Komisja Egzaminacyjna dla Aktuariuszy

LXXVII Egzamin dla Aktuariuszy

Sesja egzaminacyjna w dniu 21 listopada 2017 r.

Modelowanie

Imię i nazwisko osoby egzaminowanej: ***Rozwiązania wersja A***

Czas trwania egzaminu: 120 minut

Uwagi

a) W zadaniach wartość zagrożona na poziomie ufności α jest definiowana jako kwantyl rzędu α rozkładu odpowiedniej zmiennej losowej, tzn.

$$VaR_\alpha(X) = \inf\{x: F_X(x) \geq \alpha\}\ .$$

b) W zadaniach zastosowano następujące oznaczenia:
$E(X)$ – wartość oczekiwana
$D(X)$ – odchylenie standardowe

c) Dystrybuanta standardowego rozkładu normalnego $\Phi(u) = P(U \leq u)$ dla $0 \leq u \leq 0{,}99$)

u	**0**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389

d) Wartości $\chi^2_{\alpha;\upsilon}$ rozkładu chi-kwadrat spełniające warunek $P(\chi^2 \geq \chi^2_{\alpha;\upsilon}) = \alpha$

υ \ α	**0,995**	**0,99**	**0,975**	**0,95**	**0,9**	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**
1	0,000	0,000	0,001	0,004	0,016	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000	34,267

Zadanie 1.

Straty dwóch rodzajów ryzyka są modelowane za pomocą zmiennych losowych X_1 i X_2 o takim samym rozkładzie Pareto, którego funkcja gęstości ma postać:

$$f(x) = \frac{1}{x^2}, x > 1.$$

a) Zakładając, że zmienne X_1 i X_2 są niezależne wyznaczyć prawdopodobieństwo $P(X_1 + X_2 \leq 43{,}451)$ (wynik zaokrąglić do dwóch miejsc po przecinku).

b) Zakładając, że zmienne X_1 i X_2 są niezależne, wyznaczyć wartość zagrożoną na poziomie ufności $\alpha = 0{,}95$ dla łącznych strat $Y = X_1 + X_2$ (czyli $VaR_{0,95}(Y)$).

c) Zakładając, że zmienne X_1 i X_2 są współmonotoniczne (*comonotonic*), wyznaczyć VaR na poziomie ufności $\alpha = 0{,}95$ dla $Y^C = X_1^C + X_2^C$ (tzn. dla łącznych strat, ale przy założeniu współmonotonicznej struktury zależności między X_1 i X_2, czyli $VaR_{0,95}(Y^C)$).

d) Porównać i skomentować uzyskane wyniki dla $VaR_{0,95}(Y)$ i $VaR_{0,95}(Y^C)$

Odpowiedzi:

Odp. a)

$P(X_1 + X_2 \leq 43{,}451)=0{,}95$

....

Odp. b)

$VaR_{0,95}(Y) = 43{,}451$

....

Odp. c)

$VaR_{0,95}(Y^C) = 40$

....

Odp. d)

$VaR_{0,95}(X_1 + X_2) > VaR_{0,95}(X_1) + VaR_{0,95}(X_2).$

VaR nie jest koherentną miarą ryzyka, nie spełnia warunku subaddytywności.

Zadanie 2.
Sformułować twierdzenie Sklara i wyznaczyć kopulę dla dwuwymiarowego logistycznego rozkładu Gumbela: $F(x,y) = (1 + e^{-x} + e^{-y})^{-1}, \ \ x, y \in \mathbb{R}$.

Odpowiedzi:

Twierdzenie Sklara:

1. Niech C będzie k-wymiarową kopulą a $F_1, ..., F_k$ dystrybuantami jednowymiarowych rozkładów prawdopodobieństwa. Wtedy, dla $\mathbf{x} = (x_1, ..., x_k) \in \mathrm{R}^k$,
$$F(x_1, ..., x_k) = C(F_1(x_1), ..., F_k(x_k)) \qquad (*)$$
jest dystrybuantą rozkładu wielowymiarowego, dla którego $F_1, ..., F_k$ są dystrybuantami rozkładów brzegowych.
2. Odwrotnie, dla dowolnej dystrybuanty F łącznego wielowymiarowego rozkładu o dystrybuantach brzegowych $F_1, ..., F_k$ istnieje kopula spełniająca warunek (*). Ponadto w przypadku, gdy $F_1, ..., F_k$ są ciągłe, istnieje tylko jedna taka kopula dana wzorem:
$$C(u_1, ..., u_k) = F(F_1^{-1}(u_1), ..., F_k^{-1}(u_k)), \ (u_1, ..., u_k) \in [0,1]^k .$$

....

