Komisja Egzaminacyjna dla Aktuariuszy

LXXXIII Egzamin dla Aktuariuszy

Sesja egzaminacyjna w dniu 5 października 2021 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,
- Df – stopnie swobody,
- Sum Sq – suma kwadratów,
- Residual standard error – odchylenie standardowe reszt .

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

e) 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

f) Wartości F_{r_1,r_2} rozkładu F spełniające warunek $P(F \geq F_{r_1,r_2}) = 0.05$

r_2 \ r_1	***1***	***2***	***3***	***4***	***5***	***6***	***7***
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103
700	3.855	3.009	2.618	2.385	2.227	2.112	2.023
800	3.853	3.007	2.616	2.383	2.225	2.110	2.021
900	3.852	3.006	2.615	2.382	2.224	2.109	2.020
1000	3.851	3.005	2.614	2.381	2.223	2.108	2.019
∞	3.844	2.998	2.607	2.374	2.216	2.101	2.012

r_2 \ r_1	***8***	***9***	***10***	***11***	***12***	***13***	***14***
1	238.883	240.543	241.882	242.983	243.906	244.690	245.364
2	19.371	19.385	19.396	19.405	19.413	19.419	19.424
3	8.845	8.812	8.786	8.763	8.745	8.729	8.715
4	6.041	5.999	5.964	5.936	5.912	5.891	5.873
5	4.818	4.772	4.735	4.704	4.678	4.655	4.636
6	4.147	4.099	4.060	4.027	4.000	3.976	3.956
7	3.726	3.677	3.637	3.603	3.575	3.550	3.529
8	3.438	3.388	3.347	3.313	3.284	3.259	3.237
9	3.230	3.179	3.137	3.102	3.073	3.048	3.025
10	3.072	3.020	2.978	2.943	2.913	2.887	2.865
100	2.032	1.975	1.927	1.886	1.850	1.819	1.792
700	1.952	1.893	1.844	1.802	1.766	1.734	1.706
800	1.950	1.892	1.843	1.801	1.764	1.732	1.704
900	1.949	1.890	1.841	1.799	1.763	1.731	1.703
1000	1.948	1.889	1.840	1.798	1.762	1.730	1.702
∞	1.941	1.882	1.833	1.791	1.755	1.723	1.694

g) 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

h) Wartości dystrybuanty w punkcie 200 (tzn. $F(200)$) dla rozkładu $Gamma(\alpha,\lambda)$ (dla wybranych parametrów α,λ).
Uwaga! Przyjęto parametryzację, w której $E(X)=\frac{\alpha}{\lambda}$, $D^2(X)=\frac{\alpha}{\lambda^2}$

α \ λ	***0.0038***	***0.0039***	***0.004***	***0.0041***	***0.0042***	***0.0043***	***0.0044***	***0.0045***	***0.0046***	***0.0047***
0.700	0.6811	0.6888	0.6963	0.7036	0.7107	0.7176	0.7244	0.7309	0.7373	0.7435
0.701	0.6806	0.6883	0.6958	0.7031	0.7102	0.7172	0.7239	0.7304	0.7368	0.7431
0.702	0.6801	0.6878	0.6953	0.7026	0.7097	0.7167	0.7234	0.7300	0.7364	0.7426
0.703	0.6796	0.6873	0.6948	0.7021	0.7093	0.7162	0.7230	0.7295	0.7359	0.7422
0.704	0.6790	0.6868	0.6943	0.7017	0.7088	0.7157	0.7225	0.7291	0.7355	0.7417
0.705	0.6785	0.6863	0.6938	0.7012	0.7083	0.7153	0.7220	0.7286	0.7350	0.7413
0.706	0.6780	0.6858	0.6933	0.7007	0.7078	0.7148	0.7216	0.7282	0.7346	0.7408
0.707	0.6775	0.6853	0.6928	0.7002	0.7074	0.7143	0.7211	0.7277	0.7341	0.7404
0.708	0.6770	0.6848	0.6924	0.6997	0.7069	0.7138	0.7206	0.7272	0.7337	0.7400
0.709	0.6765	0.6843	0.6919	0.6992	0.7064	0.7134	0.7202	0.7268	0.7332	0.7395
0.710	0.6760	0.6838	0.6914	0.6987	0.7059	0.7129	0.7197	0.7263	0.7328	0.7391
0.711	0.6755	0.6833	0.6909	0.6983	0.7054	0.7124	0.7192	0.7259	0.7323	0.7386
0.712	0.6750	0.6828	0.6904	0.6978	0.7050	0.7120	0.7188	0.7254	0.7319	0.7382
0.713	0.6745	0.6823	0.6899	0.6973	0.7045	0.7115	0.7183	0.7250	0.7314	0.7377
0.714	0.6740	0.6818	0.6894	0.6968	0.7040	0.7110	0.7178	0.7245	0.7310	0.7373
0.715	0.6735	0.6813	0.6889	0.6963	0.7035	0.7105	0.7174	0.7240	0.7305	0.7368
0.716	0.6729	0.6808	0.6884	0.6958	0.7030	0.7101	0.7169	0.7236	0.7301	0.7364
0.717	0.6724	0.6803	0.6879	0.6953	0.7026	0.7096	0.7165	0.7231	0.7296	0.7360
0.718	0.6719	0.6798	0.6874	0.6948	0.7021	0.7091	0.7160	0.7227	0.7292	0.7355
0.719	0.6714	0.6793	0.6869	0.6944	0.7016	0.7087	0.7155	0.7222	0.7287	0.7351
0.720	0.6709	0.6788	0.6864	0.6939	0.7011	0.7082	0.7151	0.7218	0.7283	0.7346
0.721	0.6704	0.6783	0.6859	0.6934	0.7006	0.7077	0.7146	0.7213	0.7278	0.7342
0.722	0.6699	0.6778	0.6854	0.6929	0.7002	0.7072	0.7141	0.7208	0.7274	0.7337
0.723	0.6694	0.6773	0.6849	0.6924	0.6997	0.7068	0.7137	0.7204	0.7269	0.7333
0.724	0.6689	0.6768	0.6845	0.6919	0.6992	0.7063	0.7132	0.7199	0.7265	0.7329
0.725	0.6684	0.6763	0.6840	0.6914	0.6987	0.7058	0.7127	0.7195	0.7260	0.7324
0.726	0.6679	0.6758	0.6835	0.6910	0.6982	0.7054	0.7123	0.7190	0.7256	0.7320
0.727	0.6674	0.6753	0.6830	0.6905	0.6978	0.7049	0.7118	0.7185	0.7251	0.7315
0.728	0.6669	0.6748	0.6825	0.6900	0.6973	0.7044	0.7113	0.7181	0.7247	0.7311
0.729	0.6664	0.6743	0.6820	0.6895	0.6968	0.7039	0.7109	0.7176	0.7242	0.7306
0.730	0.6658	0.6738	0.6815	0.6890	0.6963	0.7035	0.7104	0.7172	0.7238	0.7302
0.731	0.6653	0.6733	0.6810	0.6885	0.6959	0.7030	0.7099	0.7167	0.7233	0.7297
0.732	0.6648	0.6728	0.6805	0.6880	0.6954	0.7025	0.7095	0.7163	0.7229	0.7293
0.733	0.6643	0.6723	0.6800	0.6876	0.6949	0.7020	0.7090	0.7158	0.7224	0.7288
0.734	0.6638	0.6718	0.6795	0.6871	0.6944	0.7016	0.7085	0.7153	0.7220	0.7284
0.735	0.6633	0.6713	0.6790	0.6866	0.6939	0.7011	0.7081	0.7149	0.7215	0.7280

i) Wartości dystrybuanty w punkcie 60 (tzn. $F(60)$) dla rozkładu $Gamma(\alpha, \lambda)$ (dla wybranych parametrów α, λ).
Uwaga! Przyjęto parametryzację, w której $E(X) = \frac{\alpha}{\lambda}$, $D^2(X) = \frac{\alpha}{\lambda^2}$

