International Science Group

ISG-KONF.COM

VII
INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE "INNOVATIVE TRENDS IN SCIENCE, PRACTICE AND EDUCATION"

Munich, Germany
February 22-25, 2022

ISBN 979-8-88526-758-8

DOI 10.46299/ISG.2022.I.VII

INNOVATIVE TRENDS IN SCIENCE, PRACTICE AND EDUCATION

Proceedings of the VII International Scientific and Practical Conference

Munich, Germany
February 22 – 25, 2022

MULTIVARIABLE LINEAR REGRESSIONS OF MONTHLY MORTALITY *VS* WEATHER CONDITIONS IN ZAPORIZHZHIA REGION, UKRAINE, FROM JANUARY 2020 TO NOVEMBER 2021

Antypenko Lyudmyla
PhD of Pharmacy, Associate Professor
Scientific freelancer

Antypenko Oleksii
PhD of Pharmacy, Associate Professor
Department of Organic and Bioorganic Chemistry
Zaporizhzhia State Medical University
Ukraine

The study of the mortality data from Main Department of Statistics in the Zaporizhzhia region [1], shows the frightening positive trend of the death cases from Covid-19 (Figure 1).

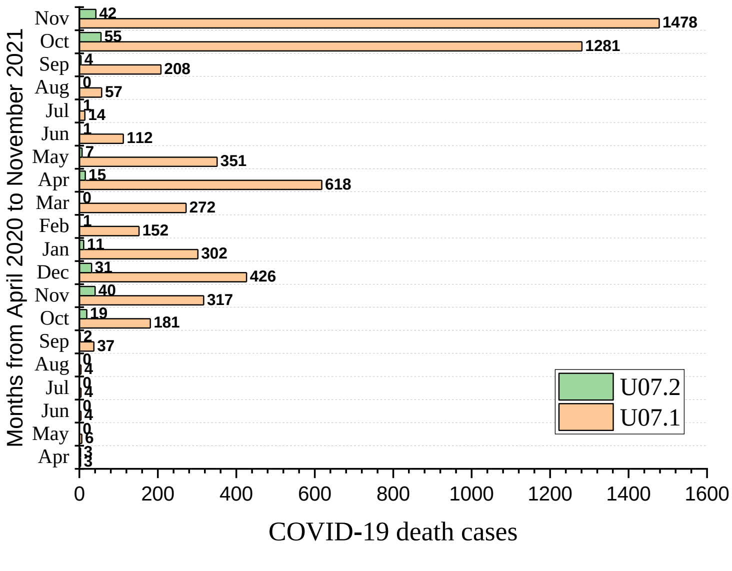

Fig. 1. Number of death cases from COVID-19 in Zaporizhzhia region, Ukraine: U07.1 – clinically identified and confirmed by tests, U07.2 – clinically identified, but not confirmed by tests, from April 2020 till November 2021 [1]

Notably, the rapid increase from 208 cases to 1281 and 1487 deaths were shown during the last months of the 2021 Autumn (Fig.1).
In the previous study [2], the correlations between the accumulated monthly mortality data statistic reports and different weather conditions were analyzed. And now it was decided to analyze the exact monthly death cases and to calculate a multivariable linear regression [3] for each disease.

The dataset of monthly average 7 weather conditions (wind direction and speed, minimum, maximum, and average temperatures, rain precipitation, atmospheric pressure) in Zaporizhzhia from January 2020 till November 2021 [4], and also month number (1-12) and year (2020 - 1, 2021 - 2) were used as the independent variables. The individual month death cases were deducted from the sum of monthly cases reported at the website of the Main Department of Statistics in the Zaporizhzhia region [1], and were used as the dependent variables. The data were analyzed using the Statistical Package for the Social Sciences software (IBM® SPSS® for Windows, v.26.0, Inc, Chicago, IL). Multivariable linear regression statistics and forecast model were calculated with a confidence interval of 95%.

Hence, during the studied period the following diseases had the lowest number of death cases for the whole studied period: F10, D50–D89, O00–O99, F01–F99, X00–X09, and M00–M99 (from 5 to 42); Q00–Q99, P07–P96, L00–L98, E10–E14, E00–E89, W65–W74, and Y00–Y09 (from 65 to 148); B20–B24, N00–N99, K70, 45, A15–A19, V01–V99, and X40–X44 (from 201 to 515) (Fig. 2, Table 1). Unfortunately, each type of mortality was slowly increasing with time, but not perfectly linear.

