

EXAMEN FINAL, ACT2040, AUTOMNE 2013

Les calculatrices¹ 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 42 questions,

- question 1 (préliminaire) sur la régression binomiale (total de 3 points)
- questions 2-20 sur la gravité d’accidents de la route (total de 20 points)
- questions 21-34 sur le provisionnement pour sinistres à payer (total de 15 points)
- questions 35-42 sur la modélisation de la durée de vie prospective (mais comptant pour 12 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 50).

Précisions sur le code R La code suivant permet de faire la régression binomiale, $Y_i \sim \mathcal{B}(N_i, p_i)$ avec $p_i = H(\mathbf{X}_i^T \boldsymbol{\beta})$, où H est la transformation logistique (standard),

```
glm(Y/N ~ X1+...+Xk, family="binomial", weights=N)
```

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.

En fait, l’examen était un peu long... j’utiliserais un barème basé sur les pondérations évoquées, mais un coefficient multiplicatif sera utilisé...

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

1. QUESTION PRÉLIMINAIRE

1. Page 1 : On dispose d'un échantillon $\{y_1, \dots, y_n\}$ i.i.d. suivant une loi $\mathcal{B}(n, p)$, avec $n = 10$. Pour estimer p , on utilise la fonction `glm()`. A partir de la sortie page 1, proposez un estimateur pour p . [un minimum d'explication sera exigé]

Ce que je voulais, c'était exploiter le fait que

$$p_i = H(\mathbf{X}_i^\top \boldsymbol{\beta}) = \frac{\exp[\mathbf{X}_i^\top \boldsymbol{\beta}]}{1 + \exp[\mathbf{X}_i^\top \boldsymbol{\beta}]}$$

et donc

$$\hat{p}_i = H(\mathbf{X}_i^\top \hat{\boldsymbol{\beta}}) = \frac{\exp[\mathbf{X}_i^\top \hat{\boldsymbol{\beta}}]}{1 + \exp[\mathbf{X}_i^\top \hat{\boldsymbol{\beta}}]}$$

Or ici, on n'a pas de variable explicative (on régresse uniquement sur la constante, aussi

$$\hat{p}_i = \hat{p} = \frac{\exp[\hat{\beta}_0]}{1 + \exp[\hat{\beta}_0]}$$

avec ici $\hat{\beta}_0 \sim -1.3714$. Bref,

$$\hat{p} = \frac{\exp[-1.3714]}{1 + \exp[-1.3714]} = 0.2023937 \sim 20\%.$$

On peut aussi utiliser les méthodes des moments, car on a les effectifs :

$$\bar{Y} = \frac{45 \cdot 0 + 133 \cdot 1 + 160 \cdot 2 + \dots + 8 \cdot 5 + 2 \cdot 6}{500} = \frac{1012}{500} = 2.024$$

Maintenant, il était noté que c'était une loi $B(10, p)$, de moyenne $10 \cdot p$. Aussi, la méthode des moments suggère d'utiliser

$$\hat{p} = \frac{\bar{Y}}{n} = \frac{2.024}{10} = 0.2024 \sim 20\%.$$

(mais c'était la méthode précédente qu'il fallait utiliser).

2. MODÉLISATION DE LA GRAVITÉ DES SINISTRES

2. Page 2, point 1 : Quel est le chiffre qui a été supprimé (derrière les **XXX**) [aucune justification]

Je vais prendre un tout petit peu de temps pour justifier, quand mme... Ici, on devait trouver le nombre de décès, soit **676** d'après la ligne précédente. Maintenant, on avait 76341 personnes dans la base (c'est le numéro des lignes). Par soustraction, le nombre qui manque est égal à $76341 - 75665$ (qui - comme par hasard - vaut 676).

2. Page 3, point **2** : Sachant qu'une personne a eu un accident, et qu'elle un taux d'alcool entre 50 et 80, quelle est la probabilité que l'accident soit mortel ? Quelle serait la probabilité si la personne avait eu un taux d'alcool entre 0 et 20 ? [*un minimum d'explication sera exigé*]

Sur les 76341 accidents, $244 + 32 + 11 + 13 = 300$ ont touché des personnes qui avaient un taux d'alcool entre 50 et 80. Sur ces 300 personnes, $1 + 2 + 2 + 11 = 16$ sont décédés. Aussi, la probabilité que l'accident soit mortel pour une personne ayant un taux d'alcool entre 50 et 80 est $16/300$, soit **5.33%**. Pour un taux d'alcool entre 0 et 20, on parle de $372 + 76 + 51 + 93 = 592$ décès parmi $3218 + 2059 + 6237 + 62433 = 73947$ accidents, soit un taux de décès de l'ordre de **0.80%**. Le taux moyen n'est pas la moyenne des taux, on en a assez parlé dans la section sur le provisionnement.

3. Page 3, point **3** : Quel est le chiffre qui a été supprimé (derrière les **XXXXXX**) [*aucune justification*]

Il s'agit de la moyenne de Y , c'est à dire le taux de décès pour l'ensemble des accidents. En utilisant les tables de la page 2, on a 676 accidents mortels, sur 76341, soit un taux de 0.88% Le chiffre manquant était ici **0.008855006**.

4. Page 3, point **4** : Quel est le chiffre qui a été supprimé (derrière les **XXXXXX**) [*un minimum d'explication sera exigé*]

Par construction, on utilise ici la méthode de biais minimal proposé par Bailey, avec ici la contrainte sur les sommes par ligne et par colonnes. Compte tenu de l'algorithme utilisé,

on a besoin d'utiliser le comptage d'accident par type de véhicules, et le nombre de décès, pour la catégorie 50-80

- bus-truck, D=1, C=13, L=3.3578117,
- moto, D=2, C=11, L=3.0102805,
- van, D=2, C=32, L=1.4996518 ,
- voiture, D=11, C=244, L=0.7585497,

A partir de là,

$$C = \frac{\sum D}{\sum C \cdot L} = \frac{1 + 2 + 2 + 11}{13 \cdot 3.35 + 11 \cdot 3.01 + 32 \cdot 1.499 + 244 \cdot 0.75} \sim \frac{16}{309.83} \sim 0.051.$$

La réponse attendue était **0.05164**. Cela dit, le vecteur C s'interprète comme le taux de décès des personnes qui conduisent avec un certain taux d'alcool (l'autre vecteur est un coefficient multiplicatif). Aussi, compte tenu des questions précédentes, il était légitime d'espérer avoir un chiffre de l'ordre de 5%. Cette justification donne la moitié des points.