α \ λ	***0.0075***	***0.0076***	***0.0077***	***0.0078***	***0.0079***	***0.008***	***0.0081***	***0.0082***	***0.0083***	***0.0084***
0.325	0.7787	0.7811	0.7834	0.7857	0.7879	0.7902	0.7924	0.7945	0.7967	0.7988
0.326	0.7780	0.7804	0.7827	0.7850	0.7873	0.7895	0.7917	0.7939	0.7960	0.7981
0.327	0.7773	0.7797	0.7820	0.7843	0.7866	0.7888	0.7910	0.7932	0.7954	0.7975
0.328	0.7766	0.7790	0.7813	0.7836	0.7859	0.7881	0.7904	0.7925	0.7947	0.7968
0.329	0.7759	0.7783	0.7806	0.7829	0.7852	0.7875	0.7897	0.7919	0.7940	0.7962
0.330	0.7752	0.7776	0.7799	0.7822	0.7845	0.7868	0.7890	0.7912	0.7934	0.7955
0.331	0.7745	0.7769	0.7792	0.7816	0.7839	0.7861	0.7884	0.7906	0.7927	0.7949
0.332	0.7738	0.7762	0.7785	0.7809	0.7832	0.7854	0.7877	0.7899	0.7921	0.7942
0.333	0.7731	0.7755	0.7778	0.7802	0.7825	0.7848	0.7870	0.7892	0.7914	0.7936
0.334	0.7724	0.7748	0.7772	0.7795	0.7818	0.7841	0.7863	0.7886	0.7907	0.7929
0.335	0.7716	0.7741	0.7765	0.7788	0.7811	0.7834	0.7857	0.7879	0.7901	0.7922
0.336	0.7709	0.7734	0.7758	0.7781	0.7805	0.7827	0.7850	0.7872	0.7894	0.7916
0.337	0.7702	0.7727	0.7751	0.7774	0.7798	0.7821	0.7843	0.7866	0.7888	0.7909
0.338	0.7695	0.7720	0.7744	0.7768	0.7791	0.7814	0.7837	0.7859	0.7881	0.7903
0.339	0.7688	0.7713	0.7737	0.7761	0.7784	0.7807	0.7830	0.7852	0.7874	0.7896
0.340	0.7681	0.7706	0.7730	0.7754	0.7777	0.7800	0.7823	0.7846	0.7868	0.7890
0.341	0.7674	0.7699	0.7723	0.7747	0.7770	0.7794	0.7816	0.7839	0.7861	0.7883
0.342	0.7667	0.7692	0.7716	0.7740	0.7764	0.7787	0.7810	0.7832	0.7855	0.7877
0.343	0.7660	0.7685	0.7709	0.7733	0.7757	0.7780	0.7803	0.7826	0.7848	0.7870
0.344	0.7653	0.7678	0.7702	0.7726	0.7750	0.7773	0.7796	0.7819	0.7841	0.7864
0.345	0.7646	0.7671	0.7695	0.7719	0.7743	0.7767	0.7790	0.7812	0.7835	0.7857
0.346	0.7639	0.7664	0.7688	0.7713	0.7736	0.7760	0.7783	0.7806	0.7828	0.7850
0.347	0.7632	0.7657	0.7682	0.7706	0.7730	0.7753	0.7776	0.7799	0.7822	0.7844
0.348	0.7625	0.7650	0.7675	0.7699	0.7723	0.7746	0.7770	0.7792	0.7815	0.7837
0.349	0.7618	0.7643	0.7668	0.7692	0.7716	0.7740	0.7763	0.7786	0.7808	0.7831
0.350	0.7611	0.7636	0.7661	0.7685	0.7709	0.7733	0.7756	0.7779	0.7802	0.7824
0.351	0.7604	0.7629	0.7654	0.7678	0.7702	0.7726	0.7749	0.7772	0.7795	0.7818
0.352	0.7597	0.7622	0.7647	0.7671	0.7696	0.7719	0.7743	0.7766	0.7789	0.7811
0.353	0.7590	0.7615	0.7640	0.7665	0.7689	0.7713	0.7736	0.7759	0.7782	0.7805
0.354	0.7583	0.7608	0.7633	0.7658	0.7682	0.7706	0.7729	0.7753	0.7775	0.7798
0.355	0.7576	0.7601	0.7626	0.7651	0.7675	0.7699	0.7723	0.7746	0.7769	0.7791
0.356	0.7569	0.7594	0.7619	0.7644	0.7668	0.7692	0.7716	0.7739	0.7762	0.7785
0.357	0.7562	0.7587	0.7612	0.7637	0.7661	0.7686	0.7709	0.7733	0.7756	0.7778
0.358	0.7555	0.7580	0.7605	0.7630	0.7655	0.7679	0.7703	0.7726	0.7749	0.7772
0.359	0.7548	0.7573	0.7599	0.7623	0.7648	0.7672	0.7696	0.7719	0.7742	0.7765
0.360	0.7541	0.7566	0.7592	0.7617	0.7641	0.7665	0.7689	0.7713	0.7736	0.7759

Zadanie 1.

W pewnym portfelu ubezpieczeń samochodowych przeprowadzono taryfikację dwumodelową i jednomodelową.

- W taryfikacji dwumodelowej liczbę szkód K_i modelowano za pomocą regresji Poissona z logarytmiczną funkcją łączącą (uogólniony model liniowy, w którym zmienna zależna ma rozkład *Poissona*), natomiast wysokość pojedynczej szkody X_i modelowano z wykorzystaniem uogólnionego modelu liniowego ze zmienną zależną o rozkładzie *gamma*.
- W taryfikacji jednomodelowej składkę czystą (*pure premium*) S_i (iloraz całkowitych roszczeń dla i-tej polisy przez jej ekspozycję) modelowano z wykorzystaniem uogólnionego modelu liniowego ze zmienną zależną o rozkładzie *Tweedie*.

We wszystkich modelach przyjęto logarytmiczną funkcję łączącą: $g(\mu) = \ln(\mu)$.
W modelach tych uwzględniono zmienne objaśniające opisane w poniższej tabeli (w różnych modelach wybrano różny zestaw zmiennych):

Zmienna	*Opis*
region	Region zamieszkania kierowcy. Zmienna jakościowa przyjmująca następujące kategorie: *R0*, *R1*, *R2*, *R3*, *R4*, *R5*, *R6*, *R7* i *R8*.
klasa.ryzyka	Zmienna jakościowa określająca klasę ryzyka. Przyjmuje następujące kategorie: *kl.1*, *kl.2*, *kl.3*, *kl.4*, *kl.5* i *kl.6* (im wyższa klasa, tym większa zniżka za bezszkodowość).
wiek.kierowcy	Zmienna jakościowa określająca wiek kierowcy. Przyjmuje następujące kategorie: *18-33*, *34-38*, *39-43*, *44-50*, *51-60*, *61-69* i *70+*.
wczes.szkody	Liczba wcześniejszych szkód (zmienna ilościowa).
prawo.jazdy	Zmienna jakościowa określająca od ilu lat kierowca posiada prawo jazdy. Przyjmuje następujące kategorie: *1*, *2*, *3*, *4*, *5*, *6*, *7* i *8+*
plec	Zmienna jakościowa określająca płeć kierowcy. Przyjmuje następujące kategorie: *K* - kobieta, *M* – mężczyzna.
stan.cywilny	Zmienna jakościowa określająca stan cywilny kierowcy. Przyjmuje następujące kategorie: *M* – małżeństwo, *S* – singiel, *R* – rozwiedziony/rozwiedziona, *W* – wdowiec/wdowa).
typ.silnika	Zmienna jakościowa określająca typ silnika. Przyjmuje następujące kategorie: *Diesel*, *Benzyna* i *LPG*.
l.siedzen	Zmienna jakościowa określająca liczbę siedzeń w samochodzie. Przyjmuje następujące kategorie: *2*, *3*, *4*, *5* i *6+*.

Poniżej podano wyniki oszacowań tych modeli:

- Model dla K_i:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.1087	0.0719	-15.4138	< 2e-16	***
regionR1	-0.5151	0.1240	-4.1530	3.28e-05	***
regionR2	-0.4158	0.1389	-2.9930	0.002763	**
regionR3	-0.3183	0.0658	-4.8350	1.33e-06	***
regionR4	-0.2878	0.0837	-3.4390	0.000585	***
regionR5	-0.1625	0.0869	-1.8700	0.061470	.
regionR6	-0.0346	0.0685	-0.5040	0.613930	
regionR7	-1.0065	0.1912	-5.2650	1.40e-07	***
regionR8	0.1009	0.0942	1.0710	0.284227	
*klasa.ryzyka*kl.*2*	-0.4934	0.2065	-2.3900	0.016847	*
*klasa.ryzyka*kl.3	-0.1229	0.0637	-1.9290	0.053778	.
*klasa.ryzyka*kl.*4*	-0.2935	0.0841	-3.4910	0.000481	***
*klasa.ryzyka*kl.5	-0.3793	0.0979	-3.8750	0.000106	***
*klasa.ryzyka*kl.*6*	-0.5877	0.0926	-6.3500	2.16e-10	***
*wiek.kierowcy*18-33	0.1564	0.0680	2.3010	0.021369	*
*wiek.kierowcy*34-38	-0.1389	0.0736	-1.8860	0.059262	.
*wiek.kierowcy*39-43	-0.1504	0.0757	-1.9870	0.046908	*
*wiek.kierowcy*51-60	-0.0060	0.0696	-0.0860	0.931725	
*wiek.kierowcy*61-69	-0.3313	0.1186	-2.7940	0.005210	**
*wiek.kierowcy*170+	-0.0836	0.2102	-0.3980	0.690740	
wczes.szkody	0.1285	0.0143	8.9890	< 2e-16	***
*prawo.jazdy*2	-0.2075	0.0686	-3.0260	0.002476	**
*prawo.jazdy*3	-0.3723	0.0809	-4.6020	4.18e-06	***
*prawo.jazdy*4	-0.3037	0.0879	-3.4550	0.000550	***
*prawo.jazdy*5	-0.4698	0.1048	-4.4810	7.43e-06	***
*prawo.jazdy*6	-0.7786	0.1360	-5.7260	1.03e-08	***
*prawo.jazdy*7	-0.5729	0.1634	-3.5070	0.000453	***
*prawo.jazdy*8+	-0.4403	0.2307	-1.9080	0.056328	.

