Energy consumption forecasting in public nursing homes using multivariable regression models

Miguel Gómez-Chaparro[1], Alajandro Prieto-Fernández[2], Manuel Botejara-Antúnez[3] and Justo García-Sanz-Calcedo[4,*]

[1]Departamento de Expresión Gráfica, Escuela de Ingenierías Industriales, Universidad de Extremadura, Avenida de Elvas, s/n, 06006, Badajoz, Spain, mgomezchaparro@unex.es

[2]Departamento de Expresión Gráfica, Escuela de Ingenierías Industriales, Universidad de Extremadura, Avenida de Elvas, s/n, 06006, Badajoz, Spain, apf@unex.es

[3]Departamento de Expresión Gráfica, Escuela de Ingenierías Industriales, Universidad de Extremadura, Avenida de Elvas, s/n, 06006, Badajoz, Spain, manuelba@unex.es

[4]Departamento de Expresión Gráfica, Escuela de Ingenierías Industriales, Universidad de Extremadura, Avenida de Elvas, s/n, 06006, Badajoz, Spain, jgsanz@unex.es

[*]Corresponding author: jgsanz@unex.es

Supplementary Material

This Supplementary Material contains the auxiliary resources of the manuscript “Energy consumption forecasting in public nursing homes using multivariable regression models”.

Regression diagnostics for the Energy Consumption model

1. Results of the multiple regression analysis

Table S1 to Table S5 present the multiple regression statistics, analysis of variance (ANOVA) results, model coefficients, residual outputs and probability outputs for the energy consumption prediction model.

Table S1. *Energy Consumption model regression statistics.*

Regression Statistics	
Multiple R	0.986921105
R^2	0.974013268
Adjusted R^2	0.970956006
Standard Error	52.87641825
Observations	20

Table S2*. Energy Consumption model Analysis of Variance (ANOVA).*

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1781501.48	890750.7399	318.5899953	3.35266E-14
Residual	17	47530.56532	2795.915607		
Total	19	1829032.045			

Table S3*. Energy Consumption model Coefficients (Significance of Individual Predictors).*

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	48	20.65905508	2.308123899	0.03383356	4.096862535	91.270455	4.096862535	91.270455
Useful floor area (m2)	0.03	0.008120481	4.197378665	0.000605036	0.016952016	0.05121745	0.016952016	0.05121745
Nº Residents	2.95	0.386822187	7.618756699	7.05002E-07	2.130980653	3.763227605	2.130980653	3.763227605

***Table S4**. Energy Consumption model Residual output.*

Observation	*Predicted values (MWh, total adjusted)*	*Residuals*	*Standard Residuals*
1	246.1659648	7.961200556	0.159172807
2	951.0540598	-34.76705975	-0.695117584
3	1131.724754	-72.15035074	-1.442542965
4	426.8366589	53.83501421	1.076354034
5	185.8970282	16.78741409	0.33564031
6	1085.252559	43.05351956	0.860793484
7	320.2757536	104.9049579	2.097424441
8	234.31428	-24.11266843	-0.482098283
9	353.499903	-37.90717812	-0.757899755
10	382.401306	-57.21705604	-1.143973118
11	905.0398932	32.53839675	0.650558658
12	294.4468776	-42.81310073	-0.855986654
13	178.8123049	-25.63551255	-0.512545371
14	451.2315255	64.34407066	1.286467571
15	344.0668973	-11.947205	-0.238867259
16	278.8300551	33.91484878	0.678078845
17	655.0508775	73.20537247	1.463636614
18	574.6692747	-51.7502401	-1.03467196
19	278.7520834	-29.29681412	-0.585747854
20	219.2853632	-42.9476094	-0.858675962

***Table S5**. Energy Consumption model Probability output.*

Percentile	*Energy Consumption values (MWh, total adjusted)*
2.5	153.1767923
7.5	176.3377538
12.5	202.6844423
17.5	210.2016115
22.5	249.4552692
27.5	251.6337769
32.5	254.1271654
37.5	312.7449038
42.5	315.5927248
47.5	325.18425
52.5	332.1196923
57.5	425.1807115
62.5	480.6716731
67.5	515.5755962
72.5	522.9190346
77.5	728.25625
82.5	916.287
87.5	937.57829
92.5	1059.574403
97.5	1128.306079

2. Assessment of regression assumptions

Linearity

The simple regression plots for the pairwise relationships between the variables are shown in Figure S1.

