

DATE :5.../...1.../...2022
 EPREUVE: Statistiques et informatique appliquées aux Sciences sociales
 ENSEIGNANT: ...Renaud Orain <renaud.orain@univ-paris1.fr>
 DIPLOME : Licence de Sciences sociales, 3^e année
 Étudiant-es de tous parcours, inscrit-es en contrôle terminal...
 DUREE : ...3. h
 DOCUMENT AUTORISE Calculatrice. Aucun document autorisé (cf. rappels en fin de sujet)

Total sur 25 points, ramenés à un maximum de 20/20 pour toutes les notes supérieures à 16/20.

Les notes reçues aux épreuves du baccalauréat anticipé se répartissent de la façon suivante parmi les candidat-es à une licence du domaine SHS à l'Université d'Île de France :

Variable d'analyse : moyneba Moyenne des notes du bac / epreuves anticipees							
Moyenne des notes du bac / epreuves anticipees	Effectif	Moyenne exacte	Amplitude	Densité	Seuils	Effectifs cumulés	Somme des carrés
]0 ; 5[3	4,69	5	0,6	0	0	327891,97
]5 ; 10[292	8,53	5	58,4	5	3	
]10 ; 12[(Passable)	478	10,91	2	239	10	295	Somme des valeurs du (rappel tableau initial)
]12 ; 14[(AB)	506	12,82	2	253	12	773	
]14 ; 16[(B)	438	14,79	2	219	14	1279	
]16 ; 18[(TB)	213	16,68	2	106,5	16	1717	
]18 ; 20[(TB)	31	18,54	2	15,5	18	1930	24809,75
Total	1961	12,65	20	98,05	20	1961	

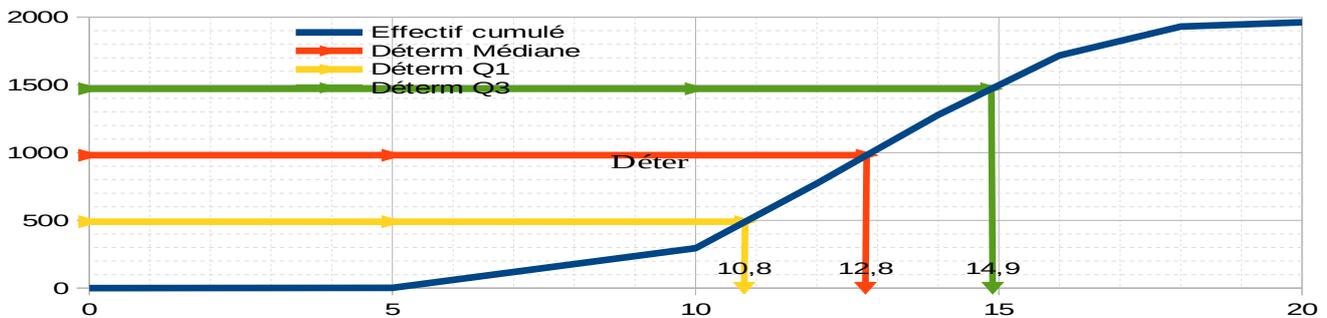
Question 1 : Évaluez la moyenne (0,5) et l'écart-type (2) de cette distribution, en complétant si besoin le tableau. Indiquez quelle est la classe modale (0,5) de cette distribution. Évaluez graphiquement au moyen d'un diagramme cumulé croissant ou d'un calcul d'interpolation la médiane (1). Un calcul similaire nous donne le premier (10,8) et le troisième quartile (14,9) . Enfin, dessinez la boîte à moustache de cette distribution (1). **Total : 5 pts**

$$\text{Moyenne} = 24809,75 / 1961 \approx 12,65 \text{ sur vingt ; écart type} = \sqrt{\frac{327891,97}{1961} - 12,65^2} \approx 2,673$$

