

IUT STID, 1^{ère} année et APPC
Statistique descriptive
 Devoir du Jeudi 13 janvier 2010

Important : Les réponses sont à donner directement sur le sujet. N'oubliez pas de noter votre nom.
Par ailleurs, les documents sont interdits et les calculatrices autorisées. Sauf indication contraire, les notations utilisées sont celles du cours.

Nom :

Exercice 1 Concentration des revenus en France

Remarque : Les réponses doivent être justifiées. Si l'espace fourni pour répondre n'est pas suffisant, vous pouvez utiliser l'espace situé page 4.

Le tableau ci-dessous donne la répartition des revenus salariaux et d'allocation chômage des 9 premiers déciles de la population française (étudiants exclus) en 2007, ainsi que divers calculs permettant la mesure de la concentration des revenus en France.

Source : INSEE (<http://www.insee.fr>)

Revenus (classe C_i)	c_i	f_i	f_i^*	$f_i c_i$	v_i	$f_i(v_i + v_{i-1})$
$[0; 4\ 786[$		11,11%				
$[4\ 786; 9\ 363[$	7 074,5	11,11%	22,22%	786	0,0648	0,0090
$[9\ 363; 12\ 974[$	11 168,5	11,11%	33,33%	1 241	0,1413	0,0229
$[12\ 974; 15\ 420[$	14 197	11,11%	44,44%	1 577	0,2386	0,0422
$[15\ 420; 17\ 512[$	16 466	11,11%	55,56%	1 830	0,3514	0,0655
$[17\ 512; 19\ 795[$	18 653,5	11,11%	66,67%	2 073	0,4791	0,0923
$[19\ 795; 22\ 585[$		11,11%				
$[22\ 585; 26\ 525[$	24 555	11,11%	88,89%	2 728	0,7924	0,1574
$[26\ 525; 34\ 081[$	30 303	11,11%	100,00%	3 367	1,0000	0,1992
Total	\emptyset	1	\emptyset	16 222	\emptyset	

1. Complétez la première et la septième lignes du tableau avec les valeurs de c_i , f_i^* , $f_i c_i$, v_i et $f_i(v_i + v_{i-1})$

manquantes. Vous détaillerez ci-dessous les calculs effectués.

2. Comment peut-on interpréter concrètement les nombres 33,33% et 0,1413 situés dans les colonnes f_i^* et v_i de la troisième ligne du tableau ?

3. Tracez, sur le quadrillage ci-dessous, la courbe de Lorentz des revenus en France.

4. Déterminez l'indice de Gini des revenus en France.¹

5. Au vu des deux dernières questions, comment peut-on qualifier la concentration des revenus en France ?

6. L'indice de Gini de la Suède en 2008 est inférieur à 0,25 alors que celui des USA est compris entre 0,45

¹Notez que, comme ces données ne sont fournies que pour les 9 premiers déciles, l'indice de Gini de la France est légèrement sous-évalué par rapport à la réalité.

et $0,50^2$. Que vous suggère, concrètement, la comparaison entre ces deux indices de Gini et celui obtenu à la question 4 ?

Espace supplémentaire pour d'éventuels détails dans les calculs :

Exercice 2 Corrélation 1 (dédié à Daniel)

Remarque : Les réponses doivent être justifiées. Si l'espace fourni pour répondre n'est pas suffisant, vous pouvez utiliser l'espace situé page 7.

On s'intéresse à la distribution conjointe de deux variables qui correspondent à deux questions posées à un échantillon de la population française lors d'une enquête sur le sentiment d'insécurité :

- Avez-vous parfois peur à votre domicile ? (réponses possibles : oui ou non)
- Avez-vous été
 - Victime d'une agression ?
 - Pas victime d'une agression mais témoins de destructions dans votre quartier ?
 - Pas victime d'une agression ni de destructions mais témoins d'une agression ?
 - Ni victime d'une agression, ni témoins de destructions ou d'une agression ?

²Source : Wikipédia, « World Map Gini coefficient » (http://commons.wikimedia.org/wiki/File:World_Map_Gini_coefficient.svg)

On notera X la variable correspondant aux réponses à la première question et Y la variable correspondant aux réponses à la seconde question. La table de contingence sur l'échantillon de Français interrogés est la suivante³ :

X	Y	Victime	Destructions dans quartier	Témoins d'agression	Rien	
Apeurés		816	1468	163	1 632	
Non apeurés		5 083	12 015	1 848	27 727	

1. Entre ces deux distributions :

- Distribution de Y conditionnellement à X
- Distribution de X conditionnellement à Y

laquelle permet de déterminer si les personnes apeurées ont plus souvent été victimes d'agressions que les personnes non apeurées ?

