## Corrigé de l'Examen Final 2020

**Exercice 1 :(05 pts):**

Une urne contient 8 boules qui portent les numéros « 0, 0, 0 ,1 , 1 , 1 , 2 , 2 »,On tire simultanément 4 boules.

1. Le nombre de tirages possibles. : $card\ \Omega = C_8^4 = \frac{8!}{4!4!} = \frac{8.7.6.5}{4.3.2} = 70$ **(01)**

2. $A =$ « Avoir des numéros qui forment le nombre 2020 » :

$$Card\ A = C_3^2\ C_2^2 = 3 \Rightarrow P(A) = \frac{card A}{card\,\Omega} = \frac{3}{70}$$

Car, les évènements élémentaires sont équiprobables **(01)**

$B =$ « La somme des numéros tirés est égale à 4»

$$Card\,B = C_3^2\ C_2^2 + C_3^2\ C_3^1\,C_2^1 = 3 + 3.3.2 = 21 \Rightarrow P(B) = \frac{21}{70}$$ **(01)**

3. Calculons $P(A \cap B)$

$A \subset B \Rightarrow A \cap B = A$

$$Card(A \cap B) = Card\ A \Rightarrow P(A \cap B) = P(A) = \frac{3}{70}.$$ **(01)**

Les évènements $A$ et $B$ ne sont pas indépendants car:

$P(A \cap B) = P(A) \neq P(A).P(B)$ **(01)**

**Exercice 2 :(05 pts):**

1. Traduisons les probabilités qui interviennent dans l'énoncé.

5% d'une population est contaminée par le covid19: $P(C) = 0.05 \Rightarrow P(\overline{C}) = 0.95$

Si le patient est contaminé, le test est positif avec une probabilité de 0.98 : $P(T/C) = 0.98$

Si le patient non contaminé, le test est positif avec une probabilité de 0.01 : $P(T/\overline{C}) = 0.01$

**(01)**

2.

$\{C, \overline{C}\}$ forment un système complet car

$C \cap \overline{C} = \Phi\ et\ C \cup \overline{C} = \Omega$ **(01)**

D'après la formule de la probabilité totale :

$$P(T) = P(C)P(T/C) + P(\overline{C})P(T/\overline{C}) = 0.05.\ 0.98 + 0.95.\ 0.01 = 0.0585 \approx 0.06$$

**(01)**

Dans 6% de cas, le test donne un résultat positif sur une population où la contamination est de 5%.

3. Les évènements « T » et « C » sont-ils indépendants ?
$P(T) \approx 0.06$ et comme $P(T/C) = 0.98$ il est claire que $P(T/C) \neq P(T)$.
Les évènements « T » et « C » ne sont pas indépendants. **(01)**

Ou bien :

$$P(T \cap C) = P(C)P(T/C) = 0.05.\ 0.98 = 0.049$$

$$P(T) = 0.06 \ et \ P(C) = 0.05 \Rightarrow P(T).P(C) = 0.06.\ 0.05 = 0.003 \neq P(T \cap C) = 0.049$$

4. Si le test est positif, la probabilité que le patient soit contaminé est :
D'après la formule de BAYES :

$$P(C/T) = \frac{P(C \cap T)}{P(T)} = \frac{P(C)P(T/C)}{P(T)} = \frac{0.05.\ 0.98}{0.058} = 0.84$$ **(01)**

**Exercice 3 :(05 pts):**

Le nombre de frères et sœurs sur un échantillon de 14 étudiants, est donné par la liste suivante :

4 3 3 1 6 0 2 1 2 2 3 5 2 4

1. Quel est le caractère étudié ? Quelle est sa nature ?
**Le caractère c'est le nombre de frères et sœurs, quantitatif discret** **(0.5)+(0.5)**
2. Le tableau statistique des effectifs et des effectifs cumulés.

| $x_i$ | $n_i$ | $\widetilde{n}_i$ | $n_i.x_i$ |
|---|---|---|---|
| 0 | 1 | 1 | 0 |
| 1 | 2 | 3 | 2 |
| 2 | 4 | 7 | 8 |
| 3 | 3 | 10 | 9 |
| 4 | 2 | 12 | 8 |
| 5 | 1 | 13 | 5 |
| 6 | 1 | 14 | 6 |
| | 14 | | 38 |

**(01)**

3. Calculer le mode, la médiane et la moyenne arithmétique. En comparant ces 3 valeurs, quelle conclusion peut-on tirer sur la forme de la distribution ?
$M_0 = 2$ **(0.5)**
Le plus grand nombre d'étudiants sont ceux qui ont 2 frères et sœurs.

$$Med = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_7 + x_8}{2} = 2.5$$ **(0.5)**

**50%** d'étudiants ont le nombre de frères et sœur inférieur à 2,5

$$\overline{X} = \frac{38}{14} = 2.71$$ **(0.5)**

$$M_0 < M_{ed} < \overline{X}$$ **(0.5)**

La distribution présente un étalement vers la droite.

4. Donner l'intervalle interquartile. Que représente cet intervalle ?
L'intervalle interquartile, c'est l'intervalle délimité par le 1° et le 3° quartile, il contient 50% d'observations centrales :

$$\alpha = 0.25 \Rightarrow n\alpha = 3.5 \notin N \Rightarrow Q_1 = x_{E(n\alpha)+1} = x_4 = 2$$ **(0.25)+ (0.25)**

$$\alpha = 0.75 \Rightarrow n\alpha = 10.5 \notin N \Rightarrow Q_3 = x_{E(n\alpha)+1} = x_{11} = 4$$

C'est-à-dire 50% des étudiants ont le nombre de frères et sœurs, compris entre 2 et 4. **(0.5)**

**Exercice 4 :(05 pts):**

**Reprenez les données de l'exercice 3.**

1. Regrouper les données de l'exercice précédent, en 3 classes de même amplitude.

$$e = X_{\max} - X_{\min} = 6 \Rightarrow a = \frac{e}{k} = \frac{6}{3} = 2$$ **(0.5)**

Les classes sont : [0 ,2[, [2 , 4[ et [4, 6]

2. Donner le tableau statistique des effectifs et des effectifs cumulés.

| Les classes | $c_i$ | $n_i$ | $\widetilde{n}_i$ | $n_i.c_i$ |
|---|---|---|---|---|
| [0,2[ | 1 | 3 | 3 | 3 |
| [2, 4[ | 3 | 7 | 10 | 21 |
| [4,6] | 5 | 4 | 14 | 20 |
| | | 14 | | 44 |

**(01)**

3. Calculer le mode, la médiane et la moyenne, comparez avec les valeurs calculées précédemment. Commentez.

$$M_0 \in [2,4[ \Rightarrow M_0 = 2 + \frac{7-3}{(7-3)+(7-4)} 2 = 3.14$$ **(0.5)**

$$M_{ed} \in [2,4[ \Rightarrow M_{ed} = 2 + \frac{7-3}{7} 2 = 3.14$$ **(0.5)**

$$\overline{X} = \frac{44}{14} = 3.14$$ **(0.5)**

$M_0 = M_{ed} = \overline{X}$ : La distribution est symétrique.

Après le regroupement, on remarque qu'on ne trouve pas les mêmes valeurs trouvées précédemment, dans le 1° cas on calcule avec les valeurs exactes, alors qu'avec le regroupement, on applique les formules d'interpolation, c'est-à-dire, des approximations, et par conséquent on perd de l'information. **(0.5)**

4. Donner l'expression de la fonction de répartition dans la classe médiane. Déduire le pourcentage d'étudiants ayant plus de 3 frères et sœurs.

$$x \in [a_{i-1}, a_i[ \; : \; F(x) = \sum_{j=1}^{i-1} f_j + \frac{x - a_{i-1}}{a_i - a_{i-1}} f_i$$

$$x \in [2, 4[ \; : \; F(x) = \frac{3}{14} + \frac{x-2}{4-2} \cdot \frac{7}{14} = 0.21 + \frac{x-2}{4} = \frac{1}{4}x - 0.29$$ **(01)**

Pour $x = 3$ : $F(3) = 0.46$ .C'est-à-dire 46% d'étudiants ont le nombre de frères et sœurs inférieur ou égale à 3, donc 54% d'étudiants ont plus de 3 frères et sœurs.