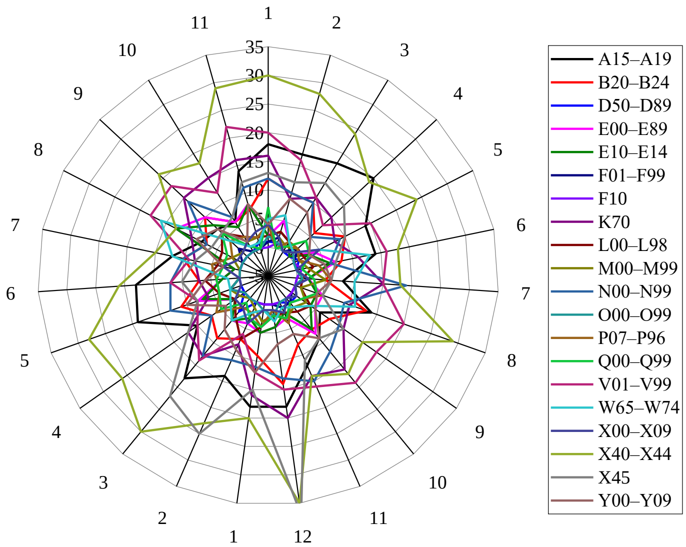

Fig. 2. Low level of monthly mortality (about 1-35 cases) from diseases by international classification
from January - December of 2020 till January - November of 2021

The moderate amount of deaths were caused by A00–B99, G00–G9, and X60–X84 (617-724); I42 (1226), K00–K92 and R00–R99 (2032-2175), and V01–Y89 (3060) (Fig. 3, Table 1). Among them, a positive trend of mortality with time was observed for K00–K92, and R00–R99.

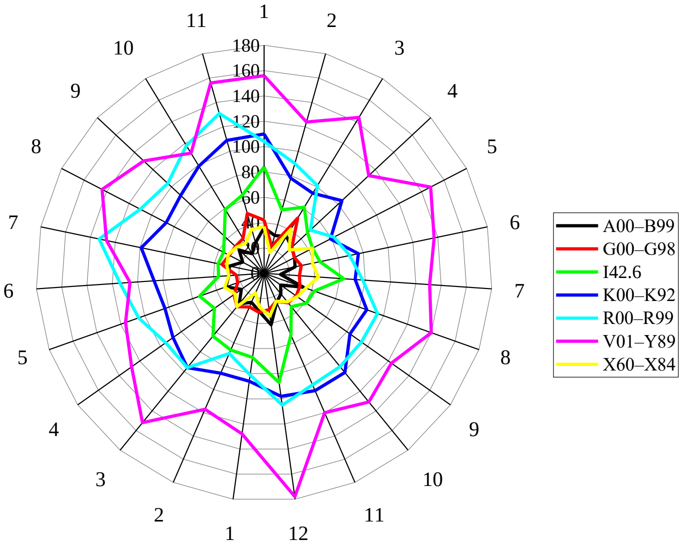

Fig. 3. Moderate level of monthly mortality (about 13-180 cases) from diseases by international classification
from January - December of 2020 till January - November of 2021

The constant high mortality in the studied period was observed due to: U00-U85 (6056), C00–C97 (8375), C00–D48 (8444), and I60–I69 (10182) (Fig. 4, Table 1). And the maximum was caused by diseases of the circulatory system, I00–I99 (36221), namely, coronary heart disease, I20–I25 (21774), increasing in the 10th-12th months. The rapid growth in COVID-19 (U00-U85) mortality is seen in the latest months.

