

**Komisja Egzaminacyjna dla Aktuariuszy**

**LXXXIX Egzamin dla Aktuariuszy**

**Sesja egzaminacyjna w dniu 17 października 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 – oznaczają dewiancję oszacowanego modelu,
  - Null deviance – oznaczają dewiancję modelu zerowego,
  - Deviance – redukcję dewiancji po dodaniu kolejnej zmiennej objaśniającej,
  - Df – stopnie swobody,
  - Sum Sq – suma kwadratów,
  - 'log Lik.' – logarytm wiarygodności.
- c) Dystrybuanta standardowego rozkładu normalnego.

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.500	0.504	0.508	0.512	0.516	0.520	0.524	0.528	0.532	0.536
0.1	0.540	0.544	0.548	0.552	0.556	0.560	0.564	0.567	0.571	0.575
0.2	0.579	0.583	0.587	0.591	0.595	0.599	0.603	0.606	0.610	0.614
0.3	0.618	0.622	0.626	0.629	0.633	0.637	0.641	0.644	0.648	0.652
0.4	0.655	0.659	0.663	0.666	0.670	0.674	0.677	0.681	0.684	0.688
0.5	0.691	0.695	0.698	0.702	0.705	0.709	0.712	0.716	0.719	0.722
0.6	0.726	0.729	0.732	0.736	0.739	0.742	0.745	0.749	0.752	0.755
0.7	0.758	0.761	0.764	0.767	0.770	0.773	0.776	0.779	0.782	0.785
0.8	0.788	0.791	0.794	0.797	0.800	0.802	0.805	0.808	0.811	0.813
0.9	0.816	0.819	0.821	0.824	0.826	0.829	0.831	0.834	0.836	0.839
1	0.841	0.844	0.846	0.848	0.851	0.853	0.855	0.858	0.860	0.862
1.1	0.864	0.867	0.869	0.871	0.873	0.875	0.877	0.879	0.881	0.883
1.2	0.885	0.887	0.889	0.891	0.893	0.894	0.896	0.898	0.900	0.901
1.3	0.903	0.905	0.907	0.908	0.910	0.911	0.913	0.915	0.916	0.918
1.4	0.919	0.921	0.922	0.924	0.925	0.926	0.928	0.929	0.931	0.932
1.5	0.933	0.934	0.936	0.937	0.938	0.939	0.941	0.942	0.943	0.944
1.6	0.945	0.946	0.947	0.948	0.949	0.951	0.952	0.953	0.954	0.954
1.7	0.955	0.956	0.957	0.958	0.959	0.960	0.961	0.962	0.962	0.963
1.8	0.964	0.965	0.966	0.966	0.967	0.968	0.969	0.969	0.970	0.971
1.9	0.971	0.972	0.973	0.973	0.974	0.974	0.975	0.976	0.976	0.977
2	0.977	0.978	0.978	0.979	0.979	0.980	0.980	0.981	0.981	0.982
2.1	0.982	0.983	0.983	0.983	0.984	0.984	0.985	0.985	0.985	0.986
2.2	0.986	0.986	0.987	0.987	0.987	0.988	0.988	0.988	0.989	0.989
2.3	0.989	0.990	0.990	0.990	0.990	0.991	0.991	0.991	0.991	0.992
2.4	0.992	0.992	0.992	0.992	0.993	0.993	0.993	0.993	0.993	0.994
2.5	0.994	0.994	0.994	0.994	0.994	0.995	0.995	0.995	0.995	0.995
2.6	0.995	0.995	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996
2.7	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997
2.8	0.997	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
2.9	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.999	0.999	0.999
3	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999

**Zadanie 1.**

W tabeli 1.1 przedstawiono roczne dane dotyczące rezygnacji z odnowienia polisy w podziale na 3 segmenty taryfowe i 2 kanały sprzedaży. Pokazano stan na początku roku oraz liczbę rezygnacji w ciągu roku (ze stanu początkowego).

