EXAMEN FINAL, ACT2040, HIVER 2013

Les calculatrices[1] sont autorisées, mais pas les téléphones 'intelligents' (qui devront être rangés pendant toute la durée de l'épreuve). Tous les documents sont interdits. Comme rappelé dans le plan de cours, la triche est interdite.

Dans les feuilles qui suivent, il y a 25 questions,

- 19 questions sur le provisionnement pour sinistres à payer (total de 22 points)
- 6 questions portant sur des aspects de tarification (mais comptant pour 15 points)

Précisions sur la notation Pour chaque question, une courte réponse doit être apportée, détaillée si demandé. Vous gagnez les points si la réponse est claire, et correcte. Si aucun détail n'est demandé, seule la bonne réponse apportera le point. Votre note finale est le total des points (sur 37).

Précisions sur l'épreuve Je ne réponds à aucune question durant l'épreuve. En cas de doute, ou de suspicion d'erreur, merci de m'en faire part dans le cahier d'examen.

1. Méthodes de provisionnement

On a observé le triangle de paiements suivant (la colonne de gauche indiquant l'année)

Triangle 1

	0	1	2	3	4	5
2007	[redacted]	9668	10563	10771	10978	11040
2008	6346	9593	10316	10468	10536	
2009	6269	9245	10092	10355		
2010	5863	8546	9268			
2011	5778	8524				
2012	6184	$\star$	$*$			

[1]BA-35, BA II Plus, TI-30X, TI-30Xa, TI-30XIIS et TI-30XIIB (cf plan de cours).

Question 1. Quel montant a été payé en 2012 pour les sinistres survenus en 2010 ? [*Le calcul sera explicité dans la feuille.*]

Question 2. Le premier montant, en haut à gauche, a été effacé. En utilisant la méthode Chain Ladder, quel montant serait-il réaliste d'avoir ? [*Le calcul sera explicité dans la feuille.*]

Question 3. Quel serait le montant qui devrait être prédit, par la méthode Chain Ladder pour 2012, en $\star$? [*Le calcul sera explicité dans la feuille.*]

Question 4. Quel serait le montant qui devrait être prédit, par la méthode Chain Ladder pour 2012, en $*$? [*Le calcul sera explicité dans la feuille.*]

Question 5. En supposant la première année close, quel serait le montant de provision que l'on devrait constituer pour l'année de survenance 2009, par la méthode Chain Ladder ? [*Le calcul sera explicité dans la feuille.*]

Question 6. En réalité, l'année 2007 n'est pas close, et selon des dires d'experts, il convient de constituer une provision de 108, pour cette année de survenance. En tenant compte de cette information, quel serait votre nouvel estimateur pour le montant de provisions pour l'année de survenance 2009, par la méthode Chain Ladder ? [*Le calcul sera explicité dans la feuille.*]

Fin 2013, on dispose de davantage d'information. En particulier, on a retrouvé le payement effectué en 2007 pour les sinistres survenus en 2007. Et une année supplémentaire de paiements a été observée

Triangle 2

	0	1	2	3	4	5	6
2007	5946	9668	10563	10771	10978	11040	11106
2008	6346	9593	10316	10468	10536	10572	
2009	6269	9245	10092	10355	10507		
2010	5863	8546	9268	9459			
2011	5778	8524	9178				
2012	6184	9013	$*$				

Question 7. Si nous avions connu ce montant fin 2012, le montant qui aurait du être prédit, par la méthode Chain Ladder pour 2012, en $\star$ serait il plus grand ou plus petit que ce qui a été calculé à la question 3. ? [*Une rapide explication sera donnée dans la feuille.*]

Question 8. Sur ce nouveau triangle, quel serait le montant qui devrait être prédit, par la méthode Chain Ladder pour 2012, en $*$? [*Le calcul sera explicité dans la feuille.*]

Question 9. Sans tenir compte du tail factor mentionné dans la question 6, et en utilisant les calculs de la question 5, avait-on constitué assez de provision fin 2012 pour payer les sinistres de 2009 en 2013 ? [*Le calcul sera explicité dans la feuille.*]

Dans les questions 10-12, on va utiliser le triangle suivant, constitué des $C_{i,j}$

Triangle 1b

	0	1	2	3	4	5
2007	5946	9668	10563	10771	10978	11040
2008	6346	9593	10316	10468	10536	
2009	6269	9245	10092	10355		
2010	5863	8546	9268			
2011	5778	8524				
2012	6184					

ainsi que les données suivantes, correspondant aux primes acquises Π_i

Primes acquises

2007	15474
2008	14882
2009	14456
2010	14055
2011	14525
2012	15026

À partir du triangle de paiements, on calcule $S_{i,j} = C_{2007,j} + C_{2008,j} + \cdots + C_{i,j}$, donné dans le triangle suivant

