

DATE : ...24.../...2.../...2022  
 EPREUVE: Statistiques et informatique appliquées aux Sciences sociales  
 ENSEIGNANT: ...Renaud Orain <renaud.orain@univ-paris1.fr> .....  
 DIPLOME : Licence de Sciences sociales, 3<sup>e</sup> année  
 Étudiant-es de tous parcours, inscrit-es en contrôle continu...  
 DUREE : ...3. h  
 DOCUMENT AUTORISE Calculatrice. Aucun document autorisé (cf. rappels en fin de sujet)

*Total sur 23 points, ramenés à un maximum de 20/20 pour toutes les notes supérieures à 16/20.*

Les temps d'attente sur la plate-forme Parcoursup se répartissent de la façon suivante parmi les candidat·es à une licence du domaine SHS à l'Université d'Ile de France :

Variable d'analyse : temps d'attente avant la proposition de la formation étudiée en semaines								
Nombre de semaines d'attente avant réponse de la formation	Effectif	Moyenne exacte	Somme des temps d'attente par classe	Amplitude de des classes	<i>Densité</i>	<i>Valeurs</i>	<i>Effectifs cumulés croissants</i>	Somme des carrés (pour le calcul de la variance et de l'écart-type)
]0 ; 1[	299	0,350	104,710	<i>1</i>	<i>299</i>	<i>0</i>	<i>0</i>	541436,65
[1 ; 2[	178	1,417	252,144	<i>1</i>	<i>178</i>	<i>1</i>	<i>299</i>	
[2 ; 4[	67	2,548	170,720	<i>2</i>	<i>33,5</i>	<i>2</i>	<i>477</i>	
[4 ; 8[	67	6,566	439,916	<i>4</i>	<i>16,75</i>	<i>4</i>	<i>544</i>	
[8 ; 12[	226	10,115	2 286,060	<i>4</i>	<i>56,5</i>	<i>8</i>	<i>611</i>	
12 et plus	1253	17,000	<i>21301</i>	<i>10</i>	<i>125,3</i>	<i>12</i>	<i>837</i>	
<b>Total</b>	<i>2090</i>	<i>11,75</i>	<i>24554,55</i>			<i>22</i>	<i>2090</i>	

NB : il ne s'agit pas ici des temps d'attentes globaux sur la plate-forme. Les candidat·es pouvaient recevoir une proposition alternative et abandonner le vœu en question. Dans ce cas leur temps d'attente pouvait être très inférieur à celui indiqué. Les temps d'attente en italique ont été imputés arbitrairement ; il s'agit de candidat·es à qui la formation n'aurait pas pu être proposée.

**Question 1 :** Évaluez la moyenne et l'écart-type de cette distribution, en complétant si besoin le tableau. Tracez un histogramme de cette distribution et indiquez quelle est la classe modale. Évaluez graphiquement au moyen d'un diagramme cumulé croissant ou d'un calcul d'interpolation les quartiles dont la médiane. Enfin, dessinez la boîte à moustache de cette distribution. **10,5 pts**

*Le calcul de la moyenne repose ici sur le remplissage du tableau ci-dessus. On a retrouvé la somme des temps d'attente dans la dernière classe en multipliant la moyenne de 17 semaines (manifestement arbitraire ici) par l'effectif de la classe, sachant que cette classe est manifestement composée de personnes n'ayant pas du tout reçu la proposition. Il était donc légitime de faire le calcul en excluant cette dernière classe et le résultat en a alors également été accepté.*