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

- Model dla X_i:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6,5957	0,0459	143,7080	< 2e-16	***
*stan.cywilny*R	-0,0466	0,2208	-0,2110	0.83297	
*stan.cywilny*S	0,4837	0,1613	2,9990	0.00275	**
*stan.cywilny*W	-0,0414	0,2457	-0,1690	0.86617	
*plec*K	0,2254	0,0999	2,2560	0.02417	*
*l.siedzen*2	0,0222	0,0744	0,2980	0.76541	
*l.siedzen*3	0,1526	0,1379	1,1060	0.26866	
*l.siedzen*4	-0,0141	0,1224	-0,1150	0.90823	
*l.siedzen*6+	0,2610	0,1832	1,4250	0.15436	
*typ.silnika*Benzyna	-0,5290	0,2925	-1,8090	0.07065	.
*typ.silnika*LPG	-0,7100	0,5266	-1,3480	0.17774	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Estimated dispersion parameter: 1.9128

- Model dla S_i:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	5.4410	0.1705	31.9170	< 2e-16	***
regionR1	-0.5456	0.2524	-2.1620	0.030629	*
regionR2	-0.6984	0.3023	-2.3100	0.020889	*
regionR3	-0.2589	0.1472	-1.7590	0.078582	.
regionR4	-0.0361	0.1792	-0.2010	0.840372	
regionR5	-0.2399	0.1987	-1.2070	0.227486	
regionR6	0.0001	0.1608	0.0010	0.999570	
regionR7	-0.8986	0.3247	-2.7670	0.005655	**
regionR8	0.1103	0.2278	0.4840	0.628163	
*klasa.ryzyka*kl.2	-1.2590	0.4698	-2.6810	0.007349	**
*klasa.ryzyka*kl.3	-0.1663	0.1475	-1.1270	0.259694	
*klasa.ryzyka*kl.4	-0.4191	0.1818	-2.3050	0.021188	*
*klasa.ryzyka*kl.5	-0.7701	0.2128	-3.6200	0.000296	***
*klasa.ryzyka*kl.6	-0.8717	0.1920	-4.5410	5.63e-06	***
*wiek.kierowcy*18-33	0.1014	0.1573	0.6450	0.519233	
*wiek.kierowcy*34-38	-0.4004	0.1637	-2.4460	0.014432	*
*wiek.kierowcy*39-43	-0.2170	0.1624	-1.3360	0.181510	
wiek.kierowcy 51-60	-0.0674	0.1519	-0.4440	0.657397	
*wiek.kierowcy*61-69	-0.1527	0.2251	-0.6780	0.497623	
*wiek.kierowcy*170+	-0.5290	0.4764	-1.1100	0.266841	
wczes.szkody	0.1670	0.0372	4.4880	7.24e-06	***
*prawo.jazdy*2	-0.2076	0.1619	-1.2820	0.199761	
*prawo.jazdy*3	-0.3129	0.1824	-1.7150	0.086331	.
*prawo.jazdy*4	-0.1653	0.2001	-0.8260	0.408898	
*prawo.jazdy*5	-0.4130	0.2270	-1.8190	0.068908	.
*prawo.jazdy*6	-1.2490	0.2825	-4.4220	9.83e-06	***
*prawo.jazdy*7	-0.1775	0.3104	-0.5720	0.567369	
*prawo.jazdy*8+	-0.2773	0.4302	-0.6450	0.519210	
*plec*K	0.2878	0.1519	1.8950	0.058093	.
*typ.silnika*Benzyna	-0.6323	0.4687	-1.3490	0.177369	
*typ.silnika*LPG	-0.9621	0.7564	-1.2720	0.203421	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Estimated dispersion parameter: 80.91
Estimated index parameter p: 1.6462

Dla kierowcy scharakteryzowanego następującymi wartościami zmiennych objaśniających: *region*: R0; *klasa.ryzyka*: kl.1; *wiek.kierowcy*: 44-50; *wczes.szkody*: 2; *prawo.jazdy*: 1; *plec*: M; *typ.silnika*: Diesel; *stan.*cywilny: M; *l.siedzen*: 5 wyznaczyć składkę czystą i jej wariancję wykorzystując:

a) ***(3p.)*** taryfikację dwumodelową,
b) ***(2p.)*** taryfikację jednomodelową.

Uwaga! Funkcja wariancji wynosi:
- dla rozkładu gamma: $V(\mu) = \mu^2$
- dla rozkładu Tweedie: $V(\mu) = \mu^p$

Odpowiedzi:

....

Odp. a)
Składka czysta: 312.3112
Wariancja: 665846.75

....

Odp. b)
Składka czysta: 322.1444
Wariancja: 1088316.3

Rozwiązanie:

ad. a)
Składka czysta: $P_i = E(Z_i) = E(K_i) \cdot E(X_i)$
Wariancja: $Var(Z_i) = Var(K_i) \cdot \left(E(X_i)\right)^2 + E(K_i) \cdot Var(X_i)$

$E(K_i) = \exp(-1.1087 + 2 \cdot 0.1285) = \exp(-0.8517) = 0.4267.$
$Var(K_i) = E(K_i) = 0.4267.$

$E(X_i) = \exp(6.5957) = 731.9411.$
$Var(X_i) = \phi \cdot \left(E(X_i)\right)^2 = 1.9128 \cdot 731.9411^2 = 1024759.1.$

$P_i = 0.4267 \cdot 731.9411 = 312.3112.$
$Var(Z_i) = 0.4267 \cdot 731.9411^2 + 0.4267 \cdot 1024759.1 = 665846.75.$

ad. b)
Składka czysta:
$P_i = E(S_i) = \exp(5.4410 + 2 \cdot 0.1670) = \exp(5.7750) = 322.1444.$

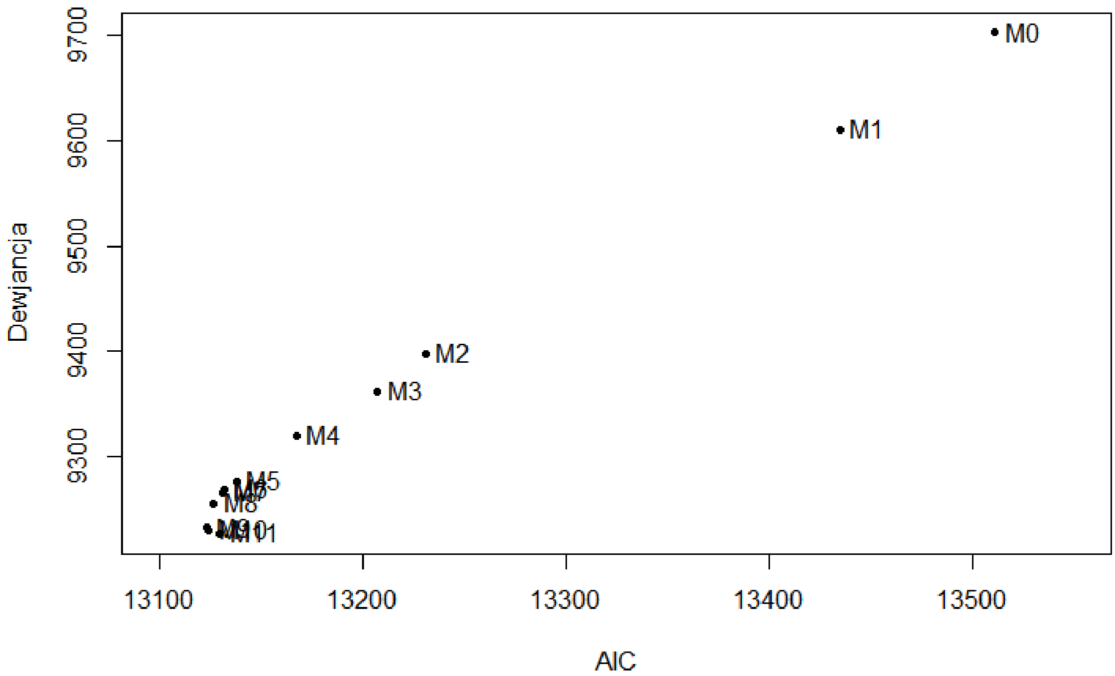
Wariancja:
$Var(S_i) = \phi \cdot \left(E(S_i)\right)^p = 80.91 \cdot (322.1444)^{1.6462} = 1088316.3.$

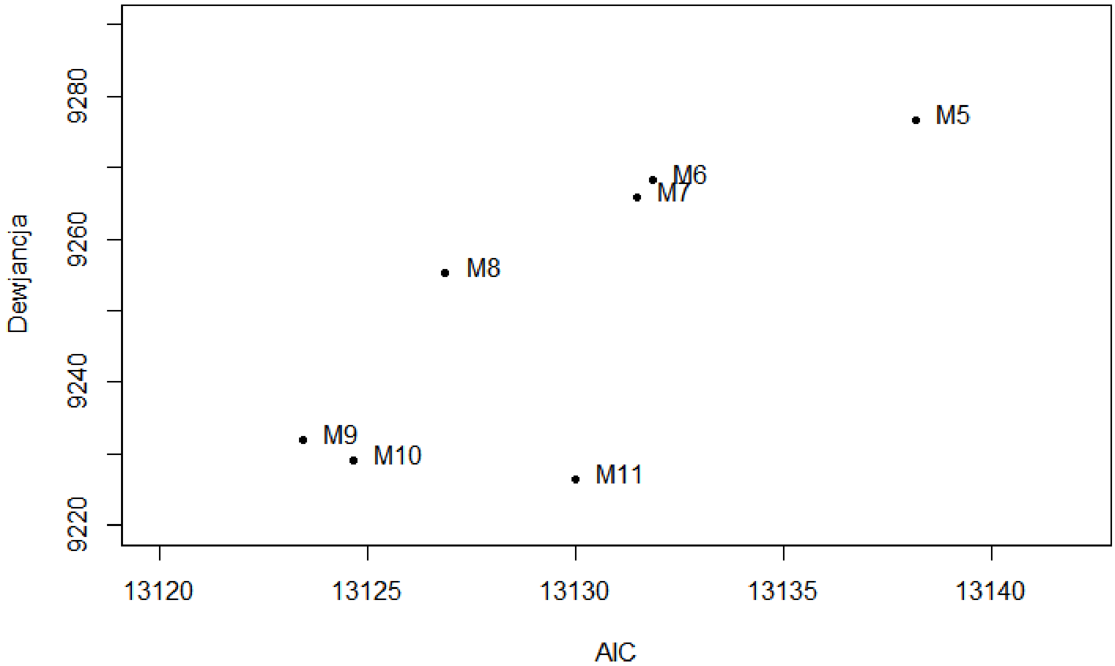
Zadanie 2.

