

DUT STID, 1^{ère} année & APPC
Statistique descriptive
 Devoir 1 du Jeudi 4 novembre 2010

Important !

- Les réponses sont à donner directement sur le sujet. N'oubliez pas de noter votre nom.
- Les documents sont interdits et les calculatrices autorisées ; les téléphones portables même utilisés comme calculatrice, sont formellement interdits sur les tables.
- Sauf indication contraire, les notations utilisées sont celles du cours.
- Il sera tenu compte de la présentation et de la justification des résultats : tout résultat non justifié ne donnera lieu à aucun point.
- Le barème fourni est seulement indicatif : il peut évoluer.

Noms : /30

Exercice 1 Fécondité des femmes de 20 à 30 ans...../12

Le tableau ci-dessous donne l'âge (âge révolu en années entières) lors de la naissance de leurs enfants des femmes âgées de 20 à 30 ans ayant eu un enfant en France en 2009.¹

Âge	20 ans	21 ans	22 ans	23 ans	24 ans	25 ans	26 ans	27 ans	28 ans	29 ans	30 ans
Fréq.	2,81%	3,75%	4,92%	6,10%	7,57%	9,38%	10,93%	12,45%	13,65%	14,07%	14,37%
f_i^*	2,81%	6,56%	11,48%	17,58%	25,15%	34,53%	45,46%	57,91%	71,56%	85,63%	1
$f_i x_i$	0,562	0,788	1,082	1,403	1,817	2,345	2,842	3,362	3,822	4,080	4,311
$f_i x_i^2$	11,20	16,32	23,72	32,27	43,78	58,75	73,68	90,40	107,41	118,58	129,60

Remarque : Vous pouvez utiliser les lignes supplémentaires du tableau ci-dessus pour détailler vos résultats. Vous pouvez également utiliser l'espace supplémentaire laissé en page 3.

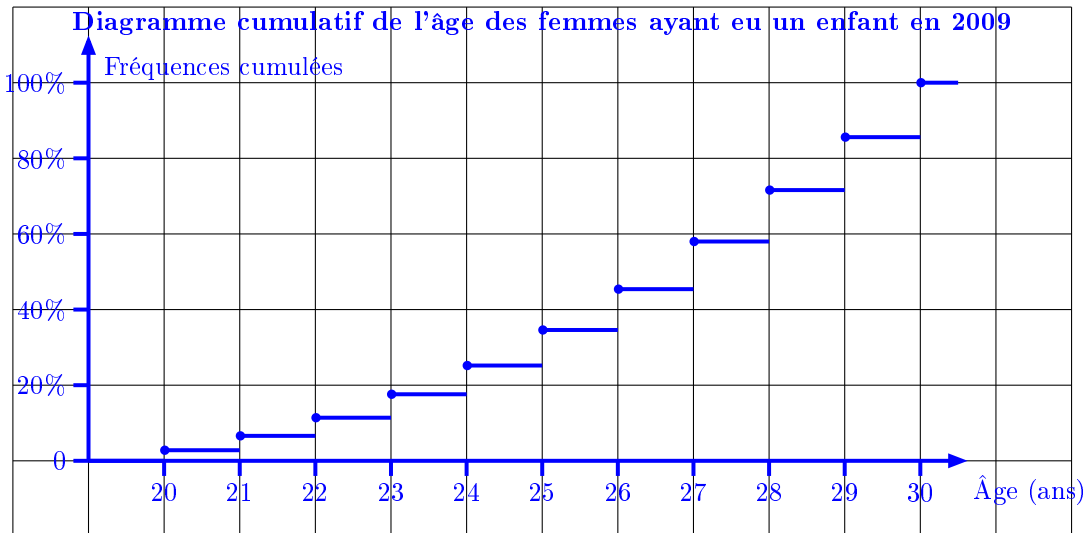
1. Quelle est la population étudiée? Quelle est la variable étudiée? Quel est le type de la variable étudiée?

Réponse :

La population étudiée est l'ensemble des femmes âgées de 20 à 30 ans ayant eu un enfant en 2009.
/1
 La variable étudiée est l'âge (âge révolu en années entières); elle est de type quantitatif discret.
/1

2. Construire, sur le quadrillage ci-dessous, le diagramme cumulatif de la distribution.

1. Source : Données INSEE, <http://www.insee.fr>.



- /1,5
3. Calculer l'âge moyen des femmes âgées de 20 à 30 ans ayant eu un enfant en 2009. Calculer l'écart type de l'âge des femmes âgées de 20 à 30 ans ayant eu un enfant en 2009.