5. Page 4, point **5** : Quels sont les chiffres qui ont été supprimés (derrière les **XXXXXX**) [*un minimum d'explication sera exigé*]

On va utiliser le calcul précédant pour prédire tous les taux. Pour bus-truck par exemple, on s'attend à voir $0.05 \cdot 3.3578$ soit 17.3% - ou pour être plus précis **0.1734**. Pour moto, $0.05 \cdot 3.0102$ soit 15,5% - ou pour être plus précis **0.1554**. Si on continue, pour van, **0.0774** et pour voiture **0.0392** (on retrouve ici les chiffres qui figuraient dans la première sortie, distribuée avant l'examen).

6. Page 4, point **6** : Quels est le chiffre qui a été supprimé (derrière les **XXXX**) [*un minimum d'explication sera exigé*]

On a vu en cours que la régression de Poisson, avec un lien logarithmique, sur des facteurs, coïncide avec l'approche de Bailey. La valeur manque est $\hat{\beta}$ de telle sorte que

$$\exp[-3.6132 + \hat{\beta}]$$

correspond au taux de décès prédit pour **bus-truck** (modalité de référence pour le type de véhicule) et **50-80** pour le modèle de Poisson. Pour la méthode de Bailey, on avait fait le calcul a la question précédente, et on avait trouvé 0.1734. Aussi,

$$\exp[-3.6132 + \hat{\beta}] = 0.1734 \text{ i.e. } \hat{\beta} = \log[0.1734] + 3.6132$$

soit $\hat{\beta} = 1.8610$. La réponse attendue était **1.8610**

6. Page 4, point **7** : Quels est le chiffre qui a été supprimé (derrière les **XXXX**) [*un minimum d'explication sera exigé*]

Pour récupérer l'écart-type, on utilise le test de type Student,

$$z = \frac{\hat{\beta}}{\sqrt{\widehat{\text{Var}}(\hat{\beta})}} \text{ soit } \sqrt{\widehat{\text{Var}}(\hat{\beta})} = \frac{\hat{\beta}}{z}$$

donc ici, l'écart-type (Std. Error) est $1.8610/7.344 = 0.253$. La réponse attendue était **0.253**.

7. Page 4, point **8** : Y-avait-il des étoiles, et si oui, combien ? (derrière les **XXXX**) [*un minimum d'explication sera exigé*]

On a obtenue une valeur z de l'ordre de 7. On peut voir dans le tableau qu'avec 5, on avait une p -value de l'ordre de 10^{-7} . Ici, on doit tre bien plus faible ! Bref, on doit avoir *******. (Pour information, la p -value était de l'ordre de 10^{-13} , cf. sortie en ligne avant l'examen).

8. Page 4 : A l'aide de la régression de Poisson, quelle serait la probabilité qu'un accident soit mortel pour une personne circulant en bus (catégorie **bus-truck**) sans avoir bu d'alcool (catégorie 0-20) ? [*un minimum d'explication sera exigé*]

Pour faire une prvision pour une personne dont les caractéristiques sont **bus-truck** (modalité de référence) et **0-20** (modalité de référence), comme on a un lien logarithmique, on a

$$\exp[(-3.6132)] \sim 0.02696$$

La réponse attendue était ici **0.0267**.

9. Page 4 : Que deviendrait cette probabilité si la personne avait bu beaucoup d'alcool (catégorie 150+) ? [*un minimum d'explication sera exigé*]

Si la personne avait beaucoup bu (modalité 150+), comme on a un lien logarithmique, on a

$$\exp[(-3.6132) + 1.4897] \sim 0.11961$$

La réponse attendue était ici **0.1196** (presque 12%). Pour ces deux questions, il est possible de faire tourner le code suivant

```
> predict(reg1,newdata=data.frame(cat.alc=c("0-20","150+"),
+ vehicule=c("bus-truck","bus-truck")),type="response")
      1      2
0.02696618 0.11961801
```

10. Pages 6-7, points **9a** et **9b** : Est-ce surprenant que les valeurs de coefficients soient proches entre les deux régressions (**reg2** Poisson avec un lien log, et **reg3** Bernoulli avec un lien logistique) ? [*un minimum d'explication sera exigé*]

Pour l'instant, dans la régression **reg2**, on utilisait un modèle

$$\begin{cases} \mathbb{E}[Y_i] = \exp[\beta_0 + \beta_1 V_i + \beta_2 A + i] \\ \text{Var}[Y_i] = \mathbb{E}[Y_i] \end{cases}$$

où (histoire de poser le modèle aussi proprement que possible), V désigne le véhicule et A le taux d'alcool. La première relation vient du lien logarithmique, et la seconde, de la loi de Poisson. Pour le modèle **reg3**, on souhaite avoir

$$\begin{cases} \mathbb{E}[Y_i] = \frac{\exp[\gamma_0 + \gamma_1 V_i + \gamma_2 A_i]}{1 + \exp[\gamma_0 + \gamma_1 V_i + \gamma_2 A_i]} \\ \text{Var}[Y_i] = \mathbb{E}[Y_i] \cdot [1 - \mathbb{E}[Y_i]] \end{cases}$$

La première propriété vient du fait que que l'on a un lien logistique, et la seconde est une proprit de la loi de Bernoulli. Comme on l'a vu, ces relations sur les deux premiers moments

caractérisent complètement notre modèle GLM (une première relation pour la fonction de lien, une seconde pour la fonction variance, i.e. la loi de la variable d'intérêt). Maintenant, notons une chose importante : les probabilités que l'on calcule sont faibles (cf toutes les questions précédentes). Aussi, $\mathbb{E}[Y_i]$ est proche de 0. Or si $u \sim 0$, notons deux choses,

$$\begin{cases} \frac{u}{1+u} \sim u \\ u \cdot (1-u) \sim u \end{cases}$$

Aussi, pour le second modèle,

$$\begin{cases} \mathbb{E}[Y_i] \sim \exp[\gamma_0 + \gamma_1 V_i + \gamma_2 A_i] \\ \text{Var}[Y_i] \sim \mathbb{E}[Y_i] \end{cases}$$

Eh, mais c'est le premier modèle précédent ! Donc oui, **il est normal d'avoir des estimations proches**, car les modèles sont proches.

11. Page 6, point **10** : Quels sont les chiffres qui ont été supprimés (derrière les **XXXX**)
[un minimum d'explication sera exigé]

Faire la prévision signifie que l'on cherche à calculer

$$\exp[-3.659 - 1.441 + \hat{\beta}]$$

où $\hat{\beta}$ dépendra du taux d'alcool, puisqu'on utilise une régression GLM avec un lien logarithmique. Aussi,

- pour 0-50, $\exp[-3.659 - 1.441] = \mathbf{0.00609}$
- pour 50-150, $\exp[-3.659 - 1.441 + 1.306] = \mathbf{0.022498}$
- pour 150+, $\exp[-3.659 - 1.441 + 1.490] = \mathbf{0.02707}$

qui sont les trois valeurs attendues (cf. sortie distribuée avant l'examen pour ceux qui veulent refaire les calculs avec R).

