

**Komisja Egzaminacyjna dla Aktuariuszy**

**LXXXVIII Egzamin dla Aktuariuszy**

**Sesja egzaminacyjna w dniu 13 czerwca 2023 r.**

**Modelowanie**

**Imię i nazwisko osoby egzaminowanej: .....**

**Czas trwania egzaminu: 120 minut**

## Uwagi

- a) W prezentowanych wynikach separatorem dziesiętnym (znakiem dziesiętnym) jest kropka „.”.
- b) W prezentowanych wynikach oszacowań modeli:
- Residual deviance i Resid. Dev – oznacza dewiancję oszacowanego modelu,
  - Null deviance – oznacza dewiancję modelu zerowego,
  - Deviance – redukcję dewiancji po dodaniu kolejnej zmiennej objaśniającej,
  - Df – stopnie swobody,
  - Sum Sq – suma kwadratów,
  - 'log Lik.' – logarytm wiarygodności.
- c) Wartości  $\chi^2_{\alpha;v}$  rozkładu chi-kwadrat spełniające warunek  $P(\chi^2 \geq \chi^2_{\alpha;v}) = \alpha$ .

$v \backslash \alpha$	0.99	0.95	0.90	0.48	0.49	0.50	0.10	0.05	0.01	0.0085	0.0086	0.0087
<b>1</b>	0.000	0.004	0.016	0.499	0.477	0.455	2.706	3.841	6.635	6.925	6.904	6.883
<b>2</b>	0.020	0.103	0.211	1.468	1.427	1.386	4.605	5.991	9.210	9.535	9.512	9.489
<b>3</b>	0.115	0.352	0.584	2.474	2.420	2.366	6.251	7.815	11.345	11.696	11.671	11.646
<b>4</b>	0.297	0.711	1.064	3.486	3.421	3.357	7.779	9.488	13.277	13.650	13.623	13.597
<b>5</b>	0.554	1.145	1.610	4.499	4.425	4.351	9.236	11.070	15.086	15.479	15.451	15.423
<b>6</b>	0.872	1.635	2.204	5.512	5.430	5.348	10.645	12.592	16.812	17.222	17.193	17.164
<b>7</b>	1.239	2.167	2.833	6.525	6.435	6.346	12.017	14.067	18.475	18.902	18.871	18.841
<b>8</b>	1.646	2.733	3.490	7.537	7.440	7.344	13.362	15.507	20.090	20.532	20.500	20.469
<b>9</b>	2.088	3.325	4.168	8.548	8.445	8.343	14.684	16.919	21.666	22.122	22.089	22.057
<b>10</b>	2.558	3.940	4.865	9.559	9.450	9.342	15.987	18.307	23.209	23.679	23.645	23.612
<b>11</b>	3.053	4.575	5.578	10.570	10.455	10.341	17.275	19.675	24.725	25.207	25.173	25.139
<b>12</b>	3.571	5.226	6.304	11.580	11.460	11.340	18.549	21.026	26.217	26.712	26.676	26.641
<b>13</b>	4.107	5.892	7.042	12.589	12.464	12.340	19.812	22.362	27.688	28.195	28.159	28.123
<b>14</b>	4.660	6.571	7.790	13.599	13.469	13.339	21.064	23.685	29.141	29.659	29.622	29.585
<b>15</b>	5.229	7.261	8.547	14.608	14.473	14.339	22.307	24.996	30.578	31.107	31.069	31.032
<b>16</b>	5.812	7.962	9.312	15.617	15.477	15.338	23.542	26.296	32.000	32.540	32.501	32.463
<b>17</b>	6.408	8.672	10.085	16.626	16.481	16.338	24.769	27.587	33.409	33.959	33.919	33.880
<b>18</b>	7.015	9.390	10.865	17.634	17.485	17.338	25.989	28.869	34.805	35.365	35.325	35.286
<b>19</b>	7.633	10.117	11.651	18.642	18.489	18.338	27.204	30.144	36.191	36.761	36.720	36.679
<b>20</b>	8.260	10.851	12.443	19.650	19.493	19.337	28.412	31.410	37.566	38.145	38.104	38.063

**Zadanie 1.**

Dla pewnego portfela ubezpieczeń badano zależność rocznej liczby szkód (zmienna *clm.count*) od wieku ubezpieczonego wyrażonego w latach (zmienna ilościowa *driver.age*) oraz płci (zmienna jakościowa *driver.gender*, przyjmująca dwie wartości: *Female*, *Male*). Oszacowano dwa modele regresji Poissona z kanonicznymi funkcjami łączącymi (linkami kanonicznymi). W obydwu modelach jako zmienną offsetową uwzględniono czas ekspozycji na ryzyko w latach (zmienna *exposure*). Uzyskano następujące wyniki:

**Model M1:**

Call:

```
glm(formula = clm.count ~ driver.age + driver.gender + offset(log(exposure)),
     family = poisson, data = zbior.uczacy)
```

Deviance Residuals:

```
   Min     1Q  Median     3Q      Max
-0.6960 -0.4767 -0.3828 -0.2566  4.8785
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.231453	0.107225	-11.485	< 2e-16 ***
driver.age	-0.009330	0.002082	-4.482	7.39e-06 ***
driver.genderMale	-0.189300	0.065907	-2.872	0.00408 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 9881.4 on 24455 degrees of freedom

Residual deviance: 9852.1 on 24453 degrees of freedom

'log Lik.' -6865.075

**Model M2** (zmienna *driver.age.kw = driver.age<sup>2</sup>*):

Call:

```
glm(formula = clm.count ~ driver.age + driver.age.kw + driver.gender +
     offset(log(exposure)), family = poisson, data = zbior.uczacy)
```

Deviance Residuals:

```
   Min     1Q  Median     3Q      Max
-0.9647 -0.4706 -0.3824 -0.2692  4.8862
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-3.647114	0.530283	-6.878	6.08e-12 ***
driver.age	0.018353	0.006276	2.924	0.00345 **
driver.age.kw	0.000493	0.000106	4.665	3.09e-06 ***
driver.genderMale	-0.187095	0.065926	-2.838	0.00454 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

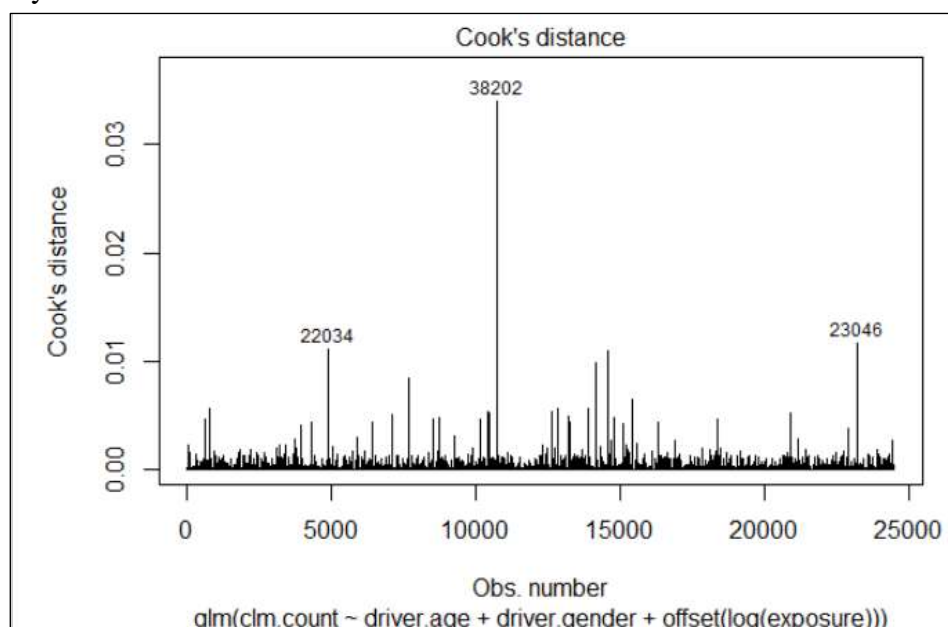
Null deviance: 9881.4 on 24455 degrees of freedom

Residual deviance: 9831.9 on 24452 degrees of freedom

'log Lik.' -6855

- a) (2p.) Wyjaśnij związek między wiarygodnością  $L$ , a kryterium informacyjnym AIC i wskaż kiedy każdy z tych mierników może być użyty do porównania różnych modeli. Który z oszacowanych modeli (tzn. M1, M2) jest lepszy? Wybór uzasadnij.
- b) (2p.) W zbiorze uczącym wykorzystanym do oszacowania obydwu modeli znajduje się 46-letnia kobieta z trzymiesięczną ekspozycją na ryzyko, w czasie której nie zgłosiła żadnej szkody. Wykorzystując model M1:
- oszacuj prawdopodobieństwo, że w ciągu jednego roku nie zgłosi ona żadnej szkody,
  - wyznacz resztę Pearsona odpowiadającą tej obserwacji.
- c) (1p.) Na rysunku 1.1 przedstawiono jeden z wykresów diagnostycznych dla modelu M1. Wyjaśnij w jakim celu wykorzystuje się tego typu wykresy. Czy dla modelu M2 uzyskamy identyczny? Odpowiedź uzasadnij.

Rys. 1.1



## Odpowiedzi

### Odp. a)

W odpowiedzi należało wskazać, że:

- $AIC = -2 \log$ arytm wiarygodności +  $2 \cdot$ liczba oszacowanych parametrów.
- wiarygodność można wykorzystać do porównywania modeli które posiadają taką samą liczbę parametrów lub są zagnieżdżone. Kryterium informacyjne AIC jest bardziej przydatne, gdy modele różnią się liczbą parametrów i są zbudowane na innym zestawie zmiennych objaśniających.
- model M2 jest lepszy.

---

.....  
**Odp. b)**

$$\hat{\mu} = \exp(-1.231453 - 0.009330 \cdot 46 - 0.189300 \cdot 0) = 0.190019$$

Prawdopodobieństwo: 0.8269433

$$\text{Reszta Pearsona: } r_p = \frac{y - \hat{y}}{\sqrt{\hat{y}}} = \frac{0 - 0.25 \cdot 0.190019}{\sqrt{0.25 \cdot 0.190019}} = -0.2179559$$

.....  
**Odp. c)**

Należało wskazać, że

- wykres „Cook's distance” jest wykorzystywany w analizie regresji jako miara wpływu poszczególnych obserwacji na wyniki regresji. Umożliwia wykrycie obserwacji, które znacząco wpływają na wyniki regresji, a tym samym pozwala zbadać ich wpływ na model.
- dla M2 uzyska się inny wykres, ponieważ w mierze Cooka uwzględnia się reszty modeli.