Szukana kopula:

$$C(u,v) = \frac{uv}{u + v - uv}$$

Zadanie 3.

Niech C_θ będzie kopulą FGM (kopula Farlie-Gumbela-Morgensterna) z parametrem θ: $C_\theta(u,v) = uv(1+\theta(1-u)(1-v))$, gdzie $-1 \leq \theta \leq 1$. Wiadomo, że straty dla dwóch rodzajów ryzyka X_1 i X_2 są modelowane za pomocą rozkładu dwuwymiarowego z kopulą FGM z parametrem $\theta = 0{,}5$ i brzegowymi o takim samym rozkładzie wykładniczym z parametrem $\lambda = 0{,}25$.

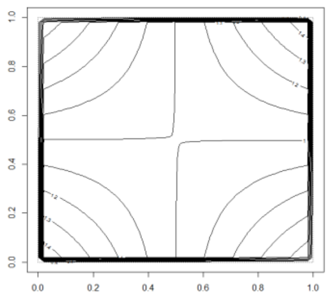
a) Wyznaczyć funkcję gęstości dla tej kopuli (tzn. FGM z parametrem $\theta = 0{,}5$) i naszkicować jej wykres konturowy.
b) Wyznaczyć prawdopodobieństwo, że straty z tych rodzajów ryzyka będą następujące: $E(X_1) < X_1 < E(X_1) + D(X_1)$ i $E(X_2) < X_2 < E(X_2) + D(X_2)$)
c) Wyznaczyć prawdopodobieństwo, że $X_1 > E(X_1) + D(X_1)$ lub $X_2 > E(X_2) + D(X_2)$.

Odpowiedzi:

Odp. a)

Funkcja gęstości: $c(u,v) = 1 + 0{,}5(1-2u)(1-2v)$

Wykres konturowy:



....

Odp. b) $E(X_1) = E(X_2) = 4$, $D(X_1) = D(X_2) = 4$

$$P(4 < X_1 < 8 \; i \; 4 < X_2 < 8) = 0{,}060746$$

....

Odp. c)

$$P(X_1 > 8 \; lub \; X_2 > 8) = 0{,}245508$$

Zadanie 4.

Wiadomo, że wymogi kapitałowe dla trzech rodzajów ryzyka (trzech linii biznesu) wynoszą odpowiednio: $\kappa(X_1) = 100$, $\kappa(X_2) = 80$ i $\kappa(X_3) = 20$. Do agregacji ryzyka zastosowano metodę wariancji-kowariancji. Współczynniki korelacji liniowej między poszczególnymi rodzajami ryzyka są następujące:

	X_1	X_2	X_3
X_1			
X_2	0,50		
X_3	0,25	0,25	

a) Wyznaczyć efekt dywersyfikacji.
b) Wyznaczyć efekt dywersyfikacji przy założeniu, że ryzyka są niezależne.
c) Wskazać ograniczenia stosowania tej metody agregacji i potencjalne skutki niewłaściwego jej wykorzystania (tzn. zastosowania mimo niespełnienia podstawowych założeń). Skutki przedstawić w kontekście wyznaczania kapitałowych wymogów wypłacalności.

Odpowiedzi:

Odp. a)

Bezwzględny efekt dywersyfikacji: $D = 36{,}9059$.
Względny efekt dywersyfikacji: $d = 0{,}184525$ (ewentualnie $d = 0{,}815475$).

....

Odp. b)

Bezwzględny efekt dywersyfikacji: $D = 70{,}3852$.
Względny efekt dywersyfikacji: $d = 0{,}3519$ (ewentualnie $d = 0{,}6481$).

....