*Somme des durées pour la dernière classe = 17 \* 1253 = 21301*

*Somme totale de toutes les durées = 104,71 + 252,144 + ... + 21301 = 24554,55*

*Effectif total = 299 + 178 + ... + 1253 = 2090*

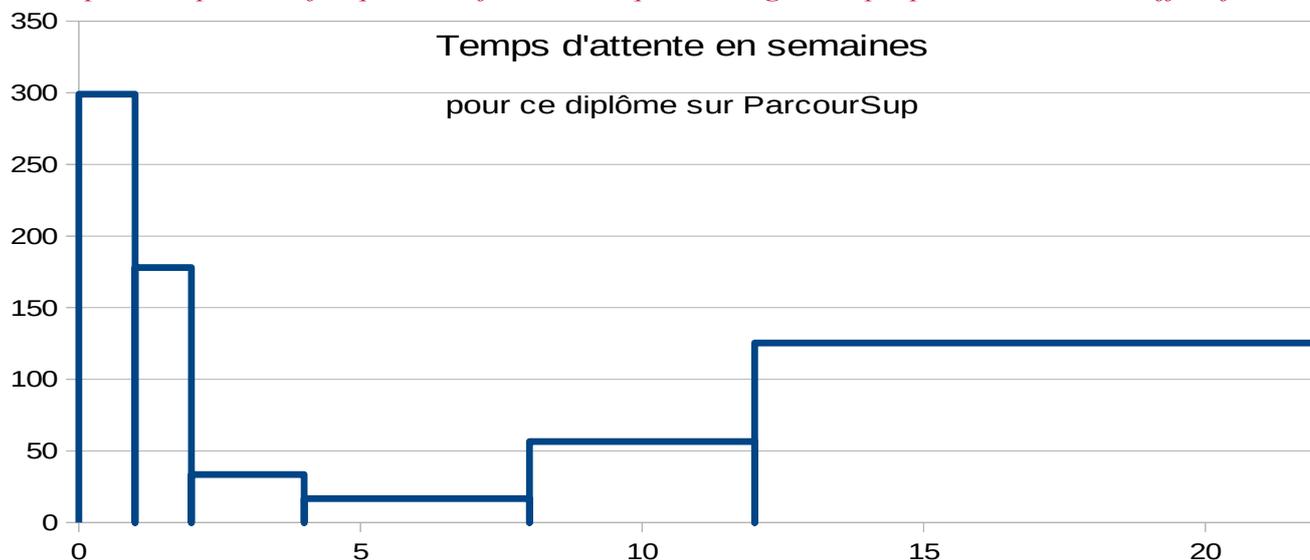
Durée d'attente moyenne en semaines =  $24554,55 / 2090 \approx 11,75$ .

Sans la dernière classe, Effectif = 837 ; Somme = 3253,55 ; Moyenne =  $3253,55/837 \approx 3,89$  semaines

<p>Variance théorique = Moyenne des carrés - Carré de la moyenne  Moyenne des carrés = <math>541436,65/2090 \approx 259,06</math>  Carré de la moyenne = <math>(24554,55/2090)^2 \approx 138,03</math>  Variance <math>\approx 259,06 - 138,03 \approx 121,03</math>  Écart-type = <math>\sqrt{\text{Variance}} \approx \sqrt{121,03} \approx 11,00</math></p>	<p>Pour évaluer la somme des carrés des valeurs sans la dernière classe on peut utiliser le tableau de la question suivante :  <math>\text{Var} \approx 27170,11/837 - 3,89^2 \approx 17,35</math></p>
--	--

Écart-type =  $\sqrt{17,35} \approx 4,17$  comme dans le tableau.

Histogramme : on utilise les densités calculées dans le tableau ci-dessus en divisant chaque effectif de classe par l'amplitude afin que la surface de chaque rectangle soit proportionnelle à cet effectif.

