



PRESIDENCE

POLYNESIE FRANÇAISE

SERVICE DU PERSONNEL
ET DE LA FONCTION PUBLIQUE

.....

CONCOURS SPECIAL EXTERNE POUR LE
RECRUTEMENT DE 2 ATTACHES
D'ADMINISTRATION DE CATEGORIE A DEVANT
ETRE AFFECTES A DES FONCTIONS DE
STATISTICIEN-ECONOMISTE RELEVANT DE LA
FONCTION PUBLIQUE DE LA POLYNESIE
FRANÇAISE

EPREUVE ECRITE DE PROBABILITES ET STATISTIQUES

Jeudi 13 janvier 2011

(Durée : 4 heures – coefficient 4)

Matériel autorisé : calculatrice.

Le sujet comporte 9 pages (page de garde incluse)

CONCOURS D'ENTREE DANS LA FONCTION PUBLIQUE TERRITORIALE DE POLYNESIE FRANÇAISE – JANVIER 2011

SUJET DE L'EPREUVE DE STATISTIQUES – PROBABILITES / DUREE 4 HEURES

LES EXERCICES SONT TOUS INDEPENDANTS LES UNS DES AUTRES. EPREUVE NOTEE SUR 60 NOTE RAMENEE SUR 20

LE SUJET COMPORTE 8 PAGES DONT UNE CONSACREE A DES TABLES DE LOIS DE PROBABILITE

Exercice I

10 points

Une enquête est réalisée sur un échantillon de 10 000 personnes âgées entre 20 et 60 ans en Polynésie française. Cet échantillon est tiré de manière stratifiée géographiquement. Le tableau suivant donne la répartition de la population âgée entre 20 et 60 ans en Polynésie française selon les zones géographiques définies pour l'enquête. Il présente également la répartition de l'échantillon selon cette même répartition.

**Répartition de la population dont l'âge est compris entre 20 et 60 ans en Polynésie française
par zones géographiques - Source : ISPF - recensement de la population de 2007**

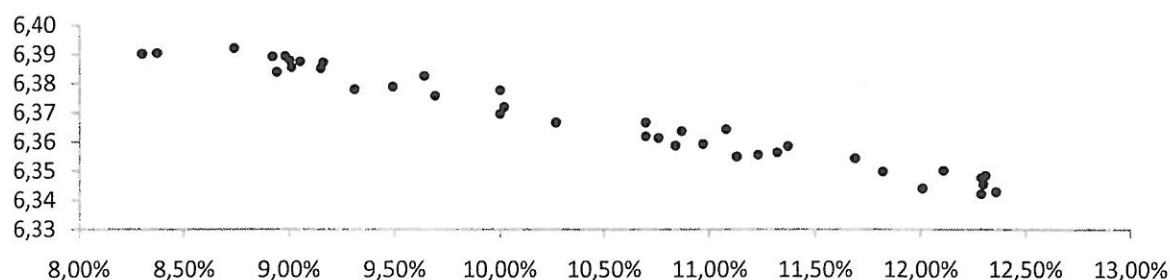
Zones géographiques	Au recensement de 2007	Echantillon de l'enquête
Zone Urbaine Est	16,0%	1 550
Zone Urbaine Ouest	27,0%	2 650
Zone Rurale	17,0%	1 650
Papeete	10,0%	1 050
Moorea-Maiao	6,0%	550
Hors îles du vent	24,0%	2 550
Total	100,0%	10 000

1. Expliquez ce que représente un tirage d'échantillon stratifié (2 point)
2. Présentez d'autres méthodes d'échantillonnage en expliquant leur mode de fonctionnement (3 points)
3. Déterminez au seuil de risque de 5 % si les conclusions de l'enquête pourront être contestées (4 points)
4. En admettant que les proportions dans l'échantillon ne bougent pas mais que la taille de l'échantillon soit divisée par 2, sans refaire le test intégralement, votre conclusion serait-elle la même (justifiez) ? (1 point)

Exercice II

10 points

Vous disposez de la représentation graphique suivante et des résultats présentés ci-après concernant l'évolution du taux de chômage en pourcentage et la consommation des ménages en milliards de F.CFP entre. Le graphique ci-dessous présente en abscisse le taux de chômage et en ordonnée la consommation des ménages en milliards.



