Komisja Egzaminacyjna dla Aktuariuszy

LXXIX Egzamin dla Aktuariuszy

Sesja egzaminacyjna w dniu 27 listopada 2018 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 uogólnionych modeli liniowych (GLM):
- Residual deviance i Resid. Dev – oznacza dewiancję oszacowanego modelu,
- Null deviance – oznacza dewiancję modelu zerowego,
- Deviance – redukcję dewiancji po dodaniu kolejnej zmiennej objaśniającej.

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

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

d) W zadaniach warunkowa wartość zagrożoną (*Conditional Value at Risk*) jest definiowana w następujący sposób:

$$CVaR_\alpha(X) = E\big(X|X > VaR_\alpha(X)\big).$$

Miara ta nazywana jest także *Tail Value-at-Risk* i oznaczana przez $TVaR_\alpha(X)$.

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

f) 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

g) Wybrane kwantyle standardowego rozkładu normalnego

α	**0.900**	**0.950**	**0.975**	**0.990**	**0.995**	**0.998**	**0.999**
u_α	1.282	1.645	1.960	2.326	2.576	2.878	3.090

Zadanie 1.

Wybór przez kierowcę opcji pełnej lub częściowej ochrony w ramach ubezpieczenia AC oferowanego przez zakład ubezpieczeń XYZ modelowano z wykorzystaniem regresji logistycznej. Przyjęto, że zmienna zależna Y, przyjmuje dwie wartości: $Y = 1$, gdy kierowca wybierze opcję z pełną ochroną oraz $Y = 0$, gdy kierowca wybierze opcję z częściową ochroną oraz uwzględniono następujące zmienne objaśniające:

- *plec:* Płeć (K - kobieta, M – mężczyzna)
- *zamieszkanie:* Miejsce zamieszkania (Miasto – miasto, Wies – wieś)
- *uzytkowanie:* Użytkowanie samochodu (Pryw – do celów prywatnych, Sluzb – do celów służbowych)
- *stan.cywilny:* Stan cywilny (C – małżeństwo, S – singiel, O – inny)
- *wiek:* Wiek kierowcy w latach
- *lojalnosc:* Liczba lat, w których kierowca był klientem firmy XYZ.

Na podstawie zbioru uczącego liczącego 3109 obserwacji, metodą największej wiarygodności uzyskano następujące oszacowania parametrów modelu z wszystkimi zmiennymi objaśniającymi:

Coefficients:

	Estimate	*Std. Error*	*p-Value*
(Intercept)	1.911196	0.181985	<2e-16
plec: M	-0.991094	0.097332	<2e-16
zamieszkanie: Wies	-1.174730	0.089374	<2e-16
uzytkowanie: Sluzb	-1.058024	0.564130	0.0607
stan.cywilny: C	-0.019311	0.110712	0.8615
stan.cywilny: O	0.242578	0.231877	0.2955
wiek	-0.059577	0.004528	<2e-16
lojalnosc	0.140203	0.007978	<2e-16

oraz tabelę analizy dewiancji (*Deviance Table*):

Model: binomial, link: logit
Response: Y
Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			3108	3967.0
plec	1	107.76	3107	3859.3
zamieszkanie	1	136.76	3106	3722.5
uzytkowanie	1	3.53	3105	3719.0
stan.cywilny	2	22.19	3103	3696.8
wiek	1	44.64	3102	3652.1
lojalnosc	1	366.96	3101	3285.2

Na podstawie podanych informacji:

a) Zinterpretować oszacowane parametry.
b) Wyznaczyć i zinterpretować $pseudo - R^2$ ($McFadden\ R^2$).
c) Jaka jest prognoza wyboru opcji ubezpieczenia (zmiennej Y) dla kierowców charakteryzujących się następującymi wartościami zmiennych objaśniających:
   - *plec:* K (kobieta),
   - *zamieszkanie:* Miasto (miasto),
   - *uzytkowanie:* Pryw (do celów prywatnych),
   - *stan.cywilny:* S (singiel),
   - *wiek:* 30 lat,
   - *lojalnosc:* 5 lat.

   Przyjąć poziom odcięcia równy 0.35. Przy jakim poziomie odcięcia prognoza ulegnie zmianie.

Odpowiedzi:

....

Odp. a)

Interpretujemy tylko parametry statystycznie istotne, przyjmując poziom istotności równy 0.05.

- Parametr stojący przy kategorii *plec: M* wynosi -0.991094 i jest statystycznie istotny. Ujemna wartość parametru wskazuje, że mężczyzna jest mniej podatny na wybór pełnej opcji ubezpieczenia. Jeżeli kierowcy różnią się tylko płcią, to mężczyzna ma o $(1 - \exp(-0.991094)) * 100\% = 62.88\%$ niższe szanse na wybór pełnej opcji ubezpieczenia.
- Parametr stojący przy kategorii *zamieszkanie: Wies* wynosi -1.174730 i jest statystycznie istotny. Ujemna wartość parametru wskazuje, że kierowcy mieszkający na wsi są mniej podatni na wybór pełnej opcji ubezpieczenia. Jeżeli kierowcy różnią się tylko miejscem zamieszkania, to mieszkający na wsi mają o $(1 - \exp(-1.174730)) * 100\% = 69.11\%$ niższe szanse na wybór pełnej opcji ubezpieczenia.
- Parametry stojące przy kategoriach: *uzytkowanie: Sluzb, stan.cywilny: C, stan.cywilny: O* są statystycznie nieistotne. Nie mają wpływu na wybór opcji ubezpieczenia.
- Parametr stojący przy zmiennej *wiek* wynosi -0.059577 i jest statystycznie istotny. Spośród kierowców identycznych pod względem innych cech zawartych w modelu, a różniący się tylko wiekiem o 1 rok, kierowca starszy jest mniej podatny na kupno pełnej opcji ubezpieczenia. Kierowca starszy ma $(1 - \exp(-0.059577)) * 100\% = 5.78\%$ niższe szanse wyboru pełnej opcji ubezpieczenia.
- Parametr stojący przy zmiennej *lojalnosc* wynosi 0.140203 i jest statystycznie istotny. Spośród kierowców identycznych pod względem innych cech zawartych w modelu, a różniący się tylko liczbą lat w firmie XYZ o 1 rok, kierowca bardziej lojalny („z większym stażem w firmie XYZ") jest bardziej podatny na kupno pełnej opcji ubezpieczenia. Kierowca bardziej lojalny ma $(\exp(0.140203) - 1) * 100\% = 15.05\%$ wyższe szanse wyboru pełnej opcji ubezpieczenia.

Odp. b)
Współczynnik
$$pseudo - R^2 = 1 - \frac{D_M}{D_{Null}},$$
gdzie
D_M – dewiancja oszacowanego modelu
D_{Null} – dewiancja modelu zerowego (tylko ze stałą)
Na podstawie tablicy dewiancji:
$D_M = 3285.2,$
lub
$D_M = 3967 - (107.76 + 136.76 + 3.53 + 22.19 + 44.64 + 366.96) = 3967 - 681.8 = 3285.2,$
$D_{Null} = 3967.$
$pseudo - R^2 = 1 - \frac{3285.2}{3967} = 0.1719$
Uwzględnione w modelu zmienne objaśniające redukują dewiancję modelu zerowego (bez zmiennych objaśniających) o 17.19%.