W celu opracowania modelu dla liczby szkód K_i zgłaszanych przez kierowców z pewnego portfela ubezpieczeń AC wykorzystano regresję Poissona (uogólniony model liniowy, w którym zmienna zależna ma rozkład Poissona). Brano pod uwagę 11 następujących zmiennych objaśniających:

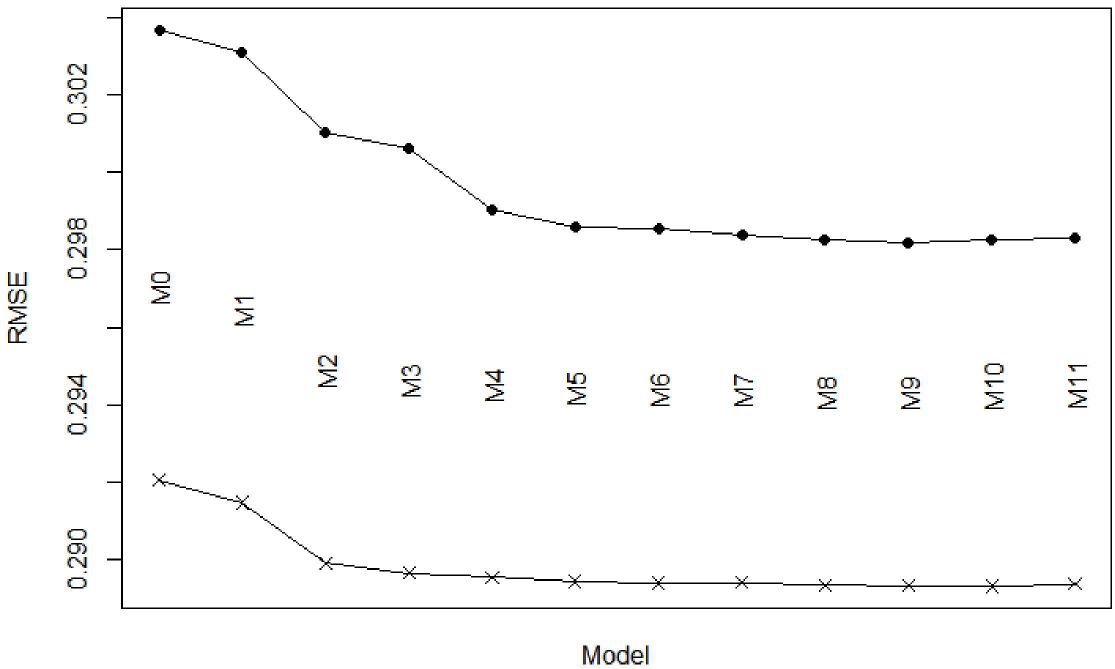
Nr	*Zmienna*	*Opis*
1	*region*	Region zamieszkania kierowcy. Zmienna jakościowa przyjmująca następujące kategorie: *R0*, *R1*, *R2*, *R3*, *R4*, *R5*, *R6*, *R7* i *R8*.
2	*klasa.ryzyka*	Zmienna jakościowa określająca klasę ryzyka. Przyjmuje następujące kategorie: *kl.1*, *kl.2*, *kl.3*, *kl.4*, *kl.5* i *kl.6* (im wyższa klasa, tym większa zniżka za bezszkodowość).
3	*wiek.kierowcy*	Zmienna jakościowa określająca wiek kierowcy. Przyjmuje następujące kategorie: *18-33*, *34-38*, *39-43*, *44-50*, *51-60*, *61-69* i *70+*.
4	*wczes.szkody*	Liczba wcześniejszych szkód (zmienna ilościowa).
5	*prawo.jazdy*	Zmienna jakościowa określająca od ilu lat kierowca posiada prawo jazdy. Przyjmuje następujące kategorie: *1*, *2*, *3*, *4*, *5*, *6*, *7* i *8+*
6	*polisa*	Zmienna jakościowa określająca, czy polisa jest nowa, czy odnowiona. Przyjmuje następujące kategorie: *np* i *op*.
7	*plec*	Zmienna jakościowa określająca płeć kierowcy. Przyjmuje następujące kategorie: *K* - kobieta, *M* – mężczyzna.
8	*stan.cywilny*	Zmienna jakościowa określająca stan cywilny kierowcy. Przyjmuje następujące kategorie: *M* – małżeństwo, *S* – singiel, *R* – rozwiedziony/rozwiedziona, *W* – wdowiec/wdowa).
9	*moc*	Zmienna jakościowa określająca moc silnika (w koniach mechanicznych). Przyjmuje następujące kategorie: *[70,75]*, *(0,65]*, *(65,70]*, *(75,85]*, *(85,90]*, *(90,100]*, *(100,105]*, *(105,110]*, *(110,120]*, *(120,140]* i *(140,200]*.
10	*typ.silnika*	Zmienna jakościowa określająca typ silnika. Przyjmuje następujące kategorie: *Diesel*, *Benzyna* i *LPG*.
11	*l.siedzen*	Zmienna jakościowa określająca liczbę siedzeń w samochodzie. Przyjmuje następujące kategorie: *2*, *3*, *4*, *5* i *6+*.

Oszacowano 12 modeli: M0, M1,..., M11. Przy czym M0 jest modelem zerowym (nie uwzględniono w nim żadnej zmiennej objaśniającej), model M1 zawiera tylko pierwszą zmienną objaśniającą *region*, model M2 – zawiera dwie pierwsze zmienne objaśniające, tj. *region* i *klasa.ryzyka*, model M3 – trzy pierwsze (*region*, *klasa.ryzyka*, *wiek.kierowcy*) itd. Ogólnie, model Mi uwzględnia wszystkie zmienne modelu M(i-1) i dodatkowo i-tą zmienną z powyższej tabeli. Na poniższych dwóch rysunkach przedstawiono wartości kryterium informacyjnego Akaike'a i dewiancje oszacowanych modeli. Drugi rysunek jest „powiększonym” fragmentem pierwszego, obejmującym modele od M5 do M11.


Dewjancja
9700
9600
9500
9400
9300
AIC
13100
13200
13300
13400
13500
M0
M1
M2
M3
M4
M5
M8


Dewjancja
9280
9260
9240
9220
AIC
13120
13125
13130
13135
13140
M5
M6
M7
M8
M9
M10
M11

Przeprowadzono także *10*-krotną walidację krzyżową tych modeli. Na poniższym rysunku przedstawiono wyniki dla średniokwadratowego błędu predykcji ex post (*root mean square error*, RMSE) uzyskane na zbiorze uczącym (symbol: „ ×”) i testowym (symbol „●”)



a) ***(2p.)*** Omówić przydatność prezentowanych wykresów w wyborze optymalnego modelu. Jakie niepożądane zjawisko można zidentyfikować w oparciu o walidację krzyżową? Nazwać i omówić to zjawisko. Czy wystąpiło ono w analizowanym przypadku?
b) ***(2p.)*** W oparciu o podane informacje wybrać najlepszy model. Wybór uzasadnić.
c) ***(1p.)*** Wskazać liczbę oszacowanych parametrów wybranego modelu.

Odpowiedzi

....

Odp. a)

Odpowiedź powinna zawierać:

- informacje o sposobie wykorzystania kryterium AIC i dewiancji w wyborze optymalnego modelu;
- wskazanie, że w oparciu o walidację krzyżową można zidentyfikować niepożądane zjawisko, jakim jest przeuczenie modelu oraz krótki opis tego zjawiska;
- informacje, że na podstawie rysunku przedstawiającego RMSE można zidentyfikować to zjawisko. (Ponieważ na rysunku nie widać wyraźnego wzrostu RMSE dla M10 i M11 w porównaniu z M9, akceptowana była także racjonalnie uzasadniona odpowiedź, że przeuczenie modelu nie występuje.)

....

Odp. b)

Należało wskazać model M9. Uzasadnienie powinno zawierać porównanie modeli M9 i M10 i informację, że na ten model wskazuje wykres przedstawiający wyniki *10*-krotnej walidacji krzyżowej.

Uwaga! Akceptowany był także inny dobrze uzasadniony wybór.

....

Odp. c)

Model M9 ma 43 parametry.

Zadanie 3.

Wykorzystując zbiór uczący liczący 2000 obserwacji, oszacowano dwa uogólnione modele liniowe (M1 i M2) dla wysokości pojedynczego roszczenia (*severity model*) w ubezpieczeniach samochodowych. W modelach tych dla zmiennej objaśnianej przyjęto rozkład gamma oraz logarytmiczną funkcję wiążącą. Liczba oszacowanych parametrów modelu M1 wynosi: 14, jego dewiancja: 1850.0 oraz oszacowany parametr dyspersji: 0.90. W modelu M2 uwzględniono wszystkie zmienne z modelu M1 oraz dodatkowo zmienną jakościową *typ samochodu* liczącą 10 kategorii. Dewiancja modelu M2 wynosi: 1835.0 a oszacowany parametr dyspersji: 0.93.

a) ***(3p.)*** Wskazać i **uzasadnić** jaki test należy użyć, aby określić, czy *typ pojazdu* powinien zostać dodany do modelu M1. Podać hipotezę zerową tego testu i określić (nazwać) rozkład statystyki testowej.
b) ***(2p.)*** Wykorzystując ten test sprawdzić na poziomie istotności 0.05, czy model M2 ma lepsze zdolności predykcyjne w porównaniu z M1.

Odpowiedzi:

....

Odp. a)

Należało wskazać na test F, który jest stosowany, gdy w modelu GLM szacowany jest parametr dyspersji. Testem porównuje się istotność redukcji dewiancji po uwzględnieniu kolejnych zmiennych objaśniających w modelach zagnieżdżonych. Hipoteza zerowa: Dodanie nowych zmiennych nie wpływa na polepszenie zdolności predykcyjnych modelu (Należy wykorzystać prostszy model).
Statystyka testowa wyraża się wzorem:

$$F = \frac{D\left(y;\hat{\theta}^P\right) - D\left(y;\hat{\theta}^Q\right)}{\hat{\phi}(q-p)},$$

gdzie:
$D\left(y;\hat{\theta}^P\right)$ – dewiancja modelu o mniejszej liczbie parametrów p,
$D\left(y;\hat{\theta}^Q\right)$ – dewiancja modelu o większej liczbie parametrów q,
$\hat{\phi}$ - oszacowanie parametru dyspersji modelu o większej liczbie parametrów.
Statystka ta ma rozkład F Snedecora o $(q-p)$ i $(n-q)$ stopniach swobody.