On pose X le taux de chômage et Y la consommation des ménages. Vous disposez alors des résultats suivants :

- $\sum X_i = 418,56$
- $\sum X_i^2 = 44,46$
- $(\sum X_i)^2 = 1\,751,92$
- $\sum Y_i = 254,70$
- $\sum Y_i^2 = 1\,621,80$
- $(\sum Y_i)^2 = 64\,871,54$
- $\sum X_i Y_i = 26,64$
- $(\sum X_i Y_i)^2 = 709,88$
- $\sum X_i^2 Y_i^2 = 18,00$
- $\sum X_i Y_i^2 = 169,60$
- $\sum X_i^2 Y_i = 2,83$
- $N^2 = 1\,600$

1. Déterminez la moyenne, la variance et l'écart type du taux de chômage (2 points)
2. Déterminez la moyenne, la variance et l'écart type de la consommation des ménages (2 points)
3. Déterminez la relation affine semblant exister entre le taux de chômage et la consommation des ménages (3 points)
4. Déterminez les paramètres nécessaires pour juger de la qualité de cette relation (2 points)
5. En considérant comme robuste cette relation, quelle serait la consommation des ménages estimée si le taux de chômage atteignait la valeur de 7,5 % ? (1 point)

Exercice III

10 points

Vous disposez des éléments suivants concernant la répartition des salaires en Polynésie française au 31 décembre 2009. Les tableaux présentés ci-après n'en forme qu'un seul mais en raison du nombre important de colonnes il a été séparé en deux.

Valeur Basse F.CFP incluse	Valeur Haute F.CFP exclue	C_i Centre de classe	$C_{mi} = \frac{C_i}{1000}$	A_i Amplitude	D_i Densité	n_i Effectifs
0	50 000	25 000	25	50 000	7,97%	3 986
50 000	100 000	75 000	75	50 000	10,04%	5 018
100 000	120 000	110 000	110	20 000	10,42%	2 084
120 000	140 000	130 000	130	20 000	10,23%	2 046
140 000	150 000	145 000	145	10 000	36,72%	3 672
150 000	160 000	155 000	155	10 000	32,56%	3 256
160 000	180 000	170 000	170	20 000	27,47%	5 493
180 000	200 000	190 000	190	20 000	20,77%	4 154
200 000	250 000	225 000	225	50 000	13,99%	6 994
250 000	350 000	300 000	300	100 000	9,06%	9 064
350 000	500 000	425 000	425	150 000	6,14%	9 209
500 000	1 000 000	750 000	750	500 000	1,74%	8 678
1 000 000	1 500 000	1 250 000	1 250	500 000	0,32%	1 608
1 500 000	2 000 000	1 750 000	1 750	500 000	0,10%	488
2 000 000	3 000 000	2 500 000	2 500	1 000 000	0,03%	288
3 000 000	4 000 000	3 500 000	3 500	1 000 000	0,02%	189
					Somme... N =	66 227

Valeur Basse F.CFP incluse	Valeur Haute F.CFP exclue	$n_i \times C_{mi}$	$n_i \times C_{mi}^2$	$f_i = \frac{n_i f_i}{N}$ Fréquence	F_i Fréquences cumulées
0	50 000	99 650	2 491 250	6,02%	6,02%
50 000	100 000	376 350	28 226 250	7,58%	13,60%
100 000	120 000	229 240	25 216 400	3,15%	16,74%
120 000	140 000	265 980	34 577 400	3,09%	19,83%
140 000	150 000	532 440	77 203 800	5,54%	25,38%
150 000	160 000	504 680	78 225 400	4,92%	30,29%
160 000	180 000	933 810	158 747 700	8,29%	38,59%
180 000	200 000	789 260	149 959 400	6,27%	44,86%
200 000	250 000	1 573 650	354 071 250	10,56%	55,42%
250 000	350 000	2 719 200	815 760 000	13,69%	69,11%
350 000	500 000	3 913 825	1 663 375 625	13,91%	83,01%
500 000	1 000 000	6 508 500	4 881 375 000	13,10%	96,11%
1 000 000	1 500 000	2 010 000	2 512 500 000	2,43%	98,54%
1 500 000	2 000 000	854 000	1 494 500 000	0,74%	99,28%
2 000 000	3 000 000	720 000	1 800 000 000	0,43%	99,71%
3 000 000	4 000 000	661 500	2 315 250 000	0,29%	100,00%
	Sommes...	22 692 085	16 391 479 475	100,00%	

- Définissez et déterminez la classe modale (1 point)
- A partir des éléments disponibles dans le tableau vous estimerez le salaire moyen, la variance et l'écart type des salaires (2 points)
- Déterminez, après les avoir définis, dans quels intervalles se situent le premier quartile, la médiane et le troisième quartile (3 points)
- Vous estimerez par interpolation, la valeur de la médiane (2 points)
- A l'aide des résultats statistiques et de vos connaissances économiques, commentez en cinq ou six lignes les résultats obtenus (2 points)

