1) Déterminer n_{34} et n_{43}.
2) a) Déterminer les distributions marginales associées à X et à Y.
b) Calculer $\bar{X}$, $\bar{Y}$, $V(X)$, $V(Y)$, $\sigma(X)$ et $\sigma(Y)$.
3) Construire le nuage des points, représentant la série statistique double donnée.
4) a) Calculer $Cov(X,Y)$.
b) Déterminer les droites de régression D et D' respectivement de Y en X et de X en Y.
c) Calculer r (coefficient de corrélation linéaire des variables X et Y) et conclure.

Exercice 9 :

Une étude faite sur certain nombre d'entreprises suivant l'investissement mensuel X et le chiffre d'affaire mensuel Y (en millions de francs) a donné les résultats suivants :

Y / X	2	6	10	14	18
6	4	2	1	0	0
8	2	5	2	0	0
10	1	6	16	5	1
12	0	2	3	6	2
14	0	1	0	1	3

1) Donner le nombre total d'entreprises.
2) Combien d'entreprises ont un chiffre d'affaire de 6 millions de francs ?
3) On considère la série conditionnelle X sachant que Y = 2 notée X / Y_1 .

X / Y_1	6	8	10	12	14
Effectif	4	2	1	0	0

Calculer sa moyenne T_1 (T_1 est donc l'investissement moyen des entreprises ayant un chiffre d'affaire mensuel de 2 millions de francs). Déterminer de la même façon pour chacune des autres séries conditionnelles X/Y_i la moyenne T_i (pour i = 2, 3, 4 et 5) puis compléter le tableau suivant :

Y_i	2	6	10	14	18
T_i					

4) Calculer la covariance de Y et T.
5) Donner le coefficient de corrélation linaire de (Y,T). Quel commentaire peut-on faire ?
6) Déterminer la droite d'ajustement de T en Y par la méthode des moindres carrées.
7) Quel devra être l'investissement moyen pour un chiffre d'affaire de 20 millions de francs ?

Exercice 10 :

Dans une maternité, on a relevé pour chacune des 20 naissances d'une journée, l'âge x de la mère (en années) et la masse y du nouveau-né (en kilogrammes).
Les résultats sont regroupés dans le tableau à double entrée ci-dessous :

y \ x	16	18	20	22	26	Totaux
2,6	0	0	0	0	1	1
2,8	1	1	0	3	0	5
3	0	2	0	2	2	6
3,2	0	0	3	1	0	4
3,4	0	2	0	0	0	2
3,6	0	0	1	0	1	2
Totaux	1	5	4	6	4	20

Donner les formules avant d'effectuer les calculs puis les réponses à 10^{-2} près par défaut.

1) a) Déterminer les séries marginales associées aux caractères x et y.
   b) Déterminer les moyennes respectives de ces séries marginales.
   c) Déterminer le coefficient de corrélation linéaire de x et y. La corrélation est-elle bonne ?
2) A la fin de la journée, une équipe de journalistes de passage pour les besoins d'un reportage désire prendre en photo un bébé.
   On suppose que les bébés ont tous les mêmes chances d'être choisis pour la photo.

Soient les événements :
A « Le bébé choisi pèse 3,2 kilogrammes » ;
B « Le bébé choisi a une maman de 22 ans » ;
C « Le bébé choisi pèse 2,8 kilogrammes ».
a) Déterminer les probabilités des événements A, B et $A \cap B$.
En déduire la probabilité $P(A \cup B)$. Justifier les résultats.
b) Déterminer la probabilité $P(C \setminus \overline{B})$. Justifier.

Exercice 11 :

63 candidats se sont présentés au baccalauréat comportant une épreuve de Mathématiques et une épreuve de Sciences Physiques : SP.
Le tableau statistique suivant donne le nombre de candidats ayant obtenu un couple de notes donné :

Note en SP \ Note en Maths	2	6	10	14	18	Totaux
6	4	2	1	0	0	7
8	2	5	2	0	0	9
10	1	6	16	5	1	29
12	0	2	3	6	2	13
14	0	1	0	1	3	5
Totaux	7	16	22	12	6	63

On appelle $X = (x_i)$ la série statistique des notes de Sciences Physiques et $Y = (y_i)$ la série statistique des notes de Mathématiques.
1) Déterminer pour chaque x_i la moyenne conditionnelle z_i de la série conditionnelle y/x_i.
2) On considère la série double (x_i, z_i).
   a) Dans le plan rapporté à un repère orthonormé construire le nuage de points $M(xi, z_i)$.
   b) Calculer le coefficient de corrélation linéaire entre la série $X = (x_i)$ et $Z = (z_i)$.
   c) Déterminer une équation de la droite d'ajustement linéaire de Z en X par la méthode des moindres carrés.
   d) Tracer cette droite.

Exercice 12 :

Une étude faite sur l'effectif X des familles d'une cité et la quantité Y de sucre en Kilogrammes consommée par mois dans chaque famille a donné les résultats ci-dessous :

Y \ X	[5; 7]	[8; 10]	[11; 13]	[14; 18]
[10; 15[	1	3	0	0
[15; 25[	5	9	8	3
[25; 35[	0	7	5	9

1) Calculer la moyenne et l'écart type des séries marginales X et Y.
2) A chaque centre x_i de classe de la série de X on associe la moyenne z_i de Y sachant que $X = x_i$.
   On obtient une série double $(x ; z)$. Déterminer cette série.