....

Odp. b)
Na poziomie istotności 0.05 nie ma podstaw do odrzucenia hipotezy, że należy wykorzystać prostszy model.

Rozwiązanie:
Statystyka testowa:
$F = \frac{D\left(y;\hat{\theta}^P\right) - D\left(y;\hat{\theta}^Q\right)}{\hat{\phi}(q-p)} = \frac{1850.0-1835.0}{0.93\cdot 9} = 1{,}7921.$
Stopnie swobody:
- licznika: 9,
- mianownika: $2000-23=1977.$

Wartość krytyczna: $F_{kr} = F_{9,\ 1977} = 1.882$ (odczytana z zamieszczonych tablic).
Zatem nie ma podstaw do odrzucenia hipotezy zerowej.

Zadanie 4.
Kupno przez kierowcę nowego produktu w pewnej firmie ubezpieczeniowej modelowano z wykorzystaniem regresji probitowej (uogólnionego modelu liniowego ze zmienną objaśnianą o rozkładzie Bernoulliego i funkcją łączącą Φ^{-1}, gdzie Φ jest dystrybuantą standaryzowanego rozkładu normalnego). Przyjęto, że zmienna zależna Y, przyjmuje dwie wartości: $Y = 1$, gdy kierowca kupi nowy produkt oraz $Y = 0$, gdy nie kupi. Uwzględniono następujące zmienne objaśniające:

- *plec:* Płeć (K - kobieta, M – mężczyzna)
- *stan.cywilny:* Stan cywilny (C – małżeństwo, S – singiel, O – inny)
- *wiek:* Zmienna jakościowa przyjmująca trzy kategorie wiekowe (trzy przedziały wiekowe w latach): [18, 25], (25, 60], (60, 95]
- *lojalnosc:* Liczba lat, w których kierowca był klientem firmy.

Na podstawie zbioru uczącego liczącego 3109 obserwacji, metodą największej wiarygodności oszacowano następujący model:

Coefficients:	*Estimate*	*Std. Error*	*p-Value*
(Intercept)	-0.0023	0.1603	0.988
*plec*M	-0.6305	0.0562	2.00E-16
*stan.cywilny*O	0.1347	0.1233	0.274
*stan.cywilny*S	0.2515	0.0625	5.68E-05
wiek.kat(25,60]	-0.6397	0.1548	3.57E-05
wiek.kat(60,95]	-1.2613	0.1731	3.15E-13
lojalnosc	0.0614	0.0040	2.00E-16

a) ***(2p.)*** Wyznaczyć prognozę kupna nowego produktu przez następującego kierowcę:
   - *plec:* K (kobieta),
   - *stan.cywilny:* C (zamężna),
   - *wiek:* [18, 25],
   - *lojalnosc:* 4 lata.

   Przyjąć punkt odcięcia równy 0.34.
b) ***(3p.)*** Na zbiorze testowym, przyjmując pewien punkt odcięcia otrzymano następującą tabelę trafności prognoz dla tego modelu:

Stan faktyczny

Prognoza	$Y_i = 0$ (N)	$Y_i = 1$ (P)
$Y_i^P = 0$ (N)	368	96
$Y_i^P = 1$ (P)	134	179

Podać współrzędne punktu na krzywej ROC, który odpowiada przyjętemu punktowi odcięcia.

Odpowiedzi

....

Odp. a)
Kierowca kupi nowy produkt.

....

Odp. b)
Współrzędne punku: $(0.2669, 0.6509)$

Rozwiązanie:

Ad a)
Predyktor liniowy: $\eta_i = -0.0023 + 4 \cdot 0.0614 = 0.2434$
Prognoza prawdopodobieństwa: $p_i = 0.595$ (Wartość odczytana z zamieszczonych tablic.)
Prognoza kupna nowego produktu: Ponieważ $p_i > 0.34$, zatem kierowca kupi nowy produkt.

Ad b)
Współrzędne punku: (1- specyficzność, czułość)

specyficzność: $\frac{368}{368+134} = 0.7331$
czułość: $\frac{179}{179+96} = 0.6509$
Współrzędne punku: $(0.2669, 0.6509)$.

Zadanie 5.

a) ***(2p.)*** Krótko scharakteryzować model z inflacją (nadmierną liczbą) zer (*Zero-Inflated Model*).

b) ***(3p.)*** Liczbę szkód K_i zgłaszanych przez kierowców w ciągu roku w pewnym portfelu ubezpieczeń AC, modelowano uwzględniając następujące zmienne objaśniające:

Zmienna	Opis
Wiek	Zmienna jakościowa przyjmująca cztery kategorie wiekowe (cztery przedziały wiekowe w latach): [18, 25], (25, 45], (45, 60], (60, 100].
Power	Moc silnika. Zmienna jakościowa przyjmująca 10 kategorii: d, e, f, g, h, i, j, k, l, m, n, o.

Oszacowano dwa modele z inflacją zer ZIP1 i ZIP0, w których zmienna zliczająca ma rozkład Poissona (*Zero-Inflated Poisson Model*). W modelu ZIP0 nie uwzględniono zmiennych objaśniających. W obu modelach dla zmiennej o rozkładzie Poissona jako funkcję łączącą przyjęto *ln*, a dla zmiennej o rozkładzie Bernoulliego *logit*. Poniżej podano wyniki oszacowania tych modeli:

Model ZIP1

```
Count model coefficients (poisson with ln link):
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)   -2.02545    0.23461  -8.633  < 2e-16 ***
Wiek(25,45]   -0.77910    0.25573  -3.047 0.002315 ** 
Wiek(45,60]   -0.72569    0.25760  -2.817 0.004845 ** 
Wiek(60,100]  -0.38475    0.26151  -1.471 0.141218    
Powere         0.21337    0.04524   4.717 2.40e-06 ***
Powerf         0.21012    0.04334   4.848 1.25e-06 ***
Powerg         0.14687    0.04490   3.271 0.001071 ** 
Powerh         0.21399    0.06577   3.254 0.001140 ** 
Poweri         0.26853    0.07362   3.648 0.000265 ***
Powerj         0.42077    0.07334   5.737 9.65e-09 ***
Powerk         0.38644    0.10099   3.826 0.000130 ***
Powerl         0.34564    0.15581   2.218 0.026529 *  
Powerm         0.62239    0.16988   3.664 0.000249 ***
Powern         0.63241    0.21262   2.974 0.002936 ** 
Powero         0.30240    0.30828   0.981 0.326640    

Zero-inflation model coefficients (binomial with logit link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.01978    0.45325  -0.044    0.965
wiek(25,45]   -0.86001    0.57567  -1.494    0.135
wiek(45,60]   -0.60889    0.54989  -1.107    0.268
wiek(60,100]   0.28825    0.49753   0.579    0.562
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 62
Log-likelihood: -2.458e+04 on 19 Df
```

Model ZIP0

```
Count model coefficients (poisson with ln link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.45703    0.06206  -39.59   <2e-16 ***

Zero-inflation model coefficients (binomial with logit link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.3634     0.1482  -2.452   0.0142 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 11
Log-likelihood: -2.464e+04 on 2 Df
```

Dla kierowcy w wieku od 18 do 25 lat i posiadającego samochód o mocy silnika z kategorii *d*:

i. ***(2p/3p.)*** Obliczyć różnicę między wartością oczekiwaną liczby zgłaszanych szkód w ciągu roku wyznaczoną za pomocą tych modeli.
ii. ***(1p/3p.)*** Korzystając z obydwu modeli wyznaczyć prawdopodobieństwo, że kierowca w ciągu roku zgłosi co najwyżej jedną szkodę.

Odpowiedzi:

....

Odp. a)

Odpowiedź powinna zawierać:

- informację, że model z inflacją (nadmierną liczbą) zer wykorzystuje się do modelowania zmiennych zliczających, w których występuje duża liczba zer, przez co klasyczne rozkłady zliczające (jak np. Poissona) nie daje zadowalających wyników. Z taką sytuacją mamy do czynienia np. w modelowaniu liczby szkód, ponieważ bardzo często portfele ubezpieczeniowe charakteryzują się tym, że w przypadku wielu polis w okresie ubezpieczenia nie występuje żadna szkoda;
- informację o postaci tego modelu (wzór lub krótki opis);
- wzór na wartość oczekiwaną.

....