Tab. 1.1

		Początek roku		Rezygnacje	
		Kanał sprzedaży		Kanał sprzedaży	
		A	B	A	B
Segment	S1	60000	100000	3000	6000
	S2	100000	20000	2500	500
	S3	50000	400000	100	20000

Wykorzystując przedstawione dane, modelowano prawdopodobieństwo rezygnacji dla indywidualnych umów w zależności od segmentu i kanału sprzedaży z wykorzystaniem uogólnionych modeli liniowych (GLM). Oszacowano dwa następujące modele:

**Model M1:**

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.95062	0.01710	-172.56	<2e-16 ***
SegmentS2	-0.74945	0.02339	-32.04	<2e-16 ***
SegmentS3	-0.20156	0.01454	-13.87	<2e-16 ***
KanałB	0.20220	0.01984	10.19	<2e-16 ***

---

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

Null deviance: 2094.257 on 5 degrees of freedom

Residual deviance: 86.814 on 2 degrees of freedom

AIC: 150.86

**Model M2:**

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.94444	0.01873	-157.190	< 2e-16 ***
SegmentS2	-0.71912	0.02759	-26.066	< 2e-16 ***
SegmentS3	-0.94738	0.10274	-9.221	< 2e-16 ***
KanałB	0.19290	0.02298	8.394	< 2e-16 ***
SegmentS2:KanałB	-0.19290	0.05468	-3.528	0.000419 ***
SegmentS3:KanałB	0.75448	0.10385	7.265	3.73e-13 ***

---

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

Null deviance: 2.0943e+03 on 5 degrees of freedom

Residual deviance: -6.3172e-11 on 0 degrees of freedom

AIC: 68.047

- a) (1p.) Wskaż jaki rozkład przyjęto dla zmiennej objaśnianej i podaj postać odpowiadającej jej kanonicznej funkcji łączącej (linku kanonicznego).
- b) (2p.) W oszacowanych modelach zastosowano kodowanie zero-jedynkowe. Wskaż kategorie referencyjne i podaj macierz modelu (*design matrix*) dla **M1** i dla **M2**.
- c) (2p.) Wykorzystując oba modele (**M1** i **M2**) oszacuj prawdopodobieństwo rezygnacji dla umowy z drugiego segmentu (S2) i kanału sprzedaży B. Porównaj wyniki z częstością rezygnacji dla tej grupy umów (tj. segment S2 i kanał B) wyznaczoną na podstawie danych (tab. 1.1). Skomentuj wynik porównania dla **M2** (tzn. wyjaśnij dlaczego uzyskano takie oszacowanie prawdopodobieństwa z wykorzystaniem tego modelu).

### Odpowiedzi

#### Odp. a)

Dla zmiennej objaśnianej przyjęto rozkład zero-jedynkowy.

Link kanoniczny:  $g(p_i) = \ln \frac{p_i}{1-p_i}$

#### Odp. b)

Kategorie referencyjne:

- dla segmentu: S1 (SegmentS1),
- dla kanału sprzedaży: A (KanałA)

Macierz modelu dla **M1**:

(Intercept)	SegmentS2	SegmentS3	KanałB
1	0	0	0
1	0	0	1
1	1	0	0
1	1	0	1
1	0	1	0
1	0	1	1

Macierz modelu dla **M2**:

(Intercept)	SegmentS2	SegmentS3	KanałB	SegmentS2: KanałB	SegmentS3: KanałB
1	0	0	0	0	0
1	0	0	1	0	0
1	1	0	0	0	0
1	1	0	1	1	0
1	0	1	0	0	0
1	0	1	1	0	1

---

.....

**Odp. c)**

**Model M1:**

Wartość predyktora liniowego:

$$\ln \frac{p_{S2,B}}{1-p_{S2,B}} = -2.95062 - 0.74945 + 0.20220 = -3.497871,$$

$$\text{stąd } p_{S2,B} = 0.02937287$$

**Model M2:**

Wartość predyktora liniowego:

$$\ln \frac{p_{S2,B}}{1-p_{S2,B}} = -2.94444 - 0.71912 + 0.19290 - 0.19290 = -3.663562,$$

$$\text{stąd } p_{S2,B} = 0.025$$

Częstość rezygnacji dla segmentu S2 i kanału B:  $\frac{500}{20000} = 0.025$ .

Prawdopodobieństwo oszacowane za pomocą modelu **M2** równa się częstości, ponieważ jest to model nasycony.

**Zadanie 2.**

- a) (2p.) Jakie są najważniejsze zalety i ograniczenia jądrowej estymacji funkcji gęstości w porównaniu z innymi technikami, takimi jak histogramy czy estymatory parametryczne?
- b) (3p.) Z pewnej polisy zbiorowej zanotowano następujące wypłaty (w tys. zł): 25, 30, 35, 35, 37, 39, 45, 47, 49, 55. Wykorzystując estymację jądrową z jądrem jednostajnym o stałej wygładzania 10, oszacuj prawdopodobieństwo wypłaty powyżej 40 tys. zł.