Réponse :

On a :

$$\bar{X} = \sum_i f_i x_i \simeq 26,414 \text{ ans}$$

d'après le tableau de la page 1..... /1,5

De plus,

$$\text{Var}(X) = \sum_i f_i x_i^2 - \bar{X}^2 = 705,71 - 26,414^2 \simeq 8,01$$

et donc $\sigma(X) = \sqrt{8,01} \simeq 2,83 \text{ ans}$ /2

4. Déterminer, par la méthode de votre choix (que vous justifierez convenablement), les quartiles de l'âge des femmes âgées de 20 à 30 ans ayant eu un enfant en 2009.

Réponse :

1^{er} quartile : La première fréquence cumulée dépassant 25% est 25,15% qui correspond à $Q_1 = 24$ ans.

Médiane : La première fréquence cumulée dépassant 50% est 57,91% qui correspond à $m = 27$ ans.

3^{ème} quartile : La première fréquence cumulée dépassant 75% est 85,63% qui correspond à $Q_3 = 29$ ans.

..... /1

5. Dans un autre pays (appelé G), l'âge moyen des mères âgées de 20 à 30 ans ayant eu un enfant en 2009 est 23,12 ans avec un écart type de 3,34 ans. Une femme de 25 ans ayant eu un enfant en 2009 est-elle plus "précoce" en France qu'une femme de 22 ans ayant eu un enfant en 2009 dans le pays G ?

Réponse :

Calculons l'âge centré réduit de ces deux femmes :

- en France, l'âge centré réduit d'une femme de 25 ans est : $\frac{25-26,414}{2,83} \simeq -0,500$;
- dans le pays G, l'âge centré réduit d'une femme de 22 ans est : $\frac{22-23,12}{3,34} \simeq -0,335$.

Comme $-0,5 < -0,335$, une femme de 25 ans en France est plus précoce qu'une femme de 22 ans dans le pays G. /2

6. Si la population des femmes âgées de 20 à 30 ans ayant eu un enfant en 2009 avait systématiquement un autre enfant 3 ans plus tard, quel serait alors l'âge moyen de ces femmes au moment de la naissance de cet enfant suivant ? Quel serait l'écart type de cet âge ? Quelle serait la médiane de cet âge ?

Réponse :

Notons Y la variable donnant l'âge des mères au moment de la naissance de l'enfant suivant. On a $Y = X + 3$. Ainsi, l'âge moyen de ces femmes au moment de la naissance de l'enfant suivant serait :

$$\bar{Y} = \bar{X} + 3 = 29,414 \text{ ans.}$$

L'écart type de leurs âges serait :

$$\sigma(Y) = \sigma(X) = 2,83 \text{ ans.}$$

et la médiane de leurs âges serait $m_Y = m_X + 3 = 30 \text{ ans.} \dots \dots \dots /2$

Exercice 2 PIB dans le monde/8

Le tableau ci-dessous donne le regroupement en classes des PIB de 182 pays répertoriés par le FMI en 2009².

PIB	n_i	n_i^*	f_i^*	c_i ($\times 10^3$)	$n_i c_i$ ($\times 10^3$)	$(n_i c_i)^*$ ($\times 10^3$)	v_i	$f_i(v_i + v_{i-1})$
[0, 1 200[17	17	9,34%	10,2	173,4	173,4	0,01%	0,000
[1 200; 4 000[18	35	19,23%	46,8	842,4	1 015,8	0,04%	0,000
[4 000; 8 000[19	54	29,67%	114	2 166	3 181,8	0,12%	0,000
[8 000; 12 000[16	70	38,46%	160	2 560	5 741,8	0,21%	0,000
[12 000; 20 000[14	84	46,15%	224	3 136	8 877,8	0,32%	0,000
[20 000; 50 000[28	112	61,54%	980	27 440	36 317,8	1,32%	0,003
[50 000; 100 000[14	126	69,23%	1 050	14 700	51 017,8	1,85%	0,002
[100 000; 225 000[18	144	79,12%	2 925	52 650	103 667,8	3,76%	0,006
[225 000; 500 000[20	164	90,11%	7 250	145 000	248 667,8	9,01%	0,014
[500 000; 15 000 000[18	182	100,00%	139 500	2 511 000	2 759 667,8	100,00%	0,108
Total	182	✓	✓	✓	2 759 667,8	✓	✓	0,133

Remarque : Vous pouvez utiliser les colonnes supplémentaires du tableau ci-dessus pour détailler vos résultats. Vous pouvez également utiliser l'espace supplémentaire laissé en page 4.