Odp. b)

ad i.: Różnica (M1-M0): 0.016075

ad. ii.:
Model M1: $P(K_i \leq 1) = 0.99598$
Model M0: $P(K_i \leq 1) = 0.99795$

Rozwiązanie:

Ad i.

$$E(K_i) = (1 - p_i)\lambda_i$$

Model M1:

$$\lambda_i = \exp(-2.02545) = 0.131934$$

$$p_i = \frac{\exp(-0.01978)}{1 + \exp(-0.01978)} = 0.495055$$

$$E(K_i) = (1 - 0.495055) \cdot 0.131934 = 0.06662$$

Model M0:

$$\lambda_i = \exp(-2.45703) = 0.085689$$

$$p_i = \frac{\exp(-0.3634)}{1 + \exp(-0.3634)} = 0.410137$$

$$E(K_i) = (1 - 0.410137) \cdot 0.085689 = 0.050545$$

Różnica (M1-M0): 0.016075

Ad ii.

$$P(K_i = k) = \begin{cases} p_i + (1 - p_i)e^{-\lambda_i}, & k = 0 \\ (1 - p_i)\dfrac{e^{-\lambda_i}\lambda_i^k}{k!}, & k = 1,2,\dots \end{cases}$$

Model M1: $P(K_i \leq 1) = P(K_i = 0) + P(K_i = 1) = 0.93759 + 0.05839 = 0.99598$

Model M0: $P(K_i \leq 1) = P(K_i = 0) + P(K_i = 1) = 0.95156 + 0.04639 = 0.99795$

Zadanie 6.
W pewnym portfelu ubezpieczeń samochodowych kierowcy są ubezpieczeni w dwóch liniach biznesu: LOB1 i LOB2 (na wypadek realizacji dwóch rodzajów ryzyka). W celach taryfowych oszacowano dwuwymiarowy model wysokość szkód. Przyjęto, że szkody z LOB1 i LOB2 są modelowane z wykorzystaniem uogólnionych modeli liniowych, w których zmienna zależna ma rozkład gamma oraz funkcja łącząca jest logarytmiczna ($g(\mu) = \ln(\mu)$). Uwzględniono w nich następujące zmienne objaśniające:

Nr	*Zmienna*	*Opis*
1	*plec*	Zmienna jakościowa określająca płeć kierowcy. Przyjmuje następujące kategorie: K - kobieta, M – mężczyzna.
2	*wiek.kierowcy*	Zmienna jakościowa określająca wiek kierowcy (w latach). Przyjmuje następujące kategorie: *18-25*, *26-35*, *36-45*, *46-55* i *+55*.
3	*wiek.samoch*	Zmienna jakościowa określająca wiek samochodu (w latach). Przyjmuje następujące kategorie: *(0,3]*, *(3,6]* i *(6,12]*.

Strukturę zależności modelowano za pomocą kopuli Gumbela-Hougaarda o generatorze $\phi(t) = (-\ln t\,)^{\theta}$ bez uwzględniania zmiennych objaśniających.
Wyniki oszacowania są następujące:
- rozkłady brzegowe:

Coefficients:

	LOB1	**LOB2**
(Intercept)	5.1395	3.2507
*plec*M	0.3145	0.3083
*wiek.kierowcy*18-25	0.6540	1.3173
*wiek.kierowcy*26-35	-0.1421	0.1212
*wiek.kierowcy*36-45	-0.2396	0.2825
*wiek.kierowcy*46-55	-0.0549	-0.1823
*wiek.*samoch(3, 6]	0.1040	0.4039
*wiek.*samoch(6, 12]	0.2331	0.7850
parametr dyspersji	1.4008	2.8858

- kopula: $\hat{\theta} = 2$.

Dla kobiety w wieku 30 lat i posiadającej czteroletni samochód:
a) ***(1p.)*** Obliczyć średnią łączną szkodę z tych dwóch linii biznesu.
b) ***(4p.)*** Obliczyć prawdopodobieństwo łączne, że szkody z LOB1 (zm. X_{LOB1}) są mniejsze od 200 oraz szkody z LOB2 (zm. X_{LOB2}) są mniejsze od 60, tzn. $P(X_{LOB1} < 200\ i\ X_{LOB2} < 60)$. **Uwaga!! W odpowiednim momencie należy skorzystać z tablic z punku *h* oraz *i* załączonych na początku zestawu.**

Odpowiedzi:

....

Odp. a)

Średnia łącznych szkód: 207.8841.

....

Odp. b)

$P(X_{LOB1} < 200 \; i \; X_{LOB2} < 60) = 0.652233.$

Rozwiązanie:

Ad. a)

Średnia dla LOB1:

$$E(X_{LOB1}) = \exp(5.1395 - 0.1421 + 0.1040) = 164.2517$$

Średnia dla LOB2:

$$E(X_{LOB2}) = \exp(3.2507 + 0.1212 + 0.4039) = 43.6324$$

Średnia łącznych szkód: $164.2517 + 43.6324 = 207.8841$

Ad. b)

Jeżeli zależność między zmiennymi X i Y jest modelowana za pomocą kopuli C, to

$$F_{(X,Y)}(x, y) = C(F_X(x), F_Y(y))$$

Wyznaczamy $F_{LOB1}(200)$

Wyznaczamy parametry rozkładu gamma dla LOB1:
$E(X_{LOB1}) = 164.2517.$
$Var(X_{LOB1}) = 1.4008 \cdot 164.2517^2 = 37791.6517.$
Parametry rozkładu gamma w parametryzacji z załączonych tablic:
$\alpha = 0.7139$
$\lambda = 0.0043$
Stąd: $F_{LOB1}(200) = 0.7110$ (z tablic)>

Wyznaczamy $F_{LOB2}(60)$

Wyznaczamy parametry rozkładu gamma dla LOB2:
$E(X_{LOB2}) = 43.6324.$
$Var(X_{LOB1}) = 2.8858 \cdot 43.6324^2 = 5493.9467.$
Parametry rozkładu gamma w parametryzacji z załączonych tablic:
$\alpha = 0.3465$
$\lambda = 0.0079$
Stąd: $F_{LOB2}(60) = 0.7730$ (z tablic).

Otrzymujemy:
$P(X_{LOB1} < 200 \; i \; X_{LOB2} < 60) = C(F_{LOB1}(200), F_{LOB2}(60)) = C(0.7110, 0.7730)$
W zadaniu C jest kopulą Archimedesa, zatem: $C(u,v) = \phi^{-1}(\phi(u) + \phi(v))$.
Stąd:
$\phi(0.7110) = (-\ln(0.7110))^2 = 0.116338$
$\phi(0.7730) = (-\ln(0.7730))^2 = 0.066294$
$\phi^{-1}(z) = \exp(-\sqrt{z})$
$C(0.7110, 0.7730) = exp\left(-\sqrt{0.116338 + 0.066294}\right) = 0.652233$

Zadanie 7.

Scharakteryzować uogólnione modele addytywne (*Generalized Additive Models* – GAM). Wskazać ogólną ideę modeli GAM i porównać z uogólnionymi modelami liniowymi (GLM).

Odpowiedź:

....

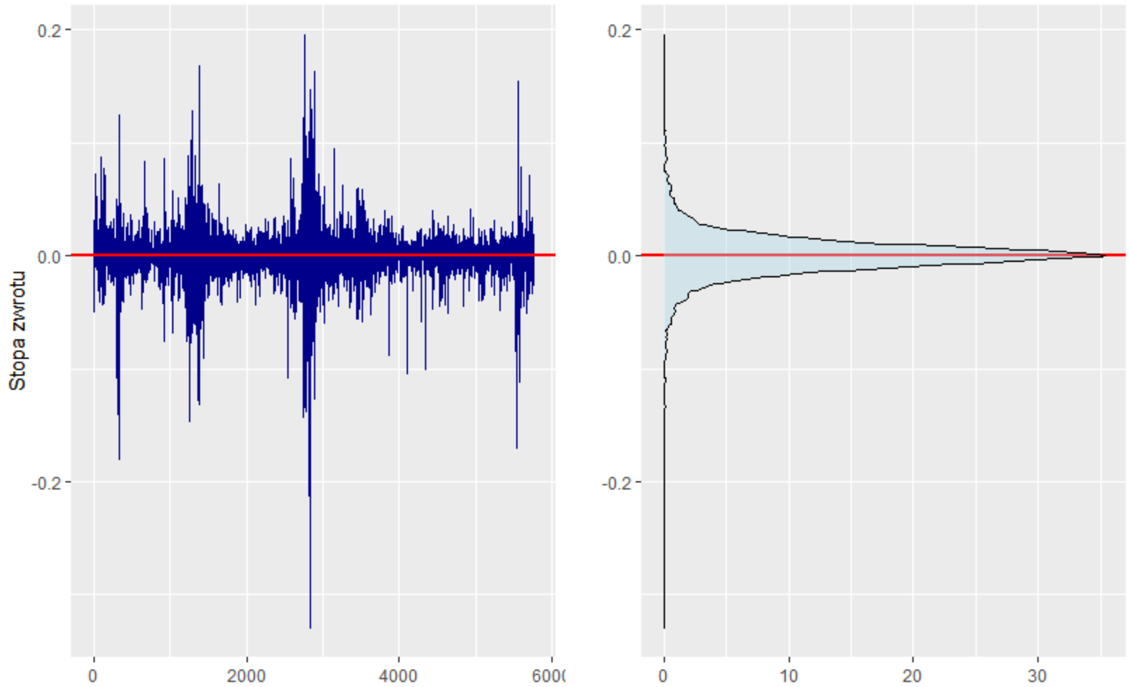
Odpowiedź powinna zawierać:

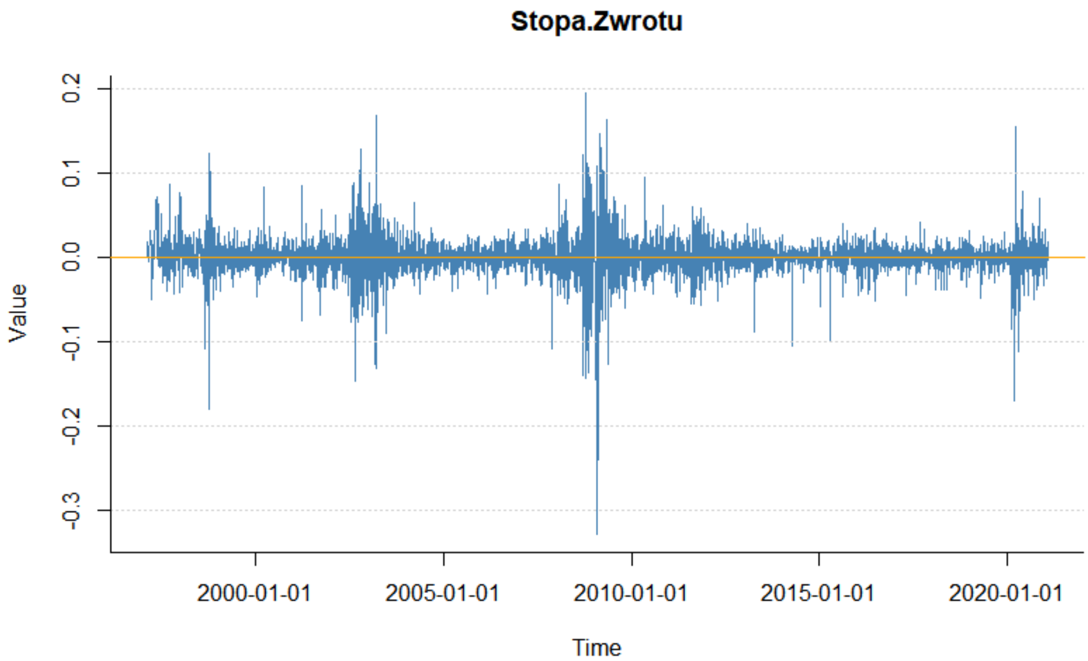
- informację o postaci modelu GAM (wzór lub krótki opis);
- wskazanie, że występujące w modelach GLM (w predyktorze) składowe $\beta_i x_i$ w modelach GAM przyjmują bardziej ogólną postać: $f_i(x_i)$, gdzie f_i są funkcjami zmiennych niezależnych. Można było (ale nie koniecznie) omówić sposoby konstrukcji tych funkcji;
- wskazanie, że modele GAM są uogólnieniem modeli GLM.

