



UNIVERSITÉ PARIS 1  
**PANTHÉON SORBONNE**

DATE : .....21/12/2020-5/1/2021.....  
 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 conseillée: ...2. h  
 DOCUMENT AUTORISE Devoir à la maison (tous documents autorisés)

**Sujet obligatoire pour les étudiant-es inscrit-es en contrôle terminal**

Ce devoir devra être rendu sous forme manuscrite, en main propre ou en version scannée ou photographiée sur le devoir. En cas d'impossibilité, prière de me contacter au moyen de l'adresse de courriel indiquée ci-dessus et de contacter votre chargé de TD en expliquant bien la situation.

Il vous est demandé de traiter les questions qui suivent, en faisant les calculs « à la main » (calculatrices autorisées, y compris sur ordinateur) en indiquant si possible le temps nécessaire pour réaliser le devoir. Cette mention ne sera pas prise en compte dans l'évaluation. Elle vise à mieux calibrer les sujets futurs.

**Distribution d'âge et sorties culturelles**

Les étudiants en Master Démographie en 2019-2021 ont construit un questionnaire portant sur les sorties et pratiques culturelles des jeunes mais ce questionnaire est accessible en ligne à quiconque.

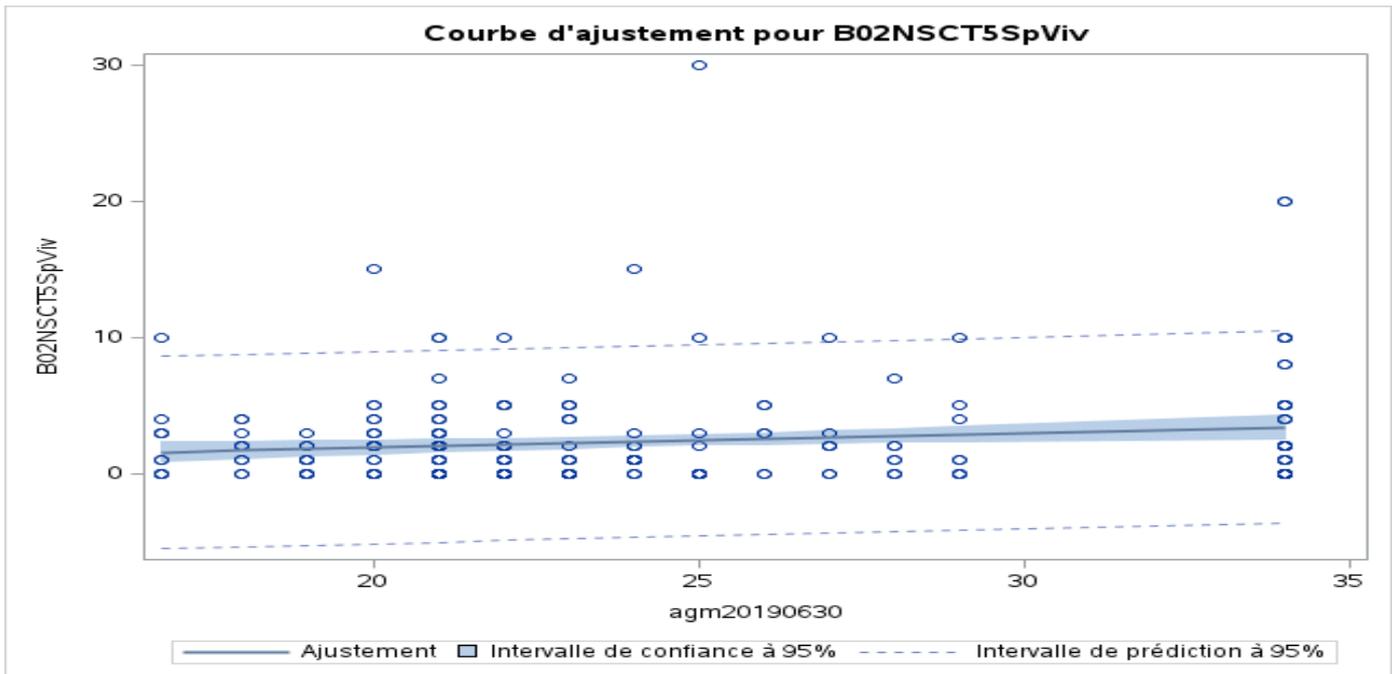
On se demande d'abord comment varie le nombre de sorties en 2019 en fonction de l'âge en milieu d'année 2019. On se restreint aux personnes ayant répondu sur leur année de naissance et on évalue les corrélations de l'âge avec un décompte de différent type de sortie faites dans l'année.