Exercice IV

5 points

- Sur un échantillon de 400 individus pris au hasard dans la population, on dénombre pas moins de 237 personnes ayant au moins le baccalauréat. Déterminez l'intervalle de confiance au seuil de risque de 5% dans lequel se trouve la proportion d'individus ayant au moins le baccalauréat dans la population (2 points)
- Sur ce même échantillon, le poids moyen des individus était de 74 kg avec une variance de 144 kg². Au seuil de risque de 5%, déterminez l'intervalle de confiance dans lequel se trouve le poids moyen de la population (3 points)

Exercice V

5 points

L'inflation mensuelle en Polynésie française suit une loi normale de moyenne 0,2% et de variance 0,000225%.

- Quelle est la probabilité pour qu'un mois donné, l'inflation soit supérieure à 0,5% ? (2 points)
- Quelle est la probabilité pour qu'elle soit comprise entre 0,05% et 0,15% ? (3 points)

Exercice VI

5 points

Vous devez constituer des échantillons de 7 personnes à partir d'un groupe de 20 individus composé de 13 femmes et 7 hommes.

1. Combien d'échantillons différents pouvez-vous constituer ? (1 point)
2. Combien d'échantillons comprenant 4 femmes et 3 hommes pouvez-vous constituer ? (2 points)
3. Quelle est la probabilité de constituer un échantillon exclusivement féminin ? (2 points)

Exercice VII

5 points

20% des conducteurs ont moins de 25 ans. 60% des moins de 25 ans et 20% des plus de 25 ans ont eu un accident durant les deux dernières années.

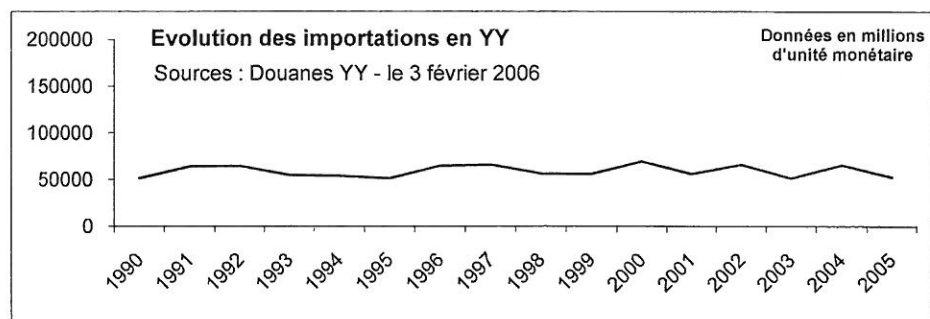
1. Quelle est la probabilité de rencontrer un conducteur ayant eu un accident dans les deux dernières années ? (2 points)
2. En cas d'accident dans les deux dernières années, quelle est la probabilité pour que le conducteur ait moins de 25 ans ? (3 points)

Exercice VIII

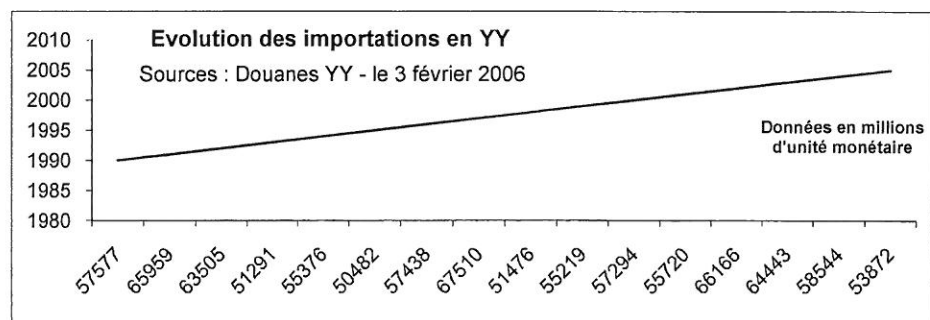
5 points

Il vous est demandé de porter un regard critique sur les graphiques suivants. En aucun cas il vous est demandé de reproduire ces éléments graphiques sur votre copie. Les graphiques sont numérotés. Veuillez simplement indiquer le numéro du graphique devant vos commentaires.

