Initiation aux Statistiques **TSA11B**

Correction Examen du 19/02/2013

Seuls les documents distribués avec le sujet sont acceptés. Les calculatrices sont acceptées.
Les exercices sont indépendants et peuvent être traités dans l'ordre que vous souhaitez.

Exercice 1

Compléter le tableau suivant puis calculer la moyenne statistique, l'écart-type et la médiane.

<u>Temps d'attente à la CAF :</u>

Moyenne : 7,25
Écart type : 5,52834
Médiane : 5,52632

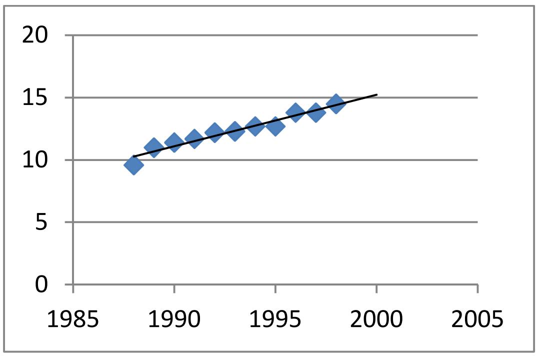
Temps en min	Effectifs: n_i	Fréquence
] 0 ; 5]	**38**	47,5 %
] 5 ; 10]	**19**	23,75 %
] 10 ; 15]	**14**	17,5 %
] 15 ; 20]	**7**	8,75 %
] 20 ; 25]	**2**	2,5 %
Totaux	80	100 %

Exercice 2

Dans le tableau ci-dessous, on donne le trafic aérien intérieur français, en milliards de voyageurs-kilomètres entre 1988 et 1998 (source : DGAC)

Année	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Trafic (en milliards de voyageurs-km)	9,6	11,0	11,4	11,7	12,2	12,3	12,7	12,7	13,8	13,8	14,5

1. a) Représenter le nuage de points de la série statistique double (année ; trafic).



   b) Peut-on estimer qu'il y a une dépendance linéaire entre l'année et le trafic ? oui

2. a) Déterminer le coefficient de corrélation linéaire. r = 0,97

   b) Permet-il de confirmer la réponse faite à la question 1b ? oui

3. Donner l'équation de la droite de régression. y = 0,41x – 810,23

Exercice 3

Dans une agence de location de voitures, le patron veut savoir quelles sont les voitures qui n'ont roulé qu'en ville pour les revendre immédiatement. Pour cela, il y a dans chaque voiture une boîte noire qui enregistre le nombre d'heures pendant lesquelles la voiture est restée au point mort, au premier rapport, au deuxième rapport,..., au cinquième rapport. On sait qu'une voiture qui ne roule qu'en ville passe en moyenne 10% de son temps au point mort, 5% en première, 30% en seconde, 30% en troisième, 20% en quatrième, et 5% en cinquième. On décide de faire un test du Chi_2 pour savoir si une voiture n'a roulé qu'en ville ou non.

1) Sur une première voiture, on constate sur 2000 heures de conduite : 210h au point mort, 94h en première, 564h en seconde, 630h en troisième, 390h en quatrième, et 112h en cinquième.

   Cette voiture n'a-t-elle fait que rester en ville ?

On veut tester l'adéquation de notre échantillon à la loi discrète : $p_0 = 0.1$, $p_1 = 0.05$, $p_2 = 0.3$, $p_3 = 0.3$, $p_4 = 0.2$, $p_5 = 0.05$. On effectue un test du χ^2. En fait, on veut tester H_0 = la voiture n'a roulé qu'en ville, contre H_1 = la voiture n'a pas roulé qu'en ville.
1) Pour la première voiture, on constate

	0	1	2	3	4	5
eff obs obs_i	210	94	564	630	390	112
eff th th_i	200	100	600	600	400	100

On calcule la distance du χ^2.

$$k_{2000} = \sum_{i=0}^{5} \frac{(th_i - obs_i)^2}{th_i} = \frac{10^2}{200} + \frac{6^2}{100} + \frac{36^2}{600} + \frac{10^2}{600} + \frac{10^2}{400} + \frac{12^2}{100} = 6.21$$

Détermination du seuil : $P[\chi_5^2 > c] = 0.05 \Longrightarrow c = 11.07$ et $\forall \alpha, \chi_5^2 \in [9;21]$
Comme $k_{2000} = 6.21 < 9$, on ne peut pas rejeter H_0 : la voiture n'a roulé qu'en ville.

