Examen d'accès - 26 Septembre 2018

Aucun document autorisé - Calculatrice fournie par le centre d'examen

Les consignes indiquées ci-dessous sont suffisamment explicites pour ne pas laisser de doute quant à leur interprétation. Les personnes surveillant l'examen ne répondront à aucune question relative à ces consignes durant l'épreuve, la bonne compréhension de ces règles faisant elle aussi partie de l'examen.

Cet examen est un questionnaire à choix multiples constitué de 50 questions. Plusieurs réponses sont proposées pour chaque question (ou ensemble de questions). Le nombre de bonnes réponses à une question peut aller de 0 à 5.

Les réponses sont à inscrire sur la feuille jointe, en cochant pour chaque question la (ou les) case(s) correspondant à la (ou les) bonne(s) réponse(s).

Toute réponse ambiguë sera considérée comme une absence de réponse.

Comptabilisation des points :

— Toute case cochée à tort entraîne une pénalité de 0,5 point.

— Toute case cochée à raison entraîne une bonification de 1 point (même si d'autres cases dans la même question auraient dû être cochées et ne l'ont pas été).

— Une case non cochée ne donne ni bonification ni malus.

PARTIE QUESTIONS GENERALES

Q 1 Soit la fonction $f(x) = (x^2 - 1)\ln\left(\dfrac{1+x}{1-x}\right)$. Le domaine de définition de f est :

A) $]-\infty, 1[$

B) $]1, +\infty[$

C) $[0, 1[$

D) $]-1, 0]$

E) $]-1, 1[$

Q 2 (Suite Q1). La fonction f est :

A) croissante puis décroissante puis croissante

B) décroissante puis croissante

C) croissante puis décroissante

Pour tout x dans le domaine de définition, on a

D) $f(x) \geq 0$

E) $f(x) \leq 0$

Q 3 (Suite Q2). La limite de la fonction f à la borne supérieure de son domaine de définition vaut :

A) $+\infty$

B) 0

C) $-\infty$

Et cette fonction est

D) impaire

E) paire

Q 4 (Suite Q3). Cette fonction f est :

A) convexe

B) concave

C) convexe puis concave

D) concave puis convexe

E) convexe puis concave puis convexe

Q 5 On se propose d'agréger B ($B > 1$) estimateurs statistiques appartenant à la même classe de modèles, indépendants les uns des autres, au cours d'une procédure statistique d'estimation. On note chaque estimateur $\hat{\pi}_b$, où $1 \leq b \leq B$.
Grâce à cette stratégie pour construire $\hat{\pi} = \frac{1}{B} \sum_{b=1}^{B} \hat{\pi}_b$,

A) le biais de l'estimateur agrégé $\hat{\pi}$ est plus faible que celui des estimateurs individuels $\hat{\pi}_b$

B) le facteur de diminution du biais vaut $1/B$

C) la variance de l'estimateur agrégé $\hat{\pi}$ est plus faible que celle des estimateurs individuels $\hat{\pi}_b$

D) le facteur de diminution de la variance vaut $1/B^2$

E) le biais et la variance de l'estimateur agrégé $\hat{\pi}$ sont moindres par rapport à ceux des estimateurs individuels $\hat{\pi}_b$

Q 6 Un fumeur, après avoir lu une série de statistiques effrayantes sur les risques de cancer, problèmes cardio-vasculaires liés au tabac, décide d'arrêter de fumer. Toujours d'après des statistiques, on estime les probabilités suivantes : si cette personne n'a pas fumé un jour J_n, alors la probabilité pour qu'elle ne fume pas le jour suivant J_{n+1} est 0.3 ; mais si elle a fumé un jour J_n, alors la probabilité pour qu'elle ne fume pas le jour suivant J_{n+1} est 0.9.
La probabilité P_{n+1} pour qu'elle fume le jour J_{n+1} en fonction de la probabilité P_n pour qu'elle fume le jour J_n vaut :

A) $P_{n+1} = -0.6P_n + 0.7$

B) $P_{n+1} = 0.3P_n + 0.1$

C) $P_{n+1} = 0.7P_n + 0.1$

Et la limte de P_n est :

D) 0

E) 0.4375

Q 7 On considère la matrice $A = \begin{pmatrix} 3 & 9 & -9 \\ 2 & 0 & 0 \\ 3 & 3 & -3 \end{pmatrix}$.