**(0.5)**

**<u>Exercice 5</u>:(05 points)**
Un certain nombre de personnes souffrant d'obésité suivent un régime d'amincissement. Le tableau suivant donnele nombre de kilogrammes perdus pendant la période de cure suivie.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X durée (en mois) | 2 | 4 | 5 | 12 | 11 | 9 | 8 | 10 | 6 | 8 | 12 | 6 |
| Y nombre de kilogrammes perdus | 2 | 3 | 4 | 9 | 8 | 7 | 6 | 8 | 5 | 7 | 10 | 5 |

On donne :

$$\sum_{i=1}^{12} x_i = 93 \; , \; \sum_{i=1}^{12} y_i = 74 \; , \; \sum_{i=1}^{12} x_i^2 = 835 \; , \; \sum_{i=1}^{12} y_i^2 = 522 \; \sum_{i=1}^{12} x_i \, y_i = 659$$

1. Calculer les moyennes et les variances marginales $\overline{X}$, $\overline{Y}$, $V(X)$ *et* $V(Y)$.

$$\overline{X} = \frac{1}{12} \sum_{i=1}^{12} x_i = \frac{93}{14} = 7.75$$

$$\overline{Y} = \frac{1}{12} \sum_{i=1}^{12} y_i = \frac{74}{12} = 6.17$$

$$V(X)=\frac{1}{12}\sum_{i=1}^{12}x_i^2-\overline{X}^2=\frac{835}{12}-\overline{X}^2=9.52, \qquad \sigma_X=3.08$$

$$V(Y)=\frac{1}{12}\sum_{i=1}^{12}y_i^2-\overline{Y}^2=\frac{522}{12}-\overline{Y}^2=5.43, \qquad \sigma_y=2.33$$

**(0.5)x4=(02)**

2. Calculer la covariance entre X et Y, déduire le coefficient de corrélation linéaire. Commentez le résultat.

$$COV(X,Y)=\frac{1}{12}\sum_{i=1}^{12}x_i\,y_i-\overline{X}.\overline{Y}=\frac{659}{12}-\overline{X}.\overline{Y}=7.09$$ **(0.5)**

Le coefficient de corrélation linéaire :

$$r(X,Y)=\frac{COV(X,Y)}{\sigma_X\sigma_Y}=\frac{7.09}{(3.08).(2.33)}=\frac{7.09}{7.17}=0.98$$ **(0.5)**

**Très forte corrélation linéaire entre les variables.** **(0.5)**

3. Donner l'équation de la droite de régression de Y en X.

$Y=aX+b$ Avec :

$$a=\frac{Cov(X,Y)}{V(X)}=\frac{7.09}{9.52}=0.74 \quad \text{Et} \quad b=\overline{Y}-a\overline{X}=0.4 \quad \text{alors :}$$

$Y=0.74X+0.4$ **(01)**

4. Donner une estimation du nombre de kilogrammes perdus après une cure 13 mois.

$Y=0.74X+0.4$

Pour $X=13$ alors $Y=0.74(13)+0.4=10.02\ kg$ **(0.5)**

# Corrigé de l'examen final Proba/Stat 2018 (MI)

**Problème :** (14 points)

Le montant du loyer exprimé en milliers de dinars (Y)

Le nombre de pièces (X)

| X\Y | [25-35[ 30 | [35-45[ 40 | [45-55[ 50 | [55-65[ 60 | $n_{i.}$ | |
|---|---|---|---|---|---|---|
| 2 | 5 300 | 2 160 | 0 | 0 | 7 | |
| 3 | 3 270 | 7 840 | 0 | 0 | 10 | |
| 4 | 0 | 1 160 | 5 1000 | 3 720 | 9 | |
| 5 | 0 | 0 | 2 500 | 2 600 | 4 | |
| $n_{.j}$ | 8 | 10 | 7 | 5 | 30 | 4550 |

**Partie A : (03 points)**

1- La nature du caractère : quantitatif discret. (0.25)

2- La distribution marginale de la variable X. (0.25) +(0.5)

| X | $n_{i.}$ | $\tilde{n}_{i.}$ | $n_{i.}x_i$ | $n_{i.}x_i^2$ |
|---|---|---|---|---|
| 2 | 7 | 7 | 14 | 28 |
| 3 | 10 | 17 | 30 | 90 |
| 4 | 9 | 26 | 36 | 144 |
| 5 | 4 | 30 | 20 | 100 |
| | | | 100 | 362 |

3- le mode $M_0 = 3$ (0.5)

la médiane : $M_{ed} = \frac{x_{15}+x_{16}}{2} = 3$ (0.5)

4- la moyenne $\overline{X} = \frac{100}{30} = 3.33$ . $V(X) = \frac{362}{30} - (3.33)^2 = 0.97$ . $\sigma_X = 0.98$ (0.5)+ (0.5)

**Partie B : (05 points)**

1- La nature du caractère Y : quantitatif continu. (0.25)

2- La distribution marginale de Y. (0.25)+(0.5)

| Y | $n_{.j}$ | $\tilde{n}_{.j}$ | $n_{.j}.y_j$ | $n_{.j}.y^2_j$ |
|---|---|---|---|---|
| [25-35[ 30 | 8 | 8 | 240 | 7200 |
| [35-45[ 40 | 10 | 18 | 400 | 16000 |
| [45-55[ 50 | 7 | 25 | 350 | 17500 |
| [55-65[ 60 | 5 | 30 | 300 | 18000 |
| | 30 | | 1290 | 58700 |

3- Le nombre de familles qui dépensent plus de 35000DA est :10+7+5=30-8=22 alors le pourcentage est $p = \frac{22}{30}.100\% = ????$ (0.5)

4- La courbe cumulative et l'intervalle interquartile graphiquement : (0.5)+ (0.5)

5- $M_0 \in [35,45[$, $M_0 = 35 + \frac{2}{2+3}.10 = 39$ (0.5)

$Med \in [35,45[$, $M_{ed} = 35 + \frac{15-8}{10}.10 = 42$ (0.5)

$\overline{Y} = \frac{1290}{30} = 43$

(0.5)

Les trois valeurs calculées : $M_0 < M_{ed} < \overline{Y}$. étalement à droite (0.25)

On peut conclure que la distribution présente un étalement à droite.

(0.25)

On peut le confirmer par le coefficient d'asymétrie : $\delta = \frac{\mu_3(Y)}{\sigma_Y^3}$ Qui sera positif.

6- La variance: $V(Y) = \frac{58700}{30} - (43)^2 = 107.66$ (0.5)

l'écart type de Y : $\sigma_Y = 10.37$



**Partie C : (06 points)**

1- Les dépenses des familles de 4 personnes : (0.5)

| Y | [25-35[ 30 | [35-45[ 40 | [45-55[ 50 | [55-65[ 60 | |
|---|---|---|---|---|---|
| $f_{Y/X=4}$ | 0 | $1/9$ | $5/9$ | $3/9$ | |
| $y.f_{Y/X=4}$ | 0 | $40/9$ | $250/9$ | $180/9$ | $m_{Y/X=4} = 470/9$ |

$m_{Y/X=4} = 52.22$. Les dépenses moyennes des familles de 4 personnes est 52220 DA (0.5)

2- $f_{11} = \frac{5}{30} \neq \frac{7}{30} \cdot \frac{8}{30} = f_{1.} * f_{.1}$. Donc X et Y ne sont pas indépendantes. (0.5)

3- La covariance entre les variables X et Y :

$\text{cov}(X,Y) = \frac{1}{n} \sum \sum n_{ij} x_i y_j - \overline{X}.\overline{Y} = \frac{4550}{30} - (3.33).(43) = 8.47.$ (01)+(01)

4- Le coefficient de corrélation linéaire. : $r(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = 0.83$. Forte corrélation.

(0.5)+ (0.5)

L'équation de la droite de régression de Y en X : Y =8.7 X+13.9 (0.5)+ (0.5)

5- Si X=6, alors Y= 66.28. une famille de 4 personnes dépense 66280 DA . (0.5)

**Exercice 1 :** (03 points)

1. S= « l'étudiant réussit le module de statistique »
A= « l'étudiant réussit le module d'analyse »

$P(S) = 0.6$ , $P(A) = 0.4$ $P(S \cup A) = 0.7$ (0.5)

2. La probabilité que l'étudiant réussisse les deux modules : (0.5)

$$P(S\cup A)=P(S)+P(A)-P(S\cap A)\Rightarrow$$

$$P(S\cap A)=P(S)+P(A)-P(S\cup A)=0.6+0.4-0.7=0.3$$

30 % des étudiants réussissent les deux modules.

3. En stat et pas en analyse :

$$P(S-A)=P(S)-P(S\cap A)=0.6-0.3=0.3 \qquad \textbf{(01)}$$

4. Quelle est la probabilité d'avoir un seul module ?

$$P(S\Delta A)=P(S).+P(A)-2P(S\cap A)=0.6+0.4-0.6=0.4 \qquad \textbf{(01)}$$

**<u>Exercice 2 :</u>** (03 points)

1-

A : « l'étudiant choisit l'itinéraire A » $P(A)=\frac{1}{3}$ **(0.5)**

B : « l'étudiant choisit l'itinéraire B » $P(B)=\frac{1}{4}$

C : « l'étudiant choisit l'itinéraire C » $P(C)=\frac{1}{12}$

D : « l'étudiant choisit l'itinéraire D »

R : « l'étudiant arrive en retard »

$P(R/A)=\frac{1}{20}$ , $P(R/B)=\frac{1}{10}$ , $P(R/C)=\frac{1}{5}$ et $P(R/D)=0$ **(0.5)**

2- La probabilité que l'étudiant choisisse l'itinéraire D :

$\{A,B,C,D\}$ forment un système complet ;

$$P(A)+P(B)+P(C)+P(D)=1$$

$$P(D)=1-[P(A)+P(B)+P(C)]=1-\left[\frac{1}{3}+\frac{1}{4}+\frac{1}{12}\right]=\frac{1}{3} \qquad \textbf{(0.5)}$$

3- **D'après la formule de la probabilité totale:**

$$P(R)=P(A)P(R/A)+P(B)P(R/B)+P(C)P(R/C)+P(D)P(R/D)$$

$$=\frac{1}{3}.\frac{1}{2}+\frac{1}{4}.\frac{1}{10}+\frac{1}{12}.\frac{1}{5}+\frac{1}{3}.0= \qquad \textbf{(01)}$$

4- **Formule de Bayes :** $P(C/R)=\dfrac{P(C)P(R/C)}{P(R)}=$ **(0.5)**

Nom : ........................................ Prénom : ........................................ Matricule : ..........................