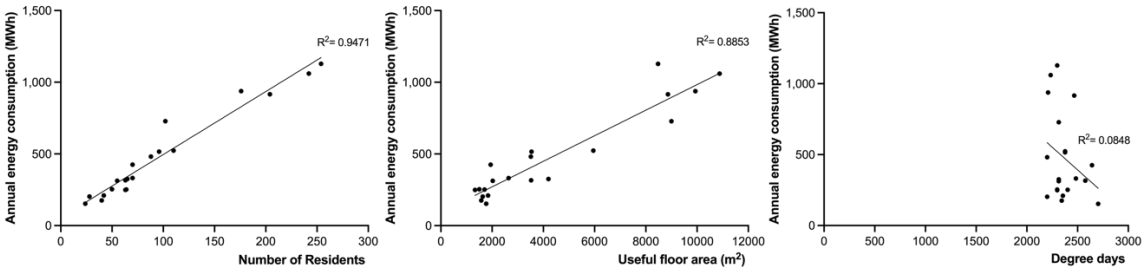


Figure S1. *Pairwise simple regression plots.*

Homoscedasticity

The assumption of homoscedasticity was evaluated using residuals versus fitted values plots (see Table S4). The corresponding plots are presented in Figure S2.

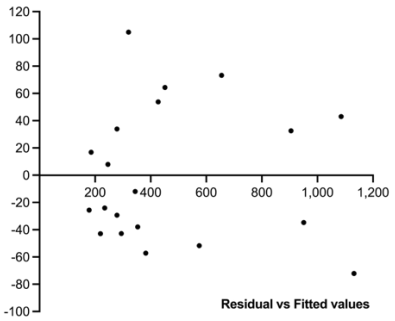


Figure S2. *Residuals vs. fitted values plot.*

Multicollinearity

Table S6 to Table S11 presents the linear regression statistics, the results of the analysis of variance (ANOVA) and the coefficients from the paired comparison of the independent variables: Useful floor area and Number of Residents. Firstly, Table S6 to Table S8 present the results of the analysis of the variable Useful floor area in relation to the variable Number of Residents.

Table S6. *Regression statistics.*

Regression Statistics	
Multiple R	0.88990318
R^2	0.79192767
Adjusted R^2	0.780368096
Standard Error	1534.772575
Observations	20

***Table S7**. Analysis of Variance (ANOVA).*

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	161373327.1	161373327,1	68.5083792	1.50423E-07
Residual	18	42399483.43	2355526,857		
Total	19	203772810.6			

***Table S8**. Coefficients (Significance of Individual Predictors).*

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	253.50	596.6582739	0.424865546	0.037597012	-1000.032975	1507.032062	-1000.032975	1507.032062
Nº Residents	42.39	5.121539682	8.276978869	1.50423E-07	31.63092013	53.15083132	31.63092013	53.15083132

Secondly, Table S9 to Table S11 present the results of the analysis of the variable Number of Residents on the variable Useful floor area.

***Table S9.** Regression statistics.*

Regression Statistics	
Multiple R	0.88990318
R^2	0.79192767
Adjusted R^2	0.780368096
Standard Error	32.21917387
Observations	20

***Table S10**. Analysis of Variance (ANOVA).*

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	71116.84703	71116.84703	68.5083792	1.50423E-07
Residual	18	18685.35297	1038.075165		
Total	19	89802.2			

***Table S11**. Coefficients (Significance of Individual Predictors).*

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	15.09	12.07500831	1.249980606	0.022731464	-10.27512489	40.46217729	-10.27512489	40.46217729
Useful floor area (m^2)	0.02	0.002257051	8.276978869	1.50423E-07	0.013939672	0.023423447	0.013939672	0.023423447

Finally, based on Equation 13 in the manuscript and the R^2 values in Table S6 and Table S9 [1,2], Table S12 presents the Variance Inflation Factor (VIF) values for both scenarios of pairwise comparisons of the independent variables Useful floor area and Number of Residents.

***Table S12**. VIF values for the independent variables.*

	Useful floor area (m2)	*N Residents*
R square	0,79192767	0,79192767
VIF	4,806021066	4,806021066

Regression diagnostics for the Energy Costs model

1. Results of the multiple regression analysis

Table S13 to Table S17 present the multiple regression statistics, analysis of variance (ANOVA) results, model coefficients, residual outputs and probability outputs for the Energy Costs prediction model.

Table S13. *Energy Costs model regression statistics.*

Regression Statistics	
Multiple R	0.988503341
R^2	0.977138855
Adjusted R^2	0.974449309
Standard Error	7338.418432
Observations	20

Table S14*. Energy Costs model Analysis of Variance (ANOVA).*

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	39130209765	19565104882	363.3099045	1.12806E-14
Residual	17	915490546.4	53852385.08		
Total	19	40045700311			

Table S15*. Energy Costs model Coefficients (Significance of Individual Predictors).*

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	10,950	2867.153177	3.81923291	0.001372334	4901.161334	16999.49021	4901.161334	16999.49021
Useful floor area (m2)	5.37	1.126995517	4.766042397	0.000179279	2.993555719	7.749061116	2.993555719	7.749061116
Nº Residents	422.01	53.68485918	7.860876911	4.63086E-07	308.7449178	535.2752222	308.7449178	535.2752222