Odp. c)

Od strony metodologicznej taki sposób oceny efektu dywersyfikacji jest poprawny, gdy:

– rodzaje ryzyka $X_{1},\ldots,X_{k}$ mają wielowymiarowy rozkład normalny,
– wymogi kapitałowe zarówno dla rodzajów ryzyka $X_{1},\ldots,X_{k}$, jak i łącznego ryzyka X, są określane na podstawie formuły:

$$\kappa(X_i) = F_{X_i}^{-1}(\alpha) - E(X_i)$$

Niewłaściwe stosowanie prowadzi do przeszacowania lub niedoszacowania kapitałowych wymogów wypłacalności (SCR-ów). Oba przypadki są niepożądane przez zakład ubezpieczeń.

Zadanie 5.

Straty pierwszej linii biznesu (LOB1) są modelowane za pomocą zmiennej losowej X_1 o rozkładzie logarytmiczno-normalnym z parametrami $\mu = 2$ i $\sigma = 1$ (dystrybuanta $F_1(x) = \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)$, gdzie Φ oznacza dystrybuantę standardowego rozkładu normalnego), natomiast drugiej (LOB2) – za pomocą zmiennej losowej X_2 o rozkładzie wykładniczym z parametrem $\lambda = \frac{1}{12}$. Struktura zależności między stratami X_1 i X_2 (wektora losowego (X_1, X_2)) jest modelowana za pomocą kopuli Gumbela z parametrem $\theta = 4$. Wymogi kapitałowe są ustalane na poziomie kapitału ekonomicznego wyznaczanego za pomocą wartości zagrożonej, tzn. $\kappa(X) = VaR_\alpha(X) - E(X)$. W celu wyznaczenia łącznego wymogu kapitałowego dla LOB1 i LOB2, tzn. wymogu dla $S = X_1 + X_2$ zastosowano symulacyjną metodę agregacji za pomocą kopuli (Gumbela z parametrem $\theta = 4$).

a) Wiadomo, że z kopuli wygenerowano następującą parę (obserwację): $(u_1, v_1) = (0{,}304,\ 0{,}528)$. Wyznaczyć odpowiadającą tej parze wartość sumy $S = X_1 + X_2$ (realizację zmiennej S).

b) Wiadomo, że kapitał ekonomiczny wyznacza się w oparciu o VaR na poziomie ufności $\alpha = 0{,}995$. Ustalić efekt dywersyfikacji, jeżeli VaR dla sumy $S = X_1 + X_2$ wyznaczono za pomocą metody symulacji historycznej w oparciu o 1000 obserwacji wygenerowanych dla S. Wartości statystyk pozycyjnych $S_{(981:1000)}, \dots S_{(1000:1000)}$ podano w poniższej tabeli.

981	*982*	*983*	*984*	*985*	*986*	*987*	*988*	*989*	*990*
100,58	101,53	102,78	103,38	107,25	109,12	113,55	119,25	123,21	126,68

991	*992*	*993*	*994*	*995*	*996*	*997*	*998*	*999*	*1000*
129,31	130,52	146,05	152,43	162,95	177,77	202,74	219,14	230,94	252,36

Odpowiedzi:

Odp. a)

$$s = 4{,}4371 + 9{,}0093 = 13{,}4464$$

....

Odp. b)

Bezwzględny efekt dywersyfikacji:

$$D = 85{,}3319 + 51{,}5798 - 138{,}7675 = -1{,}8558$$

(Brak efektu dywersyfikacji, VaR nie jest koherentną miarą ryzyka, nie spełnia warunku subaddytywności.)

Zadanie 6.

Wysokość pojedynczej szkody X dla pewnego portfela ubezpieczeń modelowano za pomocą krzywej ogiwalnej wyznaczonej na podstawie następujących danych:

Wysokość pojedynczej szkody	Liczba szkód
(0, 10]	15
(10, 30]	35
(30, 80]	25
(80, 180]	15
(180, 380]	10
powyżej 380	0

a) Oszacować prawdopodobieństwo, że wysokość pojedynczej szkody nie przekroczy 150.
b) Oszacować wartość zagrożoną na poziomie ufności $\alpha = 0{,}95$ ($VaR_{0,95}(X)$).
c) Dla wypłat odszkodowań z tego portfela zakład ubezpieczeń stosuje górną granicę odpowiedzialności na poziomie 100 (tzn. wypłata $Y = \min(X, 100)$). Oszacować wartość oczekiwaną i wariancję dla wypłaty Y.