**Exercice. 1** [04 *pts*] *(10 minutes)*

*Répondre par **Vrai** ou **Faux** aux affirmations suivantes et corriger en cas d'erreur.*

① *La covariance est toujours positive.*

**Réponse** Faux, la covariance est définie dans ℝ 0,5

② *On peut calculer dix (10) déciles.*

**Réponse** Faux, on peut en calculer neuf seulement 0,5

③ *Si le coefficient de corrélation vaut 1, alors les deux droites de régession sont confondues.*

**Réponse** Vrai 0,25

④ *Si les notes obtenues à un examen d'un groupe de 30 élèves sont toutes augmentées d'un (01) point, alors la variance sera augmentée de 1.*

**Réponse** Faux, la variance ne changera pas 0,5

⑤ *Le mode est toujours plus grand que la médiane.*

0,5 **Réponse** Faux, il n'y a pas de relation d'ordre fixe entre eux

⑥ *En regroupant en classes les observations d'une série statistique à caractère continu, on en récupère le maximum d'information possible.*

**Réponse** Faux, plutôt on perd un peu d'information 0,5

⑦ *Si le coefficient d'asymétrie est nul, la distribution statistique présente une oblique à gauche.*

**Réponse** Faux, elle est plutôt symétrique 0,5

⑧ *Le cinqième décile est la médiane.*

**Réponse** Vrai 0,25

⑨ *Le 7ème décile appartient à l'intervalle interquartile.*

**Réponse** Vrai 0,25

⑩ *Si X et Y sont indépendantes l'une de l'autre alors le coefficient de corrélation entre elles est nul.*

**Réponse** Vrai 0,25

**Exercice. 2** [11.5 *pts*] *(1 heure) Au terme d'une année universitaire, on a pris note du nombre d'absences (variable X) de chacun des étudiants d'un groupe de 30, ainsi que de leur moyenne annuelle en Mathématiques (variable Y) :*

| X | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Y | 17 | 15 | 15 | 14.5 | 14.33 | 14 | 13.25 | 12.5 | 12 | 12 | 12.33 | 11.83 | 11 | 10.5 | 10.5 |
| X | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 4 | 4 |
| Y | 10.33 | 10 | 11.83 | 10.5 | 10 | 8 | 7.5 | 7.25 | 8.5 | 6.75 | 5.25 | 3.75 | 3 | 2.75 | 2 |

**Partie 1 : *On s'intéresse au caractère Y seul.***

1. *Quelle est la nature de Y ?*

   **Réponse** Y est quantitatif continu. 0,25

2. *Compléter le tableau statistique suivant :*
   **Réponse**

| $Y$ | $[2-7[$ | $[7-12[$ | $[12-17]$ |
|---|---|---|---|
| $n_j$ | 6 | 13 | 11 |

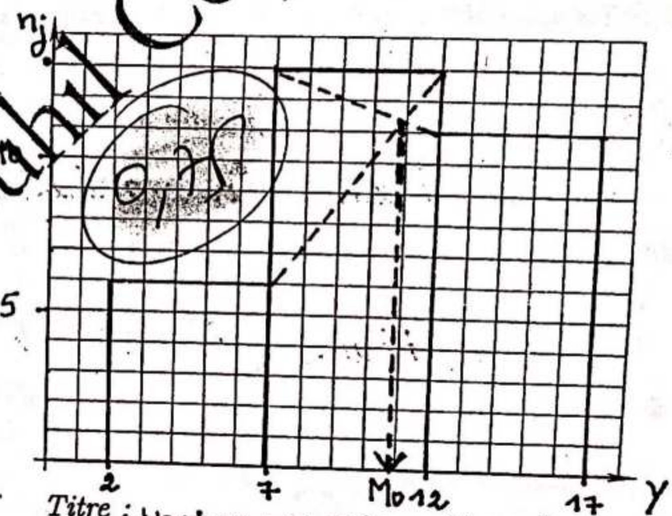
0,5

3. *Déterminer graphiquement une valeur approchée du mode.* ▷

   **Réponse**

   0,25

   $M_o \simeq 10,9$




Titre : Histogramme des effectifs en fonction de Y.

4. *Retrouver la valeur du mode par le calcul.*

   **Réponse** La classe modale est $[a_1 - a_2[ = [7 - 12[$. Donc,

   $M_o = a_1 + \left(\frac{\delta_1}{\delta_1 + \delta_2}\right) l = 7 + \frac{(13-6)}{(13-6)(13-11)} \cdot 5 = 10,89$ 0,5

5. *Calculer la valeur de la médiane et en donner une interprétation statistique.*

   **Réponse** La classe médiane est $[a_1 - a_2[ = [7 - 12[$. Donc,

   0,5 $M_e = a_1 + \frac{n/2 - \tilde{n}_1}{n_2} \times l = 7 + \frac{15-6}{13} \times 5 = 10,46$

   Interprétation :

   Au moins 50% des étudiants ont eu une moyenne annuelle plus petite ou égale à 10,46 0,5

6. **Calculer la moyenne annuelle moyenne du groupe d'étudiants en question.**

**Réponse** $\bar{Y} = \frac{1}{n} \sum_{j=1}^{3} n_j \, y_j = \frac{310}{30} = 10{,}33$ 0,5

7. **Conclure sur la forme de la distribution de Y.**

0,5

**Réponse** On a $\bar{Y} < Me < Mo$, donc la distribution de Y présente un étalement vers la gauche (oblique à droite).

## Partie 2 : On s'intéresse aux deux caractères X et Y simultanément.

1. *Compléter le tableau de contingence suivant :*

| X \ Y | [2 − 7[ | [7 − 12[ | [12 − 17] | |
|---|---|---|---|---|
| | 4,5 | 9,5 | 14,5 | $n_{i\bullet}$ |
| 0 | | | 10 | 10 |
| 1 | | 6 | 1 | 7 |
| 2 | | 6 | | 6 |
| 3 | 3 | 1 | | 4 |
| 4 | 3 | | | 3 |
| $n_{\bullet j}$ | 6 | 13 | 11 | n=30 |

1



2. *X et Y sont-elles indépendantes ? (justifier)*

0,5

**Réponse** X et Y ne sont pas indépendantes l'une de l'autre car pour $(i,j) = (1,1)$, on a $f_{11} = 0$ mais $f_{1\bullet} \cdot f_{\bullet 1} = \frac{6}{30} \times \frac{10}{30} \neq 0$

3. *Donner la distribution conditionnelle de X, sachant que $Y \in [7 - 12[$.*

**Réponse** *On doit calculer* $f_{i/j=2} = \frac{n_{i2}}{n_{\bullet 2}}$ pour $i = 1, \ldots, 5$ 0,25

*On obtient alors le tableau suivant :*

1

| X | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $f_{X/Y \in [7-12[}$ | 0 | 6/13 | 6/13 | 1/13 | $\boxed{0}$ |

4. *Que signifie la valeur encadrée dans le tableau ci-dessus ?*

0,5

**Réponse** Elle veut dire que parmi les étudiants ayant obtenu une moyenne annuelle comprise entre 2 et 7, aucun n'a fait quatre absences.

5. *Calculer la covariance entre X et Y.*

**Réponse** $COV(X,Y) = \overline{XY} - \overline{X}\cdot\overline{Y}$ tels que :

* $\overline{XY} = \frac{1}{n}\sum_{i=1}^{5}\sum_{j=1}^{3} n_{ij}\, x_i\, y_j = \frac{308,5}{30} = 10,28$

* $\overline{X} = \frac{1}{n}\sum_{i=1}^{5} n_{i\cdot}\, x_i = \frac{43}{30} = 1,43$

* $\overline{Y} = 10,33$ (déja calculée)

Donc, $COV(X,Y) = 10,28 - 1,43 \times 10,33 = -4,49$

1

6. *Calculer le coefficient de corrélation et interpréter la valeur obtenue.*

**Réponse** $\rho_{xy} = \frac{COV(X,Y)}{\sigma_x \cdot \sigma_y}$ tels que :

$\sigma_x = \sqrt{V(X)} = \sqrt{\frac{1}{n}\sum_{i=1}^{5} n_{i.}\, x_i^2 - \bar{x}^2} = \sqrt{115/30 - (1,43)^2} = \sqrt{1,78} = 1,33$.

$\sigma_y = \sqrt{V(Y)} = \sqrt{\frac{1}{n}\sum_{j=1}^{3} n_{.j}\, y_j^2 - (\bar{y})^2} = \sqrt{\frac{3607,5}{30} - (10,33)^2} = \sqrt{13,54} = 3,67$.

Donc, $\rho_{xy} = \frac{-4,49}{1,33 \times 3,67} = -0,92$

Interprétation : On a $R = \rho_{xy}^2 = 0,85$ est beaucoup plus grand que la valeur théorique qui vaut 0,13. La corrélation linéaire entre X et Y est donc significative et l'ajustement linéaire est justifié. Et comme $\rho_{xy} < 0$, on déduit que X et Y varient dans deux sens opposés. i.e. un étudiant qui fait beaucoup d'absences obtient une faible moyenne et vice versa.

0,5

1



7. *Donner l'équation de la droite de régression de Y en X.*

**Réponse** $(D_{Y/X}) : y = ax + b$ tels que :

$a = \frac{COV(X,Y)}{V(X)} = \frac{-4,49}{1,78} = -2,52$ et $b = \bar{y} - a\bar{x} = 10,33 + 2,52 \times 1,43 = 13,93$.