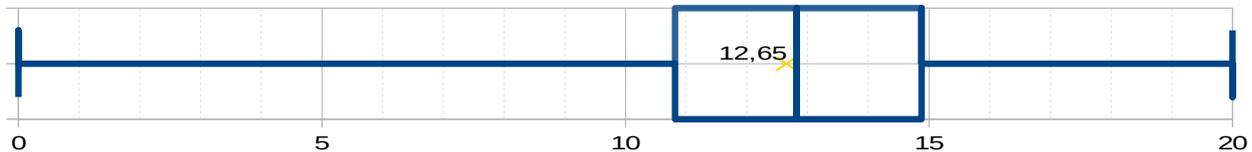
La classe modale est également la classe la plus nombreuse [12;14] car aucune classe n'a une amplitude plus faible que les deux points de celle-ci et ne pourrait donc compter plus de personnes par point.

La détermination de la médiane suppose de d'abord déterminer quelle classe contient cette médiane, à l'aide des effectifs cumulés croissants par note. La médiane correspond à un effectif cumulé croissant égal à 1961/2=980,5. Or on a 773 candidat-es avec moins de 12 et 1279 avec moins de 14 ; on peut donc en conclure que la médiane se situe entre 12 et 14.

Plus précisément, $Q_2 = 12 + \frac{980,5 - 773}{506} \cdot 2 \approx 12,82$ On trouvera l'interpolation graphique ci-dessous.



On utilise enfin ces résultats pour la boîte à moustache



Question 2 : on constate que 129 candidat·es n'ont pas de notes renseignées aux épreuves anticipées du baccalauréat et que cela est plus fréquent parmi les personnes déjà bacheliers (en réorientation ou reprise d'études) que parmi les personnes en train de passer le baccalauréat. Cependant, 36 personnes sans notes du baccalauréat anticipé ont une note moyenne du baccalauréat et 67 ont une mention renseignée. Ceci peut être lié une série professionnelle ou technologique où ces épreuves n'existaient pas.

Étudiez le tableau suivant puis évaluez la covariance (2) et la corrélation (1), pour les personnes dont à la fois les notes du baccalauréat anticipé et la moyenne au baccalauréat sont renseignées. NB : du fait de cette restriction, les moyennes et variances calculées pour cette question peuvent donc différer de celles calculées à la question 1.

Déduisez-en les coefficients (2) de la régression linéaire des notes au baccalauréat anticipé en fonction de la moyenne de l'ensemble du baccalauréat, ainsi que le coefficient de détermination (2) de la régression. **Total : 7 pts**

Variable : notes	Somme des produits		Statistiques simples					
	Bac anticipé	Moy. Bac.	N	Moyenne	Ecart-type	Somme	Min.	Max.
Bac anticipé	97412,9412	100075,7310	628	12,21012	2,45738	7668	6,000	18,66667
Moy. Bac.	100075,7310	104874,4635	628	12,73627	2,18922	7998	8,570	19,58000

On peut utiliser la formule $Covariance = Moyenne\ des\ produits - Produit\ des\ moyennes$
Moyenne produits = $100075,7/628 \approx 159,356$ et $Produit\ moyennes = 12,212 \cdot 12,736 \approx 155,511$

$Covariance \approx 159,356 - 155,511 \approx 3,845$ et $Corrélation = \frac{covariance}{Produit\ des\ écarts\ types} \approx \frac{3,845}{2,457 \cdot 2,189} \approx 0,715$

Pour la régression linéaire, on utilise $Y = a \cdot X + b + U$; $a = \frac{cov(X, Y)}{var(X)}$; $b = \bar{Y} - a \cdot \bar{X}$ avec X moyenne du baccalauréat et Y moyenne des notes du baccalauréat anticipées,

$Var(X) \approx \frac{104874,5}{628} - 12,736^2 \approx 2,189^2 \approx 4,785$ donc $a \approx \frac{3,845}{4,785} \approx 0,80$ et $b \approx 12,21 - 0,80 \cdot 12,74 \approx 2,0$