Triangle 3

	0	1	2	3	4	5
2007	5946	9668	10563	10771	10978	11040
2008	12292	19261	20879	21239	21514	
2009	18561	28506	30971	31594		
2010	24424	37052	40239			
2011	30202	45576				
2012	36386					

Question 10. À partir de ces données, calculer les facteurs de transition λ_j par la méthode Chain Ladder, avec $C_{i,j} = \lambda_j \cdot C_{i,j-1}$? [*Aucun calcul ne sera détaillé sur la feuille, un tableau avec les différentes valeurs suffira*]

Question 11. Donnez le montant de provision nécesssaire, toutes années confondues, donné par la méthode Chain Ladder. [*Aucun calcul ne sera détaillé sur la feuille, seul le montant total de provision est demandé*]

Question 12. On suppose que, par année, le ratio sinistres sur primes doit ètre constant, de l'ordre de 72%. En utilisant la charge ultime qui donnerait un tel ratio, en déduire le montant qu'il faudrait provisonner, toutes années confondues. [*Aucun calcul ne sera détaillé sur la feuille, seul le montant total de provision est demandé*]

Dans la suite des questions 13-15, on va utiliser la sortie de régression suivante, basée sur le triangle 1b :

Régression 1

```
Call:
glm(formula = Y ~ i + j, family = quasipoisson, data = B)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.77688    0.05178 169.514  < 2e-16 ***
i2008       -0.04110    0.06871  -0.598   0.5630
i2009       -0.04556    0.06912  -0.659   0.5247
i2010       -0.13654    0.07128  -1.916   0.0844 .
i2011       -0.13771    0.07331  -1.878   0.0898 .
i2012       -0.04716    0.08237  -0.573   0.5796
j1          [redacted]  [redacted]  [redacted]  [redacted]
j2          -2.04177    0.09451 -21.605 1.01e-09 ***
j3          -3.41227    0.20468 -16.672 1.26e-08 ***
j4          -3.83292    0.30632 -12.513 1.97e-07 ***
j5          -4.64975    0.64187  -7.244 2.78e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for quasipoisson family taken to be 25.37786)
```

Question 13 [$\times$2]**.** Quelle est la valeur de l'estimateur masqué sur la sortie précédante, $\widehat{\beta}_1$? [*Le calcul sera explicité dans la feuille.*]

Question 14. Donnez la prédiction du montant de paiements qui sera effectué en 2015 pour les sinistres survenus en 2011. [*Le calcul sera explicité dans la feuille.*]

Question 15. En supposant la première année close, quel serait le montant de provision que l'on devrait constituer pour l'année de survenance 2010, à l'aide de cette régression ? [*Le calcul sera explicité dans la feuille.*]

En cherchant à simplifier le modèle de manière à n'avoir que des variables significatives, on finit par considérer le modèle suivant

Régression 1b

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.74411    0.02632 332.270  < 2e-16 ***
i2010&11    -0.10544    0.04037  -2.612   0.0196 *
j1          -0.67226    0.04254 -15.802 9.29e-11 ***
j2          -2.03823    0.08161 -24.974 1.24e-13 ***
j3&4        -3.55337    0.14965 -23.745 2.59e-13 ***
j5          -4.61697    0.56126  -8.226 6.10e-07 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for quasipoisson family taken to be 19.4877)
```

Question 16. Donnez la prédiction du montant de paiements qui sera effectué en 2015 pour les sinistres survenus en 2011. [*Le calcul sera explicité dans la feuille.*]

Question 17 [$\times 2$]**.** Calculez les montants de provions induits par ce modèle [*Aucun calcul ne sera détaillé sur la feuille*]

Dans la suite des questions 18-19, on va utiliser la sortie de régression suivante, toujours basée sur le triangle 1b :

Régression 2