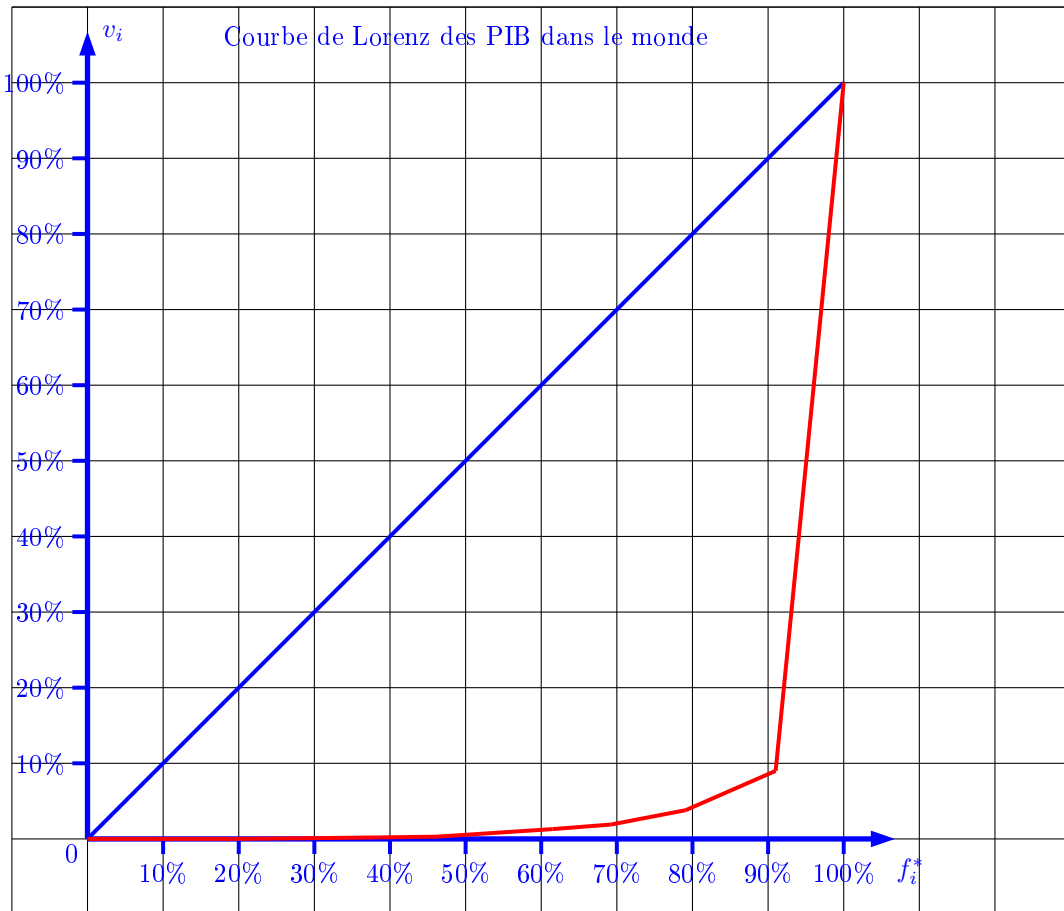
1. Quelle est la population étudiée? Quelle est la taille de cette population? Quelle est la variable étudiée? Quel est le type de cette variable?

Réponse :

La population étudiée est un ensemble de 182 pays dont le PIB a été répertorié par le FMI en 2009. Sa taille est $N = 182$/1
 La variable étudiée est le PIB de ces pays. C'est une variable de type quantitatif continu./1

2. Construire la courbe de Lorenz du PIB dans ces 182 pays sur le quadrillage ci-dessous.

2. Source : BIP 2009 nach Ländern, In der *World Economic Outlook Database*, April 2010, des Internationalen Währungsfonds.