Coefficient de détermination $R^2 = \frac{Var\ expl}{Var\ tot} = \frac{a^2 \cdot Var(X)}{Var(Y)} = correat^2$ donc $R^2 \approx 0,715^2 \approx 51\%$ (1 point)

Question 3 : On utilise les notes réelles ou imputées conformément à une méthode similaire à celle utilisée dans la question précédente et on construit le tableau détaillant la répartition des candidat·es en fonction de ces notes et de l'issue de la procédure. Complétez les différents tableaux (4) ci-dessous, calculez le PEM (0,5) et le χ^2 (1,5) du tableau initial et tracez le graphique des écarts à l'indépendance manquant (2). **Total : 8 pts**

On trouvera les tableaux complétés ci-après.

%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 %
Total	53,5 %	16,0 %	4,8 %	12,9 %	12,9 %	100,0 %

Effectifs théoriques	Renonce	Refuse	Reçue	Non reçue	Sorties de route	Ensemble
<10	161,0	48,1	14,4	38,7	38,7	301
[10 ; 12[(Passable)	279,2	83,4	25,0	67,2	67,2	522
[12 ; 14[(AB)	298,5	89,2	26,7	71,8	71,8	558
[14 ; 16[(B)	247,7	74,0	22,2	59,6	59,6	463
16 et + (TB)	131,6	39,3	11,8	31,7	31,7	246
Total	1118	334	100	269	269	2090

On note qu'aucun effectif théorique n'est inférieur à 5 personnes, ce qui permet de justifier l'emploi direct du test du Khi².

Écarts à l'indépendance	Renonce	Refuse	Reçue	Non reçue	Sorties de route	Ensemble
<10	16,0	-48,1	-14,4	19,3	27,3	0,00
[10 ; 12[(Passable)	28,8	-64,4	-3,0	30,8	7,8	0,00
[12 ; 14[(AB)	37,5	-44,2	0,3	16,2	-9,8	0,00
[14 ; 16[(B)	-1,7	43,0	11,8	-35,6	-17,6	0,00
16 et + (TB)	-80,6	113,7	5,2	-30,7	-7,7	0,00
Total	0,00	0,00	0,00	0,00	0,00	0,00

Somme des écarts positifs à l'indépendance = 357,665 environ et Somme des écarts positifs à l'indépendance pour des écarts maximaux avec une structure similaire = 1096,025 environ