Zadanie 8.

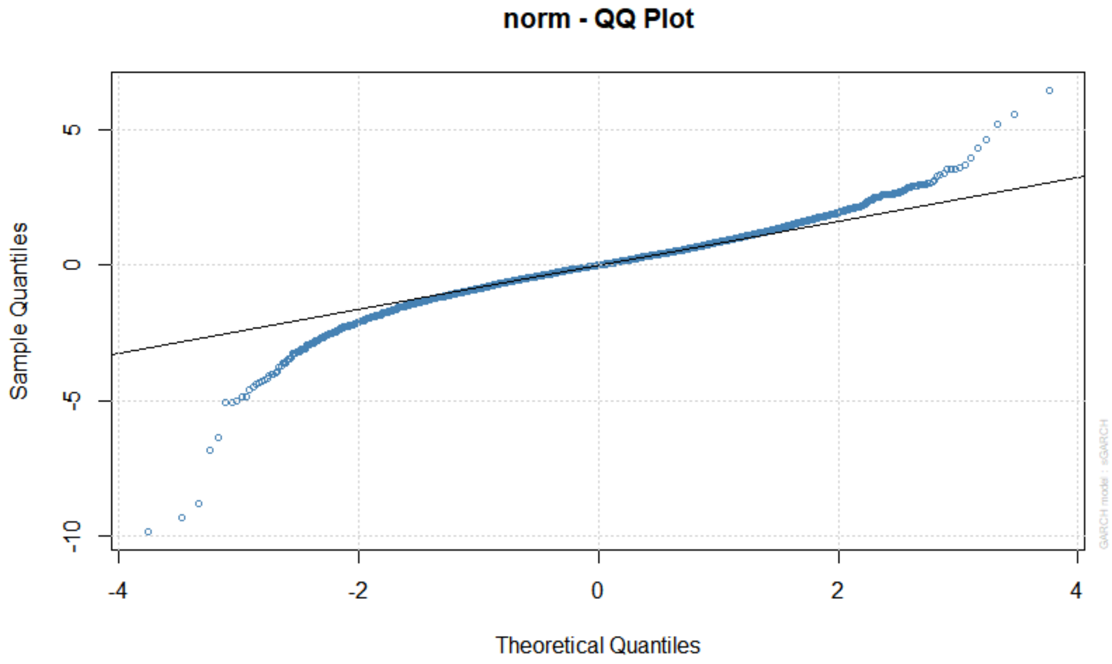
Podczas analizy szeregu czasowego logarytmicznych stóp zwrotu pewnego instrumentu finansowego skonstruowano:

1. Wykresy wartości logarytmicznych stóp zwrotu:

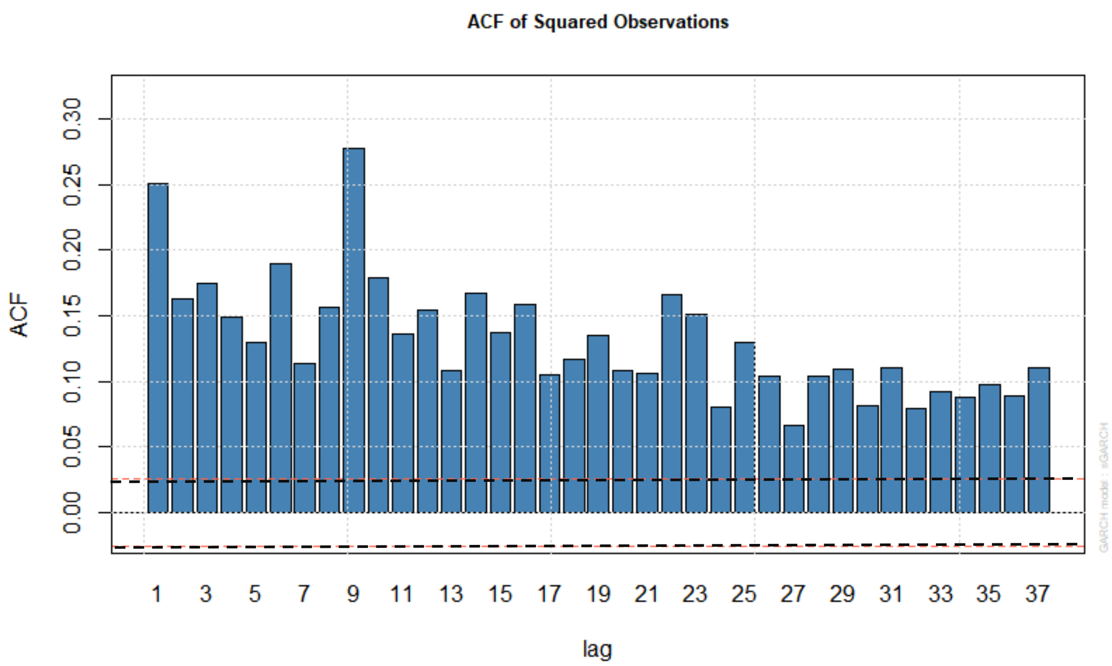




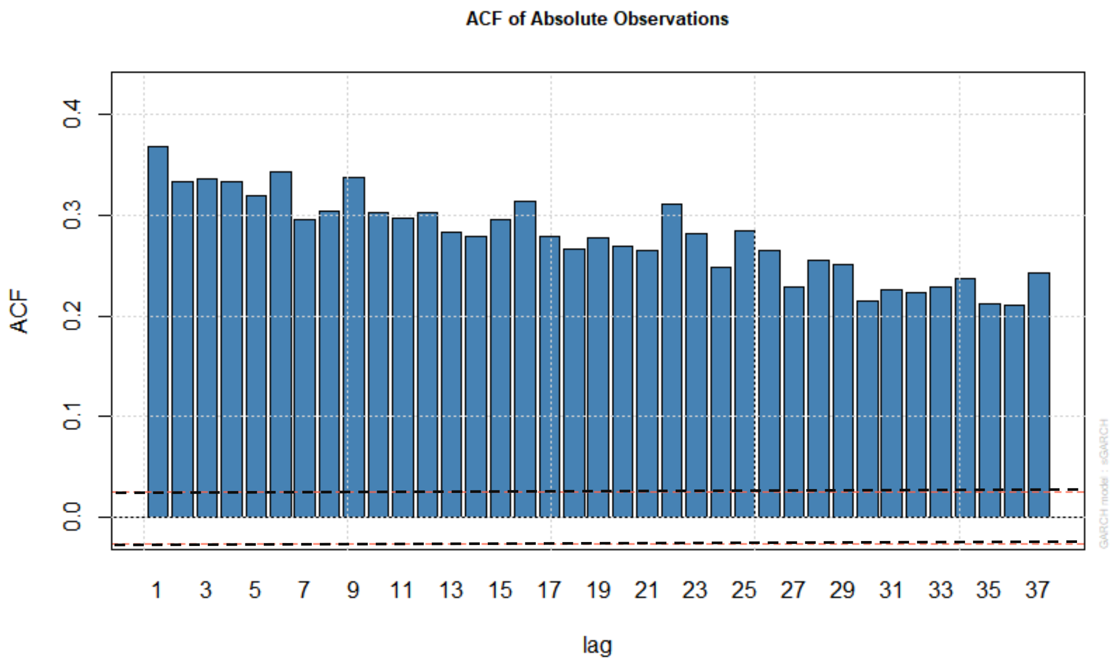
2. Wykres kwanty-kwanty logarytmicznych stóp zwrotu:



3. Wykres funkcji autokorelacji kwadratów logarytmicznych stóp zwrotu:



4. Wykres funkcji autokorelacji modułów logarytmicznych stóp zwrotu:



Wskazać i krótko scharakteryzować trzy właściwości finansowych szeregów czasowych (trzy tzw. stylizowane fakty), które można zaobserwować na przedstawionych wykresach.

Odpowiedzi:

....

Należało wskazać i krótko scharakteryzować następujące właściwości finansowych szeregów czasowych:

- występowanie zjawiska grupowania wariancji (małe, jak i duże zmiany kursu danego instrumentu występują seriami),
- leptokurtyczność rozkładów stóp zwrotu (tzw. grube ogony rozkładu),
- dodatnia autokorelacja kwadratów stóp zwrotu, (ew. dodatnia autokorelacja modułów stóp zwrotu).

Zadanie 9.

