

CPES 3 : Partiel Statistique 2020-2021

Documents et calculatrice interdits.

Dans tout le sujet, nous noterons

$$\phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt,$$

la fonction de répartition de la loi normale $\mathcal{N}(0, 1)$ et nous rappelons la définition de la densité de la loi normale $\mathcal{N}(\mu, \sigma^2)$,

$$f_{\mu, \sigma^2}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t-\mu)^2/(2\sigma^2)}.$$

Exercice 1 : questions de cours (/5)

On répondra aux questions suivantes en une phrase (sans justification).

1. Soient X_1, \dots, X_n de loi de Bernoulli de paramètre θ . Donner un estimateur de θ .
2. Soit X une variable aléatoire. Définir la médiane de la distribution de X .
3. Soit $\alpha \in]0, 1[$ et X_1, \dots, X_n des observations. Comment calcule-t'on le quantile empirique d'ordre α ?
4. Soit X_1, \dots, X_n des observations telles que $X_i \geq 0$ pour tout $i = 1, \dots, n$. Définir une médiale de X_1, \dots, X_n .
5. On compte le nombre de naissances de filles et de garçons sur N années. Si l'on part du principe qu'à chaque naissance, les probabilités que le bébé soit de sexe masculin ou de sexe féminin sont égales, quelle est la probabilité d'observer plus de naissances masculines que de naissances féminines sur toutes les années ?

Exercice 2 : niveau de vie et seuil de pauvreté

Cet exercice est basé sur l'article intitulé *70 % des personnes pauvres en 2016 le restent l'année suivante, une persistance en hausse depuis 2008* disponible sur le site de l'INSEE www.insee.fr/fr/statistiques/4802040. Les parties sont indépendantes.

Partie I : notion de seuil de pauvreté monétaire L'INSEE définit le seuil de pauvreté monétaire à 60% du niveau de vie médian. Dans cette partie (et uniquement dans cette partie), nous modélisons le niveau de vie d'une personne donnée dans la population suivant la loi normale de moyenne μ et de variance σ^2 (tous les deux inconnus).

1. Montrer que le niveau de vie médian est égal à μ , c'est-à-dire que la médiane de X correspond à sa moyenne. Nous noterons par la suite le seuil de pauvreté : $s_{0.6} = 0.6\mu$.
2. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$. Exprimer la probabilité d'être en-dessous du seuil de pauvreté

$$p_{0.6} = \mathbb{P}(X \leq s_{0.6})$$

en fonction de ϕ , μ et σ .

Partie II : estimation du taux de pauvreté Le taux de pauvreté de la population est la part θ de personnes dont le niveau de vie est en-dessous du seuil de pauvreté. Nous supposons que nous connaissons le niveau de vie de n personnes dans la population et que nous constatons que, parmi ces n personnes, p_n vivent sous le seuil de pauvreté.

1. Donner une estimation $\hat{\theta}_n$ du seuil de pauvreté en fonction de p_n et n .
2. Nous faisons l'hypothèse que les niveaux de vie des n personnes interrogées peuvent être modélisés comme des variables aléatoires indépendantes et de même loi. Montrer qu'il existe s^2 tel que

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, s^2).$$

Écrire s^2 en fonction de θ .

3. Soit $\alpha \in]0, 1[$, donner (en justifiant bien les différentes étapes) un intervalle de confiance asymptotique pour θ au niveau α en fonction d'un quantile (dont le niveau sera à préciser) de la loi normale $\mathcal{N}(0, 1)$.

4. Application numérique : en 2017, une étude de l'INSEE effectuée sur un panel de $n = 12000$ personnes a relevé un taux de pauvreté $\hat{\theta}_n = 13\%$.

- (a) Calculer l'intervalle de confiance au niveau $\alpha = 5\%$. On arrondira pour les calculs $\phi^{-1}(1 - \alpha/2) \approx 2$ et on pourra utiliser le résultat suivant :

$$\frac{\sqrt{0.13 \times (1 - 0.13)}}{\sqrt{12000}} = 0.003.$$

- (b) Combien faudrait-il interroger de personnes supplémentaires pour diviser approximativement par 10 la longueur de l'intervalle de confiance ?

Masse des niveaux de vie détenue par...	1996	2018
les 10 % les plus riches	23	25
les 20 % les plus riches	37	39
les 30 % les plus riches	49	50
les 40 % les plus riches	60	61
les 50 % les plus riches	69	70
les 60 % les plus riches	77	78
les 70 % les plus riches	85	85
les 80 % les plus riches	91	91
les 90 % les plus riches	96	97

Table 1: Part (en %) de la masse totale de niveau de vie détenue selon la tranche de niveau de vie

Partie III : évolution du niveau de vie entre 1996 et 2018

1. A partir du tableau 1, montrez que la courbe de Lorenz du niveau de vie correspondant aux années 1996 et 2018 passe par les points suivants :

$\widehat{F}(x)$	\widehat{FQ}_{1996}	\widehat{FQ}_{2018}
90	77	75
80	63	61
70	51	50
60	40	49
50	31	30
40	23	22
30	15	15
20	9	9
10	4	3

2. Construisez sur le même graphique p. 5, les courbes de Lorenz du niveau de vie correspondant aux années 1996 et 2018.
3. Que pensez-vous de l'évolution de la répartition de la masse du niveau de vie entre 1996 et 2018 ?

Tracé des courbes de concentration