Donc PEM = 357,665/1096,025 ≈ 32,6 %

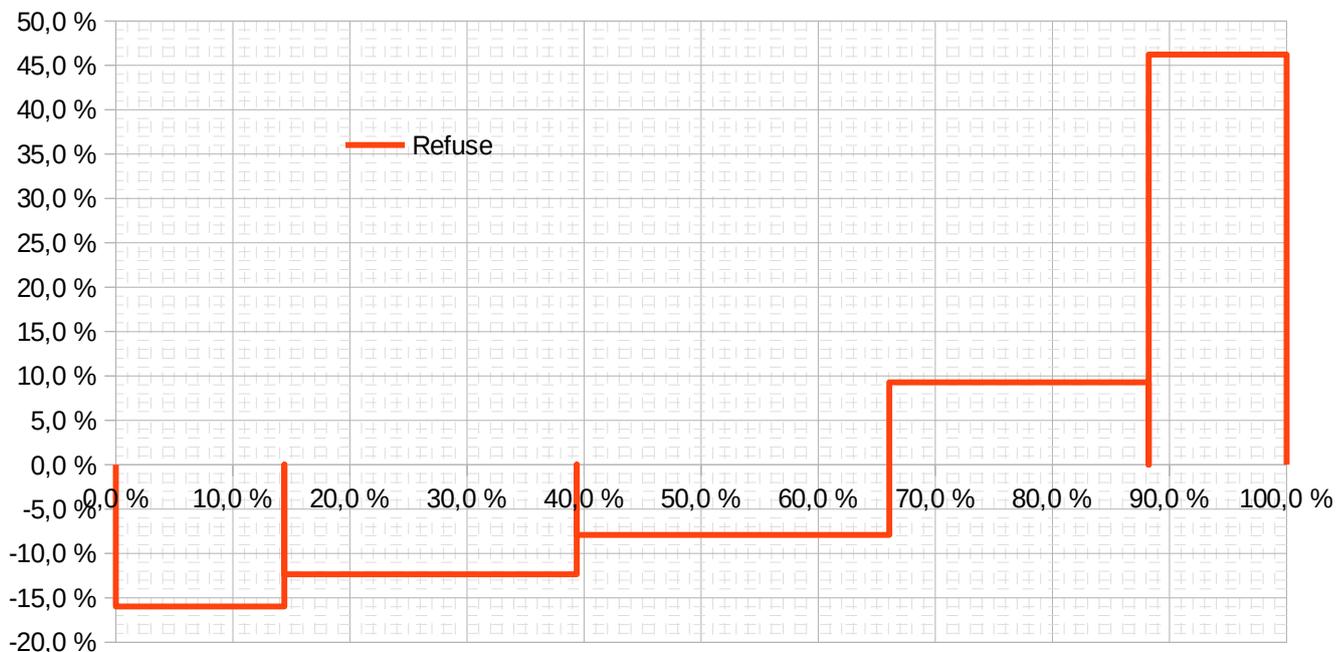
Contributions au Khi ²	Renonce	Refuse	Reçue	Non reçue	Sorties de route	Ensemble
<10	1,59	48,10	14,40	9,57	19,18	92,85
[10 ; 12[(Passable)	2,96	49,75	0,35	14,13	0,91	68,11
[12 ; 14[(AB)	4,71	21,88	0,00	3,65	1,34	31,59
[14 ; 16[(B)	0,01	25,00	6,34	21,26	5,19	57,80
16 et + (TB)	49,36	328,77	2,32	29,69	1,85	412,00
Total	58,63	473,50	23,42	78,30	28,48	

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

Graphiques des écarts à l'indépendance

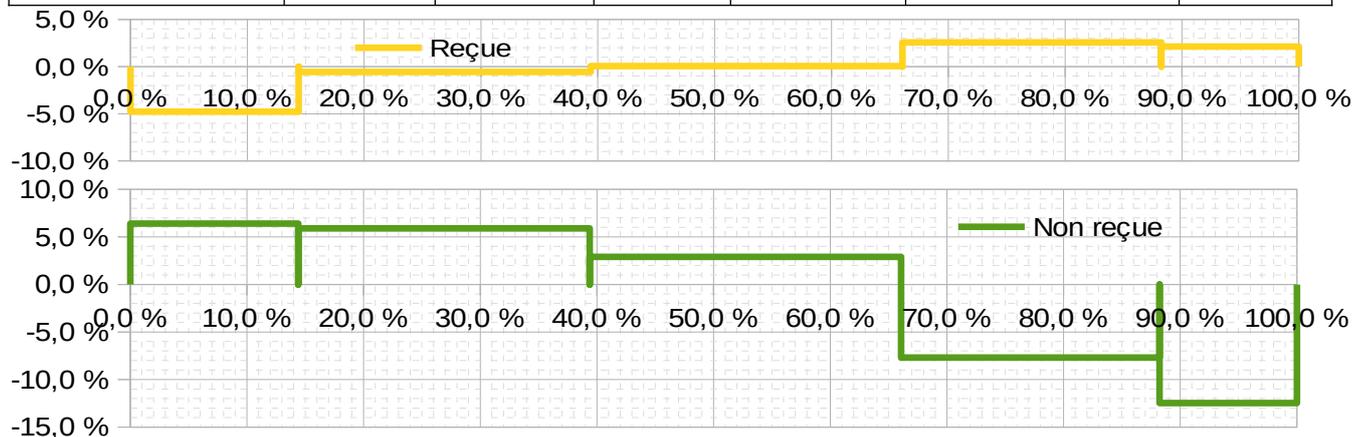
Le graphique manquant était celui des refus que l'on trouvera ci-dessous : il était nécessaire de calculer les écarts à la moyenne ainsi que les % de chaque mention afin de l'établir, comme suit :

