Exercice n° 1 :

L'observation d'une population de 250 habitats urbains a révélé les résultats suivants concernant le nombre d'habitants (X) et la superficie du logement (Y) :

X \ Y(en m^2)	60-120	120-180	180-400
2-4	22	35	39
4-8	n_{21}	38	22
8-10	20	27	30

T.A.F :

1. Calculer le nombre moyen d'habitants et la superficie moyenne de logement.
2. Calculer $\overline{x_2}$ et donner sa signification
3. Analyser la dispersion de la superficie de logement sachant que le nombre d'habitants est compris entre 4 et 8 habitants.
4. On se propose de voir s'il existe un lien entre le nombre d'habitants et la superficie des logements. Confirmer.

Exercice n° 2 :

La réparation des valeurs de production (Prod) et d'emploi (Emp) réalisés par 730 petites entreprises dans le secteur de textile est présentée comme suit (en milliers de DH) :

Emp \ Prod	500-1000	1000-1500	1500-2000
30-80	48	71	12
80-160	77	n_{22}	68
160-200	114	142	78

T.A.F :

1. Déterminer les distributions marginales et conditionnelles.
2. Calculer les moyennes et les variances marginales.
3. Calculer la moyenne et la variance de la production sachant que l'emploi est compris entre 80 et 160. Interpréter.
4. Analyser l'indépendance des deux variables.

Exercice n° 3 :

L'observation de niveau de risque d'accident effectuée sur 1000 conducteurs a permet de déterminer les proportions de conducteurs suivants les distances parcourues en Km et l'âge.

Distance Age	< 100	100-200	200-300	300-400	400-500
<26	4,4	1,6	--	--	--
26-32	7,2	8,2	4,0	2,6	--
32-38	2,4	7,2	13,6	14,4	4,4
38-42	--	--	2,4	11,6	6,0
42-48	--	--	--	4,4	5,6

T .A.F :

1. Déterminer les distributions marginales conditionnelles.
2. Calculer les moyennes et les variances marginales.
3. Calculer la moyenne et l'écart type de l'âge des conducteurs sachant que la distance parcourue est de 300 à 400 Km. Interpréter.
4. Calculer et interpréter le coefficient de corrélation

Exercice n° 4 :

Le tableau suivant donne la répartition des salariés d'une entreprise de bâtiment selon le nombre d'enfant à charge X et les dépenses mensuelles Y en milliers de DH.

Dépenses Y / Enfants à charge X	13-20	20-25	25-30
1	23	18	26
2	16	22	24
3	31	16	13
4	21	14	08

T.A.F :

1. Donner les deux distributions marginales.
2. Donner les dépenses mensuelles moyenne et médiane pour les salariés avec 4 enfants à charge.
3. Calculer la variance marginale de X.
4. Analyser la dépendance des deux variables.
5. Que peut-on dire de la corrélation des deux variables.

Exercice n° 5 :

Une étude faite sur 350 adhérents d'un club de loisir donne la répartition de ces derniers selon les variables de nombre de visites effectué par semaine (X) et dépenses (Y en dhs) comme suite :

Y / X	500-600	600—650	650-750	750-800
2-4	23	21	26	4
4-8	n_{21}	16	16	3
8-10	29	18	18	4
10-12	n_{41}	26	20	6
12-14	13	25	23	3

T.A.F :

1. Quels sont les caractères étudiés ? Quelle est leur nature ?
2. Compléter le tableau sachant qu'il y a 67 adhérents qui visitent le club entre 4 et 8 fois par semaine.

3. Chercher les dépenses moyenne et modale
4. Quelle est la moyenne des dépenses des adhérents relative aux adhérents qui effectuent entre 10 et 12 visites par semaine.
5. Quelle est le nombre de visite moyen relatif aux adhérents dont les dépenses sont comprises entre 500 et 600 dhs.
6. Calculer en interprétant $\overline{x_2}$ et $\overline{y_1}$.

Exercice n° 6 :

Le service technique d'une entreprise a présenté ses données concernant les investissements et les performances de production réalisés durant les 12 dernières années comme suit :

L'investissement	Les performances de production
2	10
4	16
10	50
14	120
18	140
24	210
28	260
34	298
40	345
46	450
52	521
62	635

TAF:

1. L'entreprise pense qu'il y'a un lien entre l'investissement (X) et les performances de production (Y), pouvez-vous le confirmer ?
2. Etablir la relation liant l'investissement et les performances de production.
3. Combien l'entreprise peut-elle espérer réaliser comme performance de production avec des dépenses d'investissement de 80 dhs ?

Exercice n° 7 :

Une entreprise commerciale a présenté ses ventes (x_i) et ses frais d'emballage (y_i) mensuels au cours de sa première année de démarrage comme suit (en 10000 DH).

Mois	Ventes	Frais d'emballage
1	400	11
2	300	08
3	420	12
4	460	14
5	440	13
6	380	11
7	340	10
8	420	20
9	480	25
10	530	30
11	500	29
12	460	23

TAF :

1. Analyser la dépendance des deux variables.
2. Déterminer une l'équation de la droite d'ajustement qui donne le montant des ventes lorsqu'on connait les frais d'emballage.
3. Quel serait le montant des ventes si les frais d'emballage atteindront 55000DH.
4. Déterminer s'il y a ou non une liaison entre les ventes et les frais d'emballage.

Exercice n° 8:

Des étudiants de 1ere annéeau centre universitaire ont eu les résultats en statistiques et en mathématiques (/100) récapitulés comme suit :

X (notes de statistique)	66	64	69	93	80	71	87	73	79	56	47
Y (note de math.)	72	70	60	94	82	68	86	82	90	55	64

TAF :

1. Préciser la droite d'ajustement de Y à X.
2. Quelle note de statistiques pouvez-vous prédire à un étudiant de ce niveau qui a eu 75 en mathématiques financières ?
3. Calculez le coefficient de corrélation.

Exercice n° 9:

Une agence de voyage a réalisé une étude pour savoir s'il existe une relation entre le prix moyen mensuel de billet (X en 1000 Dhs) et le nombre de client qui ont réservé un voyage (Y). Les résultats sont donnés ci-après :

X	7	9	9,5	11	12,5	13	15	17	19	21
Y	90	110	130	150	170	190	210	230	250	270

TAF :

1. Calculer un paramètre qui permet de mesurer l'intensité de la relation linéaire entre les deux variables étudiées. Interpréter votre résultat.
2. Déterminer l'équation de la droite donnant le nombre de client qui ont réservé un voyage en fonction du prix moyen de billet.
3. Déterminer le nombre de clients qui ont réservé pour des prix moyens de billet qui sont respectivement de 8000,00 Dhs et 10000,00 Dhs.

Exercice n° 10:

On considère un échantillon de 12 clients choisis au hasard. On note X le nombre d'article achetés et Y le nombre de visite effectué au centre commercial, de chaque client durant un trimestre. On obtient les résultats suivant :

X	34	42	53	30	50	60	46	57	32	24	36	28
Y	8	9	11	14	21	12	12	14	11	11	13	5

TAF :

Donner une mesure de la corrélation entre X et Y.