....

Odp. c)
Dla kierowców charakteryzujących się następującymi wartościami zmiennych objaśniających:

- *plec:* K (kobieta),
- *zamieszkanie:* Miasto (miasto),
- *uzytkowanie:* Pryw (do celów prywatnych),
- *stan.cywilny:* S (singiel),
- *wiek:* 30 lat,
- *lojalnosc:* 5 lat,

oszacowana wartość predyktora liniowego wynosi:
$$\hat{\eta} = 1.911196 - 0.991094 \cdot 0 - 1.174730 \cdot 0 - 1.058024 \cdot 0 - 0.019311 \cdot 0 + 0.242578 \cdot 0 - 0.059577 \cdot 30 + 0.140203 \cdot 5 = 0.824901$$
Stąd otrzymujemy:
$$\hat{p} = \frac{\exp(0.824901)}{1 + \exp(0.824901)} = \frac{2.281655}{1 + 2.281655} = 0.6953$$
Przy poziomie odcięcia równym 0.35 prognozujemy, że zmienna $Y = 1$, czyli prognozujemy, że kierowcy ci wybiorą pełną opcję ubezpieczenia. Prognoza ulegnie zmianie, gdy poziom odcięcia będzie większy 0.6953.

Rozwiązanie:

Zadanie 2.

Podobnie jak w zadaniu 1, wybór przez kierowcę opcji pełnej lub częściowej ochrony w ramach ubezpieczenia AC oferowanego przez zakład ubezpieczeń XYZ modelowano z wykorzystaniem regresji logistycznej. Przyjęto, że zmienna zależna Y, przyjmuje dwie wartości: $Y = 1$, gdy kierowca wybierze opcję z pełną ochroną oraz $Y = 0$, gdy kierowca wybierze opcję z częściową ochroną oraz uwzględniono następujące zmienne objaśniające:

- *plec:* Płeć (K - kobieta, M – mężczyzna)
- *zamieszkanie:* Miejsce zamieszkania (Miasto – miasto, Wies – wieś)
- *uzytkowanie:* Użytkowanie samochodu (Pryw – do celów prywatnych, Sluzb – do celów służbowych)
- *stan.cywilny:* Stan cywilny (C – małżeństwo, S – singiel, O – inny)
- *wiek:* Wiek kierowcy w latach
- *lojalnosc:* Liczba lat, w których kierowca był klientem firmy XYZ.

Na podstawie zbioru uczącego liczącego 3109 obserwacji, metodą największej wiarygodności oszacowano dwa modele (M1 i M2). Uzyskano następujące wyniki:

Model M1:

Coefficients:

	Estimate	*Std. Error*	*p-Value*
(Intercept)	1.906310	0.181497	<2e-16
plec: M	-1.010303	0.096792	<2e-16
zamieszkanie: Wies	-1.175562	0.089101	<2e-16
wiek	-0.059375	0.004196	<2e-16
lojalnosc	0.140010	0.007938	<2e-16

Null deviance: 3967.0 on 3108 degrees of freedom
Residual deviance: 3290.5 on 3104 degrees of freedom
AIC: 3300.5

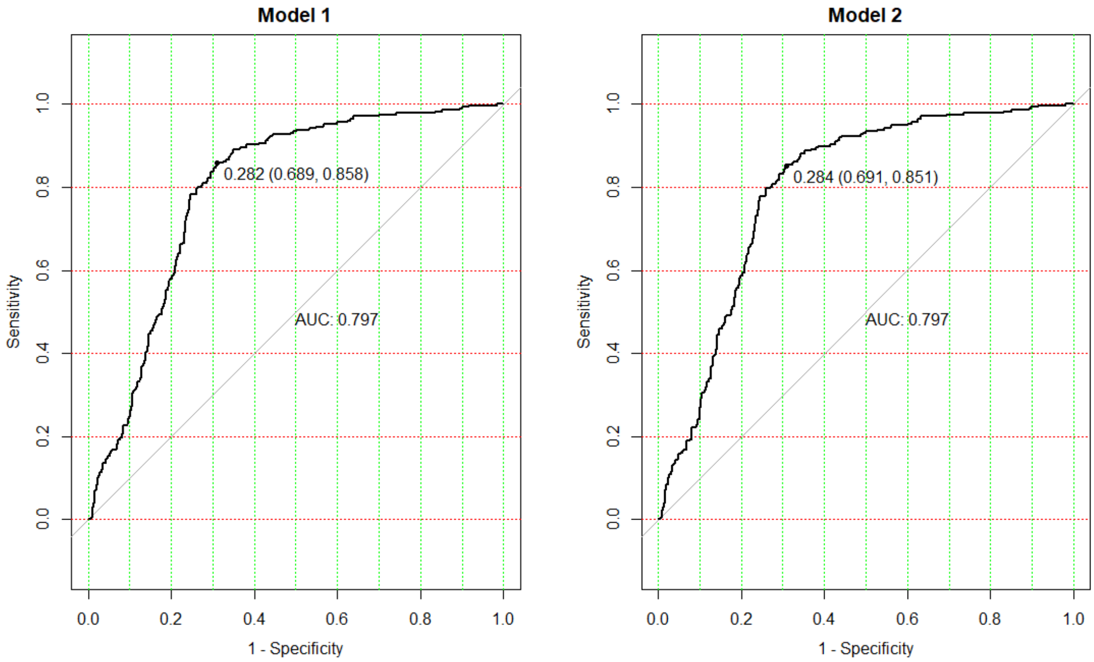
Model M2:

Coefficients:

	Estimate	*Std. Error*	*p-Value*
(Intercept)	1.916078	0.181678	<2e-16
plec: M	-1.001466	0.096887	<2e-16
zamieszkanie: Wies	-1.174699	0.089163	<2e-16
uzytkowanie: Sluzb	-1.009913	0.561654	0.0722
wiek	-0.059545	0.004197	<2e-16
lojalnosc	0.139905	0.007942	<2e-16

Null deviance: 3967.0 on 3108 degrees of freedom
Residual deviance: 3286.6 on 3103 degrees of freedom
AIC: 3298.6

Krzywe ROC (*Receiver Operator Characteristics*) wyznaczone na podstawie zbioru testowego dla oszacowanych modeli przedstawiono na poniższym rysunku (AUC oznacza pole pod krzywą ROC):



Na podstawie podanych wyników:

a) Sprawdzić na poziomie istotności 0.05, czy model M2 ma lepszą zdolność predykcyjną w porównaniu z modelem M1 (tzn. czy dodanie do M1 dodatkowej zmiennej istotnie wpływa na polepszenie zdolności predykcyjnej).
b) Jak konstruuje się i interpretuje krzywą ROC.
c) Uwzględniając wyniki z punktu a), wartości kryteriów informacyjnych *AIC* oraz wskazania krzywych ROC wybrać lepszy model. Wybór uzasadnić.

Odpowiedzi:

....

Odp. a)

Statystyka testowa:

$$T = 2(l_2 - l_1) = D_1 - D_2,$$

gdzie:

$l_2,\ l_1$ - logarytmy wiarygodności odpowiednio modelu M2 i M1,

D_2, D_1 - dewiancja odpowiednio modelu M2 i M1.