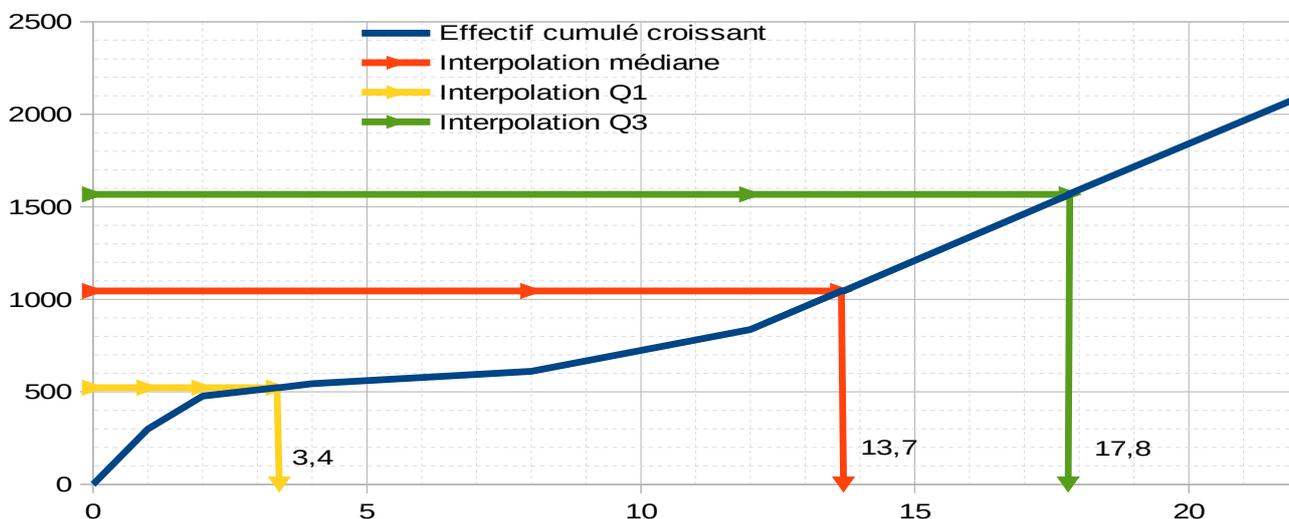


La classe modale est la première, constituée des personnes ayant attendu moins d'une semaine. Ceci tient au fonctionnement de la plate-forme : dans un premier temps, toutes les places sont proposées et on observe un certain nombre de désistements assez rapides du fait que l'afflux des propositions des autres formations est lui aussi très rapide. Dans un second temps, les nouvelles places proposées dépendent des désistements, ce qui provoque un ralentissement tendanciel. En fin de période, il y a une nouvelle accélération liée au fait que la plate-forme impose aux candidat·es de faire un choix définitif. Enfin on mesure l'importance numérique de la dernière classe, de loin la plus nombreuse mais qui ne permet pas ici de trouver le second mode du fait de la convention retenue (on a supposé que le max était à la même distance de la durée moyenne que le minimum de classe). Quoiqu'il en soit la classe modale est ici loin d'être la plus nombreuse.

Interpolation des quartiles :

On calcule  $n/2 = 1045$  (pour la médiane)  $n/4 = 522,5$  et  $3n/4 = 1567,5$  pour les quartiles

La médiane et  $Q_3$  sont donc dans l'intervalle  $[12;22]$ ,  $Q_1$  est dans l'intervalle  $[2;4]$ .

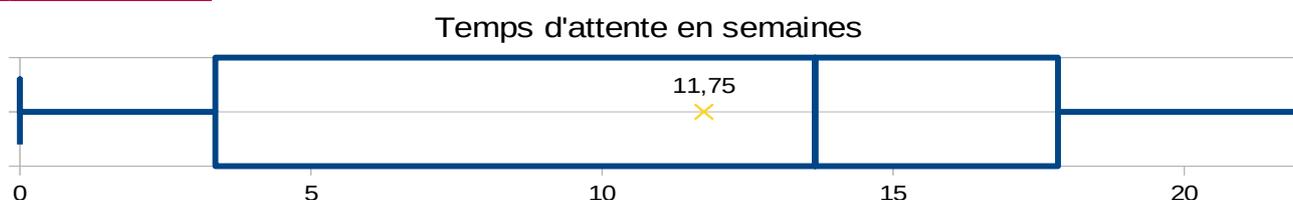


Pour le calcul d'interpolation on utilise la formule rappelée en annexe :