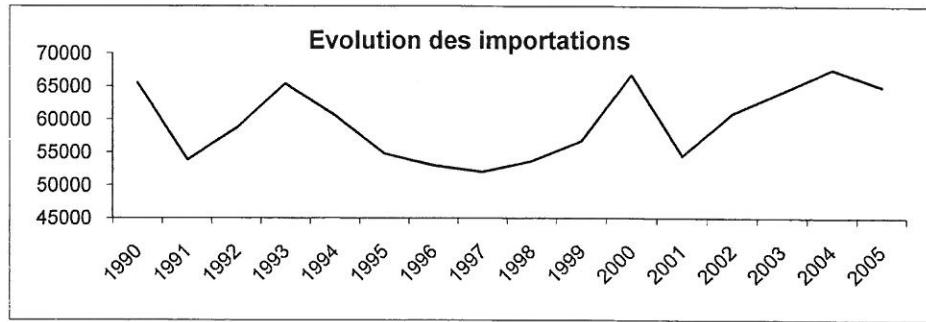
Graphique 1



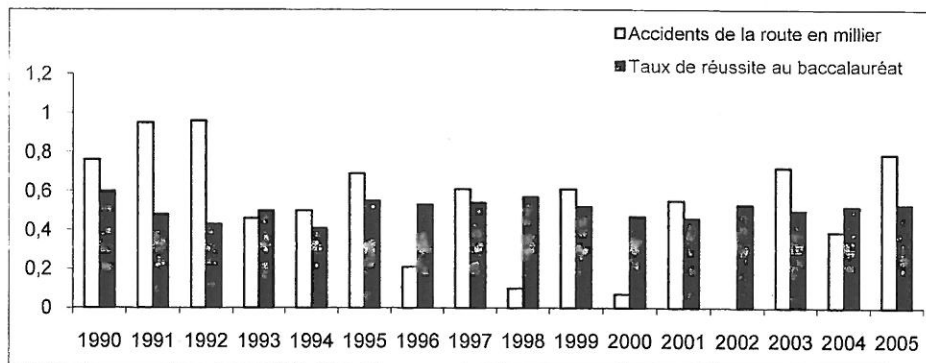
Graphique 2



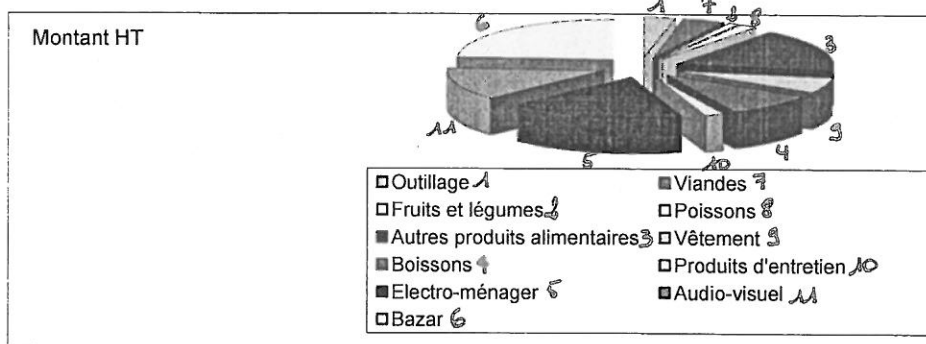
Graphique 3



Graphique 4



Graphique 5



Exercice IX

5 points

Après avoir rappelé le principe essentiel de toute méthode de redressement, décrivez en quelques lignes quelques méthodes de redressement de valeurs manquantes (1 point)

Une étude a été menée sur un échantillon de 1255 personnes. Cette étude a permis de recenser le sexe, l'âge des individus ainsi que le milieu social économique auquel ils appartiennent, le temps consacré à la télévision de manière hebdomadaire, le nombre de pages de livres lues de manière hebdomadaire également. Dans l'ensemble des données recueillies, 4 questionnaires n'ont pas été intégralement remplis. Ces questionnaires présentent les informations suivantes :

N° Quest.	Age	Sexe	Milieu socio-économique	Nombre de pages lues quotidiennement	Nombre d'heures consacrées hebdomadairement à la télévision
1		Féminin	Défavorisé	11	38,75
2	14		Aisé	52	21,00
3	67	Masculin		17	59,25
4	51	Féminin	Aisé		20,00
5	37	Masculin	Défavorisé	9	

Comme vous pouvez le constater ces 5 questionnaires présentent des informations manquantes, informations qu'il vous est demandé de combler en utilisant l'ensemble des informations disponibles obtenues à partir des autres questionnaires entièrement remplis. On pose « A » la variable « Age », « S » la variable « Sexe », « MS » la variable « Milieu socio-économique », « NP » la variable « nombre de pages lues quotidiennement » et « NH » la variable « Nombre d'heures consacrée hebdomadairement à la télévision ».