Wartość statystyki wynosi:

$$T = D_1 - D_2 = 3290.5 - 3286.6 = 3.9$$

Wartość krytyczna:

$$\chi^2_{0{,}05;1} = 3.841.$$

Ponieważ $T > 3.841$, na poziomie istotności 0.05 możemy przyjąć, że model M2 ma lepsze zdolności predykcyjne.

....

Odp. b)
Konstrukcja krzywej ROC:
1. Dla każdego z możliwych punktów odcięcia obliczmy czułość i specyficzność.
2. W układzie współrzędnych, zaznaczamy punkty o współrzędnych:
   (*1-specyficzność*, *czułość*).
3. Łączymy te punkty otrzymując krzywą ROC.

Na podstawie krzywej ROC można ustalić optymalny punkt odcięcia. Przyjmując równe koszty błędnych klasyfikacji, optymalnym punktem odcięcia jest punkt na krzywej ROC znajdujący się najbliżej punktu o współrzędnych (0.1).
Krzywą ROC wykorzystuje się także do oceny i porównywania między sobą modeli klasyfikacyjnych. Jako miarę dobroci i trafności danego modelu przyjmuje się pole pod wykresem krzywej ROC, oznaczane jako AUC. Miernik AUC przyjmuje wartości z przedziału $[0, 1]$. Im wartość tego miernika jest większy, tym model jest lepszy.

....
Odp. c)

Lepszy jest model M2:
- ma lepsze zdolności predykcyjne,
- ma mniejszą wartość AIC
- ma taką samą wartość miernika AUC.

Uwaga! Ze względu na nieistotność parametru przy zmiennej *uzytkowanie: Sluzb* w M2, podobne wartości AIC obu modeli, można było wskazać także model M1.

Rozwiązanie:

Zadanie 3.

Przedstawić *k*-krotną walidację krzyżową (*k-fold cross-validation*).

Odpowiedzi:

....

Odp.

W przypadku, gdy dysponujemy próbą o małej liczebności uniemożliwiającą określenie zbioru walidacyjnego, można zastosować tzw. *k*-krotną walidację krzyżową. Pozwala ona na wykorzystanie całej próby zarówno do uczenia jak i do walidacji. Służy do określenia jakości modelu w trakcie jego uczenia, w celu wyeliminowania problemu przeuczenia się. Sposób postępowania:

- Początkową próbę losowo dzielimy na *k* możliwie równolicznych wzajemnie niezależnych podzbiorów S_1, S_2, ..., S_k.
- Spośród podzbiorów S_1, S_2, ..., S_k wybieramy S_i (i=1,..., *k*).
- Szacujemy model wykorzystując *k*-1 pozostałych podzbiorów.
- Testujemy model na podzbiorze S_i.
- Postępując w ten sposób, szacujemy *k* modeli. Każdy z podzbiorów jest wykorzystany do oszacowania *k*-1 modeli i raz do testowania.
- Następnie *k* rezultatów uśredniamy (lub łączymy w inny sposób) w celu uzyskania jednego wyniku.

Rozwiązanie:

Zadanie 4.

Składki czyste (*pure premiums*) dla pewnego portfela ubezpieczeń AC wyznaczano wykorzystując model dla liczby szkód K zgłaszanych przez kierowców w ciągu jednego roku (*frequency model*) i model dla wysokości pojedynczej szkody (*severity model*).

Liczbę szkód K modelowano z wykorzystaniem regresji Poissona (uogólnionego modelu liniowego dla zmiennej zależnej o rozkładzie Poissona). Przyjęto następujące zmienne objaśniające:

- *klasa.ryzyka:* Zmienna jakościowa określająca klasę ryzyka, którą przypisano kierowcy. Przyjmuje następujące wartości: *kl.1, kl.2, kl.3, kl.4, kl.5 i kl.6* (im wyższa klasa, tym większa zniżka za bezszkodowość),
- *prior.szkody:* Liczba wcześniejszych szkód (zmienna ilościowa),
- *wiek:* Wiek kierowcy. Zmienna jakościowa przyjmująca następujące wartości *mlody* (młody kierowca), *doswiadczony* (doświadczony kierowca), *starszy* (starszy kierowca).

Metodą największej wiarygodności oszacowano dwa modele dla zmiennej K (KM1 i KM2). Uzyskano następujące wyniki:

Model KM1

Coefficients:

	Estimate	*Std. Error*	*p-Value*
(Intercept)	-1.82286	0.04994	<2e-16

Null deviance: 1584.4 on 2481 degrees of freedom
Residual deviance: 1584.4 on 2481 degrees of freedom
AIC: 2329

Model KM2

Coefficients:

	Estimate	*Std. Error*	*p-Value*
(Intercept)	-0.49180	0.50179	0.327043
klasa.ryzyka: kl.2	-0.79067	0.50818	0.119739
klasa.ryzyka: kl.3	-0.33024	0.12500	0.008245
klasa.ryzyka: kl.4	-0.42870	0.16878	0.011084
klasa.ryzyka: kl.5	-0.79371	0.23132	0.000601
klasa.ryzyka: kl.6	-0.71392	0.15799	6.22e-06
prior.szkody	0.09523	0.03182	0.002766
wiek: doswiadczony	-1.09807	0.50400	0.029352
wiek: starszy	-1.16895	0.53880	0.030041

Null deviance: 1584.4 on 2481 degrees of freedom
Residual deviance: 1541.8 on 2473 degrees of freedom
AIC: 2302.4

Wysokość pojedynczej szkody Y modelowano bez uwzględniania zmiennych objaśniający. Dopasowano trzy rozkłady: normalny (model MY1) gamma (model MY2) i logarytmiczno-normalny (model MY3). Uzyskano następujące wyniki:

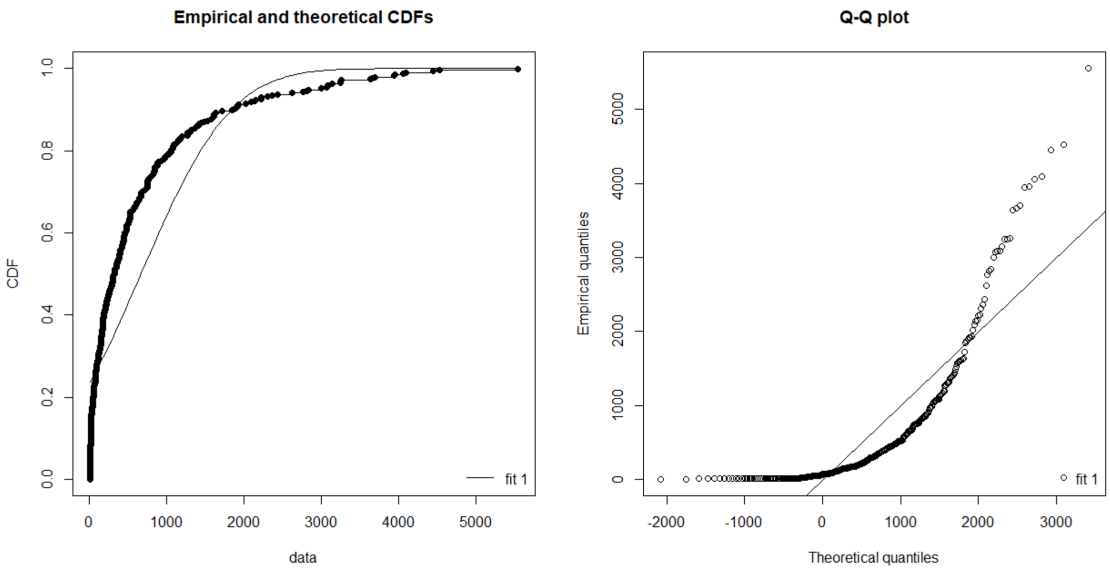
Model MY1

Fitting of the distribution ' norm ' by maximum likelihood

Parameters :