Exercice n° 11:

L'observation sur un marché de raisin de table des quantités offertes (en tonne) et des prix de vente (en Dhs) a donné les résultats suivants :

Quantité X à la vente	100	120	84	78	87	80	110	95
Prix moyen Y par Kg	10,5	9,6	11,2	11,5	10,8	11,2	9,8	10,5

TAF :

1. Calculez le coefficient de corrélation linéaire entre X et Y.
2. On suppose que le prix varie en fonction de la quantité, déterminer l'équation de la droite d'ajustement de Y en X.

SOLUTIONS

Exercice 1 :

Nous commençons par définir la population et les caractères utilisés, tels que :

La population est désignée par les 250 habitats urbains observés **n..=250** .

X représente le nombre d'habitants par habitat observé.

Alors que Y est la superficie du logement c-a-d l'habitat observé.

1-Il s'git de calculer les moyennes marginales des deux variables c-à-d « le nombre moyen d'habitants indépendamment de la superficie de logement $\bar{\bar{X}}$ »ainsi que « la superficie moyenne de logement indépendamment du nombre d'habitants $\bar{\bar{Y}}$ ».

Pour ce faire,on donne le tableau statistique suivant dans lequel nous présentons les distributions marginales de X (marg X) et de Y(marg Y) :

X/Y	**60-120G**	**120-180**	**180-400**	**marg X**
2-4	22	35	39	96
4-8	17	38	22	77
8-10	20	27	30	77
Marg Y	59	100	91	250

Vu que l'effectif n_{21} est indéfini il faut donc le calculer en fonction de l'effectif total **n..**et la somme des autres effectifs simultanés dans le tableau tel que :

$$n_{21}=n_{..}-\sum_{\substack{i=1 j=1 \\ i\neq 2 j\neq 1}}^{3} n_{ij}=17$$

On peut «éventuellement passer par les effectifs marginaux soit de X soit de Y.

Par la suite, on aura besoin de développer le calcul au niveau de chacune des deux variables. D'où les résultats suivants :

Pour la variable X :

X	**$n_{i.}$**	**C_i**	**$n_{i.}*C_i$**	**$f_{i.}$**	**$f_{i.}*C_i$**
2-4	96	3	288	0,38	1,15
4-8	77	6	462	0,31	1,85
8-10	77	9	693	0,31	2,77
T	250		1443	1	5,77

On a donc :

$\overline{\overline{X}}=\frac{\sum_{i=1}^{3} ni.Ci}{n..}=\frac{1443}{250}$=5,77 en termes d'effectifs.

$\overline{\overline{X}} = \sum_{i=1}^{3} fiCi = 5{,}77$ en termes de fréquences.

Ce qui signifie que, indépendamment de la superficie du logement, chaque habitat pris parmi les 250 observés contient un nombre moyen d'habitant d'environ 6 personnes.

Pour la variable Y :

Y	**$n_{.i}$**	**C_j**	**$n_{.j}*C_j$**	**$f_{.j}$**	**f.j*Ci**
60-120	59	90	5310	0,24	60,00
120-180	100	150	15000	0,40	105,56
180-400	91	290	26390	0,36	105,56
T	250		46700	1	186,80

Ce qui donne :

$$\overline{\overline{X}} = \frac{\sum_{J=1}^{3} nj \times, cj}{n} = \frac{46700}{250} = 186{,}80 \text{ en termes d'effectifs.}$$

$\overline{\overline{X}} = \sum_{i=1}^{3} Fj \times cj = 186{,}80$ en termes de fréquences.

Ce qui signifie que chaque habitat pris parmi les 250 observés est composé par une superficie moyenne du logement d'environ 186,80 m².et ce indépendamment du nombre d'habitant.

2-$\overline{x_2}$ représente la moyenne de X conditionnée par la 2° modalité de Y. ceci revient à calculer le nombre moyen d'habitant sachant que la superficie de logement est comprise entre 120 et 180 m².

Soit donc le tableau suivant présentant les statistiques de cette distribution conditionnelle de X :

$Y_{/j=2}$	**$n_{.i2}$**	**C_j**	**$n_{.i2}*C_j$**	**$F_{i/j=2}.j$**	**$F_{i/2}*Ci$**
2-4	35	3	105	0,35	1,05
4-8	38	6	228	0,38	2,28
8-10	27	9	243	0,27	2,43
T	100		576	1	5,76

Ce qui donne :

$\overline{X_{i\ j=2}} = \frac{\sum_{i=1}^{3} n_{i2} \times Ci}{n_2} = \frac{576}{100} = 5{,}76$en termes d'effectifs.

$\overline{X_{i\ j=2}} = \sum_{i=1}^{3} F_{i2} \times Ci = 5{,}76$en termes de fréquences.

Ce qui signifie que chaque habitat pris parmi les 100 observés et dont la superficie est exactement comprise entre 120 et 180 m², cet habitat est composé d'un nombre moyen d'habitant d'environ 6 personnes. Ce qui représente encore pratiquement la même moyenne que l'ensemble des habitats.

3- Calculer la superficie moyenne de logement sachant que le nombre d'habitant est compris entre 4 et 8 personnes revient à déterminer $\overline{Y_J/i_{=2}}$.

Ce qui représente la moyenne de Y conditionnée par la 2° modalité de X.

D'où le tableau des statistiques suivant :

$Y_{/i=2}$	$n_{.2j}$	C_j	$n_{.2j}*C_j$	$F_{j/i=2}$	$f_{\bullet j/2}*C_j$
60-120	17	90	1530	0,22	19,87
120-180	38	150	5700	0,49	74,03
180-400	22	290	6380	0,29	82,86
T	77		13610	1	176,75

Ce qui donne :

$\overline{Y_{i\ j=2}} = \frac{\sum_{J=1}^{3} n_{2J} \times CJ}{n_2} = \frac{13610}{77} = 176{,}75$en termes d'effectifs.

$\overline{Y_{i\ j=2}} = \sum_{J=1}^{3} f_{2J} \times CJ = 176{,}75$en termes de fréquences.

Ceci signifie que chaque logement pris parmi les 77 observés et dont le nombre moyen d'habitant est compris entre 4 et 8 personnes, est représenté par une superficie moyenne d'environ 177 m².

Nous travaillons dans cette question sur la même distribution que la question précédente c-à-d « la superficie de logement » conditionnée par le nombre d'habitant compris entre 4 et 8 habitants $Y_{j/i=2}$.

Pour analyser sa dispersion on reprend le mêmetableau précédent comme suit :

$Y_{/i=2}$	N_{2j}	c_j	$N_{2j}*c_j$	$F_{j/i=2}$	$F_{j/2}*c_j$	$N2_j*(c_j)^2$	$F_j/2*(c_j)^2$
60-120	17	90	1530	0,22	19,87	137700	1788,31
120-180	38	150	5700	0,49	74,03	855000	11103,90
180-400	22	290	6380	0,29	82,86	1850200	24028,57

T	77		13610	1	176,75	2842900	36920,78

Ce qui donne un écart-type de :

-En termes d'effectifs.

$$\boldsymbol{\sigma_{j\,,i=2}} = \sqrt[2]{\frac{\sum_{i=1}^{3} n_{2J} \times (CJ)^2}{n_2} - (\overline{\overline{Y_{J\ i=2}}})^{\mathbf{2}}} = \sqrt[2]{\frac{\mathbf{242900}}{\mathbf{77}} - (\mathbf{176,75})^{\mathbf{2}}} = \sqrt[2]{\mathbf{5679,07}}$$
$$= \mathbf{75,36}$$

En termes de fréquences.

$$\boldsymbol{\sigma_{j\,,i=2}} = \sqrt[2]{\sum_{J=1}^{3} f_{2J} \times CJ - (\overline{\overline{Y_{J\ i=2}}})^{\mathbf{2}}}^{\mathbf{=}\sqrt[2]{\mathbf{36920,78}\ -(\mathbf{176,75})^{\mathbf{2}}}} = \sqrt[2]{\mathbf{5679,07}} = \mathbf{75,36}$$

Pour analyser la dispersion de la superficie de logement sachant que le nombre nombre d'habitants est compris entre 4 et 8 habitants, on doit calculer le coefficient de variation de cette distribution

CV = $\frac{\boldsymbol{\sigma_{j\,,i=2}}}{\overline{\overline{Y_{i\ j=2}}}}$ = 0 ,42

On peut dire qu'il existe une moyenne dispersion

5-Pour analyser l'existence du lien entre les 2 variables X : « le nombre d'habitants » et Y : « la superficie des logements » on peut passer par les fréquences ou la covariance.