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

Zalety estymacji jądrowej funkcji gęstości:

- Metoda nieparametryczna, nie wymaga założenia konkretnego rozkładu, co pozwala na bardziej ogólną analizę danych.
- Za jej pomocą otrzymuje się gładką funkcję gęstości, która może lepiej odzwierciedlać rzeczywisty rozkład danych niż histogramy, szczególnie gdy dane są mało liczne lub mają skomplikowany rozkład.
- Dzięki stałej wygładzania (szerokości jądra) można kontrolować stopień wygładzenia estymowanej gęstości. Większa wartość stałej prowadzi do bardziej wygładzonej funkcji gęstości, podczas gdy mniejsza bardziej precyzyjnie odwzorowuje dane.
- Pozwala na porównywanie rozkładów różnych zestawów danych.

Ograniczenia jądrowej estymacji funkcji gęstości:

- Wymaga wyboru odpowiedniego jądra (np. gaussowskiego, Epanechnikowa) oraz stałej wygładzania. Dobór tych parametrów może być subiektywny i wpływać na wyniki, a niewłaściwy ich wybór może prowadzić do błędnej estymacji rozkładu danych.
- Może być wymagająca obliczeniowo, szczególnie przy dużej ilości danych.
- Może niedokładnie odwzorować ogon rozkładu danych.

**Odp. b)**

Jądro jednostajne o stałej wygładzania 10:

$$k_y(x) = \begin{cases} 0, & x < y - 10 \\ \frac{1}{2 \cdot 10}, & y - 10 \leq x \leq y + 10 \\ 0, & x > y + 10 \end{cases}$$

Stąd otrzymujemy:

$$\begin{aligned} P(X > 40) &= 2 \cdot \frac{1}{20} \cdot 5 \cdot 0.1 + \frac{1}{20} \cdot 7 \cdot 0.1 + \frac{1}{20} \cdot 9 \cdot 0.1 + \frac{1}{20} \cdot 15 \cdot 0.1 + \frac{1}{20} \cdot 17 \cdot 0.1 \\ &+ \frac{1}{20} \cdot 19 \cdot 0.1 + \frac{1}{20} \cdot 20 \cdot 0.1 = 0.485 \end{aligned}$$

Zobacz podrozdział 14.6 w: “Loss Models: From Data to Decisions”, 5th edition - S.A. Klugman, H.H Panjer, G.E. Willmot, Wiley, 2019.

**Zadanie 3.**

- a) (3p.) Przedstaw koncepcję modelu DGLM (*Double Generalized Linear Model*) oraz krótko omów sposób estymacji jego parametrów.
- b) (2p.) Wysokość pojedynczej szkody (zmienna *clm.incurred*) w pewnym portfelu ubezpieczeń AC modelowano z uwzględnieniem następujących zmiennych objaśniających:
- ccm* – pojemność silnika w  $\text{cm}^3$  (zmienna ilościowa),
  - nb.rb* – zmienna jakościowa przyjmująca dwie kategorie:
    - NB – nowa polisa;
    - RB – wznowiona polisa,
  - driver.gender* – płeć kierowcy (zmienna jakościowa: *Female, Male*).

Wstępna analiza danych wykazała, że wariancja wysokości pojedynczej szkody dla nowych polis była ponad półtora raza większa od wariancji dla polis wznowionych. W związku z tym, na podstawie zbioru uczącego oszacowano *Double Generalized Linear Model*, w którym przyjęto rozkład gamma dla zmiennej objaśnianej oraz link logarytmiczny dla obydwu modeli. Uzyskano następujące wyniki:

Call: `dglm(formula = clm.incurred ~ ccm + nb.rb + driver.gender, dformula = ~nb.rb + ccm, family = Gamma(link = "log"), data = zbior.uczacy)`

Mean Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	6.906079e+00	1.201117e-01	57.4971318	0.000000000
<i>ccm</i>	4.048551e-05	6.569563e-05	0.6162588	0.537768064
<i>nb.rb</i> RB	-1.907909e-01	6.412235e-02	-2.9754202	0.002948009
<i>driver.gender</i> Male	-2.085674e-01	6.370910e-02	-3.2737456	0.001072737

(Dispersion Parameters for Gamma family estimated as below )

Scaled Null Deviance: 3942.734 on 3168 degrees of freedom

Scaled Residual Deviance: 3918.855 on 3165 degrees of freedom

Dispersion Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	3.768577e-01	9.429130e-02	3.996739	6.422112e-05
<i>nb.rb</i> RB	-6.159272e-02	4.975300e-02	-1.237970	2.157273e-01
<i>ccm</i>	9.304519e-05	5.556936e-05	1.674397	9.405254e-02

(Dispersion parameter for Digamma family taken to be 2 )