2) Avec une autre voiture, on obtient les donnés suivantes : 220h au point mort, 80h en première, 340h en seconde, 600h en troisième, 480h en quatrième et 280h en cinquième.

Pour la seconde voiture, on constate

	0	1	2	3	4	5
eff obs obs_i	220	80	340	600	480	280
eff th th_i	200	100	600	600	400	100

On calcule la distance du χ^2.

$$k_{2000} = \sum_{i=0}^{5} \frac{(th_i - obs_i)^2}{th_i} = 458.67 >> 11.07$$

On rejette H_0 : la voiture n'a pas roulé qu'en ville. La p-valeur vaut 0. La décision ne fait pas de doute.

Exercice 4

Une étude médicale a été menée dans différents hôpitaux français sur une maladie rare et mortelle. 15 cas de cette maladie ont été recensés et les durées de vie exprimées en jours, entre le diagnostic et le décès ont été relevées :

Voiture	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Durée de vie	298	748	760*	760*	80	132	31	760*	520	300*	300*	383	550*	95*	225

Les durées de vie suivies d'une astérisque correspondent à des voitures accidentées avant les 760 semaines et à 3 moteurs encore en vie à la fin de l'expérimentation pour les valeurs 760.

1- Calculer une estimation de la fiabilité d'un moteur en utilisant la méthode de Kaplan-Meier puis celle de Johnson.

2- Justifier graphiquement l'utilisation d'une loi exponentielle pour modéliser la durée de vie de ces moteurs. Estimer graphiquement la durée de vie moyenne.

t_j (défaillances)	$R^*(t_j)$ (%) Kaplan-Meier	$R^*(t_j)$ (%) Johnson
31	93	95,5
80	87	89
132	79	82
225	72	75
298	65	68
383	56	59,2
520	46	50,5
748	35	40

La représentation graphique du *ln* des valeurs donne des points pratiquement alignés, d'où loi exponentielle.
MTTF = 715 (avec K-M) et 815 (avec Johnson)

Exercice 5

Répartition de 200 naissances. Les deux caractères observés sont le sexe et le poids à la naissance.

	Filles	Garçons	
< 3 kg	30	20	50
>= 3 kg	65	85	150
	95	105	200

Tester l'indépendance du sexe du bébé sur son poids à la naissance avec un risque de 5%, puis avec un risque de 1%,

Nous pouvons conclure au rejet de l'hypothèse d'indépendance, le seuil, au risque de 5%, est de 3,841 car $\chi^2_{1,5\%} \simeq 3{,}841$ avec une valeur du Chi2 supérieure de $\sum_{i=1}^{c}\sum_{j=1}^{l}\frac{\left(x_{ij}-\frac{x_{i\bullet}x_{\bullet j}}{n}\right)^2}{\frac{x_{i\bullet}x_{\bullet j}}{n}} \simeq 4{,}177$. Nous pouvons alors conclure qu'au risque de 5% le poids est différent à la naissance entre les filles et les garçons.

On s'aperçoit que le seuil est proche des valeurs critiques. En effet, si l'on choisit 1% comme risque d'erreur, la conclusion devient alors opposée. Au risque de 1%, le seuil est de 6,63 soit une valeur supérieure à 4,177. On ne peut plus rejeter l'hypothèse d'indépendance au risque de 1%. Nous pouvons alors conclure qu'au risque de 1% nous n'avons pas observé de différence du poids à la naissance entre les garçons et les filles.

Remarque : il est normal que le seuil augmente quand le risque diminue. En effet, si nous diminuons le risque d'affirmer à partir d'un exemple (échantillon) que l'hypothèse de départ est fausse alors que dans la réalité elle est vraie, nous sommes obligés d'accepter des valeurs plus élevées pour le Chi2.