-Pour les fréquences on peut, en effet, comparer les fréquences conditionnelles par rapport aux fréquences marginales. Si toutes les fréquences conditionnelles sont égales entre elles et sont égales à leurs fréquences marginales correspondantes alors les deux distributions seront totalement indépendantes. Si par contre il existe un seul cas de différence entre celles-ci, le lien serait inexistant.

On peut montrer dans ce cas que :

$$f1 = \frac{n_{11}}{n_1} = \frac{22}{59} = 0{,}373\ et\ \ f1 = \frac{n_1}{n.} = \frac{96}{250} = 0{,}384$$

Les fréquences sont toutes les deux différentes. Ce qui signifie que les deux variables sont dépendantes malgré la légère différence entre les deux fréquences.

-Pour la covariance , ce qui est un peu compliqué que le calcul des fréquences, on peut montrer avec l'équation de la covariance que :

$$cov(x,y) = \frac{\sum_{i=1,J=1}^{3} n_{ij}\,(c_i - \bar{x})(c_j - \bar{y})}{n_{..}}$$

$N_{ij}*(c_i-x_{bar})*(c_j-y_{bar})$	**90**	**150**	**290**	**T**
3	5898,99	3567,76	-11148,7	-1682,16
6	-378,49	-321,63	522,19	-177,93
9	-6253,28	-3209,33	10000,08	537,47
				-1322,41

D'où :

$$cov(x,y) = \frac{\sum_{i=1,J=1}^{3} n_{ij}\,(c_i - \bar{x})(c_j - \bar{y})}{n\cdot} = \frac{-1322{,}41}{250} = -5{,}29$$

Ce qui confirme effectivement la dépendance des deux variables puisque la covariance est nulle. Ces dernières sont même liées négativement puisque la covariance est négative.

Ce qui signifie que lorsque la superficie du logement augmente le nombre d'habitant diminue et vise versa.

<u>Exercice 2</u> :

De la même façon que l'exercice précédent, on a :

La population est composée par les 730 petites entreprises observées dans le secteur de textile avec **n..=730**.

X :le nombre d'emplois par entreprise.

Y : la production réalisée en milliers de Dhs.

1-On donne dans ce qui suit successivement les distributions marginales et conditionnelles des deux variables :

-On présente dans le tableau ci-dessous les distributions marginales successivement de X dans la dernière colonne et de Y dans la dernière ligne.

X/Y	**500-1000**	**1000-1500**	**1500-2000**	**Marg X**
30-80	48	71	12	131
80-160	77	120	68	265
160-200	114	142	78	334
Marg Y	239	333	158	730

L'effectif n_{22} est calculé en fonction de l'effectif total n..et la somme des autres effectifs simultanées dans le tableau tel que :

$$n_{22} = n_{..} - \sum_{\substack{i=1\cdot j=1 \\ i\neq 1\cdot j\neq 1}}^{3} n_{ij=120}$$

-Les distributions conditionnelles quant à elles sont représentées en effectifs dans toutes les colonnes du tableau(sauf la dernière colonne) pour la variable X et toutes les lignes du tableau(sauf la dernière ligne)pour la variable Y.

On peut les représenter séparément comme suit :

Pour X

$X_{i/j=1}$	**500-1000**
30-80	48
80-160	77
160-200	114
$n_{.1}$	239

$X_{i/j=2}$	1000-1500
30-80	71
80-160	120
160-200	142
$n_{.2}$	333

$X_{i/j=3}$	1500-2000
30-80	12
80-160	68
160-200	78
$n_{.3}$	158

Pour Y

$Y_{j/i=1}$	30-80
500-1000	48
1000-1500	71
1500-2000	12
$n_{1.}$	131

$Y_{j/i=2}$	80-160
500-1000	77
1000-1500	120
1500-2000	68
$n_{2.}$	265

$Y_{j/i=3}$	160-200
500-1000	114
1000-1500	142
1500-2000	78
$n_{3.}$	334

2-Le calcul des moyennes ($\bar{\bar{x}}; \bar{\bar{y}}$) et des variances ($\sigma_x$;σ_y) marginales des deux variables l'une indépendamment de l'autre donne les résultats qu'on présente successivement dans ce qui suit :

-Pour la variable X :

X	**$n_{i.}$**	**C_i**	**$n_{i.}*C_i$**	**$n_{i.}*(C_i)^2$**
30-80	131	55	7205	396275
80-160	265	120	31800	3816000
160-200	334	180	60120	10821600

T	730		99125	15033875

On a donc :

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{3} \mathrm{n_{ici}}}{n..} = \frac{99125}{370} = 135{,}79$$

Chaque petite entreprise prise parmi les 730 observés dans le secteur de textile compte un nombre moyen d'emploi d'environ 136 personnes et ce indépendamment de la production réalisée.

Var (X) = $\sum_{i=1}^{p} \frac{ni.(xi-\bar{\bar{x}})^2}{n..} = 2156.06$

$$\sigma x = \sqrt[2]{\sum_{i=1}^{p} \frac{ni.(xi-\bar{\bar{x}})^2}{n..}} = \sqrt[2]{2156.06} = 46{,}43$$

En moyenne et indépendamment de la production réalisée, le nombre d'emploi de chaque entreprise des 730 observées, s'éloigne de l'emploi moyen (égale à 135,79) positivement ou négativement ou d'environ 46 emplois.

-Pour la variable Y :

Y	**n_j.**	**C_j**	**N_j.*C_j**	**N_j.*(C_j)^2**
500-1000	239	750	179250	134437500
1000-1500	333	1250	416250	520312500
1500-2000	158	1750	276500	483875000
T	730		872000	1138625000

Ce qui donne :

$$\bar{\bar{y}} = \frac{\sum_{j=1}^{3} \mathrm{nj.} \times \mathrm{cj}}{n.} = \frac{872000}{730} = 1194.52$$

Chaque petite entreprise prise parmi les 730 observées dans le secteur de textile réalise une production d'environ 1194520 dhs et ce indépendamment du nombre d'emploi effectué.