Scaled Null Deviance: 3632.966 on 3168 degrees of freedom

Scaled Residual Deviance: 3710.548 on 3166 degrees of freedom

Minus Twice the Log-Likelihood: 48391.39

Number of Alternating Iterations: 5

Wiadomo, że podstawowe statystyki opisowe dla zmiennej  $ccm$  w zbiorze uczącym są następujące:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
970	1398	1560	1643	1868	2700

Na podstawie podanych wyników wyznacz:

- możliwe minimalne i możliwe maksymalne oszacowanie parametru dyspersji dla polis w zbiorze uczącym,
- oszacowania wartości oczekiwanej i wariancji zmiennej losowej modelującej wysokość pojedynczej szkody dla nowej polisy wystawionej na samochód o pojemności 2700 cm<sup>3</sup>, którego właścicielem jest mężczyzna.

### Odpowiedzi

#### Odp. a)

Zobacz podrozdział 7.3 w: “Effective Statistical Learning Methods for Actuaries I” - M. Denuit, D. Hainaut, J. Trufin, Springer, 2019.

#### Odp. b)

Minimalne:

$$\exp(0.3768577 - 0.06159272 + 0.00009304519 \cdot 970) = 1.500081$$

Maksymalne:

$$\exp(0.3768577 + 0.00009304519 \cdot 2700) = 1.874008$$

Oszacowanie wartości oczekiwanej:

$$\exp(6.906079 + 0.00004048551 \cdot 2700 - 0.208567) = 903.9941$$

Oszacowanie wariancji:

$$\text{Parametr dyspersji: } \exp(0.3768577 + 0.00009304519 \cdot 2700) = 1.874008$$

stąd otrzymujemy następujące oszacowanie wariancji:

$$1.874008 \cdot 903.9941^2 = 1531449$$

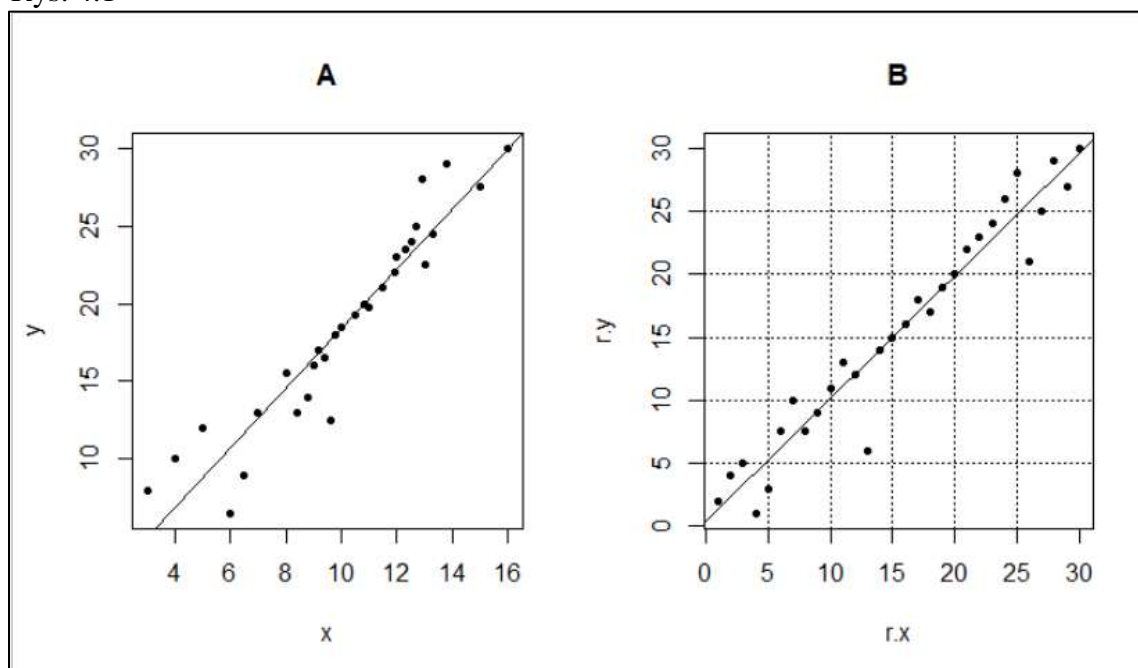


**Zadanie 4.**

Zależność między dwiema zmiennymi  $X$  i  $Y$  analizowano na podstawie 30-sto elementowej próby:  $(x_1, y_1) \dots (x_{30}, y_{30})$ . Przy czym w próbie tej nie powtarzają się zarówno wartości  $x_i$ , jak i wartości  $y_i$ . Na rysunku 4.1 przedstawiono wykresy rozrzutu (diagramy korelacyjne) dla danych (wykres **A**) oraz dla ich rang (wykres **B**).