Straty w trzech liniach biznesu modelowano z wykorzystaniem zmiennych losowych X_1, X_2, X_3 o rozkładach normalnych z następującymi parametrami:

- zmienna losowa X_1: $\mu = 10$ i $\sigma = 5$.
- zmienna losowa X_2: $\mu = 20$ i $\sigma = 10$.
- zmienna losowa X_3: $\mu = 50$ i $\sigma = 20$.

Łączny wymóg kapitałowy dla tych trzech linii wyznaczono na podstawie następującej formuły: $\kappa = VaR_{0.995}(S) - E(S)$, gdzie $S = X_1 + X_2 + X_3$ natomiast $VaR_{0.995}(\cdot)$ oznacza wartość zagrożoną na poziomie bezpieczeństwa $\alpha = 0.995$. W celu wyznaczenia rozkładu zmiennej S zastosowano dwie metody agregacji:

- **Metoda M1.** Przyjęto, że struktura zależności między X_1, X_2, X_3 opisywana jest wielowymiarowym rozkładem normalnym (tzn. wektor losowy $(X_1\ X_2\ X_3)$ ma wielowymiarowy rozkład normalny), w którym współczynnik korelacji liniowej między:
  - X_1 i X_2 jest równy 0.60,
  - X_1 i X_3 jest równy 0.50,
  - X_2 i X_3 jest równy 0.80.
- **Metoda M2.** Przyjęto, że struktura zależności między X_1, X_2, X_3 opisywana jest kopulą Gumbela z parametrem $\theta = 2.5$. Wartość zagrożoną $VaR_{0.995}(S)$ uzyskano na podstawie podanych w poniższej tabeli statystyk pozycyjnych $S_{(1081:1100)}, \dots, S_{(1100:1100)}$ dla 1100 wartości zmiennej S uzyskanych w drodze symulacji. Wartości ***A***, ***B*** i ***C*** odpowiadają następującym ciągom wygenerowanym ze zastosowanej kopuli:
  $(u_1^A, u_2^A, u_3^A) = (0.9772,\ 0.9675,\ 0.9963)$,
  $(u_1^B, u_2^B, u_3^B) = (0.9872,\ 0.9928,\ 0.9908)$,
  $(u_1^C, u_2^C, u_3^C) = (0.9913,\ 0.9916,\ 0.9920)$.

1081	*1082*	*1083*	*1084*	*1085*	*1086*	*1087*	*1088*	*1089*	*1090*
151.11	151.87	152.06	154.52	154.79	155.45	156.41	157.26	157.67	158.22

1091	*1092*	*1093*	*1094*	*1095*	*1096*	*1097*	*1098*	*1099*	*1100*
158.80	160.14	160.78	**A**	**B**	**C**	167.65	169.71	170.65	171.30

Wyznaczyć różnicę między efektem dywersyfikacji uzyskanym w wyniku zastosowania metody agregacji **M1** i **M2**.

Uwaga! Poniżej podano kwantyle standaryzowanego rozkładu normalnego:

α	*0.9772*	*0.9675*	*0.9963*	*0.9872*	*0.9928*	*0.9908*	*0.9913*	*0.9916*	*0.9920*	*0.9950*
$\Phi^{-1}(\alpha)$	1.9996	1.8455	2.6748	2.2336	2.4451	2.3592	2.3763	2.3913	2.4082	2.5758

Odpowiedź:

....

Efekt dywersyfikacji:

	Metoda M1	Metoda M2	Różnica
Bezwzględny	8.4957	7.350	1.1450
Względny	0.0942	0.0815	0.0127

Wystarczyło podać jeden z podanych wyżej wyników (tzn. albo różnicę dla efektu dywersyfikacji bezwzględnego albo dla względnego).

Rozwiązanie:

Wyznaczamy wymogi kapitałowe dla linii biznesu: $\kappa_i = VaR_{0.995}(X_i) - E(X_i)$.
Pierwsza linia: $\kappa_1 = 22.879 - 10 = 12.879$
Druga linia: $\kappa_2 = 45.758 - 20 = 20.758$
Trzecia linia: $\kappa_3 = 101.516 - 50 = 50.516$
Suma wymogów: 90.153

Łączny wymóg wyznaczony zgodnie z metodą M1
Możemy zastosować metodę wariancji-kowariancji agregacji wymogów dla poszczególnych linii lub wyznaczyć wymóg zgodnie z wzorem $\kappa = VaR_{0.995}(S) - E(S)$ ($S = X_1 + X_2 + X_3$), wyznaczając wcześniej rozkład S. Ponieważ zakładamy, że struktura zależności między X_1, X_2, X_3 opisywana jest wielowymiarowym rozkładem normalnym, zmienna S ma rozkład normalny o średniej 80
Łączny wymóg wyznaczony zgodnie z M1 wynosi: 81.657.

Efekt dywersyfikacji dla M1:

- bezwzględny: $90.1530 - 81.6573 = 8.4957$
- względny:$\frac{8.4957}{90.1530} = 0.0815$

Łączny wymóg wyznaczony zgodnie z metodą M2
Korzystamy ze wzoru $\kappa = VaR_{0.995}(S) - E(S)$
W celu oszacowania VaR-u stosujemy wzór:

$$\widehat{VaR}_\alpha(X) = X_{[n\cdot\alpha]+1:n} \,.$$

$\widehat{VaR}_{0.995}(S) = S_{[1100\cdot 0.995]+1:1100} = S_{1095:1100}$, czyli należało wyznaczyć wartość sumy dla ***B***.

$$\boldsymbol{B} = F_1^{-1}(0.9872) + F_2^{-1}(\ 0.9928) + F_2^{-1}(\ 0.9908) = 21.168 + 44.451 + 97.148 \\ = 162.803$$

Łączny wymóg

$$\kappa = VaR_{0.995}(S) - E(S) = 162.803 - (10 + 20 + 50) = 82.803$$

Efekt dywersyfikacji dla M2:

- bezwzględny: $90.153 - 82.803 = 7.350$
- względny:$\frac{7.350}{90.153} = 0.0127$

Zadanie 10.
Wykorzystując uogólnione modele liniowe opracowano dwa plany taryfikacji: **Plan A** i **Plan B.** Plany te przetestowano na zbiorze liczącym 5 obserwacji. Uzyskano następujące wyniki:

Ryzyko	Ekspozycja	Prognozy		Wartości rzeczywiste
		Plan A	Plan B	
1	0,9	21	22	20
2	0,5	3	6	5
3	1	15	15	15
4	0,8	9	5	8
5	0,8	4	3	2

a) ***(3p.)*** Narysować krzywą Lorenza dla **planu B**.
b) ***(1p.)*** Wskazać co mierzy współczynnik Giniego zastosowany w taryfikacji.
c) ***(1p.)*** Obliczyć współczynnik Giniego dla planu A i B, wiedząc, że pole pod krzywą Lorenza dla **planu A** wynosi: 0,366750 a dla **planu B**: 0,35175. W oparciu o uzyskane wyniki wskazać lepszy plan.

Odpowiedzi:

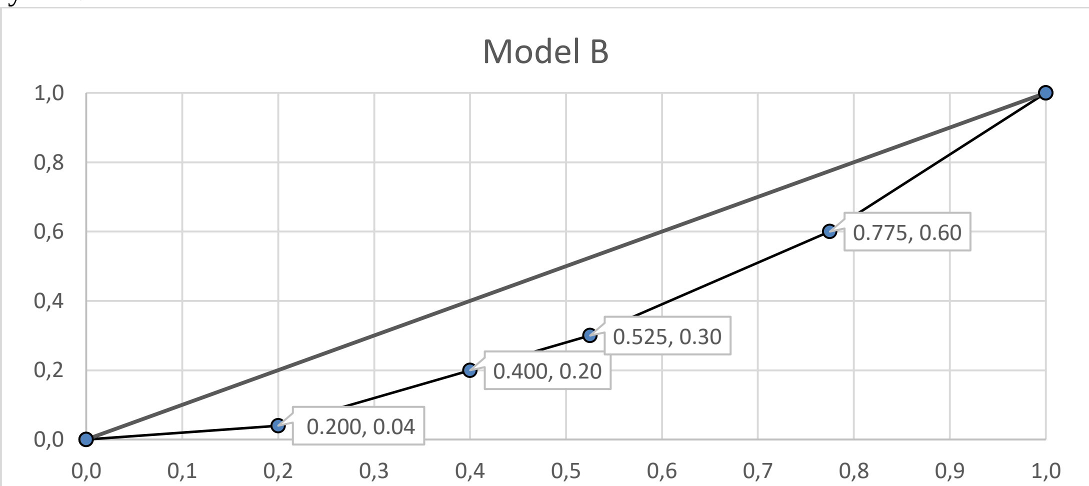
....

Odp. a)
Krzywa Lorenza dla planu B

Obliczenia pomocnicze

Ryzyko	Ekspozycja	Prognoza	Rzeczywista				
5	0.8	3	2	0.8	2	0.200	0.04
4	0.8	5	8	1.6	10	0.400	0.20
2	0.5	6	5	2.1	15	0.525	0.30
3	1.0	15	15	3.1	30	0.775	0.60
1	0.9	22	20	4	50	1.000	1.00

Wykres:



....

Odp. b)

Współczynnik Giniego zastosowany w taryfikacji mierzy poziom zróżnicowania składek w planach taryfowych. Im wyższa jego wartość, tym bardziej składki są zróżnicowane. Preferowane są plany taryfowe o wyższej wartości współczynnika Giniego.

....

Odp. c)

Współczynnik Giniego

$$G = 2 \cdot (0.5 - pole\ pod\ krzywą\ Lorenza)$$

Plan A: $G = 0.2665$
Plan B: $G = 0.2965$

Preferowany jest plan B.

Sesja egzaminacyjna w dniu 5 października 2021 r.

Modelowanie

Arkusz ocen

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