$$Qtl = b.\text{inf} + \frac{n \uparrow Qtl - n \uparrow \text{inf}}{n \uparrow \text{sup} - n \uparrow \text{inf}} \cdot (b.\text{sup} - b.\text{inf}) \quad \text{donc Médiane} = 12 + \frac{1045 - 837}{1253} \cdot 10 \approx 13,66 \text{ semaines}$$

$$\text{et } Q3 = 12 + \frac{1567,5 - 837}{1253} \cdot 10 \approx 17,83 \text{ semaines tandis que } Q1 = 2 + \frac{522,5 - 477}{67} \cdot 2 \approx 3,36 \text{ semaines.}$$

Boîte à moustache :



**Question 2 :** on étudie la façon dont les notes anticipées du baccalauréat se répercutent sur le temps d'attente par rapport à la formation du fait de la prise en compte de ces notes par la commission d'examen des vœux. Les 129 candidat·es sans notes renseignées aux épreuves anticipées du baccalauréat s'en sont vues imputer une grâce à une régression linéaire menée sur d'autres variables disponibles et on écarte de l'analyse les personnes à qui la proposition n'a finalement pas été faite.

Étudiez le tableau suivant puis évaluez la covariance et la corrélation entre le temps d'attente en semaines et les notes du baccalauréat anticipé

Déduisez-en les coefficients de la régression linéaire de ce temps d'attente en fonction des notes au baccalauréat anticipé, ainsi que le coefficient de détermination de la régression. **7,5 pts**

Variable	Somme des produits		Statistiques simples					
	Moy Bac anticipé	Temps d'attente	N	Moyenne	Ecart-type	Somme	Min.	Max.
Moy Notes Bac anticipé	182953,36	44668,22	837	14,65	2,01	12260	10	19,5
Temps d'attente	44668,22	27170,1080	837	3,89	4,17	3254	0	11,07

$$\text{Covariance} = \frac{\text{Moyenne des produits} - \text{Produit des moyennes}}{N} \approx \frac{44668,22 \cdot 12260 - 837 \cdot 3254}{837} \approx -3,58$$

$$\text{corrélation} = \frac{\text{covariance}}{\text{produit des écarts types}} \approx \frac{-3,58}{2,01 \cdot 4,17} \approx -0,428 \quad (\text{calcul fait avec moins d'arrondis})$$

$\text{temps d'attente} = a \cdot \text{notes} + b + \text{résidu}$  ;  $Y = \text{temps d'attente}$  ;  $X = \text{notes}$

$$Y = a \cdot X + b + U ; a = \frac{\text{cov}(X, Y)}{\text{var}(X)} ; b = \bar{Y} - a \cdot \bar{X} \quad ; \text{on recalcule } \text{var}(x) = 182953,4/837 - (12260/837)^2 \approx 4,03$$

donc  $a \approx -3,58 / 4,03 \approx -0,887$  et  $b = 3,89 + 0,887 \cdot 14,65 \approx 16,89$  (calculs faits avec moins d'arrondis)

$$R^2 = \frac{\text{Var expl}}{\text{Var tot}} = \frac{a^2 \cdot \text{Var}(X)}{\text{Var}(Y)} = \text{correlat}^2 \approx (-0,428)^2 \approx 18,3 \%$$

**Question 3 :** Commentez l'ensemble de vos résultats. Pour la corrélation, vous pourrez utiliser le résultat du test de Student qui indique que la probabilité d'obtenir la corrélation observée si le lien était dû au seul hasard de l'échantillonnage statistique est inférieure à 0,0001. Pour illustrer l'utilisation de ces notes imputées au baccalauréat anticipé et mieux appréhender le déroulement réel de la procédure Parcoursup, vous pourrez également vous appuyer sur le tableau ci-dessous, qui renseigne l'issue de cette procédure en fonction de ces notes. **Total : 5 pts**

Les temps d'attente sur Parcoursup sont très hétérogènes pour les 2090 personnes ayant candidaté sur la formation étudié puisque ces temps varient de moins d'une semaine pour un peu moins de trois candidat·es sur 20 et moins de 4 semaines pour plus d'un·e sur 4 à plus de 12 à 14 semaines pour plus de la moitié, les personnes concernées ayant simplement attendu jusqu'à la fin sans recevoir la proposition, ce qui conduit à un écart type de 11 semaines, du même ordre que la durée moyenne d'attente avant de recevoir la proposition.