- a) (2p.) Na wykresie **B** liczba niezgodnych par punktów wynosi 25. Oblicz współczynnik korelacji rang Kendalla.

Rys. 4.1



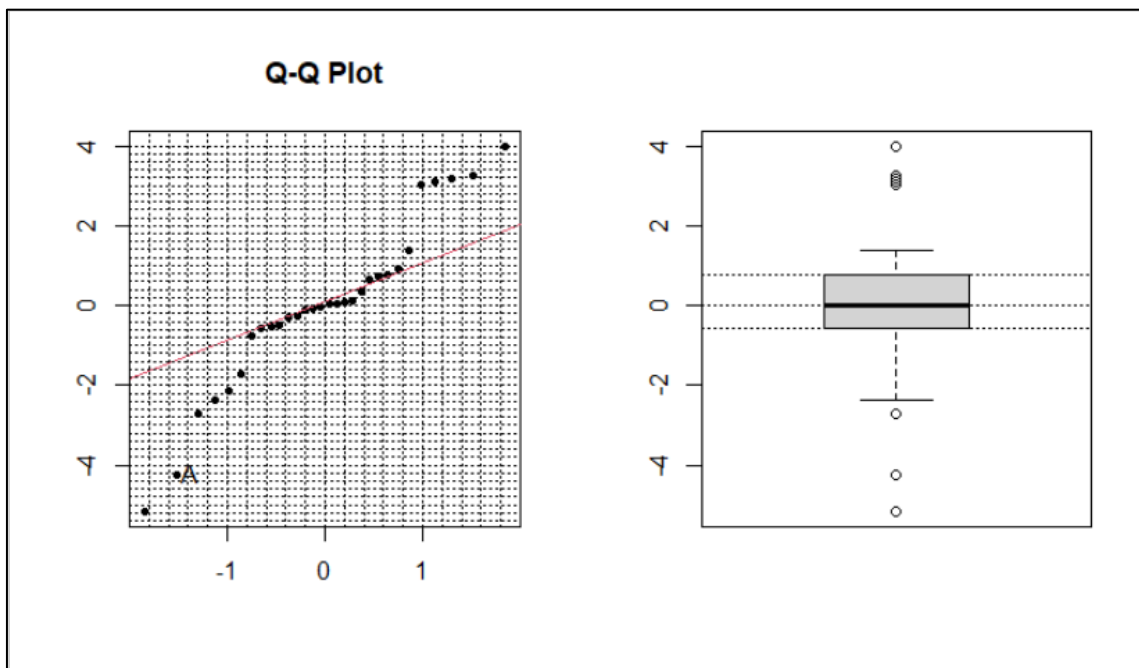
- b) (3p.) Wykorzystując próbę:  $(x_1, y_1) \dots (x_{30}, y_{30})$  oszacowano liniowy model regresji zmiennej  $Y$  względem  $X$ . Otrzymano następujące reszty z tego modelu:

$i$	1	2	3	4	5	6	7	8	9	10
$\hat{\varepsilon}_i$	3.02	3.1	3.18	-2.71	-4.24	0.33	0.91	-2.36	-2.13	-0.51
$i$	11	12	13	14	15	16	17	18	19	20
$\hat{\varepsilon}_i$	0.1	-0.78	-5.17	-0.05	0.07	-0.1	0.03	-0.56	-0.32	-0.09
$i$	21	22	23	24	25	26	27	28	29	30
$\hat{\varepsilon}_i$	0.72	0.64	0.76	1.38	3.99	-1.7	-0.28	3.26	-0.55	0.03

Na rysunku 4.2 przedstawiono wykresy kwantyl-kwantyl (przy założeniu rozkładu normalnego) oraz pudełkowy dla tych reszt.

- Wyjaśnij jakie wartości są na osiach wykresu kwantyl-kwantyl.
- Podaj współrzędne punktu A wskazanego na wykresie kwantyl-kwantyl.
- Na podstawie zamieszczonych wykresów (rys.4.2) wypowiedź się na temat normalności rozkładu reszt. Podaj jeden argument za i jeden przeciw temu założeniu.

Rys. 4.2



### Odpowiedzi

.....

**Odp. a)**

$$\tau = 2 \frac{P - Q}{n(n - 1)}$$

gdzie:

$n$  – liczba obserwacji

$P$  – liczba par zgodnych

$Q$  – liczba par niezgodnych

$$P = \frac{n(n - 1)}{2} - Q = \frac{30 \cdot 29}{2} - 25 = 410$$

$$\tau = 2 \frac{P - Q}{n(n - 1)} = 2 \frac{410 - 25}{30 \cdot 29} = 0.885$$

.....