**Zadanie 2.**

- a) (1p.) Wyjaśnij w jaki sposób przeprowadza się  $k$ -krotną walidację krzyżową.
- b) (1p.) Podaj na czym polega walidacja za pomocą metody LOOCV (*Leave-one-out cross-validation*).
- c) (2p.) Jakie są zalety i wady  $k$ -krotnej walidacji krzyżowej w porównaniu z:
- podjęciem wykorzystującym jedynie jeden zbiór walidacyjny,
  - metodą LOOCV.

W odpowiedzi uwzględnij problem kompromisu między obciążeniem a wariancją modelu.

- d) (1p.) Oszacowano model regresji liniowej na podstawie 5-ciu obserwacji. Uzyskano następujące reszty: 1.78, -1.30, 1.09, -1.89, 0.32. Wiadomo, że w analizowanym przypadku macierz daszkowa jest równa:

$$H = \begin{bmatrix} 0.29 & 0.24 & -0.03 & 0.15 & 0.35 \\ & 0.22 & 0.09 & 0.17 & 0.27 \\ & & 0.80 & 0.34 & -0.20 \\ & & & 0.23 & 0.11 \\ & & & & 0.46 \end{bmatrix}$$

Walidację tego modelu przeprowadzono z wykorzystaniem błędu średniokwadratowego MSE (*mean squared error*) za pomocą metody LOOCV. Jaki otrzymano wynik?

**Odpowiedzi****Odp. a)**

Zobacz podrozdział 5.1.3 w “An Introduction to Statistical Learning with Applications in R” - G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2021.

**Odp. b)**

Zobacz podrozdział 5.1.2 w “An Introduction to Statistical Learning with Applications in R” - G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2021.

**Odp. c)**

Zobacz podrozdział 5.1.4 w “An Introduction to Statistical Learning with Applications in R” - G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2021.

**Odp. d)**

$$CV_n = \frac{1}{n} \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{1 - h_i} \right)^2$$

$$CV_n = \frac{1}{5} \left( \left( \frac{1.78}{1 - 0.29} \right)^2 + \left( \frac{-1.30}{1 - 0.22} \right)^2 + \left( \frac{1.09}{1 - 0.80} \right)^2 + \left( \frac{-1.89}{1 - 0.23} \right)^2 + \left( \frac{0.32}{1 - 0.46} \right)^2 \right) \\ = 9.028$$

**Zadanie 3.**

- a) (**2p.**) Krótko przedstaw ideę uogólnionych modeli addytywnych (*Generalized Additive Models* – GAM). Wskaż dlaczego weszły do zestawu narzędzi aktuarusza.
- b) (**1p.**) Podaj definicję funkcji sklejaney stopnia 3 (splajnu kubicznego).
- c) (**2p.**) Liczbę roszczeń (zmienna *clm.count*) w pewnym portfelu ubezpieczeń AC modelowano z uwzględnieniem następujących zmiennych objaśniających:  
*driver.gender* – płeć kierowcy (zmienna jakościowa: *Female*, *Male*),  
*driver.age* – wiek kierowcy (zmienna ilościowa),  
*vehicle.age* - wiek samochodu (zmienna ilościowa),  
*vehicle.value* – wartość samochodu (zmienna ilościowa),  
*hp* – moc silnika (zmienna ilościowa).

Oszacowano uogólniony model addytywny, w którym przyjęto rozkład Poissona dla liczby roszczeń oraz link logarytmiczny. Zinterpretuj uzyskane wyniki (podane poniżej). W interpretacji uwzględnij także wykresy przedstawione na rysunku 3.1.

Family: poisson  
Link function: log

Formuła:

$clm.count \sim driver.gender + s(driver.age) + s(vehicle.age) + s(vehicle.value, hp) + offset(exposure)$

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.86959	0.07887	-36.382	< 2e-16 ***
driver.genderMale	-0.23469	0.08422	-2.787	0.00533 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	Chi.sq	p-value
s(driver.age)	7.026	8.014	43.38	< 2e-16 ***
s(vehicle.age)	1.001	1.002	26.86	2.41e-07 ***
s(vehicle.value, hp)	3.320	4.299	17.34	0.00224 **