*On est donc en présence d'une formation ne pouvant accueillir la totalité des personnes ayant candidaté et maintenu leur candidature. Ce type de formation est qualifié de « en tension » lorsqu'elles ne sont pas sélectives. Si l'on se centre sur les 837 personnes ayant une chance de recevoir la proposition, les temps d'attente tournent alors autour d'un peu moins de 4 semaines en moyenne, soit environ un mois, mais toujours avec de fortes disparités puisque l'écart type de ce temps d'attente est lui supérieur à 4 semaines pour ces mêmes personnes. On pourrait reprendre à ce sujet les indications données dans ce corrigé en dessous de l'histogramme : les premières réponses interviennent rapidement dans les limites de la capacité d'accueil puis les possibilités de désistements se raréfient à mesure que les candidat·es mettent de côté leurs vœux prioritaires ce qui conduit à un fort ralentissement. En fin de procédure, il y a à nouveau accélération, les candidat·es devant alors opter définitivement pour un vœu.*

*On constate une corrélation négative marquée et significative entre ces temps d'attente et les notes moyennes au baccalauréat anticipé (ou à défaut tout autre critère pouvant s'y substituer) : en moyenne, un gain d'un point sur cette moyenne tend à faire économiser plus de trois semaines et demie d'attente aux candidat·es, ce qui est considérable. On constate également que les personnes reçues le sont plus fréquemment que la moyenne avec une moyenne au baccalauréat anticipé supérieure à 14/20 et elles refusent également plus souvent avec ce type de notes, tandis que les personnes ayant des résultats inférieurs à ce seuil ont plus tendance à renoncer à ce vœu au profit d'une formation plus accessible ou à attendre en vain la proposition. Chez les personnes ayant eu moins de 12 ou de 10, le phénomène constaté est davantage un abandon de la plate-forme elle-même sans doute faute de réponses satisfaisantes.*

**Répartition des candidat·es selon l'issue de la procédure, détail des % en ligne selon la note réelle ou imputée au baccalauréat anticipé**

%ligne	Renonce	Refuse	Reçue	Non reçue	Sorties de route	Ensemble
<10	58,8 %	0,0 %	0,0 %	19,3 %	21,9 %	100,0 %
[10 ; 12[ (Passable)	59,0 %	3,6 %	4,2 %	18,8 %	14,4 %	100,0 %
[12 ; 14[ (AB)	60,2 %	8,1 %	4,8 %	15,8 %	11,1 %	100,0 %
[14 ; 16[ (B)	53,1 %	25,3 %	7,3 %	5,2 %	9,1 %	100,0 %
16 et + (TB)	20,7 %	62,2 %	6,9 %	0,4 %	9,8 %	100,0 %
<b>Total</b>	<b>53,5 %</b>	<b>16,0 %</b>	<b>4,8 %</b>	<b>12,9 %</b>	<b>12,9 %</b>	<b>100,0 %</b>

<b>Issue de la procédure Parcoursup (modalités détaillées)</b>	Khi <sup>2</sup>	662,33	
Renonce au vœu, avant d'avoir reçu la proposition	DL	16	
Refuse après réception de la proposition	Prob>Khi <sup>2</sup>	1,33E-130	
Reçue : Personne acceptée et venue s'inscrire	Significativité	5,00 %	1,00 %
Non-reçue : maintien du vœu mais sans proposition	Seuil	26,296	32,000
Sorties de route : diverses démissions de la plate-forme à différents stades dont personnes acceptées non inscrites	V de Cramer	0,28147	