	estimate	Std. Error
mean	667.5034	48.43165
sd	917.7275	34.24635

Loglikelihood: -2958.461 AIC: 5920.922 BIC: 5928.689



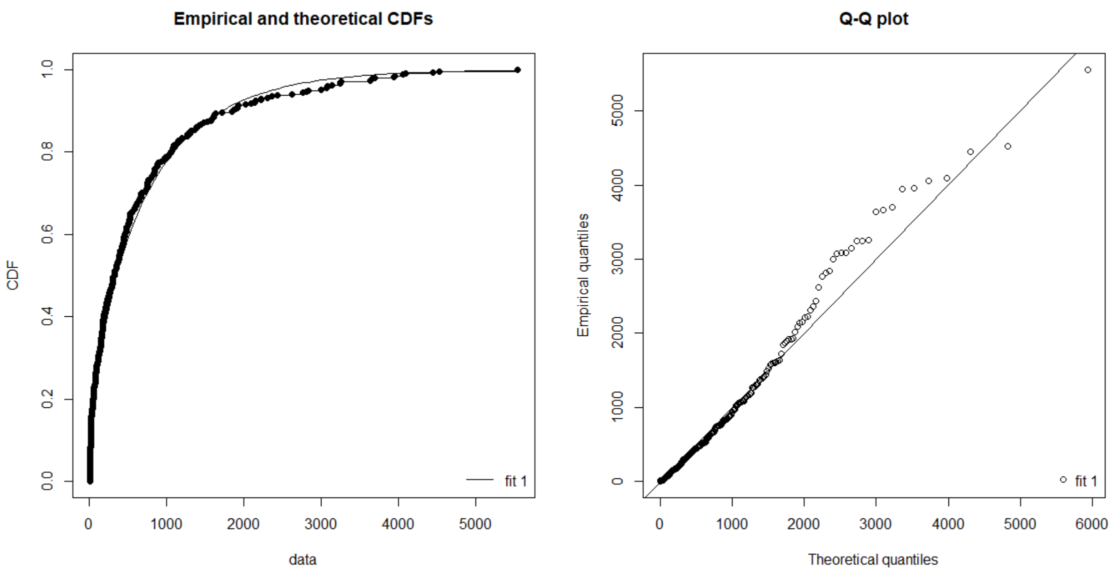
Model MY2

Fitting of the distribution ' gamma ' by maximum likelihood

Parameters :

	estimate	Std. Error
scale	1084.0501	99.15078318
shape	0.613407	0.03840023

Loglikelihood: -2659.254 AIC: 5322.508 BIC: 5330.275

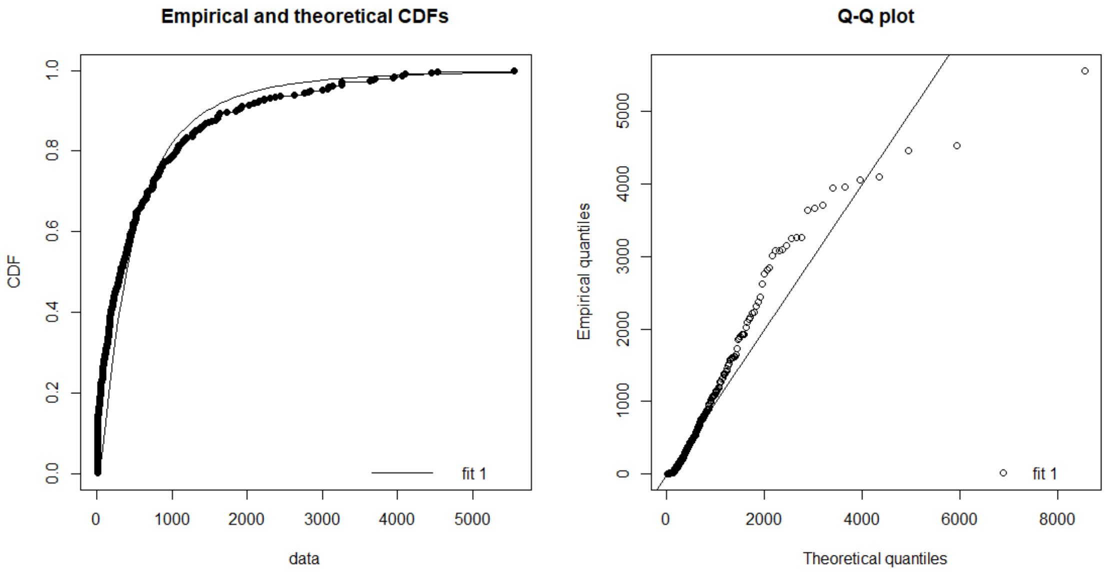


Model MY3

Fitting of the distribution ' lnorm ' by maximum likelihood
Parameters :

	estimate	Std. Error
meanlog	5.972872	0.08583359
sdlog	1.030216	0.06069341

Loglikelihood: -2800.588 AIC: 5605.177 BIC: 5612.943



Dla oszacowanych modeli otrzymano następujące wartości statystyk dopasowania:

Goodness-of-fit statistics:

	MY1	MY2	MY3
Kolmogorov-Smirnov statistic	0.2370056	0.06452082	0.1991727
Cramer-von Mises statistic	6.1474697	0.27766283	3.8102485
Anderson-Darling statistic	33.9218995	2.43076031	22.4832908

Oszacowane średnie dla poszczególnych modeli podano w poniższej tabeli:

Model	Średnia
MY1 (rozkład normalny)	667.5034
MY2 (rozkład gamma)	664.9639
MY3 (rozkład logarytmiczno-normalny)	667.5034

Na podstawie przedstawionych wyników:

a) Wybrać najlepszy model dla liczby szkód i wysokości pojedynczej szkody. Wybór uzasadnić.
b) Wykorzystując wybrane modele wyznaczyć różnicę między składką czystą dla młodych kierowców (*wiek=mlody*) z pierwszej klasy ryzyka (*kl.1*) i szóstej (*kl.6*), którzy wcześniej zanotowali po 2 szkody (*prior.szkody=2*).

Odpowiedzi:

....

Odp. a)

- Dla liczby szkód spośród podanych dwóch modeli lepszym jest model **KM2.** W porównaniu z **KM1** ma lepsze zdolności predykcyjne (wartość statystyki $T = 42.6$, wartość krytyczna dla poziomu istotności 0.05 wynosi 15.507). Ponadto w porównaniu z KM1 ma mniejszą wartość AIC.
- Dla wysokości pojedynczej szkody spośród trzech podanych modeli najlepszym jest model **MY2.** W porównaniu z pozostałymi ma najmniejszą wartość AIC oraz najmniejsze wartości statystyk testowych. Na wybór tego modelu wskazuje również wykres kwantyl-kwantyl.