A) A est nilpotente d'indice 3

B) le polynome caractéristique P de l'endomorphisme A est $P(X) = -X^3$

C) $A^2 = \begin{pmatrix} 0 & 0 & 0 \\ 6 & 18 & -18 \\ 6 & 18 & -18 \end{pmatrix}$

D) A n'est pas inversible

E) le rang de A vaut 2

Q 8 Une matrice de corrélation est une matrice :

A) composée de coefficients dans l'intervalle $[-1, 1]$

B) dont tous les éléments diagonaux valent 0

C) symétrique

D) carrée

E) inversible

Q 9 On considère une loi mélange, plus précisément un mélange Poisson-Gamma. La sinistralité S d'un portefeuille d'assurance suit cette loi, i.e. $S = \sum_{i=1}^{N} X_i$. Dans cette situation, le nombre de sinistres N suit une loi de Poisson de paramètre λ ($N \sim \mathcal{P}(\lambda)$), et les pertes individuelles X_i sont indépendantes et identiquement distribuées de loi Gamma, i.e $X_i \sim \mathcal{G}(\alpha, \delta)$ avec α le paramètre de forme et δ l'inverse du paramètre d'échelle.

A) L'espérance de la loi S vaut $\lambda\, \alpha/\delta$

B) La variance de cette loi excède son espérance

C) C'est une loi sous-dispersée

D) Ce mélange est équivalent à une loi binomiale négative

E) Le nombre de sinistres est aléatoire

Q 10 Soit un portefeuille de sinistres bris de glace dont les montants X_k de sinistres peuvent être modélisés via une loi exponentielle de paramètre λ, i.e. $X_k \sim \mathcal{E}xp(\lambda)$.

A) Le montant moyen des sinistres du portefeuille vaut λ

B) Le montant moyen des sinistres du portefeuille vaut $1/\lambda$

C) La densité de probabilité des montants de sinistres vaut $g(x) = \lambda \exp(-\lambda x)$

D) La variance des montants de sinistre est supérieure à leur moyenne

E) Cette loi admet une queue de distribution épaisse

PARTIE PROBABILITES

Q 11 On considère trois événements aléatoires A, B et C, chacun de probabilité égale à 1/3, tels que :

$$\begin{cases} A \cap B = \emptyset \\ \\ P(C/A) = \dfrac{1}{3} \quad \text{(probabilité conditionnelle de } C \text{ sachant } A\text{)} \\ \\ P(C/B) = \dfrac{1}{3} \quad \text{(probabilité conditionnelle de } C \text{ sachant } B\text{)} \end{cases} .$$

A) $P(\bar{A} \cap \bar{B}) = 1/3$.

B) $P(\bar{A} \cup \bar{B}) = 2/3$.

C) $P(A \cup B \cup C) = 7/9$.

D) $P(A \cup B \cup C) = 8/9$.

E) $P(A \cup B \cup C) = 1$.

Q 12 Soit f une densité de probabilité telle que :

$$f(x) = \begin{cases} c\,x & \text{si } 0 \leq x \leq 1 \\ c\,(2-x) & \text{si } 1 < x \leq 2 \\ 0 & \text{sinon} \end{cases} \cdot$$

On note X une variable aléatoire de densité f.

A) $c = 1$.

B) $P[\,|X-1| \leq 1] = 1$.

C) $P[\,|X-1| \geq 1] = 1$.

D) $E(X) = 1$ (espérance de X) .

E) $V(X) = 1$ (variance de X) .

Q 13 On dispose de 2 urnes, contenant chacune 1 boule noire et 1 boule blanche. On tire d'abord, au hasard, une boule dans la première urne et on la place dans la deuxième urne.
On effectue ensuite trois tirages sans remise dans cette deuxième urne, jusqu'à épuisement des boules qu'elle contient.

A) La probabilité pour qu'après un tirage dans la deuxième urne, il y reste deux boules blanches est $1/3$.

B) La probabilité pour qu'après un tirage dans la deuxième urne, il y reste deux boules noires est $1/3$.

C) La probabilité pour qu'après un tirage dans la deuxième urne, il y reste une boule blanche et une boule noire est $1/3$.

D) Lorsque la dernière boule tirée (de la deuxième urne) est blanche, la probabilité que la première boule tirée (de la première urne) soit blanche est égale à $2/3$.

E) Lorsque la dernière boule tirée (de la deuxième urne) est blanche, la probabilité que la première boule tirée (de la première urne) soit noire est égale à $2/3$.