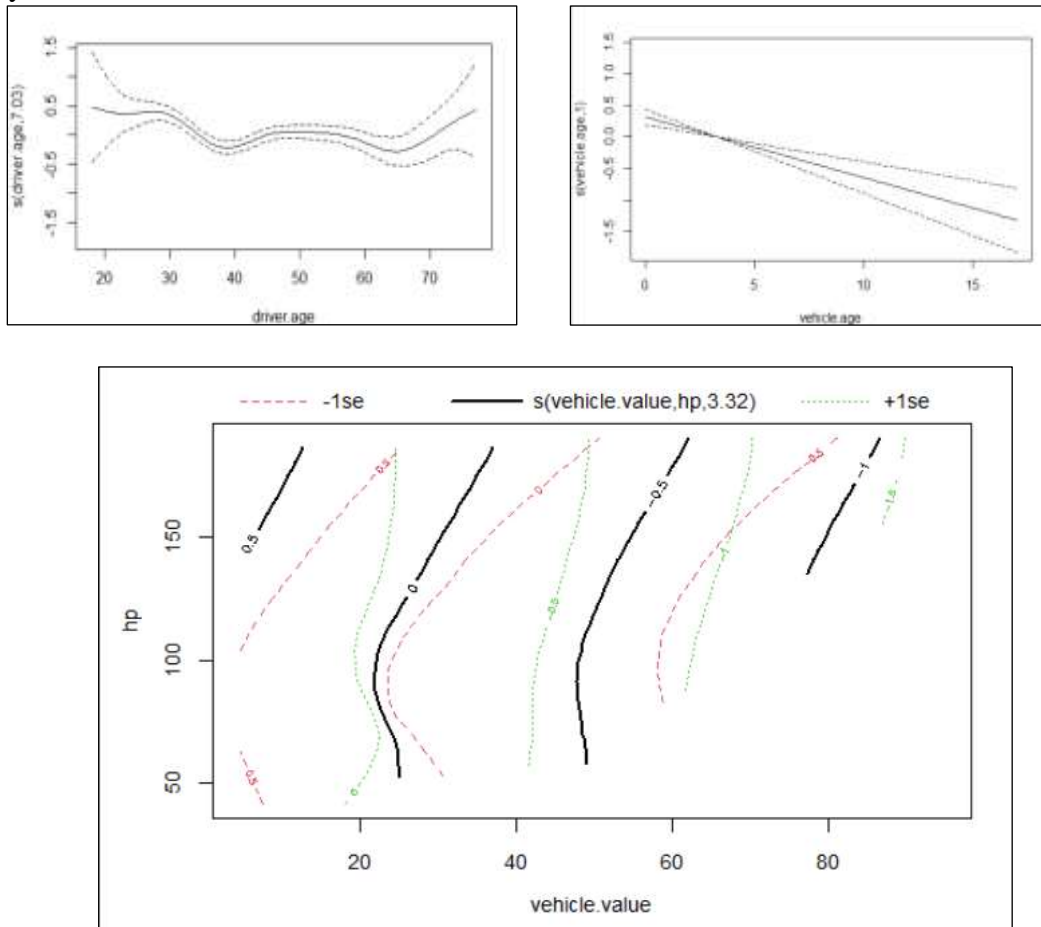
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.0264 Deviance explained = 1.66%

UBRE = -0.59697 Scale est. = 1 n = 14634

Rys. 3.1.



### Odpowiedzi

#### Odp. a)

Zobacz np.:

- Podrozdział 6.1 w: “Effective Statistical Learning Methods for Actuaries I” - M. Denuit, D. Hainaut, J. Trufin, Springer, 2019.
- Podrozdział 7.7 w “An Introduction to Statistical Learning with Applications in R” - G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2021.

#### Odp. b)

Zobacz np.:

- Podrozdział 6.3.2.2 w: “Effective Statistical Learning Methods for Actuaries I” - M. Denuit, D. Hainaut, J. Trufin, Springer, 2019.
- Podrozdział 7.4.3 w “An Introduction to Statistical Learning with Applications in R” - G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2021.



---

.....

**Odp. c)**

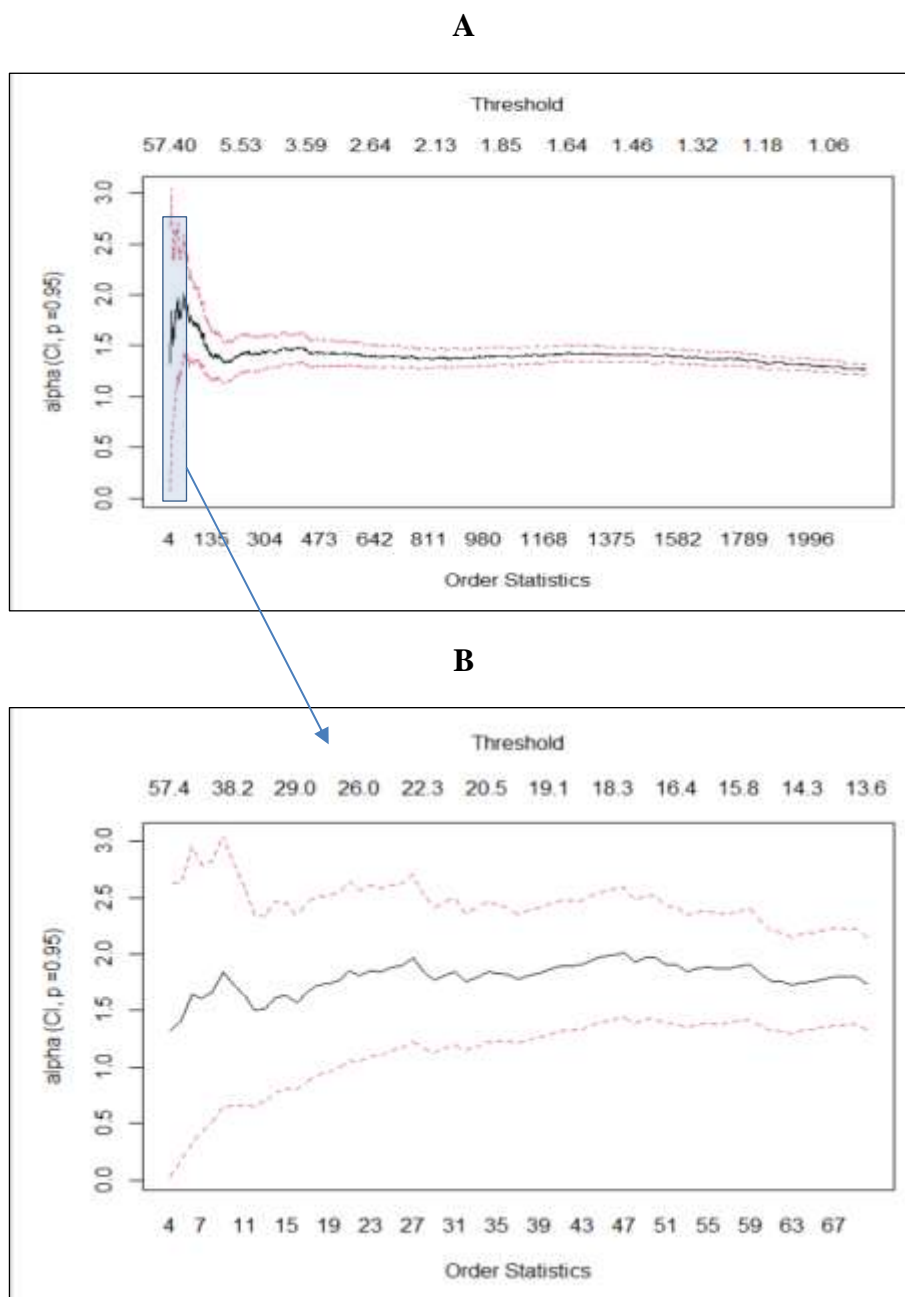
W odpowiedzi należało omówić: część liniową modelu (zwracając uwagę na istotność parametrów), część nieliniową (zwracając uwagę na istotność wpływu poszczególnych splajnów na przewidywania modelu) oraz zidentyfikować nieliniowe zależności między zmiennymi objaśniającymi a zmienną zależną (na podstawie wykresów).