Déterminer cette distribution dans le tableau ci-dessous :

X	Y	Victime	Destruction dans quartier	Témoins d'agression	Rien	
Apeurés						
Non apeurés						

2. D'après les résultats de la question précédente, les personnes apeurées ont-elles été plus souvent victimes d'agressions que les personnes non apeurées ? (Justifier en reprenant les valeurs adéquates du tableau précédent).

³Les données sont reconstituées de la publication de l'INSEE : Thomas Le Jeannic (2006) Insécurité : Perceptions et réalités. Dans *Données sociales, la société française, Édition 2006*, p637-647. Elles concernent la période 2000-2004.

3. Le tableau ci-dessous contient les effectifs théoriques d'indépendance : complétez-le en justifiant vos résultats. Coloriez ensuite en vert les paires de modalités sous-représentées et en rouge les paires de modalités sous-représentées. Que peut-on déduire du coloriage ?

Tableau des effectifs théoriques d'indépendance

X	Y	Victime	Destruction dans quartier	Témoins d'agression	Rien	Total
Apeurés		474,11	1 083,65			
Non apeurés		5 424,89	12 399,35			
Total						

4. Le tableau ci-dessous contient les contributions au χ^2 : complétez-le en justifiant vos résultats. Quelle est la paire de modalités qui contribue le plus au χ^2 . Interprétez ce phénomène.

Tableau des contributions au χ^2

X	Y	Victime	Destruction dans quartier	Témoins d'agression	Rien
Apeurés		246,54	136,33		
Non apeurés		21,55	11,91	0,00	19,61

5. Calculez le χ^2 puis le C de Cramer. Interprétez cette dernière valeur.

Espace supplémentaire pour d'éventuels détails dans les calculs :

Exercice 3 Corrélation 2

Remarque : Les réponses doivent être justifiées. Si l'espace fourni pour répondre n'est pas suffisant, vous pouvez utiliser l'espace situé page 11.

Des expériences ont été effectuées concernant la distance de freinage sur route mouillée : en fonction de la vitesse de la voiture au moment du freinage, on a mesuré la distance nécessaire pour l'arrêt complet du véhicule. Les expériences, répétées 100 fois, ont été reportées dans le tableau ci-dessous. Pour alléger la taille du tableau, seule une partie du tableau et divers calculs associés sont reproduits.⁴

⁴Les vitesses sont données en km/h et les distances en mètres.

Exp. numéro	Vitesse	Distance	Vitesse ²	Distance ²	Vitesse × Distance	Vitesse ² × Distance	Vitesse ⁴
1	102	137	10 404	18 769	13 974	1 425 348	108 243 216
2	42	20	1 764	400	840	35 280	3 111 696
3	73	71	5 329	5 041	5 183	378 359	28 398 241
4	8	3	64	9	24	192	4 096
5	62	53	3 844	2 809	3 286	203 732	14 776 336
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	120	185	14 400	34 225	22 200	2 664 000	207 360 000
99	134	227	17 956	51 529	30 418	4 076 012	322 417 936
100	20	9	400	81	180	3 600	160 000
Total	7 759	10 461	801 509	1 850 311	1 185 352	$\simeq 141\,928 \times 10^3$	$\simeq 10\,900 \times 10^6$

1. Quelle est la population étudiée? Quelle est sa taille?

2. Quelles sont les variables étudiées? Quels sont leurs types?

3. Laquelle de ces deux régressions :

- régression de la vitesse en la distance de freinage
- régression de la distance de freinage en la vitesse

vous paraît être d'intérêt dans ce problème? (Cochez la case correspondante)

4. Pour la régression choisie à la question précédente, calculez le coefficient de régression linéaire. Commentez

sa valeur.