Q 14 Soit X une variable aléatoire dont la fonction de répartition F est donnée par :

$$F(x) = \begin{cases} 0 & \text{si } x < -1 \\ 1/3 & \text{si } -1 \leq x < 0 \\ 2/3 & \text{si } 0 < x \leq 1 \\ 1 & \text{si } x \geq 1 \end{cases} .$$

A) X est une variable aléatoire à densité.

B) X est une variable aléatoire discrète.

C) La variable aléatoire $|X|$ suit la loi de Bernoulli de paramètre $1/3$.

D) La variable aléatoire $|X|$ suit la loi de Bernoulli de paramètre $1/2$.

E) La variable aléatoire $|X|$ suit la loi de Bernoulli de paramètre $2/3$.

Q 15 On considère deux variables aléatoires indépendantes M et N, suivant chacune la loi de Poisson de paramètre $\lambda > 0$.

On pose $X = 2M$, $Y = M + N$ et on note g_M, g_N, g_X, g_Y les fonctions génératrices respectives de M, N, X, Y.

A) $g_M(s) = \mathrm{e}^{\lambda(s-1)}$.

B) $g_X(s) = \mathrm{e}^{\lambda(s^2-1)}$.

C) $g_X(s) = \mathrm{e}^{\lambda(2s-1)}$.

D) $g_Y(s) = \mathrm{e}^{2\lambda(s-1)}$.

E) $g_Y(s) = \mathrm{e}^{\lambda(2s-1)}$.

Q 16 Soit g la fonction définie sur $\mathbb{R}$ par : $g(x) = \mathrm{e}^{-(x-1)^2}$.

A) $\displaystyle\int_{-\infty}^{+\infty} g(x)\,\mathrm{d}x = \sqrt{2\pi}$.

B) $\displaystyle\int_{-\infty}^{+\infty} g(x)\,\mathrm{d}x = \sqrt{\pi}$.

C) $\displaystyle\int_{-\infty}^{+\infty} g(x)\,\mathrm{d}x = 2\sqrt{\pi}$.

D) $\displaystyle\int_{-\infty}^{+\infty} (x-1)\,g(x)\,\mathrm{d}x = 0$.

E) $\displaystyle\int_{-\infty}^{+\infty} (x-1)^2\,g(x)\,\mathrm{d}x = 1$.

Q 17 On jette un dé équilibré, de manière répétée, pour déterminer une suite aléatoire de chiffres compris entre 1 et 6.
On note X_1 le nombre de jets nécessaires pour obtenir pour la première fois un chiffre impair, et X_2 le nombre total de jets nécessaires pour obtenir deux chiffres impairs.
On note S le nombre de jets nécessaires pour obtenir pour la première fois deux chiffres impairs de suite.

A) La variable aléatoire X_1 suit une loi géométrique.

B) La variable aléatoire X_2 suit une loi géométrique.

C) Les variables aléatoires X_1 et X_2 sont indépendantes.

D) $X_1 \leq S \leq X_2$.

E) Les variables aléatoires X_1 et S sont indépendantes.

Q 18 (Suite de la question Q 17) On note Y le nombre de jets nécessaires pour obtenir pour la première fois un chiffre supérieur ou égal à 4.
On note Z le nombre de jets nécessaires pour obtenir pour la première fois le chiffre 5.

A) $Z = \text{Min}\{X_1, Y\}$.

B) $Z = \text{Max}\{X_1, Y\}$.

C) Pour tout $n \in \mathbb{N}^*$, on a : $\quad P[Y = n] = (\frac{1}{2})^n$.

D) Pour tout $n \in \mathbb{N}^*$, on a : $\quad P[Z \geq n] = (\frac{5}{6})^n$.

E) Pour tout $n \in \mathbb{N}^*$, on a : $\quad P[Z > n] = (\frac{5}{6})^n$.

Q 19 Soit X une variable aléatoire qui suit la loi exponentielle de paramètre 1.
On note Y la variable aléatoire définie par : $\quad Y = \begin{cases} 0 & \text{si } X \leq 2 \\ X - 2 & \text{si } X > 2 \end{cases}$.

A) $E(X) = V(X) = 1$.

B) $E(\mathrm{e}^X) = 1$.

C) $P[Y = 0] = \mathrm{e}^{-2}$.

D) $E(Y) = \mathrm{e}^{-2}$.