3) Dans la suite on considère la série $(x\ ;\ z)$ définie par le tableau suivant :

x_i	6	9	12	16
z_i	18,75	22,5	23,85	27,5

a) Calculer le coefficient de corrélation linéaire entre x et z. Un ajustement affine est-il justifié ?
b) Déterminer une équation de la droite de régression de z en x.
c) Estimer la quantité moyenne de sucre consommée par mois pour une famille d'effectif égal à 20.

Exercice 13 :

Le tableau suivant donne la répartition d'un échantillon de 100 candidats bacheliers suivant leurs notes d'histoire (H) et de sciences (S). On ne dispose pas d'observations individuelles.
Les notes ont été regroupées par classes :

S(x) \ H(x)	[0; 4[	[4; 8[	[8; 12[	[12; 14[	[14; 18[
[0; 4[	2	1	0	0	0
[4; 8[	2	15	5	0	0
[8; 12[	0	4	40	10	0
[12; 14[	0	0	5	6	5
[14; 18[	0	0	0	4	1

On note x_i les centres de classes des notes d'histoire et y_i les centres des notes de sciences.
1) Calculer la moyenne $\bar{x}$ des notes d'histoire et la moyenne $\bar{y}$ des notes de sciences.
2) Pour chaque classe de centre x_i, calculer la moyenne $\bar{y}_i$ des notes de sciences.
Dans un repère bien choisi, représenter la série $(x_i; \bar{y}_i)$. Donner une équation de la droite de régression de $\bar{y}$ en x, par la méthode des moindres carrés, sous la forme $\bar{y} = ax + b$. La tracer.
3) Pour chaque classe de centre y_i, calculer la moyenne $\bar{x}_i$ des notes d'histoire.
Avec le même repère que ci-dessus, représenter la série $(\bar{x}_i; y_i)$. Donner une équation de la droite de régression de $\bar{x}$ en y, par la méthode des moindres carrés, sous la forme $\bar{x} = a'x + b'$. La tracer.
4) Pour mesurer l'intensité de la corrélation linéaire entre les variables x et y, on calcule le coefficient de corrélation linéaire r tel que $r^2 = aa'$, r étant positif si a et a' sont positifs.
Calculer r. La corrélation linéaire semble-t-elle bonne ?

Exercice 14 :
Les parties **A** et **B** sont indépendantes.

Partie A : Une étude du service des transports donne la distance de freinage d'une voiture sur une route en bon état en fonction de sa vitesse.

Vitesse en km/h : X	40	50	60	70	80	90	100	110	120
Distance en m : Y	8	12	18	24	32	40	48	58	72

On désigne par X la vitesse et par Y la distance de freinage.
1) Représenter le nuage de points.
On prendra 1 cm pour 10 km/h en abscisse et 1 cm pour 5 m en ordonnée.
NB : On commencera en abscisse les graduations à partir de 40km/h et en ordonnée les graduations à partir de 8 m.
2) Déterminer l'équation de la droite de régression de Y en X.
3) Déterminer le coefficient de corrélation linéaire r. Avons-nous une bonne corrélation ?

4) a) On suppose que cette évolution se poursuit. Un automobiliste roulant à 150 km/h entame un freinage à 85 m d'un obstacle immobile. Percutera-t-il l'obstacle ?

b) Quelle devrait être sa vitesse maximale au moment du freinage pour ne pas heurter l'obstacle ?

Cause des accidents : X \ Type de transport : Y	Particuliers y_1	Transporteurs en commun y_2
Accidents liés à l'excès de vitesse : x_1	440	360
Accidents à cause mécanique : x_2	110	90

Partie B : Une autre étude sur les causes des accidents donne les résultats ci-contre :

1) Déterminer l'effectif total des accidents enregistrés lors de cette étude.

2) Déterminer les fréquences conditionnelles f_{y_2/x_1} et f_{x_2/y_2} .

3) Déterminer les fréquences marginales $f_{.1}$ et $f_{2.}$.

<u>Exercice 15 :</u>

Une étude sur le nombre d'années d'exercice X, des ouvriers d'une entreprise et leur salaire mensuel Y en milliers de francs, a donné les résultats indiqués dans le tableau ci-dessous avec des données manquantes désignées par a et b.

Y \ X	2	6	10	14	18	22
75	a	5	0	0	0	0
125	0	7	1	0	2	0
175	2	0	9	8	15	4
225	0	1	0	3	b	1

1) Déterminer a et b pour que la moyenne de la série marginale de X soit égale à $\frac{596}{59}$ et celle de la série marginale de Y soit $\frac{8450}{59}$.

2) Dans la suite, on suppose que a = 40 et b = 20. A chaque valeur x_i de X on associe la moyenne m_i de la série conditionnelle : Y/ X = x_i. On obtient ainsi la série double (X, M) définie par le tableau ci-dessous. Les calculs se feront à deux chiffres après la virgule.

X	2	6	10	14	18	22
M	80	113	170	189	199	185

a) Calculer le coefficient de corrélation de X et M puis interpréter le résultat.

b) Déterminer l'équation de la droite de régression de M en X..

c) Quelle serait le salaire moyen d'un ouvrier de l'entreprise si son ancienneté était de 30 ans, si cette tendance se poursuit ?