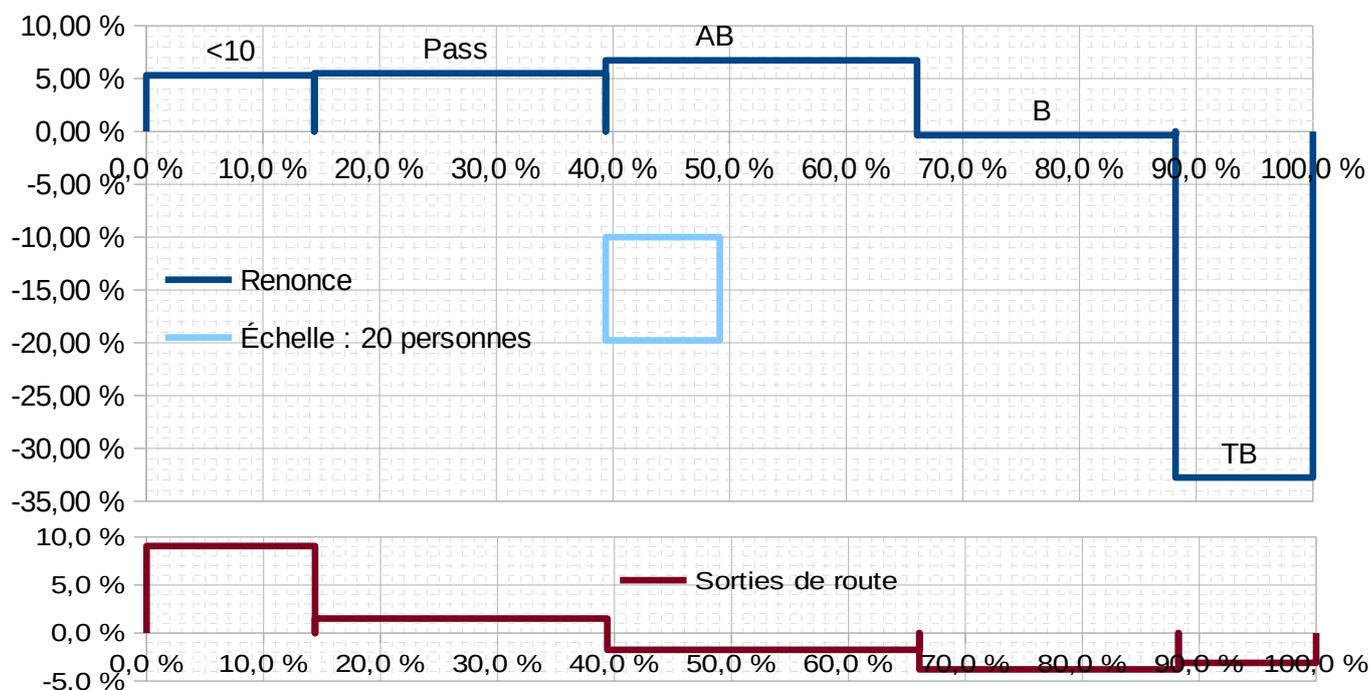
Écart à la moyenne	Renonce	Refuse	Reçue	Non reçue	Sorties de route	%col marginal
<10	5,3 %	-16,0 %	-4,8 %	6,4 %	9,1 %	14,4 %
[10 ; 12[(Passable)	5,5 %	-12,3 %	-0,6 %	5,9 %	1,5 %	25,0 %
[12 ; 14[(AB)	6,7 %	-7,9 %	0,1 %	2,9 %	-1,8 %	26,7 %
[14 ; 16[(B)	-0,4 %	9,3 %	2,6 %	-7,7 %	-3,8 %	22,2 %
16 et + (TB)	-32,8 %	46,2 %	2,1 %	-12,5 %	-3,1 %	11,8 %