12. Page 7, point **11** : Quels sont les chiffres qui ont été supprimés (derrière les **XXXX**)
[un minimum d'explication sera exigé]

Faire la prévision signifie que l'on cherche à calculer

$$\frac{\exp[-3.627 - 1.469 + \hat{\beta}]}{1 + \exp[-3.627 - 1.469 + \hat{\beta}]}$$

où $\hat{\beta}$ dépendra du taux d'alcool, puisqu'on utilise une régression GLM avec un lien logarithmique. Lançons nous:

- pour 0-50, $\frac{\exp[-3.627 - 1.469]}{1 + \exp[-3.627 - 1.469]} = \mathbf{0.00608}$
- pour 50-150, $\frac{\exp[-3.627 - 1.469 + 1.341]}{1 + \exp[-3.627 - 1.469 + 1.341]} = \mathbf{0.022857}$
- pour 150+, $\frac{\exp[-3.627 - 1.469 + 1.537]}{1 + \exp[-3.627 - 1.469 + 1.537]} = \mathbf{0.027687}$

On observe que les prévisions sont proches.

13. Page 7 : En utilisant la régression `reg3`, sachant qu'un individu conduisait une voiture et avait consommé beaucoup d'alcool (catégorie 150+) quelle est la probabilité que l'accident ait été mortel ? [*un minimum d'explication sera exigé*]

Il s'agit de la dernière probabilité calculée dans la question précédente. La réponse attendue était **0.0277**. La question peut sembler stupide, mais je crois que le but était d'insister sur l'interprétation, avec le bloc de trois questions qui viennent. Formellement, on demande $\mathbb{P}_v(Y = 1|A = 150+)$ (où l'indice v est ici mis pour rappeler qu'on ne considère que les conducteurs de voiture).

14. Page 7 : En utilisant la régression `reg3`, sachant qu'un individu conduisait une voiture et avait consommé beaucoup d'alcool (catégorie 150+) quelle est la probabilité que l'accident ait causé des dommages corporels, sans avoir été mortel ? [*un minimum d'explication sera exigé*]

Cette fois, ce qu'on demande, c'est $\mathbb{P}(Y = 0|A = 150+)$ et comme $Y \in \{0, 1\}$,

$$\mathbb{P}_v(Y = 0|A = 150+) = 1 - \mathbb{P}_v(Y = 1|A = 150+)$$

Bref, inutile de relancer la régression sur $1 - Y$ la probabilité était ici **0.9723**.

15. Page 7 : En utilisant la régression **reg3**, sachant qu'un individu conduisait une voiture a eu un accident mortel, quelle est la probabilité que l'individu ait consommé beaucoup d'alcool (catégorie 150+) ? [*un minimum d'explication sera exigé*]

Dans la question précédentes, on avait obtenu les probabilités qu'un accident soit mortel, pour toutes les catégories de niveau d'alcool. On nous demande ici $\mathbb{P}_v(A = 150+|Y = 1)$. Or on a la formule de Bayes,

$$\mathbb{P}_v(A = 150+|Y = 1) = \frac{\mathbb{P}_v(A = 150+ \cap Y = 1)}{\mathbb{P}_v(Y = 1)}$$

et en utilisant une nouvelle fois la formule de Bayes,

$$\mathbb{P}(A = 150+|Y = 1) = \frac{\mathbb{P}(A = 150+)}{\mathbb{P}(Y = 1)} \cdot \mathbb{P}(Y = 1|A = 150+)$$

Bon, on connaît la dernière probabilité (on vient de la calculer). Pour les deux autres, on a juste besoin d'utiliser les statistiques données dans les premières questions.

$$\mathbb{P}_v(Y = 1) = \frac{372 + 25 + 11 + 17}{62433 + 795 + 56 + 244 + 762} \sim 0.00661$$

(0.66% de tous les accidents corporels impliquant des voitures sont mortels) alors que

$$\mathbb{P}_v(A = 150+) = \frac{795}{62433 + 795 + 56 + 244 + 762} \sim 0.01236$$

(1.24% des accidents corporels impliquant des voitures concernaient des personnes avec un fort taux d'alcool). On obtient alors

$$\mathbb{P}_v(A = 150+|Y = 1) = \frac{795}{372 + 25 + 11 + 17} \cdot 0.0276 \sim \mathbf{0.0516}.$$

Bref, si on nous rapporte qu'une personne est décédé dans un accident de voiture, il y a 5.16% chance que la personne avait un fort taux d'alcool (si on sait juste qu'il s'agissait d'un accident corporel, on était à 1.24%). Si on utilise les tableaux de contingence donnés au début des sorties (car on a tous les effectifs). Parmi les décès (table *deces*, page 3), sur $372+25+11+17=425$ décès observés en voiture, 25 présentaient un fort taux d'alcool. Soit 5.88%. On note que la probabilité induite par notre modèle est un peu plus faible.

16. Page 8 : Dans la sortie de la régression, avait-on sous-dispersion, sur-dispersion, ou équi-dispersion ? [*un minimum d'explication sera exigé*]

Ce point a été évoqué rapidement en cours, et dans la correction de l'an passé. On note ici que l'écart-type de tous les coefficients $\hat{\beta}$ est (un peu) plus *faible* avec la régression quasipoisson qu'avec la régression Poisson, ce qui devrait se traduire par une **sous-dispersion**. Donc $\hat{\phi}$ devrait être inférieur à 1. Si on reprend la sortie complète, effectivement, $\hat{\phi} \sim 0.98071$.