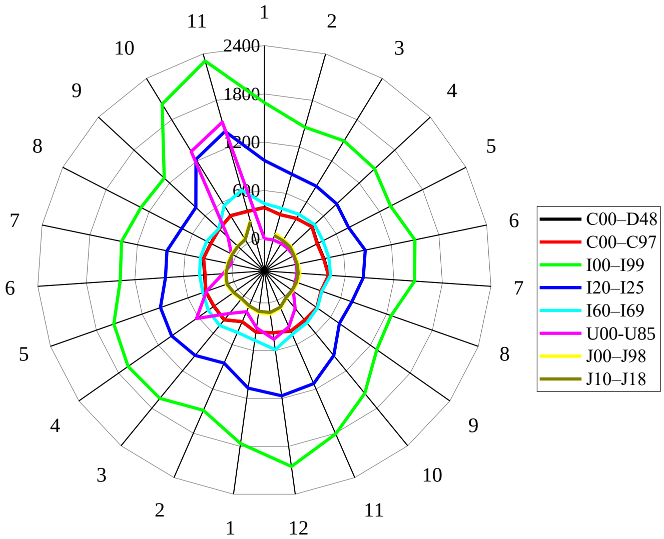

Fig. 4. High level of monthly mortality (about 3-2400 cases) from diseases by international classification
from January - December of 2020 till January - November of 2021

As the main step of study, in the Table 1, the calculated multivariable linear regression data towards 7 weather parameters mentioned above, and also year, and months order, can be found. The majority of the found regression coefficients are positive, and those, which are negative, all are insignificant.

Table 1

Multivariable linear regression results of monthly mortality *vs* 7 weather parameters, year, months in Zaporizhzhia region, Ukraine, from January 2020 till November 2021

Class	Disease	International classification of diseases (ICD)	Adjusted R^2	ANOVA, Sig.
I	Some infectious and parasitic diseases:	A00–B99	-0.061	0.579
	tuberculosis	A15–A19	0.336	0.091
	human immunodeficiency virus (HIV) disease	B20–B24	-0.463	0.984
II	Tumors:	C00–D48	0.472	0.327
	malignant tumors	C00–C97	0.124	0.303
III	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	D50–D89	0.165	0.967
IV	Endocrine diseases, eating disorders and metabolic disorders:	E00–E89	0.116	0.314
	diabetes mellitus	E10–E14	0.514	0.235
V	Mental and behavioral disorders:	F01–F99	-0.227	0.816
	mental and behavioral disorders due to alcohol use	F10	-0.106	0.650
VI	Diseases of the nervous system	G00–G98	0.064	0.387
IX	Diseases of the circulatory system:	I00–I99	0.656	0.003
	coronary heart disease	I20–I25	0.651	0.003
	alcoholic cardiomyopathy	I42.6	0.429	0.043
	cerebrovascular diseases	I60–I69	0.587	0.008
X	Respiratory diseases:	J00–J98	0.547	0.013
	influenza and pneumonia	J10–J18	0.568	0.010
XI	Digestive diseases:	K00–K92	0.478	0.027
	alcoholic liver disease	K70	0.475	0.028
XII	Diseases of the skin and subcutaneous tissue	L00–L98	0.202	0.208
XIII	Diseases of the musculoskeletal system and connective tissue	M00–M99	0.098	0.339
XIV	Diseases of the genitourinary system	N00–N98	-0.332	0.918
XV	Pregnancy, childbirth and the postpartum period	O00–O99	-0.209	0.794
XVI	Some conditions that occur in the perinatal period	P00–P96	0.231	0.177
XVII	Congenital malformations, deformities and chromosomal abnormalities	Q00–Q99	0.037	0.429

XVIII	Symptoms, signs and abnormalities found in clinical and laboratory studies are not classified in other headings	R00–R99	0.650	0.003
XX	External causes of death:	V01–Y89	-0.245	0.837
	traffic accidents	V01–V99	0.584	0.008
	accidental drowning and immersion in water	W65–W74	0.279	0.133
	accidents caused by smoke, fire and flame	X00–X09	-0.013	0.506
	accidental poisoning caused by toxic substances (except alcohol)	X40–X44, X46–X49	-0.110	0.655
	accidental poisoning and alcohol exposure	X45	0.592	0.007
	intentional self-harm	X60–X84	-0.234	0.824
	the consequences of an attack to kill or injure	X85–X99, Y00–Y09	0.266	0.145
XXII	Codes for special purposes	U00–U85	0.543	0.013
	COVID-19, virus identified	U07.1	0.527	0.016
	COVID-19, virus unconfirmed	U07.2	0.787	0.000

Thus, among 37 calculated regressions, only 15 with significance ≤0.05 were chosen for further consideration (Table 2).