Zobacz podrozdział 6.4.2.2 w: “Effective Statistical Learning Methods for Actuaries I”  
- M. Denuit, D. Hainaut, J. Trufin, Springer, 2019.

**Zadanie 4.**

- (2p.) Krótko opisz podejście Hilla do modelowania ogonów rozkładów (m.in. podaj założenia odnośnie rozkładów i przedstaw odpowiedni estymator).
- (1p.) Przedstaw konstrukcję wykresu Hilla (*Hill plot*) i wskaż w jakim celu jest wykorzystywany.
- (2p.) Analizowano straty pożarowe zarejestrowane przez Copenhagen Re. Na poniższym rysunku (Rys. 4.1.) przedstawiono skonstruowany na ich podstawie wykres Hilla (panel **B** przedstawia powiększony fragment zaznaczony na panelu **A**). Zinterpretuj otrzymane wyniki.

Rys. 4.1.



---

**Odpowiedzi**

.....

**Odp. a)**

Zobacz podrozdział 5.2.4 w: “Quantitative Risk Management: Concepts, Techniques and Tools”, revised edition - A. McNeil, R. Frey, P. Embrecht, Princeton, 2015

.....

**Odp. b)**

Zobacz podrozdział 5.2.4 w: “Quantitative Risk Management: Concepts, Techniques and Tools”, revised edition - A. McNeil, R. Frey, P. Embrecht, Princeton, 2015

.....

**Odp. c)**

W odpowiedzi należało wskazać, że przedstawione wykresy można wykorzystać do oceny indeksu ogona rozkładu  $\alpha$  (*tail index*). W przypadku analizowanych danych można przyjąć, że  $\alpha$  wynosi od 1,5 do 2, co sugeruje oszacowania dla  $\xi$  od 0,5 do 0,67, z których wszystkie odpowiadają rozkładom o nieskończonej wariancji dla strat pożarowych.

**Zadanie 5.**

- a) (2p.) Wymień etapy statystycznej analizy szeregów czasowych danych  $y_1, y_2, \dots, y_t$ . Krótko opisz jeden z nich.
- b) (2p.) Przedstaw sposób prognozowania szeregów czasowych za pomocą modeli ARMA. Podaj ogólne założenia i wskaż ideę.
- c) (1p.) Na podstawie szeregu czasowego liczącego 200 obserwacji oszacowano model ARMA(1,1). Uzyskano następujące wyniki:

Call:

arima(x = data, order = c(1, 0, 1), method = "ML")

Coefficients:

ar1	ma1	intercept
0.4039	0.5361	0.0393
s.e. 0.0788	0.0668	0.1866

sigma^2 estimated as 1.059: log likelihood = -290, aic = 588

Wartości rzeczywiste i oszacowane reszty  $\hat{\varepsilon}_t$  dla 3 ostatnich obserwacji przedstawia tabela 5.1.

Tab. 5.1

$t$	198	199	200
$x_t$	1.17510868	-0.11635671	0.06456704
$\hat{\varepsilon}_t$	-0.6482727	-0.2668962	0.2312339

Wyznacz prognozę dla tego szeregu czasowego na okres  $t = 202$ .