17. Page 9, point **13** : En utilisant la régression `reg5`, sachant qu'un individu conduisait une voiture et avait consommé beaucoup d'alcool (catégorie 150+) quelle est la probabilité que l'accident ait été mortel ? [*un minimum d'explication sera exigé*]

On est ici sur la régression multinomiale, avec un lien logistique. Donc pour une personne dont les caractéristiques sont *voiture* (modalité de référence) et *150+* (modalité de référence), comme on a un lien logarithmique, la probabilité que la personne décède est obtenue en utilisant les estimateurs de la première ligne

$$\frac{\exp[(-3.1282) + (-1.0360) + (1.1238)]}{1 + \exp[(-3.1282) + (-1.0360) + (1.1238)]} \sim \mathbf{0.045633}$$

18. Page 9 : En utilisant la régression `reg5`, sachant qu'un individu conduisait une voiture est décédé. quelle est la probabilité que la personne avait consommé beaucoup d'alcool (catégorie 150+) ? [*un minimum d'explication sera exigé*]

Tiens, on retrouve une question précédente, basée sur la formule de Bayes (on veut la probabilité inverse). Et les probabilités conditionnelles sont ici les mêmes, donc la probabilité demandée est

$$\mathbb{P}_v(A = 150+ | Y = 1) = \frac{795}{372 + 25 + 11 + 17} \cdot 0.045633 \sim \mathbf{0.08536}.$$

19. Page 10, point **14** : dans la régression reg6, la variable **sexe** est-elle significative ? La variable **anciennete** est-elle significative ? [*un minimum d'explication sera exigé*]

Pour rappel, dans les GLM, les estimateurs de coefficients $\hat{\beta}_k$ sont asymptotiquement Gaussiens (c'est une propriété des estimateurs obtenus par maximum de vraisemblance). Comme on a beaucoup d'observations, $\hat{\beta}_k$ sera supposé Gaussien. Et comme on a la valeur observé et un estimateur de l'écart-type, on peut construire un test de normalité. La méthode la plus simple et rapide est de noter qu'un intervalle de confiance à 95% pour β_k est

$$[\hat{\beta}_k \pm 2\widehat{\text{sd}}(\hat{\beta}_k)]$$

Aussi, si 0 appartient à cet intervalle de confiance, on dira que la variable n'est pas significative (c'est ce qu'on aura décidé avec un critère basé sur une p -value à 5%). Pour la variable **sexe**, l'intervalle de confiance est

$$[0.3509 \pm 2 \cdot 0.0983] = [0.1543; 0.5475]$$

pour les décès et

$$[0.2443 \pm 2 \cdot 0.01612] = [0.21209; 0.27657]$$

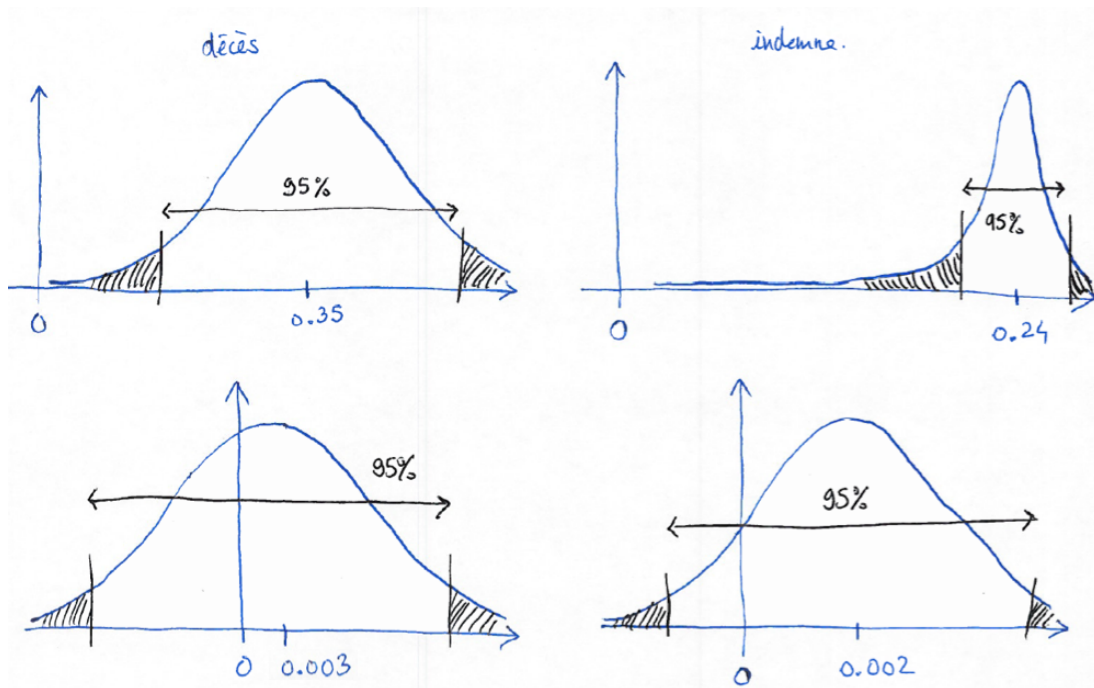
pour la probabilité d'être indemne. Aussi, la variable **sexe peut être considérée comme significative**, car elle 'explique' à la fois la probabilité d'avoir un décès et de sortir indemne. En revanche, pour la variable **anciennete**, l'intervalle de confiance est ici

$$[0.0032 \pm 2 \cdot 0.00979] = [-0.01671; 0.02321]$$

pour les décès et

$$[0.0022 \pm 2 \cdot 0.0018] = [-0.0015566; 0.0059866]$$

pour la probabilité d'être indemne. Cette fois, les deux intervalles contiennent la valeur 0 : aussi, la variable **sexe peut être considérée comme non significative**, quelle que soit la modalité que l'on cherche à expliquer.



20. Page 10, point 14 : à la question 17, nous avons calculé la probabilité que l'accident soit mortel en utilisant la régression `reg5`. Que deviendrait cette probabilité si la personne était un homme (`sexe='M'`) et si la voiture était neuve (`anciennete=0`) ? [un minimum d'explication sera exigé]

Un dernier petit calcul pour cette partie... On a toujours un modèle logistique, et ici la probabilité que l'accident soit mortel (modalité `deces`) est (pour rappel, la personne conduit une voiture et elle avait beaucoup bu 150+)

$$\frac{\exp[(-3.4747) + (-0.912) + 1.0826 + 0.35092]}{1 + \exp[(-3.4747) + (-0.912) + 1.0826 + 0.35092]} \sim \mathbf{0.049586}$$

La probabilité est ici très légèrement supérieure (ce qui était prévisible car - quelque part - si on ne prenait pas en compte le sexe, comme dans la question précédente, on avait une valeur intermédiaire entre les hommes et les femmes. Or pour les hommes, la probabilité que l'accident soit mortel est plus élevée (la variable est significativement positive)).

3. MÉTHODES DE PROVISIONNEMENT

21. Page 11, point **15** : Quel montant a été payé en 1991 pour les sinistres survenus en 1989 ? [*aucune justification ne sera exigé*]

On se lance : pour les sinistres survenus en 1989

- 5984 a été payé l'année 0, c'est à dire en 1989
- 9939-5984=3955 a été payé l'année 1, c'est à dire en 1990
- 11725-9939=1786 a été payé l'année 2, c'est à dire en 1991

La réponse attendue était **1786**.