Table 2

The multivariable linear regression data for mortality cases *vs* 7 weather conditions, year and month's number with statistically significant adjusted R^2

Short disease name	**ICD**	**Adjust. R^2**	**ANOVA**		**Cook's dist.**	**Unstandardized B**		
			F	**Sig.**		**Sig. ≤ 0.05**	**sign**	**Sig.**
COVID-19, unconfirmed	U07.2	0.787	10.040	0.000	1.026	month	posit.	0.018
Circulatory system	I00–I99	0.656	5.666	0.003	0.812	-	-	-
Coronary heart	I20–I25	0.651	5.566	0.003	0.633	-	-	-
Symptoms, signs and abnormalities	R00–R99	0.650	5.545	0.003	0.868	t. min.	posit.	0.002
						t. aver.	negat.	0.006
						year	posit.	0.009
						t. max.	posit.	0.021
Accidental poisoning and alcohol exposure	X45	0.592	4.545	0.007	1.571	wind direc.	negat.	0.042
Cerebrovascular	I60–I69	0.587	4.471	0.008	0.874	-	-	-
Traffic accidents	V01–V99	0.584	4.427	0.008	0.399	year	negat.	0.001
						wind speed	negat.	0.008
						t. min	posit.	0.048

						t. average	negat.	0.052
Influenza and pneumonia	J10–J18	0.568	4.214	0.010	1.245	t. min.	posit.	0.044
Respiratory	J00–J98	0.547	3.955	0.013	1.271	t. min.	posit.	0.051
COVID-19, all	U00–U85	0.543	3.908	0.013	0.834	year	posit.	0.057
COVID-19, identified	U07.1	0.527	3.726	0.016	0.832	year	posit.	0.056
Digestive	K00–K92	0.478	3.241	0.027	1.256	-	-	-
Alcoholic liver	K70	0.475	3.213	0.028	0.511	rain precip.	negat.	0.029
Alcoholic cardiomyopathy	I42.6	0.429	2.835	0.043	1.295	-	-	-

The significant contribution to mortality of the diseases mentioned in Table 2 was made by the weather factors in the following order of frequency: year, minimum, average and maximum temperatures, wind speed, month number, amount of rain, and wind direction. The increase in wind direction and speed, amount of rain and average temperature leads only to decrease in corresponding death cases (negative value of unstandardized B). And positive impact into the level of mortality was always seen by month number, minimum and maximum temperatures. Some death cases increased with years (R00–R99, U00–U85), and some reduced (V01–V99).

Atmospheric pressure had no influence on mortality rates. And for I00–I99, I20–I25, I60–I69, K00–K92, and I42.6 there was found no significant impact of weather conditions.

Also, only for 8 regressions, the Cook's distance was less than 1 [5]. Among which, calculated for variables "Symptoms, signs and abnormalities found in clinical and laboratory studies are not classified in other headings" (R00–R99) and "External causes of death: traffic accidents" (V01–V99), the maximum number of significant weather factors (4 for each) were found (Table 2). And, it was decided to study the first one in more detail considering COVID-19 consequences.

According to Table 3, the regression model adjusted R square has a moderate positive value of 0.650 with a high significance (.003). The strongest positive impact into the regression equation is calculated for minimum temperature (Part: 0.475), and strongest negative: respectively, for average temperature (Part: −0.417), than goes year order and maximum temperature with positive influence (Part: 0.385, 0.332).

Table 3

Regression statistics on R00-R99 death cases *versus* 7 weather conditions, year and month's number

Model Summary[b]

Model	*R*	*R^2*	*Adjusted R^2*	*Std. Error of the Estimate*	*Change Statistics*				
					R^2 Change	*F Change*	*df1*	*df2*	*Sig. F Change*
1	.891[a]	.793	.650	12.216	.793	5.545	9	13	.003

[a.] Predictors: (Constant), month, year, precipitation, t. max, wind direction, wind speed, atm. pressure, t. min, t. average.
[b.] Dependent Variable: R00–R99