Var (Y) = $\sum_{j=1}^{q} \frac{n.j\,(Yj-\bar{\bar{y}})^2}{n..} = 132880.93$

$$\sigma y = \sqrt[2]{\sum_{j=1}^{q} \frac{n.j(Yj-\bar{\bar{y}})^2}{n..}} = \sqrt[2]{132880.93} = 364.53$$

En moyenne et indépendamment du nombre d'emploi, la production réalisée par chaque entreprise des 730 observées, s'éloigne de la production moyenne (égale à 1194520dhs) positivement ou négativement d'environ 364530dhs.

3-La distribution concernée dans cette question est représentée par « la production conditionnée par l'emploi ». Si celui-ci est exactement compris entre 80 et 160 emplois, c.-à-d. la 2° modalité de X, alors la production sera donnée par $Y_{j/i=2}$.

Soit donc le tableau suivant présentant les statistiques de cette distribution conditionnelle de Y :

$Y_{j/i=2}$	$N_{2j.}$	C_j	$N_{2j.}*C_j$	$N_{2j.}*(C_j)^2$
500-1000	77	750	57750	43312500
1000-1500	120	1250	150000	187500000
1500-2000	68	1750	119000	208250000
T	265		326750	439062500

Ce qui donne :

$$\overline{\mathbf{Yj/i=2}} = \frac{\sum_{j=1}^{3} \mathrm{n2j.}\times (\mathrm{cj})^2}{n2} = \frac{326750}{265}\ 1233.02$$

Ce qui signifie que chaque petite entreprise prise parmi les 265 œuvrant dans ce secteur de textile et ayant un emploi exactement compris entre 80 et 160 personnes, réalise une production moyenne de 133020dhs.

La variance quant à elle définie comme suit :

$$V2(Y) = \frac{\sum_{j=1}^{3} \mathrm{n2j} \times (\mathrm{cj} - (\overline{\mathbf{Yj/i=2}}))^2}{n2.} = 136504.09$$

Ce qui donne un écart-type de :

$$\sigma = \sqrt[2]{\sigma^2} = \sqrt[2]{136504.09} = 369.46$$

Ceci signifie qu'en moyenne, la production de chaque petite entreprise prise parmi les 265 œuvrant dans le secteur de textile et ayant un emploi exactement compris entre 80 et 160 personnes, s'éloigne de la production moyenne (égale à 133030dhs) positivement ou négativement d'environ de 369460dhs.

6-L'analyse de l'indépendance des deux variables X : « le nombre d'emplois »

Et Y : « la production en milliers de dhs » peut se faire comme dans l'exercice précédent soit par les fréquences soit par la covariance.

On peut aisément vérifier pour lesconditionnelles comparées par rapport aux fréquences marginales que les deux variables sont effectivement dépendantes. Toutes les fréquences conditionnelles dans ce cas sont, en effet, différentes entre elles et sont différentes des fréquences marginales correspondantes.

On peut montrer par exemple que :

$$\frac{f2}{3} = \frac{n23}{n.3} = \frac{68}{158} = 0.43 \; et \; f2 = \frac{n2}{n.} = \frac{265}{730} = 0.36$$

Ces deux fréquences, la première conditionnelle et la seconde marginale, sont toutes les deux différentes. Ce qui signifie que les deux variables sont certainement liées entre elles. Ce lien peut être fort ou faible selon le degré d'influence de l'une sur l'autre.

Exercice3 :

La solution donnée à cet exercice doit se faire de la même manière que le précédent. Sauf que le tableau statistique présente dans ce cas les fréquences simultanées f_{ij} et non marginales ni conditionnelles. Ce qui représente en fait une remarque de fond dans ce cadre.

Nous avons :

La population est composée par les 1000 conducteurs avec n..=1000 .

X=l'âge des conducteurs.

Y=est la distance parcourue en KM.

1-Avant de déterminer les distributions marginales et conditionnelles et afin de rendre le calcul et le raisonnement plus simples il faut reconstituer le tableau statistique en termes des effectifs et non les fréquences relatives.

En effet, raisonner sur les fréquences (comme dans le tableau initial définit en %) peut poser un problème au niveau des autres distributions marginales et conditionnelles puisque celles-ci ne sont pas calculées sur la même base. Si les fréquences marginales sont définies par rapport à l'effectif total n.., les fréquences conditionnelles sont par contre définies par rapport aux effectifs marginaux ni. Ou n.j .

On se propose donc pour répondre à cette question, de retrouver le tableau statistique en termes des effectifs en fonction du tableau initial tel que :

Tableau initial présentant les distributions marginales en termes de fréquences %

Distance-	<100	100-	200-	300-	400-	Marg X

age		200	300	400	500	
<26	4.4	1.6	--	--	--	06.00
26-32	7.2	8.2	4.00	2.6	--	22.00
32-38	2.4	7.2	13.6	14.4	4.4	42.00
38-42	--	--	2.4	11.6	6.0	20.00
42-48	--	--	--	4.4	5.6	10.00
Marg Y	14.00	17.00	20.00	33.00	16.00	100.00

Par exemple, 4,4% de l'effectif total signifie que 44 sur les 1000 conducteurs observés ont un âge inférieur à 26 ans et parcourent une distance inférieur à 100km.

Tableau présentant les distributions marginales en termes des effectifs.

x/y	*0-100*	*100- 200*	*200-300*	*300-400*	*400-500*	*Marg X*
20-26	*44*	*16*	*0*	*0*	*0*	*60*
26-32	*72*	*82*	*40*	*26*	*0*	*220*
32-38	*24*	*72*	*136*	*144*	*44*	*420*
38-42	*0*	*0*	*24*	*116*	*60*	*200*
42-48	*200*	*0*	*0*	*44*	*56*	*100*
42-48	*140*	*170*	*200*	*330*	*160*	*1000*

N.B Les deux bornes de « 20 » pour X et « 0 » pour Y sont calculées en fonction des amplitudes fréquentes alors que les effectifs nij sont estimés en utilisant les fréquences relatives et l'effectif total tel que :

$$n_{ij} = f_{ij} \times n$$

Pour les distributions conditionnelles et comme d'habitude elles sont toujours représentées en effectifs ou en fréquences dans toutes les colonnes du tableau (sauf la dernière colonne) pour la variable X et toutes les lignes du tableau (sauf la dernière ligne) pour la variable Y.

On aura en effet autant de distributions pour chaque variable qu'il y a de modalités dans l'autre variable. Ce qui fait 5 distributions pour chaque variable qu'on peut les représenter séparément comme suit :

Pour X :

$X_{i/j=1}$	**0-100**
2026	**44**

$X_{i/j=2}$	**100-200**
2026	**16**

$X_{i/j=3}$	**200-300**
2026	**0**

26-32	72
32-38	24
38-42	0
42-48	0
$n_{.1}$	140

26-32	82
32-38	72
38-42	0
42-48	0
$n_{.2}$	170

26-32	40
32-38	136
38-42	24
42-48	0
$n_{.3}$	200

$X_{i/j=4}$	300-400
2026	0
26-32	26
32-38	144
38-42	116
42-48	44
$n_{.4}$	330

$X_{i/j=5}$	400-500
20	0
26-32	0
32-38	44
38-42	60
42-48	56
$n_{.5}$	160

Pour Y :

$Y_{j/i=1}$	20-26
0-100	44
100-200	16
200-300	0
300-400	0
400-500	0
$n_{.1}$	60

$Y_{j/i=2}$	26-32
0-100	72
100-200	82
200-300	40
300-400	26
400-500	0
$n_{.2}$	220