Coefficients de corrélation de Pearson, N = 253 Proba >  r  sous H0: Rho=0							Mo- yennes	Ec- type	Coefficients de variation
Âge Nb sorties	Age mi 2019	1:cinéma	2 :Concert	3 :Musee	4 :Monuments	5 : Cirque, théâtre			
Age mi 2019	1,00000	-0,03598 0,5689	0,12422 0,0484	-0,06295 0,3186	-0,02928 0,6430	0,15574 0,0131	24,19	5,21	0,22
1:cinéma	-0,03598 0,5689	1,00000	0,21627 0,0005	0,31075 <,0001	0,23937 0,0001	0,29990 <,0001	7,15	6,60	0,92
2 :Concert	0,12422 0,0484	0,21627 0,0005	1,00000	0,15306 0,0148	0,18197 0,0037	0,33887 <,0001	3,40	4,96	1,46
3 :Musee	-0,06295 0,3186	0,31075 <,0001	0,15306 0,0148	1,00000	0,60333 <,0001	0,42451 <,0001	6,80	7,03	1,03
4 :Monuments	-0,02928 0,6430	0,23937 0,0001	0,18197 0,0037	0,60333 <,0001	1,00000	0,29076 <,0001	5,49	5,76	1,05
5 : Cirque, théâtre	0,15574 0,0131	0,29990 <,0001	0,33887 <,0001	0,42451 <,0001	0,29076 <,0001	1,00000	2,33	3,60	1,54

1. À l'aide des sommes et sommes de produits ci-dessous, expliquer les étapes du calcul de la moyenne et de la variance de l'âge moyen en 2019, puis de la corrélation entre l'âge et le nombre de sorties au cirque ou au théâtre dans l'année.

Somme des produits	Âges mi 2019	Cirque, théâtre	Somme	Effectif
Âges mi 2019	154 924,2	15 010,8	6120,25	253
Cirque, théâtre	15 010,8	4 646,0	590	253

2. Déterminer les coefficients et le coefficient de détermination dans la régression linéaire du nombre de sorties au cirque et au théâtre sur l'âge.



Suite aux résultats qui précèdent, on décide de construire le tableau suivant pour mieux appréhender le lien entre l'âge des enquêté-es et la fréquentation des cirques et des théâtres.

Effectifs	Nombre de séances théâtre ou cirque			
Âge 2019	aucune	1-4 fois	5 et +	Total
20 ou -	17	33	4	54
]20;30]	63	69	24	156
>30	14	16	13	43
<b>Total</b>	94	118	41	253
<b>Fréquence manquante = 2</b>				

3. Faire le tableau des % en ligne. Ajouter une colonne à droite des totaux avec la répartition par âge.
4. Tracer le diagramme des écarts à l'indépendance pour les personnes qui se sont rendues 5 fois ou plus au théâtre ou au cirque. Les autres diagrammes des écarts à l'indépendance sont reproduits en annexe à la fin du sujet.
5. Calculer les effectifs à l'indépendance (théoriques).
6. Calculer directement les écarts à l'indépendance représentés graphiquement à la question 4.
7. Construire le tableau des effectifs représentant la situation où l'écart à l'indépendance est maximal en respectant le plus possible les écarts apparus dans le tableau des effectifs observés puis calculer le pourcentage de l'écart maximal (PEM) de ce tableau.
8. Calculer les écarts pondérés à l'indépendance (contributions au  $\text{Khi}^2$ ) puis calculer le  $\text{Khi}^2$  total du tableau et le degré de liberté. Si l'on admettait avoir ici un échantillon représentatif, pourrait-on en déduire que le lien entre les deux variables est significatif ? Quelles sont les cases du tableau qui contribuent le plus au  $\text{Khi}^2$  ?
9. En utilisant le V de Cramer et/ou le PEM, le lien entre les deux variables apparaît-il fort ?

**10.** Commenter l'ensemble des résultats. Ceux-ci étant issus d'une enquête encore inachevée dont l'échantillon n'est pas tiré aléatoirement à partir d'une base de sondage dans la population cible (jeunes résidant en Ile de France, mais aussi dans d'autres régions), mais d'un appel à participation à partir de plusieurs réseaux sociaux au sens large, vous pourrez vous demander ce que ces résultats vous apprennent non seulement sur la population touchée par cet appel mais aussi sur le profil des personnes les plus susceptibles de répondre.

**Annexe :** diagrammes des écarts à l'indépendance non demandés dans le sujet