Coefficients

Model		*Unstandard. Coeffic.*		*Standard. Coeffic.*	*t*	*Sig.*	*95.0% Confidence Interval for B*		*Correlations*		
		B	*Std. Error*	*Beta*			*Lower Bound*	*Upper Bound*	*Zero-order*	*Partial*	*Part*
1	(Constant)	-1298.933	1611.625		-.806	.435	-4780.636	2182.771			
	tavg	-30.141	9.111	-13.467	-3.308	.006	-49.825	-10.457	.124	-.676	-.417
	tmin	19.810	5.259	7.706	3.767	.002	8.449	31.171	.240	.722	.475
	tmax	12.482	4.735	6.359	2.636	.021	2.252	22.713	.062	.590	.332
	prcp	-2.359	2.641	-.156	-.893	.388	-8.064	3.346	.091	-.240	-.113
	wdir	-.058	.085	-.116	-.682	.507	-.242	.126	-.299	-.186	-.086
	wspd	.974	1.790	.120	.544	.595	-2.894	4.843	-.442	.149	.069
	pres	1.358	1.576	.286	.862	.404	-2.047	4.763	.209	.232	.109
	year	20.792	6.807	.514	3.055	.009	6.087	35.496	.499	.646	.385
	month	-.805	1.558	-.133	-.517	.614	-4.170	2.560	.463	-.142	-.065

Residuals Statistics

	Minimum	*Maximum*	*Mean*	*Std. Deviation*	*N*
Predicted Value	54.83	132.69	94.57	18.399	23
Std. Predicted Value	-2.160	2.072	.000	1.000	23
Standard Error of Predicted Value	5.236	10.670	7.874	1.737	23
Adjusted Predicted Value	60.50	152.35	95.31	22.561	23
Residual	-15.474	19.799	.000	9.391	23
Std. Residual	-1.267	1.621	.000	.769	23
Stud. Residual	-1.598	1.999	-.022	1.060	23
Deleted Residual	-34.355	43.499	-.749	19.268	23
Stud. Deleted Residual	-1.713	2.309	.003	1.129	23
Cook's Distance	.000	.868	.131	.217	23
Centered Leverage Value	.140	.719	.391	.185	23

All other weather conditions were insignificant in the regression (Table 3).

In the Residuals statistics (Table 3), their studied and standardized values not exceeded -3/3. And Cook's distance not exceeded 1, but was high – 0.868.

Nevertheless, the accuracy of the regression model for R00-R99 estimation could be improved by selection of the most fitting independent variables [3].

Thus, only 4 significant variables: year, maximum, minimum and average temperatures were chosen for further calculations (Table 4).

Table 4

Regression statistics on R00-R99 death cases *versus* 3 weather conditions, and year

Model Summary[b]

Model	*R*	*R^2*	*Adjusted R^2*	*Std. Error of the Estimate*	*Change Statistics*				
					R^2 Change	*F Change*	*df1*	*df2*	*Sig. F Change*
1	.856[a]	.732	.673	11.821	.732	12.296	4	18	.000

[a.] Predictors: (Constant), year, t. max, t. min, t. average.
[b.] Dependent Variable: R00–R99

Coefficients

Model		*Unstandard. Coeff.*		*Standard. Coeff.*	*t*	*Sig.*	*95.0% Confidence Interval for B*		*Correlations*		
		B	*Std. Error*	*Beta*			*Lower Bound*	*Upper Bound*	*Zero-order*	*Partial*	*Part*
	(Constant)	95.666	11.832		8.085	.000	70.807	120.524			
1	tavg	-26.107	7.854	-11.665	-3.324	.004	-42.607	-9.606	.124	-.617	-.406
	tmin	18.237	4.004	7.094	4.554	.000	9.825	26.650	.240	.732	.556
	tmax	9.484	4.140	4.832	2.291	.034	.787	18.181	.062	.475	.280
	year	14.507	5.140	.359	2.822	.011	3.708	25.306	.499	.554	.344

Residuals Statistics

	Minimum	*Maximum*	*Mean*	*Std. Deviation*	*N*
Predicted Value	55.17	134.17	94.57	17.674	23
Std. Predicted Value	-2.229	2.241	.000	1.000	23
Standard Error of Predicted Value	3.628	9.019	5.394	1.160	23
Adjusted Predicted Value	57.22	136.18	94.88	17.956	23
Residual	-19.316	22.603	.000	10.692	23
Std. Residual	-1.634	1.912	.000	.905	23
Stud. Residual	-1.717	2.125	-.011	1.005	23
Deleted Residual	-21.324	27.917	-.316	13.237	23
Stud. Deleted Residual	-1.825	2.386	.000	1.050	23
Cook's Distance	.001	.212	.047	.052	23
Centered Leverage Value	.051	.539	.174	.104	23

Hence, the adjusted R^2 increased to 0.673 (F = 12.296, Sig. = .000), the Cook's distance decreased to 0.212. All regression coefficients were statistically significant (.000 - .034). Predicted and standardized residuals were between -3/3.