$\Rightarrow (D_{Y/X}) : y = -2,52x + 13,93$.

1

8. *Si un étudiant s'absente 5 fois, quelle moyenne annuelle de Math lui prévoit-on ?*

**Réponse** On lui prévoit une moyenne que l'on estime par :

$y^* = -2,52 \times 5 + 13,93 = 1,33$

0,5

**Exercice. 3** [4.5 pts] *(20 minutes) Dans un entrepôt, un système d'alarme-incendie est mis en place. Le système est constitué de deux détecteurs de fumée. En cas d'incendie, la probabilité de déclenchement du premier détecteur est de 0.98 et la probabilité de déclenchement du second est de 0.97, tandis que la probabilité que les deux se déclenchent simultanément est estimée dans ce cas à 0.96.*

*La probabilité qu'un incendie se déclare dans l'entrpôt est estimée à 0.003.*

*La probabilité qu'un détecteur donne une fausse alerte est estimée à 0.01 pour le premier et 0.02 pour le second, tandis que la probabilité qu'une fausse alerte est déclarée par les deux en même temps est nulle.*

1. *Soient les évenements :*

$$D_j = \text{"le détecteur numéro } j \text{ se déclenche"}, \; j \in \{1,2\},$$

$$I = \text{"un incendie se déclare"}$$

*Les évenements $D_1$ et $D_2$ sont-ils indépendants ? (justifier et ne pas arrondir les valeurs calculées)*

***Réponse***

0,5 On sait que $D_1 \perp\!\!\!\perp D_2 \iff P(D_1 \cap D_2) = P(D_1) \times P(D_2)$, et on a :

0,5 * $P(D_1) = P(D_1/I)P(I) + P(D_1/\bar{I}).P(\bar{I})$

$= 0,98 \times 0,003 + 0,01 \times 0,997 = 0,01291.$

0,5 * $P(D_2) = P(D_2/I)P(I) + P(D_2/\bar{I}).P(\bar{I})$

$= 0,97 \times 0,003 + 0,02 \times 0,997 = 0,02285$

0,5 * $P(D_1 \cap D_2) = P(D_1 \cap D_2/I).P(I) + P(D_1 \cap D_2/\bar{I}).P(\bar{I})$

$= 0,96 \times 0,003 + 0 \times 0,997 = 0,00288.$

0,25 Comme $P(D_1).P(D_2) = 0,0002949 \neq P(D_1 \cap D_2)$, on déduit que $D_1$ et $D_2$ ne sont pas indépendants.

2. *Quelle est la probabilité que le système d'alarme se déclenche ?*

***Réponse***

0,5 $P(D_1 \cup D_2) = P(D_1) + P(D_2) - P(D_1 \cap D_2)$

$= 0,01291 + 0,02285 - 0,00288 = 0,03288$

3. *Sachant que le système ne se déclenche pas, quelle est la probabilité qu'un incendie se déclare ?*

***Réponse***

0,5 $P(I / \overline{D_1 \cup D_2}) = \dfrac{P(\overline{D_1 \cup D_2} \cap I)}{P(\overline{D_1 \cup D_2})}$

0,75 Avec $P(\overline{D_1 \cup D_2} \cap I) = P(\overline{D_1 \cup D_2} / I)\, P(I)$

$= [1 - P(D_1 \cup D_2 / I)]\, P(I)$

$= [1 - P(D_1/I) - P(D_2/I) + P(D_1 \cap D_2 / I)] \times P(I)$

$= [1 - 0{,}98 - 0{,}97 + 0{,}96] \times 0{,}003$

$= 0{,}0003$

0,25 et $P(\overline{D_1 \cup D_2}) = 1 - P(D_1 \cup D_2) = 1 - 0{,}03288 = 0{,}96712$

0,25 D'où, $P(I / \overline{D_1 \cup D_2}) = \dfrac{0{,}0003}{0{,}96712} = 0{,}0003101$



---

*Table de décision pour le coefficient de détermination $R^2$.*

| $n$ | $R^2$ | $n$ | $R^2$ |
|---|---|---|---|
| 3 | 99 | 22 | 18 |
| 4 | 90 | 24 | 16 |
| 5 | 77 | 26 | 15 |
| 6 | 66 | 28 | 14 |
| 7 | 57 | 30 | 13 |
| 8 | 50 | 35 | 11 |
| 9 | 44 | 40 | 10 |
| 10 | 40 | 50 | 8 |
| 11 | 36 | 60 | 6 |
| 12 | 33 | 70 | 6 |
| 13 | 31 | 80 | 5 |
| 14 | 28 | 90 | 4 |
| 15 | 26 | 100 | 4 |
| 16 | 25 | 125 | 3 |
| 17 | 23 | 150 | 3 |
| 18 | 22 | 200 | 2 |
| 19 | 21 | 300 | 1 |
| 20 | 20 | 500 | 1 |

USTHB Faculté de mathématique
Module Proba/Stat
2° Année ST-Section J- GM

28/01/2016

## Examen de rattrapage

### Exercice 1 (5 points)

On a relevé le nombre d'interventions par heure, d'une équipe de secours durant 50 heures consécutives. Le tableau suivant donne les effectifs selon le nombre d'interventions enregistrées.

| Nombre d'interventions | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Effectifs | 6 | 10 | 16 | 10 | 6 | 2 |

1. Quelle est la population statistique concernée par cette étude ?
2. Quel est le caractère étudié ?
3. Quelle est la valeur du mode ?
4. Calculer puis interpréter la valeur de la médiane
5. Calculer la moyenne arithmétique ?
6. Calculer la variance
7. Tracer la courbe cumulative des fréquences
8. Déterminer graphiquement l'intervalle interquartile. Que représente cet intervalle.



### Exercice 2 : (5 points)

La distribution des salaires noté Y, de 50 employés d'une entreprise est donnée par le tableau suivant :

| Salaire (en $10^4$ DA) | [3 , 5[ | [5 , 7[ | [7 , 9[ | [9 , 11[ | [11 ,13[ |
|---|---|---|---|---|---|
| Nombre d'employés | 4 | 6 | 20 | 12 | 8 |

1. Quelle est la nature du caractère Y.
2. Quel est le pourcentage d'employés qui perçoivent un salaire supérieur à 70.000DA ?
3. Calculer le mode, la médiane et le salaire moyen.
4. Tracer la courbe cumulative et donner graphiquement l'intervalle qui contient 90% des valeurs centrales.
5. La distribution est-elle symétrique ? par quel paramètre peut-on le confirmer ?



### Exercice 3 : (5 points)

On a relevé entre les années 2000 et 2010, la production de blé (X en tonnes/ha) et le nombre de jour de pluies par an (Y). les résultats sont donnés dans le tableau suivant :

| Valeurs de X | 13 | 13.4 | 14.5 | 15.2 | 16.5 | 16.5 | 17.2 | 17.4 | 18 | 18.3 |
|---|---|---|---|---|---|---|---|---|---|---|
| Valeurs de Y | 40 | 40 | 50 | 50 | 50 | 60 | 60 | 70 | 70 | 60 |

On donne :

$$\sum_{i=1}^{10} x_i = 160,\ \sum_{i=1}^{10} y_i = 550,\ \sum_{i=1}^{10} x_i^2 = 2591{,}84,\ \sum_{i=1}^{10} {y_i}^2 = 31300\ ,\ \sum_{i=1}^{10} x_i . y_i = 8964$$

1. Déterminer le coefficient de corrélation linéaire entre Y et X. Commenter le résultat.
2. Donner l'équation de la droite de régression de Y en X.
3. Peut-on prévoir la production de blé pour 75 jours de pluies ?
4. Le modèle est-il adapté pour prévoir la production de blé, dans le cas où il n'y a eu que 10 jours de pluies par an ?

**Exercice 4 :** (2 points)

1- Montrer que si $A$ et $B$ sont indépendants alors $\overline{A}$ et $B$ le sont aussi.

2- Soient $A$ et $B$ deux évènements tels que $P(A)=0.4$, $P(B)=0.5$ et $P(A/B)=0.6$, calculer $P(A\cup B)$ et $P(B/\overline{A})$.

**Exercice 5 :** (3 points)

Un chef d'entreprise spécialisée dans l'assemblage de ventilateurs est approvisionné en moteur électrique par deux fournisseurs. La compagnie A fournit 90% des moteurs, la compagnie B fournit le reste.

Ce chef d'entreprise sait que 5% des moteurs fournis par A sont défectueux et 3 % des moteurs fournis pas B sont défectueux.

1. Décrire correctement les évènements et leurs probabilités.
2. Ce chef d'entreprise vérifie un ventilateur au hasard, Quelle est la probabilité que son moteur se trouve défectueux.
3. Le moteur étant défectueux, quelle est la probabilité qu'il provienne de la compagnie A.

2° Année ST-Section J- GM

## Corrigé de l'examen de rattrapage 2016

**Exercice 1** (4 points)

| $x_i$ | $n_i$ | $\tilde{n}_i$ | $\tilde{f}_i$ | $n_i.x_i$ | $n_i.x_i^2$ |
|---|---|---|---|---|---|
| 0 | 6 | 6 | | 0 | 0 |
| 1 | 10 | 16 | | 10 | 10 |
| 2 | 16 | 32 | | 32 | 64 |
| 3 | 10 | 42 | | 30 | 90 |
| 4 | 6 | 48 | | 24 | 96 |
| 5 | 2 | 50 | | 10 | 50 |
| | 50 | | | 106 | 310 |

1. le caractère étudié : le nombre d'interventions par heure .Sa nature : quantitatif discret.
2. Le Mode $M_0 = 2$
3. la médiane $M_{ed} = \frac{x_{25} + x_{26}}{2} = 2.$
4. $\overline{X} = \frac{106}{50} = 2.12$
5. la variance $V(X) = \frac{310}{50} - (2.12)^2 = 1.7$
6. La courbe cumulative des fréquences
7. Graphiquement l'intervalle interquartile :
   Il contient 50% de valeurs centrales.