Odpowiedzi:

Odp. a)

$$P(X \leq 150) = 0{,}855$$

....

Odp. b)

$$VaR_{0,95}(X) = 280$$

....

Odp. c)

$$E(Y) = 46{,}2, \quad D^2(Y) = 1274{,}56$$

Zadanie 7.
W poniższej tabeli przedstawiono dane dotyczące liczby wypadków, płci (K - kobieta, M - mężczyzna), doświadczenia (D – doświadczony, N – niedoświadczony) oraz ekspozycji na ryzyko (1 – oznacza rok) dla 12 kierowców.

Nr kierowcy	Ekspozycja	Płeć	Doświadczenie	Liczba wypadków
1	1	M	D	0
2	1	K	D	0
3	0,8	M	D	1
4	0,6	K	N	1
5	0,8	K	D	0
6	0,6	K	N	1
7	1	M	N	0
8	1	M	N	2
9	0,6	M	D	0
10	0,4	K	D	1
11	0,8	M	N	1
12	1	K	D	0

Na podstawie tych danych oszacowano dwa uogólnione modele liniowe (GLM) dla liczby wypadków, przyjmując rozkład Poissona z logarytmiczną funkcją wiążącą (*link function*). Model zerowy (w skrócie M0), w którym nie uwzględniono żadnego regresora oraz model z obydwoma regresorami (w skrócie M1)

a) Oszacować parametry Modelu M0.
b) Wiedząc, że otrzymano następujące oszacowania parametrów modelu M1

(Intercept)	plec M	doswiadczenie N
-0,9387	-0,2268	1,3150

wyznaczyć różnicę między oszacowaniem za pomocą modelu M0 i M1 prawdopodobieństwa, że w ciągu roku mężczyzna, który jest doświadczonym kierowcą spowoduje co najwyżej jeden wypadek.

Odpowiedzi:

Odp. a)

$$\lambda_0 = \frac{7}{9{,}6} = 0{,}7292$$

....

Odp. b)

$$P^{M1}(K \leq 1) - P^{M0}(K \leq 1) = 0{,}9604 - 0{,}834 = 0{,}1264$$

Zadanie 8.

Na podstawie poniższych danych (zmienna X_3 przyjmuje dwie kategorie „D” i „Z”) oszacowano liniowy model regresji, w którym zmienną objaśnianą jest Y, a zmiennymi objaśniającymi są X_1, X_2 i X_3.

Nr obserwacji	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Y	6	10	12	7	9	9	11	10	13	8
X_1	12	8	5	11	10	9	8	9	5	11
X_2	25	19	15	23	17	22	14	16	10	23
X_3	Z	D	Z	D	D	Z	D	D	Z	D

Otrzymano następujące wyniki:

(Intercept)	X_1	X_2	X_3 Z
18,1422	-0,6349	-0,1577	-0,3828

Ponadto wiadomo, że suma kwadratów reszt (RSS) wynosi 0,81632 oraz elementy leżące na głównej przekątnej macierzy „daszkowej” (*hat matrix)* H są równe:

0,68894 0,45208 0,53916 0,26537 0,32897 0,34334 0,28986 0,23839 0,58853 0,26537.

Dla pierwszej obserwacji:

a) Wyznaczyć i zinterpretować wskaźnik wpływu (*leverage*).

b) Wyznaczyć i zinterpretować miarę Cooka (*Cook's distance*).

Odpowiedzi:

Odp. a) Wskaźnik wpływu dla pierwszej obserwacji, to $h_{11} = 0{,}68894$

Dany element diagonalny (h_{ii}) reprezentuje odległość pomiędzy wartościami macierzy *X* dla *i*-tej obserwacji, a wartościami średnimi wszystkich wartości macierzy *X*. Wartości te wskazują na to czy wartości macierzy *X* dla danej obserwacji są odstające. Element diagonalny jest określany terminem *wskaźnika wpływu*. Duża wartość *wskaźnika wpływu* oznacza, że i-ta obserwacja jest odległa od środka obserwacji w macierzy *X* . Inaczej mówiąc *wskaźnik wpływu* dla i-tej obserwacji opisuje wpływ tej obserwacji na wartość teoretyczną $\hat{y}_i$. Wskaźniki wpływu są bardzo użyteczne w diagnostyce modelu. Ich średnia wartość to $\frac{k+1}{n}$. W literaturze obserwację uznaje się za wpływową, jeżeli jej wskaźnik wpływu przekracza dwu lub trzykrotnie średnią.