Calculs /2
 Graphique /1

3. Calculer l'indice de Gini des PIB dans ces 182 pays.

Réponse :

D'après le tableau,

$$G = 1 - 0,133 = 0,867.$$

..... /2

4. Commenter la courbe de Lorenz et la valeur de l'indice de Gini.

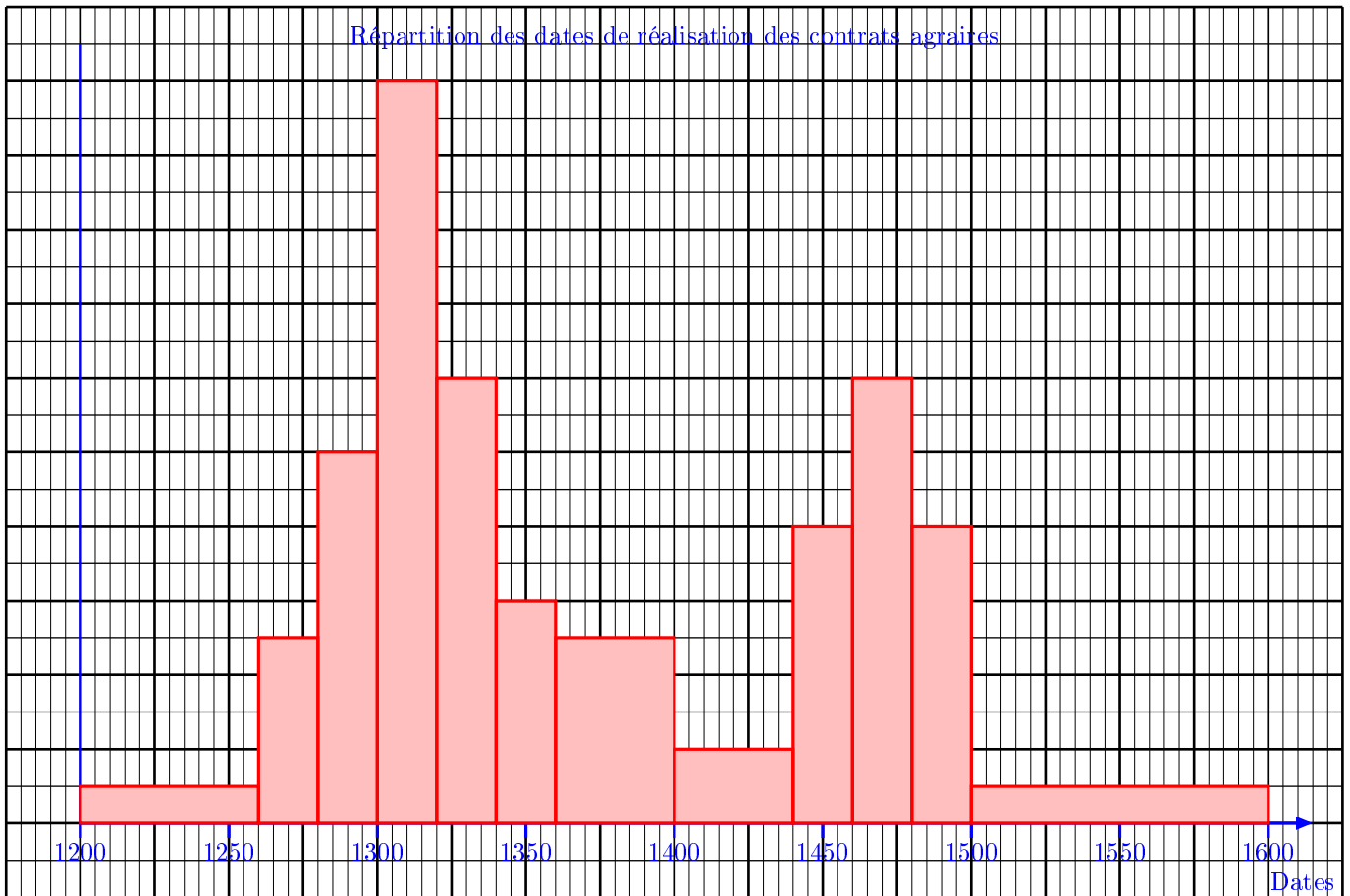
Réponse :