$Y_{j/i=3}$	32-38
0-100	24
100-200	72
200-300	136
300-400	144
400-500	44
$n_{.3}$	420

$Y_{j/i=4}$	38-42
0-100	0
100-200	0
200-300	24
300-400	116
400-500	60
$n_{.4}$	200

$Y_{j/i=5}$	42-48
0-100	0
100-200	0
200-300	0
300-400	44
400-500	56
$n_{.5}$	100

2-Le calcul des moyennes ($\bar{\bar{x}}, \bar{\bar{y}}$) et des variances ($\sigma_x$,$\sigma_y$) marginales des deux variables l'une indépendamment de l'autre est présenté successivement comme suit :

-Pour la variable X :

X	$n_i.$	C_i	$n_i.*C_i$	$n_i.*(C_i)^2$
20 26	60	23	1380	31740
26-32	220	29	6380	185020
32-38	420	35	14700	514500
38-42	200	40	8000	320000
42-48	100	45	4500	202500
T	1000		34960	1253760

On a donc :

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{5} ni\ ci}{n..} = \frac{34960}{1000} = 34.96$$

Chaque conducteur observé dans cet échantillon de 1000 est agé en moyen d'environ 35 ans et ce indépendamment de la distance parcourue.

Var (X) = $\sum_{i=1}^{p} \frac{ni.(xi - \bar{\bar{x}})^2}{n..} = 31,58$

$$\sigma_x = \sqrt[2]{\sum_{i=1}^{p} \frac{ni.(xi - \bar{\bar{x}})^2}{n..}} = 5.62$$

En moyenne et indépendamment de la distance parcourue, l'âge de chaque conducteur pris dans cet échantillon, s'écarte de l'âge moyen (égale à 34,96) positivement ou négativement d'environ 6 ans.

-Pour la variable Y :

y	$n_i.$	C_i	$n_i.*C_i$	$n_i.*(C_i)^2$
0-100	140	50	7000	350000
100-200	170	150	25500	3825000
200-300	200	250	50000	12500000
300-400	330	350	115500	40425000
400-500	160	450	72000	32400000
T	1000		270000	89500000

Ce qui donne :

$$\bar{\bar{y}} = \frac{\sum_{j=1}^{5} nj\ \times cj}{n..} = \frac{270000}{1000} = 270.000$$

Chaque conducteur observé dans cet échantillon de 1000 parcourt une distance moyenne d'environ 270km et ce indépendamment de l'âge.

Var (Y) = $\sum_{j=1}^{q} \frac{n.j\,(Yj-\bar{\bar{y}})^2}{n..} = 16599{,}74$

$$\sigma_y = \sqrt[2]{\text{Var (Y)} \;=\; \sum_{j=1}^{q} \frac{n.j(Yj-\bar{\bar{y}})^2}{n..}} = \sqrt[2]{16599{,}74} = 128.84$$

En moyenne et indépendamment de l'âge, la distance parcourue par chaque conducteur observé dans cet échantillon, s'écarte de la distance moyenne (égale à 270km) positivement ou négativement d'environ 129km.

3-Il s'agit de calculer la moyenne et l'écart type de la distribution représentée par « l'age des conducteurs conditionné par la distance parcourue ». Si la distance est exactement comprise entre 300 et 400 km, c.-à-d. la 4° modalité de Y, alors la distribution sera présentée par X/j=4.

Soit donc le tableau suivant présentant les statistiques de la distribution conditionnelles de X :

$X_{i/j=4}$	n_{i4}	C_i	$n_{i4}*C_i$	$n_{i4}*(C_i)^2$
20 26	0	23	0	0
26-32	26	29	754	21866
32-38	144	35	5040	176400
38-42	116	40	4640	185600
42-48	44	45	1980	89100
T	330		12414	472966

Ce qui donne :

$$\overline{x_{i/j4}} = \frac{\sum_{i=1}^{5} ni\;4 \times ci}{n.4} = \frac{12414}{330} = 37.62$$

Ce qui signifie que chaque conducteur pris parmi les 330 conducteurs ayant parcouru une distance comprise entre 300 et 400km, possède un age moyen de 38 ans.

Le calcul de l'écart-type sera comme suit :

$$\sigma_{i/j4} = \sqrt[2]{\frac{\sum_{i=1}^{5} ni4\; \times (ci)^2}{n..} - (\overline{y_{j/i2}})^2} = \sqrt[2]{\frac{472966}{330} - (37.62)^2} = 4.25$$

Ceci signifie qu'en moyenne, l'age de chaque conducteur observé parmi les 330 ayant parcouru une distance exactement comprise entre 300 et 400km, s'éloigne de l'age moyen (égale à 37,62) d'environ 4 ans soit positivement soit négativement.

4-L'analyse de quelques fréquences conditionnelles comparées par rapport aux autres fréquences conditionnelles de la meme variable ou bien aux fréquences marginales peut facilement révéler que les deux variables sont effectivement dépendantes.

En effet, on peut montrer par exemple que :

$$f\,{}^{2}/_{3} = \frac{n23}{n.3} = \frac{40}{200} = 0.2 \; et \; f2/4 = \frac{n24}{n.4} = \frac{26}{330} = 0.079$$

Ces deux fréquences conditionnelles sont calculées pour la meme variable qui est X : »age des conducteurs » alors qu'elles sont, par contre, différentes. Ce qui signifie que l'information apportée par la réalisation de la seconde variable change le résultat. Selon que la distance parcourue par les conducteurs soit comprise entre 200 et 300 ou bien entre 300 et 400km, cela infecte le pourcentage de ces derniers qui est respectivement de 20% et 7,9%. Les deux variables sont donc bel et bien dépendantes l'une de l'autre.

Exercice 4 :

Nous avons :

La population est composée par les salariés d'une entreprise de batiment et dont l'effectif est égal à 232 ce qui donne n..=232.

X : « Le nombre d'enfants à charge ». c'est un caractère quantitatif discret.

Y : « les dépenses mensuelles présentées sur le tableau en milliers de DH » en tant que caractère continu.

1-les distributions marginales des deux caractères sont présentées dans le tableau statistique suivant n termes des effectifs tel que :

X/Y	13-20	20-25	25-30	Marg X n.i
1	23	18	26	67
2	16	22	24	62
3	31	16	13	60
4	21	14	8	43
Marg Y n.i	91	70	71	232

On a, par exemple, 67 salariés qui ont un seul enfant à charge, et ce quelque soit les dépenses mensuelle. Généralement, ces derniers dépensent entre 13000dhs et 30000dhs.

Par contre, 91 des salariés sur 232 qui sont observés en total dépensent entre 13000dhs et 20000dhs par mois, et ce quelque soit le nombre d'enfant à charge. Ce qui représente 39,22 de l'échantilon.

2-Il suffit de calculer la moyenne et la médiane de la distribution représentée par « les dépenses mensuelles conditionnée par le nombre d'enfants à charge ». Ce qui représente Yj/i=4 avec la 4° modalité de X qui définit la condition.

D'où le tableau suivant :

$y_{j/i=4}$	N_{4j}	C_j	$n_{4j}*C_i$	$n_{4j}CC$
13-20	**21**	**16 ,5**	**346,5**	**21**
20-25	**14**	**22,5**	**315**	**35**
25-30	**8**	**27,5**	**220**	**43**
T	**43**	**--**	**881.5**	**--**

Ce qui donne une dépense mensuelle moyenne de :

$$\overline{y_{j/i4}} = \frac{\sum_{i=1}^{3} n4j \times cj}{n4.} = \frac{881.5}{43} = 20.50$$

Cela représente une dépense mensuelle moyenne d'environ 20500dhs pour chaque salarié pris parmi les 43 qui ont exactement 4 enfants à charge.