22. Page 11 : En utilisant la méthode Chain-Ladder, proposez une prévision pour la valeur **XXXX** [*un minimum d'explication sera exigé*]

Pour le coefficient de transition, on utilise

$$\lambda_7 = \frac{12199 + 13109 + 15720}{12120 + 13034 + 15637} \sim 1.0058$$

Pour prédire la valeur $C_{1991,7}$, on utilise le fait que

$$\hat{C}_{1991,7} = C_{1991,6} \cdot \lambda_7 = 13872 \cdot 1.0058 \sim 13952.6$$

La réponse attendue était **13953**.

23. Page 11, point **16** : Quels sont les chiffres qui a été supprimés (derrière les **XXXX**) [*aucune justification ne sera exigé*]

Opps, j'ai un peu raté ma question... on veut calculer $Y_{1990,2}$ qui est ici $C_{1990,2}$ (on n'a pas encore fait tourné le code pour définir les incréments). Bref, la réponse logique a la question était **14171**. Si on avait attendu d'avoir effectivement construit la matrice d'incrément, on aurait obtenu 1750.

24. Page 11 : De combien augmenterait le montant de provisions si on utilisait un *tail factor* de 5% (i.e. $\lambda_\infty = 1.05$ en utilisant les notations du cours) ? [*un minimum d'explication sera exigé*]

Avec ce *tail factor*, les charges ultimes (Ultimate) vont augmenter de 5%. Bon, ben il faut se lancer (on va commencer par les années les plus récentes)

- $7704 \cdot 1.05 - 3689 = 4399.7$
- $10529 \cdot 1.05 - 8202 = 2853.3$
- $10088 \cdot 1.05 - 8984 = 1608.4$
- $9923 \cdot 1.05 - 9325 = 1094.3$
- $9039 \cdot 1.05 - 8757 = 734.4$
- $11453 \cdot 1.05 - 11282 = 743.8$
- $13964 \cdot 1.05 - 13872 = 789.8$
- $13732 \cdot 1.05 - 15720 = 799.0$
- $13113 \cdot 1.05 - 13113 = 655.6$
- $12215 \cdot 1.05 - 12215 = 610.7$

et si on somme tout, on obtient **14288.71**, ce qui constitue une hausse de 66% du montant des provisions ! Mon point ici était d'insister sur le fait qu'augmenter *un peu* l'estimation de charge (+5%) peut impacter fortement l'estimation des réserves (+66%). En tapant la correction, je vais croire que mon examen était peut tre un peu long...

25. Page 12, point **17** : Avec reg1 , quels montant payera-t-on en 2000 pour les sinistres survenus en 1994 ? [*un minimum d'explication sera exigé*]

Point important pour les questions qui viennent : on demande combien on va payer pour une année donnée, pour une année de survenance donnée. Ceci correspond à ce qu'on a appelé 'incrément', et que l'on a noté Y . On cherche ici une prévision de Y . On veut une prévision pour des sinistres de 1994 (valeur de la variable A) pour l'année 2000, c'est à dire l'année de développement $B=6$ (puisque $1994+6=2000$, c'est pour cette raison qu'on

commence les développement en 0). Bref, ici la prévision pour $\log Y$ sera

$$146.48 + 1994 \cdot (-0.06916) + 6 \cdot (-0.74997) \sim 4.071325$$

Compte tenu des propriétés de la loi lognormale, la valeur moyenne de Y (correspondant à notre prévision) sera

$$\exp\left(4.071325 + \frac{0.4881}{2}\right) \sim 66.0522$$

car l'estimateur de σ (l'écart type des résidus dans le modèle linéaire sur $\log Y$, appelé Residual standard error) est 0.4881. La réponse attendue était ici **66.05**.

26. Page 13, point **18** : Avec `reg2`, quels montant payera-t-on en 2000 pour les sinistres survenus en 1994 ? [*un minimum d'explication sera exigé*]

C'est un peu comme la question d'avant, sauf que cette fois, on a des facteurs. Il faut donc utiliser $\hat{\alpha}_{1994}$ et $\hat{\beta}_6$ (avec les notations 'intuitives'). Aussi, ici la prévision pour $\log Y$ sera

$$8.8241 + (-0.4954) + (-4.6344) \sim 3.694$$

Compte tenu des propriétés de la loi lognormale, la valeur moyenne de Y (correspondant à notre prévision) sera

$$\exp\left(3.694314 + \frac{0.4385}{2}\right) \sim 44.276$$

car les résidus sont ici de variance un peu plus faible. La réponse attendue était ici **44.27**.

27. Page 14, point **19** : Avec `reg3`, quels montant payera-t-on en 2000 pour les sinistres survenus en 1994 ? [*un minimum d'explication sera exigé*]

Cette fois, on modélise non plus le logarithme des incréments de paiements, mais les paiements. Avec un modèle de Poisson (et surtout un lien logarithmique). Aussi, ici la prévision pour Y , pour $A=2000$ et $B=6$ sera

$$\exp[8.674 + (-0.20779) + (-4.0501)] = \exp[4.382516] \sim 80.0391$$

La réponse attendue était ici **80.04**.

28. Page 15, point **20** : Avec **reg4**, quels montant payera-t-on en 2000 pour les sinistres survenus en 1994 ? [*un minimum d'explication sera exigé*]

Allez, la dernière valeur... On va aller vite, la prévision sera

$$\exp[11.3872 + 1994 \cdot (-0.051564) + 6 \cdot (-0.6978)] = \exp[4.416152] \sim 82.777$$

La réponse attendue était ici **82.78**.

29. Page 16, point **21** : Quel modèle prédit le montant de provisions pour sinistres à payer le plus important ? [*un minimum d'explication sera exigé*]

Pour les premières valeurs, comme on l'a rappelé dans les questions précédentes, il faut faire un peu attention. Pour rappel, la prévision sur le logarithme des incréments ne permet pas d'avoir une prévision des incréments, puisque

$$\mathbb{E}(Y) \neq \exp[\mathbb{E}(\log Y)]$$

Par contre, comme on a une loi lognormale, on sait que

$$\mathbb{E}(Y) = \exp \left[\mathbb{E}(\log Y) + \frac{\sigma^2}{2} \right] = \exp[\mathbb{E}(\log Y)] \cdot \exp \left[\frac{\sigma^2}{2} \right]$$

Après, pour le montant total de provision, on utilise la linéarité de l'espérance mathématique, puisque

$$\mathbb{E} \left(\sum Y_{i,j} \right) = \exp \left[\frac{\sigma^2}{2} \right] \cdot \sum \exp[\mathbb{E}(\log Y_{i,j})]$$

il faut corriger avec le facteur (multiplicatif) $\exp[\sigma^2/2]$. Aussi, on a

- modèle **reg1** : $7643.561 \cdot \exp(0.4881/2) = 9756.304$
- modèle **reg2** : $8060.825 \cdot \exp(0.4385/2) = 10036.88$
- modèle **reg3** : 8177.528
- modèle **reg4** : 7600.721

On a maintenant les quatre valeurs à comparer. Le montant le plus important est obtenu avec le second modèle, il fallait donc répondre **reg2**.