Visually, the standardized residuals were not perfectly but normally distributed, fit into plots, and correlated to predicted values (Fig. 5).

Thus, the multivariable linear regression model for monthly R00-R99 death cases is y = 95.666 - 26.107*average temperature in ºC + 18.237*minimum temperature in ºC + 9.484* maximum temperature in ºC + 14.507*year number (2020 – 1, 2021 – 2, etc.) with SE of 11.832 (Table 4).

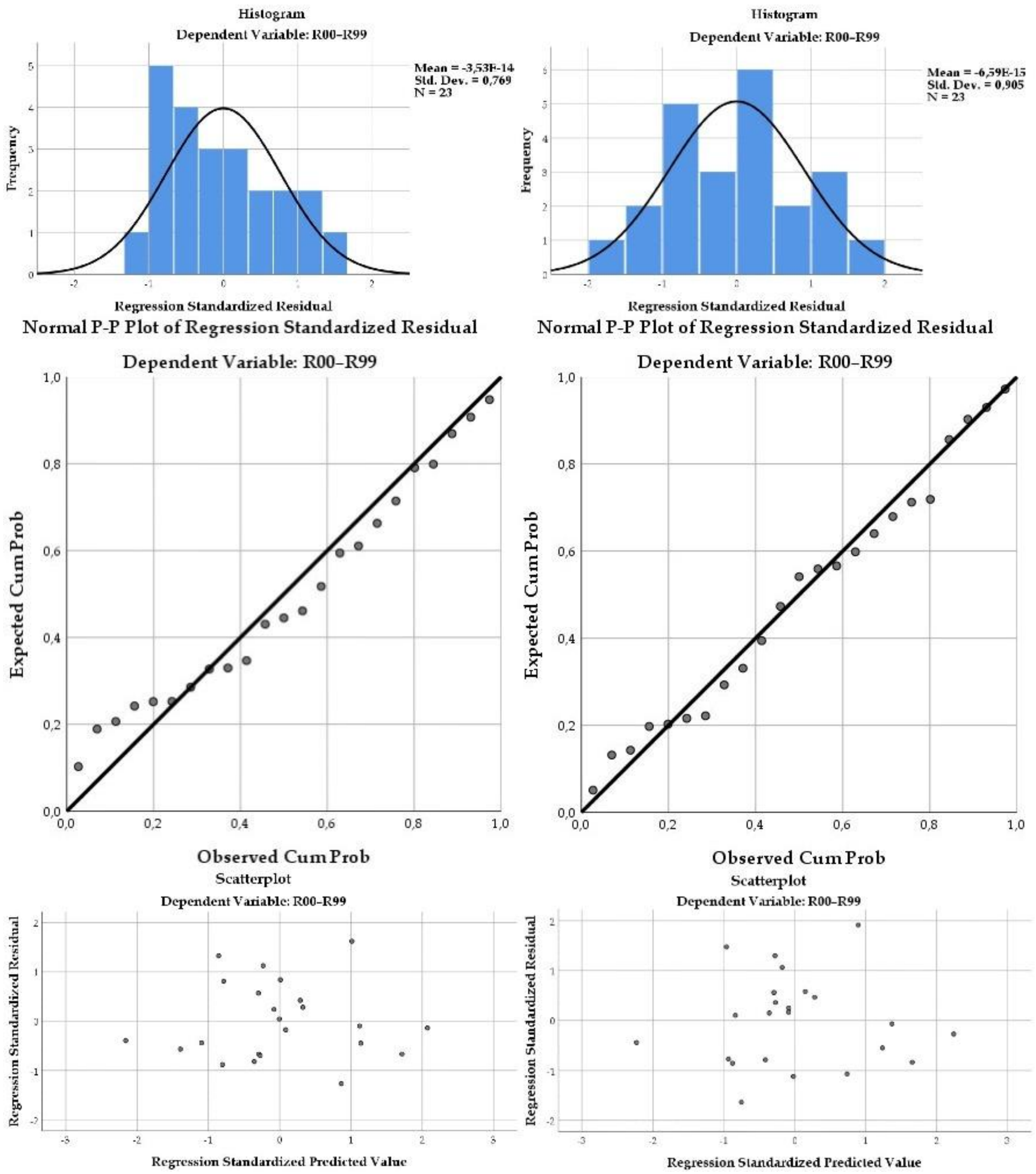

Fig. 5. Histograms, normal probability plots and scatterplots of R00-R99 multivariate linear regression models (to the right – *vs* 9 variables; to the left – *vs* 4 variables)

Besides, Winters' multiplicative model of forecasting was used for prediction of the next year mortality rate of R00-R99 (Table 5). The model's R^2 was found to be weak positive (0.243) and statistically significant (.007).