E) $E_{[X>2]}(Y) = \mathrm{e}^{-2}$ (espérance conditionnelle de Y sachant que $X > 2$).

Q 20 On considère trois variables aléatoires centrées X, Y et Z, admettant chacune une variance, dont la matrice de variance-covariance est :

$$M = \begin{pmatrix} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 1 \end{pmatrix} \quad .$$

On note $U = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ et V la matrice transposée de U.

A) M est inversible.

B) $M = E(UV)$.

C) $M = E(VU)$.

D) $V(X + Y) = V(X) + V(Y)$.

E) $V(X - 2Z) = 0$.

PARTIE MATHEMATIQUES FINANCIERES

Pour les 10 questions suivantes, on adopte la convention de taux annuel, les intérêts étant payés en fin de période.

Q 21 Un particulier fait 3 placements annuels de 10 000 euros au taux de 1,25%. Quelle est la valeur de son placement à la fin de la 3ème année ?

A) 30 556,29 euros

B) 30 756,27 euros

C) 31 052,57 euros

D) 31 261,28 euros

Q 22 Quelle est l'annuité d'un crédit de 15 000 euros d'une durée de 3 ans au taux de 1,3% ?

A) 5 130,56 euros

B) 5 160,85 euros

C) 5 201,32 euros

Le coût du crédit est

D) 482,55 euros

E) 391,68 euros

Q 23 Soit un crédit de 30 000 euros d'une durée de 3 ans au taux de 1,6%. La première annuité est de 11 000 euros. La seconde annuité est de 10 000 euros. Quel est le montant de la troisième annuité ?

A) 9 998,98 euros

B) 10 042,27 euros

C) 9 948,35 euros

Le montant total des intérêts est de

D) 1 042,27 euros

E) 948,35 euros

Q 24 Quel est le montant d'un crédit de 3 ans dont 1 an de différé de remboursement, sachant que le taux est de 0,75% et que l'annuité est de 7 641,36 euros ?

A) 13 000 euros

B) 14 500 euros

C) 15 000 euros

Le coût du crédit est

D) 282,72 euros

E) 782,72 euros

Q 25 Un prêt de 10 000 euros sur 3 ans a des annuités constantes de 3 433,83 euros. Le taux du prêt est de

A) 1%

B) 1,5%

C) 2%

Le capital restant dû après la 2ème annuité est de

D) 3 283,83 euros

E) 3 383,08 euros

Q 26 Quel est le taux de rendement à maturité d'une obligation zéro coupon 2 ans de nominal 100 et qui cote 100,5 ?

A) -0,25%

B) -0,5%

C) 0,25%

Sa duration est

D) 2

E) 1,75

Q 27 Quelle est la valeur actuelle d'une obligation de maturité 2 ans de nominal 100 payant un coupon de 2,5 avec un taux de marché du 2 ans de 0,5% ?

A) 103,97

B) 106,05

C) 99,63

Sa duration est

D) 1,98

E) 2

Q 28 Quel est le prix à terme d'échéance 2 ans d'une action qui cote 50 aujourd'hui, sachant que le taux de marché du 2 ans est de 1,5% ?

A) 48,53

B) 49,26

C) 50,75

D) 51,51

Q 29 On considère un marché financier sans arbitrage et deux contrats définis respectivement par les paiements X_T et Y_T pour une même maturité T, et de prix X_0 et Y_0. Si $X_T < Y_T$ dans tous les états du monde, alors

A) $X_0 > Y_0$

B) $X_0 \geq Y_0$

C) $X_0 = Y_0$

D) $X_0 < Y_0$

Q 30 On considère un modèle binomial à 2 périodes, de taux sans risque $r = 0.2$, caractérisé par le coefficient de hausse $u = 1.5$, et de baisse $d = 0.8$, ($S_1 = uS_0$ en cas de hausse, et $S_1 = dS_0$ en cas de baisse). La probabilité risque-neutre est caractérisée par une probabilité de hausse de

A) 0.43

B) 0.57

C) 1.75

Une option d'achat (ou call) donne le droit (et non l'obligation) d'acheter en $T = 2$ l'action au prix $K = 3$. Le prix coté de cette action en $t = 0$ est $S_0 = 4$. La valeur de cette option d'achat en $t = 0$ est

D) 1.97

E) 2.83

PARTIE STATISTIQUE ET ANALYSE DE DONNEES

Q 31 Un bon estimateur minimise le risque quadratique qui se décompose en :