30. Page 17, points **22** et **20**: Est-ce suprenant que les coefficients soient les mêmes pour les facteurs B ? [*un minimum d'explication sera exigé*]

Avec le premier modèle, on avait

$$Y_{i,j} \sim \mathcal{P}(\lambda_{i,j} \text{ avec } \lambda_{i,j} = \exp[A_i + B_j])$$

alors qu'avec le second, on a une variable offset, de telle sorte que

$$Y_{i,j} \sim \mathcal{P}(P_i \cdot \tilde{\lambda}_{i,j} \text{ avec } \lambda_{i,j} = \exp[\tilde{A}_i + \tilde{B}_j])$$

Aussi, ce second modèle peut aussi s'écrire

$$Y_{i,j} \sim \mathcal{P}(\mu_{i,j}) \text{ avec } \mu_{i,j} = P_i \cdot \exp[\tilde{A}_i + \tilde{B}_j] = \exp[\underbrace{\log P_i + \tilde{A}_i}_{A_i} + \underbrace{\tilde{B}_j}_{B_j}]$$

Les deux modèles sont rigoureusement équivalents. Pour les coefficients de type A, il y aura une translation de l'ordre du logarithme de la prime, mais pour les coefficients de type B, on a **exactement les mêmes valeurs**.

31. Page 17, point **22** : Avec reg5, quels montants payera-t-on en 2000 pour les sinistres survenus en 1994 ? [*aucune justification ne sera exigée*]

Comme on l'a dit dans la question précédente, non seulement les coefficients de type B sont identiques, les modèles sont rigoureusement équivalents : ils vont donc prédire exactement la même chose. Sans calcul, on peut garantir que la prévision sera $\exp[4.416152]$ soit 82.777. Mais on peut vérifier ! La prime pour 1994 était de 15083 (page 11), aussi, la prévision sera

$$15083 \cdot \exp[(-0.9989) + (-0.15611) + (-4.050147)] = 15083 \exp[-5.205171] \sim 82.777$$

La réponse attendue était ici **82.78**.

32. Page 18, point **23** : Avec reg6, quels montants payera-t-on en 2000 pour les sinistres survenus en 1994 ? [*aucune justification ne sera exigée*]

Cette fois, on a remplacé le facteur ligne par les primes, toujours avec un modèle de Poisson, i.e.

$$Y_{i,j} \sim \mathcal{P}(P_i \cdot \exp[B_j]) \text{ ou } Y_{i,j} \sim \mathcal{P}(\exp[\log P_i + B_j])$$

Bon, reste à obtenir notre prévision :

$$15083 \cdot \exp[(-1.0538) + (-3.951)] = 15083 \exp[-5.0048] \sim 101.127$$

La réponse attendue était ici **101.13**.

33. Page 18, point **23** : En 1998, on s'est aperçu que les primes de 1997 devait être surévaluée. Quel serait le montant total de provision à constituer si les primes acquises pour 1997 étaient 15000 ? [*un minimum d'explication sera exigé*]

Le montant total de provision est ici 9511.738, d'après la sortie. Le surplus de provision sera ici le 'surplus' de $Y_{1997,j}$. En effet, initialement, on avait prévu payer $P_{1997} \cdot \exp[\beta_0 + B_j]$, mais si la prime change, on devrait prévoir payer $(P_{1997} + \Delta P) \cdot \exp[B_j]$. Aussi, la variation de provision est

$$\Delta R = \sum_{j=1}^n \Delta P \cdot \exp[\beta_0 + B_j] = \underbrace{\Delta P \cdot \exp[\beta_0]}_{[15000-11217] \cdot 0.3485884} \cdot \sum_{j=1}^n \exp[B_j]$$

La somme commence en 1 car on ne prévoit que les valeurs futures, i.e. pour $j = 1, 2, \dots$. C'est un peu pénible, j'en conviens, mais il va falloir sommer les sommes des exponentielles de tous les coefficients. On obtient pour la somme 1.104977. La variation de réserves est alors

$$3783 \cdot 0.3485884 \cdot 1.104977 \sim 1457.142$$

Le nouveau montant de réserve est alors 9511.738+1457.142. La réponse attendue était **10968.9**. On peut le vérifier avec l'ordinateur

```
> df2=df
> df2$P[df2$A==1997]=15000
> sum(exp(predict(reg6,newdata=df2)[futur]))
[1] 10968.88
```

La question était un peu longue, j'en conviens..

34. Page 19, point **24** : Dans ce modèle, on suppose que la charge totale ne peut dépasser la prime acquise. En moyenne, quel ratio sinistre / prime observe-t-on, sur cette branche d'activité ? [*un minimum d'explication sera exigé*]

On attaque enfin la dernière question sur le provisionnement, avec la loi binomiale. La phrase écrite dans la question n'était pas tout à fait correcte, puisque l'on modélise les incréments : pour une année (et une année de survenance donnée), on ne peut pas payer plus que la prime encaissée pour cette année de survenance. La charge totale pourrait dépasser la prime acquise. Mais cela n'impacte pas la réponse à la question... On suppose ici que $Y_{i,j} \sim \mathcal{B}(P_i, \pi_{i,j})$ avec $\pi_{i,j}$ fonction de j seulement. Mais peu importe... On nous donne ici

$$\sum_{j \in \text{futur}} \widehat{Y}_{i,j} = 9511.738$$

Or la charge totale (apparaissant dans la notion de *sinistre*) est

$$\sum_j \widehat{Y}_{i,j} = \sum_{j \in \text{passé}} Y_{i,j} + \sum_{j \in \text{futur}} \widehat{Y}_{i,j}$$

La somme pour le passé, on l'a, page 12 : c'est la valeur Latest dans la sortie de Chain Ladder, soit 105159. Bref, la charge totale prévue de sinistres pour les 10 années observées est

$$105159 + 9511.738 \sim 114670.7$$

Le montant total de prime touché, on le connaît. Certes, il faut faire une somme, mais on peut l'obtenir facilement, à savoir 156276. Aussi, le ratio sinistre sur prime est de l'ordre de

$$\frac{114670}{156276} \sim 0.7337$$

La réponse attendue était ici **73.4%**.

4. MÉTHODES MORTALITÉ PROSPECTIVE

Pour les questions qui suivent, on admettra que si $E=1$, la prévision obtenu s'interprètera comme un taux de décès, noté $p_x^{(t)}$, au sens où $D_{x,t} = E_{x,t} \cdot p_x^{(t)}$, où x correspond à l'âge, et t l'année. $p_x^{(t)}$ sera alors aussi appelé probabilité de décès à l'âge x , l'année t .