**Exercice 2 :** (4 points)
La distribution des salaires noté Y, de 50 employés d'une entreprise est donnée par le tableau suivant :

| Les classes | | $n_i$ | $\tilde{n}_i$ | $\tilde{f}_i$ | $n_i.x_i$ |
|---|---|---|---|---|---|
| [3,5[ | 4 | 4 | 4 | 0.08 | 16 |
| [5, 7[ | 6 | 6 | 10 | 0.2 | 36 |
| [7,9[ | 8 | 20 | 30 | 0.6 | 160 |
| [9,11[ | 10 | 12 | 42 | 0.84 | 120 |
| [11, 13[ | 12 | 8 | 50 | 1 | 96 |
| | | 50 | | | 428 |

1. La nature du caractère Y : quantitatif continu

2. Le pourcentage d'employés qui perçoivent un salaire supérieur à 70.000DA :

20+12+8=40 donc $P=\frac{40}{50}.100\%=80\%$

3. Le mode $M_0 \in [7,9[ \Rightarrow M_0=7+\frac{20-6}{(20-6)+(20-12)}2=7.27$

La médiane $M_{ed} \in [7,9[ \Rightarrow M_{ed}=7+\frac{25-10}{20}2=8.5$

La moyenne $\overline{X}=\frac{428}{50}=8.56$.

4. La courbe cumulative.
   La médiane graphiquement :
5. $M_0 < M_{ed} < \overline{X}$ : Alors on a un étalement vers la droite, et on pourra confirmer par le calcul du coefficient d'asymétrie qui sera positif.

**Exercice 3 :** (6 points)

La production de blé (X en tonnes/ha) et le nombre de jour de pluies par an (Y).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Valeurs de X | 13 | 13.4 | 14.5 | 15.2 | 16.5 | 16.5 | 17.2 | 17.4 | 18 | 18.3 |
| Valeurs de Y | 40 | 40 | 50 | 50 | 50 | 60 | 60 | 70 | 70 | 60 |

$$\sum_{i=1}^{10} x_i=160,\ \sum_{i=1}^{10} y_i=550,\ \sum_{i=1}^{10} x_i^2=2591{,}84,\ \sum_{i=1}^{10} y_i^{\,2}=31300,\ \sum_{i=1}^{10} x_i.y_i=8964$$

1. Le coefficient de corrélation linéaire entre Y et X.

$$\overline{X}=\frac{160}{10}=16,\quad V(X)=\frac{2591.84}{10}-\overline{X}^2=3.18,\qquad \sigma_X=1.78$$

$$\overline{Y}=\frac{550}{10}=55\quad V(Y)=\frac{31300}{10}-\overline{Y}^2=105,\qquad \sigma_z=10.25$$

$$COV(X,Y)=\frac{8964}{10}-\overline{X}.\overline{Y}=16.4$$

$$r(X,Y)=\frac{COV(X,Y)}{\sigma_X\sigma_Y}=\frac{16.4}{(1.78).(10.25)}=0.899\approx0.9$$

**Très forte corrélation linéaire.**

2. L'équation de la droite de régression de Y en X :

$a=\frac{Cov(X,Y)}{V(X)}=5.16$ Et $b=\overline{Y}-a\overline{X}=-27.51$ alors :

$Y=5.16X-27.51$

3. Pour prévoir la production de blé pour 75 jours de pluies, on détermine l'équation de la droite de régression de X en Y :

$$\alpha=\frac{Cov(X,Y)}{V(Y)}=0.16 \quad \text{Et} \quad \beta=\overline{X}-\alpha\overline{Y}=7.41 \ \text{alors :}$$

$$X=0.16Y+7.41$$

Si $Y=75$ jours de pluies alors la production de blé est : $X=19.41$ tonnes/ha

---

**<u>Exercice 4 :</u>** (3 points)

1- Si $A$ *et* $B$ sont indépendants alors $P(A\cap B)=P(A).P(B)$.

Montrons que: $P(\overline{A}\cap B)=P(\overline{A}).P(B)$.

$$P(\overline{A}\cap B)=P(B-A)=P(B)-P(A\cap B)=P(B)-P(A)P(B)$$
$$=P(B)[1-P(A)]=P(\overline{A})P(B)$$

2. Soient $A$ et $B$ deux évènements tels que $P(A)=0.4$, $P(B)=0.5$ et $P(A/B)=0.6$,

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$
$$=P(A)+P(B)-P(B).P(A/B)=0.4+0.5-0.5.0.6=0.6$$

$$P(B/\overline{A})=\frac{P(\overline{A}\cap B)}{P(\overline{A})}=\frac{P(B-A)}{1-P(A)}=\frac{P(B)-P(A\cap B)}{1-P(A)}$$
$$=\frac{P(B)-P(B).P(A/B)}{1-P(A)}=\frac{P(B)[1-P(A/B)]}{1-P(A)}=\frac{0.5\,[1-0.6]}{1-0.4}=0.33$$

**<u>Exercice 5 :</u>** (3 points)

1. On donne les évènements et leurs probabilités.

Les évènements : A= « le moteur provient du fournisseur A »
B= « le moteur provient du fournisseur B »
D= « Le moteur est défectueux »

$P(A)=0.9$, $P(D/A)=0.05$ .
$P(B)=0.1$, $P(D/B)=0.03$ ,

2. Ce chef d'entreprise vérifie un ventilateur au hasard, la probabilité que son moteur se trouve défectueux: $\{A,B\}$ forment un système complet d'évènements, la formule de la probabilité totale nous donne : $P(D)=P(A)\,P(D/A)+P(B).P(D/B)=0.048$

3. Le moteur étant défectueux, la probabilité qu'il provienne de la compagnie A.

$$P(A/D)=\frac{P(A)P(D/A)}{P(D)}=\frac{0.1.0.03}{0.048}=0.0625$$

(Formule de Bayes).

U.S.T.H.B
Faculté de Mathématiques
1ère année – Lic- Inf – sections 1 et 2

Année universitaire 2012/2013

## Epreuve finale du module de Statistiques

**Exercice 1 (6 points) :**

50 femmes ont été soumises à une enquête. Pour chacune d'entre elles, on a noté le nombre d'enfants ayant obtenu le baccalauréat. Les résultats sont regroupés dans le tableau suivant

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 0 | 0 | 3 | 2 | 1 | 2 | 3 | 2 | 1 | 1 | 0 | 2 | 1 | 2 | 0 | 3 | 1 | 1 | 2 | 2 | 3 | 2 | 1 |
| 3 | 2 | 3 | 2 | 0 | 1 | 2 | 4 | 1 | 4 | 4 | 0 | 4 | 2 | 4 | 0 | 2 | 1 | 1 | 0 | 2 | 0 | 1 | 2 | 3 |

1. Donner le tableau de la distribution des fréquences de cette série.
2. Donner une représentation graphique adéquate.
3. Donner la fonction cumulative puis tracer la courbe cumulative.
4. Déterminer le mode, la médiane et les quartiles.
5. Calculer la variance et l'écart interquartiles.

**Exercice 2 (14 points) :**

On a relevé le poids Y et la taille X d'un certain nombre d'écoliers. Les résultats sont donnés dans le tableau de contingence suivant où il manque quelques données.

| X \ Y | [25 ; 35[ | [35 , 45[ | [45 , 55[ | Total |
|---|---|---|---|---|
| [85 , 95[ | ? | 0 | 0 | 6 |
| [95 , 105[ | 6 | ? | 0 | 11 |
| [105 , 115[ | ? | 3 | ? | 7 |
| [115 , 125[ | ? | 2 | ? | ? |
| [125 , 135[ | 0 | ? | 2 | 2 |
| Total | 15 | 10 | 5 | ? |

1. Retrouver les effectifs manquants.
2. Donner les distributions marginales des variables statistiques X et Y.
3. Calculez les pourcentages suivants :
   a) celui des enfants dont la taille est inférieure à 120 cm.
   b) celui des enfants dont le poids est compris entre 45 et 52 kilogrammes.
4. Calculer pour la variable statistique X la médiane et le deuxième décile.
5. Calculer les moyennes et les variances des variables X et Y.
6. Donner $f_{\cdot 3}$ , $f_{4,3}$ et $f_{(Y\in[35,45[/X\in[105,115[)}$. Donner la distribution conditionnelle de Y sachant que $X\in[105 , 115[$. En déduire la moyenne et la variance de cette distribution conditionnelle.
7. Y'a-t-il une forte corrélation linéaire entre X et Y ? Commenter le résultat.
8. Déterminer l'équation de la droite de régression linéaire de Y en X.

