

Uniquement les calculatrices non scientifiques sont autorisées et une feuille recto-verso de notes

Durée : 2 heures

Attention toute réponse devra être précisément justifiée. Une formule ne sera pas suffisante pour avoir les points.

[1] Une étudiante en sociologie veut analyser, dans le cadre d'un projet de TER, s'il existe une relation linéaire entre la densité de population dans les régions métropolitaines et le taux de criminalité correspondant dans ces régions. Le taux de criminalité (Y) est indiqué en nombre de crimes par 10 000 habitants et la densité de population (X) est mesurée en milliers d'habitants par km^2 :

Région	1	2	3	4	5	6	7	8	9	10	11	12
x_i	7,7	5,8	11,5	2,1	3,7	3,6	7,5	4,2	3,8	10,3	8,6	7,2
y_i	12	9	15	4	4	2	10	3	5	11	10	11

On obtient les valeurs suivantes :

$$\sum_{i=1}^{12} x_i = 76 \text{ et } \sum_{i=1}^{12} x_i^2 = 576,46, \sum_{i=1}^{12} y_i = 96 \text{ et } \sum_{i=1}^{12} y_i^2 = 962 \text{ et } \sum_{i=1}^{12} y_i x_i = 732,6.$$

- Identifier la variable à expliquer et la variable explicative. Pour chaque variable calculer la moyenne observée et la variance observée. Donner également la valeur de la covariance et du coefficient de corrélation. Commenter.
- Faire un graphique de Y en fonction de X . Que peut-on en conclure ? Une modélisation linéaire est-elle pertinente ?
- Ecrire l'équation de la droite de régression obtenue en utilisant les valeurs estimées des coefficients a et b .
- Estimer le taux de criminalité le plus plausible pour une densité de population de 7500 habitants par km^2 .

[2] On veut étudier la proportion p de gens qui vont au théâtre chaque mois. On prend donc un échantillon de taille $n = 100$. Soit N_n le nombre de personnes dans l'échantillon qui vont au théâtre mensuellement.

- Quelle est la loi de N_n ? Par quelle loi peut-on l'approcher et pourquoi ? En déduire une approximation de la loi de $F = N_n/n$.
- On observe une proportion f_n de gens qui vont chaque mois au théâtre. Donner la forme d'un intervalle de confiance pour p , de niveau de confiance $1 - \alpha$.
- Applications numériques pour calculer l'intervalle de confiance asymptotique pour p , de niveau de confiance $1 - \alpha$ $f_n = 0,2$ et $1 - \alpha = 92\%, 95\%, 97\%$.

[3] Soit le modèle d'échantillonnage suivant : $X_1, \dots, X_n$ i.i.d. de densité $\theta^2 e^{-\theta^2 x} \mathbb{1}_{\{x \geq 0\}}$ où $\theta > 0$

Le paramètre d'intérêt est le paramètre de ce modèle.

- Écrire le modèle sous la forme d'un triplet, espace d'observation, tribu, famille de probabilités.
- Que vaut $\mathbb{E}(X_1)$? Proposer un estimateur de θ grâce à la méthode des moments.
- Le modèle est-il identifiable ?
- Est-ce que l'estimateur du maximum de vraisemblance de θ existe ? Si oui le donner.
- On suppose que le modèle est régulier, que vaut l'information de Fisher ?
- Soit $\hat{\theta} = \frac{1}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i}}$ un estimateur de θ . Quelle est la vitesse de $\hat{\theta}$?
- En déduire un intervalle de confiance asymptotique à 90 pourcent de θ . Avec les valeurs numériques : $n = 100$, $\frac{1}{n} \sum_{i=1}^n x_i = 0.9$.
- Question Bonus : Est-ce que c'est un estimateur asymptotiquement efficace de θ ?

$$\Phi(t) = P(X \leq t) \text{ pour } X \sim \mathcal{N}(0, 1)$$

t	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,5279	0,53188	0,53586
0,1	0,53983	0,5438	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,6293	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,6591	0,66276	0,6664	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,7054	0,70884	0,71226	0,71566	0,71904	0,7224
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,7549
0,7	0,75804	0,76115	0,76424	0,7673	0,77035	0,77337	0,77637	0,77935	0,7823	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,8665	0,86864	0,87076	0,87286	0,87493	0,87698	0,879	0,881	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,9032	0,9049	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,9222	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,9452	0,9463	0,94738	0,94845	0,9495	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,9608	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,9732	0,97381	0,97441	0,975	0,97558	0,97615	0,9767
2	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,9803	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,983	0,98341	0,98382	0,98422	0,98461	0,985	0,98537	0,98574
2,2	0,9861	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,9884	0,9887	0,98899
2,3	0,98928	0,98956	0,98983	0,9901	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,9918	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,9943	0,99446	0,99461	0,99477	0,99492	0,99506	0,9952
2,6	0,99534	0,99547	0,9956	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,9972	0,99728	0,99736
2,8	0,99744	0,99752	0,9976	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861

Table pour les grandes valeurs

3	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	4
0,99865	0,99903	0,99931	0,99952	0,99966	0,99977	0,99984	0,99989	0,99993	0,99995	0,99997