W rozważanym przypadku: $\frac{k+1}{n} = 0{,}4$, więc pierwszą obserwację nie można uznać za wpływową.

....

Odp. b) Miara Cooka: $D_1 = 0{,}51329$

Wykazuje ona różnicę między wyznaczonymi wartościami współczynników β modelu, a wartościami wyznaczonymi przy wyłączeniu danego przypadku z obliczeń. Wszystkie odległości powinny być tego samego rzędu. Jeśli nie są, to można przypuszczać, że dany przypadek (przypadki) miał istotny wpływ na obciążenie współczynników równania regresji. Nie ma ustalonej granicy, której przekroczenie wskazuje na istotny wpływ. W literaturze przyjmuje się np. następujące: 1, $\frac{4}{n-(k+1)}$ lub $\frac{4}{n}$

W rozważanym przypadku: $\frac{4}{n-(k+1)} = \frac{4}{10-4} = 0{,}66666$

$\frac{4}{n} = \frac{4}{10} = 0{,}4$ (według tego kryterium obserwacja nr 1 ma wpływ na obciążenie współczynników równania regresji).

Zadanie 9.

Poniżej podano informacje dotyczące dwóch modeli oszacowanych na podstawie danych z poprzedniego zadania (zadanie 8) (**Model 1** jest taki sam jak w zadaniu 8, w **Modelu 2** uwzględniono tylko dwie zmienne objaśniające X_1 i X_2).

Model 1

formula: $Y \sim X_1 + X_2 + X_3$

(Intercept)	X_1	X_2	X_3 Z
18,1422	-0,6349	-0,1577	-0,3828

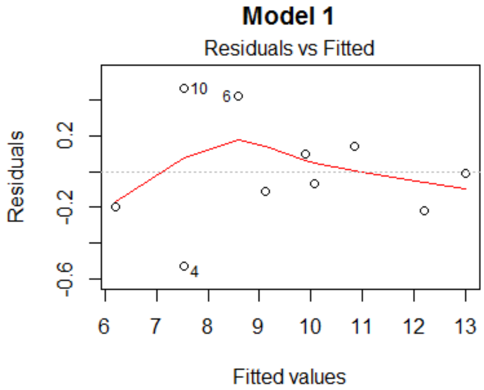
Residual standard error: 0,3689 on 6 degrees of freedom
Multiple R-squared: 0,9808,
Adjusted R-squared: 0,9712
F-statistic: 102,1 on 3 and 6 DF, p-value: 1,539e-05
AIC= 13,32344; BIC= 14.83636

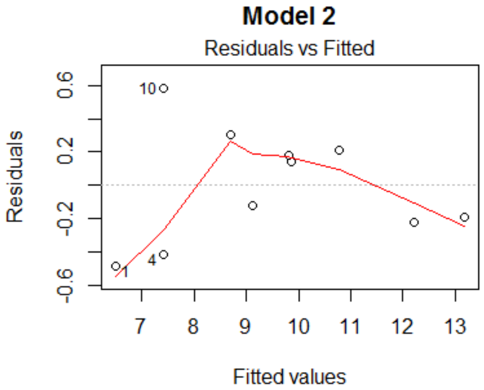
Model 2

formula: $Y \sim X_1 + X_2$

(Intercept)	X_1	X_2
17,8382	- 0,5425	-0,1937

Residual standard error: 0,3857 on 7 degrees of freedom
Multiple R-squared: 0,9755,
Adjusted R-squared: 0,9685
F-statistic: 139,3 on 2 and 7 DF, p-value: 2,304e-06
AIC= 13,76001; BIC= 14,97035