La dépense mensuelle médiane est calculée comme suit :

$$Me = I_0 + \frac{a\left[\frac{n}{2} - n_{i-1}cc\right]}{n_i} = 20 + \frac{5.(21.5 - 21)}{14} = 20.18$$

Avec la classe médiane égale à [20-25] puisque la moitié de cette population est égale à $\frac{n}{2}$ =21,5.

On a donc la moitié de cette population c-a-dd environ 22 salariés sur 43 d'entre eux ayant 4 enfants à la charge qui ont une dépense mensuelle inférieure à 20180dhs. L'autre moitié de cette population dépense plus que 20180dhs par mois.

3-Pour calculer la variance marginale de X on donne le tableau suivant :

X	$n_{i.}$	$n_{i.}*C_i$	$n_{i.}*(c_i)^2$
1	67	67	67
2	62	124	248
3	60	180	540
4	43	172	688
T	232	543	1543

On vérifie d'abord la moyenne :

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{5} ni\ xi}{n..} = \frac{543}{232}\ 2.34$$

Indépendamment des dépenses mensuelles effectuées, chaque salarié observé dans cette population de 232 salariés à un nombre moyen d'environ 2 enfants à charge.

Puis la variance :

$$\sigma_x = \sqrt[2]{\frac{\sum_{i=1}^{4} ni\ \times (ci)^2}{n..} - \ (\bar{\bar{x}})^2} = \sqrt[2]{\frac{1543}{232} - \ (2.34)^2} = 1.08$$

En moyenne et indépendamment des dépenses mensuelles effectuées, le nombre d'enfants à charge de chaque salarié observé dans cette population de 232 s'écarte du nombre moyen d'environ plus ou moins l'enfant.

4-Dans le cadre de la statistique descriptive à deux dimensions, une règle générale veut que « la moyenne marginale de chaque variable est égale à la moyenne de ses moyennes conditionnelles ». C'est exactement ce qu'on cherche à vérifier dans cette question, tel que :

$$\bar{\bar{x}} = \sum_{j=1}^{p} f.j\ \times \bar{x}\ \ et\, \bar{\bar{y}} = \sum_{i=1}^{n} fi\ \times \bar{y}$$

Avec :

$\bar{\bar{x}}$et$\bar{\bar{y}}$: définissent les moyennes marginales respectivement de X et de Y.

$\bar{x_j}\ et\ \bar{y_i}$:définissent les moyennes conditionnelles de X et de Y respectivement.

$f_i\ et\ \ f_j$:sont les fréquences marginales respectivement de X et de Y.

Il faut donc calculer toutes les moyennes conditionnelles et toutes les fréquences marginales corréspondantes à la même variable. On aura autant de moyennes et de fréquences pour une variable qu'il y a de modalités dans l'autre.

Dans ce cas, il faut avoir 3 distributions conditionnelles de X et donc 3 moyennes conditionnelles et 3 fréquences marginales correspondantes à X (puisqu'on a 3 modalités dans Y). Comme on doit avoir 4distributions conditionnelles de Y et donc 4 moyennes conditionnelles et 4 fréquences marginales correspondantes à Y (puisqu'on a 4 modalités dans X).

Le développement des calculs se fait comme suit :

-Pour la variable X :

On récapitule les résultats des moyennes conditionnelles et des fréquences marginales correspondantes dans le tableau suivant :

MoyCond .		**Fréq – marg.(f.j)**	
$\overline{x_{j=1}}$	**2.55**	**f.1**	**0.39**
$\overline{x_{j=2}}$	**2.37**	**f.2**	**0.30**
$\overline{x_{j=3}}$	**2.04**	**f.3**	**0.31**
			1.00

On peut vérifier par exemple que :

$$\overline{x_{i/j1}} = \frac{\sum_{i=1}^{4} n_{i1} \times x_i}{n_{\cdot 1}} = \frac{(23 \times 1) + (16 \times 2) + (31 \times 3) + (21 \times 4)}{91} = 2.55$$

Et :$f_1 = \frac{n_{\cdot 1}}{n_{\cdot\cdot}} = \frac{91}{232} = 0.39$

On peut donc vérifier si on calcule la moyenne des ces moyennes conditionnelles pondérées par leurs fréquences marginales correspondantes que :

$$\bar{\bar{x}} = \sum_{j=1}^{3} f.j \times \overline{x_j} = (2.55 \times 0.39) + (2.37 \times 0.30) + (2.04 \times 0.31) = 2.34$$

Ce qui représente le meme résultat qu'on a trouvé dans la question précédente puisque :

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{4} nixi}{n..} = \frac{543}{232} = 2.34$$

-Pour la variable Y :

Nous calculons d'abord la moyennes marginales comme :

$$\bar{\bar{y}} = \frac{\sum_{j=1}^{4} ni \times xi}{n..} = \frac{5029}{232} = 21.68$$

y	**n_j**	**C_j**	**$N_j.*c_j$**
13-20	91	16.5	1501.5
20-25	70	22.5	1575
25-30	71	27.5	1952.5
T	232		5029

Ensuite, les moyennes conditionnelles sont données ci-après :

MoyCond .		**Fréq – marg.(f.j)**	
$\overline{Y_{i=1}}$	22.38	F_1	0.29

$\overline{Y_{i=2}}$	22.89	F_2	0.27
$\overline{Y_{i=3}}$	20.48	F_3	0.26
$\overline{Y_{i=4}}$	20.50	F_4	0.19
			1.00

Là aussi à titre de vérification, on peut montrer par exemple que :

$$\overline{y_{j/i1}} = \frac{\sum_{i=1}^{3} n_{1j} \times cj}{n_{\cdot 1}} = \frac{(23 \times 16.5) + (18 \times 22.5) + (26 \times 27.5)}{67} = 22.38$$

Alors que : $f_1 = \frac{n_{\cdot 1}}{n_{\cdot\cdot}} = \frac{67}{232} = 0.29$

Enfin, le calcul de la moyenne de ces moyennes conditionnelles pondérées par leurs fréquences marginales correspondantes pour la variable Y donne :

$$\bar{\bar{y}} = \sum_{i=1}^{4} f_i \times \bar{y}_i$$

Donc la moyenne marginales de Y est égale à la moyenne de ses moyennes conditionnelles pondérées par ses fréquences marginales.

4- L'analyse de la corrélation des deux variables peut éventuellement se faire en utilisant le coefficient de corrélation R tels que :

$$R = \frac{cov(x.y)}{\sigma_x \sigma_y}$$

Pour cela on donne les résultats suivants de la covariance :

Nij	16,5	22,5	27,5	T
1	159,65	-19,75	-202,77	-63,90
2	28,18	-6,13	-47,49	-25,44
3	-105,98	8,66	49,94	-47,38
4	-180,57	19,06	77,30	-84,22
				-219,94

Avec :

$$cov(x.y) = \frac{\sum_{i=1.j=1}^{3.4} n_{ij}\,(x_i - \bar{x})\big(c_j - \bar{y}\big)}{n_{\cdot\cdot}} = \frac{-219.94}{232} = -0.95$$

Ce qui confirme la dépendance négative opposée des deux variables puisque la covariance est nulle. Cela signifie aussi que lorsque le nombre des enfants à charge augment la dépense mensuelle des salariés diminue et vise versa.