Répartition des candidat·es selon la note réelle ou imputée au baccalauréat anticipé ainsi que selon l'issue de la procédure (tableau initial)

Effectifs	Renonce	Refuse	Reçue	Non reçue	Sorties de route	Ensemble
<10	177	0	0	58	66	301
[10 ; 12[(Passable)	308	19	22	98	75	522
[12 ; 14[(AB)	336	45	27	88	62	558
[14 ; 16[(B)	246	117	34	24	42	463
16 et + (TB)	51	153	17	1	24	246
Total	1118	334	100	269	269	2090





Question 4 : 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. **Total : 5 pts**

On étudie ici en premier lieu les notes obtenues par les personnes candidatant sur Parcoursup à une formation du domaine SHS (Sciences humaines et sociales) aux épreuves anticipées du baccalauréat. Il nous est indiqué que la plupart (94%) de ces candidat-es ont une note moyenne renseignée soit un effectif de 1961 personnes dont la plupart avait au moins 10/20, près des trois-quarts dépassant la note de 11/20 tandis que moins d'un quart de ces personnes avaient dépassé la note de 15/20, avec une moyenne de 12,6, un écart type de 2,7 points et une médiane proche de 13/20.

De manière assez prévisible, ces résultats anticipés, qui interviennent dans le calcul de la note globale au baccalauréat, apparaissent fortement et significativement corrélés positivement avec ce résultat définitif pour les 628 personnes dont on connaît les deux à la fois (avec une corrélation positive de 0,715) : on peut noter au passage que cette information n'est connue que pour environ un tiers des personnes candidates, essentiellement suite à une réorientation. À l'inverse, on a sans doute une proportion de néo-bacheliers de l'ordre des deux tiers dans ce vivier, dont les résultats définitifs ne sont donc pas encore connus.

Ceci étant précisé, on entreprend ici d'effectuer une imputation d'une note au baccalauréat anticipé pour ces personnes déjà bachelières pour lesquelles cette note est inconnue ou n'a pas de sens (pour certains diplômes étrangers ou pour les baccalauréats professionnels par exemple), sachant que la moyenne générale au baccalauréat permet de prédire 51 % des variations des notes au baccalauréat anticipé lorsque les deux sont connues et qu'un gain d'un point sur la note globale se traduit en moyenne par un gain de 0,8 points prévisible sur la moyenne des épreuves anticipées. Ceci permet donc de construire ce qu'on appelle une proxy de la note du baccalauréat anticipé pour les personnes qui n'ont pas une telle note, qui a l'avantage d'être déjà connue avec précision pour l'essentiel de l'échantillon pour des cas résiduels où l'on a d'autres informations, plus complètes, à savoir le résultat global à l'examen.

Il apparaît assez probable que la commission examinant les vœux sur cette formation ait pris en compte ces notes ou une proxy à défaut puisque l'issue de la procédure apparaît significativement et fortement liée à ce résultat. On constate en effet que les personnes reçues le sont plus fréquemment que la moyenne avec une moyenne au baccalauréat anticipé supérieure à 14/20 mais c'est pour le cas où elles ont les moyens de refuser la formation considérée grâce à ce type de notes que le contraste apparaît le plus marqué. À l'inverse, 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. Finalement, seule une minorité ayant accès à la formation, c'est ce contraste entre les personnes qui ont pu refuser le vœu plutôt que d'y renoncer parce qu'elles ne l'avaient pas reçu qui apparaît le plus marqué, chez les personnes ayant eu plus de 16/20. Chez les personnes ayant eu moins de 12 voire de 10, le phénomène constaté est davantage un abandon de la plate-forme elle-même sans doute faute de réponses satisfaisantes. Une partie pourrait correspondre à un échec à l'examen lui-même.