### Odpowiedzi

.....  
**Odp. a)**

Zobacz podrozdział 4.1.4 w: "Quantitative Risk Management: Concepts, Techniques and Tools", revised edition - A. McNeil, R. Frey, P. Embrecht, Princeton, 2015

.....  
**Odp. b)**

Zobacz podrozdział 4.1.4 w: "Quantitative Risk Management: Concepts, Techniques and Tools", revised edition - A. McNeil, R. Frey, P. Embrecht, Princeton, 2015

.....  
**Odp. c)**

Prognoza na okres  $t = 201$ :

$$x_{201}^P = 0.0393 + 0.4039 \cdot (0.06456704 - 0.0393) + 0.5361 \cdot 0.2312339 \\ = 0.1734699$$

Prognoza na okres  $t = 202$ :

$$x_{202}^P = 0.0393 + 0.4039 \cdot (0.1734699 - 0.0393) = 0.0934912$$

**Zadanie 6.**

- a) (2p.) Podaj definicję danych prawostronnie cenzurowanych (*right censoring*). Wskaż i omów co najmniej dwie sytuacje, w których aktuariusz analizuje tego typu dane.
- b) Wykorzystując dane zwarte w tabeli 6.1, gdzie symbolem (\*) oznaczono obserwacje cenzurowane z góry:
- (2p.) Skonstruuj estymator Kaplana–Meiera dla funkcji przeżycia  $S(x)$ .
  - (1p.) Oszacuj wariancję estymatora Kaplana–Meiera dla  $S(2)$ .

Tab. 6.1

1	2	3*	4	4	4*	4*	5	7*	8	8	8	8	9	9	9	9	10*	12	12	15*
---	---	----	---	---	----	----	---	----	---	---	---	---	---	---	---	---	-----	----	----	-----

**Odpowiedzi:****Odp. a)**

Zobacz podrozdział 14.3 w: “Loss Models: From Data to Decisions”, 5th edition - S.A. Klugman, H.H Panjer, G.E. Willmot, Wiley, 2019.

**Odp. b)**

Ad. i

$i$	$y_i$	$s_i$	$r_i$	$\hat{S}_n(y_i)$
1	1	1	20	$1 - \frac{1}{20} = 0.950$
2	2	1	19	$0.950 \cdot \left(1 - \frac{1}{19}\right) = 0.900$
3	4	2	17	$0.900 \cdot \left(1 - \frac{2}{17}\right) = 0.794$
4	5	1	13	$0.794 \cdot \left(1 - \frac{1}{13}\right) = 0.733$
5	8	3	11	$0.733 \cdot \left(1 - \frac{3}{11}\right) = 0.533$
6	9	4	8	$0.533 \cdot \left(1 - \frac{4}{8}\right) = 0.267$
7	12	2	3	$0.267 \cdot \left(1 - \frac{2}{3}\right) = 0.089$

Ad. ii

$$\widehat{Var}(S_{20}(2)) = 0.900^2 \cdot \left(\frac{1}{20 \cdot 19} + \frac{1}{19 \cdot 18}\right) = 0.0045$$

Szczegóły Ad. i. oraz Ad. ii. w podrozdziale 14.3 w: “Loss Models: From Data to Decisions”, 5th edition - S.A. Klugman, H.H Panjer, G.E. Willmot, Wiley, 2019.

**Zadanie 7.**

Przedstaw wytyczne Krajowego Standardu Aktuarnego w zakresie stosowania modeli (tj. wyboru, tworzenia, modyfikowania i przeliczania modeli) dotyczące:

- a) (1p.) ryzyka modelu,
- b) (2p.) walidacji modeli,
- c) (2p.) wykorzystania wyników przebiegu modelu.

**Odpowiedzi:**

.....  
**Odp. a)**

Zobacz podrozdział 2.10 w: „Krajowy Standard Aktuarny Polskiego Stowarzyszenia Aktuariuszy – Praktyka Aktuarna”, 2022,

.....  
**Odp. b)**

Zobacz podrozdział 2.10 w: „Krajowy Standard Aktuarny Polskiego Stowarzyszenia Aktuariuszy – Praktyka Aktuarna”, 2022,

.....  
**Odp. c)**

Zobacz podrozdział 2.10 w: „Krajowy Standard Aktuarny Polskiego Stowarzyszenia Aktuariuszy – Praktyka Aktuarna”, 2022,

**Zadanie 8.**

a) (3p.) Przedstaw ideę i sposób konstrukcji wykresów PDP (*Partial Dependence Plot*).

b) (2p.) Liczbę roszczeń  $K_i$  w pewnym portfelu ubezpieczeń AC modelowano z uwzględnieniem następujących zmiennych objaśniających:

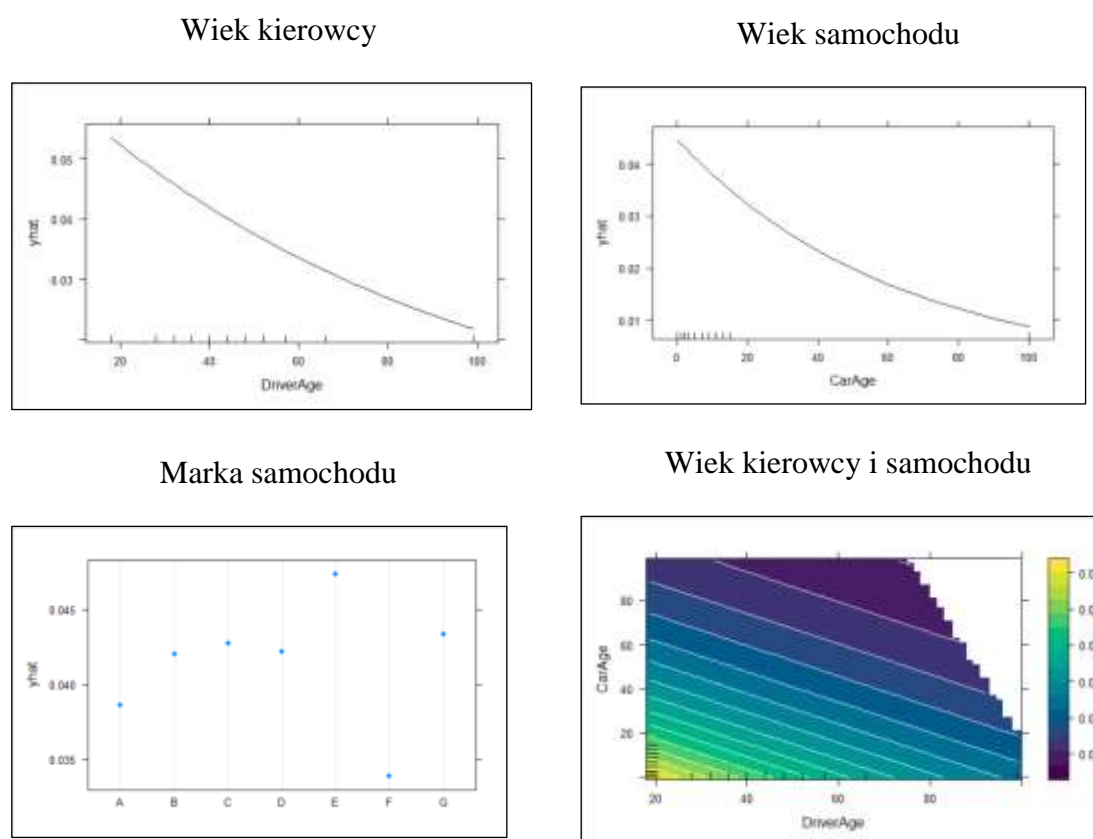
*DriverAge* – wiek kierowcy (w latach),

*CarAge* - wiek samochodu (w latach),

*Brand* – marka samochodu. Zmienna jakościowa przyjmująca następujące kategorie: *A, B, C, D, E, F* i *G*.

Oszacowano uogólniony model liniowy, w którym przyjęto rozkład Poissona dla  $K_i$  oraz link kanoniczny. Dla tego modelu skonstruowano wykresy PDP przedstawione na rysunku 8.1. Podaj interpretację tych wykresów.

Rys. 8.1

**Odpowiedzi:**

.....  
**Odp. a)**

Zobacz podrozdział 4.6.2 w: “Effective Statistical Learning Methods for Actuaries II” - M. Denuit, D. Hainaut, J. Trufin, Springer, 2019.

---

.....

**Odp. b)**

W odpowiedzi należało wskazać:

- czy zmienne *DriverAge*, *CarAge* oraz *Brand* mają wpływ na prognozowaną liczbę roszczeń (na wynik modelu),
- czy zależność między określoną zmienną ilościową a prognozowaną liczbą roszczeń jest liniowa, nieliniowa,
- czy istnieją interakcje między zmiennymi.



**Zadanie 9.**

- a) (3p.) Przedstaw ideę i konstrukcję testu ilorazu wiarygodności. Zapisz hipotezę zerową i alternatywną i wskaż czy różnią się one od hipotez (zerowej i alternatywnej) stawianych w testach zgodności (np. chi-kwadrat, Kołmogorowa-Smirnowa). Podaj postać statystyki testowej i jej rozkład.
- b) (2p.) Wiadomo, że wysokość szkód w pewnym portfelu ubezpieczeń ma rozkład Pareto z parametrem  $\alpha = 2$  i nieznanym parametrem  $\theta$ . Z portfela wylosowano 20 szkód i oszacowano  $\theta$  metodą największej wiarygodności, uzyskując wartość 7.0 ( $\hat{\theta} = 7.0$ ). Następnie z wykorzystaniem testu ilorazu wiarygodności testowano hipotezę zerową  $H_0: \theta = 3.1$ . Wyznacz prawdopodobieństwo testowe (*p-Value*) dla tego testu.

Uwaga!  $\sum_{i=1}^{20} \ln(x_i + 7.0) = 49.01$ ;  $\sum_{i=1}^{20} \ln(x_i + 3.1) = 39.30$

Funkcja gęstości rozkładu Pareto ma postać:  $f(x) = \frac{\alpha \theta^\alpha}{(x+\theta)^{\alpha+1}}$ .

**Odpowiedzi:****Odp. a)**

Zobacz podrozdział 15.4.4 w: "Loss Models: From Data to Decisions", 5th edition - S.A. Klugman, H.H Panjer, G.E. Willmot, Wiley, 2019.

**Odp. b)**

Funkcja wiarygodności wynosi:

$$L(\alpha, \theta; x_i) = \frac{\alpha^{20} \theta^{20\alpha}}{\prod_{i=1}^{20} (x_i + \theta)^{\alpha+1}},$$

stąd logarytm wiarygodności jest równy:

$$l(\alpha, \theta; x_i) = 20 \ln(\alpha) + 20\alpha \ln(\theta) - (\alpha + 1) \sum_{i=1}^{20} \ln(x_i + \theta)$$

Wartość statystyki testowej:

$$T = 2(l^{(H_1)} - l^{(H_0)}) = 2 \cdot (-55.3307 - (-58.7810)) = 6.901,$$

gdzie  $l^{(H_0)}$ ,  $l^{(H_1)}$  - oznacza logarytm wiarygodności przy założeniu odpowiednio hipotezy zerowej i alternatywnej.