U.H.B, Faculté de Maths. Juin 2013 — MI - Sects 1 et 2

# Corrigé de l'E.F de STAT

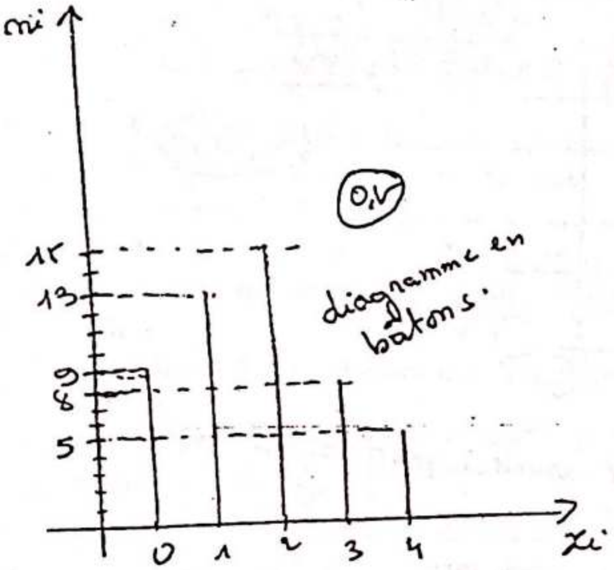
**EX01 :** (6pts)

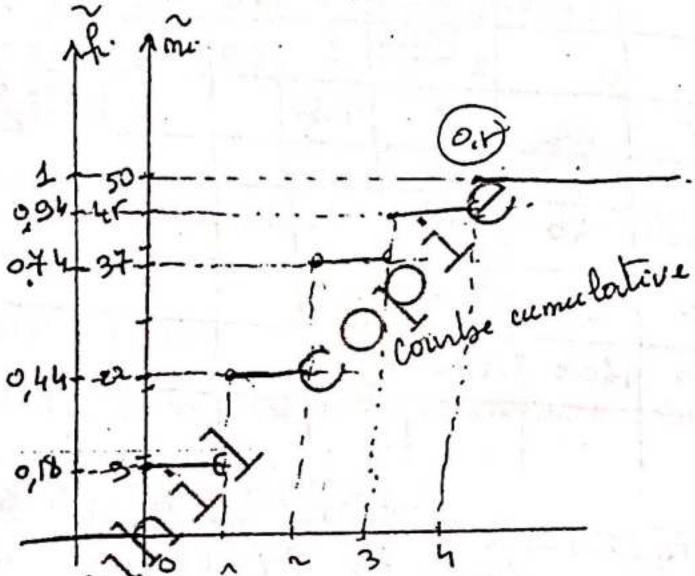
1°/

| $x_i$ | 0 | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|---|
| $n_i$ | 9 | 13 | 15 | 8 | 5 | $N=50$ |
| $f_i$ | 0,18 | 0,26 | 0,30 | 0,16 | 0,10 | 1 |
| $\tilde{n}_i$ | 9 | 22 | 37 | 45 | 50 | |
| $n_i x_i$ | 0 | 13 | 30 | 24 | 20 | $\sum n_i x_i = 87$ |
| $n_i x_i^2$ | 0 | 13 | 30 | 72 | 80 | $\sum n_i x_i^2 = 225$ |

(1)

2°/




3°/ $F_X : \mathbb{R} \longrightarrow [0,1]$

$$x \longmapsto F_X(x) = \begin{cases} 0 & \text{si } x < x_1 = 0 \\ \tilde{f}_i = \frac{\tilde{n}_i}{N} & \text{si } x_i \le x < x_{i+1} \quad i=1,\dots,k-1 \\ 1 & \text{si } x \ge x_5 = 4 \end{cases}$$

(0,5)

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0,18 & \text{si } 0 \le x < 1 \\ 0,44 & \text{si } 1 \le x < 2 \\ 0,74 & \text{si } 2 \le x < 3 \\ 0,90 & \text{si } 3 \le x < 4 \\ 1 & \text{si } x \ge 4 \end{cases}$$

4°/ $Mod(X) = 2$ enfants (0,5)
C'est la modalité $x_i$ qui admet le plus grand effectif

$$Med(X) = \begin{cases} \dfrac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2} & \text{si } \frac{N}{2} \in \mathbb{N} \\ x_{[N/2]+1} & \text{si } N/2 \notin \mathbb{N} \end{cases}$$

on a $\frac{N}{2} = 25 \in \mathbb{N}$ $\Longrightarrow Med(X) = \frac{x_{25} + x_{26}}{2} = \frac{2+2}{2} = 2$ enfants (0,5)

$Q_1(X) = q_{\frac{1}{4}} = x_{([12,5]+1)} = x_{13} = 1$ enfant car $\frac{N}{4} = 12,5 \notin \mathbb{N}$. (0,5)

$Q_3(X) = q_{\frac{3}{4}} = x_{([37,5]+1)} = x_{38} = 3$ enfants car $\frac{3N}{4} = 37,5 \notin \mathbb{N}$. (0,5)

5°/ intervalle Ecart interquartiles $= Q_3 - Q_1 = 3 - 1 = 2$ (0,25)

5/ $\overline{X}=\frac{1}{N}\sum_{i=1}^{5} n_i x_i = \frac{87}{50} = 1,74$ (0,25) et $Var(X)=\frac{1}{N}\sum_{i=1}^{5} n_i x_i^2 - \overline{X}^2$

$= \frac{225}{50} - (1,74)^2 = 1,47$ (0,5)

Exo 2 (14 points)

| x\y | [25-35[ | [35-45[ | [45-55[ | $n_{i.}$ | $\tilde{n}_{i.}$ | $x_i$ | $n_{i.}x_i$ | $n_{i.}x_i^2$ |
|---|---|---|---|---|---|---|---|---|
| [85-95[ | 6 $\boxed{16200}$ | 0 | 0 | 6 | 6 | 90 | 540 | 48600 |
| [95-105[ | 6 $\boxed{18000}$ | 5 $\boxed{20000}$ | 0 | 11 | 17 | 100 | 1100 | 11 000 |
| [105-115[ | 3 $\boxed{9900}$ | 3 $\boxed{13200}$ | 1 $\boxed{5500}$ | 7 | 24 | 110 | 770 | 84700 |
| [115-125[ | 0 | 2 $\boxed{9600}$ | 2 $\boxed{12000}$ | 4 | 28 | 120 | 480 | 57600 |
| [125-135[ | 0 | 0 | 2 $\boxed{13000}$ | 2 | 30 | 130 | 260 | 33800 |
| | | | | | | | $\sum n_{i.}x_i = 3180$ | $\sum n_{i.}x_i^2 = 334700$ |
| $n_{.j}$ | 15 | 10 | 5 | N=30 | | | | |
| $\tilde{n}_{.j}$ | 15 | 25 | 30 | | | | | |
| $y_j$ | 30 | 40 | 50 | | | | | |
| $n_{.j}y_j$ | 450 | 400 | 250 | $\sum n_{.j}y_j = 1100$ | | | | |
| $n_{.j}y_j^2$ | 13500 | 16000 | 12500 | $\sum n_{.j}y_j^2 = 42000$ | | | | |

3 × (0,5)



3/ b) on a $F_y(52)$ = % d'individus pour lesquels le caractère y est inférieur ou égal à 52 ($Y \le 52$)

Comme $52 \in [b_2; b_3[ = [45-55[$ ($j=3$) on a :

$F_y(52) = \tilde{f}_{.2} + \frac{52-45}{55-45} \times f_{.3} = \frac{25}{30} + \frac{7}{10}\times\frac{5}{30} = 0,833 + 0,116 = 94,9\%$ (1)

De même on a $F_y(45) = \frac{15+10}{30} = 0,833 = 83,3\%$

D'où le % demandé $= F_y(52) - F_y(45) = 94,9\% - 83,3\% = 11,6\%$

4/ $Med(X) = ?$

on a $\frac{N}{2} = \frac{30}{2} = 15$ donc $\underset{\tilde{n}_{1.}}{6} < 15 < \underset{\tilde{n}_{2.}}{17} \Rightarrow Med(X) \in [a_1, a_2[ = [95-105[$

$\Rightarrow Med(X) = q_{\frac{1}{2}} = a_1 + \left(\frac{a_2-a_1}{n_{2.}}\right)\left(\frac{N}{2} - \tilde{n}_{1.}\right) = 95 + \frac{10}{11}(15-6) = 103,18$ (0,75)

$N\alpha = 30 \times \frac{2}{10} = 6$

$D_2(X) = q_{0,2} \in [85-95[$ (0,75)

$D_2(X) = 85 + \frac{10}{6}(6-0) = 95$

$\overline{X} = \frac{1}{N}\sum_{i=1}^{5} n_i x_i = \frac{3180}{30} = 105$ (0,5)

5/ $V(X) = \frac{1}{N}\sum_{i=1}^{5} n_i x_i^2 - \overline{X}^2 = \frac{334700}{30} - 105^2 = 131,67 \Rightarrow \sigma(X) = 11,47$ (0,75)

$\overline{Y} = \frac{1}{N}\sum_{j=1}^{3} n_{.j}y_j = \frac{1100}{30} = 36,67$ (0,5)

$V(Y) = \frac{1}{N}\sum_{j=1}^{3} n_{.j}y_j^2 - \overline{Y}^2 = \frac{42000}{30} - (36,67)^2 = 55,31 \Rightarrow \sigma(Y) = \sqrt{V(Y)} = 7,44$ (0,75)

U.S.T.H.B
Faculté de Mathématique
Coordination.

15/01/2015

## Examen Final

### Problème :

**I.**

Dans une population de 40 travailleurs d'une entreprise, on a relevé le nombre de jours du congé de maladie pendant une période donnée. Les résultats sont donnés dans le tableau suivant :

| $X_i$ | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|
| $n_i$ | 5 | 9 | 13 | 5 | 5 | 3 |

1. Quelle est la nature du caractère X.
2. Quel est le nombre de travailleurs ayant demandé plus de 5 jours de congé de maladie ?
3. Calculer le mode et les quartiles.
4. Calculer la moyenne arithmétique et la variance.