**Odp. b)**

Na osi X są kwantyle standaryzowanego rozkładu normalnego  $u_{\frac{i}{31}}$ , gdzie  $i = 1, \dots, 30$ , natomiast na osi Y są uporządkowane reszty.

Współrzędne punktu A:

–  $x_A = u_{\frac{2}{31}}$ . Korzystając z zamieszczonej tablicy C, otrzymujemy  $x_A = -1.54$ .

–  $x_B = -4.54$

Zatem otrzymujemy:  $A = (-1.54, -4.54)$

---

Reszty nie wykazują rozkładu normalnego, co jest sygnalizowane przez odmienny kształt wykresu kwantyl-kwantyl w ogonach w porównaniu do kształtu oczekiwanego. Jako pewien argument za przyjęciem rozkładu normalnego można podać względnie symetryczny wykres pudełkowy oraz charakterystyczny kształt wykresu kwantyl-kwantyl poza obszarami ogonów.

**Zadanie 5.**

- a) (1p.) Podaj definicję procesu GARCH( $p, q$ ).
- b) (2p.) Do jakich celów służą modele klasy GARCH?
- c) (2p.) Na podstawie szeregu czasowego liczącego 1837 obserwacji stóp zwrotu ( $r_t$ ) PZU (od 2016-06-06 do 2023-10-05) oszacowano model ARMA(0,0)-GARCH(1,1) z gaussowskimi innowacjami. Uzyskano następujące wyniki:

## Optimal Parameters

	Estimate	Std. Error	t value	Pr(> t )
mu	0.000366	0.000386	0.94631	0.343988
omega	0.000021	0.000006	3.65831	0.000254
alpha1	0.077897	0.014377	5.41818	0.000000
beta1	0.854787	0.028546	29.94420	0.000000

Stopy zwrotu  $r_t$  i oszacowane warunkowe odchylenia standardowe  $\hat{\sigma}_t$  dla 3 ostatnich obserwacji przedstawia tabela 5.1.

Tab. 5.1

$t$	1835	1836	1837
$r_t$	0.00497761	0.00692729	-0.00049322
$\hat{\sigma}_t$	0.01856621	0.01781931	0.01720450

Na podstawie powyższych informacji oszacuj na okres  $t = 1838$ :

- warunkowe odchylenie standardowe,
- jednodniowy VaR (wartość zagrożona), przyjmując poziom tolerancji  $\alpha = 0.05$ .

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

Zobacz podrozdział 4.2 w: “Quantitative Risk Management: Concepts, Techniques and Tools”, revised edition - A. McNeil, R. Frey, P. Embrechst, Princeton, 2015

**Odp. b)**

Zobacz podrozdział 4.2 w: “Quantitative Risk Management: Concepts, Techniques and Tools”, revised edition - A. McNeil, R. Frey, P. Embrechst, Princeton, 2015

**Odp. c)**

Warunkowe odchylenie standardowe:

$$\begin{aligned}\hat{\sigma}_{1838}^2 &= 0.00002122216 + 0.07789685 \cdot (-0.00049322 - 0.0003655178)^2 \\ &\quad + 0.8547871 \cdot 0.01720450^2 = 0.0002742922 \\ \hat{\sigma}_{1838} &= \sqrt{0.0002742922} = 0.01656177\end{aligned}$$

Jednodniowy VaR:

$$VaR_{0.05} = 0.000366 + 0.01656177 \cdot (-1.64) = -0.0267953$$

**Zadanie 6.**

Wykorzystując dane zwarte w tabeli 6.1, gdzie symbolem (\*) oznaczono obserwacje cenzurowane z góry:

- (3p.) Skonstruuj estymator Nelsona-Åalena dla funkcji przeżycia  $S(x)$ . Uwzględnij poprawkę Kleina-Moeschbergera, przyjmując  $\gamma = 22$ .
- (2p.) Oszacuj wariancję estymatora Nelsona-Åalena dla  $S(2)$ .

Tab. 6.1

$i$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$x_i$	1	2	3*	4	4	4*	4*	5	7*	8	8	8	9	9	9	9	10*	12	12	15*

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

Zobacz podrozdział 14.3 (Przykład 14.8) w: "Loss Models: From Data to Decisions", 5th edition - S.A. Klugman, H.H Panjer, G.E. Willmot, Wiley, 2019.