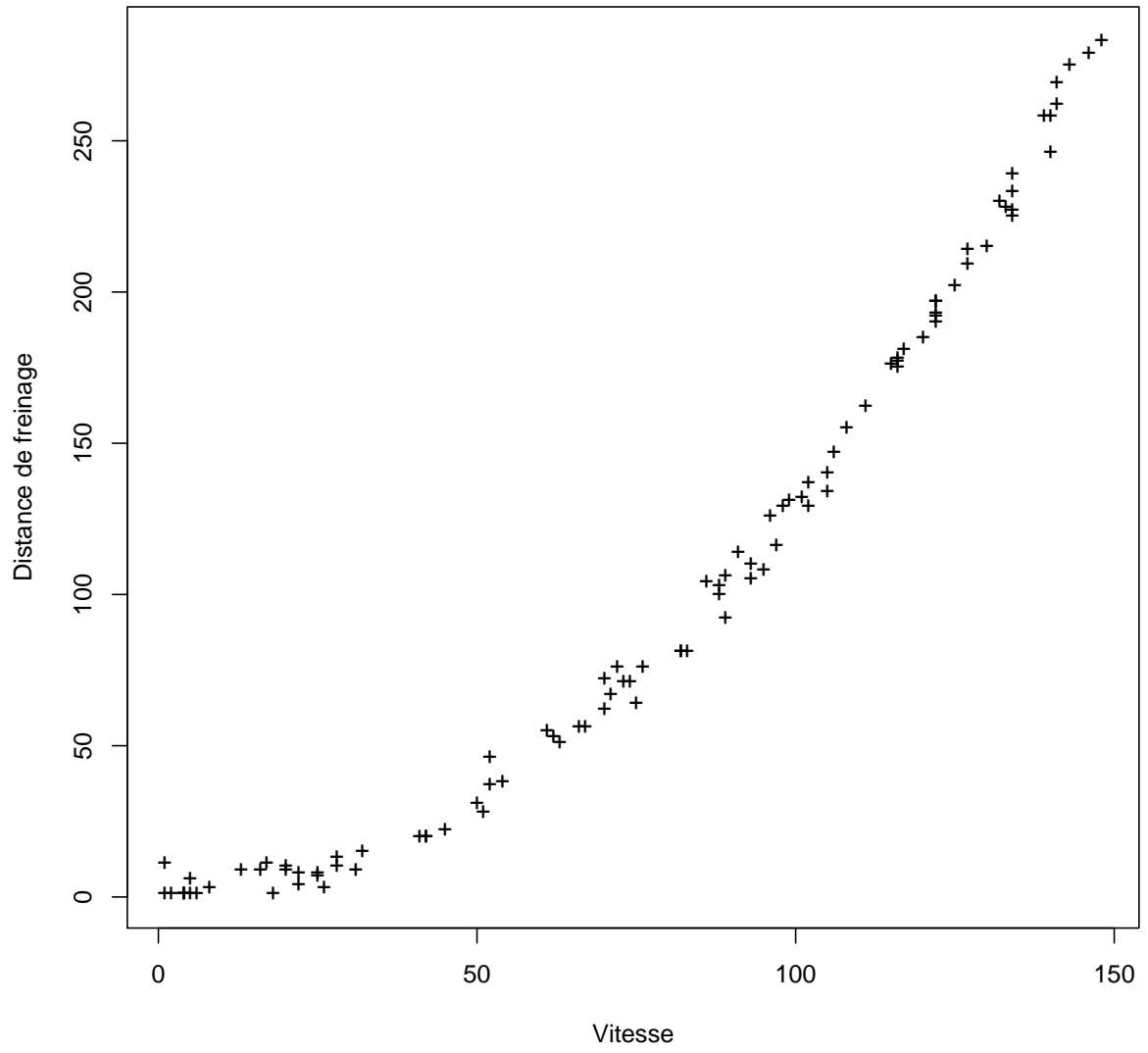
5. On s'intéresse, à présent, à la régression de la distance de freinage en la vitesse au carré ($Vitesse^2$). Pour cette régression, calculez le coefficient de régression linéaire. Commentez sa valeur en la comparant à celle trouvée à la question précédente.

6. Déterminez l'équation de la droite de régression de la distance de freinage en la vitesse au carré. En déduire

la relation obtenue entre la vitesse et la distance de freinage.

7. Quelle est la distance de freinage estimée pour une vitesse de 100 km/h ?

8. Sur la figure de la page suivante, représentant le nuage de points des deux variables étudiées, tracez la courbe de la relation trouvée dans la question 6. On fera apparaître quelques points de construction.



Espace supplémentaire pour d'éventuels détails dans les calculs :