D'où : $R = \frac{COV(X.Y)}{\sigma_x \sigma_y} = \frac{-0.95}{1.08 \times 4.58} = -19.21\%$

5- Les deux variables sont donc négativement corrélées mais faiblement puisque la valeur de R est aussi faible.

Exercice 5 :

1-Nous avons :

La population est constituée par les 350 adhérents d'un club de loisir avec **n..**=350.

Alors que les caractères sont représentés par :

X : « le nombre de visites effectuées par semaine par les adhérents du club ».C'est un caractère quantitatif continu.

Y : « les dépenses en dhs réalisées par les adhérents du club ».C'est un caractère aussi quantitatif continu.

2-Pour compléter le tableau on a besoin de calculer les deux effectifs marquants notamment **n**21=32 et **n**41=24.Pour cela on donne le tableau statistique suivant :

X/Y	500-600	600-650	650-750	750-800	**Marg X**
2-4	23	21	26	4	74
4-8	32	16	16	3	67
8-10	29	18	18	4	69
10-12	24	26	20	6	76
12-14	13	25	23	3	64
Marg Y	121	106	103	20	350

Les effectifs n21 et n41 sont calculés en fonction de l'effectif total **n..**et la somme des autres effectifs simultanés le tableau tel que :

N21=n 2. $-\sum_{j=2}^{4} n$2j =n2.−n22 −n23 −n24

N21=67-16-16-3=32

Pour **n**41on aura : n41=n..-$\sum_{\substack{i=1,j=1 \\ i\neq 4, j\neq 1}}^{54}$ nijnij=24

3.Il s'agit de calculer la moyenne et le mode marginaux de la seule distribution **Y** représentée par « les dépenses effectuées par les adhérents du club ».

Pour ce faire on reprend tableau se référant aux statistiques marginales de Y suivant :

Y	**n.j**	**Cj**	**n.j*cj**	**n.j corrigés**
500-600	121	550	66550	605
600-650	106	625	66250	106
650-750	103	700	72100	103
750-800	20	775	15500	20
T	350		220400	

On vérifie d'abord que la dépense est égale à :

$\bar{\bar{y}}=\frac{\sum_{i=1}^{5} \quad \mathrm{n.i}\times\mathrm{cj}}{n_{..}}=\frac{220400}{350}$=629,71

On dit donc que, indépendamment du nombre de visite effectué par les 350 adhérents du club, ces derniers dépensent par semaine une somme moyenne d'environ 630,dhs.

Pour la dépense modale, et comme on doit le remarquer, on doit passer d'abord par la correction des effectifs (**n.j corrigés)** vu que la première classe présente une amplitude qui représente le double de l'amplitude habituelle. Son effectif se trouve alors divisé par

2.La classe modale devient alors égale à la classe [600-650] et le mode sera définit par l'équation algébrique suivante :

Mo=l0+ [a.$\frac{\mathrm{d1}}{\mathrm{d1+d2}}$]=600+$\left(50.\frac{(106-60,5)}{(106-60,5)+(106-103)}\right)$=646,91

Ce qui signifie aussi que, indépendamment du nombre de visite effectué par les 350 adhérents du club, la majorité de ces derniers dépensent par semaine une somme d'environ 647,00dhs.

4. La distribution concernée dans cette question est représentée par « les dépenses réalisées par des adhérents ayant effectué entre 10 et 12 visites par semaine ».Ce qui est représenté par Yj/i=4 avec la 4° modalité de X qui est réalisée en premier lieu(la condition).

D'où le tableau suivant :

Yj/i=4	**N4j**	**cj**	**N4j*cj**
500-600	24	550	13200
600-650	26	625	16250

650-750	20	700	14000
750-800	6	775	4650
T	76		48100

Ce qui donne une dépense mensuelle moyenne de :

$\overline{yj/i = 4} = \frac{\sum_{j=1}^{4} n4j \times nj}{n4.} = \frac{48100}{76} = 632,89$

Cela représente une dépense par semaine moyenne d'environ 632,89 Dhs pour chaque adhérent pris parmi les 76 qui ont exactement effectué un nombre de visite compris entre 10 et 12 par semaine.

5. La distribution définit dans cette question est représentée par X conditionnée par Y et notamment : « le nombre de visite relatif aux adhérents dont les dépenses sont comprises entre 500 et 600 dhs ».Ce sera alors la distribution Xi/j=1.

Soit donc le tableau suivant présentant les statistiques de la distribution conditionnelle de X :

Xi/j=4	**Ni4**	**ci**	**Ni4*ci**
2-4	23	3	69
4-8	32	6	192
8-10	29	9	261
10-12	24	11	264
12-14	13	12	156
T	121		942

Ce qui donne :

$\overline{xi/j = 4} = \frac{\sum_{i=1}^{5} n\ i4 \times ci}{n\ .4} = \frac{492}{121} = 7,79$

Ce qui signifie que chaque adhérents pris parmi les 121 ayant réalisé une dépense comprise entre 500 et 600 dhs, effectue un nombre moyen de visite d'environ 8 fois par semaine.

6. Par définition on a : $\overline{x2} = \overline{xi/j = 2} = \frac{\sum_{i=1}^{5} ni\,2 \times ci}{n.2}$ c'est la moyenne de X conditionnée par la réalisation à priori de Y. Ce qui définit le nombre moyen des visites effectuée par semaine par les adhérents qui réalisent une dépense comprise entre 600 et 650 dhs.

Le tableau sera récapitulé ainsi :

Xi/j=2	**ni2**	**Ci**	**Ni2*ci**
2-4	21	3	63
4-8	16	6	96
8-10	18	9	162
10-12	26	11	286
12-14	25	12	300
T	106		907

Ce qui donne :

$\overline{x2}=\frac{\sum_{i=2}^{5} ni\,2\times ci}{n.2}=\frac{907}{106}=8{,}56$

Les adhérents ayant réalisé une dépense comprise entre 600 et 650 dhs et dont l'effectif est de 106 adhérents, ont effectué individuellement un nombre moyen de visite d'environ 9 fois par semaine.

De la même manière, on a : $\overline{y1} = \overline{yj/i = 1}=\frac{\sum_{j=1}^{4} n1j\times c\,j}{n1.}$ ce qui représente la moyenne de Y conditionnée par la réalisation à priori de X. En fait, c'est la somme moyenne des dépenses sachant que le nombre des visites effectué par les adhérents est compris entre 2 et 4.

Le tableau sera récapitulé comme :

Y j/i=1	**nij**	**cj**	**N1j*cj**
500-600	23	550	12650
600-650	21	625	13125
650-750	26	700	18200
750-800	4	775	3100
T	74		47075

Ce qui donne :

$\overline{y1}=\frac{\sum_{j=1}^{4} n1j\times cj}{n1.}=\frac{47075}{74}=636{,}15$

Cela représente une dépense par semaine moyenne de 636,15 dhs pour chaque adhérent pris parmi les 74 qui ont exactement effectué un nombre de visite par semaine compris entre 2 et 4.