```
Call:
lm(formula = log(Y) ~ i2008 + j, data = B)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.7531     0.1071  81.765  < 2e-16 ***
i2008        -0.2611     0.1354  -1.928 0.072953 .
j1           -0.6772     0.1553  -4.361 0.000559 ***
j2           -2.0118     0.1658 -12.131 3.73e-09 ***
j3&4         -3.5515     0.1584 -22.421 5.99e-13 ***
j5           -4.6259     0.2778 -16.653 4.40e-11 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.2563 on 15 degrees of freedom
```

Question 18. Quelle est le montant que l'on pense payer, avec ce modèle, en 2013 pour les sinistres survenus en 2012 ? [*Le calcul sera explicité dans la feuille.*]

Question 19 [×2]**.** En supposant la première année close, quel serait le montant de provision que l'on devrait constituer à l'aide de cette régression, pour toutes les années ? [*Le calcul sera explicité dans la feuille.*]

2. Tarification

Trois exercices de tarifications sont proposés. Ils peuvent se traîter indépendamment. Il y a toutefois une logique à les faire dans l'ordre.

2.1. **Tarification, partie 1.** On suppose que pour un assuré, sa fréquence annuelle de sinistres suit une loi de Poisson de moyenne λ. Toutefois, on sait que la moitié des sinistres seulement dépasseront la franchise d. On sait aussi que si un sinistre dépasse d, il aura pour coût Y de distribution conditionnelle

$$\mathbb{P}(Y > y | Y > d) = \left(\frac{d}{x}\right)^{\alpha}, \text{ pour } y \geq d, \text{ et sinon } 1.$$

Question 20 [×2]**.** Quelle sera la prime pure d'un tel contrat ?

Question 21 [×2]**.** Afin d'attirer de nouveau client, on souhaite augmenter la franchise d de manière à proposer un contrat deux fois moins cher. À quel niveau faudra-t-il placer la nouvelle franchise ?

2.2. **Tarification, partie 2.** Dans la section suivante, on cherche à construire un tarif en assurance automobile, partir de 2 variables explicative: l'âge d'un assuré et son sexe. On dispose d'un ensemble de données et plusieurs modèles sont ajustés. Tous les assurés sont restés un an dans la base. On suppose de plus quand les variables sexe et âge sont indépendantes. On suppose aussi disposer des statistiques suivantes

	total	homme	femme
nombre moyen d'accident, par an	0.0793	0.0838	0.07242
coût moyenne d'un accident	579.669	622.4477	504.2472
moyenne du log(coût) d'un accident	6.173865	6.249597	6.040344

Régression A

```
Call:
glm(formula = N ~ age + sexe, family = poisson)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.111191   0.127413 -16.570  < 2e-16 ***
age         -0.010135   0.002281  -4.442 8.89e-06 ***
sexeH        [redacted]     [redacted]    2.069   0.0385 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)
```

Régression B

```
Call:
glm(formula = N ~ cut(age, c(0, 30, 100)) + sexe, family = poisson,
    data = basen)

Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)              -2.37019    0.11272 -21.027  < 2e-16 ***
age (30,100]             -0.28704    0.10988  -2.612  0.00899 **
sexeH                     0.14917    0.07391   2.018  0.04356 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)
```

Régression C

```
Call:
lm(formula = log(Y) ~ sexe, data = basec)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.04034    0.03601 167.734  < 2e-16 ***
sexeH        [redacted]    [redacted]   4.642 4.05e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.6101 on 791 degrees of freedom
Multiple R-squared: 0.02651,        Adjusted R-squared: 0.02528
F-statistic: 21.54 on 1 and 791 DF,  p-value: 4.046e-06
```

Régression D

```
Call:
glm(formula = Y ~ sexe, family = Gamma(link = "log"), data = basec)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.22307    0.03791 164.146  < 2e-16 ***
sexeH        0.21059    0.04746   4.437 1.04e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for Gamma family taken to be 0.4125056)
```

On suppose que plusieurs compagnies sont en compétition sur le marché.

- compagnie 1 : pas de segmentation
- compagnie 2 : modèle A pour les nombres, C pour les coûts
- compagnie 3 : modèle B pour les nombres, C pour les coûts
- compagnie 4 : modèle A pour les nombres, D pour les coûts
- compagnie 5 : modèle B pour les nombres, D pour les coûts

Question 22 [×4]**.** Un jeune homme de 20 ans souhaite s'assurer. Quelle compagnie lui propose la prime pure la plus faible ? Quel est le montant proposé ? [*Les calculs seront détaillés dans la feuille.*]

Question 23 [×2]**.** Il trouve la prime trop chère, et demande à sa mère (qui a 50 ans) si elle peut assurer la voiture à son nom. Quelle compagnie proposera la prime pure la plus faible ? Quel est le montant proposé ? [*Les calculs seront détaillés dans la feuille.*]

2.3. **Tarification, partie 3.**

Une autre compagnie arrive sur le marché, et propose d'utiliser une **loi exponentielle** pour les coûts de sinistres

Régression E

```
Call:
glm(formula = Y ~ sexe, family = Gamma(link = "log"), data = basec)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.22307    0.05903  105.42  < 2e-16 ***
sexeH        0.21059    0.07390    2.85  0.00437 **
---

(Dispersion parameter for Gamma family taken to be 1)
```

On a alors

- compagnie 6 : modèle A pour les nombres, E pour les coûts

Question 24. Un homme de 30 ans souhaite s'assurer. Quelle serait sa prime pure ? [*Les calculs seront détaillés dans la feuille.*]

Question 25 [×4]**.** Il trouve la prime trop chère, mais cette fois, l'assureur lu propose un contrat avec une franchise de 100. En supposant les coûts et la fréquence indṕendants, quelle sera la prime pure proposée ? [*Les calculs seront détaillés dans la feuille.*]