**Odp. b)**

Zobacz podrozdział 14.3 (Przykład 14.9) w: "Loss Models: From Data to Decisions", 5th edition - S.A. Klugman, H.H Panjer, G.E. Willmot, Wiley, 2019.

**Zadanie 7.**

- a) (**1p.**) Krótko przedstaw ideę statystycznych metod uczenia zespołowego (*Ensemble Statistical Learning*).
- b) (**1p.**) Wymień co najmniej trzy takie metody wykorzystujące drzewa (*Tree Ensemble Methods*)
- c) (**3p.**) Opisz jedną metodę spośród wymienionych w punkcie b).

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

Zobacz np. podrozdział 8.2 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 8.2 w “An Introduction to Statistical Learning with Applications in R” - G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2021

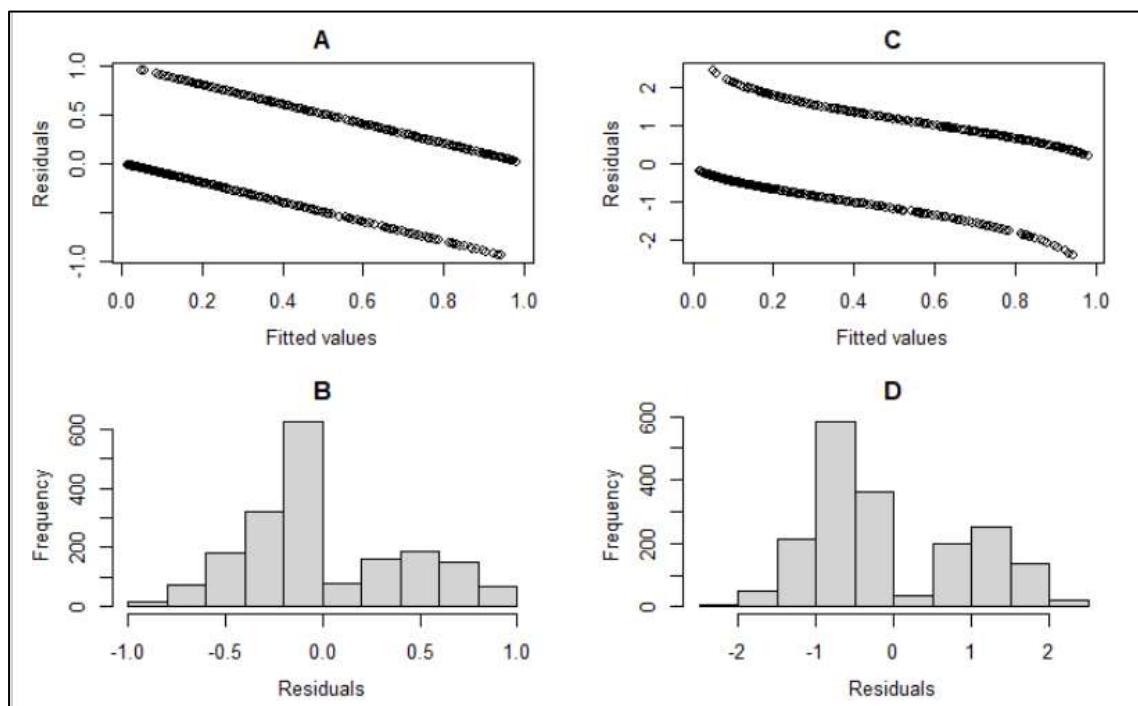
**Odp. c)**

Zobacz np. podrozdział 8.2 w “An Introduction to Statistical Learning with Applications in R” - G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2021

**Zadanie 8.**

- a) (1p.) W jaki sposób oblicza się reszty dewiancyjne (*deviance residuals*)?
- b) (2p.) Dlaczego są one często preferowaną formą reszt w modelach GLM? Jakie są główne zalety ich wykorzystania w porównaniu z innymi rodzajami reszt?
- c) (2p.) Na rysunku 8.1 przedstawiono wykresy reszt zwykłych i dewiancyjnych w zależności od wartości dopasowanych oraz ich histogramy. Reszty te odpowiadają temu samemu oszacowanemu uogólnionemu modelowi liniowemu. Wskaż i uzasadnij, jaki rozkład miała zmienna objaśniana w tym modelu oraz któremu rodzajowi reszt odpowiadają poszczególne rysunki A, B, C i D.

Rys. 8.1

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

Zobacz podrozdział 4.9.2 w: “Effective Statistical Learning Methods for Actuaries I” - M. Denuit, D. Hainaut, J. Trufin, Springer, 2019.