***Table S16**. Energy Costs model Residual output.*

Observation	*Predicted values (€, total adjusted)*	*Residuals*	*Standard Residuals*
1	40107.7919	782.5656016	0.112738077
2	144657.0292	-13872.73372	-1.998535738
3	171516.5983	-4007.380933	-0.577311882
4	66967.36102	8815.225758	1.269940307
5	31543.32569	2319.700218	0.334180983
6	163684.2076	4646.742963	0.669419746
7	50938.22554	13278.80141	1.912972577
8	38579.44143	-1327.506758	-0.191243467
9	56470.82235	-4603.429408	-0.663179902
10	60940.47567	-8634.261873	-1.243870262
11	138593.4185	9785.733343	1.409753709
12	47122.42241	-5418.572018	-0.780611093
13	30596.52597	-4146.148174	-0.597302988
14	70472.35298	5458.93363	0.786425673
15	54687.39882	87.70705505	0.012635266
16	45043.15048	6139.707623	0.884499433
17	102337.1287	5805.731119	0.836386062
18	89330.71856	-4909.32179	-0.707247414
19	44691.54299	-3076.508777	-0.443208445
20	36295.91064	-7124.985273	-1.026440642

***Table S17**. Energy Costs model Probability output.*

Percentile	*Energy Costs values (€, total adjusted)*
2.5	26450.37779
7.5	29170.92536
12.5	33863.0259
17.5	37251.93468
22.5	40890.3575
27.5	41615.03422
32.5	41703.85039
37.5	51182.8581
42.5	51867.39295
47.5	52306.2138
52.5	54775.10587
57.5	64217.02696
62.5	75782.58678
67.5	75931.28661
72.5	84421.39676
77.5	108142.8598
82.5	130784.2955
87.5	148379.1519
92.5	167509.2174
97.5	168330.9506

2. Assessment of regression assumptions

Linearity

The simple regression plots for the pairwise relationships between the variables are shown in Figure S3.

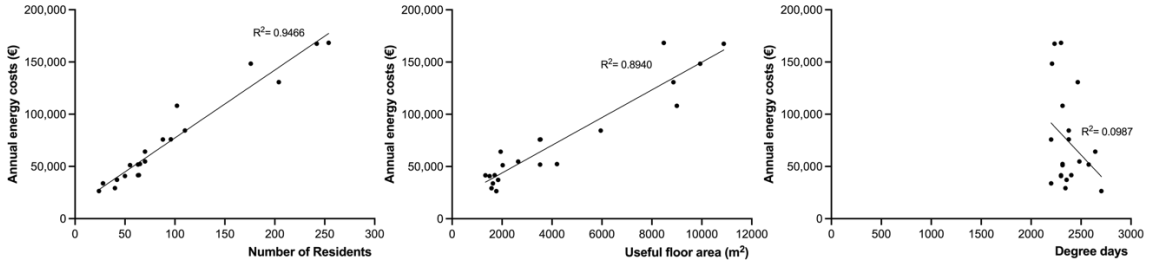


Figure S3. *Pairwise simple regression plots.*

Homoscedasticity

The assumption of homoscedasticity was evaluated using residuals versus fitted values plots (see Table S16). The corresponding plots are presented in Figure S4.

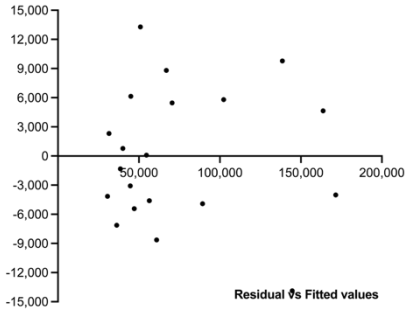


Figure S4. *Residuals vs. fitted values plot.*

Multicollinearity

The multicollinearity analysis is identical for the three predictive models, as the same independent variables (Useful floor area and Number of Residents) were used. Therefore, the VIF values obtained for the Energy Consumption model, reported in Table S12, are adopted.

Regression diagnostics for the CO_2 Emissions model

1. Results of the multiple regression analysis

Table S18 to Table S22 present the multiple regression statistics, analysis of variance (ANOVA) results, model coefficients, residual outputs and probability outputs for the CO_2 Emissions prediction model.