Sur l'ensemble des questionnaires intégralement remplis il a été obtenu que :

	Age moyen	Nombre moyen de pages lues quotidiennement	Nombre moyen d'heures hebdomadaires consacrées à la télévision
Féminin	38,39	77,48	24,89
Aisé	38,22	126,34	15,18
Défavorisé	38,57	26,66	34,99
Masculin	38,05	75,72	25,37
Aisé	38,32	126,84	15,98
Défavorisé	37,79	26,17	34,47
Moyenne globale	38,22	76,58	25,14

Par ailleurs vous avez connaissance des relations suivantes :

Toutes modalités confondues :

$$NP = 1,52 \times A + 18,54 \text{ avec } \rho = 0,43 \text{ et } \rho^2 = 0,19$$

$$NH = 0,38 \times A + 10,49 \text{ avec } \rho = 0,46 \text{ et } \rho^2 = 0,21$$

$$NH = 0,11 \times NP + 33,76 \text{ avec } \rho = -0,48 \text{ et } \rho^2 = 0,22$$

En ce qui concerne les personnes aisées

$$NP = 2,98 \times A + 12,44 \text{ avec } \rho = 0,99 \text{ et } \rho^2 = 0,99$$

$$NH = 0,02 \times A + 14,67 \text{ avec } \rho = 0,05 \text{ et } \rho^2 = 0,002$$

$$NH = 0,009 \times NP + 14,50 \text{ avec } \rho = 0,05 \text{ et } \rho^2 = 0,003$$

En ce qui concerne les personnes défavorisées

$$NP = 0,04 \times A + 28,04 \text{ avec } \rho = 0,05 \text{ et } \rho^2 = 0,003$$

$$NH = 0,77 \times A + 5,51 \text{ avec } \rho = 0,98 \text{ et } \rho^2 = 0,95$$

$$NH = 0,05 \times NP + 42,29 \text{ avec } \rho = 0,05 \text{ et } \rho^2 = 0,002$$

A partir des informations ici présentées, complétez les éléments manquants de la manière la plus précise qui soit. (4 points)

Table de la fonction intégrale de la loi de LAPLACE GAUSS (dite LOI NORMALE)
Probabilité de trouver une valeur inférieure à x

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
0,4	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
0,5	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
0,6	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
0,7	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
0,8	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
0,9	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
1,0	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
1,1	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
1,2	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
1,3	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
1,4	0,919	0,921	0,922	0,924	0,925	0,926	0,928	0,929	0,931	0,932
1,5	0,933	0,934	0,936	0,937	0,938	0,939	0,941	0,942	0,943	0,944
1,6	0,945	0,946	0,947	0,948	0,949	0,951	0,952	0,953	0,954	0,954
1,7	0,955	0,956	0,957	0,958	0,959	0,960	0,961	0,962	0,962	0,963
1,8	0,964	0,965	0,966	0,966	0,967	0,968	0,969	0,969	0,970	0,971
1,9	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977
2,0	0,977	0,978	0,978	0,979	0,979	0,980	0,980	0,981	0,981	0,982
2,1	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985	0,986
2,2	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989	0,989
2,3	0,989	0,990	0,990	0,990	0,990	0,991	0,991	0,991	0,991	0,992
2,4	0,992	0,992	0,992	0,992	0,993	0,993	0,993	0,993	0,993	0,994
2,5	0,994	0,994	0,994	0,994	0,994	0,995	0,995	0,995	0,995	0,995
2,6	0,995	0,995	0,996	0,996	0,996	0,996	0,996	0,996	0,996	0,996
2,7	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997
2,8	0,997	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998
2,9	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,999	0,999	0,999

Table de Pearson ou de KHI^2

Valeur de KHI^2 ayant la probabilité $P = 5\%$ d'être dépassées

Degré de liberté	1	2	3	4	5	6	7	8	9	10
Valeur critique	3,84	5,99	7,81	9,49	11,07	12,59	14,07	15,51	16,92	18,31

Degré de liberté	11	12	13	14	15	16	17	18	19	20
Valeur critique	19,68	21,03	22,36	23,68	25,00	26,30	27,59	28,87	30,14	31,41

Degré de liberté	21	22	23	24	25	26	27	28	29	30
Valeur critique	32,67	33,92	35,17	36,42	37,65	38,89	40,11	41,34	42,56	43,77