**Odp. b)**

Zobacz podrozdział 4.9.2 w: “Effective Statistical Learning Methods for Actuaries I” - M. Denuit, D. Hainaut, J. Trufin, Springer, 2019.

**Odp. c)**

Zmienna objaśniana ma rozkład zero-jedynkowy.

Wykresy A i B odpowiadają zwykłymi resztom, a C i D resztom dewiancyjnym. W rozważanym przypadku, zwykłe reszty nie mogą być większe od 1, ani mniejsze od -1 (są to różnice między zaobserwowanymi wartościami zmiennej zależnej tj. 0 lub 1, a oszacowanymi prawdopodobieństwami).

**Zadanie 9.**

- a) (3p.) Opisz koncepcję redukcji wariancji w metodzie Monte Carlo za pomocą metody próbkowania ważonego (metody *IS - importance sampling*).
- b) (2p.) Zmienna losowa  $X$  ma standardowy rozkład normalny ( $X \sim N(0, 1)$ ). Wykorzystując przesunięty rozkład wykładniczy z parametrem  $\lambda = 1$  i odpowiednio dobranym parametrem przesunięcia, wyznacz estymator Monte Carlo wykorzystujący próbkowanie ważne (*importance sampling estimator*) dla prawdopodobieństwa  $P(X > 3.5)$ .

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

Zobacz podrozdział 9.2.2 w “Reinsurance: Actuarial and Statistical Aspects” - H. Albrecher, J. Beirlant, J. Teugels, Wiley, 2017

**Odp. b)**

*Importance sampling estimator:*

$$\frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{g(X_i)} h(X_i),$$

gdzie próba  $X_1, X_2, \dots, X_n$  jest losowana z rozkładu  $g$ .

W zadaniu

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$$g(x) = e^{-(x-3.5)}$$

$$h(X_i) = \mathbb{I}_{X_i > 3.5} = 1.$$

Stąd

$$\frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{g(X_i)} h(X_i) = \frac{1}{n} \sum_{i=1}^n \frac{e^{-\frac{1}{2}X_i^2 + X_i - 3.5}}{\sqrt{2\pi}}.$$



**Zadanie 10.**

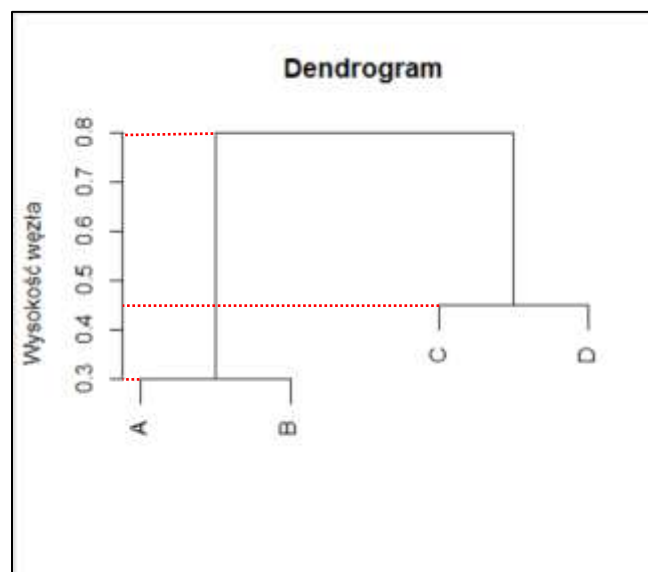
- a) (3p.) Przedstaw przebieg procesu grupowania hierarchicznego według algorytmu aglomeracyjnego.
- b) (2p.) Analizowano podobieństwo profili ryzyka czterech portfeli ubezpieczeń: A, B, C i D z wykorzystaniem aglomeracyjnego grupowania hierarchicznego, w którym odległość między skupieniami (klastrami) mierzono metodą pełnego wiązania (*complete linkage*). Każdy portfel opisano pięcioma zmiennymi diagnostycznymi. W oparciu o zestandaryzowane wartości tych zmiennych otrzymano następującą macierz niepodobieństwa:

	A	B	C	D
A		0.3	0.4	0.7
B	0.3		0.5	0.8
C	0.4	0.5		0.45
D	0.7	0.8	0.45	

Narysuj dendrogram, będący wynikiem tego grupowania. Zaznacz na nim wysokość węzłów i podpisz liście.

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

Zobacz np. podrozdział 12.4.2 w “An Introduction to Statistical Learning with Applications in R” - G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2021

**Odp. b)**

---

**Sesja egzaminacyjna w dniu 17 października 2023 r.****Modelowanie****Arkusz ocen**

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