35. Page 20, point **25** : En 2010, quelle proportion des personnes de 108 ans n'auront pas célébré leur 109ème anniversaire ? [*aucune justification ne sera exigé*]

Pour 2010, on regarde la première sortie de la page 20 : 20 personnes étaient en vie en début de période (variable E) à 108 ans, et 15 sont décédés (variable D). La proportion de décès est alors de 15/20 soit **75%**.

36. Page 20, points **26** : Entre 1950 et 2010, par combien a été divisée la probabilité de décès à 20 ans ? [*un minimum d'explication sera exigé*]

On utilise le raisonnement précédant pour calculer le taux de décès pour les deux années demandées

- en 1950, 343 décès pour 220766 personnes en vie, soit un taux de 0.001553
- en 2010, 238 décès pour 466476 personnes en vie, soit un taux de 0.000510

Aussi, le taux de décès a été divisé par 3.045 en 60 ans. La réponse attendue était **3**.

37. Page 20, points **27** : Entre 1950 et 2010, par combien a été divisée la probabilité de décès à 60 ans ? [*un minimum d'explication sera exigé*]

C'est la mme chose qu'auparavant. Le but de la question est de savoir si le amélioration des conditions de vie (qui se traduisent sur les probabilités de décès) sont les mmes à tous les ges, ou pas

- en 1950, 1974 décès pour 105298 personnes en vie, soit un taux de 0.018747
- en 2010, 2740 décès pour 409177 personnes en vie, soit un taux de 0.006696

Aussi, le taux de décès a été divisé par 2.7995 en 60 ans. La réponse attendue était **2.8**. Bref, le gain est (sensiblement) du mme ordre de grandeur.

38. Pages 20-21, points **28** : Afin de parler d'amélioration des conditions de vie, on souhaite écrire $p_x^{(t+1)} = (1 - u) \cdot p_x^{(t)}$, où u sera interprété comme un gain annuel. A l'aide de la régression `reg1`, donnez un estimateur de u . [*un minimum d'explication sera exigé*]

On a ici une régression de Poisson. Aussi, $D_{x,t} \sim P(E_{x,t} \cdot \lambda_{x,t})$, de telle sorte que

$$p_x^{(t)} = \lambda_{x,t} = \exp[\beta_0 + A_x + \beta_1 \cdot t],$$

l'année t - noté **Y** dans la régression - est ici prise en tant que variable continue, alors que l'âge x - noté **A** - est pris en tant que facteur. Aussi,

$$p_x^{(t+1)} = \exp[\beta_0 + A_x + \beta_1 \cdot (t + 1)] = \exp[\underbrace{\beta_0 + A_x + \beta_1 \cdot t}_{p_x^{(t)}} + \beta_1] = p_x^{(t)} \cdot \exp[\beta_1]$$

et donc, le ratio entre $p_x^{(t+1)}$ et $p_x^{(t)}$ vaut

$$\exp[\beta_1] = \exp[-1.544 \cdot 10^{-2}] = \exp[-0.01543] \sim 0.9846884$$

Aussi,

$$p_x^{(t+1)} = (1 - u) \cdot p_x^{(t)} \text{ avec } u \sim 0.01531$$

La réponse attendu était **1.53%** de gain annuel sur les taux de mortalité.

39. Pages 20-21, points **28** : Avec la régression `reg1`, par combien a été divisée la probabilité de décès à 60 ans entre 1950 et 2010 ? [*un minimum d'explication sera exigé*]

On va utiliser la question précédente, en notant que

$$p_x^{(2010)} = \exp[\beta_0 + A_x + \beta_1 \cdot (1950 + 60)] = \exp[\underbrace{\beta_0 + A_x + \beta_1 \cdot 1950}_{p_x^{(50)}} + \beta_1 \cdot 60] = p_x^{(50)} \cdot (\exp[\beta_1])^{60}$$

Aussi, pour avoir le coefficient multiplicatif, il suffit de calculer $(\exp[\beta_1])^{60}$,

$$(\exp[-0.01543])^{60} \sim 0.9846884^{60} \sim 0.3962143 = 2.5238^{-1}$$

de telle sorte que la probabilité de décès (à n'importe quel ge ici) a été divisé par **2.52** (ce qui reste sensiblement comparable aux valeurs obtenues dans les questions précédantes).

40. Page 21, point **29** : Compte tenu de l'évolution du taux de décès par âge (modèle **reg1**) on tente une régression linéaire par morceaux, continue. Quel nombre de décès prévoit-on avec le modèle **reg2** pour les personnes de 20 ans, en 1980 ? Quel nombre de décès prévoit-on avec le modèle **reg2** pour les personnes de 60 ans, en 1980 ? [*un minimum d'explication sera exigé*]

Pour faire notre prévision, il faut bien comprendre comment s'écrit le modèle. En l'occurrence, avec nos splines linéaires, on a

$$p_x^{(t)} = \exp[\beta_0 + \beta_1 x + \beta_2(k_2 - x)_+ + \beta_3(k_3 - x)_+ + \beta_4(k_4 - x)_+ + \beta_5 \cdot t]$$

avec ici des ruptures k_i en 12, 20 et 30, respectivement. Bref, pour répondre à la première partie de la question, pour un ge de 20 (valeur de x), en 1980 (valeur de t)

$$\hat{p}_{20}^{(1980)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 20 + \hat{\beta}_2 \underbrace{(12 - 20)_+}_{=0} + \hat{\beta}_3 \underbrace{(20 - 20)_+}_{=0} + \hat{\beta}_4 \underbrace{(30 - 20)_+}_{=10} + \hat{\beta}_5 \cdot 1980]$$

ce qui donne, numériquement

$$\hat{p}_{20}^{(1980)} = \exp[20.25 + 0.0932 \cdot 20 + 0.1631 \cdot \underbrace{(30 - 20)_+}_{=10} + (-0.01531) \cdot 1980] \sim 0.0014$$

soit **0.1%** de taux de décès, à 20 ans, en 1980. Pour une personne de 60 ans (valeur de x), en 1980 (valeur de t)

$$\hat{p}_{60}^{(1980)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 60 + \hat{\beta}_2 \underbrace{(12 - 60)_+}_{=0} + \hat{\beta}_3 \underbrace{(20 - 60)_+}_{=0} + \hat{\beta}_4 \underbrace{(30 - 60)_+}_{=0} + \hat{\beta}_5 \cdot 1980]$$

ce qui donne, numériquement

$$\hat{p}_{60}^{(1980)} = \exp[20.25 + 0.0932 \cdot 60 + (-0.01531) \cdot 1980] \sim 0.0114$$

soit **1.1%** de taux de décès, à 60 ans, en 1980.