....

Odp. b)

$$|491.9602 - 240.9231| = 251.0371$$

Rozwiązanie:

Składka czysta dla kierowcy z klasy ryzyka *kl.1*:

- *klasa.ryzyka: kl.1,*
- *prior.szkody* = 2
- *wiek: mlody*

Oszacowanie średniej liczby szkód:

$$\hat{\mu}_{kl.1} = \exp(-0.49180 \ + 0.09523 \cdot 2) = 0.73983$$

Składka czysta:

$$S_{kl.1} = 0.73983 \cdot 664.9639 = 491.9602$$

Składka czysta dla kierowcy z klasy ryzyka *kl.6*:

- *klasa.ryzyka: kl.6,*
- *prior.szkody* = 2
- *wiek: mlody*

Oszacowanie średniej liczby szkód:

$$\hat{\mu}_{kl.1} = \exp(-0.49180 - 0.71392 \cdot 1 \ + 0.09523 \cdot 2) = 0.36231$$

Składka czysta:

$$S_{kl.1} = 0.36231 \cdot 664.9639 = 240.9231$$

Różnica:

$$|491.9602 - 240.9231| = 251.0371$$

Zadanie 5.

Roczne straty dla dwóch linii biznesu (LOB1 i LOB2) mają dwuwymiarowy rozkład normalny. Średnie straty dla LOB1 i LOB2 wynoszą odpowiednio 20 mln. zł i 30 mln. zł, a odchylenia standardowe dla LOB1 i LOB2 są odpowiednio równe 10 mln. zł i 5 mln. zł. Wiadomo też, że współczynnik korelacji między stratami z tych linii biznesu jest równy $\rho = 0.6$. (Uwaga! Dla zmiennych losowych X, Y o dwuwymiarowym rozkładzie normalnym: $\sigma_{Y|X}^2 = \sigma_Y^2(1-\rho^2)$.)

a) Obliczyć prawdopodobieństwo, że starty dla LOB2 będą mniejsze od 40.304 mln. zł pod warunkiem, że straty dla LOB1 są równe 20 mln. zł.
b) Obliczyć roczną wartość zagrożoną na poziomie ufności 0.995 ($VaR_{0.995}$) dla strat LOB2 pod warunkiem, że straty dla LOB1 są równe 20 mln. zł.

Odpowiedzi:

……

Odp. a)

Prawdopodobieństwo, że starty dla LOB2 będą mniejsze od 40.304 mln. zł pod warunkiem, że straty dla LOB1 są równe 20 mln. zł jest równe 0.995.

……

Odp. b)

$$VaR_{0.995}(Y|X=20) = 40.304$$

Rozwiązanie:

a)

Niech X i Y oznaczają straty odpowiednio linii LOB1 i LOB2. Z danych wynika, że $X \sim N(20, 10)$ i $Y \sim N(30, 5)$, czyli: $\mu_X = 20$, $\sigma_X = 10$, $\mu_Y = 30$, $\sigma_Y = 5$ oraz współczynnik korelacji liniowej między X i Y wynosi $\rho = 0.6$.

W przypadku, gdy zmienne X i Y mają dwuwymiarowy rozkład normalny, zmienna $Y|X = x$ ma rozkład normalny o wartości oczekiwanej równej $\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X)$ i odchyleniu standardowym $\sigma_{Y|X} = \sigma_Y\sqrt{1-\rho^2}$.

Zatem zmienna losowa $Y|X = 20$ ma rozkład normalny $N(30,4)$, więc prawdopodobieństwo, że starty dla LOB2 będą mniejsze od 40.304 mln. zł pod warunkiem, że straty dla LOB1 są równe 20 mln. zł jest równe:

$P\left(U < \frac{40.304-30}{4}\right) = \Phi(2.576) = 0.995.$

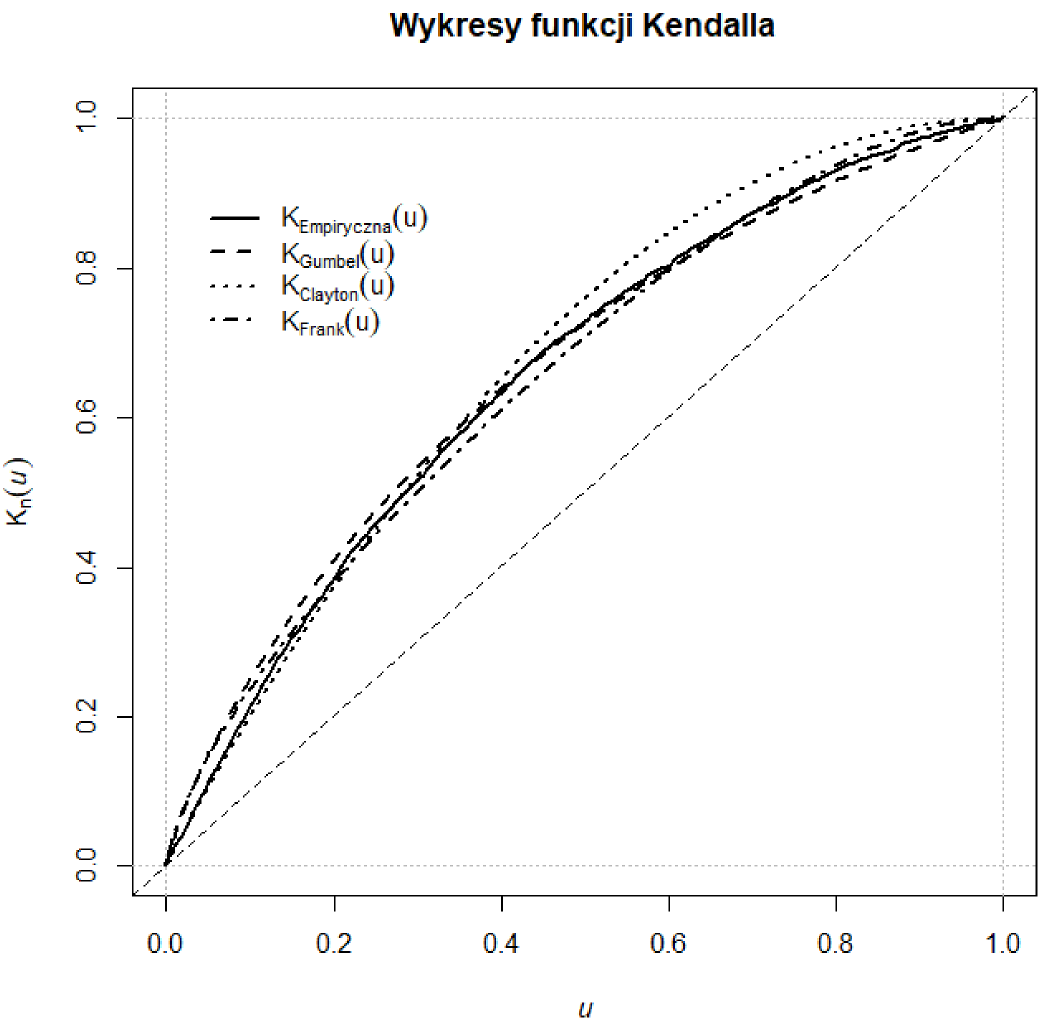
b) Na podstawie wyników z punktu a) wartość zagrożona $VaR_{0.995}(Y|X = 20) = 40.304$

Zadanie 6.