A) carré du biais et variance de l'estimateur

B) espérance de l'estimateur et variance de l'estimateur

C) précision et exactitude

La qualité de l'estimation augmente généralement :

D) avec la taille de l'échantillon

E) lorsque les variables sont gaussiennes

Q 32 Estimer un intervalle de confiance nécessite de :

A) connaître l'espérance de l'estimateur

B) connaître la variance de l'estimateur

C) connaître la loi de l'estimateur

On peut calculer des intervalles de confiance asymptotiques :

D) en utilisant le TCL (théorème central limite)

E) par la méthode du maximum de vraisemblance

Q 33 Lors d'un vote, on demande à 1800 personnes lors de leur sortie du bureau de vote s'ils ont voté pour le candidat A ou le candidat B. 1200 se prononcent pour le candidat A, et 600 pour le candidat B. On considère que la probabilité de voter pour le candidat A suit une loi de Bernoulli de paramètre p, et on cherche un intervalle de confiance à 95% pour p. On pourra procéder à une approximation normale, en supposant que cette loi est une loi normale d'espérance p et de variance $p(1-p)$, et on rappelle que le quantile 0,95 d'une loi normale standard est 1,64, et le quantile 0,975 est 1,96.
L'intervalle de confiance pour p est

A) 0,49 - 0,84

B) 0,61 - 0,72

C) 0,64 - 0,69

Utiliser un échantillon de 900 personnes aurait pour effet de :

D) multiplier par deux environ la taille de l'intervalle

E) multiplier par 1,414 environ la taille d'intervalle

Q 34 Dans un problème de test d'une hypothèse H_0 contre une hypothèse H_1, on fixe l'erreur de première espèce. A quoi correspond cette valeur ?

A) la probabilité de refuser H_0 alors que H_0 est vraie

B) la probabilité de refuser H_1 alors que H_1 est vraie

C) l'inverse de la puissance du test

La région critique d'un test est composée de l'ensemble des valeurs qui

D) conduisent à refuser H0, que H_0 soit vraie ou non

E) conduisent à refuser H0 lorsque H0 est vraie

Q 35 On veut savoir si les commerciaux d'une agence font signer significativement plus de contrats que ceux des autres agences d'une entreprise. Pour cela, on considère que le nombre moyen de contrats obtenu par commercial dans une agence suit une loi qu'on peut approcher par une loi normale d'espérance m. Dans cette entreprise, le nombre moyen de contrats est de 50. S'il s'avère que cette performance est significativement meilleure que celle du reste de l'entreprise, la direction est prête à verser une prime.

Le statisticien chargé de l'étude va poser comme test :

A) $H_0 : m > 50$ contre $H_1 : m \leq 50$

B) $H_0 : m \leq 50$ contre $H_1 : m > 50$

Pour calculer ce test en pratique, le statisticien :

C) se ramènera à une hypothèse simple pour H_0 et utilisera le fait que la statistique du test suit une loi de Gauss

D) se ramènera à une hypothèse simple pour H_0 et utilisera le fait que la statistique du test suit une loi de Student

E) utilisera le fait que la statistique du test suit une loi de Student

Q 36 L'analyse factorielle de données (ACP, AFC, ACM) consiste à :

A) déformer un nuage de points pour mieux le visualiser

B) faire une rotation du nuage et le projeter sur des droites ou des plans

Les valeurs propres (inerties expliquée par les axes) et vecteurs propres (axes d'inertie) résultent :

C) de la diagonalisation de la matrice d'inertie

D) de la linéarisation de la matrice d'inertie

E) de l'inversion de la matrice d'inertie

Q 37 En ACP, les composantes principales sont de nouvelles variables intéressantes car :

A) elles sont de variance minimale

B) elles sont sans biais

C) elles sont non corrélées deux à deux

D) elles sont les plus corrélées, en moyenne, aux variables de départ

E) elles sont plus simples que les variables de départ

Q 38 En ACP, le cercle des corrélations permet :

A) de visualiser les corrélations entre les variables de départ et les composantes principales

B) de visualiser les corrélations entre les variables de départ et les composantes principales, les corrélations entre les variables de départ, et les corrélations entre les composantes principales

C) de visualiser les corrélations entre les variables de départ et les composantes principales, et les corrélations entre les variables de départ

Ce cercle est obtenu :