Exercice 6 :

Nous avons :

La population est constituée par les 12 années de production correspondant à une entreprise. Comme cela est présenté par son service commercial, il s'agit des données individuelles définies par an. Donc tous les effectifs simultanés correspondent à l'unité c-a-d les individus statistiques **i** et **j** sont alors tous identiques. D'où la notation **n..=n=12 années .**

Les caractères utilisés sont définis par :

X : « les investissements réalisés par l'entreprise en question durant les 12 dernières années en 10^6 DH ».

Y : « les performances de production réalisées durant les 12 dernières années en 10^6 DH ».

1. Afin de détecter si un lien existe entre l'investissement (X) et les performances de production (y), nous pouvons tout simplement utiliser la notion de la covariance des deux variables X et Y³.

Pour cela on donne le tableau suivant et dans lequel on récapitule les calculs nécessaires.

Y_i	X_i
10	2
16	4
50	10
120	14
140	18
210	24
260	28
298	34
345	40
450	46
521	52
635	62
3055	**334**

D'où, la covariance suivante :

$cov(x,y) = \frac{\sum_{i=1,J=1}^{3} n_{ij}\,(c_i - \bar{x})(c_j - \bar{y})}{n\cdot}$ (ce qu’on appelle la moyenne des produits moins le produit des deux moyennes).

$\Rightarrow cov\,(x.y) = 3598,11$

Avec :

$$\bar{x} = \frac{\sum_{i=1}^{12} xi}{n} = \frac{334}{12} = 27.83 \; et \; \bar{y} = \frac{\sum_{i=1}^{12} y_1}{n} = \frac{3055}{12} = 254.58$$

Puisque la covariance est positive cela confirme effectivement l’existence d’un lien positif entre les deux variables. Ce qui signifie aussi que lorsque le montant des investissements engagés augmente la performance de production de cette entreprise suit également.

2. selon la méthode des moindres carrés, la relation liant l’investissement et les performances de production est donnée dans une forme linéaire telle que :

Y= a+bY

Prise dans cette équation, les deux variables sont supposées donc être dépendantes linéairement avec une signification particulière des paramètres, dite aussi endogène ou bien à expliquer (y dans ce cas) alors que le second définit le poids de la variable indépendante, dite aussi exogène ou bien explicative (x dans ce cas) sur la précédente.

On sait aussi que les deux paramètres a و b sont estimés parrlaméme méthode comme suit :

$$b = \frac{cov(x,y}{\sigma_x^2} = \frac{3598.11}{x^2 - x^{-2}} = \frac{3598.11}{1118.33 - (27.83)^2} = 10.46$$

Ce qui signifie une sensibilité de y de 10.46 % pour toute augmentation de 1% de x

$$a = y - bx = 254.58 - (10.46 * 27.83) = -36.66$$

La production sera donc impossible dans cette entreprise si un effort d’investissement n’est pas effectué. Le chiffre négatif ici reflète la perte de l’entreprise en millions de dhs lorsque l’investissement est nul

L’équation des moindres carrés ordinaires MCO sera par conséquent donnée par :

$$y = a + bX = -36.66 + 10.46.X$$

Pour prévoir la performance de production avec une dépense d’investissement de 80 000 000.00 dhs, il suffit d’utiliser le résultat de la droit d’ajustement qui nous sert généralement de prévision telle que :

$$Y = a + bX = -36.66 + 10.46(80) = 811.34$$

Si l'entreprise se fise un engagement annuel d'investissement de 80 000 000.00 dhs, la production atteindre alors une valeur annuelle de 811 340 000.00 dhs

Exercice7 :

Nous avons la population est constituée par les 12 mois de ventes correspondant à une entreprise commerciale. Comme il s'agit des donné es individuelles définies par mois alors les unités statistique i et j seront identique. D'où la nation .= n= 12 mois

Les variables utilisées sont représentées par :

X : « les frais d'emballage effectués par l'entreprise en question durant les 12 dernières mois en 10 dh ». nc 'est la variable explicative qui infecte l'évolution de la variable endogène y.

Y : l'analyse de la dépendance des deux variables peut toujours se faire en utilisant la covariance des deux variables x et y

D'où le tableau qui récapitule les calculs nécessaires.

Y_i	X_I	X_i^2	Y_IX_I	Y_i^2
400	11	121	4400	160000
300	8	64	2400	90000
420	12	144	5040	176400
460	14	196	6440	211600
440	13	169	5720	193600
380	11	121	4180	144400
340	10	100	3400	115600
420	20	400	8400	176400
480	25	625	12000	230900
530	30	900	15900	280900
460	23	529	10580	211600
5130	206	4210	92960	2240900

Et tels que les résultats suivants :

D'abord les moyennes :

$$x = \sum_{i=1}^{12} xi = \frac{206}{12}\ 17.17\ et\ Y = \sum_{i=1}^{12} yi = \frac{5130}{12} = 427.50$$

L'entreprise a donc réalisé pendant les 12 derniers mois un montant mensuel moyen de vente d'environ 4 275 000.00 dhs avec un montant mensuel moyen de frais d'emballage d'environ 171700.00 dhs.

D'où , la covariance suivant :

$$cov(x, y) = cov(x, y) = \frac{\sum_{i=1, J=1}(c_i - \bar{x})(c_j - \bar{y})}{n} = 406.49$$

Ce qui signifie que les deux variables sont dépendantes positivement toutes les deux évoluent donc dans le meme sens. Lorsque le montant des frais d'emballage augmente il s'en suit une augmentation aussi des ventes de cette entreprise

2. la fonction linéaire qui donne le montant des ventes lorsqu'on connait les frais d'emballage est représentée par la droit des moindres carrés ordinaires telle que : y= a+bX

Avec les valeurs estimées des deux paramètres a et b :

$$b = \frac{cov(x, y)}{\sigma^2{}_x} = \frac{406.49}{56.02} = 7.26$$

Ce qui signifie que lorsque les frais d'emballage augmentent de 1% les ventes augmentent aussi de 7.26%

$$\text{à} = y^{-} - bx^{-} = 427.5 - (7.26 \times 17.17) = 302.92$$

L'entreprise réalisera donc un chiffre d'affaire mensuel de 3 029 200.00 dhs même si aucun effort en termes d'emballage n'est effectué

D'où l'équation linéaire qui donne le montant des ventes en fonction des frais d'emba   llage donnée par :

$$Y = a + bX = 302.92 + 7.26\,X$$

3. si les frais d'emballage atteindront 55000 dh alors on aura un chiffre de ventes égal à :

$$Y = 302.92 + 7.26. (5.5) = 342.85$$

Soit donc un montant de vente mensuel de 3 428 500.00 dhs.

4. la liaison entre les ventes et les frais d'emballage est automatiquement définit d'ores et déjà par la covariance et la droite de régression donnée dans les questions précédentes. Ce qui renferme en quelque sorte une dépendance entre les deux. Cette relation est fonctionnelle du moment ou la variation de x influence ipso facto l'évolution qui, elle varie dans le mémé sens mais par dans le même rythme (la valeur de b).

On peut éventuellement préciser cette fonctionnalité entre les deux variables avec les coefficients de détermination et de corrélation qui précisent l'intensité de la corrélation tels que :

$$R = \frac{cov(x.y)}{\sigma_x \sigma_y} = 86\%$$

La dépendance est tout à fait forte entre les variables de ventes et de frais d'emballage.