Modelując strukturę zależności między stopami zwrotu z dwóch indeksów branżowych WGPW wzięto pod uwagę trzy kopule, tj. Gumbela, Claytona i Franka. W celu wyboru najlepszej spośród nich wykorzystano wykresy funkcji Kendalla (K) oraz wartości statystyki S_n:

$$S_n = n \int_{[0,1]^d} \left\{\hat{C}_n(u_1, \dots, u_d) - C_{\theta_n}(u_1, \dots, u_d)\right\}^2 d\hat{C}_n(u_1, \dots, u_d),$$

gdzie:
$\hat{C}_n(u_1, \dots, u_d)$ – kopula empiryczna (oszacowana na podstawie n-elementowej próby),
$C_{\theta_n}(u_1, \dots, u_d)$ – testowana kopula, której parametry θ_n oszacowano na podstawie n-elementowej próby.
Uzyskano następujące wyniki:



Wartości statystyki S_n

Kopula	*Wartość statystyki*
Gumbela	0.2639
Claytona	0.7198

Franka	0.2165

a) Co to jest funkcja Kendalla (*Kendall's function*) i w jaki sposób można ją wykorzystać w wyborze kopuli?
b) Na podstawie podanych wyników wybrać najlepszą kopulę (spośród trzech rozważanych). Wybór uzasadnić.

Odpowiedzi:

....

Odp. a)

Niech C będzie d-wymiarową kopulą oraz zmienne losowe $U_1, \dots U_d$ mają rozkłady jednostajne na $[0, 1]$. Jeżeli $\boldsymbol{U} = (U_1, \dots U_d) \sim C$, wówczas funkcją Kendalla nazywamy dystrybuantę zmiennej $C(\boldsymbol{U})$:

$$K_C(t) = P(C(\boldsymbol{U}) \leq t; \boldsymbol{U} \sim C).$$

W wyborze kopuli można ją wykorzystać w następujący sposób:

- Porównując empiryczną funkcję Kendalla $K_n(t)$ z funkcjami Kendalla wyznaczonymi dla rozważanych (możliwych do wyboru) kopuli. Wybieramy wtedy kopulę, dla której funkcja Kendalla jest „najbliższa" empirycznej funkcji Kendalla.
- Stosując testy zgodności dla kopuli bazujące na funkcji Kendalla.

....

Odp. b)

Kopula Franka

Wartości statystyki S_n (bazującej na porównaniu odległości danej kopuli od kopuli empirycznej) wskazują, że spośród rozważanych kopuli najlepszą jest kopula Franka. Można także uznać, że funkcja Kendalla dla kopuli Franka jest „najbliższa" empirycznej funkcji Kendalla.

Jeżeli jednak jesteśmy zainteresowani modelowaniem zależności w ogonach, to należałoby wskazać (na podstawie wykresów funkcji Kendalla):

- kopulę Claytona, w przypadku dolnego ogona;
- kopulę Gumbela, w przypadku górnego ogona.

Rozwiązanie:

Zadanie 7.

W celu analizy lojalności klientów zakładu ubezpieczeń XYZ kupujących ubezpieczenie OC zastosowano model proporcjonalnego hazardu Coxa dla tych, którzy przenieśli się do innej firmy w ciągu pierwszych dwóch lat. Wykorzystano w nim zmienne, które wydają się mieć wpływ na taką decyzję. Na podstawie danych uzyskano następujące oszacowania tego modelu:

Zmienna	*Kategorie*	*Parametr*
Płeć	Mężczyzna	-0.30
	Kobieta	0
Zamieszkanie	Wieś	-0.10
	Miasto	0
Stan cywilny	Singiel	0.20
	Inny	0.10
	Małżeństwo	0

a) Zapisać model, podając i definiując wszystkie jego składowe i zmienne objaśniające.
b) Podać cechy osoby, której dotyczy hazard bazowy (hazard odniesienia).
c) Prawdopodobieństwo, że mężczyzna będący singlem i mieszkający w mieście zrezygnuje z umowy przed upływem dwóch lat wynosi 0.60. Obliczyć prawdopodobieństwo pozostania przez co najmniej dwa lata w firmie XYZ dla żonatego mężczyzny mieszkającego na wsi.

Odpowiedzi:

....

Odp. a)

Postać modelu:

$$h(t;\ x_1, x_2, x_3, x_4) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4),$$

gdzie:

$h(t;\ x_1, x_2, x_3, x_4)$ - oznacza funkcję hazardu w momencie t przy danych zmiennych objaśniających,

$h_0(t)$ – hazard bazowy (odniesienia),

$\beta_1, \beta_2, \beta_3, \beta_4$ – parametry modelu,

x_1- zmienna objaśniająca przyjmująca wartość 1, jeśli klient jest mężczyzną oraz 0, w przeciwnym przypadku,

x_2- zmienna objaśniająca przyjmująca wartość 1, jeśli klient mieszka na wsi oraz 0, w przeciwnym przypadku,

x_3- zmienna objaśniająca przyjmująca wartość 1, jeśli klient jest singlem oraz 0, w przeciwnym przypadku,

x_4- zmienna objaśniająca przyjmująca wartość 1, jeśli stan cywilny klienta jest inny niż singiel i inny niż żonaty oraz 0, w przeciwnym przypadku.

....

Odp. b)

Hazard bazowy dotyczy żonatej kobiety mieszkającej w mieście.

....

Odp. c)

$$0.50722$$

Rozwiązanie:
Prawdopodobieństwo, że mężczyzna będący singlem i mieszkający w mieście zrezygnuje z umowy przed upływem dwóch lat wynosi 0.60. Funkcja hazardu dla tego mężczyzny ma postać $h_0(t)\exp(-0.30+0.20)=h_0(t)\exp(-0.10)$.
Prawdopodobieństwo, że mężczyzna ten nie zrezygnuje z umowy jest równe:

$$0.4=\exp\left\{-\int_0^2 h_0(t)\exp(-0.10)\,dt\right\}=\exp\left\{-\exp(-0.10)\int_0^2 h_0(t)dt\right\},$$

stąd:

$$\int_0^2 h_0(t)dt=\frac{-\ln(0.4)}{\exp(-0.10)}$$

Funkcja hazardu dla żonatego mężczyzny mieszkającego na wsi ma postać:

$$h_0(t)\exp(-0.30-0.10)=h_0(t)\exp(-0.40),$$

zatem prawdopodobieństwo, że mężczyzna ten pozostania przez co najmniej dwa lata w firmie XYZ jest równe:

$$\exp\left\{-\int_0^2 h_0(t)\exp(-0.40)\,dt\right\}=\exp\left\{-\exp(-0.40)\frac{-\ln(0.4)}{\exp(-0.10)}\right\}=0.50722$$

Zadanie 8.
W poniższej tabeli dla dwóch segmentów ubezpieczeń innych niż ubezpieczenia na życie podano:

– miarę wielkości ryzyka składki ($V_{(prem,i)}, i = s, t$),
– miarę wielkości ryzyka rezerw ($V_{(res,s)}, i = s, t$),
– odchylenie standardowe ryzyka składki ($\sigma_{(prem,i)}, i = s, t$),
– odchylenie standardowe ryzyka rezerw($\sigma_{(res,i)}, i = s, t$),

oraz parametr zależności ryzyka składki i rezerw między tymi segmentami ($CorrS_{s,t}$).