D) en effectuant l'analyse factorielle du nuage des variables

E) en effectuant une analyse factorielle des correspondances

Q 39 Une AFC est une analyse factorielle des profils des lignes et des profils des colonnes. Le vecteur caractéristique d'une modalité de X est composé :

A) des fréquences de Y dans la population

B) des fréquences de Y dans la sous-population prenant cette modalité X

C) des fréquences jointes de X et de Y

On utilise la métrique du χ^2 :

D) pour ne pas donner plus d'importance aux modalités les moins fréquentes

E) pour ne pas donner plus d'importance aux modalités les plus fréquentes

Q 40 En classification ascendante hiérarchique (CAH), la stratégie d'agrégation qui permet de regrouper les classes :

A) utilise des points caractéristiques des classes

B) représente une mesure de l'inertie du nuage

C) minimise une mesure de l'écart entre deux classes

D) centre et réduit les variables

E) utilise une AFC

PARTIE ECONOMETRIE

Q 41 On considère la variable aléatoire Y = nombre de sinistres dans l'année d'un assuré. A partir de $n = 452009$ observations, on ajuste un modèle de régression linéaire de Y en fonction des variables "VehAge" (âge du véhicule), "DrivAge" (âge du conducteur), "BonusMalus" (niveau de bonus-malus), "Density" (densité de population dans la zone de l'assuré). On trouve la sortie suivante :

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.322e-02  2.524e-03 -17.125   <2e-16 ***
VehAge      -1.009e-03  6.350e-05 -15.893   <2e-16 ***
DrivAge      7.815e-04  2.879e-05  27.145   <2e-16 ***
BonusMalus   1.129e-03  2.620e-05  43.104   <2e-16 ***
Density      2.109e-07  9.113e-08   2.314   0.0207 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2399 on 452004 degrees of freedom
Multiple R-squared:  0.004741,  Adjusted R-squared:  0.004732
F-statistic: 538.2 on 4 and 452004 DF,  p-value: < 2.2e-16
```

On considère un assuré ayant les caractéristiques suivantes :
— conducteur de 30 ans
— véhicule de 2 ans
— Bonus Malus de 50
— densité de population 15.
Si on en croit le modèle, quel est le nombre moyen d'accidents que cet assuré aura l'année prochaine ?

A) -0,04

B) 0,03

C) 0,04

D) 1,04

E) 1,03

Q 42 (Suite de la question précédente) On suppose qu'un accident, s'il survient, a pour coût moyen 500 euros. La prime pure correspondant à l'assuré précédent est

A) 15 euros si le nombre Y d'accidents est une variable indépendante de leurs coûts $C_1, ..., C_Y$

B) 20 euros si le nombre Y d'accidents est une variable indépendante de leurs coûts $C_1, ..., C_Y$

C) 15 euros même si le nombre Y d'accidents est une variable non indépendante de leurs coûts $C_1, ..., C_Y$

D) 20 euros même si le nombre Y d'accidents est une variable non indépendante de leurs coûts $C_1, ..., C_Y$

E) inférieure à la prime commerciale

Q 43 (Suite de la question précédente) On souhaite tester la significativité des variables présentes dans le modèle.

A) La variable VehAge est significative au niveau 5%

B) La variable Density est significative au niveau 5%

C) La variable Density est non significative au niveau 5%

D) La variable Density est significative au niveau 1%

E) La variable Density est non significative au niveau 1%

Q 44 (Suite de la question précédente) On souhaite effectuer une méthode de régression backward à partir des résultats précédents. Dans ce cadre, la variable que l'on va retirer afin de considérer un modèle de dimension inférieure est

A) Bonus Malus

B) Density

On retire une de ces deux variables, et le R^2 devient 0.004601. Le nouveau modèle est appelé M2, celui de la première question est noté M1.

C) M1 est préférable à M2

D) M2 est préférable à M1

E) je n'ai pas suffisamment d'éléments pour conclure que l'un est préférable à l'autre.

Q 45 (Suite de la question précédente) On considère à nouveau le modèle M1 de la première question. On note β_4 le coefficient associé à la variable Density. On souhaite tester $H_0 : \beta_4 = 0$ contre $H_1 : \beta > 0$.

A) on rejette H_0 au niveau 1%

B) on ne rejette pas H_0 au niveau 1%

C) on rejette H_0 au niveau 2%

D) on rejette H_0 au niveau 0,05%

E) on ne rejette pas H_0 au niveau 0,05%

Q 46 On souhaite plutôt modéliser les mêmes données par un modèle linéaire généralisé. Soit $X \in \mathbb{R}^d$ et $Y \in \mathbb{R}$ on rappelle qu'un modèle linéaire généralisé consiste à supposer que :

— la loi de Y sachant $X = x$ appartient à une famille de loi de probabilité $\{\mathbb{P}_{\theta,\phi} : \theta \in \Theta, \phi \in \mathbb{R}\}$ qui est une famille exponentielle.

— $g(E[Y|X = x]) = \beta_0 + \beta_1 x_1 + ... + \beta_d x_d$ où g est une fonction strictement monotone fixée.

On rappelle qu'une famille de loi est dite exponentielle (de paramétrage canonique) si, dans le cas d'une variable continue (resp. discrète), sa densité $f_\theta(y)$ (resp. $f_\theta(y) = \mathbb{P}_\theta(Y = y)$) est de la forme

$$f_\theta(y) = \exp(\frac{\theta y - a(\theta)}{\phi})c_\phi(y).$$

On modélise les données précédentes par un modèle linéaire généralisé où la famille exponentielle choisie est la famille des lois de Poisson. Dans ce cas, la fonction a est

A) $a(\theta) = \log(\theta)$

B) $a(\theta) = \exp(\theta)$

C) $a(\theta) = 1/\theta$

Si l'on utilise la paramétrisation canonique de la loi de Poisson (i.e. la paramétrisation où θ est tel que $f_\theta(y) = \exp(\frac{\theta y - a(\theta)}{\phi})c_\phi(y)$, à ne pas confondre avec la paramétrisation naturelle), l'espérance de la loi de Poisson en fonction de θ est

D) θ

E) $\exp(\theta)$

Q 47 (Suite de la question précédente) On utilise donc un modèle linéaire généralisé sur les données de la question 41, en supposant que la distribution de $Y|X = x$ est Poisson, et en prenant la fonction de lien $g(y) = \log y$. La sortie du logiciel R est la suivante :

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.495e+00  4.137e-02 -108.65   <2e-16 ***
VehAge      -1.985e-02  1.207e-03  -16.45   <2e-16 ***
DrivAge      1.307e-02  4.850e-04   26.96   <2e-16 ***
BonusMalus   1.771e-02  3.906e-04   45.34   <2e-16 ***
Density      3.129e-06  1.511e-06    2.07   0.0384 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 144913  on 452008  degrees of freedom
Residual deviance: 142804  on 452004  degrees of freedom
AIC: 188871
```

