Statistiques appliquées – L3

Vendredi 22 décembre 2006

Interrogation de contrôle continu n°2

Durée : 1h00

Aucun document n'est autorisé

Questions de cours

1. La densité d'une loi exponentielle d'espérance θ est

   a. $f(x) = \begin{cases} = \frac{1}{\theta}e^{-\frac{x}{\theta}} & \text{si } x \succeq 0 \\ = \quad 0 & \text{sinon} \end{cases}$ b. $f(x) = \begin{cases} = \frac{\theta^x}{x!}e^{-x\theta} & \text{si } x \succeq 0 \\ = \quad 0 & \text{sinon} \end{cases}$

   c. $f(x) = \begin{cases} = \theta e^{-x\theta} & \text{si } x \succeq 0 \\ = \quad 0 & \text{sinon} \end{cases}$

2. On considère un modèle de régression multiple. $Y = Xa + \varepsilon$.

   Donnez l'expression de $\hat{a}$ en fonction de X et Y.

3. Quelles sont les propriétés des estimateurs des MCO?

Exercice 1

Soient X_1, ... X_n, un n-échantillon d'une loi exponentielle de paramètre a. Ecrire la vraisemblance de a. Montrer que l'estimateur du maximum de vraisemblance de a est la moyenne empirique.

Exercice 2

Trouve-t-on son premier emploi dans l'année qui suit la fin du master? Quel est, en moyenne, le salaire d'embauche?

Pour répondre à ces deux questions, on fait un sondage parmi 27 individus, choisis de manière aléatoire, qui ont cherché du travail après un master, et sont aujourd'hui employés. On veut savoir s'ils ont trouvé leur emploi dans l'année qui suit la fin du master ou non. 80% ont trouvé leur emploi dans l'année qui suit la fin du master.

On demande également aux individus interrogés leur premier salaire. Le salaire annuel moyen à l'embauche est de 22506 €. La variance dans la population n'est pas connue. La variance sans biais estimée dans l'échantillon est 9000^2.

1. Si on voulait construire un intervalle de confiance pour la proportion des étudiants qui ont trouvé un emploi, quel loi utiliserait-on? Pourrait-on facilement obtenir les valeurs des bornes?

2. Donnez un intervalle de confiance à 95% pour la moyenne du salaire.

Exercice 3

Faut-il faire payer aux plus pauvres les produits domestiques dont l'utilisation peut sauver la vie? Cette question est l'objet de controverses. Pour certains, il est injuste de faire payer les produits d'hygiène les plus indispensables, cela exclut les plus pauvres. D'autres, au contraire, estiment que les produits distribués gratuitement sont peu ou pas utilisés. Par exemple, l'eau peut souvent transmettre des maladies. Faut-il distribuer gratuitement la solution qui permet de purifier l'eau, ou la proposer à un prix très faible?

En Zambie, il existe un produit, Clorin, dont la bouteille, vendue entre 800 et 1000 kwacha (800 kw = 0.15€) dans les commerces, permet de désinfecter l'eau utilisée par une famille pendant un mois.

Une expérience est menée, sous la forme d'une fausse campagne de marketing pour le produit. Une équipe de chercheurs va voir 1000 ménages, en leur donnant une enveloppe, qui contient le prix auquel ils peuvent acheter une bouteille de désinfectant. Il y a 10 prix possibles; l'enveloppe est tirée au sort. Chaque type de prix est en 100 exemplaires, ainsi chaque prix a le même poids. Pour chaque prix donné, le ménage dit s'il veut ou non acheter une bouteille. En additionnant les réponses pour chaque prix possible, on constate combien de bouteilles sont vendues, et on reporte les résultats dans le tableau suivant.

Groupe	Prix p_i	Bouteilles vendues q_i	$\mathbf{ln}\,(p_i)$	$\mathbf{ln}\,(q_i)$	$[\mathbf{ln}\,(p_i)]^{\mathbf{2}}$	$[\mathbf{ln}\,(q_i)]^{\mathbf{2}}$	$[\mathbf{ln}\,(p_i)][\mathbf{ln}\,(q_i)]$
1	260	82	5.56	4.41	30.9	19, 4	24.5
2	300	81	5.70	4.39	32.5	19.3	25.1
3	400	78	5.99	4.36	35.9	19	26.1
4	450	69	6.11	4.23	37.3	17.9	25.9
5	500	63	6.21	4.14	38.6	17.2	25.7
6	600	48	6.4	3.87	40.9	15	24.8
7	700	51	6.55	3.93	42.9	15.5	25.8
8	800	48	6.68	3.87	44.7	15	25.9
9	950	35	6.86	3.56	47	12.6	24.4
10	1100	23	7	3.14	49	9.8	22
Somme*	**6010**	**577**	**63.06**	**40**	**400**	**161.7**	**250.6**

* pour i variant de 1 à 10

1. On veut modéliser le problème par un modèle linéaire simple, du type $y = ax + b + \varepsilon$

   a. Donnez l'expression de l'estimateur des MCO de a en fonction de x et y.

   b. Ici, quel modèle va-t-on considérer? (que seront y et x?)

   c. Calculez l'estimateur des MCO de a.

   On donne :

$6.306\ ^2 = 39,8$; $4 * 6,306 = 25,2$; $16,17 * 25,06 = 405.2$; $25,06^2 = 628$

Comment interprétez vous le résultat obtenu?

2. L'équipe de chercheurs s'intéresse aussi à l'impact du prix payé par le ménage sur l'utilisation effective de la solution purifiante. Il y a une deuxième étape dans l'expérience: une fois que le ménage a dit si oui ou non il achetait le Clorin pour un prix donné, on sort une autre enveloppe, qui contient un bon de réduction supplémentaire. Le prix payé est donc différent; certains ménages ne payent rien du tout. L'équipe revient deux semaines après dans les familles, et teste la présence de Clorin dans l'eau que la famille utilise (il est très simple de savoir si l'eau est purifiée ou non).

   On peut alors connaître l'impact du prix payé sur la probabilité d'utiliser du Clorin. On mesure deux effets différents : savoir si payer quelque chose plutôt que de recevoir une bouteille gratuite incite à utiliser le produit; savoir si payer un peu ou beaucoup change l'utilisation du produit.

   Le fait de payer lorsque on passe d'un prix nul (gratuité) à un prix positif accroît la probabilité d'utiliser du Clorin d'un coefficient de 0.085. L'écart type du coefficient est 0.04.

   Le fait de payer 100 kw de plus lorsque le prix est déjà positif accroît la probabilité d'utiliser du Clorin de 0.01. L'écart type du coefficient est 0.012.

   a. Qu'est-ce que le t de Student? Comment l'interprète-t-on?

   b. Les coefficients sont-ils significatifs, c'est-à-dire a-t-on 95% de chance qu'ils aient bien le signe trouvé?

3. Que concluez-vous des expériences menées? Quel est l'impact du prix sur l'achat et l'utilisation de Clorin? Quel conseil donneriez-vous à des ONG qui souhaitent améliorer la qualité de l'eau consommée par les familles les plus pauvres de Zambie?