Segment *s*	**Segment *t***
$V_{(prem,s)} = 1.0$	$V_{(prem,t)} = 1.0$
$V_{(res,s)} = 1.2$	$V_{(res,t)} = 1.2$
$\sigma_{(prem,s)} = 0.10$	$\sigma_{(prem,t)} = 0.08$
$\sigma_{(res,s)} = 0.09$	$\sigma_{(res,t)} = 0.08$
$CorrS_{s,t} = 0.5$	

a) Wyznaczyć łączny wymóg kapitałowy dla ryzyka składki i rezerw tych segmentów, wykorzystując standardową formułę Solvency II.
b) Wyznaczyć ten wymóg przy założeniu, że parametr zależności ryzyka składki i rezerw między tymi segmentami wynosi zero (tzn. $CorrS_{s,t} = 0$).
c) Wyznaczyć ten wymóg przy założeniu, że parametr zależności ryzyka składki i rezerw między tymi segmentami wynosi 1 (tzn. $CorrS_{s,t} = 1$). Czy w tym przypadku otrzymujemy górne ograniczenie wymogu kapitałowego dla ryzyka składki i rezerw? Odpowiedź uzasadnić.

Odpowiedzi:

....

Odp. a)

0,86565

....

Odp. b)

0,70841

....

Odp. c)

0,99842

Odpowiednio uzasadniając (tzn. wskazując, czy ryzyka składki i rezerw między tymi segmentami mają dwuwymiarowy rozkład normalny, czy też nie), na pytanie można odpowiedzieć zarówno, że otrzymujemy jak i, że nie otrzymujemy górnego ograniczenia.

Rozwiązanie:
Formuła standardowa na SCR dla ryzyka składki rezerw jest następująca (zob. Art. 115 [ROZPORZĄDZENIE DELEGOWANE..., 2015]):

$$SCR_{nl\,prem\,res} = 3 \cdot \sigma_{nl} \cdot V_{nl}$$

gdzie σ_{nl} - odchylenie standardowe ryzyka składki i rezerw w ubezpieczeniach innych niż na życie wyznaczone zgodnie z art. 117 [ROZPORZĄDZENIE DELEGOWANE..., 2015]; V_{nl} - miara wielkości ryzyka składki i rezerw w ubezpieczeniach innych niż na życie, wyznaczone zgodnie z art. 116 [ROZPORZĄDZENIE DELEGOWANE..., 2015]. Przy czym V_{nl} jest równe:

$$V_{nl} = \sum_s V_s,$$

gdzie V_s oznacza miarę wielkości ryzyka składki i rezerw dla segmentu s (wykaz segmentów podano w Aneksie II [ROZPORZĄDZENIE DELEGOWANE..., 2015].
Odchylenie standardowe σ_{nl} jest wyznaczane następująco (zob. Art. 117, Par.1 [ROZPORZĄDZENIE DELEGOWANE..., 2015]):

$$\sigma_{nl} = \frac{1}{V_{nl}} \cdot \sqrt{\sum_{s,t} CorrS_{s,t} \cdot \sigma_s \cdot V_s \cdot \sigma_t \cdot V_t}$$

gdzie
$CorrS_{s,t}$- oznacza parametr zależności ryzyka składki i rezerw w ubezpieczeniach innych niż na życie dla segmentów s i t określony w Aneksie IV [ROZPORZĄDZENIE DELEGOWANE..., 2015]; σ_s, σ_t- odchylenia standardowe ryzyka składki i rezerw w ubezpieczeniach innych niż na życie odpowiednio dla segmentów s i t; V_s, V_t - miara wielkości ryzyka składki i rezerw odpowiednio dla segmentów s i t, o których mowa w art. 116 [ROZPORZĄDZENIE DELEGOWANE..., 2015]. Odchylenie standardowe ryzyka składki i rezerw dla konkretnego segmentu s jest równe (zob. Art. 117, Par.2 [ROZPORZĄDZENIE DELEGOWANE..., 2015]):

$$\sigma_s = \frac{\sqrt{\sigma^2_{(pr,s)} \cdot V^2_{(pr,s)} + \sigma_{(pr,s)} \cdot V_{(pr,s)} \cdot \sigma_{(res,s)} \cdot V_{(res,s)} + \sigma^2_{(res,s)} \cdot V^2_{(res,s)}}}{V_{(pr,s)} + V_{(res,s)}}$$

gdzie: $\sigma_{(pr,s)}$ - odchylenie standardowe ryzyka składki dla segmentu s wyznaczone zgodnie z Art. 117, Par.3 [ROZPORZĄDZENIE DELEGOWANE..., 2015]); $\sigma_{(res,s)}$ - odchylenie standardowe ryzyka rezerw dla segmentu s określone w Aneksie II [ROZPORZĄDZENIE DELEGOWANE..., 2015]; $V_{(pr,s)}$, , $V_{(res,s)}$ – odpowiednio miara wielkości ryzyka składki i ryzyka rezerw dla segmentu s, o której mowa w art. 116 [ROZPORZĄDZENIE DELEGOWANE..., 2015].

[ROZPORZĄDZENIE DELEGOWANE..., 2015] - ROZPORZĄDZENIE DELEGOWANE KOMISJI (UE) 2015/35 z dnia 10 października 2014 r. uzupełniające dyrektywę Parlamentu Europejskiego i Rady 2009/138/WE w sprawie podejmowania i prowadzenia działalności ubezpieczeniowej i reasekuracyjnej (Wypłacalność II)

Stąd odchylenie standardowe ryzyka składki i rezerw:

- dla segmentu s jest równe: $\sigma_s = 0.0819$
- dla segmentu t jest równe: $\sigma_t = 0.0694$

Odchylenie standardowe σ_{nl} jest równe:

- $\sigma_{nl} = 0.06558$ - w przypadku $CorrS_{s,t} = 0.5$

- $\sigma_{nl} =$ 0,05367- w przypadku $CorrS_{s,t} = 0$
- $\sigma_{nl} =$ 0,07564 - w przypadku $CorrS_{s,t} = 1$

Na tej podstawie łączny wymóg kapitałowy dla ryzyka składki i rezerw rozważnych segmentów wynosi:

- $SCR_{nl\,prem\,res} =$ 0,86565 - w przypadku $CorrS_{s,t} = 0.5$
- $SCR_{nl\,prem\,res} =$ 0,70841- w przypadku $CorrS_{s,t} = 0$
- $SCR_{nl\,prem\,res} =$ 0,99842 - w przypadku $CorrS_{s,t} = 1$

Zadanie 9.