Table 5

Winters' multiplicative model statistics for monthly R00-R99 death cases

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		R^2	Statistics	DF	Sig.	
R00–R99-Model_1	0	.243	31.800	15	.007	0

At Figure 6, its weak positive trend can be seen for 2022 if nothing changes during this year towards such matters.

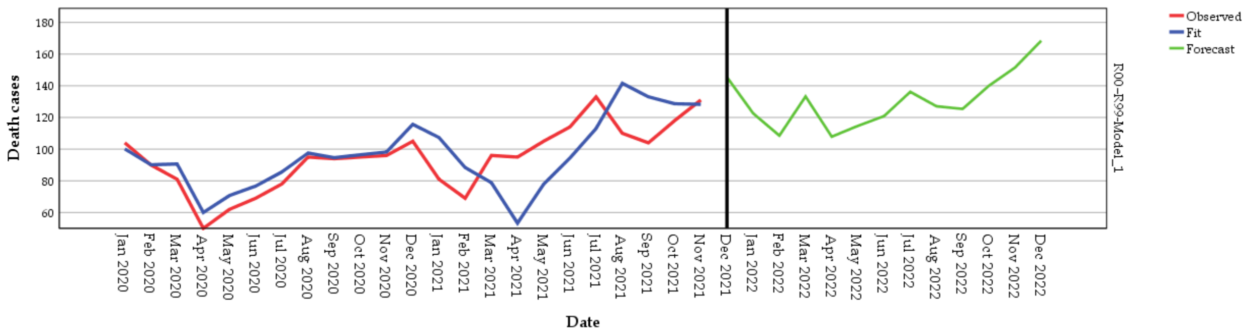

Fig. 6. Winters' multiplicative model forecast for R00-R99 death cases in 2022

Hence, analysis of the monthly mortality rate in Zaporizhzhia region, Ukraine, has shown that majority of diseases have positive trend with time. The multivariate linear regression models have found a significant impact of some weather parameters on the following illnesses: accidental poisoning and alcohol exposure, abnormalities, traffic accidents, influenza and pneumonia, respiratory, and alcoholic liver diseases.

And, unfortunately, COVID-19 mortality is still increasing with months as well as "Symptoms, signs and abnormalities found in clinical and laboratory studies". And for the latter regression equation is calculated and a weak positive growth of death cases is predicted for 2022.

References:

1. *Демографічна та соціальна статистика. Кількість померлих за причинами смерті у Запорізькій області.* (2022). Вилучено з: http://www.zp.ukrstat.gov.ua/index.php/statystychna-informatsiia/38-statistichna-informatsiya/arkhiv-statystychnoi-informatsii/demohrafichna-ta-sotsialna-statystyka/463-1-1-7

2. Antypenko, L. & Antypenko, O. Mortality in Zaporizhzhia region: trend of the 21 months of 2020 – 2021 and relationship with weather conditions. *Current issues of science, prospects and challenges: coll. of scientific papers «Scientia» with Proceedings of the I International Scientific and Theoretical Conference.* (Vol. 3, p. 44-50). December 17, 2021, Sydney, Australia, "European Scientific Platform". Retrieved from: https://www.researchgate.net/publication/357118729_Mortality_in_Zaporizhzhia_region_trend_of_the_21_months_of_2020_-_2021_and_relationship_with_weather_conditions

3. *Applied Regression Analysis. Stepwise Regression.* (2018). Retrieved from: https://online.stat.psu.edu/stat462/node/196/

4. *Meteostat Zaporizhia*. (2022). Retrieved from: https://meteostat.net/en/place/3VM2UJ?t=2021-01-01/2021-10-01

5. *Applied Regression Analysis. Identifying Influential Data Points*. (2018). Retrieved from: https://online.stat.psu.edu/stat462/node/173/