Table S18. *CO_2 Emissions model regression statistics.*

Regression Statistics	
Multiple R	0.988377118
R^2	0.976889328
Adjusted R^2	0.974170425
Standard Error	26.37814681
Observations	20

Table S19*. CO_2 Emissions model Analysis of Variance (ANOVA).*

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	500000.3032	250000.1516	359.2954439	1.23711E-14
Residual	17	11828.71269	695.8066289		
Total	19	511829.0159			

Table S20*. CO_2 Emissions model Coefficients (Significance of Individual Predictors).*

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	13.87	10.30606092	1.345785212	0.001960467	-7.8741435	35.61363226	-7.8741435	35.61363226
Useful floor area (m2)	0.02	0.004051016	4.363492032	0.000423082	0.00912968	0.026223475	0.00912968	0.026223475
Nº Residents	1.58	0.192971702	8.181622183	2.68509E-07	1.171686856	1.985956264	1.171686856	1.985956264

Table S21. *CO_2 Emissions model Residual output.*

Observation	*Predicted values (tCO_2, total adjusted)*	*Residuals*	*Standard Residuals*
1	119.3256889	5.369765867	0.215210476
2	492.6522035	-40.49586103	-1.623000659
3	588.2657268	38.72545762	1.552046102
4	214.9392121	4.436567632	0.177809584
5	86.96027595	15.6874773	0.628725637
6	564.7701226	-26.34780531	-1.055972248
7	158.7681971	6.52706063	0.261592751
8	112.7758591	4.520570394	0.181176262
9	175.6454389	-0.4986875	-0.01998649
10	190.7347719	-28.65831068	-1.148573112
11	467.3768145	63.55528746	2.547180646
12	145.0705657	-11.75296368	-0.471037467
13	83.08435742	3.393958235	0.13602369
14	227.9940225	-13.57430571	-0.544033552
15	171.1064483	19.90757938	0.797859673
16	136.5176765	17.24982924	0.691341868
17	333.9987424	-35.1730784	-1.409673186
18	292.715753	-20.61923524	-0.82638155
19	136.8807041	5.347806203	0.214330373
20	104.8808932	-7.601112405	-0.304638798

Table S22. *CO_2 Emissions model Probability output.*

Percentile	*CO_2 Emissions values (tCO_2, total adjusted)*
2.5	86.47831566
7.5	97.27978076
12.5	102.6477533
17.5	117.2964295
22.5	124.6954547
27.5	133.317602
32.5	142.2285103
37.5	153.7675058
42.5	162.0764613
47.5	165.2952578
52.5	175.1467514
57.5	191.0140277
62.5	214.4197168
67.5	219.3757798
72.5	272.0965178
77.5	298.825664
82.5	452.1563425
87.5	530.932102
92.5	538.4223172
97.5	626.9911844

2. Assessment of regression assumptions

Linearity

The simple regression plots for the pairwise relationships between the variables are shown in Figure S5.

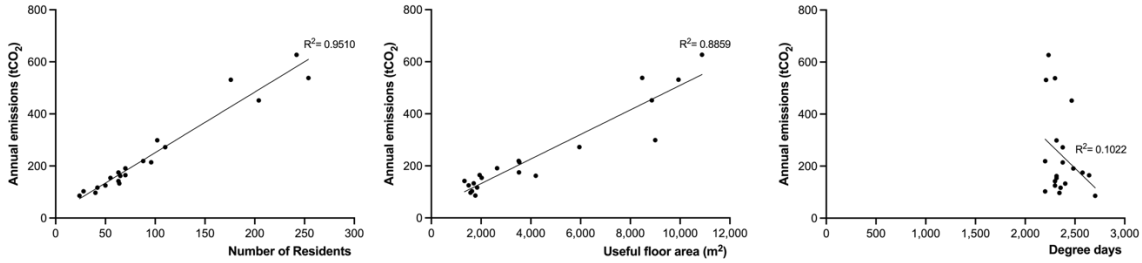


Figure S5. *Pairwise simple regression plots.*

Homoscedasticity

The assumption of homoscedasticity was evaluated using residuals versus fitted values plots (see Table S21). The corresponding plots are presented in Figure S6.

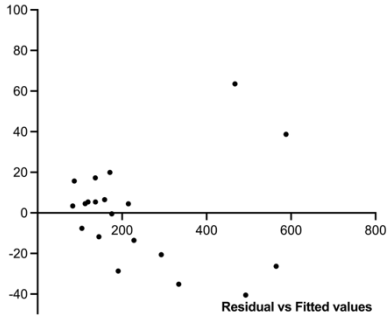


Figure S6. *Residuals vs. fitted values plot.*

Multicollinearity

The multicollinearity analysis is identical for the three predictive models, as the same independent variables (Useful floor area and Number of Residents) were used. Therefore, the VIF values obtained for the Energy Consumption model, reported in Table S12, are adopted.

References

[1] Miles J. Tolerance and Variance Inflation Factor. Wiley StatsRef Stat. Ref. Online, Wiley; 2014. https://doi.org/10.1002/9781118445112.stat06593.

[2] Chan JY-L, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong Z-W, et al. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. Mathematics 2022;10:1283. https://doi.org/10.3390/math10081283.