Statystyka  $T$  ma rozkład Chi-kwadrat z jednym stopniem swobody, zatem *p-Value* (odczytane z tabeli zamieszczonej na str.2 ) jest równe 0.0086.

**Zadanie 10.**

- a) (2p.) Wskaż co najmniej cztery reguły określające, kiedy węzeł w drzewie regresyjnym jest przyjmowany za końcowy (jest uznawany za liść).
- b) (1p.) Na czym polega i w jakim celu stosuje się przycinanie drzewa regresyjnego?
- c) (2p.) Liczbę roszczeń  $K_i$  w pewnym portfelu ubezpieczeń AC modelowano z uwzględnieniem następujących zmiennych objaśniających:

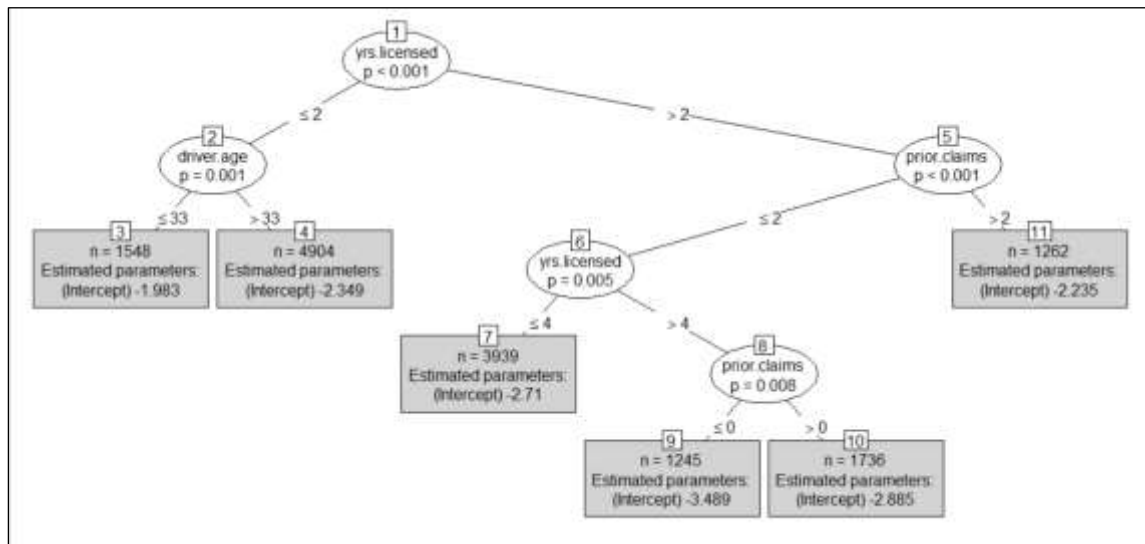
*driver.age* – wiek kierowcy w latach (zmienna ilościowa),

*prior.claims* – liczba wcześniej zgłoszonych roszczeń (zmienna ilościowa),

*yrs.licensed* – okres posiadania prawa jazdy w latach (zmienna ilościowa).

Przyjęto dla  $K_i$  rozkład Poissona i skonstruowano binarne drzewo GLM (*Generalized Linear Model Tree*) przedstawione na rysunku 10.1. Dla liści podano oszacowania modeli regresji Poissona z linkiem kanonicznym. Opisz grupę kierowców, która średnio rocznie zgłasza najwięcej szkód i grupę, która średnio rocznie zgłasza najmniej szkód. Oszacuj dla tych grup prawdopodobieństwa wystąpienia co najmniej jednego roszczenia.

Rys. 10.1

**Odpowiedzi:****Odp. a)**

Zobacz podrozdział 3.2.3 w: “Effective Statistical Learning Methods for Actuaries II” - M. Denuit, D. Hainaut, J. Trufin, Springer, 2019.

**Odp. b)**

Zobacz podrozdział 3.3 (wprowadzenie) w: “Effective Statistical Learning Methods for Actuaries II” - M. Denuit, D. Hainaut, J. Trufin, Springer, 2019.

**Odp. c)**

Grupa kierowców, która średnio rocznie zgłasza:

- **najwięcej szkód:** posiadają prawo jazdy nie dłużej niż 2 lata i mają nie więcej niż 33 lata (niezależnie od liczby zgłoszonych wcześniej szkód).

Dla tej grupy  $\lambda = \exp(-1,983) = 0,13765565$ , stąd prawdopodobieństwa wystąpienia co najmniej jednego roszczenia wynosi: **0,128601294**.

- **najmniej szkód:** posiadają prawo jazdy powyżej cztery lata i wcześniej nie zgłosili żadnej szkody (niezależnie od wieku).

Dla tej grupy  $\lambda = \exp(-3,489) = 0,030531388$ , stąd prawdopodobieństwa wystąpienia co najmniej jednego roszczenia wynosi: **0,030070013**.

---

**Sesja egzaminacyjna w dniu 13 czerwca 2023 r.****Modelowanie****Arkusz ocen**

Zadanie nr	Punktacja
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	