On considère un assuré ayant les mêmes caractéristiques que l'assuré de la question 1. Si on en croit le modèle, quel est le nombre moyen d'accidents que cet assuré aura l'année prochaine ?

A) -3,256

B) 3,256

C) $\exp(3, 256)$

D) $\log(-3, 256)$

E) $\log(3, 256)$

Q 48 (Suite de la question précédente) On souhaite tester la significativité des variables présentes dans le modèle.

A) La variable VehAge est significative au niveau 5%

B) La variable Density est significative au niveau 5%

C) La variable Density est non significative au niveau 5%

D) La variable Density est significative au niveau 3%

E) La variable Density est non significative au niveau 3%

Q 49 (Suite de la question précédente) On note β_4 le coefficient associé à la variable Density. On souhaite tester $H_0 : \beta_4 = 0$ contre $H_1 : \beta > 0$.

A) on rejette H_0 au niveau 1%

B) on ne rejette pas H_0 au niveau 1%

C) on rejette H_0 au niveau 2%

D) on rejette H_0 au niveau 0,05%

E) on ne rejette pas H_0 au niveau 0,05%

Q 50 On se place dans un modèle bayésien. On définit le facteur de risque θ, qui a pour distribution a priori un loi Gamma de paramètres r et λ, et Y, le nombre de sinistres d'un assuré, qui, conditionnellement à θ, suit une loi de Poisson de paramètre θ (paramètre naturel, i.e. paramétrisation usuelle). La loi Gamma de paramètres r et λ a pour densité

$$f_{r,\lambda}(y) = \frac{\lambda^r y^{r-1} \exp(-\lambda y)}{\Gamma(r)}.$$

Dans ce cas,

A) l'espérance de Y est égale à θ

B) l'espérance de Y est égale à r/λ

C) la variance de Y est égale à r/λ^2

D) la variance de Y est égale à θ

E) la variance de Y est égale à $r(\lambda + 1)/\lambda^2$