41. Page 22, points **29** : A l'aide du modèle **reg2**, quels seront les taux de décès

- en 2015 pour une personne de 90 ans ?
- en 2016 pour une personne de 91 ans ?
- en 2017 pour une personne de 92 ans ?

- en 2018 pour une personne de 93 ans ?
- en 2019 pour une personne de 94 ans ?

Quelle sera finalement la probabilité pour qu'une personne qui a 90 ans en 2015 vive 5 ans de plus ? [*un minimum d'explication sera exigé*]

On attaque ici la partie amusante... car si on veut modéliser correctement les probabilités de décès, dans l'avenir, il faut corriger par les gains sur la mortalité. Bref, on va utiliser ce qu'on a vu dans la question précédente, les probabilités de décès sont

- en 2015 pour une personne de 90 ans : $\hat{p}_{90}^{(2015)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 \underline{90} + 0 + \hat{\beta}_5 \cdot 201\underline{5}] \sim 0.11053$
- en 2016 pour une personne de 91 ans : $\hat{p}_{91}^{(2016)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 \underline{91} + 0 + \hat{\beta}_5 \cdot 201\underline{6}] \sim 0.11948$
- en 2017 pour une personne de 92 ans : $\hat{p}_{92}^{(2017)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 \underline{92} + 0 + \hat{\beta}_5 \cdot 201\underline{7}] \sim 0.12916$
- en 2018 pour une personne de 93 ans : $\hat{p}_{93}^{(2018)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 \underline{93} + 0 + \hat{\beta}_5 \cdot 201\underline{8}] \sim 0.13963$
- en 2019 pour une personne de 94 ans : $\hat{p}_{94}^{(2019)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 \underline{94} + 0 + \hat{\beta}_5 \cdot 201\underline{9}] \sim 0.16317$

On a ici les probabilités de décès, pour avoir les probabilités de survies, on utilise

$$\prod_{h=0}^4 (1 - \hat{p}_{90+h}^{(2015+h)})$$

car l'âge, comme l'année, varie. Ici, la probabilité de survie est

$$0.8894672 \cdot 0.8805135 \cdot 0.8708345 \cdot 0.8603715 \cdot 0.8490609 \cdot 0.8368341 \sim 0.4169$$

Aussi, la probabilité de survie est de l'ordre de **41.7%**

42. Page 23, points **30** : A l'aide du modèle reg3, quels seront les taux de décès

- en 2015 pour une personne de 90 ans ?
- en 2016 pour une personne de 91 ans ?
- en 2017 pour une personne de 92 ans ?
- en 2018 pour une personne de 93 ans ?
- en 2019 pour une personne de 94 ans ?

Quelle sera finalement la probabilité pour qu'une personne qui a 90 ans en 2015 vive 5 ans de plus ? [*un minimum d'explication sera exigé*]

Cette fois, on utilise une régression binomiale, de telle sorte que $D_{x,t} \sim \mathcal{B}(E_{x,t}, p_{x,t})$ où

$$p_{x,t} = \frac{\exp[\beta_0 + \beta_1 x + \beta_2(k_2 - x)_+ + \beta_3(k_3 - x)_+ + \beta_4(k_4 - x)_+ + \beta_5 \cdot t]}{1 + \exp[\beta_0 + \beta_1 x + \beta_2(k_2 - x)_+ + \beta_3(k_3 - x)_+ + \beta_4(k_4 - x)_+ + \beta_5 \cdot t]}$$

Pour simplifier un peu les notation, posons

$$p_{x,t} = \frac{\eta_{x,t}}{1 + \eta_{x,t}}$$

et ensuite, il faut se lancer

- en 2015 pour une personne de 90 ans : $\hat{\eta}_{90}^{(2015)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 \underline{90} + 0 + \hat{\beta}_5 \cdot 2015] \sim 0.12444$
- en 2016 pour une personne de 91 ans : $\hat{\eta}_{91}^{(2016)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 \underline{91} + 0 + \hat{\beta}_5 \cdot 2016] \sim 0.13499$
- en 2017 pour une personne de 92 ans : $\hat{\eta}_{92}^{(2017)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 \underline{92} + 0 + \hat{\beta}_5 \cdot 2017] \sim 0.14642$
- en 2018 pour une personne de 93 ans : $\hat{\eta}_{93}^{(2018)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 \underline{93} + 0 + \hat{\beta}_5 \cdot 2018] \sim 0.15883$
- en 2019 pour une personne de 94 ans : $\hat{\eta}_{94}^{(2019)} = \exp[\hat{\beta}_0 + \hat{\beta}_1 \underline{94} + 0 + \hat{\beta}_5 \cdot 2019] \sim 0.17228$

avec les valeurs de la sortie page 24. Pour les probabilités, on obtient alors

- en 2015 pour une personne de 90 ans : $\hat{p}_{90}^{(2015)} = \hat{\eta}_{90}^{(2015)} / [1 + \hat{\eta}_{90}^{(2015)}] \sim 0.12444$
- en 2016 pour une personne de 91 ans : $\hat{p}_{91}^{(2016)} = \hat{\eta}_{91}^{(2016)} / [1 + \hat{\eta}_{91}^{(2016)}] \sim 0.11893$
- en 2017 pour une personne de 92 ans : $\hat{p}_{92}^{(2017)} = \hat{\eta}_{92}^{(2017)} / [1 + \hat{\eta}_{92}^{(2017)}] \sim 0.12772$
- en 2018 pour une personne de 93 ans : $\hat{p}_{93}^{(2018)} = \hat{\eta}_{93}^{(2018)} / [1 + \hat{\eta}_{93}^{(2018)}] \sim 0.13705$
- en 2019 pour une personne de 94 ans : $\hat{p}_{94}^{(2019)} = \hat{\eta}_{94}^{(2019)} / [1 + \hat{\eta}_{94}^{(2019)}] \sim 0.14696$

Et comme dans la question précédente, pour avoir les probabilités de survies, on utilise

$$\prod_{h=0}^4 (1 - \hat{p}_{90+h}^{(2015+h)})$$

soit ici (pour tre en vie au bout de 5 ans, il faut avoir survécu à chaque année)

$$0.8893259 \cdot 0.8810659 \cdot 0.8722779 \cdot 0.8629416 \cdot 0.8530377 \sim 0.50312$$

Aussi, la probabilité de survie est de l'ordre de **50.3%** qui est sensiblement différente de celle obtenue auparavant ! (20% plus élevée...). Intéressant, non ?