Kierując się powyższymi informacjami wybrać lepszy model (**wybór uzasadnić**) i za jego pomocą wyznaczyć prognozę zmiennej Y przy założeniu, że $X_1 = 10$, $X_2 = 10$ oraz X_3 przyjmuje kategorię „D". Ocenić błąd tej prognozy. Wiadomo, że macierze $(X^T X)^{-1}$ (X oznacza macierz obserwacji) dla tych modeli są następujące:

Model 1

$$\left|\begin{matrix} 2,21018 & -0,19646 & -0,00949 & -0,51680 \\ -0,19646 & 0,10635 & -0,04360 & 0,15705 \\ -0,00949 & -0,04360 & 0,02270 & -0,06117 \\ -0,51680 & 0,15705 & -0,06117 & 0,65072 \end{matrix}\right|$$

Model 2

$$\left|\begin{matrix} 1,79974 & -0,07173 & -0,05807 \\ -0,07173 & 0,06845 & -0,02884 \\ -0,05807 & -0,02884 & 0,01695 \end{matrix}\right|$$

Odpowiedzi:

Uzasadnienie wyboru modelu:

Model 1 jest lepszy:

- Mniejszy błąd standardowy reszt.
- Większy skorygowany współczynnik determinacji.
- Mniejsze wartości kryteriów informacyjnych (AIC, BIC)
- Rozkład reszt nie wykazuje trendu

Wartość prognozy:

$$y^P = 18,1422 - 0,6349 \cdot 10 - 0,1577 \cdot 10 - 0,3828 \cdot 0 = 10,2162$$

....

Ocena błędu prognozy:

Błąd bezwzględny: $s_D = 0{,}6677$

Błąd względny: $V_D = \frac{0{,}6677}{10{,}2162} = 0{,}065356$

Zadanie 10.
Poniższa tabela zawiera informacje dotyczące siedmiu oszacowanych uogólnionych modeli liniowych dla liczby szkód. W każdym z nich przyjęto, że zmienna objaśniana ma rozkład Poissona z logarytmiczną funkcją wiążącą (*link function*).

Model	Postać modelu (zmienne objaśniające)	Log-Likelihood	Liczba stopni swobody dla reszt (df.residual)	AIC
M0	wyraz wolny (null model)	-18056,99	67460	36115,98
M1	w.samoch+m.zamieszk+w.kierowcy	-18000,41	67447	36028,82
M2	w.samoch+m.zamieszk	-18033,85	67452	36085,70
M3	w.samoch+w.kierowcy	-18007,70	67452	36033,40
M4	m.zamieszk+w.kierowcy	-18013,32	67450	36048,64
M5	w.samoch	-18043,16		
M6	m.zamieszk	-18047,43		
M7	w.kierowcy	-18020,86		

w.samoch – wiek samochodu (zmienna jakościowa przyjmująca k_1 kategorii)
m.zamieszk – miejsce zamieszkania (zmienna jakościowa przyjmująca k_2 kategorii)
w.kierowcy – wiek kierowcy (zmienna jakościowa przyjmująca k_3 kategorii)

a) Ustalić liczbę kategorii dla poszczególnych zmiennych objaśniających (k_1, k_2, k_3).
b) Uporządkować wszystkie modele od najlepszego do najgorszego według kryterium AIC.
c) Sprawdzić na poziomie istotności 0,05, czy model M1 ma lepszą zdolność predykcyjną w porównaniu z modelem M4 (tzn. czy dodanie do M4 zmiennej w.samoch istotnie wpływa na polepszenie zdolności predykcyjnej).

Odpowiedzi:

Odp. a)

$$k_1 = 4,\ k_2 = 6,\ k_3 = 6$$

....

Odp. b)

M1, M3, M4, M7, M2, M5, M6, M0

....

Odp. c)
Wartość statystyki w teście ilorazu wiarygodności:

$$T = 2(-18000{,}41 + 18013{,}32) = 25{,}82$$

Wartość krytyczna: $\chi^2_{0,05;14-11} = \chi^2_{0,05;\,3} = 7{,}815$
Model M1 ma lepszą zdolność predykcyjną w porównaniu z M4.

Egzamin dla Aktuariuszy
Sesja egzaminacyjna w dniu 21 listopada 2017r.

Modelowanie

Arkusz ocen

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