W celu oceny ryzyka łącznych strat $S = X_1 + X_2$ dwóch linii biznesu postanowiono zastosować warunkową wartość zagrożoną (oczekiwany niedobór) $CVaR_\alpha(S)$ na poziomie ufności $\alpha = 0.995$ i wyznaczyć ją na podstawie próby liczącej 1000 obserwacji z rozkładu S, uzyskanej metodą symulacji. Wiadomo, że:

- Straty dla pierwszej linii biznesu (LOB1) są modelowane za pomocą zmiennej losowej X_1 o rozkładzie logarytmiczno-normalnym z parametrami $\mu = 1$ i $\sigma = 1$ (dystrybuanta $F_1(x) = \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)$, gdzie Φ oznacza dystrybuantę standardowego rozkładu normalnego).
- Straty dla drugiej linii biznesu (LOB2) są modelowane za pomocą zmiennej losowej X_2 o rozkładzie Pareto z dystrybuantą $F_2(x) = 1 - \frac{1}{x^2}, x > 1$.
- Struktura zależności między stratami X_1 i X_2 (wektora losowego (X_1, X_2)) jest modelowana za pomocą kopuli Galambos z parametrem $\theta = 1.285$.

Wartości statystyk pozycyjnych $S_{(981:1000)}, \dots, S_{(1000:1000)}$ dla uzyskanej próby podano w poniższej tabeli, przy czym wartość $\boldsymbol{B}$ odpowiada następującej parze wygenerowanej z zastosowanej kopuli: $(u, v) = (0.998\,,\ 0.999)$.

981	*982*	*983*	*984*	*985*	*986*	*987*	*988*	*989*	*990*
31,90	32,50	32,75	33,03	33,20	33,30	34,17	34,64	35,11	36,98

991	*992*	*993*	*994*	*995*	*996*	*997*	*998*	*999*	*1000*
37,04	40,05	40,06	43,59	45,12	51,78	58,64	78,06	***B***	87,78

a) Oszacować $VaR_{0.995}(S)$

b) Oszacować $CVaR_{0.995}(S)$.

c) Czy jest możliwe, aby $CVaR_{0.995}(S) > CVaR_{0.995}(X_1) + CVaR_{0.995}(X_2)$? Czy taka nierówność może być prawdziwa w przypadku wartości zagrożonej? Odpowiedź uzasadnić.

Odpowiedzi:

....

Odp. a)

$$VaR_{0.995}(S) = 45.12$$

....

Odp. b)

$$CVaR_{0.995}(S) = 71.242$$

....

Odp. c)

W przypadku miary CVaR taka nierówność jest niemożliwa (CVaR jest koherentną miarą ryzyka).

W przypadku VaR może być prawdziwa (VaR nie jest miarą subaddytywną)

Rozwiązanie:

VaR:
$VaR_{0.995}(S) = 45.12$

CVaR:
Należy wyznaczyć $\boldsymbol{B}$
$\boldsymbol{B} = F_1^{-1}(0.998\ ) + F_2^{-1}(\ 0.999)$

$F_1(x) = u$, czyli $\Phi^{-1}(u) = \frac{\ln(x)-\mu}{\sigma}$, stąd $F_1^{-1}(u) = \exp(\sigma\,\Phi^{-1}(u) + \mu)$
$F_1^{-1}(0.998) = 48.33$

$F_2(x) = 1 - \frac{1}{x^2}$, stąd $F_2^{-1}(v) = \sqrt{\frac{1}{1-v}}$, zatem $F_2^{-1}(0.999) = 31{,}62$
Zatem otrzymujemy $\boldsymbol{B} = 79.95$

Na tej podstawie:
$CVaR_{0.995}(S) = 71.242$

Zadanie 10.
Oszacowano metodą największej wiarygodności następujące dwa modele dla liczby szkód K zgłaszanych przez kierowców w ciągu jednego roku.

Model: MPoisson
W modelu tym założono, że liczba szkód podlega rozkładowi Poissona. Uzyskano następujące wyniki oszacowania:

Fitting of the distribution ' pois ' by maximum likelihood
Parameters :

	estimate	Std. Error
lambda	1.003128	0.00391259

Loglikelihood: -86810.15 AIC: 173622.3 BIC: 173631.4

Model: MZPoisson
W tym modelu założono, że liczba szkód ma zaliczany do klasy rozkładów (*a,b,1*) zmodyfikowany w zerze rozkład Poissona (*ZM Poisson*). Uzyskano następujące oszacowania:

Fitting of the distribution ' poisZM ' by maximum likelihood

Parameters :

	estimate	Std. Error
prob	0.4009587	0.001914532
lambda	1.1378461	0.006520259

Loglikelihood: -86417.2 AIC: 172838.4 BIC: 172856.6

a) Scharakteryzować klasę rozkładów (*a,b,1*). Podać co najmniej 5 rozkładów zaliczanych do tej klasy.
b) Na podstawie przedstawionych wyników wybrać lepszy model dla liczby szkód. Wybór uzasadnić.
c) Obliczyć jaka jest różnica dla następujących prawdopodobieństw: $P(K=0)$, $P(K=1)$ i $P(K>1)$ wyznaczonych za pomocą tych dwóch modeli, tzn. $\left|P^{Poisson}(K=0)-P^{MZPoisson}(K=0)\right|$, itd.

Odpowiedzi:

....

Odp. a)
Niech $p_k = P(K=k),\ k=0,1,\ldots$ będzie funkcją prawdopodobieństwa zmiennej losowej K przyjmującej nieujemne wartości całkowite. Powiemy, że rozkład zmiennej K należy do klasy rozkładów typu (*a, b, 1*), jeżeli istnieją stałe a i b takie, że:

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad k = 2,3,4,\ldots$$

Wartości funkcji prawdopodobieństwa są wyrażone w sposób rekurencyjnie począwszy od $k=1$.
Przykłady: rozkład Poissona, dwumianowy, ujemny dwumianowy, geometryczny, ucięty w zerze Poissona, zmodyfikowany w zerze Poissona, ucięty w zerze dwumianowy, zmodyfikowany w zerze dwumianowy,

....

Odp. b)

Na podstawie kryterium AIC, należy wybrać zmodyfikowany w zerze rozkład Poissona.

....

Odp. c)

$$\left|P^{Poisson}(K=0)-P^{MZPoisson}(K=0)\right| = 0.0342287$$
$$\left|P^{Poisson}(K=1)-P^{MZPoisson}(K=1)\right| = 0.04637$$
$$\left|P^{Poisson}(K>1)-P^{MZPoisson}(K>1)\right| = 0.04841$$

Rozwiązanie:

Dla rozkładu Poissona:

$P^{Poisson}(K=0) = 0.36673$

$P^{Poisson}(K=1) = 0.36788$

$P^{Poisson}(K>1) = 1 - P^{Poisson}(K \le 1) = 0.26539$

Dla zmodyfikowanego w zerze rozkładu Poissona:

$P^{MZPoisson}(K=0) = 0.4009587$

$$p_k^{MZPoisson} = \frac{1-p_0^{MZPoisson}}{1-p_0} \cdot p_k, \; k = 1,2,\ldots.$$

$p_0 = \exp(-1.1378461) = 0.32051$

$p_1 = \frac{(1.1378461)^1}{1}\exp(-1.1378461) = 0.36469$

$p_0^{MZPoisson} = 0.4009587$

Stąd:

$P^{MZPoisson}(K=1) = \frac{1-0.4009587}{1-0.32051} \cdot 0.36469 = 0.32151$

$P^{MZPoisson}(K>1) = 1 - P^{MZPoisson}(K \le 1) = 0.3138$