La courbe de Lorenz est éloignée de la diagonale et l'indice de Gini est fort : c'est le signe d'une forte concentration des PIB et donc d'une répartition très inégalitaire des richesses mondiales entre les pays. /1

Exercice 3 Dates d'enregistrement de contrats agraires /10

L'histogramme ci-dessous donne la répartition de la date de réalisation de contrats agraires tous issus d'un même corpus de documents rédigés dans le Lot (seigneurie de Castelnau Montratier) entre 1200 et 1600.³

³ Ces documents sont conservés aux archives du Lot et les données sont adaptées du projet GraphComp : <http://graphcomp.univ-tlse2.fr>.



Dates	n_i	Largeur	Hauteur	Aire	n_i^*	c_i	$n_i(c_i - \bar{X})^4$ ($\times 10^9$)
[1200; 1260[120	2,4	0,5	1,2	120	1230	52,61
[1260; 1280[$\frac{2 \times 120}{1,2} = 200$	0,8	2,5	2	320	1270	24,03
[1280; 1300[400	0,8	5	4	720	1290	20,59
[1300; 1320[800	0,8	10	8	1 520	1310	14,02
[1320; 1340[480	0,8	6	4,8	2 000	1330	1,92
[1340; 1360[240	0,8	3	2,4	2 240	1350	0,09
[1360; 1400[400	1,6	2,5	4	2 640	1380	0,00
[1400; 1440[160	1,6	1	1,6	2 800	1420	0,67
[1440; 1460[320	0,8	4	3,2	3 120	1450	10,29
[1460; 1480[480	0,8	6	4,8	3 600	1470	39,59
[1480; 1500[320	0,8	4	3,2	3 920	1490	56,55
[1500; 1600[200	4	0,5	2	4 120	1550	188,87
Total	4 120	8 cm	✓	41,2 cm ²	✓	✓	409,23

Remarque : Vous pouvez utiliser les colonnes supplémentaires du tableau ci-dessus pour détailler vos résultats. Vous pouvez également utiliser l'espace supplémentaire laissé en page 6.

1. Quelle est la population étudiée? Quelle est la variable étudiée? Quel est son type?

Réponse :

La population étudiée est un ensemble de contrats agraires..... /1
La variable étudiée est l'année de réalisation du contrat, de type quantitative discrète. /1

2. À partir de l'histogramme et du tableau ci-dessous, complétez les deux colonnes du tableau avec les classes et leurs effectifs respectifs. *Indication : Pour cela, on pourra, par exemple, déterminer l'aire, en cm² de chaque classe sur l'histogramme.*

Réponse :

Les résultats sont donnés directement dans le tableau..... /3

3. Déterminer la médiane de cette série statistique.

Réponse :

On a : $\frac{N}{2} = \frac{4120}{2} = 2060$. Le premier effectif cumulé à dépasser 2 060 et 2 240 qui correspond à la classe [1340; 1360[. Ainsi, on a :

$$m = 1340 = (1360 - 1340) \times \frac{2060 - 2000}{2240 - 2000} = 1345.$$

..... /2

4. On donne : $\bar{X} = 1374,7$ et $\sigma(X) = 83,8$. Calculer le coefficient d'aplatissement de cette série statistique. Commentez celui-ci en vous appuyant sur l'histogramme.

Question bonus! (Hors barème mais donnant droit à un point supplémentaire et un bonbon) : Pouvez-vous trouver une explication historique à ce phénomène?

Réponse :

Le moment centré d'ordre 4 de la série statistique est :

$$\mu_4 = \frac{409,23 \times 10^9}{4120} \simeq 99,33 \times 10^6$$

donc le coefficient d'aplatissement est égal à

$$\gamma_2 = \frac{99,33 \times 10^6}{83,8^4} \simeq 2,01$$

..... /2

Le coefficient d'aplatissement est celui d'une distribution proche de la loi uniforme. Cependant, cela ne correspond pas du tout à l'histogramme trouvé : cela s'explique par le fait que la distribution est bimodale (deux pics présents) et que le coefficient d'aplatissement n'est pas adapté pour l'interprétation dans ce cas..... /1

La raison de la bimodalité de la distribution est la guerre de cent ans (1337-1453) durant laquelle le nombre de contrats réalisés a brutalement chuté..... /1