**II.**

Pour la même population, on s'intéresse à l'âge de ces 40 travailleurs, on a obtenu le tableau suivant :

| Age (Y) | [25 , 35[ | [35 , 45[ | [45 , 55[ | [55 , 65[ |
|---|---|---|---|---|
| Nombre d'employés | 4 | 8 | 16 | 12 |

1. Quelle est la nature du caractère Y.
2. Quel est le pourcentage d'employés ayant 45 ans et plus ?
3. Calculer le mode, la médiane et l'âge moyen. Comparer ces trois valeurs. La distribution est-elle symétrique ? par quel paramètre peut-on le confirmer ?
4. Tracer la courbe cumulative et déduire l'intervalle [a,b[ contenant 84% de valeurs centrales.
5. Calculer la variance de Y.

**III.**

On désire étudier maintenant le lien qui peut exister entre les deux caractères étudiés précédemment, pour la même population. Les résultats sont donnés le tableau suivant :

| X \ Y | [25 , 35[ | [35, 45[ | [45 , 55[ | [55 , 65[ | |
|---|---|---|---|---|---|
| 3 | 2 | 0 | 3 | 0 | |
| 4 | 2 | 4 | 3 | 0 | |
| 5 | 0 | 3 | 10 | 0 | |
| 6 | 0 | 1 | 0 | 4 | |
| 7 | 0 | 0 | 0 | 5 | |
| 8 | 0 | 0 | 0 | 3 | |
| | | | | | |

1. Déterminer la distribution conditionnelle de Y sachant X=4. Déduire l'âge moyen des travailleurs ayant demandé 4 jours de congé de maladie
2. Les variables X et Y sont elles indépendantes ? justifier.
3. Calculer la covariance entre les variables X et Y. Déduire le coefficient de corrélation linéaire entre X et Y. Commenter.
4. Donner l'équation de la droite de régression de Y en X.
5. Peut-on estimer le nombre de jours de congé pour un travailleur âgé de 66 ans ?

**Exercice 1 (02 points):**

Dans une usine, on utilise deux machines M1 et M2 pour fabriquer des pièces. Pour une période donnée, leurs probabilités de tomber en panne sont respectivement 0, 01 et 0.08. De plus la probabilité de l'évènement "la machine M2 est en panne sachant que M1 est en panne" est égale à 0, 4.

1. Quelle est la probabilité d'avoir les deux machines en panne au même moment ?
2. Les deux machines fonctionnent elles indépendamment ?
3. Quelle est la probabilité d'avoir une seul machine en panne.
4. Quelle est la probabilité d'avoir au moins une machine qui fonctionne ?

**Exercice 2 :( 4 points)**

La direction de l'hôpital considère que 10 % de la population venant consulter est déjà contaminée par un certain virus. Un test de dépistage est disponible et on sait que :

- Si un patient n'est pas contaminé, le test sera négatif 9 fois sur 10.
- Si un patient est contaminé, le test sera positif 8 fois sur 10.

On considère les événements suivants :

**C** : « le patient est contaminé ».
**T** : « le test effectué est positif ».

1. Donner la probabilité que le patient soit contaminé.
2. Donner les probabilités : et
3. Calculer la probabilité que le test soit positif.
4. Sachant que le test est positif, quelle est la probabilité que le patient soit contaminé ?
5. Pourquoi la direction de l'hôpital peut-elle envisager de renoncer à ce test de dépistage ?

**Problème :(14 points)** **Corrigé de l'examen Final**

**I. (04 points)**

| $X_i$ | 3 | 4 | 5 | 6 | 7 | 8 | |
|---|---|---|---|---|---|---|---|
| $n_i$ | 4 | 6 | 12 | 7 | 6 | 5 | |
| $\tilde{n}_i$ | 4 | 10 | 22 | 29 | 35 | 40 | |
| $n_i \cdot x_i$ | 12 | 24 | 60 | 42 | 42 | 40 | 220 |
| $n_i \cdot x_i^2$ | 36 | 96 | 300 | 252 | 294 | 320 | 1298 |

1. Quantitatif discret.**(0.5)**
2. Le nombre de travailleurs ayant demandé plus de 5 jours de congé de maladie 7+6+5=18. alors le pourcentage P=45%. .**(0.5)**
3. Le Mode $M_0 = 5$ . .**(0.5)**
4. l'intervalle interquartiles : $[q_1, q_3]$ contient 50% de valeurs centrales. **(0.5)**
5. $\alpha = 0.25 \Rightarrow n\alpha = 10 \in N \Rightarrow q_1 = \frac{x_{10} + x_{11}}{2} = \frac{4+5}{2} = 4.5$ **(0.5)**
6. $\alpha = 0.75 \Rightarrow n\alpha = 30 \in N \Rightarrow q_3 = \frac{x_{30} + x_{31}}{2} = \frac{7+7}{2} = 7$ **(0.5)**
7. la moyenne arithmétique : $\bar{X} = \frac{220}{40} = 5.5$ **(0.5)**
8. la variance : $V(X) = \frac{1298}{40} - (5.5)^2 = 2.2 \Rightarrow \sigma_X = 1.48$ . **(0.5)**

**II. (05 points)**

| Age (Y) | [25 , 35[ | [35 , 45[ | [45 , 55[ | [55 , 65[ | |
|---|---|---|---|---|---|
| Nombre d'employés | 4 | 8 | 16 | 12 | |
| $\tilde{n}_i$ | 4 | 12 | 28 | 40 | |
| $n_i \cdot y_i$ | 120 | 320 | 800 | 720 | 1960 |
| $n_i \cdot y_i^2$ | 3600 | 12800 | 40000 | 43200 | 99600 |

1. Quantitatif continu. **(0.5)**
2. Le nombre d'employés ayant 45 ans et plus = 16+12=28 .Alors le pourcentage= 70%. **(0.5)**
3. le mode : $M_0 \in [45, 55[ \Rightarrow M_0 = 45 + \frac{8}{8+4} 10 = 51.66$ **(0.5)**

La médiane : $M_{ed} \in [45, 55[ \Rightarrow M_{ed} = 45 + \frac{20-12}{16} 10 = 50$ **(0.5)**

L'âge moyen : $\bar{Y} = \frac{1960}{40} = 49$ . **(0.5)**

$M_0 > M_{ed} > \bar{Y}$ Alors on a un étalement vers la gauche, et on pourra confirmer par le calcul du coefficient d'asymétrie qui sera négatif. **(0.25)+ (0.25)**

4. La courbe cumulative , l'intervalle [a,b] contenant 64% de valeurs centrales. **(0.5)**

$\alpha = 0.18 \Rightarrow a \cong 36$ et $\alpha = 0.82 \Rightarrow b \cong 50$. **(0.5)**

C'est-à-dire 64% des employés sont âgés entre 39 ans et 59 ans

5. La variance de Y : $V(Y) = \dfrac{99600}{40} - (49)^2 = 89 \Rightarrow \sigma_Y = 9.43$. **(01)**

**III. (05 points)**

| X \ Y | [25 , 35[ 30 | [35, 45[ 40 | [45 , 55[ 50 | [55 , 65[ 60 | $n_{i.}$ | |
|---|---|---|---|---|---|---|
| 3 | **4** 360 | **0** | **0** | **0** | 4 | |
| 4 | **0** | **6** 960 | **0** | **0** | 6 | |
| 5 | **0** | **2** 400 | **10** 2500 | **0** | 12 | |
| 6 | **0** | **0** | **4** 1200 | **3** 1080 | 7 | |
| 7 | **0** | **0** | **2** 700 | **4** 1680 | 6 | |
| 8 | **0** | **0** | **0** | **5** 2400 | 5 | |
| $n_{.j}$ | 4 | 8 | 16 | 12 | 40 | 11280 |

1. Déterminer la distribution conditionnelle de Y sachant X=5. Déduire l'âge moyen des travailleurs ayant demandé 5 jours de congé de maladie. **(01)**

| Y | [25 , 35[ 30 | [35, 45[ 40 | [45 , 55[ 50 | [55 , 65[ 60 | |
|---|---|---|---|---|---|
| $f_{Y_j / X=5}$ | $\frac{0}{12}$ | $\frac{2}{12}$ | $\frac{10}{12}$ | $\frac{0}{12}$ | 1 |
| $Y_j \cdot f_{Y_j / X=5}$ | 0 | $\frac{80}{12}$ | $\frac{500}{12}$ | 0 | $\frac{580}{12}$ |

$m_{Y/X=5} = 48.33\, ans$

2. $f_{11} = \dfrac{4}{40} \neq f_{1.} * f_{.1}$ alors X et Y ne sont pas indépendantes. **(0.5)**

3. La covariance : $Cov(X,Y) = \dfrac{11280}{40} - (5.5)(49) = 12.5$ **(01)**

Le coefficient de corrélation: $r(X,Y) = \dfrac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y} = \dfrac{12.5}{(1.48)(9.43)} = 0.89$. **(0.5)**

On a une très forte corrélation linéaire, donc un ajustement linéaire est justifié. **(0.5)**

4. L'équation de la droite de régression de X en Y : $X = \alpha Y + \beta$ **(0.5+0.5)**

$\alpha = \dfrac{Cov(X,Y)}{V(Y)} = \dfrac{12.5}{89} = 0.14$. $\beta = \overline{X} - \alpha \overline{Y} = -1.36$

$X = 0.14Y - 1.36$

5. On peut estimer le nombre de jours de congé pour un travailleur âgé de 70 ans :
Y=70 alors X= 8.44 ( pour un travailleur âgé de 70 ans, on estime qu'il a demandé 9 jours de congé de maladie durant cette période. **(0.5)**

**Exercice 1 : (02 points)**

$P(M_1)=0.01$. $\quad P(M_2)=0.08 \quad P(M_2/M_1)=0.4$ **(0.25)**

1. La probabilité d'avoir les deux machines en panne au même moment ?
$P(M_1 \cap M_2)=P(M_1).P(M_2/M_1)=0.01.\,0.4=0.004$ **(05)**

2. Les deux machines fonctionnent-elles indépendamment ?
$P(M_1 \cap M_2) \neq P(M_1).P(M_2)$ Alors $M_1$ et $M_2$ ne sont pas indépendants, c'est-à-dire que les deux machines ne fonctionnent pas indépendamment. **(0.25)**

3. Quelle est la probabilité d'avoir au moins une machine qui fonctionne ?
$P(\overline{M}_1 \cup \overline{M}_2)=1-P(M_1 \cap M_2)=1-0.004=0.996$ **(05)**

4. Quelle est la probabilité d'avoir une seule machine en panne? **(0.5)**
$P(M_1 \Delta M_2)=P(M_1).+P(M_2)-2P(M_1 \cap M_2)=0.01+0.08-2.0.004=0.082$

**Exercice 2 :( 04 points)**

**C** : « le patient est contaminé ».
**T** : « le test effectué est positif ».

10 % de la population est contaminée par un certain virus : $P(C)=0.1 \Rightarrow P(\overline{C})=0.9$

- Si un patient n'est pas contaminé, le test sera négatif 9 fois sur 10 :
$P(\overline{T}/\overline{C})=0.9$ **(0.25)**
- Si un patient est contaminé, le test sera positif 8 fois sur 10 :
$P(T/C)=0.8$ **(0.25)**

Donner la probabilité que le patient ne soit pas contaminé : $P(\overline{C})=0.9$ **(0.25)**

1. Si le patient n'est pas contaminé, la probabilité que le test soit positif?
$P(T/\overline{C})=1-P(\overline{T}/\overline{C})=1-0.9=0.1$ **(0.5)**

2. La probabilité que le test soit positif : $P(T)$
Comme $\{C,\overline{C}\}$ forment un système complet d'évènements, la formule de la probabilité totale nous donne :
$P(T)=P(C)P(T/C)+P(\overline{C})P(T/\overline{C})=0.1.\,0.8+0.1\,.0.9=0.17$ **(01.5)**

3. Sachant que le test est positif, la probabilité que le patient soit contaminé :
$P(C/T)=\dfrac{P(C)P(T/C)}{P(T)}=\dfrac{0.1.0.8}{0.17}=0.47$ (formule de Bayes) **(01)**

4. La direction de l'hôpital peut-envisager de renoncer à ce test de dépistage ? car parmi les patient dont le test est positif 47% sont contaminé par le virus, donc parmi les patients dont le test est positif 53% ne sont pas contaminés, alors on peut dire que ce test de dépistage n'est pas fiable. **(0.25)**

Faculté de Mathématiques
Dept : Prob-Statistiques
MI ($1^{ere}$ annee)

2 Juin 2014
Section (7)

Examen de fin de Semestre (2)
Statistiques

**Exercice (1) :** 1+3+3(7pts)

Soit la série statistique correspondant à la distribution des salaires mensuels des ouvriers d'une entreprise (salaires en milliers de dinars)

| Salaires(X) | [15 17[ | [17 19[ | [19 21[ | [21 23[ | [23 25] |
|---|---|---|---|---|---|
| Effectifs | 140 | 160 | 280 | 70 | 50 |

1. Calculer la moyenne et l'écart type de X.
2. Déterminer la proportion d'ouvriers dont le salaire se situe entre 20 et 24.

Les salaires de cette entreprise sont augmentés de 5% et chaque ouvrier perçoit une prime de 1000 DA. Soit Z la nouvelle variable exprimant le nouveau salaire. Exprimer Z en fonction de X et en déduire sa moyenne et son écart type



**Exercice (2) :** 2+4(6pts)

Soit $\{E_1, E_2\}$ une partition d'une population $E$ à un caractère. On désigne respectivement par $n_1$ et $n_2$ les effectifs de $E_1$ et $E_2$, $m_1$, $m_2$ et $m$ les moyennes respectives de $E_1$, $E_2$ et $E$ et $V_1$, $V_2$ et $V$ les variances respectives de $E_1$, $E_2$ et $E$
Montrer que :

1. $m = \dfrac{n_1 m_1 + n_2 m_2}{n_1 + n_2}$
2. $V = \dfrac{n_1 v_1 + n_2 v_2}{n_1 + n_2} + \dfrac{n_1 n_2 (m_1 - m_2)^2}{(n_1 + n_2)^2}$

**Exercice(3):** 2+2+3 (7pts)

Le tableau suivant donne les notes en algèbre (X) et en statistique (Y) obtenues par un groupe de 10 étudiants d'une section.

| X | 10 ✓ | 8 | 9 | 11 | 7 | 12 | 7 | 10 ✓ | 11 | 8 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Y | 13 | 12 | 13 | 15 | 11 | 14 | 12 | 12 | 13 | 11 | ? |

- Déterminer la moyenne de X sachant Y appartenant à [12 13]
- Déterminer la variance de Y sachant X=10
- Quelle note un onzième étudiant espérera t-il avoir en statistique s'il a obtenu 13 en algèbre ?

Faculté de mathématiques, USTHB
Module:PS, MI section 7, second semestre 2015- 2016

# Examen final ;Durée: 1H30 heures

**Exercice 1** ( 4 *points*). Soit $X$ la variable quantitative mesurée sur une population de taille 10. Les observations de cette variable sont regroupées dans le tableau suivant

$$\begin{bmatrix} classes & [0;4[ & [4;6[ & [6;8] \\ n_i & 1 & 5 & 4 \end{bmatrix}$$

1. Donner une valeur approchée du mode de cette variable (1pt)
2. A partir de ce tableau, donner l'équation de la fonction de répartition $F$, la tracer dans le plan et en déduire une valeur approchée du premier quartile, de la médiane et du 42 $^{ième}$ centile de la série par la méthode dite: *d'interpolation*.(3 *pts*)

**Exercice 2** (6 *points*). On considére la table de contingence suivante associée au croisement des codages respectifs de 2 variables quantitatives $X$ et $Y$.

| $X \backslash Y$ | $[1;5[$ | $[5;9]$ |
|---|---|---|
| $[0;2[$ | 5 | 0 |
| $[2;4[$ | 3 | 2 |

- En déduire la distribution $D(X,Y)$ du couple $(X,Y)$ et les 2 les distributions marginales.(1 *pt*)
- Donner la distribution de $X$ sachant que $Y$ est dans la classe $[5;9]$ (1 *pt*)
- A partir de la table de contingence, calculer le coefficient de corrélation linéaire entre $X$ et $Y$ .(2 *pts*)
- Ecrire l'équation de la droite de regression de $X$ en $Y$ et la tracer dans un répére orthonormé.(2 *pts*)

**Exercice 3** ( 10 *points*).

Dans un elevage de chévres, on estime que 30% sont atteintes par une maladie. on dispose d'un test pour cette maladie. Si une chévre n'est pas malade, il y a 9 chances sur 10 d'avoir une réaction négative du test. Si elle est atteinte il y a 8 chances sur 10 d'avoir une réaction positive. On soumet toutes les chévres au test.

On pose : $M=$ "la chévre est malade" ,$T =$ " la chévre a une réaction positive au test"

Faculté de Mathématiques
Dpt: Prob-Statistiques
M.I : Section 6

18.06. 2010.

Epreuve de Rattrapage
Module: Statistiques.

## Exercice n° 1 :

Un échantillon de 70 poissons est soumis à la consommation d'une substance chimique A qui agit sur leur taille en fonction du temps.
Le temps (X) est exprimé en mois et la taille (Y) est exprimée en (mm). Les résultats des calculs préliminaires sont:

$$\sum x = 125, \quad \sum y = 1160, \quad \sum x^2 = 255, \quad \sum xy = 2730.$$

-I) Quelle sera la meilleure taille possible au bout de 4 mois de consommation de la substance A.
Une autre substance chimique B a été administrée à ce même échantillon de poissons. L'équation de la droite de regression est alors $y = 7,331x + 9,403$

-II) Quelle est celle des deux substances qui favorise le mieux la croissance.

## Exercice n° 2:

On a procédé à l'ajustement d'un nuage de points d'un couple de variables statistiques (X, Y). Les équations des deux droites de regression obtenues sont;

$(\Delta_1)$ : $y = x + 30$ et $(\Delta_2)$ : $x = \frac{1}{4}y + 60$.

- I) Rappeler les expressions mathématiques donnant les coefficients a et b et a' et b' respectivement de $\Delta_1$ et $\Delta_2$
- II) En déduire la valeur du coefficient de corrélation linéaire
- III) Calculer les moyennes arithmétiques des variables X et Y
- IV) Calculer la covariance entre X et Y, sachant que la variance de Y est égale à 40

Fin